

**PROPUESTA DE UN MODELO DE MUESTREO ESTRATIFICADO DOBLE  
APOYADO EN EL CONTRASTE DE LOS MODELOS DE DISEÑO DE  
EXPERIMENTOS CLÁSICOS Y DE MUESTREO ALEATORIO SIMPLE Y  
ESTRATIFICADO ALEATORIO**

Autor: Ing. Mairett Yuri Rodríguez Balza  
Tutor: Prof. Oscar Haddad

Maracay, Abril de 2006

UNIVERSIDAD CENTRAL DE VENEZUELA  
FACULTAD DE AGRONOMÍA  
COMISIÓN DE ESTUDIOS DE POSTGRADO  
POSTGRADO DE ESTADÍSTICA  
MAESTRÍA EN ESTADÍSTICA

**PROPUESTA DE UN MODELO DE MUESTREO ESTRATIFICADO  
DOBLE APOYADO EN EL CONTRASTE DE LOS MODELOS DE  
DISEÑO DE EXPERIMENTOS CLÁSICOS Y DE MUESTREO  
ALEATORIO SIMPLE Y ESTRATIFICADO ALEATORIO**

Ing. Mairett Yuri Rodríguez Balza  
Tutor: Prof. Oscar Haddad

Maracay, Abril de 2006

UNIVERSIDAD CENTRAL DE VENEZUELA  
FACULTAD DE AGRONOMÍA  
COMISIÓN DE ESTUDIOS DE POSTGRADO  
POSTGRADO DE ESTADÍSTICA  
MAESTRÍA EN ESTADÍSTICA

**PROPUESTA DE UN MODELO DE MUESTREO ESTRATIFICADO  
DOBLE APOYADO EN EL CONTRASTE DE LOS MODELOS DE  
DISEÑO DE EXPERIMENTOS CLÁSICOS Y DE MUESTREO  
ALEATORIO SIMPLE Y ESTRATIFICADO ALEATORIO**

Ing. Mairett Yuri Rodríguez Balza  
Tutor: Prof. Oscar Haddad  
Prof. Isabel Díaz (Miembro Comité Consejero)  
Prof. Marisela Ascanio (Miembro Comité Consejero)

Maracay, Abril de 2006

## VEREDICTO

Quienes suscriben, Miembros del Jurado designado por el Consejo de la Facultad de Agronomía de la Universidad Central de Venezuela, para examinar el Trabajo de Grado presentado por la ciudadana MAIRETT Y. RODRÍGUEZ B., C.I. No. 13.542.025, titulado **"PROPUESTA DE UN MODELO DE MUESTREO ESTRATIFICADO DOBLE APOYADO EN EL CONTRASTE DE LOS MODELOS DE DISEÑO DE EXPERIMENTOS CLASICOS Y DE MUESTREO ALEATORIO SIMPLE Y ESTRATIFICADO ALEATORIO"**, a los fines de cumplir con el requisito académico exigido para optar al Título de *Magister Scientiarum en Estadística*, dejan constancia de lo siguiente:

1. Leído como fue dicho Trabajo de Grado, por cada uno de los Miembros del Jurado, éste fijó el día viernes 21 de abril de 2006, a las 9:00 a.m., para que la autora lo defendiera en forma pública, lo que ésta hizo, en el Salón Principal del Postgrado en Estadística, mediante un resumen oral de su contenido, luego de lo cual respondió satisfactoriamente a las preguntas que le fueron formuladas por el Jurado, todo ello conforme a lo dispuesto en los artículos 45, 47, 50, 51, 52 y 53 del Reglamento de Estudios de Postgrado de la Universidad Central de Venezuela y a las Normas para la Elaboración de Tesis de Grado de la Facultad de Agronomía de la misma Universidad.
2. Finalizada la defensa pública del Trabajo de Grado, el Jurado decidió aprobarlo, por considerar sin hacerse solidario de las ideas expuestas por el autor, que se ajusta a lo dispuesto y exigido por el Reglamento de Estudios de Postgrado. Para dar este veredicto el jurado estimó que dicho trabajo representa un importante aporte a la experimentación agrícola con la propuesta de un modelo de muestreo estratificado doble, así mismo es un aporte teórico en el campo del muestreo y del diseño de experimento.

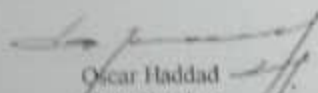
Página 2 del Veredicto del Trabajo de Grado de la ciudadana MAIRETT Y. RODRIGUEZ B.

En fe de lo cual se levanta la presente acta, a los veintidós días del mes de abril de dos mil seis, dejándose también constancia de que, conforme a lo dispuesto en la normativa jurídica vigente actuó como Coordinador del Jurado el Tutor, Prof. Oscar Haddad.

Firma del Jurado Examinador:

  
Dra. Margarita Coto  
C.I. No. 282.384

4  
MSc. Isabel Diaz  
C.I. No. 3.752.495

  
Oscar Haddad  
Coordinador  
C.I. No. 1.846.617

## **AGRADECIMIENTOS**

A Dios Todo Poderoso, por su bendición y guía en mi vida. Además por tener tantas personas, a las cuales agradecerles todo lo que han hecho por mí...

A mi Ángel de la Guarda que nunca me abandona ni a mis seres queridos.

A mi hija por quererme tanto, por estar a mi lado, por su gran apoyo, comprensión y palabras de aliento, que su corta edad (2 años) no le impide hacerlo en el momento más oportuno.

A mi madre por su amor, apoyo y comprensión. Y por creer siempre en mí. Por estar siempre allí cuando la necesito.

A mis hermanas Airin y Airimar, por siempre apoyarme y ayudarme cuando las necesito.

A mi "Abue" Francisca, por su amor y confianza. Por ser otra madre para mí.

A Javier Rodríguez por todos los maravillosos momentos que compartimos en estos 10 años, por estar a mi lado, y especialmente por darme una hija tan hermosa e inteligente. Gracias por creer en mí...

A mis amigos por estar a mi lado apoyándome y colaborando en lo que ha estado a sus alcances y por brindarme su amistad, en especial a: Brenda Zerpa, Yohnny Torres, José Angel Reyes, Luis Pérez, Hadid Fernández y Yolimar Goatache.

A mi familia y a la de Javier por siempre apoyarme y creer en mí...

Al CDCH por la beca suministrada, ya que me incentivó a culminar el trabajo de grado.

A la Universidad Central de Venezuela, Facultad de Agronomía, por convertirse en mi segundo hogar durante más de 10 años y de la cual me siento orgullosa.

A la Coordinación de Investigación de la Facultad de Agronomía, por su colaboración en la reproducción de este trabajo, especialmente al Prof. Leonardo Taylhardat, Mercedes Ascanio y María Carlina Pino, que además me apoyaron en otros aspectos relacionados con la beca.

A los empleados y obreros de la Facultad de Agronomía, especialmente a: Debbie Bello, Miriam Morán, Deyanira Altuve, Grettys Kwiecien, Jackeline González, Karina Rivas, Ericka Prieto, Laura Navarro, Noemí Hortelano, Luis Santana, Yesenia Lozano, entre otros, por su granito de arena aportados para la realización de este trabajo.

Muy especialmente al Prof. Oscar Haddad, por su invaluable labor como tutor, guiándome, apoyando, animando, incentivando y alentándome en los momentos más difíciles de la realización de este trabajo, sin él no sé si hubiese podido hacerlo en tan corto tiempo. Además por compartir sin egoísmo sus conocimientos y experiencias, desde el punto de vista especialmente del muestreo. Y por su humildad y cariño...

A la Profa. Isabel Díaz, por su valiosa colaboración como asesora y comportándose además como co-tutora. Por ofrecerme su amistad sincera, apoyándome en lo personal y en lo concerniente al Trabajo de Grado. Por su tiempo y dedicación en la elaboración del Modelo propuesto. Y por ser tan sencilla, humilde y cariñosa.

A la Profa. Marisela Ascanio, por su valioso asesoramiento para encaminar esta investigación. Además, por su asesoría en gran parte de mi carrera de Ingeniero y en la Maestría. Y por ser siempre tan receptiva.

Al Prof. Alberto Camardiel, por su grandiosa colaboración en el inicio de la investigación y en parte del Modelo propuesto. Por suministrar el primer artículo de Brakel, J and Renssen, 1998, el cual presenta en gran medida los términos de la comparación teórica, especialmente a nivel de principios del Diseño de Experimentos y del Muestreo. Además, por mostrar interés y ser tan receptivo y dispuesto a ayudar.



Al Prof. Félix Seijas, por su disposición a colaborar y por aclararme puntos importantes de la investigación.

Al Prof. Lino Lugo, por orientarme en muchas oportunidades en la investigación y en el título de la tesis. Además por ser siempre tan atento y colaborador...

Al Prof. Miguel Balza, por sus valiosas correcciones y por ser siempre tan receptivo...

A la Profa. Judith García, por sus valiosas correcciones; además por la traducción del resumen.

A la Profa. Margarita Cobo, por sus asesorías tan desinteresadas y sus valiosas correcciones. Además por ser un valuarte y ejemplo a seguir.

A la Profa. Harú Martínez por su comprensión, inestimables consejos y aportes en lo relacionado al Trabajo de Grado y a mi desempeño como docente...

A la Profa. Isa de Benítez, por sus consejos, su apoyo, por creer en mí y por su amistad sincera...

A la Profa. Isis Vivas, por su amistad, por estar presente en toda mi carrera, y por estar siempre dispuesta a ayudar.

A todos mis compañeros de trabajo de la Cátedra de Matemáticas por apoyarme: José Mosquera, Nelson Briceño, Omar Núñez, Rafael Crespo, Nicolás Crespo, Ruth Martínez, Milagros Borregos y Raiza Spirrow.

## **DEDICATORIA**

A Dios...

A mi hija que la adoro, es la bendición más grande que he tenido y estoy orgullosísima de ser su madre. Espero que sea mucho mejor que yo y que alcance sus sueños; además le pido a Dios estar mucho tiempo a su lado para compartirlos...Te amo hija.

A mi madre, quiero ser tan buena madre como ella, sacrificada por sus hijas y sé que mis hermanas también lo reconocerán y valorarán. Este otro logro, es suyo también. Que Dios le dé mucha salud...Te amo mami.

A mi "Abue" Francisca, que Dios le dé mucha salud. Este es también su logro por ayudar a mi madre a formar mi carácter, mis ansias de salir adelante y porque siempre me enseñó a valorar lo poco o mucho que se tenga...Te amo "abue".

A mis hermanas, que son como mis hijas, espero que me sigan viendo como ejemplo a seguir. Quiero que siempre cuenten conmigo...Las amo

A Javier, espero que Dios nos de fuerzas para afrontar las tempestades que aún se avecinan. Me encantaría envejecer a su lado...Te amo.

A mi familia que aún permanece bastante unida y le pido a Dios que mantenga esa unión.

A mi bisabuelita, que siempre me quiso mucho y me recordaba con tanto cariño y ternura. Se que desde el cielo está feliz por mi logro. Te llevo en mi corazón. Que en paz descanses.

A mis abuelos en el cielo, a mi prima Geovana, a Newman y a Rafael Cedeño (otro padre), que se que me cuidan y ayudan. Que en paz descansen.

## RESUMEN

Se propone un Modelo de Muestreo Estratificado Doble apoyado en el contraste de los modelos de diseño de experimentos clásicos y diseños de muestreo aleatorio simple y estratificado aleatorio. El Modelo Mixto propuesto es el siguiente:  $y_{ijk} = \mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk}$   $i = 1, 2, \dots, a$   $j = 1, 2, \dots, b$   $k = 1, 2, \dots, N_{ij}$

Donde:  $y_{ijk}$  es la k-ésima observación en el i-ésimo estrato T y el j-ésimo estrato B,  $\mu$  es un efecto común a todos los estratos que corresponden a las unidades experimentales en un Diseño de Bloques al Azar,  $\alpha_i$  es el efecto del i-ésimo nivel del estrato T (tratamientos),  $\beta_j$  es el efecto del j-ésimo nivel del estrato B (bloques),  $\rho_{ij}$  es el efecto de la interacción del i-ésimo nivel del estrato T y el j-ésimo nivel del estrato B,  $\varepsilon_{ijk}$  es el efecto aleatorio del i-ésimo nivel del estrato T y el j-ésimo nivel del estrato B, en la k-ésima observación,  $N_{ij}$  es el número de observaciones en el estrato formado por el i-ésimo nivel del estrato T y el j-ésimo nivel del estrato B, N es el número total de observaciones  $N = \sum_{i=1}^a \sum_{j=1}^b N_{ij}$ . Este Modelo está basado en la premisa de que se

puede aplicar un Modelo de Muestreo para simular un Diseño de Experimentos en Bloques al Azar, o bien para caracterizar una población bajo estudio. En presencia de datos balanceados, la Esperanza de los Cuadrados Medios (ECM) coincide con la de Modelo Estándar presentado por Montgomery (1991), en el cual se consideran los supuestos

$\sum_{i=1}^a \alpha_i = 0$ ,  $\sum_{i=1}^a \rho_{ij} = 0$ ,  $V(\rho_{ij}) = \frac{(a-1)}{a} \sigma_p^2$ ,  $Cov(\rho_{ij}; \rho_{i'j}) = -\frac{\sigma_p^2}{a}$   $i \neq i'$ , el resto de las

covarianzas se asumen cero. En el caso de datos desbalanceados, en el Modelo propuesto, la ECM es afectada por los efectos aleatorios del estrato B y el efecto interacción del estrato T y Estrato B. Este Modelo fue ejemplificado con datos reales, tomados de un experimento en Bloques al Azar, realizado en Musáceas, en la Estación experimental Samán Mocho, Edo. Carabobo, donde se evaluaron las variables: Tiempo de Desarrollo de la Enfermedad (Sigatoka Negra) en días y Número de Frutos para 6 variedades analizadas. Con esta información se contrastaron los Modelos del Diseño Completamente al Azar (DCA) vs Muestreo Estratificado Aleatorio (MEA); DCA vs Muestreo Aleatorio Simple (MAS); y Bloques al Azar (DBA) vs Muestreo Estratificado Doble (MED). Los estimadores en general coincidieron y las estimaciones fueron similares, a excepción del contraste del DCA vs MEA, donde se apreció mayores diferencias por efectos del pequeño tamaño de muestra. Tal como se señaló anteriormente, en el caso del Modelo propuesto las estimaciones coinciden con las del DBA, en tal sentido, este Modelo puede aplicarse en la práctica en ocasiones donde el DBA no pueda realizarse por razones físicas o por ser muy onerosa su realización. Se emplean las raíces comunes del Diseño de Experimentos y el Muestreo, para realizar

contrastes entre estas dos Teorías, en cuanto a la aleatorización, repetición vs tamaño de muestra y control local.

***Palabras Claves: Modelos de Muestreo, Modelos de Diseño, Muestreo Estratificado Doble, Muestreo Aleatorio Simple, Muestreo Estratificado Aleatorio, Diseño Completamente al Azar, Diseño de Bloques al Azar, Esperanza de los Cuadrados Medios, Principios Básicos del Diseño de Experimentos y Criterios de Muestreo.***

## ABSTRACT

A Model of Double Stratified Sampling supported on the contrast of the models of classic experimental design of and designs of aleatory simple and stratified sampling. The proposed Mixed Pattern is the following:  $y_{ijk} = \mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk}$   $i = 1, 2, \dots, a$   $j = 1, 2, \dots, b$   $k = 1, 2, \dots, N_{ij}$ . In which:  $y_{ijk}$  it is the  $k$ -th observation in the  $i$ -th stratum T and the  $j$ -th stratum B,  $\mu$  it is a common effect to all the strata that correspond to the experimental units in a Block Design,  $\alpha_i$  it is the effect of the  $i$ -ésimo level of the stratum T (treatments),  $\beta_j$  it is the effect of the  $j$ -ésimo level of the stratum B (blocks),  $\rho_{ij}$  it is the effect of the interaction of the  $i$ -th level of the stratum T and the  $j$ -th level of the stratum B,  $\varepsilon_{ijk}$  it is the aleatory effect of the  $i$ -ésimo level of the stratum T and the  $j$ -ésimo level of the stratum B, in the  $k$ -ésima observation,  $N_{ij}$  it is the number of observations in the stratum formed by the  $i$ -th level of the stratum T and the  $j$ -th level of the stratum B,  $N = \sum_{i=1}^a \sum_{j=1}^b N_{ij}$  is the total number of observations. This

Model is based on the premise that can be applied a Model of Sampling to simulate a Design of Experiments such as Blocks at random, or to characterize a population under study. In presence of balanced data, the Expected Square Means (ECM) coincides with that of Standard Model presented by Montgomery (1991), in which are considered the suppositions

$$\sum_{i=1}^a \alpha_i = 0, \sum_{i=1}^a \rho_{ij} = 0, \forall (p_{ij}) = \frac{(a-1)}{a} \sigma_p^2, \text{Cov}(p_{ij}; p_{i'j}) = -\frac{\sigma_p^2}{a} \text{ if } i \neq i', \text{ the rest of the}$$

covariances are assumed zero. In the case of data unbalanced, in the proposed model, the ECM is affected by the random effects of the stratum B and the effect interaction of the stratum T and Stratum B. This Model was exemplified with real data, taken of an experiment in Blocks at random, carried out in Musáceas, in the experimental Station Samán Mocho, Edo. Carabobo, where the following variables were evaluated: Time of Development of the Illness (Sigatoka) in days and Number of Fruits for 6 varieties. With this information the Design Models were contrasted. They were: Totally at random (DCA) vs Aleatory Stratified Sampling (MEA); DCA vs Simple Aleatory Sampling (MAS); and Blocks at random (DBA) vs Sampling Stratified Double (MED). The estimators in general coincided and the estimates were similar, to exception of the contrast of the DCA vs MEA, where it was appreciated bigger differences by effects of the small sample size. Just as it was pointed out previously, in the case of the proposed model the estimates coincide with those of the DBA, for that, this Model can be applied in practice in occasions where the DBA cannot be carried out for physical reasons or high costs. The common roots of the Design of Experiments and Sampling are used, to carry out contrasts among these two Theories, as for the aleatorización, repetition vs sample size and local control.

**Key words:** *Sampling Models, Design Models, Double Stratified Sampling, Simple Aleatory Sampling, Aleatory Stratified Sampling, I Design Totally at random, Design of Blocks at random, Expected of the Square Means, Basic Principles of Design of Experiments and Sampling Criteria.*



# INDICE DE CONTENIDO

	<b><u>Página</u></b>
<b>AGRADECIMIENTOS</b>	i
<b>DEDICATORIA</b>	vi
<b>RESUMEN</b>	viii
<b>ABSTRACT</b>	x
<b>INTRODUCCIÓN</b>	1
<b>OBJETIVO GENERAL</b>	4
<b>OBJETIVOS ESPECÍFICOS</b>	4
<b>METODOLOGÍA</b>	6
<b>CAPITULO I</b>	
<b>ASPECTOS TEÓRICOS PRELIMINARES</b>	
1. MODELO	14
Modelo Lineal General	14
2. ASPECTOS TEÓRICOS DE LAS FUNCIONES ESTIMABLES	16
3. GENERALIDADES SOBRE MODELOS	19
<b>CAPITULO II</b>	
<b>ASPECTOS TEÓRICOS DEL MUESTREO</b>	
1. GENERALIDADES SOBRE MUESTREO	26
Ventajas del Muestreo en Comparación con un Censo	27
2. MUESTREO ALEATORIO SIMPLE	28
Notación	30
Teoría de Aleatorización	31
Tamaño de Muestra para el Muestreo Aleatorio Simple	32
Modelo para el Muestreo Aleatorio Simple	35
3. MUESTREO ESTRATIFICADO ALEATORIO	40
Clasificación de las unidades en la muestra	42
Notación	43

## **Página**

Criterios para Formar los Estratos	45
Número de Estratos	46
Tamaño de la Muestra y Afijación de la Muestra a los Estratos	48
Modelo para el Muestreo Estratificado	51

### **CAPITULO III**

#### **ASPECTOS TEÓRICOS DEL DISEÑO DE EXPERIMENTOS**

1. DISEÑO COMPLETAMENTE AL AZAR (D.C.A.)	53
Ventajas	53
Desventajas	54
Modelo Lineal para el Diseño Completamente al Azar	54
Intervalos de Confianza	68
2. DISEÑO DE BLOQUES AL AZAR (D.B.A.)	
Ventajas	70
Desventajas	70
Modelo Lineal para el Diseño de Bloques al Azar	70

### **CAPITULO IV**

#### **MODELO PARA EL MUESTREO ESTRATIFICADO DOBLE**

1. PLANTEAMIENTO DEL MODELO	83
2. ESTIMACIÓN DE LOS PARÁMETROS UTILIZANDO EL MÉTODO DE LOS MÍNIMOS CUADRADOS	85
3. PRUEBAS DE HIPÓTESIS	87
4. ANÁLISIS DE VARIANZA	88
5. ESPERANZA DE LOS CUADRADOS MEDIOS	90
6. ESTIMACIÓN DE LOS PARÁMETROS A TRAVÉS DEL MUESTREO	113

### **CAPITULO V**

#### **CONTRASTE ENTRE EL DISEÑO/ANÁLISIS DE EXPERIMENTOS Y LA TEORÍA DEL MUESTREO**

1. PRINCIPIOS BÁSICOS DEL DISEÑO DE EXPERIMENTOS Y LOS CRITERIOS DEL DISEÑO DE INVESTIGACIÓN	123
--	-----

	<b><u>Página</u></b>
Principios básicos del diseño de experimentos	126
Criterios del diseño de Investigación	136
2. OTROS CONTRASTES DE INTERÉS ENTRE EL DISEÑO DE EXPERIMENTOS Y EL MUESTREO	141
Objetivos	141
Validez externa e interna	143
Modelos Estadísticos	144
Ventajas del Diseño de experimentos sobre el muestreo de encuestas	144
<b>RESULTADOS</b>	147
<b>DISCUSIÓN DE RESULTADOS</b>	195
<b>CONCLUSIONES</b>	222
<b>RECOMENDACIONES</b>	226
<b>REFERENCIAS BIBLIOGRÁFICAS</b>	227
<b>APÉNDICE</b>	231
A. CARACTERIZACIÓN DEL ENSAYO EN MUSÁCEAS	231
B. Diseño Completamente al Azar	235
C. Muestreo Aleatorio Simple	242
D. Muestreo Estratificado Aleatorio	249
E. Muestreo Estratificado Doble	252

## INDICE DE CUADROS

	<b><u>Página</u></b>
1. Análisis de Varianza para el Diseño Completamente al Azar (D.C.A.)	64
2. Esperanza de los Cuadrados Medios para el Modelo de Efectos Fijos en el Diseño Completamente al Azar	65
3. Esperanza de los Cuadrados Medios para el Modelo de Efectos Aleatorios en el Diseño Completamente al Azar	68
4. Análisis de Varianza para el Diseño de Bloques al Azar (D.B.A.)	76
5. Esperanza de los Cuadrados Medios para el Modelo de Efectos Fijos en el Diseño de Bloques al Azar	68
6. Esperanza de los Cuadrados Medios para el Modelo de Efectos Aleatorios en el Diseño de Bloques al Azar	81
7. Esperanza de los Cuadrados Medios para el Modelo Mixto Estratificado Doble	112
8. Esperanza de los Cuadrados Medios para el Modelo Estándar y Modelo Alterno	113
9. Coincidencias entre el diseño de experimentos y la teoría del muestreo	122
10. Diferencias entre el diseño y análisis de experimentos y la teoría del muestreo	122
11. Esperanza de los Cuadrados Medios para el Modelo Mixto Estratificado Doble	149
12. Estadística descriptiva del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, para cada variedad, en el diseño completamente al azar	151
13. Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (TDE), en Sigatoka Negra, bajo el Diseño Completamente al Azar	152
14. Análisis de Varianza de variedades susceptibles, para el Tiempo de Desarrollo de la Enfermedad (TDE), en Sigatoka Negra, bajo el Diseño Completamente al Azar	154
15. Estadística descriptiva del Número de Frutos (dedos), en Sigatoka Negra, para cada variedad, en el diseño completamente al azar	155

16. Análisis de Varianza, para el Número de frutos (dedos), bajo el Diseño Completamente al Azar	155
17. Estadística descriptiva del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, para las variedades dentro de cada bloque, bajo el diseño de bloques al azar	157
18. Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (días), bajo el Diseño de Bloques al Azar	159
19. Análisis de Varianza, de variedades susceptibles, para el Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, bajo el Diseño de Bloques al Azar	160
20. Estadística descriptiva del Número de frutos (dedos), para las variedades, en cada bloque, bajo el diseño de Bloques al Azar	161
21. Análisis de Varianza, para el Número de frutos (dedos), bajo el Diseño de Bloques al Azar.	162
22. Tamaños de muestra, para cada una de las variedades, en la variable Tiempo de Desarrollo de la Enfermedad (días), con un error de 4 días	164
23. Estimaciones de la media y varianza del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, por variedad, para el modelo del muestreo aleatorio simple	165
24. Tamaños de muestra, para cada una de las variedades, en la variable Número de frutos (dedos) con un error de 6 dedos	167
25. Estimaciones de la media y varianza del Número de Frutos (dedos), por variedad, para el modelo del muestreo aleatorio simple	168
26. Estimaciones de la media y varianza del Número de Frutos (dedos), por variedad, para el modelo del muestreo aleatorio simple, considerando $n=25$	169
27. Estimaciones de la media y varianza del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, por variedad, para el modelo del muestreo estratificado aleatorio	171
28. Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad, bajo el Modelo de Muestreo Estratificado Aleatorio	172
29. Afijación de la muestra en el muestreo estratificado aleatorio en cada estrato tipo T (variedades), para la variable Número de Frutos	173

30. Estimaciones de la media y varianza del Número de Frutos (dedos) en Sigatoka Negra, por variedad, para el modelo del muestreo estratificado aleatorio	174
31. Análisis de Varianza, para el Número de Frutos (dedos), bajo el Modelo de Muestreo Estratificado Aleatorio	175
32. Afijación de la muestra en el muestreo estratificado aleatorio en cada estrato tipo T (variedades), para la variable Número de Frutos, considerando la mayor varianza.	177
33. Estimaciones de la media y varianza del Número de Frutos (dedos), por variedad, para el modelo del muestreo estratificado aleatorio, considerando $n=38$	177
34. Análisis de Varianza, para el Número de Frutos (dedos), bajo el Modelo de Muestreo Estratificado Aleatorio con $n=38$	178
35. Estadística descriptiva del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, para cada variedad, en cada Estrato B, para el modelo del muestreo estratificado doble	180
36. Estimaciones de la media y varianza del Tiempo de Desarrollo de la Enfermedad (días) por variedad, para el modelo del muestreo estratificado doble	181
37. Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (días), bajo el Modelo de Muestreo Estratificado Doble	182
38. Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (días), bajo el Modelo de Muestreo Estratificado Doble sin la interacción	182
39. Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (días) en Sigatoka Negra para las variedades susceptibles, bajo el Modelo de Muestreo Estratificado Doble	183
40. Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (días) para las variedades susceptibles, bajo el Modelo de Muestreo Estratificado Doble sin la interacción	184
41. Afijación de la muestra en el muestreo estratificado doble en cada estrato tipo T (variedades), para la variable Número de Frutos	185
42. Estimaciones de la media y varianza del Número de frutos (dedos) por variedad, para el modelo del muestreo estratificado doble	186

43. Estadística descriptiva del Número de frutos (dedos), para las variedades, en cada bloque, bajo el modelo del muestreo estratificado doble	187
44. Análisis de Varianza, para el Número de frutos (dedos), bajo el Modelo de Muestreo Estratificado Doble	188
45. Análisis de Varianza, para el Número de frutos (dedos), bajo el Modelo de Muestreo Estratificado Doble, sin la interacción	189
46. Afijación de la muestra en el muestreo estratificado doble en cada estrato tipo T (variedades), para la variable Número de Frutos, considerando la mayor varianza	190
47. Estimaciones de la media y varianza del Número de frutos (dedos) por variedad, para el modelo del muestreo estratificado doble con n mayor	191
48. Estadística descriptiva del Número de frutos (dedos), para las variedades, en cada bloque, bajo el modelo del muestreo estratificado doble con n mayor	192
49. Análisis de Varianza, para el Número de frutos (dedos), bajo el Modelo de Muestreo Estratificado Doble con n mayor	193
50. Análisis de Varianza, para el Número de Frutos (días), bajo el Modelo de Muestreo Estratificado Doble, con la muestra mayor	194
51. Resumen de resultados de los Modelos del Diseño Completamente al Azar (D.C.A), Muestreo Aleatorio Simple (M.A.S) y Muestreo Estratificado Aleatorio (M.E.A), para la variable Tiempo de Desarrollo de la Enfermedad	201
52. Resumen de resultados de los Modelos del Diseño de Bloques al Azar (D.B.A) y Muestreo Estratificado Doble	202
53. Intervalos de Confianza al 95 % para la estimación de $\mu$ , en cada Variedad, para la variable Tiempo de Desarrollo de la Enfermedad; en cada uno de los Modelos de Muestreo	205
54. Estadística descriptiva del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, para las variedades dentro de cada bloque, bajo el Diseño de Bloques al Azar (D.B.A) y Muestreo estratificado Doble (M.E.D)	206
55. Análisis de Varianza para las variedades susceptibles en Tiempo de Desarrollo de la Enfermedad	209

56. Resumen de resultados de los Modelos del Diseño Completamente al Azar (D.C.A), Muestreo Aleatorio Simple (M.A.S) y Muestreo Estratificado Aleatorio (M.E.A), para la variable Número de Frutos	211
57. Intervalos de Confianza al 95 % para la estimación de $\mu$ , en cada Variedad, para la variable Número de frutos; en cada uno de los Modelos de Muestreo	213
58. Resumen de resultados de los Modelos del Diseño Completamente al Azar (D.C.A), Muestreo Aleatorio Simple (M.A.S) y Muestreo Estratificado Aleatorio (M.E.A), para la variable Número de Frutos con la muestra mayor	214
59. Intervalos de Confianza al 95 % para la estimación de $\mu$ , en cada Variedad, para la variable Número de frutos; en cada uno de los Modelos de Muestreo, considerando la muestra mayor	216
60. Resumen de resultados de los Modelos del Diseño de Bloques al Azar (D.B.A), Muestreo Estratificado Doble (M.E.D) para la variable Número de Frutos	217
61. Estadística descriptiva del Número de Frutos (dedos), para las variedades dentro de cada bloque, bajo el Diseño de Bloques al Azar (D.B.A) y Muestreo estratificado Doble (M.E.D)	219
62. Estadística descriptiva del Número de Frutos (dedos), para las variedades dentro de cada bloque, bajo el Diseño de Bloques al Azar (D.B.A) y Muestreo estratificado Doble (M.E.D), con n mayor	220



## INDICE DE FIGURAS

	<b><u>Página</u></b>
1. Gráfica de $\frac{1.96s}{\sqrt{n}}$ , contra n, para dos valores posibles de la desviación estándar "s"	34
2. Efectos de tres tipos de variables extrañas (C, A, P) sobre las variables explicativas	125
3. Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades estudiadas, en el diseño completamente al azar	153
4. Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades susceptibles, en el diseño completamente al azar	154
5. Promedios de Números de Frutos (dedos), para las variedades estudiadas, en el diseño completamente al azar	156
6. Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades estudiadas, en el diseño de Bloques al Azar	158
7. Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades susceptibles, en el diseño de bloques al azar	159
8. Promedios de Números de Frutos (dedos), para las variedades estudiadas, en el diseño de bloques al azar	162
9. Promedios de Tiempo de Desarrollo de la Enfermedad (días), para las variedades estudiadas, en el Modelo de Muestreo Estratificado Aleatorio	171
10. Promedios del Número de Frutos (dedos), para las variedades estudiadas, en el Modelo de Muestreo Estratificado Aleatorio	175
11. Promedios del Número de Frutos (dedos), para las variedades estudiadas, en el Modelo de Muestreo Estratificado Aleatorio con n=38	178
12. Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades estudiadas, en el modelo de muestreo estratificado doble sin interacción	183
13. Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades susceptibles, en el modelo de muestreo estratificado doble	184

**Página**

14. Promedios de Números de frutos (dedos) para las variedades estudiadas, en el modelo de muestreo estratificado doble	188
15. Promedios de Números de frutos (dedos) para las variedades estudiadas, en el modelo de muestreo estratificado doble	189
16. Promedios de Números de frutos (dedos) para las variedades estudiadas, en el modelo de muestreo estratificado doble con n mayor	193
17. Promedios de Números de frutos (dedos) para las variedades estudiadas, en el modelo de muestreo estratificado doble, con la muestra mayor	194

## **INTRODUCCIÓN**

El muestreo probabilístico es una herramienta de gran utilidad en la investigación científica, el cual permite seleccionar muestras representativas de la población en estudio, cualquiera que sea el caso. Sin embargo en algunos campos de la investigación ha sido poco explorado; empleándose con mayor frecuencia el diseño de experimentos.

En la mayoría de los estudios estadísticos, se aplican por separado la Metodología Estadística del Diseño de Experimentos y la Teoría del Muestreo. Sin embargo ellas pueden combinarse a fin de mejorar la calidad y eficiencia de los resultados.

Fienberg and Tanur (1983) consideran que las coincidencias entre las encuestas por muestreo y el diseño de experimentos son diversas. Muchos investigadores han trabajado en ambas áreas, como Cochran, Yates, Finney, Hartley, Mahalanobis y Neyman, entre otros. Pero, alrededor de 1950, los investigadores de ambas corrientes han desarrollado sus trabajos separadamente. De hecho, se puede observar en la bibliografía actual que los investigadores ignoran esas coincidencias. Lo cual trae como consecuencia, que se desconozca la herencia común de esas dos metodologías, así como la falta de exploración de nuevas coincidencias potenciales cuando se estudian problemas en cualquiera de las dos áreas.

Además, el hecho de que el muestreo y el diseño de experimentos, tengan raíces teóricas comunes no es algo generalmente aceptado. Es por ello que este trabajo pretende indagar modelos de muestreo, al mismo tiempo, contrastar la teoría del muestreo con la teoría del diseño de experimentos, a través de los criterios del diseño de investigación y los principios básicos del diseño de experimentos; a fin de presentar algunas similitudes y diferencias entre ambas teorías. También, se ejemplifica con datos reales esta situación.

Por otro lado, según Patiño (2004), los modelos han servido para estudiar propiedades de los estimadores, para compararlos o para analizar en que situaciones resultan más adecuados que otros. Entre muchas otras aplicaciones del muestreo se encuentra la de representar la falta de respuesta, efecto del investigador, etc. (Mirás, 1985, citado por Patiño, 2004).

En la experimentación agrícola, especialmente en estudios de mecanización, riego, entre otros; se realiza muestreo y luego se aplica el análisis de varianza para un modelo de diseño, obteniendo en la mayoría de los casos un F aproximado.

De acuerdo a lo anterior, en el presente trabajo, se propone un modelo de muestreo basado en la premisa de que se puede realizar un muestreo para simular un diseño de experimentos, es decir, cuando en el estudio el objetivo es comparar efectos de tratamientos. Pero dicho modelo, difiere del correspondiente al diseño de experimentos. Por otra parte, si el objetivo es

estimar parámetros en una población o bien caracterizarla se puede aplicar el modelo de muestreo.

Una vez desarrollado el modelo en cuestión, se emplea para el estudio de datos reales en la experimentación agrícola. Lográndose verificar la funcionalidad del modelo en este tipo de datos.

## **OBJETIVO GENERAL**

- Proponer un modelo de muestreo estratificado doble apoyado en el contraste de los modelos de diseño de experimentos clásicos y diseños de muestreo aleatorio simple y estratificado aleatorio.

## **OBJETIVOS ESPECÍFICOS**

- Proponer un modelo matemático para un muestreo con estratificación doble.
- Ejemplificar el modelo matemático propuesto para la estratificación doble, mediante la utilización de datos reales.
- Contrastar los principios básicos del diseño completamente al azar con los criterios del muestreo aleatorio simple, para la presentación de similitudes y diferencias.
- Contrastar los principios básicos del diseño de bloques al azar con los criterios del muestreo estratificado aleatorio, para la presentación de similitudes y diferencias.

- Relacionar los resultados obtenidos mediante la aplicación del diseño completamente al azar y de un modelo del muestreo aleatorio simple, en datos reales, para la estimación de sus parámetros.
  
- Relacionar los resultados obtenidos mediante la aplicación del diseño de bloques al azar y del modelo propuesto, en datos reales, para la estimación de sus parámetros.

## **METODOLOGÍA**

### **Materiales**

Se empleó parte de la información de un experimento en Musáceas, financiado por FONACIT y CDCH; bajo la dirección del Prof. Oscar Haddad, cuya finalidad era medir la resistencia de variedades de *Musa* a la Sigatoka Negra (***Mycosphaerella fijiensis*, Morelet**); así como de otras variables de producción y desarrollo. Para el caso específico de este trabajo se tomaron en consideración dos variables de estudio: Tiempo de Desarrollo de la enfermedad (TDE) y Números de frutos por racimo (dedos).

En el Apéndice A, se presenta la información detallada sobre la ubicación del ensayo, características de la localidad, diseño experimental del ensayo, tratamientos evaluados (Variedades), y la distribución en campo del experimento.

### **Métodos:**

- Se contrastó el Diseño Completamente al Azar con el Muestreo Aleatorio Simple a través de los Principios fundamentales del Diseño de



Experimentos (repetición, aleatorización y control local) y del Muestreo (representatividad, realismo, aleatorización y control local).

- Se contrastó el Diseño de Bloques al Azar con el Muestreo Estratificado Aleatorio a través de los Principios fundamentales del Diseño de Experimentos (repetición, aleatorización y control local) y del Muestreo (representatividad, realismo, aleatorización y control local).

- Para el Modelo Matemático del Diseño Completamente al Azar

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, r_i$$

Donde:  $y_{ij}$  es la observación obtenida con la aplicación del i-ésimo tratamiento en la j-ésima repetición

$\mu$  es un efecto común a todas las unidades experimentales, denominado media global

$\alpha_i$  es el efecto del i-ésimo tratamiento

$a$ : número de tratamientos a evaluar

$r_i$ : número de repeticiones en el i-ésimo tratamiento

$\varepsilon_{ij}$  es el efecto aleatorio del i-ésimo tratamiento en la j-ésima repetición

Se utilizó el Método de Mínimos Cuadrados Ordinarios, a fin de obtener las estimaciones de los parámetros: media general y promedio por tratamientos,

tanto para el modelo I (efectos fijos) como para el modelo II (efectos aleatorios).

- Con los datos del ensayo en Musáceas, empleando Statistix v.7 y SAS v.8:
  - ✓ se obtuvo la estadística descriptiva para todas las observaciones (60 plantas efectivas), en cada variedad, bajo el Diseño Completamente al Azar, para cada una de las variables bajo estudio.
  - ✓ se realizó un análisis de varianza, a las dos variables evaluadas, para el modelo del Diseño Completamente al Azar sin considerar los bloques. Pero, es importante señalar, que utilizar una metodología a un diseño para analizar datos provenientes de un experimento planificado bajo otro esquema de diseño, no es adecuado; sin embargo, para efectos prácticos en esta investigación se emplearon los datos provenientes del Diseño de Bloques al Azar.

- Para el Modelo Matemático del Diseño de Bloques al Azar

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, b$$

Donde:  $y_{ij}$  es la observación obtenida con la aplicación del i-ésimo

tratamiento en el j-ésimo bloque

$\mu$  es un efecto común a todas las unidades experimentales

$\alpha_i$  es el efecto del i-ésimo nivel del factor A o tratamiento

$\beta_j$  es el efecto del j-ésimo bloque

$\varepsilon_{ij}$  es el efecto aleatorio del  $i$ -ésimo tratamiento en el  $j$ -ésimo bloque

Se empleó el Método de Mínimos Cuadrados Ordinarios, para obtener las estimaciones de los parámetros: media general, promedio por tratamientos y por bloques, tanto para el modelo I (efectos fijos) como para el modelo II (efectos aleatorios).

- Con los datos del ensayo en Musáceas, empleando Statistix v.7 y SAS v.8:
  - ✓ se obtuvo la estadística descriptiva para todas las observaciones (60 plantas efectivas), en cada variedad, bajo el Diseño de Bloques al Azar  $r$ , para cada una de las variables bajo estudio.
  - ✓ se realizó un análisis de varianza, a las dos variables evaluadas, para el modelo del Diseño de Bloques al Azar.
- Se consideró el Modelo planteado por Lohr (2000) para el Muestreo Aleatorio Simple,

$$Y_1, Y_2, \dots, Y_N \text{ independientes con } E_M(Y_i) = \mu \text{ y } V_M(Y_i) = \sigma^2$$

Así como los estimadores para la media, el total y varianza.

- Para ilustrar como funciona el Modelo de Muestreo Aleatorio Simple dentro de un Diseño Completamente al Azar: Se emplearon los datos del experimento en Musáceas, bajo el arreglo de un Diseño Completamente

al Azar, y se obtuvieron las estimaciones a partir del Modelo de Muestreo Aleatorio Simple, empleando Statistix v.7 .

- Se consideró el Modelo planteado por Lohr (2000) para el Muestreo Estratificado Aleatorio,

$$Y_{hj} = \mu_h + \varepsilon_{hj}$$

Así como los estimadores para la media, el total y varianza.

- Para ilustrar como funciona el Modelo de Muestreo Estratificado Aleatorio dentro de un Diseño Completamente al Azar: Se emplearon los datos del experimento en Musáceas, bajo el arreglo de un Diseño Completamente al Azar, y se obtuvieron las estimaciones a partir del Modelo de Muestreo Estratificado Aleatorio, empleando Statistix v.7 y SAS v.8.
- Se propuso un Modelo Matemático para el Muestreo Estratificado Doble, basándose en el modelo para el Muestreo Estratificado, presentado por Lohr (2000),

$$Y_{hj} = \mu_h + \varepsilon_{hj}$$

donde, los  $\varepsilon_{hj}$  son independientes con media 0 y varianza  $\sigma_h^2$

$\mu_h$  es el promedio de las observaciones en el estrato h.

A partir del Modelo de Muestreo Estratificado Doble, se obtuvieron: las sumas de cuadrados, cuadrados medios, esperanzas de los cuadrados medios, etc.

Además desde el punto de vista de muestreo, se generó las estimaciones para la media, el total y la varianza.

- Para ilustrar como funciona el Modelo de Muestreo Estratificado Doble dentro de un Diseño de Bloques al Azar, bajo las condiciones del ensayo (ver Apéndice A): Se emplearon los datos del ensayo en Musáceas, bajo el arreglo del Diseño de Bloques al Azar, y se obtuvieron las estimaciones a partir del Modelo de Muestreo Estratificado Doble, empleando Statistix v.7 y SAS v.8.
- Para las aplicaciones de los Modelos de Muestreo Aleatorio Simple, Muestreo Estratificado Aleatorio y Muestreo Estratificado Doble, en el ensayo de Musáceas, se determinó el tamaño de la muestra a analizar, empleando información suministrada por el grupo de investigadores de musáceas de la UCV - FAGRO, Maracay (Haddad y col., 2005), el cual consiste en el empleo de la fórmula:

$$n_{\infty} = \frac{K^2 S^2}{e^2} \quad (1)$$

Donde;  $S^2$  es la cuasivarianza

$K$  es el multiplicador del error estándar que determina el nivel de confianza. En este trabajo se empleó  $K=2$ , es decir, un nivel de confianza de 95 %.

$e$  es el error máximo admisible en valor absoluto preestablecido

Los autores, indican que el tamaño de muestra puede ser redefinido en atención al tamaño (N) de la población bajo estudio, siguiendo el siguiente criterio:

Si  $\frac{n_{\infty}}{N} \leq 0.05$ , se acepta  $n_{\infty}$  como tamaño de muestra deseable

Si  $\frac{n_{\infty}}{N} > 0.05$ , hay que redefinir a  $n_{\infty}$ , de acuerdo a la siguiente expresión:

$$n = \frac{n_{\infty}}{1 + \frac{n_{\infty}}{N}}$$

Donde n es el tamaño de muestra redefinido, debiendo cumplirse que:

$$n < n_{\infty}$$

Para la aplicación de la fórmula anterior, es necesario estimar la cuasivarianza cuando ésta se desconoce; en este caso, debe conjeturarse a partir de una muestra piloto. Pero, para efectos de este trabajo, ya que se cuenta con toda la información del experimento, se empleó la obtenida para todas las observaciones, en cada variedad.

Para los Métodos de Estratificación, la afijación del tamaño de muestra se realizó de la siguiente forma:

- En el caso de la variable Tiempo de Desarrollo de la Enfermedad, para el Muestreo Estratificado Aleatorio y Muestreo Estratificado Doble, se realizó la afijación mediante el método proporcional, pero como las parcelas tienen igual tamaño la afijación condujo a una afijación igual.

- Para la variable Número de Frutos, en el Muestreo Estratificado Aleatorio y Muestreo Estratificado Doble, fue preciso utilizar la afijación de Neyman, dado que se consideraría tanto el tamaño de los estratos como la variabilidad de los mismos.

# CAPÍTULO I

## ASPECTOS TEÓRICOS PRELIMINARES

### 1. MODELO

Graybill (1976) define cuatro modelos que pueden emplearse para representar un gran número de situaciones reales; los cuales se clasifican como:

- *Modelos Cuantitativos:* Modelo Lineal General y Modelo de Regresión Lineal.
- *Modelos Cualitativos:* Modelo de Diseño y Modelo de los Componentes de Varianza.

En vista de que los modelos que se presentarán en este trabajo son una particularidad del Modelo Lineal general; se describen a continuación algunos aspectos del mismo.

#### **Modelo Lineal General:**

$$Y = X\beta + \varepsilon \quad (1.1.1)$$

Donde:  $Y_{n \times 1}$ : es un vector aleatorio observable

$X_{n \times p}$ : es una matriz de rango  $k$  de variables no aleatorias;  $n > p > k$



$\beta_{p \times 1}$ : es un vector de parámetros desconocidos

$\varepsilon_{n \times 1}$ : es un vector aleatorio no observable

$$\varepsilon \sim \text{NID}(\varepsilon : 0, \sigma_\varepsilon^2)$$

$$Y_{n \times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad X_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}; \quad \beta_{p \times 1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}; \quad \varepsilon_{n \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

Martínez (1988) señala que de acuerdo con Zyzkind (1975 citado por el autor) el enfoque clásico al problema de la estimación del vector fijo  $\beta$ , bajo el modelo anterior, fue iniciado por Legendre y Gauss. El procedimiento utiliza el método de mínimos cuadrados, el cual consiste en minimizar la suma de cuadrados de los errores, es decir, se obtiene el vector  $\beta$  que hace mínima  $\varepsilon'\varepsilon$ . El procedimiento a seguir es: partiendo del modelo (I.1.1) se obtiene que

$$\varepsilon = Y - X\beta \quad (\text{I.1.2})$$

Luego se multiplica por el vector  $\varepsilon'$  en ambos lados de la igualdad, obteniéndose:

$$\varepsilon'\varepsilon = \varepsilon'(Y - X\beta) \Rightarrow (Y - X\beta)'(Y - X\beta)$$

$$\varepsilon'\varepsilon = Y'Y - 2(X\beta)'Y + (X\beta)'(X\beta)$$

derivando en función de  $\beta$  e igualando a cero (0), se obtienen las ecuaciones normales dadas por:

$$(X'X)\hat{\beta} = X'Y \quad (\text{I.1.3})$$

Cabe destacar que si la matriz  $X_{n \times p}$  es de rango completo por columnas; es decir,  $r(X) = p$ . Existe, solución única de este sistema dada por:

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (1.1.4)$$

este estimador tiene la propiedad de que  $E(\hat{\beta}) = \beta$ , cualquiera que sea el valor verdadero de las componentes de  $\beta$ .

En caso de que la matriz  $X$ , sea de rango incompleto, es decir  $r(X) < p$ , es necesario realizar una serie de procedimientos para resolver el sistema (1.1.3), los cuales se basan en la teoría de las funciones estimables, de la cual se presentarán los aspectos teóricos de mayor relevancia.

## 2. ASPECTOS TEÓRICOS DE LAS FUNCIONES ESTIMABLES

Definición 1: "Función lineal estimable" (FLE); es una función lineal de los parámetros  $\beta$ , denotada como  $\ell'\beta$ ; la cual será estimable si y sólo si existe un estimador insesgado de  $\ell'\beta$ , que es una función lineal de los  $Y_i$  en  $Y_{n \times 1}$ . Donde  $\ell = x'a$ ;  $a_{n \times 1}$ : vector de constantes.

Además, para determinar todas las funciones lineales estimables de  $\beta$  y encontrar estimadores óptimos de todas esas funciones (Graybill, 1976).

Definición 2: "Función lineal estimable de  $\hat{\beta}$  linealmente independientes": consiste en un conjunto  $M$  de FLE de  $\beta$ , digamos  $\ell_1'\beta, \ell_2'\beta, \dots, \ell_M'\beta$ , definida como el conjunto  $M$  de FLE linealmente independientes de  $\beta$ , si y sólo si: a) cada  $\ell_m'\beta$  es una función estimable para  $m=1,2,\dots,M$ , b) los  $p \times 1$  vectores  $\ell_1, \ell_2, \dots, \ell_M$  son linealmente independientes; o el  $r(L)=M$ ,  $M \leq p$  (Graybill, 1976)

Definición 3: Dada una matriz de constantes  $L_{p \times M} = [\ell_1, \ell_2, \dots, \ell_M]$ , donde  $\ell_m'\beta$  es una función estimable para cada  $m=1,2,\dots,M$ . Entonces  $L'\beta$  se define como:

1. un conjunto  $M$  de funciones lineales estimables de  $\beta$
2. un conjunto completo de funciones lineales estimables de  $\beta$ , si  $r(L)=k$  ( $r(x)=k$ )
3. un conjunto base de funciones lineales estimables si  $M=k$  y  $r(L)=k$  (Graybill, 1976).

Teorema 1: Sea el modelo de diseño  $Y = X\beta + \varepsilon$ ,  $\varepsilon \sim \text{NID}(\varepsilon: 0, \sigma_1^2)$  y las ecuaciones normales  $(X'X)'\beta = X'Y$

1. Si  $\ell'\beta$  es una FLE de  $\beta$ , entonces  $\ell'\beta$  es invariante para alguna solución  $\hat{\beta}$  de las ecuaciones normales, y  $\ell'\hat{\beta} = \ell'X^{-}Y$

2. Si  $\ell'\beta$  es una FLE de  $\beta$ , entonces  $\ell'\hat{\beta}$  es el mejor estimador lineal insesgado de  $\ell'\beta$ , para alguna solución  $\hat{\beta}$  de las ecuaciones normales.
3.  $\hat{\sigma}^2 = (n-K)^{-1}(Y'Y - \hat{\beta}'X'Y)$  es invariante para alguna solución  $\hat{\beta}$  de las ecuaciones normales
4.  $\hat{\sigma}^2$  es el mejor estimador lineal insesgado de  $\sigma^2$
5. Cualquier fila de  $X\beta$  es una función estimable
6. El mejor estimador lineal insesgado de alguna función estimable  $\ell'\beta$  debe ser una combinación lineal de los mejores estimadores lineal insesgado de de todo conjunto base de funciones estimables (y de todo conjunto completo de funciones estimables) (Graybill, 1976).

Definición 4: una combinación lineal  $\ell'\beta$ , se llama “contraste de los parámetros en  $\beta$ , si y sólo si la suma de los coeficientes de la combinación

lineal es nula:  $\sum_{i=1}^p \ell_i = 0$

Por otro lado se presenta el Teorema de Cochran, tomado de Montgomery (1991)

Teorema 2: “Sean  $Z_i$  variables aleatorias NID  $(0,1)$  para  $i=1,2,\dots,v$  y

$$\sum_{i=1}^v Z_i^2 = Q_1 + Q_2 + \dots + Q_s$$

Donde,  $s \leq v$  y  $Q_i$  tiene  $v$  grados de libertad ( $i=1,2,\dots,s$ ). Entonces  $Q_1, Q_2, \dots, Q_s$  son variables aleatorias independientes con distribución ji-cuadrada y  $v_1, v_2, \dots, v_s$  grados de libertad si y sólo si  $v = v_1 + v_2 + \dots + v_s$ .

Como la suma de los grados de libertad de  $SC_{\text{trat}}$  y de  $SC_{EE}$  es igual a  $n-1$ , es decir, el total de grados de libertad, el Teorema de Cochran implica que  $\frac{SC_{\text{trat}}}{\sigma^2}$  y  $\frac{SC_{EE}}{\sigma^2}$  son variables aleatorias independientes con distribución ji-cuadrada. Por tanto, si la hipótesis nula de la igualdad de medias de los tratamientos es verdadera, la razón:

$$F_c = \frac{SC_{\text{trat}}/a-1}{SC_{EE}/n-a} = \frac{CM_{\text{trat}}}{CM_{EE}} \quad \square \quad F_{1-\alpha; a-1, n-a} \leq F_{\text{tab } (1-\alpha)}$$

Pero si  $F_c \geq F_{\text{tab } (1-\alpha)}$ , se rechaza la hipótesis nula de la igualdad de medias

### 3. GENERALIDADES SOBRE MODELOS

Patiño (2004) define los modelos como una esquematización de la realidad, que puede tener como finalidad comprenderla mejor, interpretar algunas situaciones o resultados, predecir su comportamiento futuro, controlarla dentro de unas cotas o incluso integrarla dentro de una construcción más general. Se puede realizar de muchas maneras, sin ánimo de ser exhaustivo, de modo verbal, lógico, informático, material, gráfico, contable o matemático.

Señala el autor que la utilización de modelos en estadística es antigua, aunque eran menos empleados que en la actualidad y posiblemente aún menos que en el futuro. Desde casi los primeros momentos del desarrollo de la teoría de muestreo en poblaciones finitas se han empleado modelos, cuando menos de un modo implícito – la propia estratificación se puede considerar como una modelización naïf – y, desde luego, claramente a partir del momento en que se introducen los estimadores de razón y de regresión.

El autor considera que existen dos puntos de vista desde los que se puede ver la población finita:

1. *Perspectiva población fija*: con cada unidad de la población está asociada una constante real desconocida pero fija, el valor de la variable de estudio (bajo la hipótesis de que no hay errores ajenos del muestreo).
2. *Perspectiva superpoblación*: con cada unidad de la población está asociada una variable aleatoria para la que se especifica una estructura estocástica, el valor real asociado con una unidad de la población se trata como el resultado de esta variable aleatoria.

Las diferencias conceptuales entre ambos puntos de vista son importantes, de manera resumida, los argumentos empleados por la perspectiva población fija son que al realizar el análisis, la población está fijada, no es el resultado de un proceso aleatorio. Desde la otra perspectiva, la población finita es la realización en el momento de estudio de un determinado modelo

de superpoblación, modelo que describe un mecanismo o proceso en el mundo real.

Por tanto, el mismo autor considera, que un modelo de superpoblación para una población particular es un modelo estocástico que la caracteriza, formalizando las concepciones y el conocimiento que se tienen sobre ella. Los respectivos valores que toma el carácter que interesa para cada uno de los elementos de la población se suponen que son realizaciones de variables aleatorias con una distribución probabilística conjunta particular. El modelo de superpoblación se entiende que es un intento de caracterizar el modo en que la población – y posiblemente otras como ella – ha devenido.

Teniendo presente los puntos de vista antes mencionados, se consideran en relación con las encuestas por muestreo, tres aproximaciones (Lombardía y col., 2003, citado por Patiño, 2004):

1. Estimación basada en el diseño
2. Estimación basada en el modelo de predicción
3. Estimación asistida por el modelo

La divergencia entre dichas aproximaciones se manifiesta especialmente al tratar de la aleatorización previa a la selección o toma de muestras al azar. En la estimación basada en el diseño – el enfoque tradicional de las encuestas por muestreo – es el propio diseño muestral quien determina el modo de estimar la variabilidad del muestreo. Mientras que en la perspectiva

basada en el modelo de predicción, es este último el que determina la forma de estimar la variabilidad, y el diseño no es importante.

Patiño (2004) acota, que los estadísticos que prefieren basar totalmente las inferencias sobre poblaciones finitas en modelos de superpoblación presupuestos (aproximación 2) argumentan que la aleatorización en la selección de la muestra es innecesaria y puede no ser deseable excepto, quizás, para evitar polémicas sobre sesgo de selección. Los autores que defienden esta aproximación sostienen que el conocimiento que el estadístico tiene sobre las características de la población en estudio le permite afirmar que el modelo presupuesto constituye la fuente de aleatoriedad más importante, ya que describe la citada población. En este tipo de inferencias no importa como se ha realizado el plan de muestreo, ni que el estimador tenga en cuenta el diseño, la inferencia se hace sobre la base del modelo, y la aleatorización aportada por la naturaleza se considera suficiente. Pero en estadística no siempre se puede hablar de aleatorización aportada por la naturaleza, sólo en algunos casos como en agricultura, ganadería, selvicultura, medio ambiente y patrimonio natural.

Kish (1995) añade que las teorías de muestreo basadas en modelos gozan de gran atractivo matemático y de estrechos vínculos con las teorías estadísticas clásicas sobre variables aleatorias, y de modo muy destacado y conciso con la estadística matemática.



Por otra parte, señala Patiño (2004), que en muestreo probabilístico tradicional la fuente de aleatoriedad está en el mecanismo de selección de los elementos de las muestras sobre los que se va a basar la estimación. Para los que defienden la visión tradicional basada en el diseño, la estimación predictiva está sujeta a mayores sesgos en la estimación y mayores errores por especificación incorrecta del modelo y, por tanto, la suposición de un modelo puede únicamente determinar el plan de muestreo y el estimador óptimo para ese modelo.

El autor considera, que la tercera aproximación (modelo asistida) puede considerarse en un intento de síntesis de las dos primeras. Consideran el modelado de la población para la construcción de los estimadores pero dentro de la inferencia tradicional basada en el diseño. Las inferencias permanecen basadas en este último y de aquí “robustas” a la mala especificación del modelo.

Lohr (2000) añade que en muchos procedimientos de muestreo habituales, si se adopta un modelo consistente con los motivos por los cuales se realiza determinada estrategia muestral, los estimadores puntuales obtenidos son semejantes a los clásicos basados en el diseño de muestreo, y que en algunos casos coinciden.

Font (1999, citado por Patiño, 2004) expone que las metodologías basadas en el diseño de muestreo y en el modelo de predicción obtienen, en la

ausencia de características auxiliares, resultados de diseño muy "similares". En cambio la coincidencia entre los resultados basados en las dos aproximaciones al problema se "rompe" al introducir en la inferencia la información derivada de una característica auxiliar, defendiendo la primera un muestreo aleatorio y la segunda un muestreo intencionado.

Patiño (2004) señala que los modelos de mayor complejidad necesitan registros administrativos de calidad. El proceso de modelización en ellos implica una selección de variables y, por tanto, cada estadístico podría realizar una diferente. Otra de las críticas, cuando se utilizan covariables, se basa en la simplificación que habitualmente se usa en la búsqueda de las relaciones funcionales, concediendo un excesivo protagonismo a la linealidad, hipótesis difícil de mantener en numerosas ocasiones.

El autor añade que el proceso de selección del modelo y el de verificación final del mismo se consideran como algunas de las principales componentes de la actividad estadística. En la actualidad los defensores de los modelos desde luego no suponen que las observaciones son independientes e idénticamente distribuidas, hipótesis que no se sostendrían en la mayoría de las poblaciones finitas con las que se trabaja en estadística. En la aproximación basada en el diseño, para tratar con este tipo de poblaciones se recurre a diseños complejos.

Por otro lado, señala que la aplicación del modelo correcto es algo que no se puede verificar, como mucho, podría someterse a una validación interna y externa, a partir de la propia muestra o de fuentes ajenas a ella. En esta crítica del modelo, y en el caso de la validación interna...cabe preguntarse, si se puede juzgar la bondad de ajuste del modelo desde una óptica estrictamente verosimilista, o hay que tener en cuenta ya el propio diseño de la investigación en el caso de que este diseño sea probabilístico.

Agrega, que hay dos hipótesis en la metodología basada en modelos que hay que tener presentes por la dificultad de su tratamiento. Una es la de independencia, difícil de contrastar, y la otra es la implícita de que la muestra y la co-muestra (unidades de la población que no están en la muestra) siguen el mismo modelo. Es decir, se está suponiendo que el modelo es válido para las unidades de la población finita que no han sido elegidas en la muestra. En esta última, como en otras hipótesis, la comprobación en fuentes diferentes puede ser una práctica recomendable.

Särndal y col. (1992, citado por Brakel and Renssen, 1998) expresa que en el modelo asistido se asume que el valor de cada elemento de la población finita con respecto a una variable objetivo es una realización de una variable estocástica. Esta variable estocástica es modelada, por ejemplo, de acuerdo a un modelo de regresión lineal con los valores de la variable auxiliar como covariable.

## **CAPÍTULO II**

### **ASPECTOS TEÓRICOS DEL MUESTREO**

Muchos fueron los trabajos revisados concernientes al tema a tratarse, la perseverante búsqueda de los mismos señaló poca información, en cuanto a modelos en muestreo y a la comparación entre la teoría del muestreo y la teoría del diseño de experimentos; sin embargo, dentro de lo relacionado al tema se consideran diferentes conceptos.

#### **1. GENERALIDADES SOBRE MUESTREO**

El **Muestreo**, se refiere al proceso de selección de los elementos, sujetos o casos de una población. Para que el muestreo sea el correcto, es necesario que los elementos escogidos representen bien a toda la población. Si la muestra no cumple este cometido se dice que la muestra es sesgada o tendenciosa; en el caso contrario, cuando la muestra ha sido bien seleccionada, se habla de muestra representativa.

Cochran (1985) clasifica las encuestas por muestreo en dos categorías principales: descriptivas y analíticas.

*Descriptivas:* en una encuesta descriptiva el único objetivo es obtener cierta información respecto a grandes grupos; por ejemplo, el rendimiento promedio de algún cultivo, el número de plantas infectadas por un hongo, etc.

*Analíticas:* en estas se hacen comparaciones entre varios subgrupos de una población, para averiguar si existen ciertas diferencias entre ellos y formular o verificar hipótesis sobre sus causas.

Además acota que la distinción entre encuestas analíticas y descriptivas no es perfectamente clara. Muchas encuestas proporcionan datos que sirven para ambos propósitos. Por ejemplo encuestas sobre el estado de los dientes de los niños antes y después de la fluorización del agua, entre otros.

### **Ventajas del Muestreo en Comparación con un Censo:**

El mismo autor presenta las siguientes ventajas:

- *Costo reducido:* Si los datos se obtienen de una pequeña fracción del total, los gastos son menores que los que se realizarían si se llevara a cabo un censo completo. En poblaciones muy grandes se pueden obtener resultados lo suficientemente exactos cuando se analizan muestras que representan sólo una pequeña fracción de la población.

- *Mayor rapidez:* Por la misma razón, los datos pueden ser recolectados más rápidamente con una muestra que con una enumeración completa. Especialmente cuando se necesita con urgencia la información.
- *Más posibilidades:* Para obtener la información en ciertos tipos de encuestas, se utilizan los servicios de personal altamente calificado o equipo muy especializado de disponibilidad limitada. Por lo tanto, en estos casos el censo completo es impracticable y como alternativa a la obtención de datos por muestreo, sólo existe la de no obtenerlos.
- *Mayor exactitud:* Debido a que al reducir el volumen del trabajo se puede emplear personal más capacitado y someterlo a un entrenamiento intensivo y debido también a que en estas ocasiones será factible la supervisión cuidadosa del trabajo de campo, una muestra puede producir resultados más exactos que la enumeración completa.

## **2. MUESTREO ALEATORIO SIMPLE**

Según Yamane (1967) es el método de selección de  $n$  unidades de la población, tal que cada una de las  $\binom{N}{n}$  muestras posibles tienen la misma probabilidad de ser seleccionadas. Donde  $N$ : tamaño de la población y  $n$ : tamaño de la muestra.

Lohr (2000) lo define como la forma más sencilla de muestreo probabilístico y proporciona la base teórica de las formas más complejas. Señala que existen dos formas de extraer una muestra aleatoria simple: con reemplazo y sin reemplazo.

El mismo autor, indica que para extraer una muestra aleatoria simple, se necesita una lista de todas las unidades de observación que pertenecen a la población, denominado marco de muestreo. En este tipo de muestras la unidad de observación y la unidad de muestreo coinciden. Cada unidad tiene asignado un número, y se elige una muestra de modo que cada unidad tenga la misma posibilidad de aparecer en la muestra y la selección de una unidad no tenga influencia de las demás unidades ya elegidas.

El muestreo aleatorio simple es muy eficiente cuando la población es homogénea.

Según Kish (1972) en el muestreo irrestricto aleatorio existen  $\binom{N}{n}$

combinaciones igualmente probables. Entre ellas hay  $\binom{N-1}{n-1}$  que contienen

un elemento específico. Por lo tanto, su probabilidad de selección es

$$\frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}.$$

**Notación:****Población:**

$N$  denota el tamaño de la población

$S^2$  denota la cuasi varianza poblacional

$$\text{Media poblacional: } \mu = \frac{\sum_{j=1}^N Y_j}{N} \quad (\text{II.2.1})$$

$$\text{Cuasi Varianza poblacional: } S^2 = \sum_{i=1}^N \frac{(Y_i - \mu)^2}{N-1} \quad (\text{II.2.2})$$

$$\text{Varianza de la media: } V(\bar{y}) = \sigma_{\bar{y}}^2 = \frac{S^2}{n} \left(1 - \frac{n}{N}\right) \quad (\text{II.2.3})$$

$$\text{Total poblacional: } \tau = N\mu \quad (\text{II.2.4})$$

$$\text{Varianza del Total: } V(\tau) = \sigma_{\tau}^2 = N^2 V(\bar{y}) = N^2 \left(1 - \frac{n}{N}\right) \frac{S^2}{n} \quad (\text{II.2.5})$$

**En la muestra:**

$$\text{Media muestral: } \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (\text{II.2.6})$$

$$\text{Varianza muestral: } s^2 = \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1} \quad (\text{II.2.7})$$

$$\text{Varianza estimada de la media: } \hat{V}(\bar{y}) = \hat{\sigma}_{\bar{y}}^2 = \frac{s^2}{n} \left(1 - \frac{n}{N}\right) \quad (\text{II.2.8})$$

$$\text{Estimación del total: } \hat{\tau} = N\bar{y} \quad (\text{II.2.9})$$

$$\text{Varianza estimada del Total: } \hat{V}(\hat{\tau}) = N^2 \left(1 - \frac{n}{N}\right) \frac{s^2}{n} \quad (\text{II.2.10})$$



## Teoría de Aleatorización

Se presenta como funciona la teoría de aleatorización de Cornfield (1994, citado por Lohr, 2000) para la deducción de propiedades de la media muestral en el muestreo aleatorio simple, basada en el diseño de muestreo; se considera como "S" a la muestra seleccionada

$$Z_i = \begin{cases} 1 & \text{si la unidad } i \text{ está en la muestra} \\ 0 & \text{en caso contrario} \end{cases}$$

Entonces

$$\bar{y} = \sum_{i \in S} \frac{y_i}{n} = \sum_{i \in S} Z_i \frac{y_i}{n}$$

Las  $Z_i$  son las únicas variables aleatorias en la ecuación anterior, pues, de acuerdo con la teoría de aleatorización, las  $y_i$  son cantidades fijas. Al extraer una muestra aleatoria simple de  $n$  unidades entre las  $N$  unidades de la población,  $\{Z_1, Z_2, \dots, Z_N\}$  son variables aleatorias de tipo Bernoulli, que se distribuyen de manera idéntica con:

$$\pi_i = P(Z_i = 1) = P(\text{elegir la unidad } i \text{ en la muestra}) = \frac{n}{N}$$

Cuando el muestreo aleatorio simple se realiza con reemplazo, la distribución que sigue la variable aleatoria es Binomial, y cuando se lleva a cabo sin reemplazo su distribución es Hipergeométrica. Mientras más grande sea la población la distribución Hipergeométrica tenderá a una distribución Binomial.

$$P(z_i = 1) = P(\text{elegir la unidad } i \text{ en la muestra}) = \frac{n}{N}.$$

$$E(\bar{y}) = \mu$$

La varianza de  $\bar{y}$  también se calcula al utilizar las propiedades de las variables aleatorias  $Z_1, \dots, Z_N$ . Por lo que

$$V(Z_i) = \sigma_{Z_i}^2 = \frac{n}{N} \left( 1 - \frac{n}{N} \right),$$

$$V(\bar{y}) = \sigma_{\bar{y}}^2 = \left( 1 - \frac{n}{N} \right) \frac{S^2}{n}.$$

### Tamaño de Muestra para el Muestreo Aleatorio Simple

Azorín (1972) señala que el diseño óptimo de la muestra, en particular la determinación previa de su tamaño óptimo, sólo podría conseguirse a partir del conocimiento de la población. Para obtener el tamaño de la muestra ( $n$ ), se parte de la ecuación:  $e = k \cdot \sigma_{\bar{y}}$ , expresa el error absoluto como un producto del error de muestreo por una desviación  $k$  correspondiente al nivel de confianza fijado.

$$\begin{aligned} \text{Como } \sigma_{\bar{y}} &= \sqrt{\frac{N-n}{N} \frac{S^2}{n}} & e &= k \cdot \sigma_{\bar{y}} = \sqrt{\frac{N-n}{N} \frac{S^2}{n}}, \\ e^2 &= k^2 \left( \frac{N-n}{N} \right) \frac{S^2}{n} \\ n &= \frac{Nk^2S^2}{Ne^2 + k^2S^2} = \frac{\frac{k^2S^2}{e^2}}{1 + \frac{1}{N} \frac{k^2S^2}{e^2}} \end{aligned} \quad (II.2.11)$$

Donde:  $k$  es la desviación correspondiente al nivel de confianza fijado,

$e$  es el error máximo admisible

$N$  el tamaño de la población a muestrear

$S^2$  la cuasi varianza poblacional

$e$  y  $k$  se fijan de acuerdo a los objetivos planteados, el tamaño de la población, se supone conocido;  $S^2$  habrá que conjeturarlo o apreciarlo a partir de algún conocimiento que se tenga de la población o a partir de una muestra piloto, previamente seleccionada

Mosquera (1974) presenta la siguiente formulación para determinar el tamaño de muestra para la media y el total

### ***Tamaño de muestra para la media***

Siendo  $e = k \cdot \sigma_{\bar{y}}$ , se tiene:  $e^2 = k^2 \cdot \sigma_{\bar{y}}^2$  sustituyendo a  $\sigma_{\bar{y}}^2$  por su valor

$$e^2 = k^2 \left( \frac{N-n}{N} \right) \frac{\hat{S}^2}{n} \Rightarrow Nne^2 = k^2 N \hat{S}^2 - k^2 n \hat{S}^2 \Rightarrow Nne^2 + k^2 n \hat{S}^2 = k^2 N \hat{S}^2$$

$$n(Ne^2 + k^2 \hat{S}^2) = k^2 N \hat{S}^2 \Rightarrow n = \frac{k^2 N \hat{S}^2}{Ne^2 + k^2 \hat{S}^2} \quad (\text{II.2.12})$$

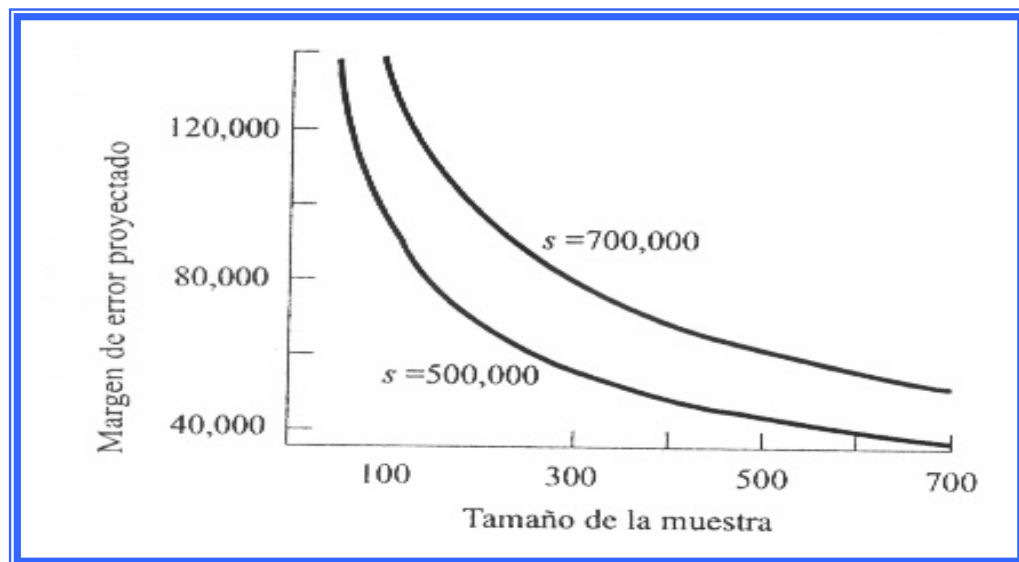
### ***Tamaño de muestra para el total***

Siendo  $e = k \cdot \sigma_{\bar{t}}$ , se tiene:  $e^2 = k^2 \cdot \sigma_{\bar{t}}^2$  sustituyendo a  $\sigma_{\bar{t}}^2$  por su valor

$$e^2 = k^2 \left[ N^2 \left( \frac{N-n}{N} \right) \frac{\hat{S}^2}{n} \right] \Rightarrow e^2 = k^2 \left[ N(N-n) \frac{\hat{S}^2}{n} \right] \Rightarrow ne^2 + k^2 N n \hat{S}^2 = k^2 N^2 \hat{S}^2$$

$$n(e^2 + k^2 N \hat{S}^2) = k^2 N^2 \hat{S}^2 \Rightarrow n = \frac{k^2 N^2 \hat{S}^2}{e^2 + k^2 N \hat{S}^2} \quad (\text{II.2.13})$$

Por otro lado Lohr (2000) señala que se pueden observar algunas de las posibles consecuencias del uso de diversos tamaños de muestra de manera gráfica. La figura 1 muestra el valor de  $\frac{1.96s}{\sqrt{n}}$ , para un rango de tamaños de muestra entre 50 y 700, y para dos valores posibles de la desviación estándar  $s$ . La gráfica muestra que si se ignora la corrección para poblaciones finitas y si la desviación estándar es cercana a 500000, una muestra de tamaño 300 dará un margen de error cercano a 60000.



**Figura 1. Gráfica de  $\frac{1.96s}{\sqrt{n}}$ , contra  $n$ , para dos valores posibles de la desviación estándar “s” (Lohr, 2000)**

Agrega el autor que la determinación del tamaño de muestra es uno de los primeros pasos de una investigación; no existe una fórmula mágica que diga el tamaño de muestra perfecto para la investigación. Mientras más grande sea la muestra, menor será el error de muestreo. Sin embargo, en la mayoría de las encuestas se debe preocupar por dichos errores y se tienen que

programar recursos para controlar los sesgos de selección y de medición. En muchos casos, los errores que no son de muestreo son mayores al considerar una muestra mayor; con una muestra más grande es fácil introducir otras fuentes de error o relajar la exigencia sobre el sesgo de selección.

### **Modelo para el Muestreo Aleatorio Simple**

A continuación se presenta una breve descripción del modelo según Lohr (2000).

Las variables aleatorias en la teoría de la aleatorización no consideran las respuestas  $y_i$ ; simplemente son variables aleatorias que expresan si la  $i$ -ésima unidad está o no en la muestra. En un enfoque basado en el diseño de muestreo (o teoría de aleatorización), la única relación entre las unidades muestreadas y las no muestreadas es que estas últimas podrían haber estado en la muestra de haber utilizado un valor inicial distinto para el generador de números aleatorios.

Por medio de la teoría de aleatorización:  $y_1, y_2, \dots, y_N$  son considerados como valores fijos,  $\bar{y}$  es insesgado porque el promedio de  $\bar{y}_s$  para todas las muestras posibles " $S$ " es igual a  $\mu$ . Las únicas probabilidades utilizadas para determinar el valor esperado y la varianza de  $\bar{y}$  son las probabilidades de que las unidades estén en la muestra.

En estadística, se tiene un enfoque distinto de la inferencia; las variables aleatorias ( $Y_i$ ) siguen cierta distribución de probabilidad y los valores reales de la muestra son realizaciones de esas variables. Así,  $Y_1, Y_2, \dots, Y_n \sim \text{IIDN}(\mu, \sigma^2)$ , y utiliza las propiedades de las variables aleatorias independientes y la distribución normal para determinar los valores esperados de diversos estadísticos.

Se puede extender este enfoque del muestreo considerando que las variables aleatorias  $Y_1, Y_2, \dots, Y_N$  son generadas a partir de cierto **modelo**. Los valores reales para la población finita,  $y_1, y_2, \dots, y_N$  son una realización de las variables aleatorias. La distribución de probabilidad conjunta de  $Y_1, Y_2, \dots, Y_N$  proporciona este vínculo entre las unidades que están en la muestra ( $y_i, i \in S$ ) y las que no lo están ( $y_i, i \notin S$ ), en este enfoque basado en el modelo; este vínculo es el que falta en el enfoque de la aleatorización. En este caso, se muestrea  $y_i, i \in S$ , y se utilizan esos datos para predecir los valores no observados  $y_i, i \notin S$ . Así, los problemas en el muestreo con poblaciones finitas pueden pensarse como problemas de predicción.

En una muestra aleatoria simple, se puede adoptar un modelo sencillo como el siguiente:

$$Y_1, Y_2, \dots, Y_N \text{ independientes con } E_M(Y_i) = \mu \text{ y } V_M(Y_i) = \sigma^2 \quad (\text{II.2.14})$$

El subíndice  $M$  indica que la esperanza utiliza la distribución del modelo y no la correspondiente a la aleatorización, mencionada anteriormente. En este

caso  $\mu$  y  $\sigma^2$  representan parámetros desconocidos de una población infinita, no las cantidades correspondientes en una población finita, como en la teoría de aleatorización. Se extrae una muestra  $S$  y se observan los valores  $y_i$ , para  $i \in S$ ; es decir, se ven realizaciones de la variable aleatoria  $Y_i$  para  $i \in S$ . Las otras observaciones en la población ( $y_i, i \notin S$ ) también son realizaciones de variables aleatorias, pero no se ven éstas.

El total de la población finita  $\tau$ , se puede escribir como:

$$\tau = \sum_{i=1}^N y_i = \sum_{i \in S} y_i + \sum_{i \notin S} y_i \text{ y es un valor posible de la variable aleatoria}$$

$$T = \sum_{i=1}^N Y_i = \sum_{i \in S} Y_i + \sum_{i \notin S} Y_i$$

Se conocen los valores  $\{y_i, i \in S\}$ . Para estimar  $\tau$  a partir de la muestra, se determinan las estimaciones de los valores  $y_i$  que no están en la muestra. Aquí es donde entra el modelo de la media común  $\mu$ . El estimador por mínimos cuadrados de  $\mu$  a partir de la muestra es  $\bar{y}_S = \sum_{i \in S} \frac{Y_i}{n}$  y éste es el mejor predictor lineal insesgado (bajo este modelo) de los valores no observados, de modo que

$$\hat{T} = \frac{N}{n} \sum_{i \in S} Y_i \quad (II.2.15)$$

Es el estimador insesgado de  $T$  para el modelo: si el modelo es cierto, entonces el promedio de  $\hat{T} - T$  bajo varias realizaciones de la población es

$$E_M(\hat{T} - T) = \frac{N}{n} \sum_{i \in S} E_M(Y_i) - \sum_{i=1}^N E_M(Y_i) = 0.$$

Por otro lado, la diferencia entre determinar las esperanzas bajo el enfoque basado en el modelo y bajo el enfoque basado en el diseño de muestreo, es que, en el primero, las  $Y_i$  son variables aleatorias y la muestra no presenta información para calcular los valores esperados. En el diseño de muestreo, las variables aleatorias están contenidas en la muestra  $S$ .

El error cuadrático medio se calcula como el cuadrado de la desviación promedio entre la estimación y el total de la población finita. Para cualquier realización dada de las variables aleatorias el error cuadrado es:

$$E_M\left[(\hat{T} - T)^2\right] = E_M\left[\frac{N}{n} \sum_{i \in S} Y_i - \sum_{i=1}^N Y_i\right]^2 = N^2 \frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right) \quad (\text{II.2.16})$$

En la práctica, si se adopta el modelo (II.2.14), se tiene que estimar  $\sigma^2$  mediante la varianza de la muestra  $s^2$ . Así, los enfoques basados en el diseño o en el modelo conducen a la misma estimación de la población total (II.2.9), la misma estimación de la varianza (II.2.8) y al mismo intervalo de confianza para la media. Sin embargo, los intervalos de confianza tienen distintas interpretaciones. El intervalo de confianza para  $\mu$ , se puede interpretar:



- Basado en el diseño de muestreo: si se consideran todas las muestras aleatorias posibles de tamaño  $n$  a partir de la población finita de tamaño  $N$  y se construye un intervalo de confianza al 95 % para cada muestra, el 95 % de todos los intervalos de confianza contruidos de esta forma incluirán al verdadero valor de  $\mu$ . Así, el intervalo de confianza basado en el diseño de muestreo tiene una interpretación de muestreo repetido.
- Basado en el modelo: el procedimiento del intervalo de confianza produce dos variables aleatorias:  $LI = \bar{y}_s - 1.96(S/\sqrt{n})$  y  $LS = \bar{y}_s + 1.96(S/\sqrt{n})$ . Entonces, al utilizar el modelo para inferir que  $\bar{Y}_s$  tiene aproximadamente una distribución normal con media  $\mu$  y varianza  $S^2/n$ .

$$P(LI \leq \mu \leq LS) = 0.95$$

Aunque los dos tipos de intervalos de confianza se pueden interpretar mediante muestras repetidas, existe una diferencia entre ellos. El nivel de confianza basado en el diseño da la proporción esperada del intervalo de confianza que contendrá a  $\mu$ , a partir del conjunto de todos los intervalos de confianza que se pueden construir extrayendo una muestra aleatoria simple de tamaño  $n$  de la población finita de valores fijos  $(y_1, y_2, \dots, y_N)$ . El nivel de confianza basado en el modelo da la proporción esperada de intervalos de confianza que incluirán a  $\mu$ , a partir del conjunto de todas las muestras que se puedan generar mediante el modelo planteado anteriormente.

### 3. MUESTREO ESTRATIFICADO ALEATORIO

Yamane (1967) lo define como el proceso en el cual se divide la población en estratos, se selecciona una muestra aleatoria simple dentro de cada estrato, las cuales se combinan para formar la muestra completa, a fin de estimar los parámetros de la población.

Según Lorh (2000) en el muestreo estratificado se divide la población de  $N$  unidades de muestreo en  $H$  "capas" o estratos, con  $N_h$  unidades de muestreo en el estrato  $h$ . Para que funcione el muestreo estratificado, se deben conocer los valores de  $N_1, N_2, \dots, N_h$  y se debe tener:

$N_1 + N_2 + \dots + N_H = N$  (número total de unidades en toda la población).

Chou (1977) opina que el muestreo estratificado al azar es uno de los métodos aleatorios que, usando la información disponible sobre la población, trata de diseñar una muestra más eficiente que la obtenida por el procedimiento simple al azar. El proceso de estratificación requiere que la población sea dividida en grupos o clases denominados estratos.

Kish (1972) señala que el muestreo estratificado consiste en los siguientes pasos:

- División de la población completa en subpoblaciones distintas llamadas estratos.
- Selección de una muestra separada dentro de cada estrato.

- Cálculo de una media de cada estrato (u otro estadístico), luego una ponderación de las medias de los estratos.
- Cálculo de las varianzas dentro de cada estrato y posteriormente una estimación combinada para la población.

En el muestreo aleatorio estratificado, que es la forma más sencilla del muestreo estratificado, se toma una muestra aleatoria simple de manera independiente en cada estrato, de modo que se elija al azar  $n_h$  observaciones de las unidades de la población en el estrato  $h$ . Se define  $S_h$  como el conjunto de  $n_h$  unidades en la muestra aleatoria simple para el estrato  $h$ .

Cabe destacar que la estratificación es más eficiente cuando las medias del estrato difieren ampliamente; en este caso, la suma de cuadrados entre los estratos es grande y la variabilidad dentro de los estratos será menor. En consecuencia, al construir los estratos se quiere que las medias de los mismos sean lo más distintas posibles.

Kish (1972) indica que hay tres *razones principales por las que comúnmente se recurre a la estratificación*:

1. La estratificación se utiliza para disminuir las varianzas de las estimaciones de la muestra. La varianza disminuye según el grado en que difieran las medias por estrato y el grado de homogeneidad existente dentro de los estratos.

2. Se pueden formar los estratos para utilizar diferentes métodos y procedimientos dentro de ellos; por ejemplo: Si la distribución física de algunas partes de la población difiere radicalmente, tal vez resulte útil adaptar diferentes procedimientos a las diferentes partes; Puede haber diferencias en las listas disponibles (marcos de muestreo), o preferidas para diversas partes de la población.
3. Los estratos pueden establecerse porque las subpoblaciones dentro de ellos también se definen como dominios de estudio. Un *Dominio* es una parte de la población para la que se han planeado estimaciones separadas en el diseño de la muestra.

Además, Yamane (1967) considera que generalmente es difícil estratificar con respecto a la variable de interés (por ejemplo las ventas en supermercados grandes, medianos o pequeños), generalmente por razones físicas o de costo. En muchos casos la población es estratificada de acuerdo a grupos administrativos, grupos geográficos o características naturales (edad, sexo, etc.).

### **Clasificación de las unidades en la muestra:**

Kish (1972) señala que para cualquier variable que se utilice para la estratificación, debe tenerse información acerca de todas las unidades de muestreo en la población. La información disponible para sólo una pequeña parte de las unidades de muestreo no es, en general, útil para la estratificación. Sin embargo;

- Si no hay información de una pequeña proporción de las unidades de muestreo, éstas pueden simplemente arrojarse en un estrato misceláneo.
- A veces no existe una variable única disponible o preferible para todas las unidades de muestreo en la población, pero tal vez se encuentre una variable relevante para cada porción de la población completa, que se pueda emplear eficientemente.
- Cuando sea demasiado costoso estratificar a la población completa, se puede hacer con base en una muestra mediante el método de muestreo doble.

**Notación** a emplear, tomada de Haddad y Salas (2003):

**Población:**

$h$  denota el estrato

$j$  denota la unidad dentro del estrato

$N_h$  número total de unidades del  $h$ –ésimo estrato

$n_h$  número de unidades en la muestra del  $h$ –ésimo estrato

$W_h = \frac{N_h}{N}$  ponderación del estrato  $h$ –ésimo

$f_h = \frac{n_h}{N_h}$  fracción de muestreo en el estrato  $h$ –ésimo

$y_{hj}$ : valor de la  $j$ –ésima unidad en el  $h$ –ésimo estrato

$\tau_h = \sum_{j=1}^{N_h} Y_{hj}$ : total correspondiente al estrato  $h$ –ésimo

$$\text{Media en el estrato } h\text{–ésimo: } \mu_h = \frac{\sum_{j=1}^{N_h} Y_{hj}}{N_h} \quad (\text{II.3.1})$$

$$\text{Media poblacional: } \mu = \frac{\tau}{N} = \frac{\sum_{h=1}^H \sum_{j=1}^{N_h} Y_{hj}}{N} \quad (\text{II.3.2})$$

$$\text{Total poblacional: } \tau = \sum_{h=1}^H \tau_h = \sum_{h=1}^H Y_h \quad (\text{II.3.3})$$

$$\text{Varianza de la población en el estrato h-ésimo: } \sigma_h^2 = \sum_{j=1}^{N_h} \frac{(Y_{hj} - \bar{Y}_h)^2}{N_h - 1} \quad (\text{II.3.4})$$

**En la muestra:**

$$\text{Media muestral para el estrato h-ésimo: } \bar{y}_h = \frac{\sum_{j=1}^{n_h} y_{hj}}{n_h} \quad (\text{II.3.5})$$

Estimación del total del estrato h-ésimo a través de la muestra:

$$\hat{\tau}_h = N_h \bar{y}_h = N_h \frac{\sum_{j=1}^{n_h} y_{hj}}{n_h} \quad (\text{II.3.6})$$

$$\text{Varianza muestral en el estrato h-ésimo: } s_h^2 = \sum_{j=1}^{n_h} \frac{(y_{hj} - \bar{y}_h)^2}{n_h - 1} \quad (\text{II.3.7})$$

**Estimadores de la Media y el Total del muestreo estratificado**

$$\text{Estimador de la media: } \bar{y} = \frac{\sum_{h=1}^H N_h \bar{y}_h}{N} = \sum_{h=1}^H W_h \bar{y}_h \quad (\text{II.3.8})$$

$$\text{Estimador del total: } \hat{\tau} = N \bar{y} \quad (\text{II.3.9})$$

$$\text{Varianza del estimador de la media: } V(\bar{y}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) \left(\frac{N_h}{N}\right)^2 \frac{s_h^2}{n_h} \quad (\text{II.3.10})$$

$$\text{Varianza del estimador del total: } V(\hat{\tau}) = N^2 V(\bar{y}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) N_h^2 \frac{s_h^2}{n_h} \quad (\text{II.3.11})$$

### **Criterios para Formar los Estratos**

Cochran (1985) y Sanchez (1979) coinciden en que existen tres criterios que se deben tener presentes cuando se esta planeando utilizar el muestreo estratificado:

- Los estratos deben formarse de tal manera que se garantice la independencia entre los estratos. Es decir, los estratos deben ser completamente independientes en el proceso de selección y de estimación.
- Las mediciones dentro de los estratos deben ser homogéneas (baja variabilidad). Las mediciones entre estratos deben ser heterogéneas (alta variabilidad).
- Una vez determinados los estratos, se extrae una muestra de cada uno, las extracciones deben hacerse independientemente en los diferentes estratos; esto permite la aplicación simultánea de métodos de muestreo diferentes de acuerdo con la información de que se disponga, el costo y las razones que motivaron la estratificación. Los tamaños de muestras dentro de los estratos se denotan con  $n_1, n_2, \dots, n_H$ ; respectivamente.

Además Cochran (1985) menciona que la estratificación puede dar lugar a una ganancia en la precisión de las estimaciones de características de la población total. Quizá sea posible dividir una población heterogénea en

subpoblaciones, en las que cada una sea internamente homogénea. Esto es lo que sugiere nombre de estratos, con su implicación de una división en capas. Si cada estrato es homogéneo, en cuanto a que las medidas varíen ligeramente de una unidad a otra, una estimación precisa de cualquier media de estrato se puede obtener a partir de una pequeña muestra en dicho estrato. Y posteriormente podrán combinarse estas estimaciones en una estimación precisa para toda la población.

### **Número de Estratos:**

En cuanto al número de estratos que deben formarse para cualquier variable aislada de estratificación Kish (1972) señala que la alternativa existe cuando la variable de estratificación es continua, por ejemplo, ingreso en efectivo; o cuando su división puede proseguir durante un largo periodo, como la división geográfica de un país en regiones, luego en estados, después en municipios, etc.

Considera Kish (1972) que: a) los estratos muy pequeños, contribuyen muy poco a las ganancias de la estratificación; estas ganancias son proporcionales a la ponderación  $W_h$  del estrato. b) la formación de sólo unos cuantos estratos producirá típicamente la mayoría de las ganancias posibles a partir de una variable. Las subdivisiones adicionales de estas no resultan más que unas ganancias pequeñas adicionales. Quizá sea suficiente tener entre 3 y 10 estratos para cualquier variable aislada.



Según Lohr (2000) el número de estratos elegidos depende de muchos factores; por ejemplo, la dificultad para construir un marco de muestreo con la información estratificada y el costo de la estratificación. Una regla general a tener en mente es la siguiente: mientras menos información exista, menos estratos se deben utilizar. Así, se debe emplear una muestra aleatoria simple si de antemano se tiene poca información sobre la población objetivo.

Además señala Kish (1972) que el efecto de aumentar el número  $H$  de estratos puede representarse por el modelo:  $\frac{R^2}{H^2} + (1-R^2)$ . Aquí  $R^2$  es la porción de la varianza afectada por la estratificación y corresponde a la relación entre las variables de estratificación y de la encuesta. La cual disminuye con el cuadrado del número de estratos. Pero la porción  $(1-R^2)$  de la varianza no relacionada con la variable de estratificación no se afecta al aumentar los estratos. Así, la varianza se aproxima a este nivel después de la creación de un número moderado de estratos. Por ejemplo,  $R^2 = 0.64$  representa una correlación vigorosa de  $R = 0.8$  entre las variables de estratificación y de encuesta. Sin embargo, con  $H = 6$  estratos se reducirá la varianza a  $0.018 + 0.36 = 0.378$ ; la duplicación del número de estratos a  $H' = 12$  producirá una disminución adicional de sólo  $0.004 + 0.36 = 0.364$ . Este modelo se ha desarrollado para afijación óptima con regresión lineal de la variable de estratificación. También parece que se comporta bien en pruebas con variables no métricas y para afijación proporcional.

### Tamaño de la Muestra y Afijación de la Muestra a los Estratos:

Azorín (1972) señala que en el caso más general del muestreo estratificado aleatorio, sin especificar el tipo de afijación empleado, si se fija el error máximo admisible y  $k$  (dependiendo del nivel de confianza); se puede emplear la siguiente ecuación para determinar el tamaño de la muestra:

#### Para estimar la media

$$e = k \cdot \sigma_{\bar{y}} \Rightarrow e^2 = k^2 \cdot \sigma_{\bar{y}}^2 \Rightarrow e^2 = k^2 \frac{1}{N^2} \sum_{h=1}^H N_h (N_h - n_h) \frac{S_h^2}{n_h}$$

$$e^2 = k^2 \left( \sum_{h=1}^H \frac{w_h S_h^2}{n_h} - \frac{\sum_{h=1}^H w_h S_h^2}{N} \right) \text{ haciendo ahora a } n_h = n w_h \text{ y dividiendo por } k^2$$

$$\frac{e^2}{k^2} = \sum_{h=1}^H \frac{w_h S_h^2}{n w_h} - \frac{\sum_{h=1}^H w_h S_h^2}{N} \Rightarrow \frac{e^2}{k^2} + \frac{\sum_{h=1}^H w_h S_h^2}{N} = \frac{1}{n} \sum_{h=1}^H \frac{w_h S_h^2}{w_h}$$

$$n = \frac{\sum_{h=1}^H \frac{w_h S_h^2}{w_h}}{\frac{e^2}{k^2} + \frac{\sum_{h=1}^H w_h S_h^2}{N}} \quad (\text{II.3.12})$$

#### Para estimar el total

$$e = k \cdot \sigma_{\bar{t}} \Rightarrow e^2 = k^2 \cdot \sigma_{\bar{t}}^2 \Rightarrow e^2 = k^2 \sum_{h=1}^H N_h (N_h - n_h) \frac{S_h^2}{n_h}$$

$$e^2 = k^2 \left( \sum_{h=1}^H \frac{N_h^2 S_h^2}{n_h} - \sum_{h=1}^H N_h S_h^2 \right) \text{ haciendo ahora a } n_h = n w_h$$

$$n = \frac{\sum_{h=1}^H \frac{N_h^2 S_h^2}{w_h}}{\frac{e^2}{k^2} + \sum_{h=1}^H N_h S_h^2} \quad (\text{II.3.13})$$

Se debe recordar que el objetivo del muestreo es proporcionar estimadores con alta precisión al menor costo posible. En el muestreo estratificado también se presenta el problema... ¿cual es el tamaño de muestra  $n$ ?, existen muchas maneras de dividir  $n$  entre los tamaños de muestra de los estratos:  $n_1, n_2, \dots, n_h$  (afijación de la muestra).

Sánchez (1979) y Haddad y Salas (2003), llaman afijación de la muestra a la distribución o reparto de las  $n$  unidades de la muestra, entre los  $h$  estratos, de manera que  $n=n_1+n_2+\dots+n_h$ ; y entre los tipos de afijación mencionan:

- **Afijación proporcional o afijación Bowley:** consiste en repartir el tamaño de muestra  $n$  en forma proporcional al tamaño de los estratos de la población. Las  $n$  unidades de la muestra se distribuyen proporcionalmente a los tamaños de los estratos.
- **Afijación uniforme o igual:** consiste en asignar el mismo número de unidades muestrales a cada estrato.

- **Afijación de mínima varianza o afijación de Neyman:** consiste en determinar los valores de  $n_h$  de forma que para un tamaño de muestra fijo igual a  $n$  la varianza sea mínima.

El tamaño de la muestra ( $n_h$ ) para cada estrato es proporcional al producto del tamaño del estrato correspondiente ( $N_h$ ) por su desviación ( $S_h$ : cuasidesviación estándar).

Es de hacer notar que no sólo toma en cuenta el tamaño del estrato como en la afijación proporcional, sino también la variabilidad del estrato, lo cual genera resultados más precisos. Si  $S_h$  fuese la misma para cada estrato ( $S_1=S_2=\dots=S_H$ ), la afijación sería proporcional.

Para la obtención del tamaño de muestra en cada estrato, se emplea la siguiente ecuación:

$$n_h = \frac{N_h S_h}{\sum_{h=1}^H N_h S_h} n \quad (II.3.14)$$

En algunas ocasiones se define la afijación de Neyman, considerando  $n_h \propto$

$$N_h \sigma_h$$

- **Afijación óptima para costos variables:** consiste en minimizar la varianza para un costo fijo.

- **Afijación arbitraria:** se selecciona un número cualquiera de elementos de cada estrato, sin considerar el tamaño ( $N_h$ ) ni la variabilidad ( $S_h$ ) de cada uno de ellos, pero debe cumplirse que  $n_1+n_2+\dots+n_H=n$

### Modelo para el Muestreo Estratificado.

Lohr (2000) plantea que el modelo de análisis de varianza en un sentido con efectos fijos proporciona una estructura subyacente para el muestreo estratificado. En este caso,

$$Y_{hj} = \mu_h + \varepsilon_{hj} \quad (\text{II.3.15})$$

donde, los  $\varepsilon_{hj}$  son independientes con media 0 y varianza  $\sigma_h^2$ . Entonces, el estimador por mínimos cuadrados de  $\mu_h$  obtenido a partir de las unidades de la muestra es el promedio de las observaciones de la muestra en el estrato  $h$ .

La variable aleatoria  $T_h = \sum_{j=1}^{N_h} Y_{hj}$ , representa el total del estrato  $h_j$  y la variable

$T = \sum_{h=1}^H Y_h$ , representa el total global.

El mejor estimador lineal insesgado de  $T_h$  es  $\hat{T}_h = \frac{N_h}{n_h} \sum_{j \in S_h} Y_{hj}$  (II.3.16)

$$E_M[\hat{T}_h - T_h] = 0$$

$$E_M\left[(\hat{T}_h - T_h)^2\right] = N_h^2 \frac{\sigma_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right)$$

Como se obtienen muestras independientes en los estratos,

$$E_M \left[ \left( \hat{T} - T \right)^2 \right] = E_M \left\{ \left[ \sum_{h=1}^H \left( \hat{t}_h - T_h \right) \right]^2 \right\}$$

$$E_M \left[ \left( \hat{T} - T \right)^2 \right] = E_M \left[ \sum_{h=1}^H \left( \hat{t}_h - T_h \right)^2 + \sum_{h=1}^H \sum_{k \neq h} \left( \hat{t}_h - T_h \right) \left( \hat{t}_k - T_k \right) \right]$$

$$E_M \left[ \left( \hat{T} - T \right)^2 \right] = E_M \left[ \sum_{h=1}^H \left( \hat{t}_h - T_h \right)^2 \right]$$

$$E_M \left[ \left( \hat{T} - T \right)^2 \right] = \sum_{h=1}^H \left( 1 - \frac{n_h}{N_h} \right) N_h^2 \frac{\sigma_h^2}{n_h} \quad (II.3.17)$$

Se puede estimar la varianza  $\sigma_h^2$  mediante  $s_h^2$ . La adopción de este modelo produce para  $\tau$  y su varianza las mismas estimaciones, que se obtienen en el caso de la teoría de aleatorización (II.3.9 y II.3.11)

## **CAPÍTULO III**

### **ASPECTOS TEÓRICOS DEL DISEÑO DE EXPERIMENTOS**

#### **1. DISEÑO COMPLETAMENTE AL AZAR (D.C.A.)**

Como lo señalan Cochran y Cox (1997), es el tipo de diseño más sencillo en el que los tratamientos son asignados completamente al azar a las unidades experimentales. Además, la aleatorización puede ser apropiada:

- Donde el material experimental es homogéneo
- Donde es probable que una parte apreciable de las unidades se destruyan o no respondan
- En experimentos pequeños en donde la mayor precisión de otros diseños no compensa la pérdida de grados de libertad del error.

Martínez (1988) señala que tal condición de homogeneidad en el material experimental, permite que cada tratamiento pueda ensayarse con el número de repeticiones que se desee.

Chacín (2000) presenta las siguientes ventajas y desventajas:

#### **Ventajas:**

1. Tanto la planificación como el análisis son los más simples si se le compara con los otros diseños conocidos.

2. Produce el máximo número de grados de libertad para el error experimental, lo que es muy útil en pequeños ensayos
3. Cuando existe un número desigual de replicaciones por tratamientos no es causa de complicaciones en el análisis estadístico.
4. Es de fácil manejo en el campo y en el laboratorio

**Desventajas:**

1. Sólo puede usarse con material experimental homogéneo
2. Su utilidad es muy restringida a los experimentos de campo, debido a la heterogeneidad del suelo, que puede ser muy grande y para evitar que la misma enmascare los resultados del experimento, debe recurrirse a experimentos más complejos.
3. Presenta desventaja cuando el material experimental es heterogéneo, ya que no se puede aumentar mucho el tamaño del experimento, porque esto motiva a variaciones altas que enmascaran el efecto del experimento.

**Modelo Lineal para el Diseño Completamente al Azar**

De acuerdo a lo establecido por los autores Neter (1974), Federer (1955), Martínez (1988), Graybill (1976), Montgomery (1991), Chou (1977), Chacín (2000), Kuehl (2001), Steel y Torrie (1988), Wang y Chow (1994) entre otros; se presenta un resumen del desarrollo para el Modelo Lineal Aditivo, que es usualmente referido como, un modelo para una vía de clasificación, o clasificación en un sentido; porque sólo se investiga un factor:



$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, r_i \quad (\text{III.1.1})$$

$$a > 1 \quad r_i \geq 1 \quad \forall i; \quad \sum_{i=1}^a r_i = n$$

Debe destacarse, que el factor a estudiar en este modelo puede ser de efectos fijos o de efectos aleatorios. Los cuales se describen a continuación:

**Modelo de Efectos Fijos (Modelo I):**

Las conclusiones que se generan a partir de este modelo (III.1.1) son pertinentes sólo a los tratamientos incluidos en el estudio. El Modelo I es relevante cuando los tratamientos son escogidos porque existe un particular interés en ellos y no son considerados como una muestra de una población más grande.

Donde:  $y_{ij}$  es la observación obtenida con la aplicación del i-ésimo tratamiento en la j-ésima repetición

$\mu$  es un efecto común a todas las unidades experimentales, denominado media global

$\alpha_i$  es el efecto del i-ésimo tratamiento

$a$ : número de tratamientos a evaluar

$r_i$ : número de repeticiones en el i-ésimo tratamiento

$\varepsilon_{ij}$  es el efecto aleatorio del i-ésimo tratamiento en la j-ésima repetición

Si  $r_1 = r_2 = \dots = r_a$ , se denomina Modelo Lineal Balanceado.

Por otro lado, es importante señalar:

- El parámetro del  $i$ -ésimo tratamiento es definido como  $\alpha_i$
- La media del  $i$ -ésimo tratamiento es definida como  $\mu + \alpha_i$
- El efecto principal del  $i$ -ésimo tratamiento es definido como  $\alpha_i - \bar{\alpha}_.$ ,

$$\text{donde } \bar{\alpha}_. = \frac{1}{a} \sum_{i=1}^a \alpha_i$$

El Modelo del D.C.A, puede expresarse de manera matricial (I.1.1), como:

$$Y_{nx1} = X_{n \times p} \beta_{p \times 1} + \varepsilon_{nx1} \quad \varepsilon_{nx1} \sim \text{NIID}(0, \sigma^2)$$

Donde;

$$Y_{nx1} = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1r_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2r_2} \\ \vdots \\ y_{a1} \\ \vdots \\ y_{ar_a} \end{bmatrix} \quad X_{n \times p} = \begin{array}{c} \begin{matrix} \mu & \alpha_1 & \alpha_2 & & \alpha_a \end{matrix} \\ \begin{matrix} r_1 \\ r_2 \\ \vdots \\ r_a \end{matrix} \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} \end{array}$$

$$\beta_{p \times 1} = \begin{bmatrix} \mu \\ \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_a \end{bmatrix}_{(a+1) \times 1}$$

Como se puede observar, los elementos de la matriz  $X_{n \times p}$ , son ceros (0) y unos (1). Además,  $X_{n \times p}$  presenta las siguientes características:

- Puede ser particionada como  $X_{n \times p} = [x_0 \ x_1 \ \dots \ x_q]$ ; donde  $x_j$  tiene orden  $n \times p_j$   $j = 0, 1, \dots, q$ ; además  $p_0 + p_1 + p_2 + \dots + p_q = p$
- Cada fila de  $x_j$  contiene un solo uno (1) y el resto es ceros (0)
- Cada columna de  $x_j$  contiene al menos un uno (1)

El Modelo de Diseño es un Modelo Lineal Estadístico de Rango Incompleto; ya que las  $p$  columnas de  $X_{n \times p}$  no son linealmente independientes, porque su primera columna es la suma del resto de las columnas. De allí que  $r(x) = a$ .

Para obtener los estimadores de los parámetros  $\mu, \alpha_1, \dots, \alpha_a$ ; se empleará el "Método de los Mínimos Cuadrados":

$$Q(\mu, \alpha_1, \dots, \alpha_a) = \sum_{i=1}^a \sum_{j=1}^{r_i} (y_{ij} - \mu - \alpha_i)^2$$

$$\frac{\partial Q}{\partial \mu} = -2 \sum_{i=1}^a \sum_{j=1}^{r_i} (y_{ij} - \hat{\mu} - \hat{\alpha}_i) = 0$$

$$\frac{\partial Q}{\partial \alpha_i} = -2 \sum_{j=1}^{r_i} (y_{ij} - \hat{\mu} - \hat{\alpha}_i) = 0 \quad i = 1, \dots, a$$

Desarrollando estas ecuaciones se obtiene:

$$\sum_{i=1}^a \sum_{j=1}^{r_i} y_{ij} - \sum_{i=1}^a \sum_{j=1}^{r_i} \hat{\mu} - \sum_{i=1}^a \sum_{j=1}^{r_i} \hat{\alpha}_i = 0$$

$$y_{..} - n\hat{\mu} - \sum_{i=1}^a r_i \hat{\alpha}_i = 0 \Rightarrow n\hat{\mu} + \sum_{i=1}^a r_i \hat{\alpha}_i = y_{..}$$

$$\sum_{j=1}^{r_i} y_{ij} - \sum_{j=1}^{r_i} \hat{\mu} - \sum_{j=1}^{r_i} \alpha_i = 0 \quad i = 1, 2, \dots, a,$$

$$y_{i.} - r_i \hat{\mu} - r_i \hat{\alpha}_i = 0 \Rightarrow r_i \hat{\mu} + r_i \hat{\alpha}_i = y_{i.}$$

De aquí se generan las ecuaciones normales del modelo

$$\left\{ \begin{array}{lcl} n\hat{\mu} + \sum_{i=1}^a r_i \hat{\alpha}_i & = & y_{..} \\ r_1 \hat{\mu} + r_1 \hat{\alpha}_1 & = & y_{1.} \\ r_2 \hat{\mu} + r_1 \hat{\alpha}_1 & = & y_{2.} \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ r_a \hat{\mu} + r_a \hat{\alpha}_a & = & y_{a.} \end{array} \right\}$$

Resultando matricialmente (I.1.3)

$$\text{Donde: } (X'X)_{p \times p} = \begin{bmatrix} n & r_1 & r_2 & \dots & r_a \\ r_1 & r_1 & 0 & \dots & 0 \\ r_2 & 0 & r_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_a & 0 & 0 & \dots & r_a \end{bmatrix} \quad \hat{\beta}_{p \times 1} = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_a \end{bmatrix} \quad X'Y = \begin{bmatrix} \sum_{i=1}^a \sum_{j=1}^{r_i} y_{ij} \\ \sum_{j=1}^{r_1} y_{1j} \\ \sum_{j=1}^{r_2} y_{2j} \\ \vdots \\ \sum_{j=1}^{r_a} y_{aj} \end{bmatrix} = \begin{bmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ \vdots \\ y_{a.} \end{bmatrix}$$

Las soluciones a las ecuaciones normales del modelo se presentan con mayor detalle en el Apéndice B, de allí se obtiene:

Para  $\mu$ :

$$y_{..} = n\hat{\mu} + \sum_{i=1}^a r_i \hat{\alpha}_i \quad (\text{III.1.2})$$

Para  $\alpha_i$ :

$$y_{i.} = r_i \hat{\mu} + r_i \hat{\alpha}_i \quad i = 1, 2, \dots, a \quad (\text{III.1.3})$$

Donde;

$$y_{..} = \sum_{i=1}^a \sum_{j=1}^{r_i} y_{ij} \quad \text{gran total}$$

$$y_{i.} = \sum_{j=1}^{r_i} y_{ij} \quad \text{total para el } i\text{--ésimo tratamiento}$$

$i = 1, 2, \dots, a$

En vista de que la matriz X es de rango incompleto, se utilizan los contrastes de las funciones estimables (definición 4; pág. 8), resultando:

$$\sum_{i=1}^a r_i \hat{\alpha}_i = 0$$

En virtud de lo anterior, las soluciones de las ecuaciones normales son:

$$n\hat{\mu} + \sum_{i=1}^a r_i \hat{\alpha}_i = y_{..} \Rightarrow n\hat{\mu} = y_{..} \Rightarrow \hat{\mu} = \frac{y_{..}}{n}$$

$$\hat{\mu} = \bar{y}_{..} \quad (\text{estimador para } \mu) \quad (\text{III.1.4})$$

$$r_i \hat{\mu} + r_i \hat{\alpha}_i = y_{i.} \Rightarrow r_i \hat{\alpha}_i = y_{i.} - r_i \hat{\mu} \Rightarrow \hat{\alpha}_i = \frac{y_{i.} - r_i \hat{\mu}}{r_i}$$

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..} \quad (\text{estimador para } \alpha_i) \quad i = 1, \dots, a \quad (\text{III.1.5})$$

$$\hat{\varepsilon}_{ij} = y_{ij} - \hat{\mu} - \hat{\alpha}_i \Rightarrow \hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_{..} - (\bar{y}_{i.} - \bar{y}_{..})$$

$$\hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_{i.} \quad (\text{estimador para el error}) \quad (\text{III.1.6})$$

Por lo tanto el vector

$$\hat{\beta}_{px1} = \begin{bmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \bar{y}_{2.} - \bar{y}_{..} \\ . \\ . \\ \bar{y}_{a.} - \bar{y}_{..} \end{bmatrix}$$

En el Modelo del D.C.A.:

➤ El estimador insesgado de la media del i-ésimo tratamiento, es decir,

$$\mu + \alpha_i \text{ es } \hat{\mu} + \hat{\alpha}_i = \bar{y}_{i.}$$

- El estimador insesgado del efecto principal del i-esimo tratamiento, es

$$\text{decir, } \alpha_i - \bar{\alpha} = \hat{\alpha}_i - \hat{\bar{\alpha}} = \bar{y}_{i\cdot} - \frac{1}{a} \sum_{i=1}^a \bar{y}_{i\cdot}.$$

- El estimador de  $\sigma^2$  es  $\hat{\sigma}^2 = \frac{1}{n-a} \sum_{i=1}^a \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i\cdot})^2$ , que matricialmente

$$\text{resultaría de la forma: } \hat{\sigma}^2 = \frac{1}{n-a} (Y'Y - \hat{\beta}'X'Y)$$

### **Pruebas de Hipótesis:**

Una vez obtenidos los estimadores de los parámetros del modelo, se procede a realizar la Prueba de Hipótesis de interés entre los efectos principales de los i-ésimos tratamientos, las cuales quedan planteadas de la siguiente forma:

$$H_0: \alpha_1 - \bar{\alpha} = \alpha_2 - \bar{\alpha} = \dots = \alpha_a - \bar{\alpha} = 0$$

vs

$$H_a: (\alpha_i - \bar{\alpha}) \neq (\alpha_{i'} - \bar{\alpha}) \text{ para algún } i \neq i'$$

ó

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_a$$

vs

$$H_a: \alpha_i \neq \alpha_{i'} \text{ para algún } i \neq i'$$

Matricialmente

$$H_0: H_{q \times p} \beta_{p \times 1} = h_{q \times 1}$$

vs

$$H_a: H_{q \times p} \beta_{p \times 1} \neq h_{q \times 1}$$

$$H_0: \begin{bmatrix} 0 & 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & 0 & \dots & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \vdots \\ \alpha_a \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$$

$\begin{matrix} \mu & \alpha_1 & \alpha_2 & \alpha_3 & \dots & \alpha_a \end{matrix}$

*Análisis de Varianza:*

A fin de realizar las pruebas de hipótesis, se aplica el Análisis de Varianza (ANAVAR)

A partir de (III.1.1) se obtiene que:

$$y_{ij} = \mu + (\alpha_{i\cdot} - \bar{\alpha}) + (y_{ij} - \alpha_{i\cdot})$$

$$y_{ij} - \mu = (\alpha_{i\cdot} - \bar{\alpha}) + (y_{ij} - \alpha_{i\cdot})$$

Como  $\mu$  y  $\alpha_{i\cdot}$  se desconocen, se consideran sus estimaciones (III.1.4 y III.1.5), obteniéndose:

$$y_{ij} - \bar{y}_{..} = (\bar{y}_{i\cdot} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i\cdot})$$

Si se elevan al cuadrado todos los términos de la ecuación anterior y se suman en "i" y en "j" se obtiene:

$$\sum_{i=1}^a \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^{r_i} (\bar{y}_{i\cdot} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i\cdot})^2$$

Esta ecuación muestra que la suma total de cuadrados corregida, puede descomponerse en:

$$SC_{\text{total}} = SC_{\text{Trat}} + SC_{\text{EE}}$$



donde;

$SC_{\text{trat}}$ : es la suma de cuadrados debida a los tratamientos, es decir, entre tratamientos. Tiene  $a-1$  grados de libertad, porque tiene  $a$  niveles de tratamientos.

$SC_{\text{Total}}$ : es la suma de cuadrados debida a la variabilidad total. Tiene  $n-1$  grados de libertad porque hay un total de  $n$  observaciones.

$SC_{\text{EE}}$ : es la suma de cuadrados debida al error, es decir, dentro de tratamientos. Tiene  $n-a$  grados de libertad.

Resultando los Cuadrados Medios de la forma:

$$CM_{\text{EE}} = \frac{SC_{\text{EE}}}{n-a}$$

Si no hay diferencias entre las medias de  $a$  tratamientos, puede usarse la variación de los promedios de los tratamientos con respecto al promedio general para estimar  $\sigma^2$ .

$$\frac{SC_{\text{trat}}}{a-1} = \frac{\sum_{i=1}^a r_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{a-1}$$

Esta es una estimación de  $\sigma^2$  si las medias de los tratamientos son iguales.

$$CM_{\text{trat}} = \frac{SC_{\text{trat}}}{a-1}$$

Por el Teorema de 2 (pág. 8):  $F_c = \frac{CM_{trat}}{CM_{EE}} \square F_{1-\alpha; \alpha-1, n-\alpha}$

El procedimiento antes indicado se resume en el cuadro 1

**Cuadro 1: Análisis de Varianza para el Diseño Completamente al Azar (D.C.A)**

F. de V.	g. de L.	SC	CM	F
Tratam.	$\alpha-1$	$\sum_{i=1}^{\alpha} \frac{(y_{i\cdot})^2}{r_i} - \frac{(y_{\cdot\cdot})^2}{n}$	$\frac{SC_{Trat}}{\alpha-1}$	$\frac{CM_{trat}}{CM_{EE}}$
EE	$n-\alpha$	$\sum_{i=1}^{\alpha} \sum_{j=1}^{r_i} (y_{ij})^2 - \sum_{i=1}^{\alpha} \frac{(y_{i\cdot})^2}{r_i}$	$\frac{SC_{EE}}{n-\alpha}$	
Total	$n-1$	$\sum_{i=1}^{\alpha} \sum_{j=1}^{r_i} (y_{ij})^2 - \frac{(y_{\cdot\cdot})^2}{n}$		

*Esperanza de los Cuadrados Medios (para mayor detalle ver Apéndice B):*

Para Tratamientos:

$$E(CM_{trat}) = E\left(\frac{SC_{Trat}}{\alpha-1}\right) = \frac{1}{\alpha-1} E(SC_{Trat})$$

$$E(CM_{trat}) = \frac{1}{\alpha-1} E\left[\sum_{i=1}^{\alpha} \sum_{j=1}^{r_i} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2\right]$$

desarrollando esta suma y aplicando las propiedades de la esperanza se obtiene:

$$E(CM_{\text{trat}}) = \sigma^2 + \frac{\sum_{i=1}^a \sum_{j=1}^{r_i} (\alpha_i)^2}{a-1}$$

Para el Error:

$$E(CM_{\text{EE}}) = E\left(\frac{SC_{\text{EE}}}{n-a}\right) = \frac{1}{n-a} E(SC_{\text{EE}})$$

$$E(SC_{\text{EE}}) = E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i\cdot})^2\right]$$

desarrollando esta suma y aplicando las propiedades de la esperanza se obtiene:

$$E(CM_{\text{EE}}) = \frac{1}{n-a} (n-a) \sigma^2 = \sigma^2$$

Se resumen las Esperanzas de los Cuadrados Medios en el cuadro 2.

**Cuadro 2: Esperanza de los Cuadrados Medios para el Modelo de Efectos Fijos en el Diseño Completamente al Azar.**

F. de V.	g. de L.	E(CM)
Tratam.	a-1	$\sigma^2 + \frac{\sum_{i=1}^a \sum_{j=1}^{r_i} (\alpha_i)^2}{a-1}$
EE	n-a	$\sigma^2$
Total	n-1	

### **Modelo de Efectos Aleatorios (Modelo I I):**

Se dice que un *factor* es *aleatorio* si se selecciona aleatoriamente "a" niveles del factor. Se obtienen conclusiones de toda la población de niveles del factor, porque los niveles usados en el experimento fueron seleccionados al azar.

El Modelo de Efectos Aleatorios se define de igual forma al de efectos fijos, a diferencia de lo que se plantea a continuación.

#### *Supuestos del Modelo*

$$\alpha_i \sim \text{NIID}(0, \sigma_\alpha^2) \quad \varepsilon_{ij} \sim \text{NIID}(0, \sigma^2)$$

$\alpha_i$  y  $\varepsilon_{ij}$  son independientes

Si  $\alpha_i$  tiene una varianza  $\sigma_\alpha^2$  y es independiente de  $\varepsilon_{ij}$ , la varianza de cualquier observación es:

$$V(y_{ij}) = \sigma_\alpha^2 + \sigma^2$$

Las varianzas  $\sigma_\alpha^2$  y  $\sigma^2$  se conocen como componentes de varianza y el Modelo (III.1.1), se denomina Modelo de Componentes de Varianza o de Efectos Aleatorios.

$$\begin{aligned} E(\mu) &= \mu \quad \text{y} \quad E(\mu^2) = \mu^2; \\ E(\alpha_i) &= 0; \quad E(\alpha_i^2) = \sigma_\alpha^2 \quad \text{y} \quad E(\alpha_i \alpha_{i'}) = 0, \quad i \neq i'; \\ E(\varepsilon_{ij}) &= 0; \quad E(\varepsilon_{ij}^2) = \sigma^2 \quad \text{y} \quad E(\varepsilon_{ij} \varepsilon_{i'j'}) = 0, \quad i \neq i'; \end{aligned}$$

Se admite también que  $\text{Cov}(\alpha_i, \varepsilon_{ij}) = 0 \quad \forall i, j$

Las Hipótesis planteadas, para determinar los efectos principales de los  $i$ -ésimos tratamientos, se plantean de la siguiente forma, para el Modelo de Efectos Aleatorios:

$$H_0: \sigma_\alpha^2 = 0$$

vs

$$H_a: \sigma_\alpha^2 \neq 0$$

Entonces,

$$SC_{Total} = \sum_{i=1}^a \sum_{j=1}^{r_i} (\bar{y}_{i\cdot} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{..})^2$$

*Esperanza de los Cuadrados Medios (para mayor detalle ver Apéndice B):*

Para los tratamientos:

$$E(CM_{trat}) = E\left(\frac{SC_{trat}}{a-1}\right) = \frac{1}{a-1} E(SC_{trat})$$

$$E(CM_{trat}) = \frac{1}{a-1} E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} (\bar{y}_{i\cdot} - \bar{y}_{..})^2\right]$$

desarrollando esta suma y aplicando las propiedades de la esperanza y los supuestos para este modelo, se obtiene:

$$E(CM_{trat}) = \sigma^2 + \frac{\sum_{i=1}^a r_i - \left[\sum_{i=1}^a (r_i)^2 / \sum_{i=1}^a r_i\right]}{a-1} \sigma_\alpha^2 = \sigma^2 + k\sigma_\alpha^2$$

Donde  $k = \frac{\sum_{i=1}^a r_i - \left[ \frac{\sum_{i=1}^a (r_i)^2}{\sum_{i=1}^a r_i} \right]}{a-1}$

Para el Error:

$$E(CM_{EE}) = E\left(\frac{SC_{EE}}{n-a}\right) = \frac{1}{n-a} E(SC_{Trat})$$

$$E(SC_{EE}) = E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i\cdot})^2\right]$$

desarrollando esta suma y aplicando las propiedades de la esperanza y los supuestos para este modelo, se obtiene:

$$E(CM_{EE}) = \frac{1}{n-a} (n-a) \sigma^2 = \sigma^2$$

**Cuadro 3: Esperanza de los Cuadrados Medios para el Modelo de efectos aleatorios en el Diseño Completamente al Azar.**

F. de V.	g. de L.	E(CM)
Tratam.	a-1	$\sigma^2 + k\sigma_\alpha^2$
EE	n-a	$\sigma^2$
Total	n-1	

**Intervalos de Confianza:**

Tanto para el caso del Modelo de Efectos Fijos como para el Modelo de Efectos aleatorios; si se supone que los errores están normalmente distribuidos, las  $\bar{y}_{i\cdot}$  son NID  $(\bar{\alpha}, \sigma^2/n)$ . Entonces, podría usarse la distribución normal para definir el intervalo de confianza buscado, si se conoce  $\sigma^2$ . Al

usar el  $CM_{EE}$  como estimador de  $\sigma^2$ , el intervalo de confianza se debe basar en la distribución t-student. Por tanto, un intervalo de confianza, al nivel de confianza  $(1-\alpha)$  para la media del i-ésimo tratamiento  $\bar{\alpha}_i$ , es:

$$IC(\bar{\alpha}_i)_{1-\alpha} = \left( \bar{y}_{i\cdot} \pm t_{(\alpha/2, n-\alpha)} \sqrt{CM_{EE}/n} \right)$$

Un intervalo de confianza para la diferencia de las medias de dos tratamientos cualesquiera, por ejemplo  $\bar{\alpha}_i - \bar{\alpha}_{i'}$ ;  $i \neq i'$  es:

$$IC(\bar{\alpha}_i - \bar{\alpha}_{i'})_{1-\alpha} = \left( \bar{y}_{i\cdot} - \bar{y}_{i'\cdot} \pm t_{(\alpha/2, n-\alpha)} \sqrt{2CM_{EE}/n} \right)$$

## 2. DISEÑO DE BLOQUES AL AZAR (D.B.A.)

Según Cochran y Cox (1997), la esencia de este diseño estriba en que el material experimental se divide en grupos, cada uno de los cuales constituye una repetición (bloque).

Martínez (1988) acota que las unidades experimentales deben ser homogéneas dentro de cada bloque, salvo por variaciones aleatorias. Dos unidades experimentales de bloques distintos pueden exhibir heterogeneidad.

Chacín (2000) presenta las siguientes ventajas y desventajas :

**Ventajas:**

1. Permite controlar una fuente de variación extraña al experimento
2. Es un diseño muy flexible desde el punto de vista de su análisis.

**Desventajas:**

1. Cuando las condiciones son homogéneas los grados de libertad para el error experimental son menores que los del completamente aleatorizado y por consiguiente es menos precisa la estimación del error experimental.
2. Cuando se usa un gran número de tratamientos las variabilidades dentro del bloque se pueden elevar de tal manera que el diseño se hace poco eficiente.

**Modelo Lineal para el Diseño de Bloques al Azar**

Al igual que en el caso del D.C.A. se presenta a continuación un resumen para el Modelo Lineal Aditivo, según lo señalado por los mismos autores mencionados para el D.C.A. Este Modelo es usualmente referido como, modelo para dos vías de clasificación, o clasificación en dos sentidos:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, b; \quad k = 1, \dots, n \quad (\text{III.2.1})$$

Donde:  $y_{ijk}$  es la k-ésima observación obtenida con la aplicación del i-ésimo tratamiento en el j-ésimo bloque

$\mu$  es un efecto común a todas las unidades experimentales

$\alpha_i$  es el efecto del i-ésimo nivel del factor A o tratamiento



$\beta_j$  es el efecto del j-ésimo bloque

$\varepsilon_{ijk}$  es el efecto aleatorio de la k-ésima observación del i-ésimo

tratamiento en el j-ésimo bloque

### **Modelo de Efectos Fijos (Modelo I):**

Se aplica en este diseño, si los tratamientos y bloques son de efectos fijos.

Presenta dos restricciones:

$$\sum_{i=1}^a \alpha_i = 0$$

$$\sum_{j=1}^b \beta_j = 0$$

Además;

$$E(\varepsilon_{ij}) = 0; \quad E(\varepsilon_{ij}^2) = \sigma^2; \quad E(\varepsilon_{ij}\varepsilon_{i'j'}) = 0;$$

Para obtener los estimadores de los parámetros  $\mu, \alpha_1, \dots, \alpha_a, \beta_1, \dots, \beta_b$ ; se emplea el “Método de los Mínimos Cuadrados”:

$$Q(\mu, \alpha_1, \dots, \alpha_a, \beta_1, \dots, \beta_b) = \sum_{i=1}^a \sum_{j=1}^b (\hat{y}_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2$$

$$\frac{\partial Q}{\partial \mu} = -2 \sum_{i=1}^a \sum_{j=1}^b (\hat{y}_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j) = 0$$

$$\frac{\partial Q}{\partial \alpha_i} = -2 \sum_{j=1}^b (\hat{y}_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j) = 0 \quad i = 1, \dots, a$$

$$\frac{\partial Q}{\partial \beta_j} = -2 \sum_{i=1}^a (\hat{y}_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j) = 0 \quad j = 1, \dots, b$$

Desarrollando se generan las ecuaciones normales del modelo

$$\left\{ \begin{array}{l} n\hat{\mu} + b\sum_{i=1}^a \hat{\alpha}_i + a\sum_{j=1}^b \hat{\beta}_j = y_{..} \\ a\hat{\mu} + \sum_{i=1}^a \hat{\alpha}_i + a\hat{\beta}_j = y_{.j} \\ b\hat{\mu} + b\hat{\alpha}_i + \sum_{j=1}^b \hat{\beta}_j = y_{i.} \end{array} \right\}$$

Donde;

$$y_{..} = \sum_{i=1}^a \sum_{j=1}^b y_{ij} \quad \text{gran total}$$

$$y_{i.} = \sum_{j=1}^b y_{ij} \quad \text{total para el } i\text{--ésimo tratamiento}$$

$i = 1, 2, \dots, a$

$$y_{.j} = \sum_{i=1}^a y_{ij} \quad \text{total para el } j\text{--ésimo bloque}$$

$j = 1, 2, \dots, b$

Cabe resaltar que si se suman las últimas ecuaciones normales se obtiene la primera de ellas. Por lo que las ecuaciones normales no son independientes y no hay una solución única para  $\mu, \alpha_1, \dots, \alpha_a, \beta_1, \dots, \beta_b$ . Hay una gran variedad de métodos para solucionar este problema, una manera es emplear los contrastes que generan las restricciones, presentadas anteriormente.

En virtud de lo anterior, las soluciones de las ecuaciones normales son:

$$n\hat{\mu} + b\sum_{i=1}^a \hat{\alpha}_i + a\sum_{j=1}^b \hat{\beta}_j = y_{..} \Rightarrow n\hat{\mu} = y_{..} \Rightarrow \hat{\mu} = \frac{y_{..}}{n}$$

$$\hat{\mu} = \bar{y}_{..} \quad (\text{estimador para } \mu) \quad (\text{III.2.2})$$

$$a\hat{\mu} + \sum_{i=1}^a \hat{\alpha}_i + a\hat{\beta}_j = y_{\cdot j} \Rightarrow a\hat{\beta}_j = y_{\cdot j} - a\hat{\mu} \Rightarrow \hat{\beta}_j = \frac{y_{\cdot j} - a\hat{\mu}}{a}$$

$$\hat{\beta}_j = \bar{y}_{\cdot j} - \bar{y}_{..} \quad (\text{estimador para } \beta_j) \quad j = 1, \dots, b \quad (\text{III.2.3})$$

$$b\hat{\mu} + b\hat{\alpha}_i + \sum_{j=1}^b \hat{\beta}_j = y_{i\cdot} \Rightarrow b\hat{\alpha}_i = y_{i\cdot} - b\hat{\mu} \Rightarrow \hat{\alpha}_i = \frac{y_{i\cdot} - b\hat{\mu}}{b}$$

$$\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}_{..} \quad (\text{estimador para } \alpha_i) \quad i = 1, \dots, a \quad (\text{III.2.4})$$

$$\hat{\varepsilon}_{ij} = y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j \Rightarrow \hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_{..} - (\bar{y}_{i\cdot} - \bar{y}_{..}) - (\bar{y}_{\cdot j} - \bar{y}_{..})$$

$$\hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{..} \quad (\text{estimador para el error}) \quad (\text{III.2.5})$$

Por tanto el vector

$$\hat{\beta}_{px1} = \begin{bmatrix} \bar{y}_{..} \\ \bar{y}_{1\cdot} - \bar{y}_{..} \\ \cdot \\ \cdot \\ \cdot \\ \bar{y}_{a\cdot} - \bar{y}_{..} \\ \bar{y}_{\cdot 1} - \bar{y}_{..} \\ \cdot \\ \cdot \\ \cdot \\ \bar{y}_{\cdot b} - \bar{y}_{..} \end{bmatrix}$$

### **Pruebas de Hipótesis:**

Una vez obtenidos los estimadores de los parámetros del modelo, se procede a realizar la Prueba de Hipótesis de interés entre los efectos principales, las cuales quedan planteadas de la siguiente forma:

Tratamientos:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a$$

vs

$$H_a : \alpha_i \neq \alpha_{i'}, \text{ para algún } i \neq i'$$

Bloques:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_a$$

vs

$$H_a : \beta_j \neq \beta_{j'}, \text{ para algún } j \neq j'$$

*Análisis de Varianza:*

A fin de realizar las pruebas se aplicará el Análisis de Varianza (ANAVAR), el cual consiste en dividir la  $SC_{Total}$  en  $SC_{Tratamientos}$ ,  $SC_{Bloques}$  y  $SC_{EE}$ . Para el caso que se presenta (efectos fijos), se tiene:

$$y_{ij} - \bar{y}_{..} = (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$$

Si se elevan al cuadrado todos los términos de la ecuación anterior y se suman en "i" y en "j" se obtiene:

$$\sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

Esto se expresa como:

$$SC_{total} = SC_{Trat} + SC_{Bloq} + SC_{EE}$$

donde;

$SC_{\text{trat}}$ : es la suma de cuadrados debida a los tratamientos, es decir, entre tratamientos. Tiene  $a-1$  grados de libertad, porque existe  $a$  niveles de tratamientos.

$SC_{\text{bloq}}$ : es la suma de cuadrados debida a los bloques. Tiene  $b-1$  grados de libertad, porque existe  $b$  bloques.

$SC_{\text{Total}}$ : es la suma de cuadrados debida a la variabilidad total. Tiene  $n-1$  grados de libertad porque hay un total de  $n$  observaciones.

$SC_{\text{EE}}$ : es la suma de cuadrados debida al error, es decir, dentro de tratamientos. Tiene  $(a-1)(b-1)$  grados de libertad.

Resultando los Cuadrados Medios de la forma:

$$CM_{\text{EE}} = \frac{SC_{\text{EE}}}{(a-1)(b-1)} \quad CM_{\text{trat}} = \frac{SC_{\text{trat}}}{a-1}$$

Por el Teorema 2 (pág. 8):

$$F_c = \frac{CM_{\text{trat}}}{CM_{\text{EE}}} \leq F_{1-\alpha; a-1, (a-1)(b-1)} \quad \text{y} \quad F_c = \frac{CM_{\text{bloq}}}{CM_{\text{EE}}} \leq F_{1-\alpha; b-1, (a-1)(b-1)}$$

De lo anteriormente planteado expuesto el respectivo Análisis de Varianza se plantea en el cuadro 4

**Cuadro 4: Análisis de Varianza para el Diseño de Bloques al Azar (D.B.A.)**

F.deV.	g.de L.	SC	CM	F
Trat.	a-1	$\sum_{i=1}^a \frac{(y_{i.})^2}{b} - \frac{(y_{..})^2}{n}$	$\frac{SC_{\text{Trat}}}{a-1}$	$\frac{CM_{\text{trat}}}{CM_{EE}}$
Bloq	b-1	$\sum_{j=1}^b \frac{(y_{.j})^2}{a} - \frac{(y_{..})^2}{n}$	$\frac{SC_{\text{Bloq}}}{b-1}$	$\frac{CM_{\text{Bloq}}}{CM_{EE}}$
EE	(a-1)(b-1)	$\sum_{i=1}^a \sum_{j=1}^b (y_{ij})^2 - \sum_{i=1}^a \frac{(y_{i.})^2}{b} - \sum_{j=1}^b \frac{(y_{.j})^2}{a} + \frac{(y_{..})^2}{n}$	$\frac{SC_{EE}}{(a-1)(b-1)}$	
Total	n-1	$\sum_{i=1}^a \sum_{j=1}^b (y_{ij})^2 - \frac{(y_{..})^2}{n}$		

*Esperanza de los Cuadrados Medios:*

Para los tratamientos:

$$E(CM_{\text{trat}}) = E\left(\frac{SC_{\text{Trat}}}{a-1}\right) = \frac{1}{a-1} E(SC_{\text{Trat}})$$

$$E(CM_{\text{trat}}) = \frac{1}{a-1} E\left[\sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{i.} - \bar{y}_{..})^2\right]$$

desarrollando esta suma y aplicando las propiedades de la esperanza se obtiene:

$$E(CM_{\text{trat}}) = \sigma^2 + \frac{b \sum_{i=1}^a (\alpha_i)^2}{a-1}$$

Para los bloques:

$$E(CM_{\text{Bloq}}) = E\left(\frac{SC_{\text{Bloq}}}{b-1}\right) = \frac{1}{b-1} E(SC_{\text{Bloq}})$$

$$E(CM_{\text{Bloq}}) = \frac{1}{b-1} E\left[\sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{\cdot j} - \bar{y}_{..})^2\right]$$

desarrollando esta suma y aplicando las propiedades de la esperanza se obtiene:

$$E(CM_{\text{Bloq}}) = \sigma^2 + \frac{\alpha \sum_{j=1}^b (\beta_j)^2}{b-1}$$

Para el error:

$$E(CM_{\text{EE}}) = E\left(\frac{SC_{\text{EE}}}{(a-1)(b-1)}\right) = \frac{1}{(a-1)(b-1)} E(SC_{\text{EE}})$$

$$E(SC_{\text{EE}}) = E\left[\sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{..})^2\right]$$

desarrollando esta suma y aplicando las propiedades de la esperanza se obtiene:

$$E(CM_{\text{EE}}) = \sigma^2$$

Se muestran en el cuadro 5, las esperanzas de los cuadrados medios obtenidas

**Cuadro 5: Esperanza de los Cuadrados Medios para el Modelo de Efectos Fijos en el Diseño de Bloques al Azar**

F. de V.	g. de L.	E(CM)
Tratam.	a-1	$\sigma^2 + \frac{b \sum_{i=1}^a (\alpha_i)^2}{a-1}$
Bloques	b-1	$\sigma^2 + \frac{a \sum_{j=1}^b (\beta_j)^2}{b-1}$
EE	(a-1)(b-1)	$\sigma^2$
Total	n-1	

**Modelo de Efectos Aleatorios (Modelo I I):**

El Modelo de Efectos Aleatorios se define de igual forma al de efectos fijos, a diferencia de lo que se plantea a continuación.

*Supuestos del Modelo*

$$\alpha_i \sim \text{NIID} (0, \sigma_\alpha^2)$$

$$\beta_j \sim \text{NIID} (0, \sigma_\beta^2)$$

$$\varepsilon_{ij} \sim \text{NIID} (0, \sigma^2)$$

$\alpha_i, \beta_j$  y  $\varepsilon_{ij}$  son independientes

$$V(y_{ij}) = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2$$

Las varianzas  $\sigma_\alpha^2$  y  $\sigma^2$  se desconocen como componentes de varianza y el Modelo (III.2.1), se denomina Modelo de Componentes de Varianza o de Efectos Aleatorios.



$$\begin{aligned}
E(\mu) &= \mu \quad \text{y} \quad E(\mu^2) = \mu^2; \\
E(\alpha_i) &= 0; \quad E(\alpha_i^2) = \sigma_\alpha^2 \quad \text{y} \quad E(\alpha_i \alpha_{i'}) = 0, i \neq i'; \\
E(\beta_j) &= 0; \quad E(\beta_j^2) = \sigma_\beta^2 \quad \text{y} \quad E(\beta_j \beta_{j'}) = 0, j \neq j'; \\
E(\varepsilon_{ij}) &= 0; \quad E(\varepsilon_{ij}^2) = \sigma^2 \quad \text{y} \quad E(\varepsilon_{ij} \varepsilon_{i'j'}) = 0, i \neq i';
\end{aligned}$$

$$\text{Cov}(\alpha_i, \beta_j) = 0 \quad \forall i, j$$

Se admite también que  $\text{Cov}(\alpha_i, \varepsilon_{ij}) = 0 \quad \forall i, j$

$$\text{Cov}(\beta_j, \varepsilon_{ij}) = 0 \quad \forall i, j$$

### ***Pruebas de Hipótesis:***

Las Hipótesis para determinar los efectos principales de los  $i$ -ésimos tratamientos, y los  $j$ -ésimos bloques, se plantean de la siguiente forma, para el Modelo de Efectos Aleatorios:

Tratamientos:

$$H_0: \sigma_\alpha^2 = 0$$

vs

$$H_a: \sigma_\alpha^2 \neq 0$$

Bloques:

$$H_0: \sigma_\beta^2 = 0$$

vs

$$H_a: \sigma_\beta^2 \neq 0$$

Entonces,

$$SC_{\text{Total}} = \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{i\cdot} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{\cdot j} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{..})^2$$

*Esperanza de los Cuadrados Medios:*

*Para los tratamientos:*

$$E(CM_{\text{trat}}) = E\left(\frac{SC_{\text{trat}}}{a-1}\right) = \frac{1}{a-1}E(SC_{\text{trat}})$$

$$E(CM_{\text{trat}}) = \frac{1}{a-1}E\left[\sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{i\cdot} - \bar{y}_{..})^2\right]$$

desarrollando esta suma y aplicando las propiedades de la esperanza y los supuestos para este modelo, se obtiene:

$$E(CM_{\text{trat}}) = \sigma^2 + b\sigma_{\alpha}^2$$

*Para los bloques:*

$$E(CM_{\text{bloq}}) = E\left(\frac{SC_{\text{bloq}}}{b-1}\right) = \frac{1}{b-1}E(SC_{\text{bloq}})$$

$$E(CM_{\text{bloq}}) = \frac{1}{b-1}E\left[\sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{\cdot j} - \bar{y}_{..})^2\right]$$

desarrollando esta suma y aplicando las propiedades de la esperanza y los supuestos para este modelo, se obtiene:

$$E(CM_{\text{bloq}}) = \sigma^2 + a\sigma_{\beta}^2$$

*Para el Error:*

$$E(CM_{\text{EE}}) = E\left(\frac{SC_{\text{EE}}}{(a-1)(b-1)}\right) = \frac{1}{(a-1)(b-1)}E(SC_{\text{trat}})$$

$$E(SC_{EE}) = E \left[ \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2 \right]$$

desarrollando esta suma y aplicando las propiedades de la esperanza y los supuestos para este modelo, se obtiene:

$$E(CM_{EE}) = \sigma^2$$

**Cuadro 6: Esperanza de los Cuadrados Medios para el Modelo de Efectos Aleatorios en el Diseño de Bloques al Azar**

F. de V.	g. de L.	E(CM)
Tratam.	a-1	$\sigma^2 + b\sigma_{\alpha}^2$
Bloques	b-1	$\sigma^2 + a\sigma_{\beta}^2$
EE	(a-1)(b-1)	$\sigma^2$
Total	n-1	

**Modelo de Efectos Mixtos (Modelo III):**

El Modelo de Efectos Mixtos se obtiene con las combinaciones según el caso.

## **CAPÍTULO IV**

### **MODELO PARA EL MUESTREO ESTRATIFICADO DOBLE**

Para el Muestreo Estratificado Doble se presenta el Modelo (IV.1.1); el cual será un Modelo de efectos Fijos, si ambas variables de estratificación se consideran fijas, un Modelo Aleatorio si son variables aleatorias y un Modelo Mixto si alguna de las variables de estratificación es fija y la otra aleatoria, como es el caso de los datos evaluados en este trabajo. Este Modelo se desarrollará para el caso en el cual se considere el supuesto de la presencia de interacción; debido a que en ausencia de la misma se comportará similar al Modelo del Diseño de Bloques al Azar (III.2.1)

Hocking (1973) presenta tres Modelos Mixtos de dos factores; considerando el Modelo I, descrito por Scheffé (1956a y 1959, citados por Hocking, 1973) como base de comparación en su artículo. Modelos II y III, relacionados al primero y que representan casos especiales del mismo.

Como lo señala Montgomery (1991), por la diversidad de Modelos Mixtos, al emplearlos ocurre la pregunta de ¿Cuál usar?. El autor indica que generalmente se emplea un Modelo que denomina “Estándar” que es el que

mas aparece en la literatura. Sin embargo; acota que cualquier modelo es apropiado si la correlación entre los componentes aleatorios no es grande.

Agrega que si la correlación entre los datos es grande, se puede emplear el Modelo de Scheffé. Considera que la elección entre los modelos debe ser dictada por los datos.

Para efectos de este trabajo, se desarrolla un Modelo Mixto desbalanceado, basándose en un Modelo Alterno presentado por Montgomery (1991)

## 1. PLANTEAMIENTO DEL MODELO

$$y_{ijk} = \mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk} \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, b \quad k = 1, 2, \dots, N_{ij} \quad (\text{IV.1.1})$$

Donde:  $y_{ijk}$  es la k-ésima observación en el i-ésimo nivel del estrato T

(tratamiento) y el j-ésimo nivel del estrato B (bloque)

$\mu$  es un efecto común a todas las unidades experimentales

$\alpha_i$  es el efecto del i-ésimo nivel del estrato T (tratamientos)

$\beta_j$  es el efecto del j-ésimo nivel del estrato B (bloques)

$\rho_{ij}$  es el efecto de la interacción del i-ésimo nivel del estrato T y el j-ésimo nivel del estrato B

$\varepsilon_{ijk}$  es el efecto aleatorio del i-ésimo nivel del estrato T y el j-ésimo nivel del estrato B, en la k-ésima observación

$N_{ij}$  es el número de observaciones en el estrato formado por el  $i$ -ésimo nivel del estrato T y el  $j$ -ésimo nivel del estrato B

$N$  es el número total de observaciones  $N = \sum_{i=1}^a \sum_{j=1}^b N_{ij}$

Para el estudio del modelo, se definen los siguientes parámetros:

*Medias poblacionales:*

$$\text{Media general: } \bar{Y} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk}}{N} \quad (\text{IV.1.2})$$

$$\text{Media para la interacción: } \bar{Y}_{ij\cdot} = \frac{\sum_{k=1}^{N_{ij}} Y_{ijk}}{N_{ij}} \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, b \quad (\text{IV.1.3})$$

$$\text{Media para el estrato T: } \bar{Y}_{i..} = \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk}}{\sum_{j=1}^b N_{ij}} \quad i = 1, 2, \dots, a \quad (\text{IV.1.4})$$

$$\text{Media para el estrato B: } \bar{Y}_{\cdot j} = \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} Y_{ijk}}{\sum_{i=1}^a N_{ij}} \quad j = 1, 2, \dots, b \quad (\text{IV.1.5})$$

*Totales Poblacionales:*

$$\text{Gran total : } \tau = N\bar{Y} \quad \tau = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk} = Y... \quad (\text{IV.1.6})$$

$$\text{Total para la interacción: } \begin{cases} \tau_{ij\cdot} = N_{ij} \bar{Y}_{ij\cdot} \\ \tau_{ij\cdot} = \sum_{k=1}^{N_{ij}} Y_{ijk} = Y_{ij\cdot} \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, b \end{cases} \quad (\text{IV.1.7})$$

$$\text{Total para el estrato T: } \begin{cases} \tau_{i..} = \sum_{j=1}^b N_{ij} \bar{Y}_{i..} \\ \tau_{i..} = \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk} = Y_{i..} \quad i = 1, 2, \dots, a \end{cases} \quad (\text{IV.1.8})$$

$$\text{Total para el estrato B: } \begin{cases} \tau_{.j.} = \sum_{i=1}^a N_{ij} \bar{Y}_{.j.} \\ \tau_{.j.} = \sum_{i=1}^a \sum_{k=1}^{N_{ij}} Y_{ijk} = Y_{.j.} \quad j = 1, 2, \dots, b \end{cases} \quad (\text{IV.1.9})$$

Considerando el estrato T, como un efecto fijo y el estrato B, como un efecto aleatorio; se supone que:

$$\begin{aligned} E(\mu) &= \mu & y & \quad E(\mu^2) = \mu^2 \\ E(\alpha_i) &= \alpha_i & y & \quad E(\alpha_i^2) = \alpha_i^2 \\ E(\beta_j) &= 0 \quad , \quad E(\beta_j^2) = \sigma_\beta^2 & y & \quad E(\beta_j \beta_{j'}) = 0 \quad j \neq j' \\ E(\rho_{ij}) &= 0 \quad , \quad E(\rho_{ij}^2) = \sigma_\rho^2 & y & \quad E(\rho_{ij} \rho_{i'j'}) = 0 \quad i \neq i' \quad j \neq j' \\ E(\varepsilon_{ijk}) &= 0 \quad , \quad E(\varepsilon_{ijk}^2) = \sigma^2 & y & \quad E(\varepsilon_{ijk} \varepsilon_{i'j'k'}) = 0 \quad i \neq i' \quad j \neq j' \\ \text{Cov}(\alpha_i, \beta_j) &= 0 \quad , \quad \text{Cov}(\alpha_i, \rho_{ij}) = 0 \quad , \quad \text{Cov}(\alpha_i, \varepsilon_{ijk}) = 0 \\ \text{Cov}(\beta_j, \rho_{ij}) &= 0 \quad , \quad \text{Cov}(\beta_j, \varepsilon_{ijk}) = 0 \quad , \quad \text{Cov}(\rho_{ij}, \varepsilon_{ijk}) = 0 \end{aligned}$$

## 2. ESTIMACIÓN DE LOS PARÁMETROS UTILIZANDO EL MÉTODO DE LOS MÍNIMOS CUADRADOS

A partir del Modelo

$$Y = X\beta + \varepsilon$$

Para obtener los estimadores de los parámetros  $\bar{Y}$ ,  $\bar{Y}_{ij}$ ,  $\bar{Y}_{i..}$ ,  $\bar{Y}_{.j.}$ ,  $\tau$ ,  $\tau_{ij..}$ ,  $\tau_{i..}$  y  $\tau_{.j.}$ ; se empleó el "Método de los Mínimos Cuadrados" generándose las ecuaciones normales para el modelo:

$$(X'X)\hat{\beta} = X'Y$$

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk} &= N\hat{\mu} + \sum_{i=1}^a \sum_{j=1}^b N_{ij}\hat{\alpha}_i + \sum_{i=1}^a \sum_{j=1}^b N_{ij}\hat{\beta}_j + \sum_{i=1}^a \sum_{j=1}^b N_{ij}\hat{\rho}_{ij} \\ \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk} &= \sum_{j=1}^b N_{ij}\hat{\mu} + \sum_{j=1}^b N_{ij}\hat{\alpha}_i + \sum_{j=1}^b N_{ij}\hat{\beta}_j + \sum_{j=1}^b N_{ij}\hat{\rho}_{ij} \quad i = 1, \dots, a \\ \sum_{i=1}^a \sum_{k=1}^{N_{ij}} Y_{ijk} &= \sum_{i=1}^a N_{ij}\hat{\mu} + \sum_{i=1}^a N_{ij}\hat{\alpha}_i + \sum_{i=1}^a N_{ij}\hat{\beta}_j + \sum_{i=1}^a N_{ij}\hat{\rho}_{ij} \quad j = 1, \dots, b \\ \sum_{k=1}^{N_{ij}} Y_{ijk} &= N_{ij}\hat{\mu} + N_{ij}\hat{\alpha}_i + N_{ij}\hat{\beta}_j + N_{ij}\hat{\rho}_{ij} \quad i = 1, \dots, a \quad j = 1, \dots, b \end{aligned}$$

A través del principio de los contrastes (definición 4, pág. 8), se generaron las siguientes restricciones, para hallar las soluciones de las ecuaciones normales:

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij}\hat{\alpha}_i = 0 \quad \sum_{i=1}^a N_{ij}\hat{\alpha}_i = 0$$

De donde, las soluciones de las ecuaciones normales obtenidas son:

$$\hat{\mu} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} y_{ijk}}{N} - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}\hat{\beta}_j}{N} - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}\hat{\rho}_{ij}}{N} = \bar{y} \dots - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}\hat{\beta}_j}{N} - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}\hat{\rho}_{ij}}{N} \quad (IV.2.1)$$

(estimador para  $\mu$ )



$$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...} + \frac{\sum_{j=1}^b \sum_{k=1}^a N_{ijk} \hat{\beta}_j}{N} + \frac{\sum_{j=1}^b \sum_{k=1}^a N_{ijk} \hat{\rho}_{ij}}{N} - \frac{\sum_{j=1}^b N_{ij.} \hat{\beta}_j}{\sum_{j=1}^b N_{ij.}} - \frac{\sum_{j=1}^b N_{ijk} \hat{\rho}_{ij}}{\sum_{j=1}^b N_{ij.}} \quad (\text{IV.2.2})$$

(estimador para  $\alpha_i$ )  $i = 1, \dots, a$

$$\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...} + \frac{\sum_{i=1}^a \sum_{k=1}^a N_{ijk} \hat{\beta}_j}{N} + \frac{\sum_{i=1}^a \sum_{k=1}^a N_{ijk} \hat{\rho}_{ij}}{N} - \frac{\sum_{i=1}^a N_{ij.} \hat{\rho}_{ij}}{\sum_{i=1}^a N_{ij.}} \quad (\text{IV.2.3})$$

(estimador para  $\beta_j$ )  $j = 1, \dots, b$

$$\hat{\rho}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} - \frac{\sum_{k=1}^a \sum_{l=1}^b N_{ijkl} \hat{\beta}_j}{N} + \frac{\sum_{k=1}^a N_{ijk} \hat{\beta}_j}{\sum_{k=1}^a N_{ijk}} + \frac{\sum_{k=1}^a N_{ijk} \hat{\rho}_{ij}}{\sum_{k=1}^a N_{ijk}} + \frac{\sum_{k=1}^a N_{ijk} \hat{\rho}_{ij}}{\sum_{k=1}^a N_{ijk}} - \frac{\sum_{k=1}^a \sum_{l=1}^b N_{ijkl} \hat{\rho}_{ij}}{N} \quad (\text{IV.2.4})$$

(estimador para  $\rho_{ij}$ )  $i = 1, \dots, a \quad j = 1, \dots, b$

$$\hat{\epsilon}_{ijk} = y_{ijk} - \bar{y}_{ij.} \quad (\text{estimador para el error}) \quad (\text{IV.2.5})$$

### 3. PRUEBAS DE HIPÓTESIS

Una vez obtenidos los estimadores de los parámetros del modelo, se procede a realizar la Prueba de Hipótesis de interés entre los efectos principales, las cuales quedan planteadas de la siguiente forma:

Estrato T:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_a$$

vs

$$H_a: \alpha_i \neq \alpha_{i'} \text{ para algún } i \neq i'$$

No hay diferencias entre los niveles del estrato T. En el Diseño de Bloques al Azar se plantea si hay diferencias entre tratamientos

Estrato B:

$$H_0: \sigma_\beta^2 = 0$$

vs

$$H_a: \sigma_\beta^2 \neq 0$$

Interacción:

$$H_0: \sigma_\rho^2 = 0$$

vs

$$H_a: \sigma_\rho^2 \neq 0$$

#### **4. ANÁLISIS DE VARIANZA**

A fin de realizar las pruebas de hipótesis, se aplicó el Análisis de Varianza (ANAVAR)

$$Y_{ijk} - \bar{y}_{...} = (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (y_{ijk} - \bar{y}_{ij.})$$

Donde la descomposición de la Suma de Cuadrados total, queda de la forma:

$$SC_{total} = SC_{estratoT} + SC_{estratoB} + SC_{interacción} + SC_{EE}$$

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (y_{ijk} - \bar{y} \dots)^2 &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{y}_{i..} - \bar{y} \dots)^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{y}_{.j.} - \bar{y} \dots)^2 \\ &+ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots)^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (y_{ijk} - \bar{y}_{ij.})^2 \end{aligned}$$

donde;

$SC_{estratoT}$ : es la suma de cuadrados debida al estrato T. Tiene  $a-1$  grados de libertad, porque tiene a niveles.

$SC_{estratoB}$ : es la suma de cuadrados debida al estrato B. Tiene  $b-1$  grados de libertad, porque tiene b niveles.

$SC_{interacc}$ : es la suma de cuadrados debida a la interacción entre los estratos. Tiene  $(a-1)(b-1)$  grados de libertad.

$SC_{Total}$ : es la suma de cuadrados debida a la variabilidad total. Tiene  $N-1$  grados de libertad porque hay un total de N observaciones.

$SC_{EE}$ : es la suma de cuadrados debida al error. Tiene  $N-ab$  grados de libertad.

Resultando los Cuadrados Medios de la forma:

$$CM_{EE} = \frac{SC_{EE}}{N-ab}$$

$$CM_{estratoT} = \frac{SC_{estratoT}}{a-1}$$

$$CM_{\text{estratoB}} = \frac{SC_{\text{estratoB}}}{b-1}$$

$$CM_{\text{int}} = \frac{SC_{\text{int}}}{(a-1)(b-1)}$$

## 5. ESPERANZA DE LOS CUADRADOS MEDIOS:

Para el estrato T:

$$E(CM_{\text{estratoT}}) = E\left(\frac{SC_{\text{est.T}}}{a-1}\right) = \frac{1}{a-1}E(SC_{\text{est.T}})$$

$$E(CM_{\text{est.T}}) = \frac{1}{a-1}E\left[\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{y}_{i..} - \bar{y}_{...})^2\right]$$

Donde

$$\begin{aligned}\bar{y}_{i..} &= \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk}}{\sum_{j=1}^b N_{ij}} = \frac{1}{\sum_{j=1}^b N_{ij}} \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk}) \\ \bar{y}_{i..} &= \frac{1}{\sum_{j=1}^b N_{ij}} \left( \mu \sum_{j=1}^b N_{ij} + \alpha_i \sum_{j=1}^b N_{ij} + \sum_{j=1}^b N_{ij} \beta_j + \sum_{j=1}^b N_{ij} \rho_{ij} + \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right) \\ \bar{y}_{i..} &= \mu + \alpha_i + \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} + \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} + \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}}\end{aligned}$$

$$\bar{y}_{...} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk}}{N} = \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk})$$

$$\bar{y}_{...} = \frac{1}{N} \left( \mu \sum_{i=1}^a \sum_{j=1}^b N_{ij} + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij} + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right)$$

Dada la restricción  $\sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i = 0$

$$\bar{y}_{...} = \mu + \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} + \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} + \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N}$$

$$(\bar{Y}_{i..} - \bar{y}_{...}) = \mu + \alpha_i + \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} + \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} + \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} - \mu - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} - \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N}$$

Elevando al cuadrado y desarrollando, se obtiene:

$$\begin{aligned} (\bar{Y}_{i..} - \bar{y}_{...})^2 = & \alpha_i^2 + \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right)^2 + \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right)^2 + \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right)^2 + \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right)^2 \\ & + \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right)^2 + \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)^2 + 2\alpha_i \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) + 2\alpha_i \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) \\ & + 2\alpha_i \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) - 2\alpha_i \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2\alpha_i \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2\alpha_i \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) \end{aligned}$$

$$\begin{aligned}
& + 2 \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) + 2 \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) - 2 \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \\
& - 2 \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2 \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) + 2 \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \\
& - 2 \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2 \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) \\
& - 2 \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2 \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) \\
& + 2 \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) + 2 \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) + 2 \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)
\end{aligned}$$

Aplicando la triple sumatoria, se obtiene:

$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{Y}_i \dots - \bar{Y} \dots)^2 &= \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right)^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right)^2 \\
&+ \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right)^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right)^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right)^2
\end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)^2 + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) \\
& + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \\
& - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) \\
& + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \\
& - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \\
& - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \\
& - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \\
& - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)
\end{aligned}$$

$$\begin{aligned}
& -2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \\
& - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \\
& + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)
\end{aligned}$$

Aplicando las propiedades de la esperanza, los supuestos y restricciones para este modelo, se obtiene:

$$E \left( \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i^2 \right) = \sum_{i=1}^a \sum_{j=1}^b N_{ij} E(\alpha_i^2) = \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i^2$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right)^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\left( \sum_{j=1}^b N_{ij} \right)^2} E \left( \sum_{j=1}^b N_{ij} \beta_j \right)^2 = \sum_{i=1}^a \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \sigma_{\beta}^2$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right)^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\left( \sum_{j=1}^b N_{ij} \right)^2} E \left( \sum_{j=1}^b N_{ij} \rho_{ij} \right)^2 = \sum_{i=1}^a \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \sigma_{\rho}^2$$



$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right)^2 &= \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\left( \sum_{j=1}^b N_{ij} \right)^2} E \left( \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right)^2 = \sum_{i=1}^a \frac{1}{\sum_{j=1}^b N_{ij}} \sum_{j=1}^b \sum_{k=1}^{N_{ij}} E(\varepsilon_{ijk}^2) \\
&= \sum_{i=1}^a \frac{1}{\sum_{j=1}^b N_{ij}} \sum_{j=1}^b N_{ij} \sigma^2 = \alpha \sigma^2
\end{aligned}$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right)^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{(N)^2} \sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2 \sigma_\beta^2 = \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \sigma_\beta^2$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right)^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{(N)^2} \sum_{i=1}^a \sum_{j=1}^b N_{ij}^2 \sigma_\rho^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2$$

$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)^2 &= \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{(N)^2} E \left( \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right)^2 = \frac{1}{N} E \left( \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}^2 \right) \\
&= \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} E(\varepsilon_{ijk}^2) = \sigma^2
\end{aligned}$$

$$\begin{aligned}
& \sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\sum_{j=1}^b N_{ij} \cdot N} E \left( \sum_{j=1}^b N_{ij} \beta_j \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \right) \\
& = \frac{1}{N} E \left[ \sum_{i=1}^a \left( \sum_{j=1}^b N_{ij} \beta_j \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \right) \right] = \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \sigma_\beta^2
\end{aligned}$$

$$\begin{aligned}
& \sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\sum_{j=1}^b N_{ij} \cdot N} E \left( \sum_{j=1}^b N_{ij} \rho_{ij} \sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij} \right) \\
& = \frac{1}{N} E \left[ \sum_{i=1}^a \left( \sum_{j=1}^b N_{ij} \rho_{ij} \sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij} \right) \right] = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2
\end{aligned}$$

$$\begin{aligned}
& \sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\sum_{j=1}^b N_{ij} \cdot N} E \left( \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right) \\
& = \frac{1}{N} E \left[ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right] = \frac{1}{N} E \left( \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right)^2 \\
& = \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} E(\varepsilon_{ijk}^2) = \sigma^2
\end{aligned}$$

$$\begin{aligned}
E \left[ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{Y}_{i..} - \bar{Y}_{...})^2 \right] &= \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i^2 + \sum_{i=1}^a \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \sigma_\beta^2 + \sum_{i=1}^a \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \sigma_\rho^2 + a \sigma^2 \\
&+ \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \sigma_\beta^2 + \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2 + \sigma^2 - 2 \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \sigma_\beta^2 \\
&- 2 \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2 - 2 \sigma^2
\end{aligned}$$

$$E\left[\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{Y}_{i..} - \bar{Y}_{...})^2\right] = \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i^2 + \sum_{i=1}^a \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \sigma_\beta^2 + \sum_{i=1}^a \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \sigma_\rho^2 + (a-1) \sigma^2$$

$$- \frac{\sum_{j=1}^b \left(\sum_{i=1}^a N_{ij}\right)^2}{N} \sigma_\beta^2 - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2$$

Sustituyendo en  $E(CM_{\text{est. T}})$ , se obtiene:

$$E(CM_{\text{est. T}}) = \sigma^2 + \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \right] \frac{\sigma_\rho^2}{a-1} + \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \frac{\sum_{j=1}^b \left(\sum_{i=1}^a N_{ij}\right)^2}{N} \right] \frac{\sigma_\beta^2}{a-1} \quad (\text{IV.5.1})$$

$$+ \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i^2}{a-1}$$

Como puede observarse en la ecuación IV.5.1, cuando se toma un tamaño de muestra desigual en cada estrato formado por las variables de estratificación T y B, la esperanza de los cuadrados medios para el estrato de efectos fijos, está afectada por el efecto aleatorio de bloques e interacción. Pero si la muestra es igual en cada estrato, digamos  $N_{ij} = M$ , se obtiene:

$$E(CM_{\text{est. T}}) = \sigma^2 + M \sigma_\rho^2 + \frac{bM \sum_{i=1}^a \alpha_i^2}{a-1} \quad (\text{IV.5.2})$$

Para el estrato B:

$$E(CM_{\text{estrato B}}) = E\left(\frac{SC_{\text{est. B}}}{b-1}\right) = \frac{1}{b-1} E(SC_{\text{est. B}})$$

$$E(CM_{\text{est.B}}) = \frac{1}{b-1} E \left[ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{y}_{\cdot j} - \bar{y} \dots)^2 \right]$$

Donde:

$$\begin{aligned} \bar{y}_{\cdot j} &= \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} Y_{ijk}}{\sum_{i=1}^a N_{ij}} = \frac{1}{\sum_{i=1}^a N_{ij}} \sum_{i=1}^a \sum_{k=1}^{N_{ij}} (\mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk}) \\ \bar{y}_{\cdot j} &= \frac{1}{\sum_{i=1}^a N_{ij}} \left( \mu \sum_{i=1}^a N_{ij} + \sum_{i=1}^a N_{ij} \alpha_i + \beta_j \sum_{i=1}^a N_{ij} + \sum_{i=1}^a N_{ij} \rho_{ij} + \sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right) \\ \bar{y}_{\cdot j} &= \mu + \beta_j + \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} + \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \end{aligned}$$

$$\begin{aligned} \bar{y} \dots &= \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk}}{N} = \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk}) \\ \bar{y} \dots &= \frac{1}{N} \left( \mu \sum_{i=1}^a \sum_{j=1}^b N_{ij} + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij} + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right) \end{aligned}$$

Dada la restricción  $\sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i = 0$

$$\begin{aligned} \bar{y} \dots &= \mu + \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} + \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} + \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \\ (\bar{y}_{\cdot j} - \bar{y} \dots) &= \mu + \beta_j + \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} + \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} - \mu - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} - \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \end{aligned}$$

Elevando al cuadrado y desarrollando, se obtiene:

$$\begin{aligned}
 (\bar{Y}_{i..} - \bar{Y}...) ^2 &= \beta_j^2 + \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right)^2 + \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right)^2 + \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) + \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right)^2 \\
 &+ \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)^2 + 2\beta_j \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) + 2\beta_j \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) - 2\beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \\
 &- 2\beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2\beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) + 2 \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) \\
 &- 2 \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2 \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) \\
 &- 2 \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2 \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) \\
 &+ 2 \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) + 2 \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) \\
 &+ 2 \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)
 \end{aligned}$$

Aplicando la triple sumatoria, se obtiene:

$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{Y}_i \dots - \bar{Y} \dots)^2 &= \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right)^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right)^2 \\
&+ \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right)^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)^2 \\
&+ 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \\
&- 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) \\
&+ 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \\
&- 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)
\end{aligned}$$

$$\begin{aligned}
& -2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \\
& - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \\
& + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)
\end{aligned}$$

Aplicando las propiedades de la esperanza, los supuestos y restricciones para este modelo, se obtiene:

$$E \left( \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j^2 \right) = \sum_{i=1}^a \sum_{j=1}^b N_{ij} E(\beta_j^2) = \sum_{i=1}^a \sum_{j=1}^b N_{ij} \sigma_\beta^2$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{i=1}^a \sum_{j=1}^b N_{ij}} \right)^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\left( \sum_{i=1}^a \sum_{j=1}^b N_{ij} \right)^2} E \left( \sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij} \right)^2 = \sum_{j=1}^b \frac{\sum_{i=1}^a N_{ij}^2}{\sum_{i=1}^a N_{ij}} \sigma_\rho^2$$

$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a \sum_{j=1}^b N_{ij}} \right)^2 &= \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\left( \sum_{i=1}^a \sum_{j=1}^b N_{ij} \right)^2} E \left( \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right)^2 = \sum_{j=1}^b \frac{1}{\sum_{i=1}^a N_{ij}} \sum_{i=1}^a \sum_{k=1}^{N_{ij}} E(\varepsilon_{ijk}^2) \\
&= \sum_{j=1}^b \frac{1}{\sum_{i=1}^a N_{ij}} \sum_{i=1}^a \sum_{k=1}^{N_{ij}} \sigma^2 = b \sigma^2
\end{aligned}$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right)^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{(N)^2} \sum_{i=1}^a \left( \sum_{j=1}^b N_{ij} \right)^2 \sigma_\beta^2 = \frac{\sum_{i=1}^a \left( \sum_{j=1}^b N_{ij} \right)^2}{N} \sigma_\beta^2$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right)^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{(N)^2} \sum_{i=1}^a \sum_{j=1}^b N_{ij}^2 \sigma_\rho^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2$$

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)^2 &= \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{(N)^2} E \left( \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right)^2 = \frac{1}{N} E \left( \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}^2 \right) \\ &= \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} E(\varepsilon_{ijk}^2) = \sigma^2 \end{aligned}$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \beta_j \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{N} E \left( \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \right)^2 = \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \sigma_\beta^2$$

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) &= \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\sum_{i=1}^a N_{ij} \cdot N} E \left( \sum_{i=1}^a N_{ij} \rho_{ij} \sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij} \right) \\ &= \frac{1}{N} E \left[ \sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \rho_{ij} \sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij} \right) \right] = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2 \end{aligned}$$



$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) &= \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\sum_{i=1}^a N_{ij} \cdot N} E \left( \sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right) \\
&= \frac{1}{N} E \left[ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right] = \frac{1}{N} E \left( \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right)^2 \\
&= \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} E(\varepsilon_{ijk}^2) = \sigma^2
\end{aligned}$$

$$\begin{aligned}
E \left[ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{Y}_{i\cdot j} - \bar{Y}_{\dots})^2 \right] &= \sum_{i=1}^a \sum_{j=1}^b N_{ij} \sigma_\beta^2 + b \sigma^2 + \sum_{j=1}^b \frac{\sum_{i=1}^a N_{ij}^2}{\sum_{i=1}^a N_{ij}} \sigma_\rho^2 + \frac{\sum_{i=1}^a \left( \sum_{j=1}^b N_{ij} \right)^2}{N} \sigma_\beta^2 + \sigma^2 \\
&\quad + \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2 - 2 \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \sigma_\beta^2 - 2 \sigma^2 - 2 \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2
\end{aligned}$$

Sustituyendo en  $E(CM_{\text{est. B}})$ , se obtiene:

$$E(CM_{\text{est. B}}) = \sigma^2 + \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{i=1}^a N_{ij}} \right) - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \right] \frac{\sigma_\rho^2}{b-1} + \left[ \sum_{i=1}^a \sum_{j=1}^b N_{ij} - \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \right] \frac{\sigma_\beta^2}{b-1} \quad (\text{IV.5.3})$$

Si se toma un tamaño de muestra igual en cada estrato interacción, digamos

$N_{ij} = M$ , se obtiene:

$$E(CM_{\text{est. B}}) = \sigma^2 + M \sigma_\rho^2 + a M \sigma_\beta^2 \quad (\text{IV.5.4})$$

Para la interacción:

$$E(CM_{int}) = E\left(\frac{SC_{int}}{(a-1)(b-1)}\right) = \frac{1}{(a-1)(b-1)} E(SC_{int})$$

$$E(CM_{int}) = \frac{1}{(a-1)(b-1)} E\left[\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot})^2\right]$$

$$\bar{y}_{ij\cdot} = \frac{\sum_{k=1}^{N_{ij}} Y_{ijk}}{N_{ij}} = \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} (\mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk})$$

$$\bar{y}_{ij\cdot} = \mu + \alpha_i + \beta_j + \rho_{ij} + \frac{\sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N_{ij}}$$

$$\bar{y}_{\cdot\cdot\cdot} = \mu + \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} + \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} + \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N}$$

$$\begin{aligned}
(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}...) &= \mu + \alpha_i + \beta_j + \rho_{ij} + \frac{\sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N_{ij}} \\
&- \mu - \alpha_i - \frac{\sum_{j=1}^b N_{ij} \beta_j}{\sum_{j=1}^b N_{ij}} - \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} - \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \\
&- \mu - \beta_j - \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} - \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \\
&+ \mu + \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} + \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} + \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N}
\end{aligned}$$

Elevando al cuadrado y desarrollando, se obtiene:

$$\begin{aligned}
(\bar{Y}_{i..} - \bar{Y}...) &= \beta_j^2 + \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right)^2 + \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right)^2 + \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right)^2 + \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right)^2 \\
&+ \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)^2 + 2\beta_j \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) + 2\beta_j \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) - 2\beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \\
&- 2\beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2\beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) + 2 \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) \\
&- 2 \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2 \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)
\end{aligned}$$

$$\begin{aligned}
& -2 \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2 \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) \\
& + 2 \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) + 2 \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) \\
& + 2 \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)
\end{aligned}$$

Aplicando la triple sumatoria, se obtiene:

$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{Y}_i \dots - \bar{Y} \dots)^2 &= \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right)^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right)^2 \\
&+ \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right)^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right)^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)^2 \\
&+ 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \left( \frac{\sum_{j=1}^b N_{ij} \rho_{ij}}{\sum_{j=1}^b N_{ij}} \right) + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \\
&- 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)
\end{aligned}$$

$$\begin{aligned}
& + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \\
& - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) \\
& - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \\
& - 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{j=1}^b N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \\
& + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) + 2 \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right)
\end{aligned}$$

Aplicando las propiedades de la esperanza, los supuestos y restricciones para este modelo, se obtiene:

$$E \left( \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j^2 \right) = \sum_{i=1}^a \sum_{j=1}^b N_{ij} E(\beta_j^2) = \sum_{i=1}^a \sum_{j=1}^b N_{ij} \sigma_\beta^2$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right)^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\left( \sum_{i=1}^a N_{ij} \right)^2} E \left( \sum_{i=1}^a N_{ij} \rho_{ij} \right)^2 = \sum_{j=1}^b \frac{\sum_{i=1}^a N_{ij}^2}{\sum_{i=1}^a N_{ij}} \sigma_\rho^2$$

$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \epsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right)^2 &= \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\left( \sum_{i=1}^a N_{ij} \right)^2} E \left( \sum_{i=1}^a \sum_{k=1}^{N_{ij}} \epsilon_{ijk} \right)^2 = \sum_{j=1}^b \frac{1}{\sum_{i=1}^a N_{ij}} \sum_{i=1}^a \sum_{k=1}^{N_{ij}} E(\epsilon_{ijk}^2) \\
&= \sum_{j=1}^b \frac{1}{\sum_{i=1}^a N_{ij}} \sum_{i=1}^a \sum_{k=1}^{N_{ij}} \sigma^2 = b \sigma^2
\end{aligned}$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right)^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{(N)^2} \sum_{i=1}^a \left( \sum_{j=1}^b N_{ij} \right)^2 \sigma_\beta^2 = \frac{\sum_{i=1}^a \left( \sum_{j=1}^b N_{ij} \right)^2}{N} \sigma_\beta^2$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right)^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{(N)^2} \sum_{i=1}^a \sum_{j=1}^b N_{ij}^2 \sigma_\rho^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2$$

$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \epsilon_{ijk}}{N} \right)^2 &= \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{(N)^2} E \left( \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \epsilon_{ijk} \right)^2 = \frac{1}{N} E \left( \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \epsilon_{ijk}^2 \right) \\
&= \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} E(\epsilon_{ijk}^2) = \sigma^2
\end{aligned}$$

$$\sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \beta_j \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j}{N} \right) = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{N} E \left( \sum_{i=1}^a \sum_{j=1}^b N_{ij} \beta_j \right)^2 = \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \sigma_\beta^2$$

$$\begin{aligned}
& \sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a N_{ij} \rho_{ij}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij}}{N} \right) = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\sum_{i=1}^a N_{ij} \cdot N} E \left( \sum_{i=1}^a N_{ij} \rho_{ij} \sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij} \right) \\
& = \frac{1}{N} E \left[ \sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \rho_{ij} \sum_{i=1}^a \sum_{j=1}^b N_{ij} \rho_{ij} \right) \right] = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2
\end{aligned}$$

$$\begin{aligned}
& \sum_{i=1}^a \sum_{j=1}^b N_{ij} E \left( \frac{\sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{\sum_{i=1}^a N_{ij}} \right) \left( \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N} \right) = \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}}{\sum_{i=1}^a N_{ij} \cdot N} E \left( \sum_{i=1}^a \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right) \\
& = \frac{1}{N} E \left[ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right] = \frac{1}{N} E \left( \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \right)^2 \\
& = \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} E(\varepsilon_{ijk}^2) = \sigma^2
\end{aligned}$$

$$\begin{aligned}
E \left[ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (\bar{Y}_{\cdot j \cdot} - \bar{Y} \dots)^2 \right] &= \sum_{i=1}^a \sum_{j=1}^b N_{ij} \sigma_\beta^2 + b \sigma^2 + \sum_{j=1}^b \frac{\sum_{i=1}^a N_{ij}^2}{\sum_{i=1}^a N_{ij}} \sigma_\rho^2 + \frac{\sum_{i=1}^a \left( \sum_{j=1}^b N_{ij} \right)^2}{N} \sigma_\beta^2 + \sigma^2 \\
&+ \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2 - 2 \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \sigma_\beta^2 - 2 \sigma^2 - 2 \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \sigma_\rho^2
\end{aligned}$$

Sustituyendo en  $E(CM_{int})$ , se obtiene:

$$\begin{aligned}
E(CM_{int}) = & \left[ N - \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \sum_{j=1}^b \left( \frac{\sum_{i=1}^a N_{ij}^2}{\sum_{i=1}^a N_{ij}} \right) - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} + 2 \sum_{i=1}^a \sum_{j=1}^b \frac{N_{ij}^3}{\sum_{i=1}^a N_{ij} \sum_{j=1}^b N_{ij}} \right] \frac{\sigma_p^2}{(a-1)(b-1)} \\
& + \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \right] \frac{\sigma_\beta^2}{(a-1)(b-1)} \\
& + \left[ ab - a - b - 1 + 2 \sum_{i=1}^a \sum_{j=1}^b \frac{N_{ij}^2}{\sum_{i=1}^a N_{ij} \sum_{j=1}^b N_{ij}} \right] \frac{\sigma^2}{(a-1)(b-1)}
\end{aligned} \tag{IV.5.5}$$

Si se toma un tamaño de muestra igual en cada estrato interacción, digamos

$N_{ij} = M$ , se obtiene:

$$E(CM_{int}) = \sigma^2 + M\sigma_p^2 \tag{IV.5.6}$$

Para el error:

$$E(CM_{EE}) = E\left( \frac{SC_{EE}}{ab(N_{ij} - 1)} \right) = \frac{1}{ab(N_{ij} - 1)} E(SC_{EE})$$

$$E(SC_{EE}) = E\left[ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (y_{ijk} - \bar{y}_{ij\cdot})^2 \right]$$

$$(Y_{ijk} - \bar{Y}_{ij\cdot}) = \mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk} - \mu - \alpha_i - \beta_j - \rho_{ij} - \frac{\sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N_{ij}}$$



$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk}^2 - \sum_{i=1}^a \sum_{j=1}^b N_{ij} \left( \frac{\sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N_{ij}} \right)^2 - 2 \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} \varepsilon_{ijk} \left( \frac{\sum_{k=1}^{N_{ij}} \varepsilon_{ijk}}{N_{ij}} \right)$$

$$E \left[ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2 \right] = N\sigma^2 + ab\sigma^2 - 2ab\sigma^2 = (N - ab)\sigma^2 = ab(N_{ij} - 1)\sigma^2$$

$$E(CM_{EE}) = \frac{ab(N_{ij} - 1)}{ab(N_{ij} - 1)} \sigma^2$$

$$E(CM_{EE}) = \sigma^2 \quad (IV.5.7)$$

En el cuadro 7, se muestran las Esperanzas de los Cuadrados Medios obtenidas para cada uno de los estratos.

Tal como se indicó al inicio del Capítulo, existen varios Modelos Mixtos, a continuación se presentan las Esperanzas de los Cuadrados Medios (cuadro 8), para los Modelos Alterno y Estándar con datos balanceados, presentados por Montgomery (1991) y John (1971).

Para el Modelo Estándar se consideran los supuestos:

$$\sum_{i=1}^a \alpha_i = 0 \quad \sum_{i=1}^a \rho_{ij} = 0 \quad v(\rho_{ij}) = \frac{(a-1)}{a} \sigma_p^2$$

$$\text{Cov}(\rho_{ij}; \rho_{i'j'}) = -\frac{\sigma_p^2}{a} \quad i \neq i'$$

El resto de las covarianzas se asumen cero.

**Cuadro 7: Esperanza de los Cuadrados Medios para el Modelo Mixto Estratificado Doble**

FdeV	g.de L.	E(CM)
Est.T	a-1	$\sigma^2 + \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \right] \frac{\sigma_p^2}{a-1} + \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \right] \frac{\sigma_\beta^2}{a-1}$ $+ \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i^2}{a-1}$
Est.B	b-1	$\sigma^2 + \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \right] \frac{\sigma_p^2}{b-1} + \left[ \sum_{i=1}^a \sum_{j=1}^b N_{ij} - \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \right] \frac{\sigma_\beta^2}{b-1}$
Int.	(a-1)(b-1)	$\left[ N - \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \sum_{j=1}^b \left( \frac{\sum_{i=1}^a N_{ij}^2}{\sum_{i=1}^a N_{ij}} \right) - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} + 2 \sum_{i=1}^a \sum_{j=1}^b \frac{N_{ij}^3}{\sum_{i=1}^a N_{ij} \sum_{j=1}^b N_{ij}} \right] \frac{\sigma_p^2}{(a-1)(b-1)}$ $+ \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \right] \frac{\sigma_\beta^2}{(a-1)(b-1)} + \left[ ab - a - b - 1 + 2 \sum_{i=1}^a \sum_{j=1}^b \frac{N_{ij}^2}{\sum_{i=1}^a N_{ij} \sum_{j=1}^b N_{ij}} \right] \frac{\sigma^2}{(a-1)(b-1)}$
EE	N-ab	$\sigma^2$
Total	N-1	

Si en la Esperanza de los Cuadrados Medios del cuadro 7, se consideran datos balanceados ( $N_{ij} = M$ ), se obtienen las Esperanzas de los Cuadrados Medios del Modelo alterno, mostradas en el cuadro 8.

**Cuadro 8: Esperanza de los Cuadrados Medios para el Modelo Estándar y Modelo Alternativo**

F.deV.	g.de L.	Modelo Estándar	Modelo Alternativo
Efecto Fijo	$\alpha-1$	$\sigma^2 + M\sigma_{\alpha\beta}^2 + \frac{bM\sum_{i=1}^{\alpha}\alpha_i^2}{\alpha-1}$	$\sigma^2 + M\sigma_{\alpha\beta}^2 + \frac{bM\sum_{i=1}^{\alpha}\alpha_i^2}{\alpha-1}$
Factor Aleatorio	$b-1$	$\sigma^2 + \alpha M\sigma_{\beta}^2$	$\sigma^2 + M\sigma_{\alpha\beta}^2 + \alpha M\sigma_{\beta}^2$
Interacc.	$(\alpha-1)(b-1)$	$\sigma^2 + M\sigma_{\alpha\beta}^2$	$\sigma^2 + M\sigma_{\alpha\beta}^2$
EE	$ab(M-1)$	$\sigma^2$	$\sigma^2$
Total	$N-1$		

Fuente: Montgomery (1991)

## 6. ESTIMACIÓN DE LOS PARÁMETROS A TRAVÉS DEL MUESTREO

Basándose en Lorh (2000), se obtienen los siguientes estimadores del Total, Media y Varianza.

Sea “S” el conjunto de elementos que pertenecen a la muestra; entonces:

**Total de la población** se puede escribir como:

$$\tau = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k \in S} Y_{ijk} + \sum_{i=1}^a \sum_{j=1}^b \sum_{k \notin S} Y_{ijk}$$

Y es un valor posible de la variable aleatoria:

$$T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k \in S} Y_{ijk} + \sum_{i=1}^a \sum_{j=1}^b \sum_{k \notin S} Y_{ijk}$$

Utilizando el estimador por Mínimos Cuadrados de  $\bar{Y}$ , se obtiene:

### Estimaciones para el Total:

✓ *Total General*

$$\begin{aligned}\hat{T} &= n\bar{y} + (N-n)\bar{y} \\ \hat{T} &= N\bar{y} \\ \hat{T} &= \frac{N}{n} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} Y_{ijk}\end{aligned} \quad (\text{IV.6.1})$$

$\hat{T}$  es un estimador sesgado de T para el modelo; cuyo sesgo está relacionado con el efecto del estrato T, es decir;

$$E_M[\hat{T} - T] = b \left[ \sum_{i=1}^a \left( \frac{N}{n} n_{ij} - N_{ij} \right) \alpha_i \right]$$

$$\begin{aligned}V(\hat{T}) &= E_M \left[ (\hat{T} - T)^2 \right] = E_M \left[ \left( \frac{N}{n} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} Y_{ijk} - \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk} \right)^2 \right] \\ V(\hat{T}) &= E_M \left[ (\hat{T} - T)^2 \right] = b^2 \sum_{i=1}^a \left( \frac{N}{n} n_{ij} - N_{ij} \right)^2 \alpha_i^2 + \alpha^2 \sum_{j=1}^b \left( \frac{N}{n} n_{ij} - N_{ij} \right)^2 \sigma_{\beta_j}^2 \\ &\quad + \sum_{i=1}^a \sum_{j=1}^b \left( \frac{N}{n} n_{ij} - N_{ij} \right)^2 \sigma_{ij}^2 + \frac{N^2}{n^2} \sum_{i=1}^a \sum_{j=1}^b n_{ij} \sigma_{ij}^2 \\ &\quad - 2 \frac{N}{n} \sum_{i=1}^a \sum_{j=1}^b n_{ij} \sigma_{ij}^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \sigma_{ij}^2\end{aligned} \quad (\text{IV.6.2})$$

✓ *Total en la interacción*

$$T_{ij\cdot} = \sum_{k=1}^{N_{ij}} Y_{ijk} = \sum_{k \in S} Y_{ijk} + \sum_{k \notin S} Y_{ijk}$$

$$\hat{T}_{ij\cdot} = n_{ij} \bar{Y}_{ij\cdot} + (N_{ij} - n_{ij}) \bar{Y}_{ij\cdot}$$

$$\hat{T}_{ij\cdot} = N_{ij} \bar{y}_{ij\cdot}$$

$$\hat{T}_{ij\cdot} = \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$$

(IV.6.3)

$\hat{T}_{ij\cdot}$  es un estimador insesgado de  $T_{ij\cdot}$  para el modelo, es decir;

$$E_{\mathcal{M}}[\hat{T}_{ij\cdot} - T_{ij\cdot}] = 0$$

$$V(\hat{T}_{ij\cdot}) = E_{\mathcal{M}}\left[(\hat{T}_{ij\cdot} - T_{ij\cdot})^2\right] = E_{\mathcal{M}}\left[\left(\frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk} - \sum_{k=1}^{N_{ij}} Y_{ijk}\right)^2\right]$$

$$V(\hat{T}_{ij\cdot}) = E_{\mathcal{M}}\left[(\hat{T}_{ij\cdot} - T_{ij\cdot})^2\right] = \frac{N_{ij}^2}{n_{ij}} \sigma_{ij}^2 \left(1 - \frac{n_{ij}}{N_{ij}}\right)$$

(IV.6.4)

✓ *Total en el estrato T*

$$T_{i\cdot\cdot} = \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk} = \sum_{j=1}^b \sum_{k \in S} Y_{ijk} + \sum_{j=1}^b \sum_{k \notin S} Y_{ijk}$$

$$\hat{T}_{i\cdot\cdot} = \sum_{j=1}^b n_{ij} \bar{Y}_{ij\cdot} + \sum_{j=1}^b (N_{ij} - n_{ij}) \bar{Y}_{ij\cdot}$$

$$\hat{T}_{i\cdot\cdot} = \sum_{j=1}^b N_{ij} \bar{y}_{ij\cdot}$$

$$\hat{T}_{i\cdot\cdot} = \sum_{j=1}^b \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$$

(IV.6.5)

$\hat{T}_{i..}$  es un estimador insesgado de  $T_{i..}$  para el modelo, es decir;

$$E_M[\hat{T}_{i..} - T_{i..}] = 0$$

$$\begin{aligned} V(\hat{T}_{i..}) &= E_M[(\hat{T}_{i..} - T_{i..})^2] = E_M\left[\left(\sum_{j=1}^b \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk} - \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk}\right)^2\right] \\ V(\hat{T}_{i..}) &= E_M[(\hat{T}_{i..} - T_{i..})^2] = \sum_{j=1}^b \frac{N_{ij}^2}{n_{ij}} \sigma_{ij}^2 \left(1 - \frac{n_{ij}}{N_{ij}}\right) \end{aligned} \quad (\text{IV.6.6})$$

✓ Total en el estrato B

$$T_{.j.} = \sum_{i=1}^a \sum_{k=1}^{N_{ij}} Y_{ijk} = \sum_{i=1}^a \sum_{k \in S} Y_{ijk} + \sum_{i=1}^a \sum_{k \notin S} Y_{ijk}$$

$$\hat{T}_{.j.} = \sum_{i=1}^a n_{ij} \bar{y}_{ij.} + \sum_{i=1}^a (N_{ij} - n_{ij}) \bar{y}_{ij.}$$

$$\hat{T}_{.j.} = \sum_{i=1}^a N_{ij} \bar{y}_{ij.} \quad (\text{IV.6.7})$$

$$\hat{T}_{.j.} = \sum_{i=1}^a \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$$

$\hat{T}_{.j.}$  es un estimador insesgado de  $T_{.j.}$  para el modelo, es decir;

$$E_M[\hat{T}_{.j.} - T_{.j.}] = 0$$

$$\begin{aligned} V(\hat{T}_{.j.}) &= E_M[(\hat{T}_{.j.} - T_{.j.})^2] = E_M\left[\left(\sum_{i=1}^a \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk} - \sum_{i=1}^a \sum_{k=1}^{N_{ij}} Y_{ijk}\right)^2\right] \\ V(\hat{T}_{.j.}) &= E_M[(\hat{T}_{.j.} - T_{.j.})^2] = \sum_{i=1}^a \frac{N_{ij}^2}{n_{ij}} \sigma_{ij}^2 \left(1 - \frac{n_{ij}}{N_{ij}}\right) \end{aligned} \quad (\text{IV.6.8})$$

### Estimaciones para la Media:

✓ *Media General*

$\bar{y}$  es un estimador sesgado de  $\bar{Y}$  para el modelo; cuyo sesgo está relacionado con el efecto del estrato T, es decir;

$$E_M[\bar{y} - \mu] = b \left[ \sum_{i=1}^a \left( \frac{n_{ij}}{n} - \frac{N_{ij}}{N} \right) \alpha_i \right]$$

$$V(\bar{y}) = \frac{V(\hat{T})}{N^2}$$

$$V(\bar{y}) = \frac{1}{N^2} \left[ b^2 \sum_{i=1}^a \left( \frac{N}{n} n_{ij} - N_{ij} \right)^2 \alpha_i^2 + \alpha^2 \sum_{j=1}^b \left( \frac{N}{n} n_{ij} - N_{ij} \right)^2 \sigma_{\beta_j}^2 + \sum_{i=1}^a \sum_{j=1}^b \left( \frac{N}{n} n_{ij} - N_{ij} \right)^2 \sigma_{ij}^2 \right. \quad (IV.6.9)$$

$$\left. + \frac{N^2}{n^2} \sum_{i=1}^a \sum_{j=1}^b n_{ij} \sigma_{ij}^2 - 2 \frac{N}{n} \sum_{i=1}^a \sum_{j=1}^b n_{ij} \sigma_{ij}^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \sigma_{ij}^2 \right]$$

✓ *Media en la interacción*

$\bar{y}_{ij\cdot}$  es un estimador insesgado de  $\bar{Y}_{ij\cdot}$  para el modelo, es decir;

$$E_M[\bar{y}_{ij\cdot} - \mu_{ij\cdot}] = 0$$

$$V(\bar{y}_{ij\cdot}) = \frac{V(\hat{T}_{ij\cdot})}{N^2}$$

$$V(\bar{y}_{ij\cdot}) = \frac{N_{ij}^2}{N^2 n_{ij}} \sigma_{ij}^2 \left( 1 - \frac{n_{ij}}{N_{ij}} \right) \quad (IV.6.10)$$

✓ *Media en el estrato T*

$\bar{y}_{i..}$  es un estimador insesgado de  $\bar{Y}_{i..}$  para el modelo, es decir;

$$E_M[\bar{y}_{i..} - \mu_{i..}] = 0$$

$$\begin{aligned} V(\bar{y}_{i..}) &= \frac{V(\hat{t}_{i..})}{N^2} \\ V(\bar{y}_{i..}) &= \sum_{j=1}^b \frac{N_{ij}^2}{N^2 n_{ij}} \sigma_{ij}^2 \left(1 - \frac{n_{ij}}{N_{ij}}\right) \end{aligned} \quad (\text{IV.6.11})$$

✓ *Media en el estrato B*

$\bar{y}_{.j.}$  es un estimador insesgado de  $\bar{Y}_{.j.}$  para el modelo, es decir;

$$E_M[\bar{y}_{.j.} - \mu_{.j.}] = 0$$

$$\begin{aligned} V(\bar{y}_{.j.}) &= \frac{V(\hat{t}_{.j.})}{N^2} \\ V(\bar{y}_{.j.}) &= \sum_{i=1}^a \frac{N_{ij}^2}{N^2 n_{ij}} \sigma_{ij}^2 \left(1 - \frac{n_{ij}}{N_{ij}}\right) \end{aligned} \quad (\text{IV.6.12})$$

Donde, un estimador insesgado de  $\sigma_{ij}^2$  es  $s_{ij}^2$

$$s_{ij}^2 = \sum_{k=1}^{n_{ij}} \frac{1}{n_{ij} - 1} \left[ \left( \mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk} \right) - \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} \left( \mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk} \right) \right]^2$$

$$s_{ij}^2 = \frac{1}{n_{ij} - 1} \left[ \sum_{k=1}^{n_{ij}} \varepsilon_{ijk}^2 - n_{ij} \sum_{k=1}^{n_{ij}} \frac{\varepsilon_{ijk}^2}{n_{ij}} \right]$$

$$s_{ij}^2 = \sigma_{ij}^2$$



**Observación:** si el tamaño del estrato formado por los estratos T y B, es constante, es decir,  $N_{ij} = M$  y el tamaño de muestra en los estratos es la misma, es decir,  $n_{ij} = m$ . Se tiene que las estimaciones tanto para el total general, como para la media general son insesgadas; tal y como puede observarse a continuación:

✓ *Total General*

$$E_M[\hat{T} - T] = b \left[ \sum_{i=1}^a \left( \frac{N}{n} m - M \right) \alpha_i \right] = b \left[ \sum_{i=1}^a \left( \frac{abM}{abm} m - M \right) \alpha_i \right] = 0$$

$$\begin{aligned} V(\hat{T}) &= E_M \left[ (\hat{T} - T)^2 \right] = b^2 \sum_{i=1}^a \left( \frac{N}{n} m - M \right)^2 \alpha_i^2 + a^2 \sum_{j=1}^b \left( \frac{N}{n} n - M \right)^2 \sigma_{\beta_j}^2 \\ &\quad + \sum_{i=1}^a \sum_{j=1}^b \left( \frac{N}{n} n - M \right)^2 \sigma_{ij}^2 + \frac{N^2}{n^2} \sum_{i=1}^a \sum_{j=1}^b m \sigma_{ij}^2 \\ &\quad - 2 \frac{N}{n} \sum_{i=1}^a \sum_{j=1}^b m \sigma_{ij}^2 + \sum_{i=1}^a \sum_{j=1}^b M \sigma_{ij}^2 \\ V(\hat{T}) &= \frac{M^2}{m} \sum_{i=1}^a \sum_{j=1}^b \sigma_{ij}^2 - 2M \sum_{i=1}^a \sum_{j=1}^b \sigma_{ij}^2 + \sum_{i=1}^a \sum_{j=1}^b M \sigma_{ij}^2 \end{aligned}$$

$$V(\hat{T}) = \frac{M^2}{m} \left( 1 - \frac{m}{M} \right) \sum_{i=1}^a \sum_{j=1}^b \sigma_{ij}^2 \quad (\text{IV.6.13})$$

✓ *Media General*

$$E_M[\bar{y} - \mu] = b \left[ \sum_{i=1}^a \left( \frac{m}{n} - \frac{M}{N} \right) \alpha_i \right] = b \left[ \sum_{i=1}^a \left( \frac{m}{abm} - \frac{M}{abM} \right) \alpha_i \right]$$

$$E_M[\bar{y} - \mu] = 0$$

$$V(\bar{y}) = \frac{1}{(abM)^2} \left[ b^2 \sum_{i=1}^a \left( \frac{abM}{abm} m - M \right)^2 \alpha_i^2 + a^2 \sum_{j=1}^b \left( \frac{abM}{abm} m - M \right)^2 \sigma_{\beta_j}^2 \right. \\ \left. + \sum_{i=1}^a \sum_{j=1}^b \left( \frac{abM}{abm} m - M \right)^2 \sigma_{ij}^2 + \frac{(abM)^2}{(abm)^2} \sum_{i=1}^a \sum_{j=1}^b m \sigma_{ij}^2 \right. \\ \left. - 2 \frac{abM}{abm} \sum_{i=1}^a \sum_{j=1}^b m \sigma_{ij}^2 + \sum_{i=1}^a \sum_{j=1}^b M \sigma_{ij}^2 \right]$$

$$V(\bar{y}) = \frac{1}{(ab)^2 m} \sum_{i=1}^a \sum_{j=1}^b \sigma_{ij}^2 - \frac{1}{(ab)^2 M} \sum_{i=1}^a \sum_{j=1}^b \sigma_{ij}^2$$

Al extraer como factor común  $\left(\frac{M}{N}\right)^2 \frac{1}{m}$  ; donde  $N = abM$ , de la expresión anterior, se obtiene:

$$V(\bar{y}) = \sum_{i=1}^a \sum_{j=1}^b \left(1 - \frac{m}{M}\right) \left(\frac{M}{N}\right)^2 \frac{\sigma_{ij}^2}{m} \quad (\text{IV.6.14})$$

Puede observarse que estas estimaciones y las estimaciones para el estrato T y el B son semejantes a las estimaciones presentadas para el diseño de muestreo estratificado (pág. 34).

## **CAPÍTULO V**

### **CONTRASTE ENTRE EL DISEÑO/ANÁLISIS DE EXPERIMENTOS Y LA TEORÍA DEL MUESTREO**

El contraste del diseño de experimentos y el muestreo, se realizará tomando en consideración:

- Los principios básicos del diseño de experimentos, repetición, aleatorización y control local y los criterios del diseño de investigación: realismo, aleatorización y representatividad, como los declara Kish (1995)
- Los objetivos del diseño de experimentos y del muestreo
- Validez externa e interna
- Modelos estadísticos
- Ventajas

La base de estos contrastes, se plantea en los cuadros 9 y 10, que reflejan algunas coincidencias y diferencias entre el diseño de experimentos y la teoría del muestreo expuestas por Brakel and Renssen (1998).

**Cuadro 9: Coincidencias entre el diseño de experimentos y la teoría del muestreo**

Diseño de Experimentos	Teoría del Muestreo
Asignación al azar de los tratamientos a las unidades experimentales	Muestreo aleatorio de unidades de muestreo de una población finita
Replicación de los tratamientos	Tamaño de muestra
Diseño de bloques aleatorizados	Diseño de muestreo estratificado
Diseño cuadrado latino	Muestreo de Lattice o Estratificación profunda
Diseño de parcelas divididas	Diseño de muestreo en dos etapas
Análisis de covarianza	Estimador de regresión lineal

Fuente: Brakel and Renssen (1998)

**Cuadro 10: Diferencias entre el diseño y análisis de experimentos y la teoría del muestreo**

Diseño de Experimentos	Teoría del Muestreo
La estocasticidad es introducida porque las observaciones se asumen como resultados de una variable aleatoria	La estocasticidad es introducida porque se toma una muestra aleatoria de una población finita.
Tradicionalmente dependiente del Modelo	Tradicionalmente basado en el diseño de muestreo
Los métodos estadísticos principalmente intentan garantizar una validez interna suficiente.	Los métodos estadísticos principalmente intentan garantizar una validez externa suficiente.
Inferencia sobre una superpoblación infinita hipotética	Inferencia sobre una población finita.

Fuente: Brakel and Renssen (1998)

A continuación se describen cada uno de los ítems establecidos anteriormente para el contraste

## 1. PRINCIPIOS BÁSICOS DEL DISEÑO DE EXPERIMENTOS Y LOS CRITERIOS DEL DISEÑO DE INVESTIGACIÓN

Es importante resaltar una reflexión que presenta R. A. Fisher en su texto, *Design of Experiments*, citada por Kish (1995), “El experto en estadística no puede eludir la obligación de tener claros los principios de la inferencia estadística, pero tampoco las demás personas pensantes pueden evitar tal obligación”.

Por lo anterior, para comprender las similitudes y diferencias entre la teoría de muestreo y el diseño de experimentos es necesario definir; los principios básicos del diseño de experimentos: aleatorización repetición y control local; y los tres criterios del muestreo: realismo, aleatorización y representatividad.

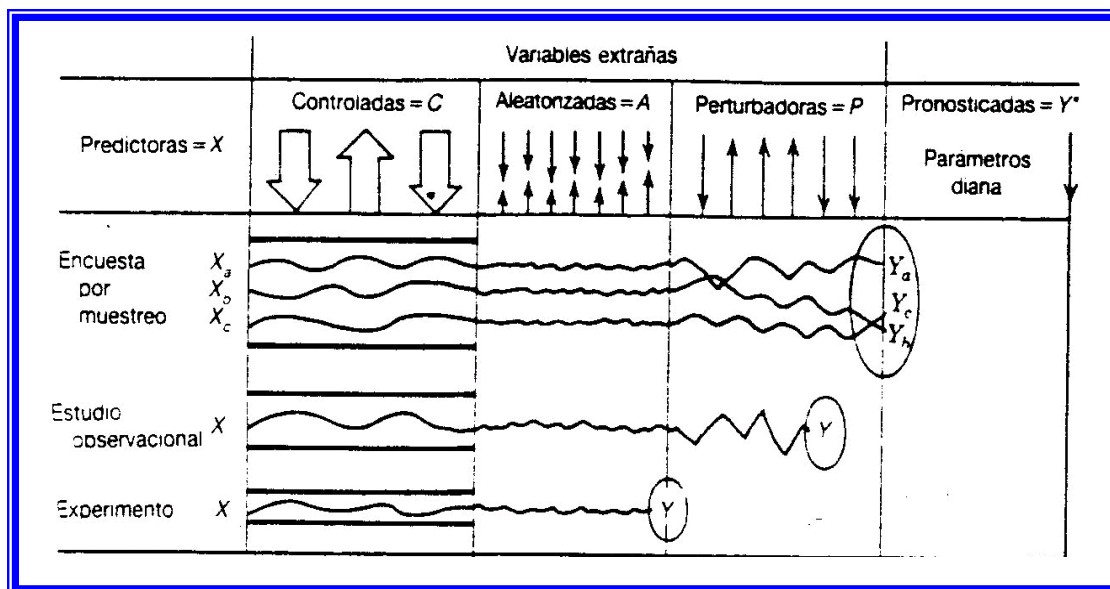
Antes, se especificarán **cuatro tipos de variables** que según Kish (1995), intervienen en la investigación empírica. Las mismas, sirven para relacionar el muestreo de encuestas y el diseño de experimentos.

- **Variables explicativas (E):** aquellas variables que expresan los objetivos del diseño de la investigación, y entre las cuales el investigador desea encontrar y medir alguna relación preestablecida. Las variables explicativas comprenden dos grupos diferenciados: las predictoras (X) incluyen las causas buscadas de las relaciones, y las pronosticadas (Y) describen los efectos de la predicción. Se han utilizado otros nombres

para estos grupos de variables: independientes y dependientes. Las variables explicativas (E), las predictoras (X) y las pronosticadas (Y), materializan los objetivos de la investigación, y los otros tres tipos de variables son extraños a esos objetivos (figura 2).

- **Variables controladas (C):** comprenden aquellas variables extrañas que pueden ser adecuadamente controladas mediante el diseño de investigación. Puede ejercerse el control bien a través del diseño de los procedimientos de selección, o bien mediante técnicas de estimación en el análisis estadístico, o quizás empleando ambos procedimientos simultáneamente. Las técnicas para controlar las variables extrañas tienen por objetivo disminuir los errores aleatorios (A), o disminuir los efectos del sesgo de variables perturbadoras (P), o ambos.
- **Variables perturbadoras (P):** son aquellas extrañas controladas que pueden ser confundidas con las variables explicativas (E). La principal desventaja de los diseños no experimentales reside en el fallo de trasladar todas estas variables P bien al tipo C de variables controladas, o bien al tipo A de variables aleatorias.
- **Variables aleatorizadas (A):** son extrañas no controladas que son tratadas como errores aleatorios. En los experimentos “ideales” o diseños de experimentos, se hace que sean de hecho y operacionalmente aleatorizadas, pero en las encuestas y en otras investigaciones sólo se

suponen que están aleatorizadas. Puede considerarse la aleatorización como una forma de control experimental, aunque distinta de las formas utilizadas para las variables de tipo C.



**Figura 2: Efectos de tres tipos de variables extrañas (C, A, P) sobre las variables explicativas (Kish, 1995)**

La figura anterior, muestra que los tres tipos de variables extrañas pueden obstaculizar el paso de las variables explicativas (E), desde las predictoras (X) hasta las pronosticadas (Y). Algunas variables son controladas (C). Otras son aleatorizadas en los experimentos, o tratadas como aleatorizadas en las encuestas. Sin embargo, algunos efectos permanecen como variables de perturbación (P) en las encuestas, pero ninguno de estos efectos permanece en los diseños de experimentos.

Las flechas hacia arriba y hacia abajo representan los efectos contrapuestos de las variables extrañas. Las variables controladas aparecen con un trazo

más grueso que las variables perturbadoras, porque se confía que las variables más fuertes sean controladas por el diseño. Las variables aleatorizadas se trazan como numerosas, delgadas y actuando en ambas direcciones, tendiendo a neutralizarse entre ellas; sus efectos sobre el curso ( $X \rightarrow Y$ ) se muestra en la figura 2 como relativamente débiles.

### ***Principios básicos del diseño de experimentos:***

Brakel and Renssen (1998), Chacín (2000), Steel y Torrie (1988), entre otros autores manifiestan que Los principios básicos del diseño de experimentos son: Aleatorización, Replicación y Control Local.

- **Aleatorización:** Neter y col (1985) señalan que la aleatorización en los experimentos es una idea relativamente reciente, la cual fue introducida primero por Fisher. Antes de su empleo se asignaban los tratamientos a las unidades experimentales, de manera sistemática o subjetivamente.

Según Chacín (2000) consiste en una asignación al azar de los tratamientos a las unidades experimentales, lo cual genera que sea válida la independencia de los errores de las observaciones. La aleatorización sólo asegura que el efecto de la asociación sobre cualquier comparación entre los tratamientos se haga tan pequeña como sea posible. Sin embargo, siempre existe la correlación entre los errores- asociados con unidades experimentales adyacentes en espacio o tiempo; y no puede



eliminarse totalmente. Este es un principio importante, ya que es un instrumento para eliminar tendencias, sistematización y subjetividad.

Steel y Torrie (1988) señalan que los métodos estadísticos requieren que las observaciones (o los errores) sean variables aleatorias independientes. La aleatorización usualmente confirma esta suposición. Además, las observaciones obtenidas constituyen muestras aleatorias de una (o de varias) población(es).

Neter y col. (1985), acotan que la aleatorización no es sólo para asignar los tratamientos a las unidades experimentales, sino también permite evitar los efectos sistemáticos que puedan estar presentes, que no son controlados por el investigador.

Kish (1995) agrega que el objetivo de la aleatorización en los experimentos reside en colocar todas las variables del tipo perturbadoras (P) dentro del las del tipo aleatorizadas (A). En los experimentos no se dejan variables del tipo P. Todas las variables extrañas han sido, o bien controladas dentro del tipo C, o bien aleatorizadas dentro del tipo A.

Adicionalmente el autor menciona que la aleatorización de los sujetos en los tratamientos es la estrategia para eliminar los sesgos en la medición de los efectos del tratamiento que son debidos a la selección realizada entre los sujetos del experimento.

- **Replicación o Repetición:** Chacín (2000) lo define como la repetición de las diferentes unidades experimentales que conforman un experimento.

Steel y Torrie (1988) consideran que tiene dos propiedades importantes: En primer lugar permite al experimentador obtener una estimación del error experimental. Tal estimación se convierte en la unidad básica para determinar si las diferencias observadas en los datos son estadísticamente significativas. En segundo lugar, el uso de réplicas permite al experimentador calcular una estimación más precisa del efecto de un factor en el experimento si se usa la media de la muestra como una estimación de dicho efecto; esto es así porque si  $\sigma^2$  es la variancia de los datos y hay  $n$  réplicas, entonces la variancia de la media de la muestra es

$$\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}.$$

Neter y col (1985) agregan que las repeticiones permiten un control sobre la precisión de las estimaciones o sobre el poder de la prueba a través de la manipulación del número de repeticiones.

Chacín (2000) acota, que en vista de que la variancia de la media es igual a la variancia de las observaciones entre el número de replicaciones ( $\sigma_{\bar{y}}^2 = \sigma^2 / n$ ), a medida que las replicaciones aumenten disminuye el efecto promedio del factor.

Steel y Torrie (1988) agregan entre las funciones de las repeticiones: Mejorar la precisión de un experimento mediante la reducción de la desviación estándar de una media de tratamiento. Aumentar el alcance de la inferencia del experimento a través de la selección del uso apropiado de unidades experimentales más variables. Y ejercer control sobre la varianza del error.

Según Montgomery (1991) el objetivo de la **inferencia estadística** es obtener conclusiones acerca de una población usando una muestra de la misma. La inferencia estadística hace uso de cantidades calculadas con las observaciones de una muestra. Una técnica de inferencia estadística llamada prueba de hipótesis ayuda al experimentador para determinar si las diferencias observadas en los datos son estadísticamente significativas. Una hipótesis estadística es una afirmación acerca de los parámetros de una población. Por ejemplo, si se quiere probar que no hay diferencias entre tratamientos, se expresa de la forma:  $H_0: \mu_i = \mu_j \quad i \neq j$ . La afirmación  $H_0$  se conoce como hipótesis nula, mientras que  $H_a: \mu_i \neq \mu_j \quad i \neq j$  se denomina hipótesis alterna.

Agrega el autor, que para probar una hipótesis hay que diseñar un procedimiento que permita seleccionar una muestra aleatoria y calcular un estadístico de prueba apropiado con objeto de rechazar o no la hipótesis nula  $H_0$ . Parte de este procedimiento consiste en especificar el conjunto de valores del estadístico de prueba que conduzcan al rechazo

de  $H_0$ . Este conjunto de valores se denomina *región crítica* o *región de rechazo* de la prueba.

Neyman y Pearson (1928), citados por Canavos (1988) conceptualizaron dos tipos de error que el investigador debe asumir, el error tipo I, de rechazar la hipótesis nula cuando ésta es cierta y el error tipo II, que se comete cuando no se rechaza la hipótesis nula y esta es falsa. El error tipo I ocurre con una probabilidad igual a  $\alpha$  y el error tipo II con una probabilidad igual a  $\beta$ . Cualquier investigador desearía que ambos errores sean igual a cero, sin embargo, factores como el tiempo, espacio y recursos disponibles para la realización de su experimento lo obligan aceptar valores de  $\alpha$  y  $\beta$  mayores de cero. Las probabilidades de cometer estos errores reciben un símbolo especial:

$$\alpha = P(\text{error tipo I}) = P(\text{rechazar } H_0 / H_0 \text{ es verdadera}) \quad (\text{V.1.1})$$

$$\beta = P(\text{error tipo II}) = P(\text{norechazar } H_0 / H_0 \text{ es falsa}) \quad (\text{V.1.2})$$

Steel y Torrie (1988) señalan que el interés es poder detectar  $H_1$  cuando  $H_1$  es verdadera. Es claro que esto supone no cometer el error de tipo II, así que la probabilidad de que eso ocurra es  $1 - \beta$ . Esta habilidad para detectar  $H_1$  cuando  $H_1$  es verdadera se llama **potencia de la prueba**.

Expresado de otra manera:

$$\text{Poder de la prueba} = 1 - \beta = P(\text{rechazar } H_0 / H_0 \text{ es falsa})$$

Canavos (1988) indica que para un  $\alpha$  fijo y una hipótesis alternativa dada, puede disminuirse el tamaño del error tipo II ( $\beta$ ) si se incrementa el tamaño de la muestra. Por lo tanto, se desprende que la función potencia aumentará conforme aumenta el tamaño de la muestra.

Montgomery (1991) agrega que el procedimiento general de la prueba de hipótesis consiste en especificar un valor  $\alpha$ , la probabilidad de cometer el error tipo I, llamado a menudo nivel de significación de la prueba, para después diseñar un experimento que asegure un valor pequeño para la probabilidad de error tipo II. La probabilidad de error tipo II depende de la diferencia verdadera que hay entre las medias ( $\delta = \mu_1 - \mu_2$ ). La gráfica de  $\beta$  contra  $\delta$ , en el caso de un valor particular del tamaño de la muestra, se conoce como curva característica de operación. El error  $\beta$  también constituye una función del tamaño de la muestra. Generalmente, en un valor dado de  $\delta$ , el error  $\beta$  disminuye a medida que el tamaño de la muestra se incrementa. Esto significa que una diferencia específica del nivel medio es más fácil de detectar en muestras grandes que en pequeñas.

El número de réplicas en un estudio de investigación afecta la precisión de las estimaciones de las medias de los tratamientos y la potencia de las pruebas.

Según Steel y Torrie (1988) la precisión, sensibilidad o cantidad de información, se mide por el inverso de la varianza de una media. Si  $I$  representa la cantidad de información, entonces  $I = \frac{1}{\sigma_{\bar{y}}^2} = \frac{n}{\sigma^2}$ .

Por lo tanto, indican que la magnitud de un intervalo de confianza y el poder de una prueba dependen en definitiva de  $\sigma_{(\bar{y})} = \frac{\sigma^2}{n}$ . A medida que el número de repeticiones aumenta, las estimaciones de las medias poblacionales, esto es, las medias observadas de los tratamientos, se hacen más precisas. Si se detecta una diferencia de cinco unidades usando 4 repeticiones, un experimento de aproximadamente 16 repeticiones detectará la mitad de esa diferencia, o sea 2,5 unidades, pues las desviaciones estándar están en proporción 2:1. Se usa la palabra "aproximado" porque la precisión especialmente en experimentos pequeños depende en parte de los grados de libertad disponibles para estimar el error experimental.

Agregan Steel y Torrie (1988) que el número de repeticiones de un experimento depende de varios factores, de los cuales el más importante es el grado de precisión deseada. Cuanta más pequeña sea la discrepancia con respecto a la hipótesis nula que se ha de medir o detectar, mayor será el número de repeticiones requeridas.

Federer (1955) plantea que el número de repeticiones para un experimento puede ser determinado por:

- ✓ Grado de precisión deseado
- ✓ Cantidad de variabilidad presente en el material experimental
- ✓ Disponibilidad de los recursos, incluyendo personal y equipos
- ✓ Tamaño y forma de la unidad experimental

Steel y Torrie (1988) señalan que en un experimento es importante tener una cuantía correcta de precisión. No tiene objeto usar 10 repeticiones para detectar una diferencia que se puede detectar con 4 en la mayoría de los casos; tampoco es de utilidad llevar a cabo un experimento en el que el número de repeticiones sea insuficiente para detectar diferencias importantes, excepto ocasionalmente.

Asimismo, el número de repeticiones depende de la variabilidad del material utilizado. Para la misma precisión, se necesita menos repeticiones en materiales uniformes que en los heterogéneos.

El número de tratamientos afecta la precisión de un experimento y el número de repeticiones necesarias para un grado de precisión dado. Por ejemplo, si aumentamos el número de tratamientos y mantenemos constante el número de repeticiones para cada uno, entonces

aumentamos el tamaño del experimento y el número de grados de libertad para la estimación de  $\sigma^2$ . Por tanto es mejor la precisión  $\sigma_y^2 = \frac{S^2}{n}$ .

El diseño experimental también afecta la precisión de un experimento y el número de repeticiones necesarias. Cuando el número de tratamientos es grande y es necesario usar unidades experimentales más heterogéneas, entonces aumenta el error experimental por unidad. Los diseños experimentales apropiados pueden controlar parte de esta variación.

Rojas (1997) menciona que, aún cuando hay grandes avances teóricos y computacionales, el investigador frecuentemente no se percató del error tipo II que su experimento lo conduce a cometer. En todos los experimentos realizados se podría valorar los dos posibles errores que pueden cometerse, con probabilidades  $\alpha$  y  $\beta$ .

Según Neter (1974), para los estudios de análisis de varianza, como para otros estudios estadísticos, es importante planificar el tamaño de muestra (entiéndase número de repeticiones) necesarios para protegernos de errores tipo I y error tipo II, que pueden ser obtenidos o estimados y de esta forma tener una precisión suficiente. La planificación es necesaria para garantizar un tamaño de muestra suficientemente grande para detectar diferencias importantes con una alta probabilidad. Por otro lado,



tamaños de muestra demasiado grande (más de lo necesario) hacen que el costo del estudio llegue a ser excesivo.

Planificar tamaños de muestra puede ser aproximado controlando los riesgos de cometer error tipo I y error tipo II, para esto se ha de considerar el poder de la prueba de F.

El poder de la prueba viene dado por la siguiente expresión:

$$\text{Potencia} = 1 - \beta = P\{F' > F(1 - \alpha; v_1, v_2) \mid \phi\}$$

En resumen, el número de réplicas necesario según Kuehl (2001), está influido primordialmente por cuatro factores que se requieren para los cálculos:

- ✓ La varianza ( $\sigma^2$ )
- ✓ El tamaño de la diferencia entre las dos medias ( $\delta$ )
- ✓ El nivel de significancia de la prueba ( $\alpha$ ), o la probabilidad del error tipo I
- ✓ La potencia de la prueba  $1 - \beta$ , o la probabilidad de detectar  $\delta$ , donde  $\beta$  es la probabilidad del error tipo II.

El mismo autor señala que el número de réplicas necesario generalmente aumenta si:

- ✓ aumenta la varianza, %CV
- ✓ disminuye el tamaño de la diferencia entre dos medias, %  $\delta$  o  $\delta$
- ✓ disminuye el nivel de significancia de la prueba,  $\alpha$

✓ aumenta la potencia de la prueba,  $1-\beta$

- **Control Local:** este principio se refiere a la cantidad de balances, bloqueo y agrupamiento de las unidades experimentales que se utilizan en un diseño experimental. El control local permite disminuir el error experimental con el control de uno o varios factores o fuentes de variabilidad que de no controlarse formarían parte del error experimental y por consiguiente abultarían dicho error (Chacín, 2000).

Además es todo manejo que se realice para controlar fuentes de variación extrañas.

Según Kish (1995), al colocar las variables perturbadoras dentro del tipo controladas (C), eliminamos los efectos que tendrían que estar dentro del tipo aleatorizadas (A). Si bien los errores aleatorios quedan dentro del tipo A, en los diseños de experimentos se eliminan los sesgos debidos al tipo P.

### ***Criterios del diseño de Investigación***

Kish (1995) declara los criterios del diseño de Investigación como: representatividad, realismo y aleatorización.

#### ➤ **Aleatorización y Representatividad**

La *aleatorización* según Brakel and Renssen (1998) es introducida porque una muestra probabilística es observada en lugar de la población objetivo completa. El concepto de la selección aleatoria es aplicado a

fin de hacer inferencias estadísticas sobre los parámetros de la población finita y generalizar sus resultados a la población objetivo de la cual se tomó la muestra.

Aunado a lo anterior, Kish (1995) analiza las necesidades de *aleatorización de los tratamientos y de representatividad*, como dos aspectos íntimamente relacionados del diseño de investigación. Considera que más que una relación, ambos requisitos tienen raíces teóricas y filosóficas comunes. Las cuales suelen tratarse por separado; el hecho de que los dos tipos de aleatorización tengan raíces teóricas comunes no es algo generalmente aceptado.

Por otro lado el mismo autor señala, que la práctica y la teoría del muestreo de encuestas reconocen que tienen que ocuparse de poblaciones de elementos con valores  $Y_i$  fijos, cuya distribución es desconocida; por lo que no es posible simplemente suponer que sus valores muestrales sean VADII (variables aleatorias distribuidas idéntica e independientemente). Por tanto, utilizamos la aleatorización de la población de elementos a fin de obtener selecciones probabilísticas de elementos en las muestras.

Kish (1995), da razones de por qué aleatorizar:

1. La teoría de la probabilidad y la teoría estadística se basan en las variables aleatorias.

2. Las poblaciones, en cuanto a universos físicos, no le llegan al investigador previamente aleatorizadas.
3. Por tanto, los investigadores deben llevar acabo lo que no hizo la naturaleza: aleatorizar sus muestras.
4. Esto sólo puede lograrse por medio de una aleatorización mecánica, como es el caso de las tablas de números aleatorios.
5. De esta manera, el punto 4 es consecuencia de que las elecciones al azar de un seleccionador humano no son aleatorias, aunque lo intente con sinceridad.

El autor comenta que se debe admitir, dos justificaciones básicas para la ausencia de una *representatividad* amplia en los experimentos. La primera es que resulta extremadamente difícil de conseguir, por lo que las inferencias estadísticas derivadas del contraste experimental de diversos tratamientos se limitan a la población o poblaciones incluidas en el diseño experimental. La segunda consiste en que muchos experimentos arrojan resultados útiles incluso a partir de pequeños grupos aislados.

Kish (1995) afirma que la experimentación es potente en cuanto al control de las variables explicativas por medio de la aleatorización de las variables predictoras en los sujetos (es decir, la asignación de los sujetos a los tratamientos); pero es débil en cuanto a la representatividad con respecto a las poblaciones objetivo definidas, y muchas veces también en cuanto al realismo de las mediciones. Las encuestas tienen su lado

fuerte en la representatividad, y su lado débil en el control de las variables.

### ➤ **Realismo**

Kish (1995) menciona, que en muchos casos, en el diseño de experimentos no es factible reproducir el *realismo* de un “entorno natural”.

En las encuestas las muestras probabilísticas no sólo permiten inferencias estadísticas claras a poblaciones definidas, sino que frecuentemente pueden realizarse las mediciones en los “entornos naturales” de poblaciones reales.

En aras del *realismo* y de la exhaustividad, Kish (1995) considera que se debería agregar el ambiente (o condiciones, situación, etc.) como factores condicionantes de las relaciones entre variables. Se formularía entonces el condicionamiento (dependencia) de las relaciones por los elementos (sujetos) y por el ambiente como:

(predictoras → pronosticadas) / elemento, ambiente

ó

pronosticadas = f(predictoras / elemento, ambiente)

Para simplificar, los efectos del ambiente pueden esconderse bajo el manto de las predictoras, o pueden incluirse en la definición de los elementos, pero debemos recordar siempre que los mismos elementos

pueden tener diferentes relaciones predictora-pronosticada según las distintas circunstancias.

➤ **Control Local o Control de Variables**

Autores como Fienberg and Tanur (1983) y Brakel and Renssen (1998), sugieren que dentro de los criterios de muestreo debe considerarse el control local.

Fienberg and Tanur (1983) coinciden en que el *diseño experimental busca minimizar el error a través del bloqueo*, donde en cada bloque se tiene un grupo homogéneo (la forma más simple del control local). Similarmente los diseños de encuestas por muestreo intentan minimizar el error de muestreo mediante la estratificación (formación de grupos homogéneos internamente pero diferentes entre ellos).

De acuerdo con lo anterior Brakel and Renssen (1998) mencionan que la *estratificación y el muestreo de lattices* son empleados para incrementar la precisión de los estimadores por la exclusión de la variación entre grupos homogéneos del error de muestreo. Estas técnicas *son similares al bloqueo en el diseño de bloques al azar y el cuadrado latino* (en el cual se realiza un doble bloqueo), respectivamente.

## 2. OTROS CONTRASTES DE INTERÉS ENTRE EL DISEÑO DE EXPERIMENTOS Y EL MUESTREO

### **Objetivos:**

El objetivo *fundamental de la estadística* es hacer inferencias acerca de una población con base en la información contenida en una muestra (Sheaffer, Mendenhall y Ott; 1987).

### **Objetivos del diseño de experimentos:**

Según Chacín (2000) el *objetivo del diseño de experimentos* es obtener información cuantitativa importante en cuanto a la comparación de los efectos de diferentes tratamientos sobre la variable respuesta en estudio.

Kish (1995) indica que los *diseños de experimentos* se han desarrollado para verificar y determinar las variables explicativas y para medir las relaciones entre ellas en los contrastes analíticos con los datos. De manera distinta y diferenciada, la *teoría del muestreo* de encuestas se ha desarrollado principalmente para aportar estadísticas descriptivas, medias, proporciones y totales, especialmente para grandes muestras extraídas de poblaciones mucho más grandes.

### **Objetivos del Muestreo:**

Brakel and Renssen (1998) señalan que el *objetivo del muestreo por encuestas* es reunir información sobre una población finita, a través de la

estimación de parámetros de la población, tales como: promedios, totales y proporción.

Por otro lado, Neyman (1934), citado por Brakel and Renssen (1998), presenta *el muestreo como un método que permite* obtener estimaciones válidas de parámetros de la población, basados en una muestra representativa, lo cual es una técnica más rápida y de menor costo.

Según Cochran (1985) *el propósito de la teoría del muestreo* es desarrollar métodos de selección de muestras y de estimación, que proporcionen, al menor costo posible, estimaciones con la suficiente exactitud para nuestros propósitos. Este principio de exactitud específica debe corresponder a un costo mínimo.

Kish (1995) considera, que en la práctica, se carece generalmente de los recursos necesarios para superar todas las dificultades, y por tanto, para conseguir simultáneamente la perfección en el realismo de las medidas, en la aleatorización para controlar los tratamientos y en la representatividad de las grandes poblaciones. A menudo, incluso casi siempre, no se pueden satisfacer los tres criterios (aleatorización, representatividad y realismo) simultáneamente



### **Validez externa e interna:**

Kish (1995), indica que algunos autores opinan que siempre que sea posible, debe lograrse a cualquier precio el carácter aleatorio de los tratamientos (o “*validez interna*”), antes de prestar atención a la representatividad con respecto a las poblaciones (o “*validez externa*”), o al realismo. Pero Kish considera, que no existe un supercriterio que pueda adjudicar la superioridad única sobre los demás; más bien hay que escoger y comprometerse con una estrategia de investigación, y en algunos problemas de investigación han de abordarse por separado mediante dos o incluso tres métodos de investigación (diseño de experimento, muestreo y estudios observacionales).

Brakel and Renssen (1998) resumen que los métodos estadísticos del diseño de experimento intentan garantizar principalmente la validez interna suficiente del experimento, y estimar las diferencias entre los efectos de tratamientos tan preciso como sea posible. La validez interna de un experimento es definida como la magnitud a la cual los efectos observados en un experimento pueden atribuirse a las diferencias entre los tratamientos. Ello señala la relación causa – efectos entre los tratamientos y los efectos observados dentro del experimento mismo.

Acotan que los *Métodos estadísticos del diseño de experimentos* intentan principalmente garantizar una validez interna suficiente del experimento, y estimar diferencias en los efectos de tratamientos tan precisamente como sea posible.

Además, sugieren los autores que los métodos estadísticos de la teoría del muestreo intentan garantizar principalmente una validez externa suficiente de una encuesta, por ejemplo, la medida a la cual los resultados del muestreo pueden generalizarse a la población objetivo.

### **Modelos estadísticos:**

Särndal et al (1992), citado por Brakel and Renssen (1998), señalan que los *modelos estadísticos* tradicionalmente han jugado un menor rol en el análisis de las encuestas por muestreo. En el modelo se asume que el valor de cada elemento de la población finita con respecto a la variable objetivo es una realización de la variable estocástica. Esta variable estocástica según Brakel and Renssen (1998), es modelada de acuerdo a un modelo de regresión lineal, con los valores de la variable auxiliar como covariable.

Brakel and Renssen (1998), expresan que *los Modelos estadísticos del diseño de experimentos* juegan un papel importante en la prueba hipótesis sobre la significación de los parámetros del modelo, los cuales se asume que reflejan los efectos de los tratamientos y sobre la exploración de las relaciones de las variables.

### **Ventajas del diseño de experimentos sobre el muestreo de encuestas:**

Kish (1995), expresa que los diseños de experimentos gozan de tres ventajas principales:

1. Es posible eliminar los sesgos originados por las variables perturbadoras (P) mediante la aleatorización de las variables extrañas.
2. Los controles sobre la introducción y variación de las variables predictoras clarifican tanto la dirección como la naturaleza de la causalidad existente entre las variables predictoras y las pronosticadas. Por el contrario, en el caso de las relaciones halladas en resultados de encuestas no están claras ni la dirección ni la naturaleza de la causalidad.
3. El diseño experimental moderno permite una gran flexibilidad, eficiencia, simetría y una potente manipulación estadística, mientras que el uso analítico de los datos de encuesta presenta difíciles problemas estadísticos.

El autor agrega que las ventajas de los experimentos sobre las encuestas, al permitir un mejor control de los tratamientos, son sólo relativas y no absolutas. El diseño de controles experimentales apropiados no es automático, sino que se trata de un arte que requiere conocimiento científico, creatividad a la hora de plantear el experimento, y capacidad retrospectiva al analizar los resultados. Sin embargo, la distinción entre experimentos y encuestas en cuanto al control es real e importante.

Dado que la falta de control sobre los tratamientos es la principal debilidad de las encuestas, los investigadores de encuestas deberían mejorar la recogida y el uso de variables auxiliares como controles de variables perturbadoras.

En cuanto a las pruebas estadísticas Kish (1995) afirma que se refieren a las medidas estadísticas de la variabilidad aleatoria, ya se trate de pruebas de significación, intervalos de confianza, intervalos de credibilidad o cualquier otra. La función de las pruebas estadísticas consiste en distinguir los efectos explicativos de las variables tipo E, de los efectos aleatorios de las variables de tipo A. En los diseños de experimentos se logra esta separación entre E (variables explicativas) y A (variables aleatorias) mediante la aleatorización de todas las variables extrañas dentro del tipo A, excepto las controladas que están dentro del tipo C, sin que queden, por tanto, términos de perturbación dentro de las del tipo P. Además, expone que las *inferencias estadísticas* derivadas del contraste de diversos tratamientos se limitan a la población o poblaciones incluidas en el diseño experimental, de allí que no los considera representativos.

## RESULTADOS

### 1. MODELO PROPUESTO PARA EL MUESTREO ESTRATIFICADO DOBLE

El modelo lineal aditivo para el muestreo estratificado doble, es de la forma:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk} \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, b \quad k = 1, 2, \dots, N_{ij} \quad (IV.1.1)$$

Donde:  $y_{ijk}$  es la k-ésima observación en el i-ésimo estrato tratamiento y el

j-ésimo estrato bloque

$\mu$  es un efecto común a todas las unidades experimentales

$\alpha_i$  es el efecto del i-ésimo nivel del estrato T (tratamientos)

$\beta_j$  es el efecto del j-ésimo nivel del estrato B (bloques)

$\rho_{ij}$  es el efecto de la interacción del i-ésimo nivel del estrato T y el

j-ésimo nivel del estrato B

$\varepsilon_{ijk}$  es el efecto aleatorio del i-ésimo nivel del estrato T y el j-ésimo

nivel del estrato B, en la k-ésima observación

$N_{ij}$  es el número de observaciones en el estrato formado por el i-ésimo nivel del estrato T y el j-ésimo nivel del estrato B

$N$  es el número total de observaciones  $N = \sum_{i=1}^a \sum_{j=1}^b N_{ij}$

Totales Poblacionales:

Gran total :  $\tau = N\mu$        $\tau = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk} = Y \dots$

Total para la interacción: 
$$\begin{cases} \tau_{ij\cdot} = N_{ij} \bar{Y}_{ij\cdot} \\ \tau_{ij\cdot} = \sum_{k=1}^{N_{ij}} Y_{ijk} = Y_{ij\cdot} \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, b \end{cases}$$

Total para el estrato T: 
$$\begin{cases} \tau_{i\cdot\cdot} = \sum_{j=1}^b N_{ij} \bar{Y}_{i\cdot\cdot} \\ \tau_{i\cdot\cdot} = \sum_{j=1}^b \sum_{k=1}^{N_{ij}} Y_{ijk} = Y_{i\cdot\cdot} \quad i = 1, 2, \dots, a \end{cases} \quad \wedge$$

Total para el estrato B: 
$$\begin{cases} \tau_{\cdot j\cdot} = \sum_{i=1}^a N_{ij} \bar{Y}_{\cdot j\cdot} \\ \tau_{\cdot j\cdot} = \sum_{i=1}^a \sum_{k=1}^{N_{ij}} Y_{ijk} = Y_{\cdot j\cdot} \quad j = 1, 2, \dots, b \end{cases}$$

Las medias, se obtienen a través de  $\tau = N\bar{Y}$ ,       $\tau_{ij\cdot} = N_{ij} \bar{Y}_{ij\cdot}$ ,       $\tau_{i\cdot\cdot} = \sum_{j=1}^b N_{ij} \bar{Y}_{i\cdot\cdot}$  y

$$\tau_{\cdot j\cdot} = \sum_{i=1}^a N_{ij} \bar{Y}_{\cdot j\cdot}$$

Las esperanzas de los cuadrados medios, quedan reflejados en el cuadro 11

**Cuadro 11: Esperanza de los Cuadrados Medios para el Modelo Mixto Estratificado Doble**

FdeV	g.de L.	E(CM)
Est.T	$\alpha-1$	$\sigma^2 + \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \right] \frac{\sigma_p^2}{\alpha-1} + \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \right] \frac{\sigma_\beta^2}{\alpha-1}$ $+ \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij} \alpha_i^2}{\alpha-1}$
Est.B	$b-1$	$\sigma^2 + \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} \right] \frac{\sigma_p^2}{b-1} + \left[ \sum_{i=1}^a \sum_{j=1}^b N_{ij} - \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \right] \frac{\sigma_\beta^2}{b-1}$
Int.	$(\alpha-1)(b-1)$	$\left[ N - \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \sum_{j=1}^b \left( \frac{\sum_{i=1}^a N_{ij}^2}{\sum_{i=1}^a N_{ij}} \right) - \frac{\sum_{i=1}^a \sum_{j=1}^b N_{ij}^2}{N} + 2 \sum_{i=1}^a \sum_{j=1}^b \frac{N_{ij}^3}{\sum_{i=1}^a N_{ij} \sum_{j=1}^b N_{ij}} \right] \frac{\sigma_p^2}{(\alpha-1)(b-1)}$ $+ \left[ \sum_{i=1}^a \left( \frac{\sum_{j=1}^b N_{ij}^2}{\sum_{j=1}^b N_{ij}} \right) - \frac{\sum_{j=1}^b \left( \sum_{i=1}^a N_{ij} \right)^2}{N} \right] \frac{\sigma_\beta^2}{(\alpha-1)(b-1)} + \left[ ab - a - b - 1 + 2 \sum_{i=1}^a \sum_{j=1}^b \frac{N_{ij}^2}{\sum_{i=1}^a N_{ij} \sum_{j=1}^b N_{ij}} \right] \frac{\sigma^2}{(\alpha-1)(b-1)}$
EE	$N-ab$	$\sigma^2$
Total	$N-1$	

### Estimadores de los parámetros a través del muestreo

➤  $\hat{T} = \frac{N}{n} \sum_{i=1}^a \sum_{j=1}^b \sum_{K=1}^{n_{ij}} Y_{ijk}$ , que es un estimador sesgado de T para el modelo, cuya

varianza es:

$$\begin{aligned}
V(\hat{T}) = E_M \left[ (\hat{T} - T)^2 \right] &= b^2 \sum_{i=1}^a \left( \frac{N}{n} n_{ij} - N_{ij} \right)^2 \alpha_i^2 + a^2 \sum_{j=1}^b \left( \frac{N}{n} n_{ij} - N_{ij} \right)^2 \sigma_{\beta_j}^2 \\
&+ \sum_{i=1}^a \sum_{j=1}^b \left( \frac{N}{n} n_{ij} - N_{ij} \right)^2 \sigma_{ij}^2 + \frac{N^2}{n^2} \sum_{i=1}^a \sum_{j=1}^b n_{ij} \sigma_{ij}^2 \\
&- 2 \frac{N}{n} \sum_{i=1}^a \sum_{j=1}^b n_{ij} \sigma_{ij}^2 + \sum_{i=1}^a \sum_{j=1}^b N_{ij} \sigma_{ij}^2
\end{aligned}$$

si el tamaño del estrato formado por los estratos T y B, es constante, es decir,  $N_{ij} = M$  y el tamaño de muestra en los estratos es la misma, es decir,  $n_{ij} = m$ ,

$\hat{T}$  sería insesgado y su varianza  $V(\hat{T}) = \frac{M^2}{m} \left( 1 - \frac{m}{M} \right) \sum_{i=1}^a \sum_{j=1}^b \sigma_{ij}^2$

Además de los estimadores insesgados:

➤  $\hat{T}_{ij\cdot} = \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$ ,  $\hat{T}_{ij\cdot}$ , es un estimador de  $T_{ij\cdot}$ , cuya varianza es:

$$V(\hat{T}_{ij\cdot}) = E_M \left[ (\hat{T}_{ij\cdot} - T_{ij\cdot})^2 \right] = \frac{N_{ij}^2}{n_{ij}} \sigma_{ij}^2 \left( 1 - \frac{n_{ij}}{N_{ij}} \right)$$

➤  $\hat{T}_{i\cdot\cdot} = \sum_{j=1}^b \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$ , es un estimador de  $T_{i\cdot\cdot}$ , con varianza:

$$V(\hat{T}_{i\cdot\cdot}) = E_M \left[ (\hat{T}_{i\cdot\cdot} - T_{i\cdot\cdot})^2 \right] = \sum_{j=1}^b \frac{N_{ij}^2}{n_{ij}} \sigma_{ij}^2 \left( 1 - \frac{n_{ij}}{N_{ij}} \right)$$

➤  $\hat{T}_{\cdot j\cdot} = \sum_{i=1}^a \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}$ , es un estimador de  $T_{\cdot j\cdot}$ , con varianza:

$$V(\hat{T}_{\cdot j\cdot}) = E_M \left[ (\hat{T}_{\cdot j\cdot} - T_{\cdot j\cdot})^2 \right] = \sum_{i=1}^a \frac{N_{ij}^2}{n_{ij}} \sigma_{ij}^2 \left( 1 - \frac{n_{ij}}{N_{ij}} \right)$$



Se ilustra a continuación como funcionan los Modelo de Muestreo dentro de un Diseño de Experimentos, bajo las condiciones del ensayo de Musáceas.

## 2. APLICACIÓN PRÁCTICA DEL MODELO DEL DISEÑO COMPLETAMENTE AL AZAR, EMPLEANDO LOS DATOS DEL ENSAYO EN MUSÁCEAS

### A. Variable Tiempo de Desarrollo de la Enfermedad (TDE):

#### ➤ *Estadística Descriptiva:*

Se presenta un cuadro con las características de la población bajo estudio, en cada una de las variedades evaluadas.

**Cuadro 12: Estadística descriptiva del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, para cada variedad, en el diseño completamente al azar**

Variedad	N	Promedio	Varianza	CV (%)	Valor mín.	Valor Máx.
I	0	-	-	-	-	-
II	57	51.97	137.21	22.54	28	75
III	58	48.57	131.51	23.61	27	74
IV	58	56.22	137.19	20.83	29	82
V	4	85.50	33.000	6.72	78	90
VI	0	-	-	-	-	-

N < 60 (cuadro 12), por la presencia de datos perdidos (plantas muertas)

➤ **Análisis de Varianza:**

Para el ensayo en Musáceas, realizado bajo el diseño completamente al azar, la prueba de medias de Duncan, generó dos grupos A y B (figura 3), estadísticamente diferentes entre sí.

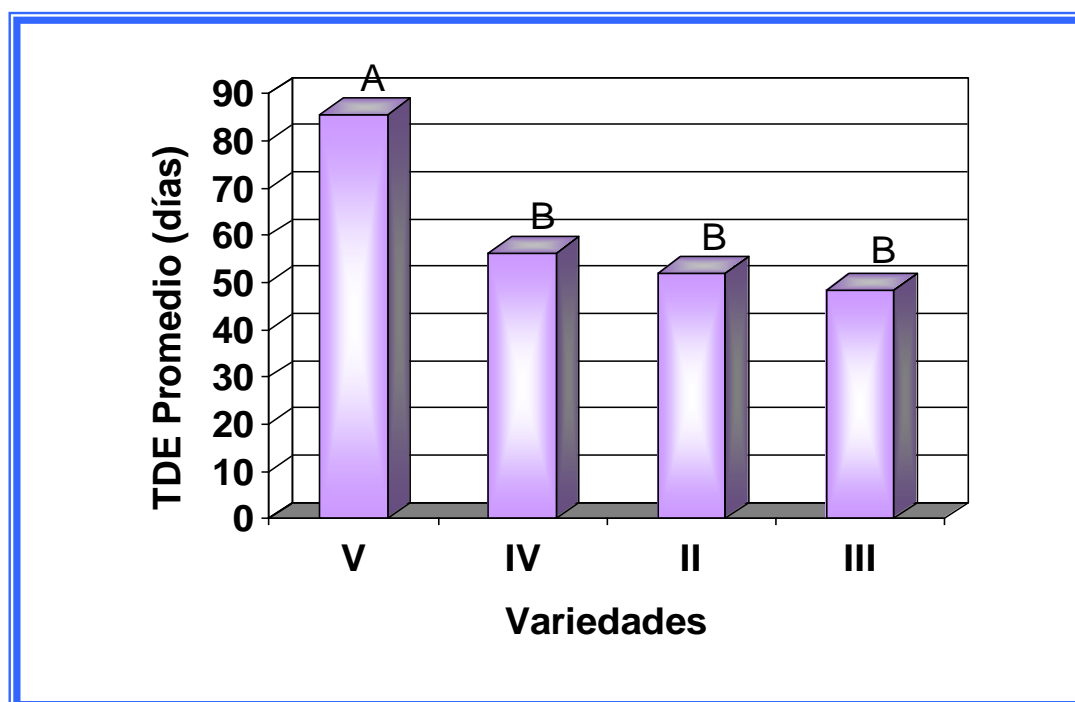
No se observaron diferencias significativas entre las variedades `Gran nain´(II), `Pisang más´(III) y `Brasileiro´(IV), es decir, el TDE fue muy similar entre estas. Pero la variedad `FHIA-02´(V) presentó mayor resistencia, ya que el TDE fue superior; además, solo se observaron 4 plantas de las 60 evaluadas, afectadas por la Sigatoka negra.

Por otro lado, las variedades `FAGRO 122´(I) y `Cien Bta-03´(VI), presentaron resistencia total a la enfermedad.

El análisis de varianza para el TDE se presenta en el cuadro 13

**Cuadro 13: Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (TDE), en Sigatoka Negra, bajo el Diseño Completamente al Azar.**

F. de V.	G de L	SC	CM	Pr
Variedades	3	6027.7541	2009.2514	<0.0001
Error	173	23099.240	133.52162	



**Figura 3. Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades estudiadas, en el diseño completamente al azar**

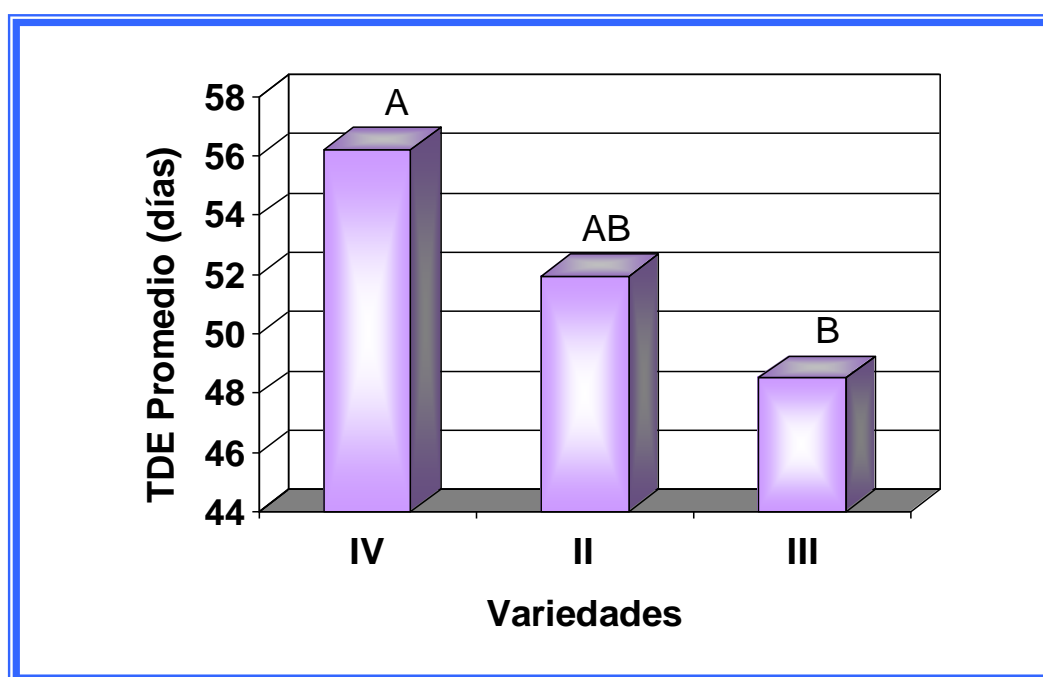
➤ **Análisis de Varianza para las variedades susceptibles:**

Como la variedad V mostró resistencia, se compararon las variedades susceptibles II, III y IV; obteniéndose, a través de la prueba de medias de Duncan tres grupos (figura 4). Sólo difieren estadísticamente las variedades IV y III

El análisis de varianza para el TDE se presenta en el cuadro 14

**Cuadro 14: Análisis de Varianza de variedades susceptibles, para el Tiempo de Desarrollo de la Enfermedad (TDE), en Sigatoka Negra, bajo el Diseño Completamente al Azar.**

F. de V.	G de L	SC	CM	Pr
Variedades	2	1706.569	853.2845	0.0023
Error	170	23000.24	135.2955	



**Figura 4. Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades susceptibles, en el diseño completamente al azar**

#### **B. Variable Número de Frutos:**

##### **➤ Estadística Descriptiva:**

Se presenta un cuadro con las características de la población bajo estudio, en cada una de las variedades evaluadas.

**Cuadro 15: Estadística descriptiva del Número de Frutos (dedos), en Sigatoka Negra, para cada variedad, en el diseño completamente al azar**

Variedad	N	Promedio	Varianza	Valor mín.	Valor Máx.
I	58	173.81	99.67	139	190
II	57	160.63	155.52	140	185
III	58	132.17	85.75	112	154
IV	58	131.12	374.25	95	168
V	53	138.68	162.07	112	165
VI	59	143.49	172.98	118	171

N < 60 (cuadro 15), por la presencia de datos perdidos (plantas muertas)

➤ **Análisis de Varianza:**

Para el ensayo en Musáceas, realizado bajo el diseño completamente al azar, la prueba de medias de Duncan, generó cuatro grupos A, B, C y D (figura 5), estadísticamente diferentes entre sí. No se observaron diferencias significativas entre las variedades V y VI, ni entre las variedades III y IV, que son las que desarrollan menos dedos. La variedad I, es la que desarrolló más frutos. El análisis de varianza para el Número de frutos se presenta en el cuadro 16

**Cuadro 16: Análisis de Varianza, para el Número de frutos (dedos), bajo el Diseño Completamente al Azar.**

F. de V.	G de L	SC	CM	Pr
Variedades	5	84026.668	16805.33	<0.0001
Error	337	59070.901	175.28	

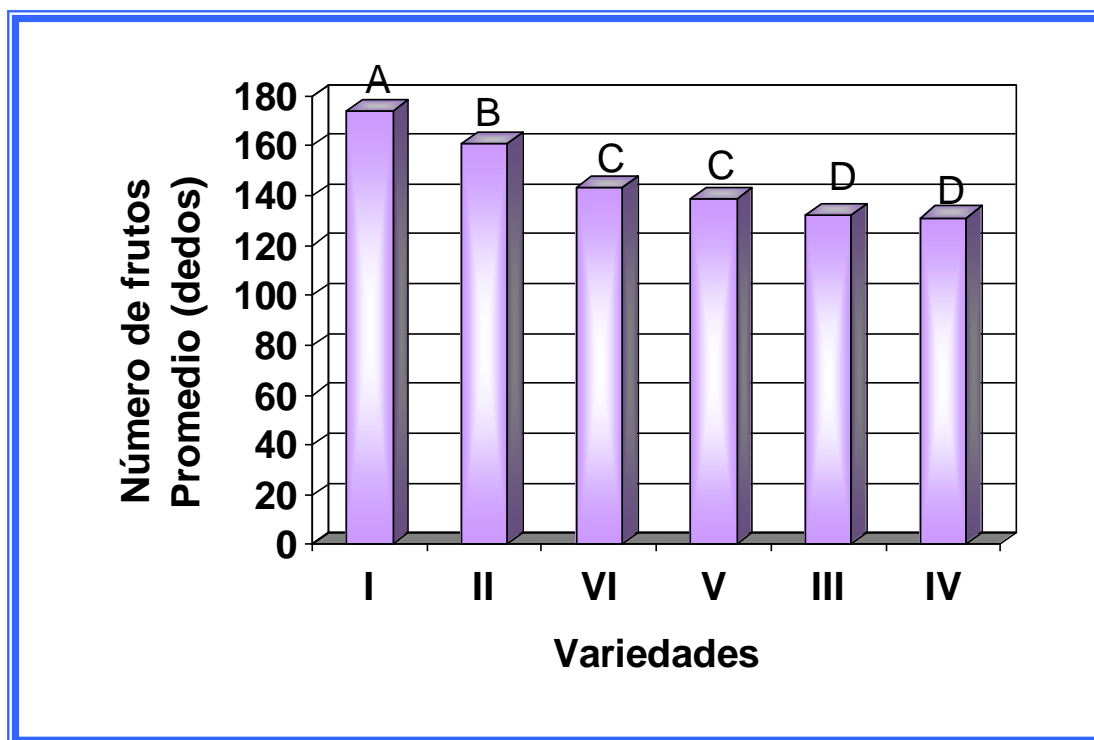


Figura 5. Promedios de Números de Frutos (dedos), para las variedades estudiadas, en el diseño completamente al azar

### 3. APLICACIÓN PRÁCTICA DEL MODELO DEL DISEÑO DE BLOQUES AL AZAR, EMPLEANDO LOS DATOS DEL ENSAYO EN MUSÁCEAS

#### A. Variable Tiempo de Desarrollo de la Enfermedad (TDE):

##### ➤ *Estadística Descriptiva*

Las características de la población bajo estudio, en cada una de las variedades evaluadas, dentro de cada bloque se presentan en el cuadro. Y los resultados por variedad se presentan en el cuadro 17, ya que se emplearon las mismas observaciones para ambos diseños.

**Cuadro 17: Estadística descriptiva del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, para las variedades dentro de cada bloque, bajo el diseño de bloques al azar**

Variedad	Bloque	N	Promedio	Varianza	Valor mín.	Valor Máx.
I		0	-	-	-	-
II	1	12	65.67	28.97	56	75
	2	11	46.91	163.09	28	72
	3	12	45.33	122.06	29	67
	4	12	58.67	14.97	48	62
	5	10	43.40	18.489	36	48
III	1	11	63.09	69.29	48	74
	2	12	50.08	92.08	36	67
	3	11	44.91	68.89	34	59
	4	12	45.83	79.606	34	63
	5	12	39.83	65.97	27	56
IV	1	12	70.42	36.63	62	82
	2	11	55.27	64.22	39	67
	3	12	51.92	71.17	29	62
	4	12	58.25	71.48	45	72
	5	11	44.18	91.16	32	63
V	1	4	85.5	33	78	90
	2	0	-	-	-	-
	3	0	-	-	-	-
	4	0	-	-	-	-
	5	0	-	-	-	-
VI		0	-	-	-	-

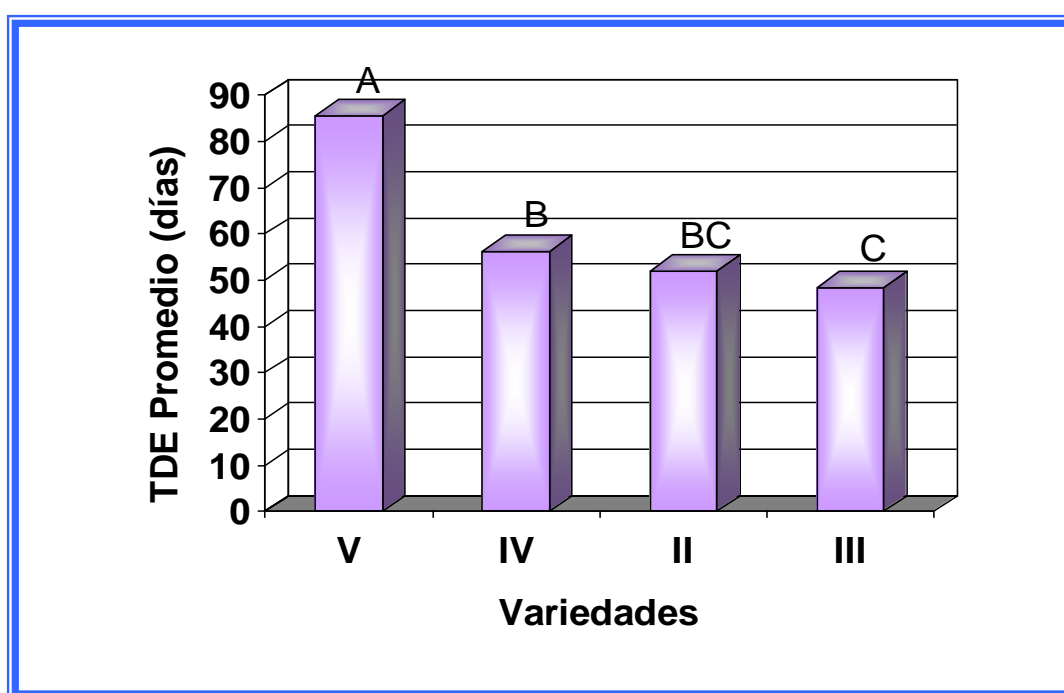
En los casos donde  $N < 12$  (cuadro 17), es por la presencia de datos perdidos (plantas muertas)

➤ **Análisis de Varianza:**

Para el ensayo en Musáceas, realizado bajo el diseño de bloques al azar, la prueba de medias de Duncan, generó cuatro grupos A, B, BC y C (figura 6), estadísticamente diferentes entre sí.

Como puede observarse en la figura 6, la variedad II se comportó similar a la variedad IV y a la III, pero las variedades III y IV si difieren estadísticamente entre sí. Por otro lado, la variedad V presentó mayor resistencia, ya que el TDE fue superior; además, solo se observaron 4 plantas de las 60 evaluadas, afectadas por la Sigatoka negra.

Las variedades `FAGRO 122'(I) y `Cien Bta-03'(VI), presentaron resistencia total a la enfermedad.



**Figura 6. Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades estudiadas, en el diseño de Bloques al Azar**

El análisis de varianza para el TDE se presenta en el cuadro 18

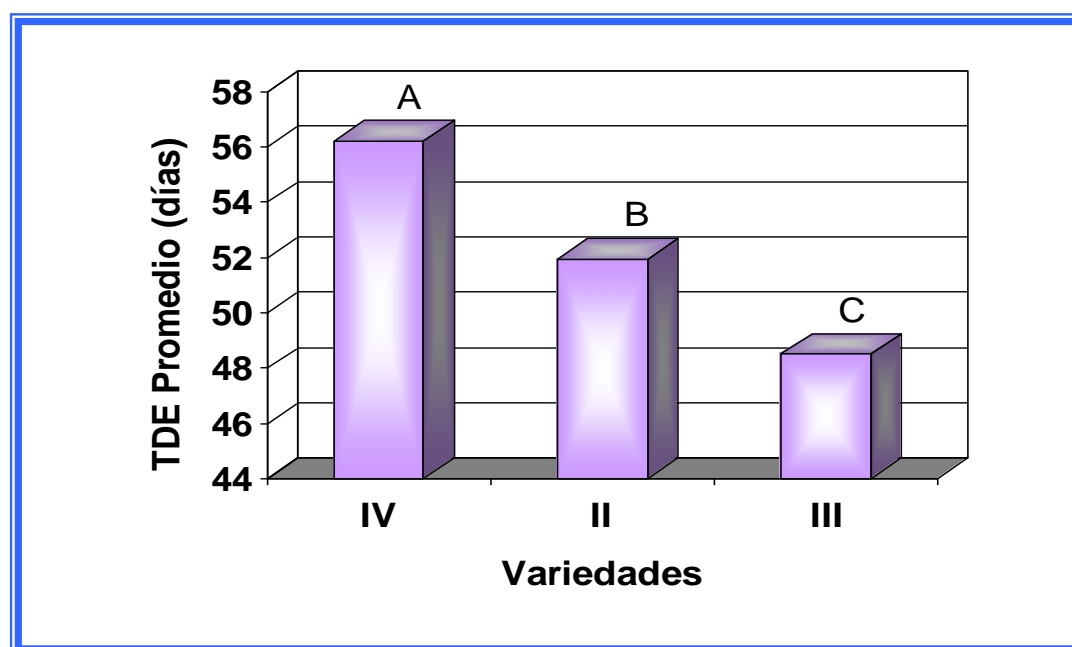


**Cuadro 18: Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (días), bajo el Diseño de Bloques al Azar.**

F. de V.	G de L	SC	CM	Pr
Variedades	3	2866.327	955.4423	<0.0001
Bloques	4	11126.28	2781.572	<0.0001
Error	169	11972.95	70.84587	

➤ **Análisis de Varianza para las variedades susceptibles:**

Como la variedad V mostró resistencia, se compararon las variedades susceptibles II, III y IV; obteniéndose, a través de la prueba de medias de Duncan tres grupos (figura 7). Las tres variedades son estadísticamente diferentes



**Figura 7. Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades susceptibles, en el diseño de bloques al azar**

El análisis de varianza para el TDE de las variedades susceptibles se presenta en el cuadro 19

**Cuadro 19: Análisis de Varianza de variedades susceptibles, para el Tiempo de Desarrollo de la Enfermedad (TDE), en Sigatoka Negra, bajo el Diseño de Bloques al Azar.**

F. de V.	G de L	SC	CM	Pr
Variedades	2	1568.481	784.2401	<0.0001
Bloques	4	11126.29	2781.572	<0.0001
Error	166	11873.95	71.52983	

## **B. Variable Número de Frutos (dedos):**

### ➤ **Estadística Descriptiva**

Las características de la población bajo estudio, en cada una de las variedades evaluadas, dentro de cada bloque se presentan en el cuadro 20. Y los resultados por variedad se presentan en el cuadro 15, ya que se emplearon las mismas observaciones para ambos diseños.

### ➤ **Análisis de Varianza:**

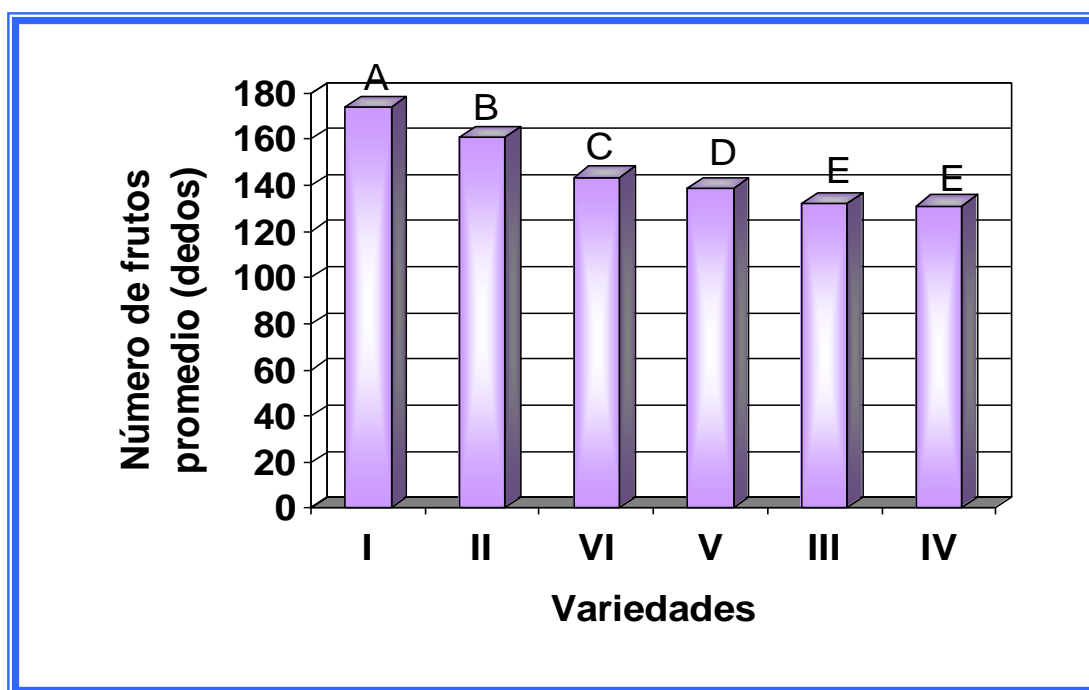
Para el ensayo en Musáceas, realizado bajo el diseño completamente al azar, la prueba de medias de Duncan, generó cinco grupos A, B, C, D y E (figura 8), estadísticamente diferentes entre sí.

**Cuadro 20: Estadística descriptiva del Número de frutos (dedos), para las variedades, en cada bloque, bajo el diseño de Bloques al Azar**

Variedad	Bloque	N	Promedio	Varianza	Valor mín.	Valor Máx.
I	1	12	171.75	255.84	139	187
	2	12	176.08	46.45	160	184
	3	11	174.18	70.96	158	186
	4	11	170.36	60.86	160	182
	5	12	176.42	64.63	164	190
II	1	12	157.17	185.24	140	184
	2	11	160.36	50.26	150	172
	3	12	151.50	88.27	143	175
	4	12	162.67	153.88	145	184
	5	10	173.60	69.82	158	185
III	1	11	129.36	49.06	118	140
	2	12	127.00	21.27	120	136
	3	11	126.91	83.69	112	144
	4	12	134.83	60.52	122	146
	5	12	142.08	62.27	130	154
IV	1	12	126.17	343.24	95	155
	2	11	131.55	371.27	104	161
	3	12	134.17	581.42	100	168
	4	12	118.25	116.75	103	137
	5	11	146.82	105.76	128	160
V	1	12	130.33	9.69	125	134
	2	11	126.55	92.87	112	140
	3	10	136.30	33.57	127	144
	4	10	154.70	78.90	142	165
	5	10	148.40	39.60	138	157
VI	1	11	132.73	49.62	122	146
	2	12	150.17	124.15	132	170
	3	12	135.42	114.27	118	152
	4	12	143.33	113.70	126	160
	5	12	154.92	136.81	135	171

En los casos donde  $N < 12$  (cuadro 20), es por la presencia de datos perdidos (plantas muertas)

No se observaron diferencias significativas entre las variedades `Pisang más´(III) y `Brasilero´(IV), las cuales desarrollan menos dedos. La variedad `FAGRO 122´(I), es la que desarrolló mas frutos, seguido por la variedad II.



**Figura 8. Promedios de Números de Frutos (dedos), para las variedades estudiadas, en el diseño de bloques al azar**

El análisis de varianza para el Número de frutos se presenta en el cuadro 21

**Cuadro 21: Análisis de Varianza, para el Número de frutos (dedos), bajo el Diseño de Bloques al Azar.**

F. de V.	G de L	SC	CM	Pr
Variedades	5	84203.425	16840.6849	<0.0001
Bloques	4	10080.614	2520.15338	<0.0001
Error	333	48990.2874	147.1180	

#### 4. APLICACIÓN PRÁCTICA DEL MODELO DE MUESTREO ALEATORIO SIMPLE, EMPLEANDO LOS DATOS DEL ENSAYO EN MUSÁCEAS

##### A. Variable Tiempo de Desarrollo de la Enfermedad (TDE):

La aplicación de la fórmula (1) presentada en la Metodología para la determinación del tamaño de muestra, en este trabajo, conlleva al siguiente desarrollo:

Tal como se señaló en el ítem correspondiente a la Metodología, cuando la cuasivarianza se desconoce, ésta debe conjeturarse a partir de una muestra piloto. Pero, para efectos de este trabajo, ya que se cuenta con toda la información del experimento, se empleó la obtenida para todas las observaciones, en cada variedad, así:

Para esta variable TDE; dado que los resultados del experimento demostraron que dos (2) variedades (variedad I: 'Fagro 122' y la variedad VI: 'Cien Bta-03'), fueron resistentes a la enfermedad, cuya varianza fue nula. Se determinó el tamaño de muestra para el resto de las variedades aplicando la fórmula (1).

$$\text{Variedad II: } S_2^2 = 137.21$$

e=4 días; de acuerdo a información suministrada por el Profesor Oscar Haddad. Por lo tanto

$$n_{II\infty} = \frac{2^2(137.21)}{4^2} = 34.3025 \approx 35$$

$$\frac{35}{60} = 0.58 > 0.05, \text{ al redefinir se obtuvo}$$

$$n = \frac{34.3025}{1 + \frac{34.3025}{60}} = 21.82 \approx 22 \text{ plantas para la variedad II}$$

Se aplicó el mismo procedimiento para cada una de las variedades y se obtuvo los tamaños de muestra presentados en el cuadro 22. Es importante señalar que se determinó el tamaño de muestra en cada variedad, ya que para el caso del Muestreo Aleatorio Simple se realizó el análisis de cada variedad por separado y en el Muestreo Estratificado Doble se consideraron como estratos los bloques, para obtener una mejor estimación, debido a que si se emplea el  $CM_{EE}$  para el Diseño de Bloques al azar, en la determinación de tamaño de muestra, la misma resultaría muy pequeña.

**Cuadro 22: Tamaños de muestra, para cada una de las variedades, en la variable Tiempo de Desarrollo de la Enfermedad (días), con un error de 4 días**

Variedad	$S^2$ (días <sup>2</sup> )	n (plantas)
II	137.21	22
III	131.51	22
IV	137.19	22

De acuerdo a los resultados del cuadro 22, se seleccionaron 22 plantas para cada variedad, incluyendo a las variedades I, V y VI. Para este caso del

Muestreo Aleatorio Simple se enumeraron las plantas del 1 al 60 en cada variedad y se seleccionó la muestra (en el Apéndice C, se presenta en forma gráfica la selección de plantas por variedad dentro del experimento) empleando la tabla de números aleatorios de Cochran (1985).

Los resultados obtenidos de la aplicación del Modelo de Muestreo Aleatorio Simple en la muestra seleccionada según lo anterior, se presentan a continuación:

➤ **Estadística Descriptiva:**

Se presenta el cuadro 23 con las estimaciones, en cada una de las variedades evaluadas.

**Cuadro 23: Estimaciones de la media y varianza del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, por variedad, para el modelo del muestreo aleatorio simple**

Variedad	n	$\bar{y}_i$	$s_i^2$	Valor mín.	Valor Máx.
I	0	-	-	-	-
II	22	50.818	154.7287	28	75
III	22	49.227	148.6692	34	74
IV	20	59.400	112.7844	38	79
V	1	84.000	-	84	84
VI	0	-	-	-	-

En los casos donde  $n < 22$  (cuadro 23), es por la presencia de datos perdidos (plantas muertas) o en el caso específico de la variedad V, se debe

a que la enfermedad, se desarrolló sólo en una planta de las que se evaluaron en la muestra.

Para las variedades I y VI, no hubo presencia de la enfermedad en la muestra

### **B. Variable Número de Frutos:**

Se determinó el tamaño de muestra para cada variedad aplicando la fórmula (1), presentada en la Metodología.

Variedad I:  $S_1^2 = 99.665$

$e=6$  frutos; de acuerdo a información suministrada por el Profesor Oscar Haddad. Por lo tanto

$$n_{1\infty} = \frac{2^2 (99.665)}{6^2} = 11.0738 \approx 12$$

$$\frac{12}{60} = 0.20 > 0.05, \text{ al redefinir se obtuvo}$$

$$n = \frac{11.0738}{1 + \frac{11.0738}{60}} = 9.35 \approx 10$$

Se aplicó el mismo procedimiento para cada una de las variedades y se obtuvo los tamaños de muestra presentados en el cuadro 24.



**Cuadro 24: Tamaños de muestra, para cada una de las variedades, en la variable Número de frutos (dedos) con un error de 6 dedos**

Variedad	$S^2$ (dedos <sup>2</sup> )	n (plantas)
I	99.665	10
II	155.52	14
III	85.759	9
IV	374.25	25
V	162.07	14
VI	172.98	15

De acuerdo a los resultados obtenidos en el cuadro 24, se enumeraron las plantas del 1 al 60 en cada variedad y se seleccionó la muestra de plantas (en el Apéndice C, se presenta en forma gráfica la selección de plantas por variedad dentro del experimento) empleando la tabla de números aleatorios de Cochran (1985).

Con la muestra seleccionada según el cuadro 24, se obtuvo el siguiente resultado:

➤ **Estadística Descriptiva:**

Se presenta el cuadro 25 con las estimaciones, en cada una de las variedades evaluadas.

**Cuadro 25: Estimaciones de la media y varianza del Número de Frutos (dedos), por variedad, para el modelo del muestreo aleatorio simple**

Variedad	n	$\bar{y}_i$	$s_i^2$	Valor mín.	Valor Máx.
I	10	171.50	55.3893	160	184
II	14	160.00	178.623	142	184
III	9	135.67	135.746	122	154
IV	24	130.63	341.473	103	155
V	11	141.91	335.696	112	165
VI	15	140.33	150.946	118	158

En los casos donde n es menor a los presentados en el cuadro 25, es por la presencia de datos perdidos (plantas muertas), en la muestra.

En vista de que las seis variedades evaluadas poseen características diferentes en las variables de producción y desarrollo, en la práctica es recomendable determinar para el **Número de Frutos**, el tamaño de muestra, considerando la mayor varianza entre las seis variedades.

Por ello, basándose en el cuadro 24, donde se puede observar que la variedad IV es la que presenta la mayor varianza y por ende el mayor tamaño de muestra, se seleccionó  $n=25$  en cada variedad. La disposición de las plantas seleccionadas, se muestra en el Apéndice C.

Los resultados obtenidos con esta muestra mayor son los siguientes:

➤ **Estadística Descriptiva considerando la mayor varianza:**

Se presenta el cuadro 26 con las estimaciones, en cada una de las variedades evaluadas.

**Cuadro 26: Estimaciones de la media y varianza del Número de Frutos (dedos), por variedad, para el modelo del muestreo aleatorio simple, considerando  $n=25$**

Variedad	n	$\bar{y}_i$	$s_i^2$	Valor mín.	Valor Máx.
I	25	172.32	111.39	139	190
II	24	160.25	156.98	142	184
III	25	133.92	86.660	122	154
IV	24	130.63	341.46	103	155
V	21	140.24	204.09	112	165
VI	25	144.28	198.88	118	170

En los casos donde n es menor a 25, es por la presencia de datos perdidos (plantas muertas), en la muestra.

## 5. APLICACIÓN PRÁCTICA DEL MODELO DE MUESTREO ESTRATIFICADO ALEATORIO

### A. Variable Tiempo de Desarrollo de la Enfermedad (TDE):

Se consideraron las variedades como estratos, por lo que se empleó el  $CM_{EE}$  obtenido con el diseño completamente al azar y se determinó el tamaño de muestra por la fórmula (1), de la siguiente manera:

e=4 días; de acuerdo a información suministrada por el Profesor Oscar Haddad. Por lo tanto

$$n_{||\infty} = \frac{2^2(133.52162)}{4^2} = 33.38 \approx 34$$

$$\frac{34}{360} = 0.094 > 0.05, \text{ al redefinir se obtuvo}$$

$$n = \frac{33.38}{1 + \frac{33.38}{360}} = 30.55 \approx 31 \text{ plantas}$$

La afijación de la muestra se llevó a cabo empleando la afijación igual, seleccionando así, 6 plantas en cada variedad, de tal manera de seleccionar una muestra que cubriera el tamaño de muestra mínimo obtenido de 31 plantas. La disposición de las plantas se muestra en el Apéndice E.

Los resultados de la aplicación del Modelo de Muestreo Estratificado Aleatorio en la muestra seleccionada según lo anterior, se presentan a continuación:

### ➤ **Estadística Descriptiva**

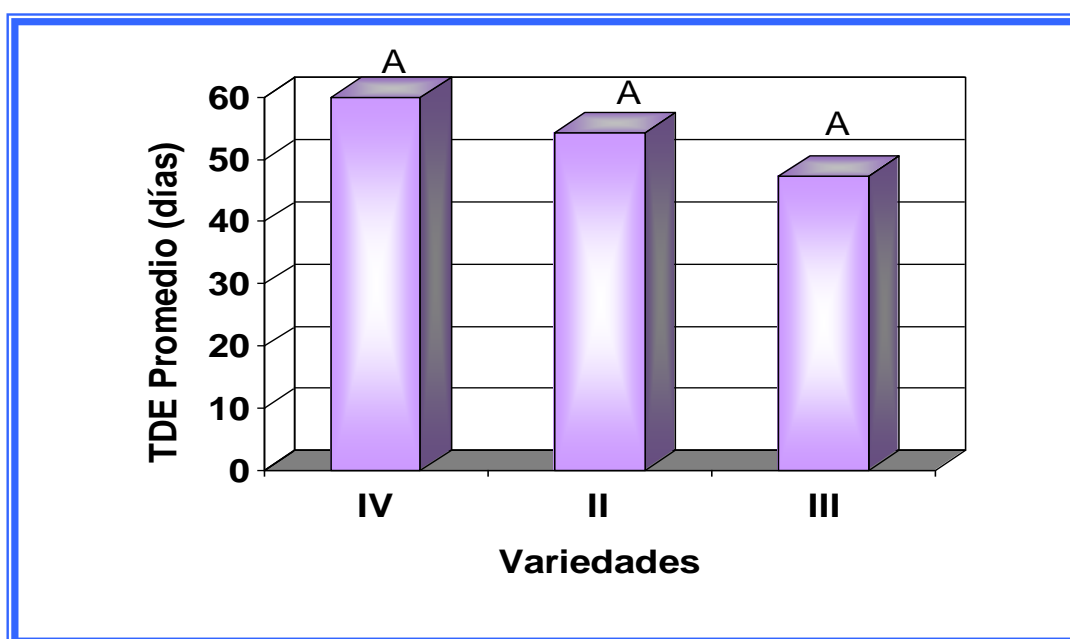
Se presenta el cuadro 27 con las estimaciones, en cada una de las variedades evaluadas.

**Cuadro 27: Estimaciones de la media y varianza del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, por variedad, para el modelo del muestreo estratificado aleatorio**

Variedad	n	$\bar{y}_i$	$s_i^2$	Valor mín.	Valor Máx.
I	0	-	-	-	-
II	6	54.333	60.6669	43	61
III	6	47.333	218.655	36	74
IV	6	60.000	199.996	39	82
V	0	-	-	-	-
VI	0	-	-	-	-

➤ **Análisis de Varianza:**

Para el ensayo en Musáceas, realizado bajo el modelo de muestreo estratificado aleatorio, las variedades II, III y IV presentaron un TDE similares, estadísticamente iguales (figura 9)



**Figura 9. Promedios de Tiempo de Desarrollo de la Enfermedad (días), para las variedades estudiadas, en el Modelo de Muestreo Estratificado Aleatorio**

Se puede observar en el cuadro 28 el análisis de varianza para el TDE, siendo los Cuadrados Medios (CM), de interés particular para este trabajo.

**Cuadro 28: Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad, bajo el Modelo de Muestreo Estratificado Aleatorio**

F. de V.	G de L	SC	CM	Pr
Variedades	2	483.1111	241.5556	0.2523
Error	15	1879.778	159.7778	

#### **B. Variable Número de frutos (dedos):**

Se consideraron las variedades como estratos, por lo que se empleó el  $CM_{EE}$  obtenido con el diseño completamente al azar para el número de frutos y se determinó el tamaño de muestra por la fórmula (1), de la siguiente manera:

$e=6$  dedos; de acuerdo a información suministrada por el Profesor Oscar Haddad. Por lo tanto

$$n_{\parallel \infty} = \frac{2^2 (175.2846)}{6^2} = 19.48 \approx 20$$

$$\frac{20}{360} = 0.055 > 0.05, \text{ al redefinir se obtuvo}$$

$$n = \frac{19.48}{1 + \frac{19.48}{360}} = 18.45 \approx 19 \text{ plantas}$$

La afijación de la muestra  $n=19$  se llevó a cabo empleando la afijación de Neyman, con la siguiente expresión

$$n_{ij} = \frac{N_{ij}S_{ij}}{\sum_{H=1}^H N_{ij}S_{ij}} n$$

De la cual se obtuvo los tamaños de muestra a seleccionar en cada variedad, como se muestra en el cuadro 29.

**Cuadro 29: Afijación de la muestra en el muestreo estratificado aleatorio en cada estrato tipo T (variedades), para la variable Número de Frutos.**

	Estrato T						
	I	II	III	IV	V	VI	N
$n_{ij}$	3	3	2	5	3	3	19

La disposición de las plantas seleccionadas, se muestra en el Apéndice D.

Los resultados obtenidos de la aplicación del Modelo de Muestreo Estratificado Aleatorio se presentan a continuación:

### ➤ **Estadística Descriptiva**

Se presenta el cuadro 30 con el promedio y varianza estimados, además de los valores mínimo y máximo, para cada variedad.

**Cuadro 30: Estimaciones de la media y varianza del Número de Frutos (dedos) en Sigatoka Negra, por variedad, para el modelo del muestreo estratificado aleatorio**

Variedad	n	$\bar{y}_i$	$s_i^2$	Valor mín.	Valor Máx.
I	3	170.00	100.000	160	180
II	3	159.33	121.330	148	170
III	2	133.00	162.001	124	142
IV	5	135.60	616.777	104	168
V	3	154.00	219.010	137	164
VI	3	133.33	82.3338	125	143

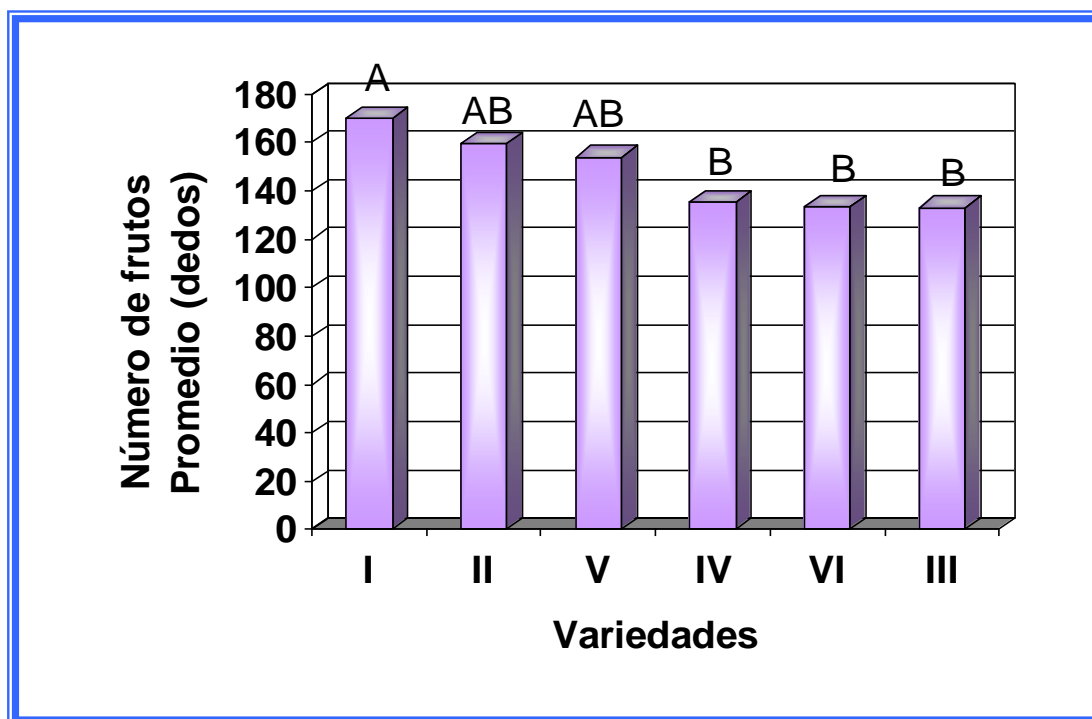
En los casos donde n es menor a 5, es por la presencia de datos perdidos (plantas muertas), en la muestra.

➤ **Análisis de Varianza:**

Para el ensayo en Musáceas, realizado bajo el modelo de muestreo estratificado aleatorio, la prueba de Duncan, señaló tres grupos (figura 10).

La variedad I presentó un mayor número de frutos y las variedades III, VI y IV, la menor cantidad, respectivamente





**Figura 10. Promedios del Número de Frutos (dedos), para las variedades estudiadas, en el Modelo de Muestreo Estratificado Aleatorio**

Se puede observar en el cuadro 31 el análisis de varianza para el Número de Frutos, siendo los Cuadrados Medios (CM), de interés particular para este trabajo.

**Cuadro 31: Análisis de Varianza, para el Número de Frutos (dedos), bajo el Modelo de Muestreo Estratificado Aleatorio**

F. de V.	G de L	SC	CM	Pr
Variedades	5	3792.414	758.4828	0.0704
Error	13	3674.533	262.6564	

Considerando que las seis variedades evaluadas poseen características diferentes en cuanto al **Número de Frutos**, se obtuvo el tamaño de muestra, considerando la mayor varianza entre las seis variedades.

e=6 dedos; de acuerdo a información suministrada por el Profesor Oscar Haddad. Por lo tanto

$$n_{11\infty} = \frac{2^2(374.25)}{6^2} = 41.58 \approx 42$$

$$\frac{42}{360} = 0.1167 > 0.05, \text{ al redefinir se obtuvo}$$

$$n = \frac{41.58}{1 + \frac{41.58}{360}} = 37.27 \approx 38 \text{ plantas}$$

La afijación de la muestra  $n=38$  se llevó a cabo empleando la afijación de Neyman, con la siguiente expresión

$$n_{ij} = \frac{N_{ij}S_{ij}}{\sum_{H=1}^H N_{ij}S_{ij}} n$$

De la cual se obtuvo los tamaños de muestra a seleccionar en cada variedad, como se muestra en el cuadro 32.

**Cuadro 32: Afijación de la muestra en el muestreo estratificado aleatorio en cada estrato tipo T (variedades), para la variable Número de Frutos, considerando la mayor varianza.**

	Estrato T						
	I	II	III	IV	V	VI	N
$n_{ij}$	5	6	5	10	6	6	38

La disposición de las plantas seleccionadas, se muestra en el Apéndice D.

Con las muestras obtenidas del ensayo en Musáceas, se obtuvo los siguientes resultados:

➤ **Estadística Descriptiva considerando la mayor varianza:**

Se presenta el cuadro 33 con las estimaciones, en cada una de las variedades evaluadas.

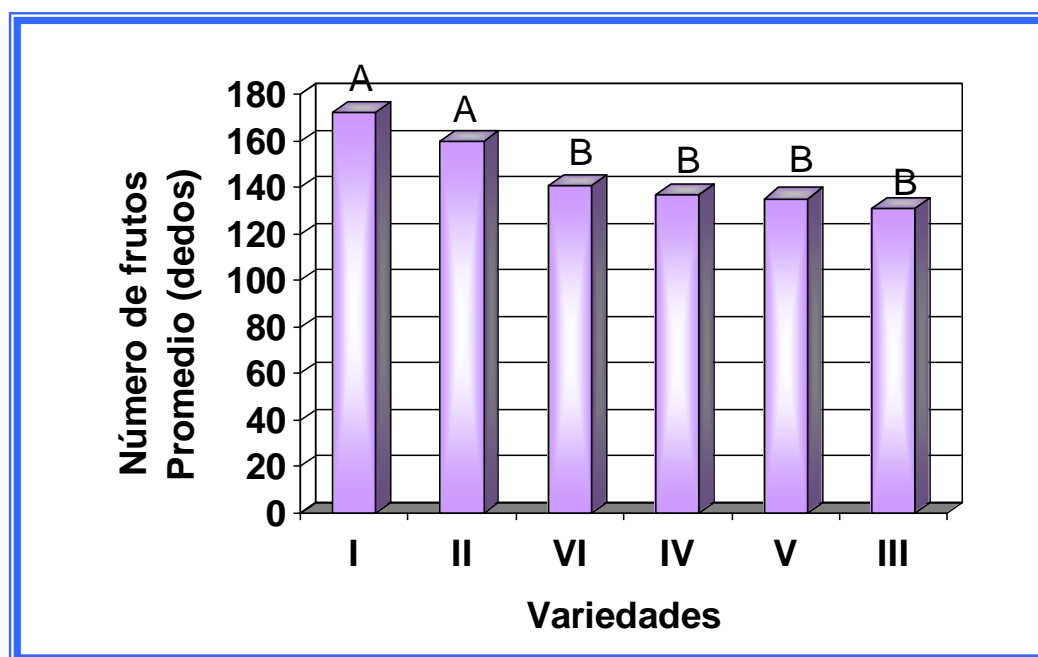
**Cuadro 33: Estimaciones de la media y varianza del Número de Frutos (dedos), por variedad, para el modelo del muestreo estratificado aleatorio, considerando  $n=38$**

Variedad	$n_i$	$\bar{y}_i$	$s_i^2$	Valor mín.	Valor Máx.
I	5	172.20	355.70	139	186
II	6	159.67	192.67	146	182
III	5	129.40	90.800	122	145
IV	10	136.80	293.51	110	161
V	6	135.17	100.17	124	150
VI	6	141.00	181.60	126	160

➤ **Análisis de Varianza considerando n mayor:**

Para el ensayo en Musáceas, realizado bajo el modelo de muestreo estratificado aleatorio, la prueba de Duncan, señaló dos grupos (figura 11).

Las variedades I y II presentaron un mayor número de frutos y las variedades VI, IV, V y III, la menor cantidad, respectivamente



**Figura 11. Promedios del Número de Frutos (dedos), para las variedades estudiadas, en el Modelo de Muestreo Estratificado Aleatorio con n=38**

El análisis de varianza se presenta en el cuadro 34.

**Cuadro 34: Análisis de Varianza, para el Número de Frutos (dedos), bajo el Modelo de Muestreo Estratificado Aleatorio con n=38**

F. de V.	G de L	SC	CM	Pr
Variedades	5	7545.73	1509.1467	0.0001
Error	32	6799.77	212.49271	

## **6. APLICACIÓN PRÁCTICA DEL MODELO DE MUESTREO ESTRATIFICADO DOBLE**

Es importante señalar que para este caso dado que es desbalanceada la muestra, no se efectuó la prueba de "F", ya que no se realizó el desarrollo del mismo en el Capítulo IV, sino para el caso balanceado.

### **A. Variable Tiempo de Desarrollo de la Enfermedad (TDE):**

En vista de que los resultados del experimento demostraron que dos (2) variedades (variedad I: 'Fagro 122' y la variedad VI: 'Cien Bta-03'), fueron resistentes a la enfermedad, la afijación de la muestra de 22 plantas según el cuadro 35, se realizó mediante el método proporcional, pero como las parcelas tienen igual tamaño la afijación condujo a una afijación igual; por la cual se seleccionaron 5 plantas para cada estrato tipo B, en cada estrato T (variedad), de tal manera de que sea igual o superior a la muestra mínima obtenida de 22 plantas. Las plantas seleccionadas por cada estrato tipo T (variedad) en cada estrato tipo B se muestran en el Apéndice E.

Los resultados obtenidos de la aplicación del Modelo de Muestreo Estratificado Doble en la muestra seleccionada según lo anterior, se presentan a continuación:

➤ **Estadística Descriptiva**

Se muestra en el cuadro 35 las estimaciones de la media y varianza, para cada variedad dentro de cada bloque.

**Cuadro 35: Estadística descriptiva del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, para cada variedad, en cada Estrato B, para el modelo del muestreo estratificado doble**

Variedad	Bloque	n	$\bar{y}_i$	$s_i^2$	Valor mín.	Valor Máx.
I		0	-	-	-	-
II	1	5	65.600	27.8003	57	71
	2	5	46.200	352.200	28	72
	3	5	44.200	173.212	29	63
	4	5	57.200	5.70016	54	60
	5	3	46.667	5.33333	44	48
III	1	5	59.800	123.210	48	74
	2	5	49.600	74.8000	40	62
	3	5	43.400	40.8002	34	50
	4	5	43.400	43.3003	36	54
	5	5	38.800	41.1997	29	45
IV	1	5	71.800	40.7005	66	82
	2	5	57.200	23.1996	51	61
	3	5	53.600	17.7999	51	61
	4	5	56.000	98.0001	45	65
	5	4	41.500	51.6673	36	52
V	1	2	84.000	8.4853	78	90
	2	0	-	-	-	-
	3	0	-	-	-	-
	4	0	-	-	-	-
	5	0	-	-	-	-
VI		0	-	-	-	-

En los casos donde  $n$  es menor a 5, es por la presencia de datos perdidos (plantas muertas) en la muestra o en el caso de la variedad V por ser una variedad resistente.

Además, se presenta el cuadro 36 con el promedio y varianza estimados, para cada variedad.

**Cuadro 36: Estimaciones de la media y varianza del Tiempo de Desarrollo de la Enfermedad (días) por variedad, para el modelo del muestreo estratificado doble**

Variedad	n	$\bar{y}_i$	$s_i^2$	Valor mín.	Valor Máx.
I	0	-	-	-	-
II	23	52.435	175.4300	28	72
III	25	47.000	108.8266	29	74
IV	24	56.625	129.9828	36	82
V	2	84.000	72.00032	78	90
VI	0	-	-	-	-

➤ **Análisis de Varianza:**

Para el ensayo en Musáceas, realizado bajo el modelo de muestreo estratificado doble, la prueba de medias de Duncan, generó dos grupos. La variedad V por ser resistente, es la que se diferencia de las variedades II, III y IV.

Se puede observar en la figura 12 que en las variedades II, III y IV, se desarrolla en menor tiempo la sigatoka negra, en relación a la variedad V. Además, en esta variedad, solo 2 plantas de las 28 observadas, presentaron la enfermedad, lo cual confirma su resistencia. El análisis de varianza se muestra en el cuadro 37.

**Cuadro 37: Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (días), bajo el Modelo de Muestreo Estratificado Doble**

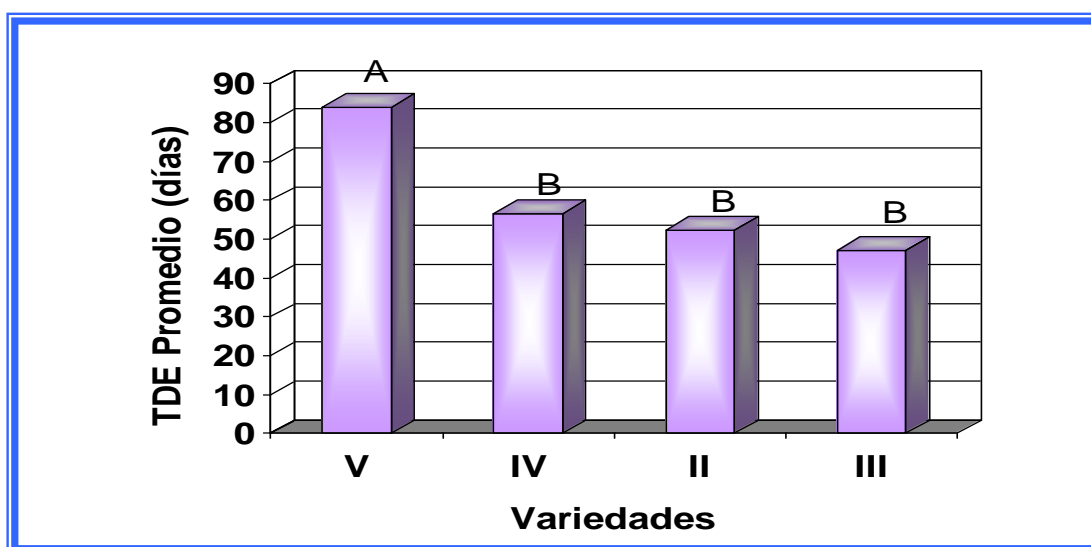
F. de V.	G de L	SC	CM
Variedades	3	1586.025	528.6751
Bloques	4	4284.735	1071.184
Interacción	8	658.5998	82.32498
Error	58	4485.267	77.33218

Se realizó el análisis de varianza sin considerar la interacción, para efectos prácticos, obteniéndose los resultados del cuadro 38. Además, la prueba de medias de Duncan, arrojó los mismos resultados.

**Cuadro 38: Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (días), bajo el Modelo de Muestreo Estratificado Doble sin la interacción**

F. de V.	G de L	SC	CM
Variedades	3	1629.898	543.29939
Bloques	4	4389.411	1097.3527
Error	66	5143.865	77.93737





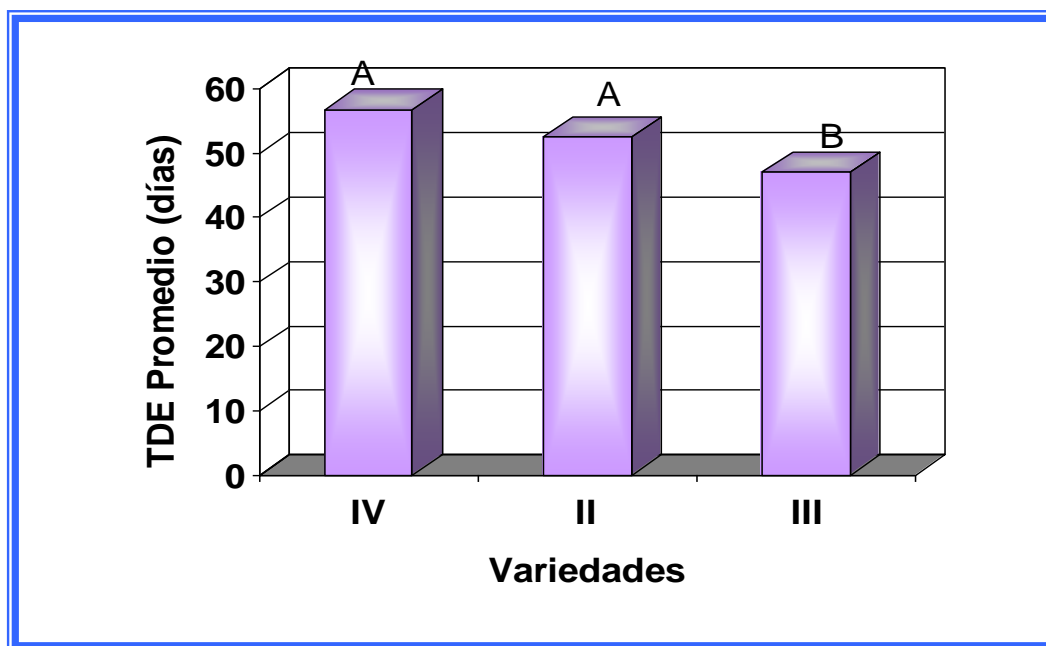
**Figura 12.** Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades estudiadas, en el modelo de muestreo estratificado doble sin interacción

➤ **Análisis de Varianza para variedades susceptibles:**

Para el ensayo en Musáceas, realizado bajo el modelo de muestreo estratificado doble, la prueba de medias de Duncan, generó dos grupos. La variedad III presenta menor resistencia a la enfermedad que las variedades IV y II (figura 13).

**Cuadro 39:** Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (días) en Sigatoka Negra para las variedades susceptibles, bajo el Modelo de Muestreo Estratificado Doble

F. de V.	G de L	SC	CM
Variedades	2	997.1940	498.59703
Bloques	4	4284.735	1071.1836
Interacción	8	658.5999	82.324982
Error	57	4413.267	77.42573



**Figura 13. Promedios de Tiempo de Desarrollo de la Enfermedad (TDE) en Sigatoka Negra para las variedades susceptibles, en el modelo de muestreo estratificado doble.**

Se realizó el análisis de varianza sin considerar la interacción obteniéndose los resultados del cuadro 40. Además, la prueba de medias de Duncan, arrojó los mismos resultados.

**Cuadro 40: Análisis de Varianza, para el Tiempo de Desarrollo de la Enfermedad (días) para las variedades susceptibles, bajo el Modelo de Muestreo Estratificado Doble sin la interacción**

F. de V.	G de L	SC	CM
Variedades	2	1041.0668	520.53341
Bloques	4	4389.4107	1097.3227
Error	65	5071.8665	78.028720

## B. Variable Número de frutos (dedos)

Se realizó la afijación de Neyman de la muestra de plantas, según los resultados del cuadro 24, a los cinco estratos tipo B, en cada variedad, como se observa en el cuadro 41, empleando la siguiente expresión:

$$n_{ij} = \frac{N_{ij}S_{ij}}{\sum_{H=1}^H N_{ij}S_{ij}} n$$

**Cuadro 41: Afijación de la muestra en el muestreo estratificado doble en cada estrato tipo T (variedades), para la variable Número de Frutos.**

	Estrato T											
	I		II		III		IV		V		VI	
	P	n <sub>ij</sub>	P	n <sub>ij</sub>	P	n <sub>ij</sub>	P	n <sub>ij</sub>	P	n <sub>ij</sub>	P	n <sub>ij</sub>
1	5	3	3	4	4	2	6	6	2	1	1	2
2	7	1	12	2	8	1	10	6	9	4	11	3
3	13	2	18	3	15	2	17	7	16	2	14	3
4	23	2	19	3	21	2	22	3	24	4	20	3
5	30	2	29	2	25	2	27	3	26	3	28	4
n		10		14		9		25		14		15

P: Parcela correspondiente a cada estrato tipo T (variedades) en cada estrato tipo B

Las plantas seleccionadas en cada parcela, se presentan en el Apéndice E

Con los datos de la muestra seleccionada se obtuvo los siguientes resultados:

### ➤ **Estadística Descriptiva**

Se muestra el cuadro 42 con el promedio y varianza estimados, para cada variedad.

**Cuadro 42: Estimaciones de la media y varianza del Número de frutos (dedos) por variedad, para el modelo del muestreo estratificado doble**

Variedad	n	$\bar{y}_i$	$s_i^2$	Valor mín.	Valor Máx.
I	10	172.70	143.353	150	186
II	14	162.21	231.405	142	185
III	9	135.56	109.035	122	152
IV	25	133.60	386.319	95	161
V	13	140.00	238.177	112	164
VI	15	147.73	190.495	118	170

Además, se presenta el cuadro 43 con las estimaciones de la media y varianza, para cada variedad dentro de cada bloque.

### ➤ **Análisis de Varianza**

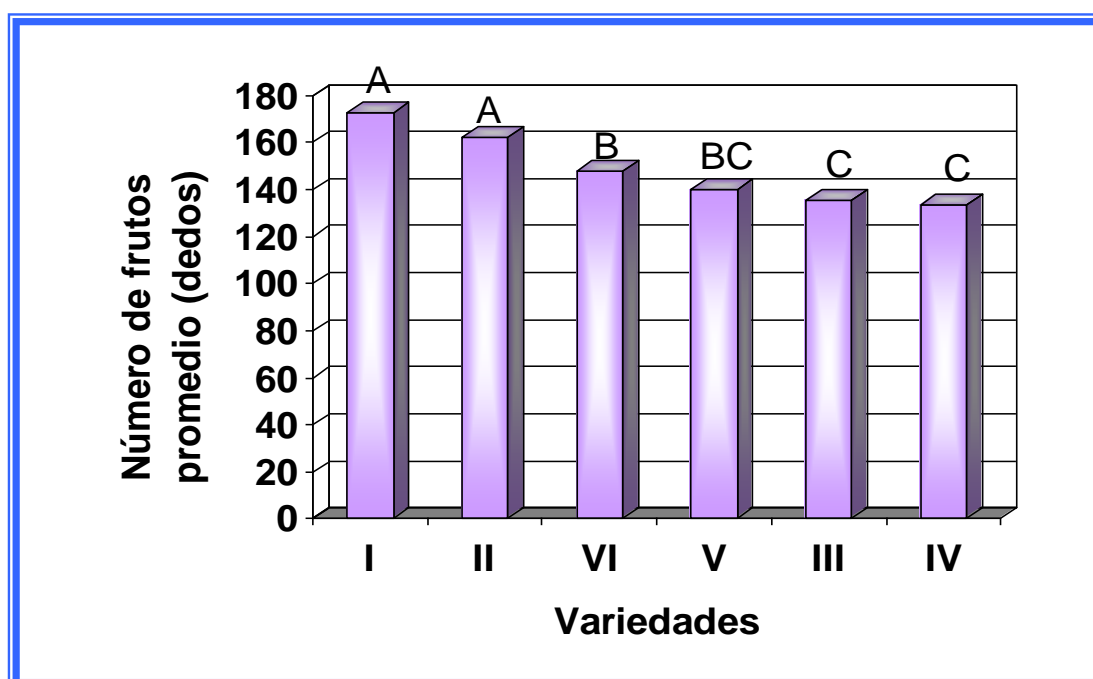
Para el ensayo en Musáceas, realizado bajo el modelo de muestreo estratificado doble, la prueba de medias de Duncan, generó cuatro grupos (figura 14). Las variedades I y II, desarrollan más frutos y las variedades IV, III y V menos frutos.

**Cuadro 43: Estadística descriptiva del Número de frutos (dedos), para las variedades, en cada bloque, bajo el modelo del muestreo estratificado doble**

Variedad	Bloque	n	Promedio	Varianza	Valor mín.	Valor Máx.
I	1	3	165.33	345.328	150	186
	2	1	170.00	-	170	170
	3	2	180.50	4.49991	179	182
	4	2	168.50	112.508	161	176
	5	2	181.50	0.49999	181	182
II	1	4	164.00	312.017	142	184
	2	2	163.00	162.002	154	172
	3	3	145.00	4.00000	143	147
	4	3	163.00	81.0000	154	172
	5	2	182.50	12.4998	180	185
III	1	2	133.00	17.9997	130	136
	2	1	130.00	-	130	130
	3	2	123.00	1.99997	122	124
	4	2	139.50	60.5004	134	145
	5	2	149.50	12.4998	147	152
IV	1	6	120.67	313.857	95	139
	2	6	137.50	349.092	108	161
	3	7	134.86	580.135	102	158
	4	3	128.33	12.3334	125	132
	5	3	154.00	7.00026	152	157
V	1	1	128.00	-	128	128
	2	4	126.50	137.007	112	140
	3	2	133.50	84.5002	127	140
	4	3	160.33	16.3337	156	164
	5	3	146.00	11.9999	142	148
VI	1	2	140.00	7.9998	138	142
	2	3	155.00	18.0228	135	170
	3	3	131.67	152.325	118	142
	4	3	154.00	3.00017	152	155
	5	4	153.50	131.676	139	167

**Cuadro 44: Análisis de Varianza, para el Número de frutos (dedos), bajo el Modelo de Muestreo Estratificado Doble**

F. de V.	G de L	SC	CM
Variedades	5	14589.815	2917.9629
Bloques	4	3982.919	995.72999
Interacción	20	4805.992	240.29959
Error	56	10931.86	195.21173



**Figura 14. Promedios de Números de frutos (dedos) para las variedades estudiadas, en el modelo de muestreo estratificado doble.**

➤ **Análisis de Varianza sin la interacción:**

El modelo de muestreo estratificado doble, sin considerar la interacción, abultó las sumas de cuadrados (cuadro 45)

Las variedades I y II, desarrollan más frutos y las variedades V, III y IV menos frutos (figura 15).

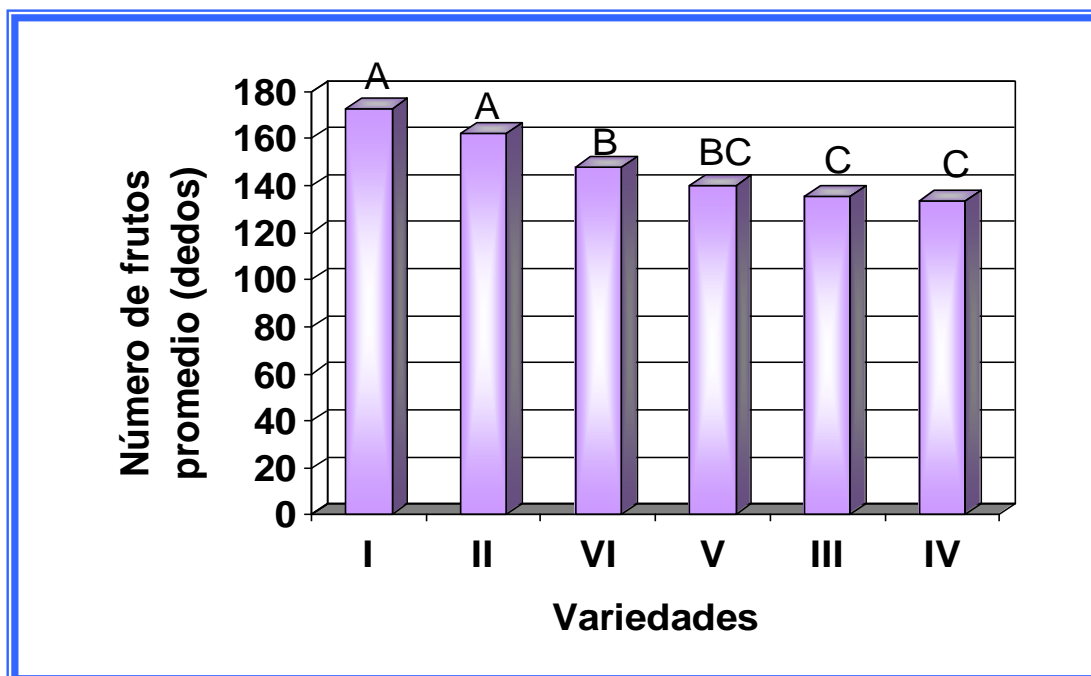


Figura 15. Promedios de Números de frutos (dedos) para las variedades estudiadas, en el modelo de muestreo estratificado doble.

Cuadro 45: Análisis de Varianza, para el Número de frutos (dedos), bajo el Modelo de Muestreo Estratificado Doble, sin la interacción

F. de V.	G de L	SC	CM
Variedades	5	15778.1824	3155.6365
Bloques	4	4229.7637	1057.4409
Error	76	15737..849	207.07696

Como se mencionó anteriormente, las seis variedades evaluadas poseen características diferentes en las variables de producción y desarrollo, por ello, en la práctica es recomendable afijar para el **Número de Frutos**, el tamaño de muestra, considerando la mayor varianza entre las seis variedades.

Por lo tanto, se afijó  $n=25$  en los estratos tipo B, para cada variedad (cuadro 46 ).

**Cuadro 46: Afijación de la muestra en el muestreo estratificado doble en cada estrato tipo T (variedades), para la variable Número de Frutos, considerando la mayor varianza.**

Estrato B	Estrato T											
	I		II		III		IV		V		VI	
	P	$n_{ij}$	P	$n_{ij}$	P	$n_{ij}$	P	$n_{ij}$	P	$n_{ij}$	P	$n_{ij}$
1	5	8	3	7	4	5	6	6	2	2	1	3
2	7	4	12	3	8	3	10	6	9	7	11	6
3	13	5	18	5	15	6	17	7	16	4	14	5
4	23	4	19	6	21	5	22	3	24	7	20	5
5	30	4	29	4	25	6	27	3	26	5	28	6
n		25		25		25		25		25		25

P: Parcela correspondiente a cada estrato tipo T (variedades) en cada estrato tipo B

Las plantas seleccionadas en cada parcela, se presentan en el Apéndice E

➤ **Estadística Descriptiva considerando la mayor varianza:**

Se presenta el cuadro 47 con las estimaciones, en cada una de las variedades evaluadas.



**Cuadro 47: Estimaciones de la media y varianza del Número de frutos (dedos) por variedad, para el modelo del muestreo estratificado doble con n mayor**

Variedad	n	$\bar{y}_i$	$s_i^2$	Valor mín.	Valor Máx.
I	25	171.48	145.01	139	187
II	25	159.76	179.36	140	185
III	24	132.63	114.33	112	152
IV	25	133.60	386.33	95	161
V	21	140.00	187.50	112	164
VI	25	146.65	185.84	118	171

Además, se muestra el cuadro 48 con las estimaciones de la media y varianza, para cada variedad dentro de cada bloque.

➤ **Análisis de Varianza considerando la mayor varianza:**

Para el ensayo en Musáceas, realizado bajo el modelo de muestreo estratificado doble, la prueba de medias de Duncan, generó cinco grupos (figura 16).

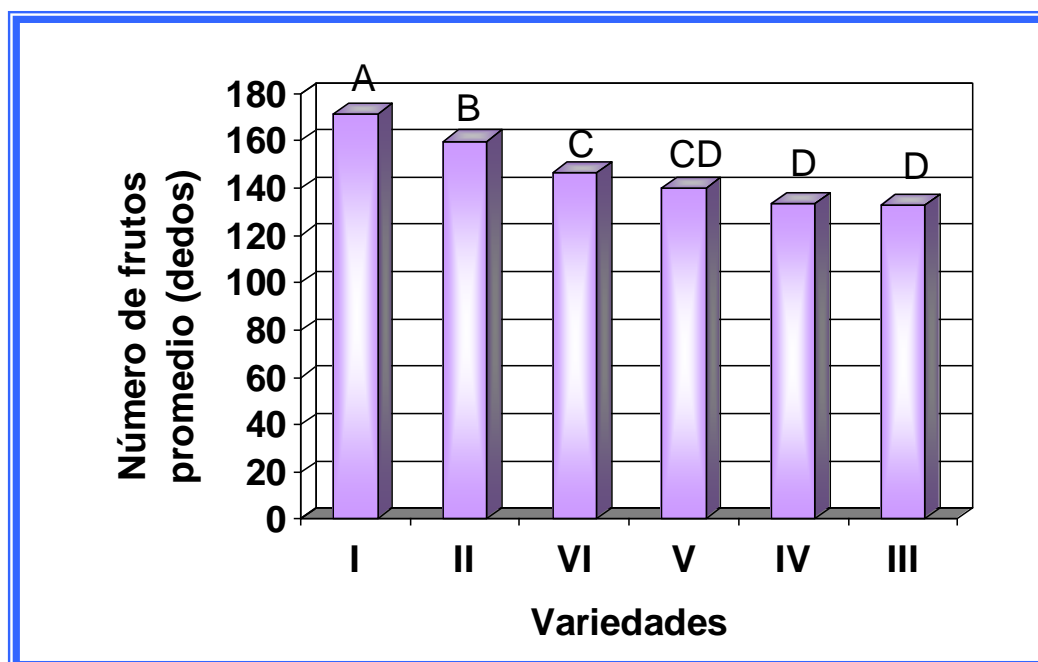
Las variedades I y II, desarrollan más frutos y las variedades V, IV y III menos frutos.

**Cuadro 48: Estadística descriptiva del Número de frutos (dedos), para las variedades, en cada bloque, bajo el modelo del muestreo estratificado doble con n mayor**

Variedad	Bloque	n	Promedio	Varianza	Valor mín.	Valor Máx.
I	1	8	170.63	341.70	139	187
	2	4	175.25	16.917	170	180
	3	5	172.20	99.200	158	182
	4	4	165.25	54.250	160	176
	5	4	174.75	71.583	164	182
II	1	7	158.29	235.24	140	184
	2	3	162.00	84.000	154	172
	3	5	147.00	11.500	143	152
	4	6	161.00	102.40	145	172
	5	4	174.75	138.25	158	185
III	1	5	128.80	43.200	122	136
	2	3	124.00	28.000	120	130
	3	5	126.40	138.80	112	144
	4	5	134.60	85.800	122	145
	5	6	143.67	52.267	132	152
IV	1	6	120.67	313.87	95	139
	2	6	137.50	349.10	108	161
	3	7	134.86	580.14	102	158
	4	3	128.33	12.333	125	132
	5	3	154.00	7.000	152	157
V	1	2	130.50	12.500	128	133
	2	6	127.83	94.567	112	140
	3	3	132.33	46.333	127	140
	4	6	153.33	87.867	142	164
	5	4	148.00	38.250	142	157
VI	1	3	138.00	16.000	134	142
	2	6	152.83	176.37	135	170
	3	5	134.60	176.80	118	152
	4	5	145.40	161.30	126	155
	5	7	154.57	124.62	139	171

**Cuadro 49: Análisis de Varianza, para el Número de frutos (dedos), bajo el Modelo de Muestreo Estratificado Doble con n mayor**

F. de V.	G de L	SC	CM
Variedades	5	28209.315	5641.86309
Bloques	4	5177.7469	1294.43672
Interacción	20	4921.5448	246.07724
Error	116	18143.025	156.40539



**Figura 16. Promedios de Números de frutos (dedos) para las variedades estudiadas, en el modelo de muestreo estratificado doble con n mayor**

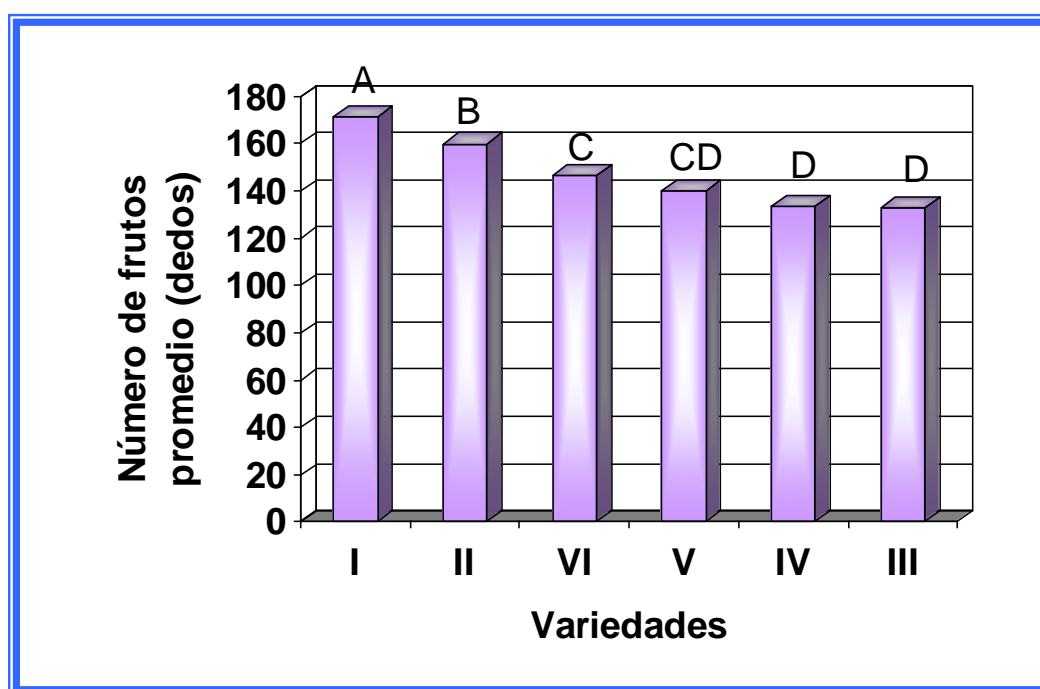
➤ **Análisis de Varianza sin la interacción:**

El modelo de muestreo estratificado doble, sin considerar la interacción, abultó las sumas de cuadrados (cuadro 50)

La prueba de medias de Duncan, generó cinco grupos (figura 17). La variedad I y II son las que desarrollaron más frutos y las variedades V, IV y III menor cantidad.

**Cuadro 50: Análisis de Varianza, para el Número de Frutos (días), bajo el Modelo de Muestreo Estratificado Doble, con la muestra mayor.**

F. de V.	G de L	SC	CM
Variedades	5	30448.628	6089.7249
Bloques	4	5017.7398	1254.4349
Error	136	23064.569	169.59243



**Figura 17. Promedios de Números de frutos (dedos) para las variedades estudiadas, en el modelo de muestreo estratificado doble, con la muestra mayor**

## **DISCUSIÓN DE RESULTADOS**

### **1. CONTRASTE ENTRE LOS PRINCIPIOS BÁSICOS DEL DISEÑO DE EXPERIMENTOS Y DEL DISEÑO DE MUESTREO**

#### **A. Diseño Completamente Aleatorizado (D.C.A.) y Muestreo Aleatorio Simple (M.A.S.)**

- **Aleatorización en el D.C.A.:** Esta aleatorización se lleva a cabo con la enumeración de los tratamientos, luego la obtención de números aleatorios a través de la tabla de números aleatorios o cualquier otro mecanismo que lo permita, para la asignación irrestricta de los tratamientos a las unidades experimentales.
  
- **En el M.A.S.:** Al igual que en el diseño de experimentos completamente al azar, se emplea la tabla de números aleatorios o cualquier otro mecanismo que permita la obtención de números aleatorios a fin de realizar la selección de la muestra al azar, para que el muestreo sea probabilístico, de no ser así sería opinático u otro.

Es importante recordar que si el muestreo no es probabilístico no se pueden realizar inferencias sobre la población con la muestra obtenida.

La aleatorización está íntimamente relacionada con la representatividad, ya que se selecciona una muestra al azar, para que ésta sea representativa de la población bajo estudio

La diferencia con el diseño de experimentos es que en este se aleatorizan los tratamientos para asignarlos a las unidades experimentales y en el muestreo lo que se obtiene es una muestra aleatoria pero ambos se basan en el principio de la aleatorización.

- **Replicación en el D.C.A.:** es el número de repeticiones que contendrá el experimento.

Como se citó anteriormente el número de repeticiones depende entre otros factores, de la variabilidad del material utilizado, por lo tanto en el diseño completamente al azar no se requieren tantas repeticiones, ya que se considera que las unidades experimentales son lo suficientemente homogéneas para aplicar este tipo de diseño. Además depende del grado de precisión deseado, dado que mientras mayor sea el número de repeticiones, las estimaciones son más precisas, siempre y cuando exista poca variabilidad ya que la precisión según Steel y Torrie (1988) está dada

$$\text{por: } l = \frac{1}{\sigma_y^2} = \frac{n}{\sigma^2}$$

- **Tamaño de la muestra M.A.S:** al igual que en el diseño completamente al azar, el tamaño de la muestra depende de la precisión deseada, pero también se deben considerar los costos, el tiempo y todo lo necesario para obtener la muestra propuesta.
- **Control local en el D.C.A.:** se realiza mediante cualquier manejo del experimento que permita las condiciones homogéneas requeridas para este diseño, al controlar variables extrañas, se eliminan los sesgos que puedan introducir las variables perturbadoras.
- **Control local en el M.A.S.:** se realiza con la adecuada aplicación del diseño de muestreo.

Las encuestas por muestreo tienen su lado fuerte en la representatividad y su lado débil en el control de variables, lo contrario al diseño de experimentos.

## **B. Diseño de Bloques al Azar (D.B.A.) y Muestreo Estratificado (M.E.A.)**

- **Aleatorización en el D.B.A.:** la aleatorización es *restringida al bloqueo*, es decir, se asignan dentro de cada bloque, aleatoriamente los tratamientos a las unidades experimentales, asegurando que todos los tratamientos se apliquen en cada bloque (figura 1, Apéndice A).

- **Aleatorización en el M.E.A.:** ocurre cuando se realiza una selección aleatoria de la muestra dentro de cada estrato.
- **Replicación en el D.B.A.:** está constituido por el número de bloques que contiene el experimento.

Para obtener precisiones similares a las obtenidas empleando el diseño completamente al azar (materiales más uniformes) se requieren más repeticiones; ya que se supone que existe una mayor heterogeneidad del material experimental, por ello se realiza el control local con el bloqueo.

- **Tamaño de la muestra en M.E.A.:** en el muestreo estratificado si los grupos o estratos son bien homogéneos, es decir con poca variabilidad, se puede tomar un tamaño de muestra pequeño.
- **Control Local en el D.B.A:** Con el bloqueo se intenta controlar una fuente de variación extraña al experimento, la cual de no ser controlada con este arreglo puede abultar el error. Se espera que exista homogeneidad dentro de bloques y heterogeneidad entre bloques.

Martínez (1988) señala que en la práctica agronómica, por ejemplo, los bloques se colocan perpendiculares a la variable a controlar con el bloqueo (figura 1 Apéndice A)



- **Control Local en el M.E.A.:** se lleva a cabo con la formación de los estratos para luego seleccionar la muestra dentro de cada uno de éstos. Además, como se mencionó en el desarrollo del muestreo estratificado, se debe garantizar la independencia de los estratos y por lo general se espera que sean homogéneos dentro de ellos y heterogéneos entre sí.

La falta de control de los tratamientos es la principal desventaja del muestreo de encuestas. El control de variables puede llevarse a cabo a través de variables auxiliares.

- **Realismo:** algunos autores consideran que no puede reproducirse en el diseño de experimentos, pero en el muestreo, generalmente, se pueden realizar las observaciones en su "entorno natural" de las poblaciones reales.

## 2. CONTRASTES TEÓRICOS GENERALES ENTRE EL DISEÑO DE EXPERIMENTOS Y EL DISEÑO DE MUESTREO

El diseño de experimentos, se realiza para evaluar o comparar los efectos de ciertos tratamientos; pero en el muestreo, lo que se persigue es estimar parámetros de la población (media, total y proporción) a partir de una muestra.

Otra diferencia básica que parece no tener mucha importancia, es que en el diseño de experimento se ha estudiado en gran medida el modelo lineal aditivo para cada tipo de diseño; pero en la teoría del muestreo han jugado un menor rol, salvo para algunos autores como los que se citaron en la revisión de literatura.

### 3. CONTRASTE DE LOS RESULTADOS ENTRE LOS MODELOS DE DISEÑO DE EXPERIMENTOS Y LOS DEL MUESTREO

#### ➤ Variable TDE

En los cuadros del 51 al 55 se presenta el resumen de la estadística descriptiva y análisis de varianza en cada uno de los modelos estudiados, para su contraste, en la variable Tiempo de Desarrollo de la Enfermedad (TDE).

**Cuadro 51: Resumen de resultados de los Modelos del Diseño Completamente al Azar (D.C.A), Muestreo Aleatorio Simple (M.A.S) y Muestreo Estratificado Aleatorio (M.E.A), para la variable Tiempo de Desarrollo de la Enfermedad**

	D.C.A					M.A.S				M.E.A				
	CMvar	2009.25								CMvar	241.556			
	CMEE	133.522								CMEE	159.778			
	CV(%)	21.74								CV(%)	23.46			
Variedades	N	$\mu$	G	S <sup>2</sup>	CV	n	$\bar{y}$	s <sup>2</sup>	CV	n	$\bar{y}$	G	s <sup>2</sup>	CV
I	0	-	-	-	-	0	-			0	-	-	-	-
II	57	51.97	B	137.21	22.54	22	50.82	154.729	24.48	6	54.33	A	60.667	14.34
III	58	48.57	B	131.51	23.61	22	49.23	148.669	24.77	6	47.33	A	218.66	31.24
IV	58	56.22	B	137.19	20.83	20	59.40	112.784	17.88	6	60.00	A	199.99	23.57
V	4	85.50	A	33.000	6.72	1	84.00	-	-	0	-	-	-	-
VI	0	-	-	-	-	0	-	-	-	0	-	-	-	-

G: grupo de medias    S<sup>2</sup>: cuasivarianza    s<sup>2</sup>: estimación de la cuasivarianza a través de la muestra.

**Cuadro 52: Resumen de resultados de los Modelos del Diseño de Bloques al Azar (D.B.A) y Muestreo Estratificado Doble**

D.B.A						M.E.D				
CMvar	955.4423					CMvar	528.6751		CMvar	543.299
CMBlo	2781.572					CMBlo	1071.184		CMBlo	1097.352
CMEE	70.84587					CMint	82.32498		CMEE	77.93737
CV (%)	15.88					CMEE	77.33218		CV (%)	16.72
						CV (%)	16.65			
Variedades	N	$\mu$	G	$S^2$	CV	N	$\bar{y}$	G	$s^2$	CV
I	0	-	-	-	-	0	-	-	-	-
II	57	51.97	BC	137.21	22.54	23	52.44	B	175.430	25.26
III	58	48.57	C	131.51	23.61	25	47.00	B	108.827	22.19
IV	58	56.22	B	137.19	20.83	24	56.63	B	129.983	20.13
V	4	85.50	A	33.000	6.72	2	84.00	A	72.0003	10.10
VI	0	-	-	-	-	0	-	-	-	-

**(M.E.D), para la variable Tiempo de Desarrollo de la Enfermedad**

G: grupo de medias

$S^2$ : cuasivarianza

$s^2$ : estimación de la cuasivarianza a través de la muestra.

En el Cuadro 51 se puede observar:

- Diferencias notorias en cuanto al  $CM_{var}$  y  $CM_{EE}$ , las cuales pueden ser consecuencias del error de muestreo.
- Los coeficientes de variación del modelo resultaron similares
- Para el M.E.A la muestra seleccionada resultó pequeña, debido a que para su determinación se empleó el  $CM_{EE}$  obtenido del diseño completamente al azar, este tamaño de muestra se afijó uniformemente a las variedades ignorando bloques.
- Menores diferencias en las estimaciones de la media y varianza para el M.A.S con respecto al M.E.A.; es de hacer notar que en ambos modelos los coeficientes de variación son aceptables.

En el Cuadro 52, se observa:

- Los  $CM_{EE}$  mostraron poca diferencia; no obstante al comparar los  $CM_{var}$  y  $CM_{bloq}$ , las diferencias fueron notorias.
- Los coeficientes de variación del modelo mostraron valores muy cercanos.

- El tamaño de muestra utilizado en el Muestreo Estratificado Doble es alrededor de un 40 % menor que el utilizado para el Diseño de Bloques al Azar, lo cual en la práctica constituiría ahorro de tiempo, recursos y facilidad de manejo
- Las estimaciones del Modelo de Muestreo Estratificado Doble para la media y varianza, puede considerarse que no difieren en gran medida.

Cabe destacar que en los Cuadros 51 y 52 el promedio de la variable TDE, presenta valores muy similares así como coeficientes de variación aceptables.

**Cuadro 53: Intervalos de Confianza al 95 % para la estimación de  $\mu$ , en cada Variedad, para la variable Tiempo de Desarrollo de la Enfermedad; en cada uno de los Modelos de Muestreo**

	Intervalo de Confianza M.A.S		Intervalo de Confianza M.E.A		Intervalo de Confianza M.E.D	
Variedades	Límite inferior	Límite superior	Límite inferior	Límite superior	Límite inferior	Límite superior
II	45.30	56.33	46.16	62.51	46.71	58.16
III	43.82	54.63	31.82	62.85	42.69	51.13
IV	54.43	64.37	45.16	74.84	51.81	61.44

En el Cuadro anterior, se puede apreciar que en el Muestreo Estratificado Aleatorio los intervalos de confianza son considerablemente amplios, lo que señala que se debe realizar un incremento del tamaño de muestra.

Además los intervalos de confianza para los tres modelos, contienen las medias del D.C.A.

**Cuadro 54: Estadística descriptiva del Tiempo de Desarrollo de la Enfermedad (días), en Sigatoka Negra, para las variedades dentro de cada bloque, bajo el Diseño de Bloques al Azar (D.B.A) y Muestreo estratificado Doble (M.E.D)**

		D.B.A				M.E.D			
Variedad	Bloque	N	$\mu$	$S^2$	CV	n	$\bar{y}$	$s^2$	CV
I		0	-	-	-	0	-	-	
II	1	12	65.67	28.97	8.20	5	65.60	27.800	8.04
	2	11	46.91	163.09	27.22	5	46.20	352.20	40.62
	3	12	45.33	122.06	24.37	5	44.20	173.21	29.78
	4	12	58.67	14.97	6.59	5	57.20	5.7002	13.20
	5	10	43.40	18.489	9.91	3	46.67	5.3333	15.65
III	1	11	63.09	69.29	13.19	5	59.80	123.21	18.56
	2	12	50.08	92.08	19.16	5	49.60	74.800	17.44
	3	11	44.91	68.89	18.48	5	43.40	40.800	14.72
	4	12	45.83	79.61	19.47	5	43.40	43.300	15.16
	5	12	39.83	65.97	20.39	5	38.80	41.199	16.54
IV	1	12	70.42	36.63	8.59	5	71.80	40.701	8.89
	2	11	55.27	64.22	14.50	5	57.20	23.199	8.42
	3	12	51.92	71.17	16.25	5	53.60	17.799	7.87
	4	12	58.25	71.48	14.51	5	56.00	98.000	17.68
	5	11	44.18	91.16	21.61	4	41.50	51.667	17.32
V	1	4	85.5	33.00	6.72	2	84.00	8.4853	10.97
	2	0	-	-	-	0	-	-	
	3	0	-	-	-	0	-	-	
	4	0	-	-	-	0	-	-	
	5	0	-	-	-	0	-	-	
VI		0	-	-	-	0	-	-	

En el cuadro se aprecia que en el Modelo de Muestreo Estratificado Doble, los resultados son concordantes en la estimación de los valores promedios; además, los coeficientes de variación fueron aceptables en la mayoría de los casos, a excepción de la variedad II ('Gran nain') en el bloque 2, cuya varianza es alta con un CV de 40.62%, y variedad IV bloque 3, cuya varianza



es baja con un CV menor. Esto puede deberse a que en la muestra se seleccionaron al azar valores muy extremos.

En el estudio realizado dos de las seis variedades evaluadas resultaron altamente resistentes (I: `Fagro 122´ y VI: `Cien Bta-03´) y otra (V: `FHIA-02´) se mostró resistente ya que presentó pocas plantas afectadas, lo que puede ser considerado por error de observación.

En virtud de lo anterior, se aplicó cada uno de los modelos de interés considerando únicamente las variedades susceptibles: II=`Gran nain´, III=`Pisang más´, y IV=`Brasilero´; cuyos resultados se muestran en el Cuadro 55

Por otro lado, al comparar los resultados del Cuadro 55 con los presentados en los Cuadros 51 y 52 se observa:

- Los  $CM_{var}$  presentan una ligera disminución en el Diseño de Bloques al Azar y el Muestreo Estratificado Doble; pero en el Diseño Completamente al Azar se observó una mayor diferencia entre los  $CM_{var}$ .
- Los  $CM_{EE}$  mostraron un ligero incremento en todos los modelos
- Los coeficientes de variación no manifestaron diferencias notorias en los distintos modelos

- En el Diseño de Bloques al Azar y el Muestreo Estratificado Doble el  $CM_{\text{bloq}}$  no sufrió variaciones
- El  $CM_{\text{int}}$  en el Modelo de Muestreo Estratificado Doble conservó su valor
- El Muestreo Estratificado Aleatorio presenta los mismos resultados, debido a que la muestra no seleccionó plantas afectadas en la variedad V para el estudio inicial.

**Cuadro 55: Análisis de Varianza para las variedades susceptibles en Tiempo de Desarrollo de la Enfermedad**

	<b>D.C.A</b>		<b>M.E.A</b>		<b>D.B.A</b>		<b>M.E.D</b>			
	CMvar	853.2845	CMvar	241.556	CMvar	784.2401	CMvar	498.597	CMvar	520.533
	CME	135.2955	CME	159.778	CM blo	2781.572	CM blo	1071.18	CM blo	1097.32
	CV(%)	22.26	CV(%)	23.46	CME	71.52983	CMint	82.3249	CME	78.0287
					CV(%)	16.19	CME	77.4257	CV(%)	17.01
							CV(%)	16.94		
Variedades	$\mu$	Grupo	$\bar{y}$	Grupo	$\mu$	Grupo	$\bar{y}$	Grupo	$\bar{y}$	Grupo
I	-	-	-	-	-	-	-	-	-	-
II	51.97	AB	54.333	A	51.97	B	52.44	A	52.44	A
III	48.57	B	47.333	A	48.57	C	47.00	B	47.00	B
IV	56.22	A	60.000	A	56.22	A	56.63	A	56.63	A
V	-	-	-	-	-	-	-	-	-	-
VI	-	-	-	-	-	-	-	-	-	-

Es importante señalar que para el modelo del M.E.D, la interacción resultó no significativa, por lo que se aplicó el modelo sin considerar la interacción, resultando que por ser esta aleatoria afecta al CM variedades y en menor grado al CM de error y CM bloques. Esto se debe al hecho de ser un modelo mixto cuya ECM contiene para las variedades el efecto de la interacción, tal como se muestra en el cuadro 8 para el Modelo Alterno

En cuanto a las Pruebas de Medias de Duncan, se obtuvo diferentes agrupaciones de las variedades susceptibles. En el D.C.A y D.B.A se formaron tres grupos. Para el M.A.E se forma un grupo y en el M.E.D dos grupos.

Los resultados del D.B.A y el M.E.D, reflejan mayor concordancia con la realidad observada en campo, dado que la variedad II (‘Gran nain’) y la variedad IV (‘Brasileiro’) son altamente susceptibles; sin embargo, la variedad III (‘Pisang más’) mostró mas susceptibilidad.

Por lo discutido anteriormente, es importante resaltar que este ejemplo con datos reales, sugiere que el modelo propuesto del M.E.D, funciona para la variable TDE, ya que no hubo correlación entre los efectos de interacción y puede emplearse en aquellos casos, en los cuales por razones prácticas no es posible aplicar el D.B.A. En la práctica, se lleva a cabo a menudo un muestreo y aplican un análisis de varianza con un F aproximado, especialmente en trabajos de mecanización, riego, entre otros, como son los de Romero (2004), Gondelles (2003) y Torrealba (2002), entre otros.

➤ **Variable Número de Frutos**

En los cuadros del 56 al 61 se presenta el resumen de la estadística descriptiva y análisis de varianza en cada uno de los modelos estudiados, para su contraste, en la variable Número de Frutos.

**Cuadro 56: Resumen de resultados de los Modelos del Diseño Completamente al Azar (D.C.A), Muestreo Aleatorio Simple (M.A.S) y Muestreo Estratificado Aleatorio (M.E.A), para la variable Número de Frutos**

Variedades	D.C.A					M.A.S				M.E.A				
	CMvar	16805.33								CMvar	758.4828			
	CMEE	175.28								CMEE	262.6564			
	CV(%)	9.024								CV(%)	11.43			
	N	$\mu$	G	$S^2$	CV	n	$\bar{y}$	$s^2$	CV	n	$\bar{y}$	G	$s^2$	CV
I	58	173.81	A	99.67	5.74	10	171.50	55.389	4.34	3	170.00	A	100.000	5.88
II	57	160.63	B	155.52	7.76	14	160.00	178.62	8.35	3	159.33	AB	121.330	6.91
III	58	132.17	D	85.75	7.01	9	135.67	135.75	8.59	2	133.00	B	162.001	9.57
IV	58	131.12	D	374.25	14.75	24	130.63	341.47	14.15	5	135.60	B	616.777	18.31
V	53	138.68	C	162.07	9.18	11	141.91	335.69	12.91	3	154.00	AB	219.010	9.61
VI	59	143.49	C	172.98	9.17	15	140.33	150.95	8.76	3	133.33	B	82.3338	21.52

G: grupo de medias

$S^2$ : cuasivarianza

$s^2$ : estimación de la cuasivarianza a través de la muestra.

En el Cuadro 56 se puede observar:

- En el D.C.A, el número de elementos estudiados varió de 53 a 58 plantas, y estos corresponden a valores poblacionales.
- En el caso del M.A.S el tamaño de la muestra se determinó para cada variedad y osciló entre 9 y 25.
- Con respecto al M.E.A, se determinó un tamaño de muestra general considerando el  $CM_{EE}$  del D.C.A, que fue afijado según Neyman.
- Los resultados en función de la media son muy similares, a excepción de las variedades V (‘FHIA-02’) y VI (‘Cien Bta-03’) en el M.E.A.
- En cuanto a las varianzas se puede observar que las mayores se presentan al aplicar M.E.A
- Analizando los CV (%) del modelo (cuadro 56), se evidencia que el CV (%) para el M.E.A de 11.43 %, indica que el modelo es apropiado.
- El  $CM_{EE}$  para el M.E.A., duplicó prácticamente al correspondiente al D.C.A, esto evidencia un deficiente tamaño de muestra.

**Cuadro 57: Intervalos de Confianza al 95 % para la estimación de  $\mu$ , en cada Variedad, para la variable Número de frutos; en cada uno de los Modelos de Muestreo**

Variedades	Intervalo de Confianza M.A.S		Intervalo de Confianza M.E.A	
	Límite inferior	Límite superior	Límite inferior	Límite superior
I	166.18	176.82	145.16	194.84
II	152.28	167.72	131.97	186.70
III	126.71	144.62	-	-
IV	122.82	138.43	104.76	166.44
V	129.60	154.22	117.24	190.76
VI	133.53	144.14	110.79	155.87

En el Cuadro 57 se puede apreciar:

Los valores correspondientes a los intervalos de confianza para la media, cuando se considera M.E.A son muy amplios, lo cual denota que existe un problema de tamaño de muestra. Obviamente los intervalos de confiando del M.A.S son mucho más cerrados y contienen la media poblacional.

Dada esta circunstancia se optó por seleccionar una muestra de mayor tamaño, tomando en consideración la mayor varianza, tal como se señaló en Metodología.

**Cuadro 58: Resumen de resultados de los Modelos del Diseño Completamente al Azar (D.C.A), Muestreo Aleatorio Simple (M.A.S) y Muestreo Estratificado Aleatorio (M.E.A), para la variable Número de Frutos con la muestra mayor**

Variedades	D.C.A					M.A.S				M.E.A				
	CMvar		16805.33							CMvar		1509.1467		
	CMEE		175.28							CMEE		212.49271		
	CV(%)		9.024							CV(%)		10.09		
	N	$\mu$	G	$S^2$	CV	n	$\bar{y}$	$S^2$	CV	n	$\bar{y}$	G	$s^2$	CV
I	58	173.81	A	99.67	5.74	25	172.32	111.39	6.12	5	172.20	A	355.70	10.95
II	57	160.63	B	155.52	7.76	24	160.25	156.98	7.82	6	159.67	A	192.67	8.69
III	58	132.17	D	85.75	7.01	25	133.92	86.660	6.95	5	129.40	B	90.800	7.36
IV	58	131.12	D	374.25	14.75	24	130.63	341.46	14.15	10	136.80	B	293.51	12.52
V	53	138.68	C	162.07	9.18	21	140.24	204.09	10.19	6	135.17	B	100.17	7.40
VI	59	143.49	C	172.98	9.17	25	144.28	198.88	9.77	6	141.00	B	181.60	9.56

G: grupo de medias

$S^2$ : cuasivarianza

$s^2$ : estimación de la cuasivarianza a través de la muestra.



En el Cuadro 58 se puede observar:

- Las estimaciones para la media, fueron similares.
- Las varianzas se mantuvieron dentro de cifras más o menos concordantes a excepción de la variedad I en M.E.A, lo cual puede deberse a la presencia de valores extremos.
- En todos los casos los CV (%) están dentro de valores aceptables.
- Se evidencia que con el M.E.A se obtienen tamaños muestrales, considerablemente menores, al compararlos con los tamaños del muestreo M.A.S y con los electos totales del D.C.A.
- Los valores correspondientes a los CM de variedades y CM del error son mucho más cercanos a los reales, por efecto del incremento del tamaño de la muestra.
- Además, el CV (%) es menor.

**Cuadro 59: Intervalos de Confianza al 95 % para la estimación de  $\mu$ , en cada Variedad, para la variable Número de frutos; en cada uno de los Modelos de Muestreo, considerando la muestra mayor**

Variedades	Intervalo de Confianza M.A.S		Intervalo de Confianza M.E.A	
	Límite inferior	Límite superior	Límite inferior	Límite superior
I	167.96	176.68	148.78	195.62
II	154.96	165.54	146.09	170.48
III	130.08	137.76	117.57	141.23
IV	122.82	138.43	122.70	146.57
V	133.74	146.74	127.15	148.57
VI	138.46	150.10	130.36	153.64

Por otro lado en el Cuadro 59, se observa:

Se evidencia que aún con el incremento del tamaño de la muestra el M.E.A se manifiesta con intervalos de confianza amplios en comparación con los obtenidos del M.A.S, bajo estas circunstancias el M.A.S se comporta como un modelo mas apropiado que el M.E.A.

**Cuadro 60: Resumen de resultados de los Modelos del Diseño de Bloques al Azar (D.B.A), Muestreo Estratificado Doble (M.E.D) para la variable Número de Frutos**

D.B.A						M.E.D					M.E.D n mayor				
CMvar		16840.69				CMvar	2917.96		CMvar	3155.64	CMvar	5641.863		CMvar	6089.73
CMblo		2520.15				CMblo	995.729		CMblo	1057.44	CMblo	1294.437		CMblo	1254.44
CMEE		147.118				CMint	240.299		CMEE	207.076	CMint	246.0772		CMEE	169.592
CV (%)		8.27				CMEE	195.212		CV(%)	9.83	CMEE	156.4054		CV(%)	8.82
						CV(%)	9.54				CV(%)	8.47			
Var	N	$\mu$	G	$S^2$	CV	n	$\bar{y}$	G	$s^2$	CV	n	$\bar{y}$	G	$s^2$	CV
I	58	173.81	A	99.67	5.74	10	172.70	A	143.35	6.93	25	171.48	A	145.01	7.02
II	57	160.63	B	155.52	7.76	14	162.21	A	231.41	9.38	25	159.76	B	179.36	8.38
III	58	132.17	E	85.75	7.01	9	135.56	C	109.04	7.70	24	132.63	D	114.33	8.06
IV	58	131.12	E	374.25	14.75	25	133.60	C	386.32	14.71	25	133.60	D	386.33	14.71
V	53	138.68	D	162.07	9.18	13	140.00	BC	238.18	11.02	21	140.00	CD	187.50	9.78
VI	59	143.49	C	172.98	9.17	15	147.73	B	190.49	9.34	25	146.65	C	185.84	9.30

En el Cuadro 60, se observa:

- Los valores promedios no son discrepantes para cada una de las variables. Los CV (%) se encuentran dentro de valores aceptables.
- EL  $CM_{EE}$  es sensiblemente menor cuando se incrementa el tamaño de muestra, sin embargo, esta condición no justifica el incremento del tamaño de muestra que en la mayoría de los casos estuvo alrededor del 40 %
- . El M.E.D es un modelo apropiado, dado que su CV (%), es bajo y cercano al del D.B.A.
- Indudablemente cuando el tamaño de muestra es mayor, aumenta la precisión, pero no se justifica, especialmente si se incurre en mayores costos.

**Cuadro 61: Estadística descriptiva del Número de Frutos (dedos), para las variedades dentro de cada bloque, bajo el Diseño de Bloques al Azar (D.B.A) y Muestreo estratificado Doble (M.E.D)**

Var	Bloq	D.B.A				M.E.D			
		N	$\mu$	$S^2$	CV(%)	n	$\bar{y}$	$s^2$	CV(%)
I	1	12	171.75	255.84	9.31	3	165.33	345.32	11.24
	2	12	176.08	46.45	3.87	1	170.00	-	-
	3	11	174.18	70.96	4.84	2	180.50	4.4999	11.75
	4	11	170.36	60.86	4.58	2	168.50	112.51	6.29
	5	12	176.42	64.63	4.56	2	181.50	0.4999	0.39
II	1	12	157.17	185.24	8.66	4	164.00	312.02	10.77
	2	11	160.36	50.26	4.42	2	163.00	162.00	7.81
	3	12	151.50	88.27	6.20	3	145.00	4.0000	13.79
	4	12	162.67	153.88	7.63	3	163.00	81.000	17.46
	5	10	173.60	69.82	4.81	2	182.50	12.499	6.13
III	1	11	129.36	49.06	5.41	2	133.00	17.999	10.09
	2	12	127.00	21.27	3.63	1	130.00	-	-
	3	11	126.91	83.69	7.21	2	123.00	1.9999	11.50
	4	12	134.83	60.52	5.77	2	139.50	60.500	17.63
	5	12	142.08	62.27	5.55	2	149.50	12.499	7.48
IV	1	12	126.17	343.24	14.68	6	120.67	313.86	14.68
	2	11	131.55	371.27	14.65	6	137.50	349.09	13.59
	3	12	134.17	581.42	17.97	7	134.86	580.14	17.86
	4	12	118.25	116.75	9.14	3	128.33	12.333	8.65
	5	11	146.82	105.76	7.00	3	154.00	7.0003	17.18
V	1	12	130.33	9.69	2.39	1	128.00	-	-
	2	11	126.55	92.87	7.62	4	126.50	137.01	9.25
	3	10	136.30	33.57	4.25	2	133.50	84.500	21.77
	4	10	154.70	78.90	5.74	3	160.33	16.334	7.97
	5	10	148.40	39.60	4.24	3	146.00	11.999	7.50
VI	1	11	132.73	49.62	5.31	2	140.00	7.9998	6.39
	2	12	150.17	124.15	7.42	3	155.00	18.023	8.66
	3	12	135.42	114.27	7.89	3	131.67	152.33	9.37
	4	12	143.33	113.70	7.44	3	154.00	3.0002	11.25
	5	12	154.92	136.81	7.55	4	153.50	131.68	7.48

➤ Las estimaciones del promedio son similares a los valores reales.

- Las varianzas en la mayoría de los casos difieren en gran medida, probablemente debido al tamaño de muestra tan pequeño.

**Cuadro 62: Estadística descriptiva del Número de Frutos (dedos), para las variedades dentro de cada bloque, bajo el Diseño de Bloques al Azar (D.B.A) y Muestreo estratificado Doble (M.E.D), con n mayor**

Var	Bloq	D.B.A				M.E.D			
		N	$\mu$	$S^2$	CV(%)	n	$\bar{y}$	$s^2$	CV(%)
I	1	12	171.75	255.84	9.31	8	170.63	341.70	10.83
	2	12	176.08	46.45	3.87	4	175.25	16.917	2.35
	3	11	174.18	70.96	4.84	5	172.20	99.200	5.78
	4	11	170.36	60.86	4.58	4	165.25	54.250	4.46
	5	12	176.42	64.63	4.56	4	174.75	71.583	4.84
II	1	12	157.17	185.24	8.66	7	158.29	235.24	9.69
	2	11	160.36	50.26	4.42	3	162.00	84.000	5.66
	3	12	151.50	88.27	6.20	5	147.00	11.500	2.31
	4	12	162.67	153.88	7.63	6	161.00	102.40	6.29
	5	10	173.60	69.82	4.81	4	174.75	138.25	6.73
III	1	11	129.36	49.06	5.41	5	128.80	43.200	5.10
	2	12	127.00	21.27	3.63	3	124.00	28.000	4.27
	3	11	126.91	83.69	7.21	5	126.40	138.80	9.32
	4	12	134.83	60.52	5.77	5	134.60	85.800	6.88
	5	12	142.08	62.27	5.55	6	143.67	52.267	5.03
IV	1	12	126.17	343.24	14.68	6	120.67	313.87	14.68
	2	11	131.55	371.27	14.65	6	137.50	349.10	13.59
	3	12	134.17	581.42	17.97	7	134.86	580.14	17.86
	4	12	118.25	116.75	9.14	3	128.33	12.333	2.74
	5	11	146.82	105.76	7.00	3	154.00	7.000	1.72
V	1	12	130.33	9.69	2.39	2	130.50	12.500	2.71
	2	11	126.55	92.87	7.62	6	127.83	94.567	7.61
	3	10	136.30	33.57	4.25	3	132.33	46.333	5.14
	4	10	154.70	78.90	5.74	6	153.33	87.867	6.11
	5	10	148.40	39.60	4.24	4	148.00	38.250	4.18
VI	1	11	132.73	49.62	5.31	3	138.00	16.000	2.90
	2	12	150.17	124.15	7.42	6	152.83	176.37	8.69
	3	12	135.42	114.27	7.89	5	134.60	176.80	9.88
	4	12	143.33	113.70	7.44	5	145.40	161.30	8.73
	5	12	154.92	136.81	7.55	7	154.57	124.62	7.22

- Mejoran las estimaciones, pero deben evaluarse los costos ya que si no es posible tomar un tamaño de muestra mayor, se corre el riesgo de que las estimaciones no sean confiables.

## **CONCLUSIONES**

Las conclusiones se presentan en función de cada uno de los Objetivos Específicos planteados:

### **Conclusiones relacionadas con el Objetivo Específico 1:**

- El Modelo Estratificado Doble propuesto, es el dado por:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \rho_{ij} + \varepsilon_{ijk} \quad i = 1, 2, \dots, a \quad j = 1, 2, \dots, b \quad k = 1, 2, \dots, N_{ij}$$

- Los estimadores de los diferentes efectos, resultaron afectados por los efectos aleatorios.
- Las ECM para el Modelo balanceado, coincide con las del Modelo Estándar presentados por Montgomery (1991).
- En el caso de desbalanceo las ECM, resultan afectadas por el estrato B y el estrato interacción. Pero cuando es balanceado, sólo están afectados por la interacción.
- En el Modelo propuesto, con datos balanceados el denominador apropiado para el cálculo de F es el  $CM_{int}$ , tanto para el estrato fijo T como para el estrato aleatorio B. Para la interacción el denominador apropiado es  $CM_{EE}$ .



- Los Estimadores de los parámetros obtenidos a través del Muestreo resultaron coincidentes, tanto con los obtenidos a partir de la Teoría de Aleatorización, como a partir del Modelo de Muestreo Estratificado Doble.

**Conclusiones relacionadas con los Objetivos Específicos 3 y 4:** Al contrastar los Principios Básicos del Diseño de Experimentos con los Criterios de Muestreo, se constató que:

- Son coincidentes en la Aleatorización, en cuanto a que permiten hacer inferencias.
- La diferencia es que en el Muestreo, se obtiene una muestra aleatoria y en el Diseño de Experimentos se aleatorizan los tratamientos a las unidades experimentales
- Son coincidentes en el Principio de Repeticiones, ya que tanto el tamaño de muestra como el número de repeticiones, dependen de la precisión y el costo.
- Son coincidentes en el Control Local, ya que en ambos se intenta controlar fuentes de variación extrañas, con el bloqueo o agrupamiento; además del manejo del experimento y el manejo del diseño de muestreo

- Difieren en que el Diseño de Experimentos permite controlar variables perturbadoras, lo cual no necesariamente ocurre en el muestreo.
- Difieren en Realismo y Representatividad

### **Conclusiones relacionadas con el Objetivo Específico 5:**

- Al contrastar el Modelo del Muestreo Aleatorio Simple con el del Diseño Completamente al Azar, se observó que los estimadores eran coincidentes.
- Las estimaciones obtenidas con la aplicación del Modelo del Muestreo Aleatorio Simple, fueron muy similares a los resultados del Diseño Completamente al Azar.
- Los estimadores del Modelo de Muestreo Estratificado Aleatorio fueron coincidentes a los del Diseño Completamente al Azar.
- En cuanto a las estimaciones obtenidas con la aplicación del Modelo de Muestreo Estratificado Aleatorio, se constató que el tamaño de muestra es fundamental para obtener una alta precisión.

**Conclusiones relacionadas con los Objetivos Específicos 2 y 6:**

- Las estimaciones obtenidas con el Modelo de Muestreo Estratificado Doble fueron muy cercanas a las resultantes del Diseño de Bloques al Azar

## **RECOMENDACIONES**

- El modelo de Muestreo Estratificado Doble puede utilizarse en la práctica cuando las condiciones de campo dificultan el manejo del experimento, y/o los costos no permiten su realización.
- Para la utilización del Modelo de Muestreo estratificado Doble, es necesario definir apropiadamente el tamaño de muestra en función de las variables bajo estudio.
- Cuando el tamaño de muestra se va a obtener en función de la información suministrada por experimentos previos, es conveniente utilizar la mayor varianza de los tratamientos; en caso contrario conviene realizar muestra pilotos para conjeturar la cuasivarianza.
- Evaluar el costo de la investigación para la selección de la muestra.
- Hallar las Sumas de Cuadrados y los valores de F, para el análisis en el Modelo de Muestreo Estratificado Doble, cuando se trabaje con datos desbalanceados.

## **REFERENCIAS BIBLIOGRÁFICAS**

Azorín, F. 1972. Curso de Muestreo y Aplicaciones. Editorial Aguilar. Madrid. 375 p.

Brakel, J. and Renssen, R. 1998. Design and Analysis of Experiments Embedded in Sample Surveys. Journal of Official Statistics, Vol. 14, N° 3, pág: 277-295.

Canavos, G. 1988. Probabilidad y Estadística, aplicaciones y métodos. McGraw-Hill Books Company, Inc., México. 651 p.

Chacín, F. 2000. Diseño y análisis de experimentos. Ediciones del Vicerrectorado Académico. Universidad Central de Venezuela. Caracas – Venezuela. 387 p.

Chou, Ya-Lun. 1977. Análisis Estadístico. Segunda Edición. Traducido por Vicente Agut Armer. Nueva Editorial Interamericana, S.A. de C.V. México. 808 p.

Cochran, W y Cox, G. 1997. Diseños Experimentales. Editorial Trillas. México. 653 p.

Cochran, W. 1985. Técnicas de Muestreo. Traducido por Sestier, A. Compañía Editorial Continental, S.A. de C.V. México. 505 p.

Federer, W. 1955. Experimental Design, Theory and Application. U.S.A. 544 p.

Fienberg, S. and Tanur, J. 1983. Reuniting the twain: remarking upon the parallels between Sample Surveys and Randomized Experiments.

International Statistical Institute, Booklet Recueil vol II. Invited and Contributed papers. Madrid. Pág. 748-751.

Gaskin, D. y Chaparro, E. 2002. Guía para la presentación escrita de los trabajos de ascenso. Universidad Central de Venezuela, Facultad de Agronomía. Maracay, Venezuela. 31 p.

Gondelles, L. 2003. Estudio del Efecto de la Interacción Suelo-Máquina en un Suelo Mecanizado de la Serie Maracay. Trabajo Ingeniero Agrónomo. Maracay, Venezuela; Facultad de Agronomía, Universidad Central de Venezuela. 74 p.

Graybill, F. 1976. Theory and Application of the Lineals Model. Editorial Drixbury Press.

Haddad, O. y Salas, Y. 2003. Técnicas de Muestreo Estratificado. U.C.V- Facultad de Agronomía. 47 p.

Haddad, O.; Muñoz, N.; Márquez, J. y Hernández, J. 2005. Tamaño de muestra para estimar el promedio de las principales variables de producción y de desarrollo en clones de *Musa*. Universidad Central de Venezuela, Facultad de Agronomía. Revista de la Facultad de Agronomía. 19 p.

Hocking, R. 1973. A Discussion of the Two-Way Mixed Model. The American Statistician, Vol. 27, N° 4, pág: 148-152.

John, Peter. 1971. Statistical Design and Analysis of Experiments. The Macmillan Company. New Cork. 356 p.

Kish, L. 1972. Muestreo de Encuestas. Editorial Trillas México. 739 p

Kish, L. 1995. Diseño Estadístico para la Investigación. Siglo XXI de España Editores. España. 362p.

Kuehl R. 2001. Diseño de Experimentos; Principios estadísticos para el diseño y análisis de investigación. 2ª. Ed. México. 665p.

Lohr, S. 2000. Muestreo: Diseño y Análisis. Traducido por: Óscar Palmas. Internacional Thomson Editores. México. 480 p.

Martínez, A. 1988. Diseños Experimentales, Métodos y Elementos de Teoría. Editorial Trillas. México. 751 p.

Montgomery,D. 1991. Diseño y Análisis de Experimentos. Traducido por:Delgado, J. Grupo Editorial Iberoamérica, S. A. México. 585 p.

Mosquera, L. 1974. Hipótesis Estadística con Aplicaciones. Segunda Edición. México. 280 p.

Neter J. And Wasserman W. 1974. Applied Linear Statistical Models. Richard D. Irwin, Inc. Illinois, 842 p.

Neter, J.; Wasserman, W. and Kutner, M. 1985. Applied Linear Statistical Model. Second Edition: Richard D. Irwin, Inc. U.S.A. 1127 p.

Patiño, C. 2004. El empleo de Modelos en la Estadística Oficial. Empleo de Modelos en la Estadística Oficial. [on line] [www.ies.jccm.es/jecas/paginas/ponenciaspanel3.htm-21k](http://www.ies.jccm.es/jecas/paginas/ponenciaspanel3.htm-21k). Instituto Galego de Estadística. 18 p.

Rojas, B. 1997. Sobre el número de repeticiones de un experimento. Serie de comunicaciones en Socioeconomía Estadística e Informática. México. 25 p.

Romero, T. 2004. Estudio del Comportamiento de Maíz (*Zea mays* L.), Bajo Riego por Aspersión y diferentes Sistemas de Labranza y Cobertura, en Turén, estado Portuguesa. Trabajo Ingeniero Agrónomo. Maracay, Venezuela; Facultad de Agronomía, Universidad Central de Venezuela. 98 p.

Sánchez, J. 1979. Curso Intensivo de Muestreo en Poblaciones Finitas. España. 192 p.

Scheaffer, R.; Mendenhall, W. y Ott, L. 1987. Elementos de Muestreo. Traducido por: Rendón, G. y Gómez, J. Grupo Editorial Iberoamérica, S. A. México. 320 p.

Steel, P y Torrie, J. 1988 Bioestadística, Principios y Procedimientos. Segunda Edición. Traducido por: Martínez Ricardo. México. 622 p.

Torrealba, L. 2002. Estudio del Comportamiento de la Soya (*Glycine máx*) en tres Tratamientos de Labranza en Parcelas Comerciales, en un Entisol en el Municipio Turén, Estado Portuguesa. Trabajo Ingeniero Agrónomo. Maracay, Venezuela; Facultad de Agronomía, Universidad Central de Venezuela. 101 p.

Wang S. y Chow S. 1994. Advanced Linear Models. Theory and Applications. Marcel Dekker, Inc. New York. 537 p.

Yamane, T. 1967. Elementary Sampling Theory. Prentice-Hall, Inc. U.S.A. 405 p.



## **APÉNDICE**

### **APÉNDICE A**

#### **Caracterización del ensayo en Musáceas**

Tiempo de Desarrollo de la enfermedad (TDE): es el tiempo (días) entre el estadio B (según escala de Brun) de la “hoja bandera” y la aparición de 10 ó más lesiones discretas, necrosadas y maduras en esa hoja.

El estadio B de BRUN corresponde al momento en que la “hoja bandera” (iniciadora superior) aún no ha abierto, pero está ya fuera del canal peciolar de la hoja anterior

El ensayo se inició el 23-11-2000 y el primer ciclo finalizó en abril del 2002.

**Ubicación:** Estación Experimental Samán Mocho, Facultad de Agronomía, Universidad Central de Venezuela. Municipio Carlos Arvelo, Sector Samán Mocho, Estado Carabobo.

**Coordenadas:** 10° 05' 58'' Latitud Norte

67° 51' 40'' Longitud Oeste

***Ambiente del sitio de la caracterización:***

- Topografía: Plana (0 – 0.5 %)
- Características Fisiográficas y geomorfológicas generales: Planicie formada por deposiciones lacustrinas con alto contenido de carbonato de calcio
- Drenaje del suelo: bien drenado
- Salinidad del suelo: moderada a alta
- Profundidad de la capa freática: > 150 cm
- pH del suelo: 7.50 a 8.50
- Erosión del suelo: baja pedregosidad (fragmentos o rocas) 0-2%
- Clase textural del suelo: franco arenoso
- Clasificación taxonómica del suelo: Mollic – Ustifluvents, franco fino mixto isohipertémico
- Disponibilidad de agua: riego en la época seca (por aspersión), se usó riego complementario cuando fue requerido
- Fertilidad del suelo: moderada
- Clima promedio anual (registro período 1967-1999):
  - ✓ Temperatura: Mínima 23.4 °C; Media 24.2 °C y Máxima 25.1 °C
  - ✓ Precipitación: Mínima 292.6 mm; Media 1082.2 mm y Máxima 2312 mm
  - ✓ Vientos: Mínimo 0.7 m/seg; Medio 0.9 m/seg y Máximo 1.2 m/seg
  - ✓ Humedad relativa: Mínima 67%; Media 76% y Máxima 82%

- ✓ Evaporación: Mínima 1184 mm; Media 1698.2 mm y Máxima 2134 mm
- ✓ Luz: abundante (soleado)

**Espaciamiento en el campo:** distancia entre plantas sobre la hilera 3 m, distancia entre hileras 2.5 m

**Diseño experimental del ensayo:** se utilizó un diseño de bloques al azar con cinco (5) repeticiones, seis (6) tratamientos y doce (12) plantas efectivas por parcela experimental. Cada parcela y cada planta dentro de parcela se enumeraron (figuras 1 y 2) y los tratamientos se aleatorizaron dentro de los bloques, tal como se muestra en la figura. El ensayo estaba constituido por un:

- ✓ Número total de plantas: 650
- ✓ Número de plantas efectivas: 360
- ✓ Número de plantas/variedad: 60
- ✓ Área efectiva del experimento: 4875 m<sup>2</sup>
- ✓ Área de contorno: 1824 m<sup>2</sup>
- ✓ Área total del experimento: 6699 m<sup>2</sup>

**Tratamientos (Variedades):**

I = 'Fagro 122' (Grupo AAA, Subgrupo Ibota)

II = 'Gran nain' (Grupo AAA, Subgrupo Cavendish)

III = 'Pisang más' (Subgrupo Sucrier)

IV = 'Brasileño' (Grupo AAA, Subgrupo Cavendish)

V = 'FHIA-02' (Grupo AAAB)

VI = 'Cien Bta-03' (Grupo AAAA, mosaico genético)

<b>B V</b>																		
		III			V			IV			VI			II			I	
			25			26			27			28			29			30
<b>B IV</b>																		
		V			I			IV			III			VI			II	
			24			23			22			21			20			19
<b>B III</b>																		
		I			VI			III			V			IV			II	
			13			14			15			16			17			18
<b>B II</b>																		
		II			VI			IV			V			III			I	
			12			11			10			9			8			7
<b>B I</b>																		
		VI			V			II			III			I			IV	
			1			2			3			4			5			6

**Figura 1: Plano en campo del experimento en Musáceas, en el cual se indica las parcelas en números arábigos y las variedades en números romanos. Bloques dispuestos horizontalmente.**

4	5	12
3	6	11
2	7	10
1	8	9

**Figura 2: Disposición de las plantas en cada parcela experimental.**

## APENDICE B

### Diseño Completamente al Azar

#### Modelo de Efectos Fijos (Modelo I):

Las soluciones a las ecuaciones normales del modelo son las siguientes:

Para  $\mu$ :

$$\begin{aligned} y_{..} = E(y_{..}) &= E \left[ \sum_{i=1}^{\alpha} \sum_{j=1}^{r_i} y_{ij} \right] = E \left[ \sum_{i=1}^{\alpha} \sum_{j=1}^{r_i} (\mu + \alpha_i + \varepsilon_{ij}) \right] \\ &= E \left[ \sum_{i=1}^{\alpha} \sum_{j=1}^{r_i} \mu + \sum_{i=1}^{\alpha} r_i \alpha_i + \sum_{i=1}^{\alpha} \sum_{j=1}^{r_i} \varepsilon_{ij} \right] \end{aligned}$$

Como  $E(\varepsilon_{ij}) = 0$

$$y_{..} = n\hat{\mu} + \sum_{i=1}^{\alpha} r_i \hat{\alpha}_i \quad (\text{III.1.2})$$

Para  $\alpha_i$ :

$$\begin{aligned} y_{i.} = E(y_{i.}) &= E \left[ \sum_{j=1}^{r_i} y_{ij} \right] = E \left[ \sum_{j=1}^{r_i} (\mu + \alpha_i + \varepsilon_{ij}) \right] \\ &= E \left[ r_i \mu + r_i \alpha_i + \sum_{j=1}^{r_i} \varepsilon_{ij} \right] \end{aligned}$$

Como  $E(\varepsilon_{ij}) = 0$

$$y_{i.} = r_i \hat{\mu} + r_i \hat{\alpha}_i \quad i = 1, 2, \dots, \alpha \quad (\text{III.1.3})$$

*Esperanza de los Cuadrados Medios:*

*Para Tratamientos:*

$$E(CM_{\text{trat}}) = E\left(\frac{SC_{\text{trat}}}{a-1}\right) = \frac{1}{a-1}E(SC_{\text{trat}})$$

$$E(CM_{\text{trat}}) = \frac{1}{a-1}E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} (\bar{y}_{i\cdot} - \bar{y}_{..})^2\right]$$

$$\begin{aligned}\bar{y}_{i\cdot} &= \frac{\sum_{j=1}^{r_i} Y_{ij}}{r_i} = \frac{1}{r_i} \sum_{j=1}^{r_i} (\mu + \alpha_i + \varepsilon_{ij}) = \frac{1}{r_i} \left( r_i \mu + r_i \alpha_i + \sum_{j=1}^{r_i} \varepsilon_{ij} \right) \\ &= \mu + \alpha_i + \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij}\end{aligned}$$

$$\begin{aligned}\bar{y}_{..} &= \frac{\sum_{i=1}^a \sum_{j=1}^{r_i} Y_{ij}}{n} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} (\mu + \alpha_i + \varepsilon_{ij}) = \frac{1}{n} \left( n\mu + \sum_{i=1}^a r_i \alpha_i + \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij} \right) \\ &= \mu + \frac{1}{n} \sum_{i=1}^a r_i \alpha_i + \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij}\end{aligned}$$

$$\text{Como } \sum_{i=1}^a r_i \alpha_i = 0$$

$$\begin{aligned}
E(SC_{\text{Trot}}) &= E \left\{ \sum_{i=1}^a \sum_{j=1}^{r_i} \left[ \left( \mu + \alpha_i + \frac{1}{r_i} \varepsilon_{i\cdot} \right) - \left( \mu + \frac{1}{n} \varepsilon_{..} \right) \right]^2 \right\} \\
&= E \left\{ \sum_{i=1}^a \sum_{j=1}^{r_i} \left[ \alpha_i + \left( \frac{1}{r_i} \varepsilon_{i\cdot} - \frac{1}{n} \varepsilon_{..} \right) \right]^2 \right\} \\
&= E \left\{ \sum_{i=1}^a \sum_{j=1}^{r_i} \left[ \alpha_i^2 + 2\alpha_i \left( \frac{1}{r_i} \varepsilon_{i\cdot} - \frac{1}{n} \varepsilon_{..} \right) + \left( \frac{1}{r_i} \varepsilon_{i\cdot} - \frac{1}{n} \varepsilon_{..} \right)^2 \right] \right\}
\end{aligned}$$

$$\left( \frac{1}{r_i} \varepsilon_{i\cdot} - \frac{1}{n} \varepsilon_{..} \right)^2 = \left( \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij} \right)^2 - 2 \left( \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij} \right) \left( \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij} \right) + \left( \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij} \right)^2$$

Sustituyendo se obtiene

$$\begin{aligned}
E(SC_{\text{Trot}}) &= E \left( \sum_{i=1}^a \sum_{j=1}^{r_i} \alpha_i^2 \right) - 2E \left[ \sum_{i=1}^a \sum_{j=1}^{r_i} \alpha_i \left( \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij} - \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij} \right) \right] \\
&\quad + E \left[ \sum_{i=1}^a \sum_{j=1}^{r_i} \left( \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij} \right)^2 \right] - 2E \left[ \sum_{i=1}^a \sum_{j=1}^{r_i} \left( \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij} \right) \left( \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij} \right) \right] + E \left[ \sum_{i=1}^a \sum_{j=1}^{r_i} \left( \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij} \right)^2 \right]
\end{aligned}$$

Dado que  $E(\varepsilon_{ij}) = 0$  y Aplicando las propiedades de la esperanza, los supuestos y restricciones para este modelo, se obtiene:

$$E \left[ \sum_{i=1}^a \sum_{j=1}^{r_i} \alpha_i \left( \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij} - \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij} \right) \right] = 0$$

$$\begin{aligned}
E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} \left(\frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij}\right)^2\right] &= \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{r_i^2} E\left(\sum_{j=1}^{r_i} \varepsilon_{ij}\right)^2 \\
&= \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{r_i^2} E\left(\varepsilon_{i1}^2 + \varepsilon_{i2}^2 + \dots + \varepsilon_{ir_i}^2 + 2\varepsilon_{i1}\varepsilon_{i2} + \dots + 2\varepsilon_{i1}\varepsilon_{ir_i} + \dots + 2\varepsilon_{ir_{i-1}}\varepsilon_{ir_i}\right) \\
&= \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{r_i^2} r_i \sigma^2 = \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{r_i} \sigma^2
\end{aligned}$$

$$\begin{aligned}
2E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} \left(\frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij}\right) \left(\frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij}\right)\right] &= 2 \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{r_i} \frac{1}{n} E\left(\sum_{j=1}^{r_i} \varepsilon_{ij}\right) \left(\sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij}\right) \\
&= 2 \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{r_i} \frac{1}{n} E\left(\varepsilon_{i1} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij} + \varepsilon_{i2} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij} + \dots + \varepsilon_{ir_i} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij}\right) \\
&= 2 \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{r_i} \frac{1}{n} (r_i \sigma^2) = 2 \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{\sigma^2}{n}
\end{aligned}$$

$$\begin{aligned}
E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} \left(\frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij}\right)^2\right] &= \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{n^2} E\left[\varepsilon_{11}^2 + \dots + \varepsilon_{1r_1}^2 + \varepsilon_{21}^2 + \dots + \varepsilon_{2r_2}^2 + \dots + \varepsilon_{a1}^2 + \dots + \varepsilon_{ar_a}^2 \right. \\
&\quad \left. + 2\left(\varepsilon_{11}\varepsilon_{12} + \varepsilon_{11}\varepsilon_{13} + \dots + \varepsilon_{11}\varepsilon_{1r_1} + \varepsilon_{11}\varepsilon_{21} + \dots + \varepsilon_{11}\varepsilon_{2r_2}\right) \right. \\
&\quad \left. + \dots + \varepsilon_{11}\varepsilon_{ar_a} + \dots + \varepsilon_{a1}\varepsilon_{ar_a}\right] \\
&= \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{n^2} n \sigma^2 = \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{\sigma^2}{n}
\end{aligned}$$

Sustituyendo en  $E(SC_{\text{Trot}})$

$$\begin{aligned}
E(SC_{\text{Trot}}) &= \sum_{i=1}^a \sum_{j=1}^{r_i} \alpha_i^2 + \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{\sigma^2}{r_i} + \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{\sigma^2}{n} - 2 \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{\sigma^2}{n} \\
&= \sum_{i=1}^a \sum_{j=1}^{r_i} \alpha_i^2 + a \sigma^2 - \sigma^2
\end{aligned}$$



$$E(CM_{\text{trat}}) = \sigma^2 + \frac{\sum_{i=1}^a \sum_{j=1}^{r_i} (\alpha_i)^2}{a-1}$$

Para el Error:

$$E(CM_{\text{EE}}) = E\left(\frac{SC_{\text{EE}}}{n-a}\right) = \frac{1}{n-a} E(SC_{\text{EE}})$$

$$E(SC_{\text{EE}}) = E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i\cdot})^2\right]$$

$$\begin{aligned} E(SC_{\text{EE}}) &= E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} \left[ (\mu + \alpha_i + \varepsilon_{ij}) - \left( \mu + \alpha_i + \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij} \right) \right]^2\right] \\ &= E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} \left( \varepsilon_{ij} - \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij} \right)^2\right] \\ &= E\left(\sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij}^2\right) - 2E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij} \left( \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij} \right)\right] + E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} \left( \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij} \right)^2\right] \end{aligned}$$

$$\begin{aligned} 2E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} \varepsilon_{ij} \left( \frac{1}{r_i} \sum_{j=1}^{r_i} \varepsilon_{ij} \right)\right] &= \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{r_i} E\left(\varepsilon_{ij} \sum_{j=1}^{r_i} \varepsilon_{ij}\right) = \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{r_i} E(\varepsilon_{ij} \varepsilon_{i1} + \varepsilon_{ij} \varepsilon_{i2} + \dots + \varepsilon_{ij} \varepsilon_{ir_i}) \\ &= \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{1}{r_i} E(\varepsilon_{ij}^2) = \sum_{i=1}^a \sum_{j=1}^{r_i} \frac{\sigma^2}{r_i} = a\sigma^2 \end{aligned}$$

$$E(SC_{\text{EE}}) = n\sigma^2 - 2a\sigma^2 + a\sigma^2 = (n-a)\sigma^2$$

$$E(CM_{\text{EE}}) = \frac{1}{n-a} (n-a)\sigma^2 = \sigma^2$$

### Modelo de Efectos Aleatorios (Modelo I I):

Esperanza de los Cuadrados Medios

Para los tratamientos:

$$E(CM_{trat}) = E\left(\frac{SC_{trat}}{a-1}\right) = \frac{1}{a-1}E(SC_{trat})$$

$$E(CM_{trat}) = \frac{1}{a-1}E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} (\bar{y}_{i\cdot} - \bar{y}_{..})^2\right]$$

$$\begin{aligned} E(SC_{trat}) &= E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} (\bar{y}_{i\cdot} - \bar{y}_{..})^2\right] \\ &= E\left\{\sum_{i=1}^a \sum_{j=1}^{r_i} \left[\left(\mu + \alpha_i + \frac{1}{r_i} \varepsilon_{i\cdot}\right) - \left(\mu + \frac{1}{n} \sum_{i=1}^a r_i \alpha_i + \frac{1}{n} \varepsilon_{..}\right)\right]^2\right\} \\ &= E\left\{\sum_{i=1}^a \sum_{j=1}^{r_i} \left[\left(\alpha_i - \frac{1}{n} \sum_{i=1}^a r_i \alpha_i\right) + \left(\frac{1}{r_i} \varepsilon_{i\cdot} - \frac{1}{n} \varepsilon_{..}\right)\right]^2\right\} \\ &= E\left\{\sum_{i=1}^a \sum_{j=1}^{r_i} \left[\left(\alpha_i - \frac{1}{n} \sum_{i=1}^a r_i \alpha_i\right)^2 + 2\left(\alpha_i - \frac{1}{n} \sum_{i=1}^a r_i \alpha_i\right)\left(\frac{1}{r_i} \varepsilon_{i\cdot} - \frac{1}{n} \varepsilon_{..}\right) + \left(\frac{1}{r_i} \varepsilon_{i\cdot} - \frac{1}{n} \varepsilon_{..}\right)^2\right]\right\} \end{aligned}$$

$$\begin{aligned} E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} \left(\alpha_i - \frac{1}{n} \sum_{i=1}^a r_i \alpha_i\right)^2\right] &= E\left\{\sum_{i=1}^a \sum_{j=1}^{r_i} \left[\alpha_i^2 - 2\alpha_i \left(\frac{1}{n} \sum_{i=1}^a r_i \alpha_i\right) + \left(\frac{1}{n} \sum_{i=1}^a r_i \alpha_i\right)^2\right]\right\} \\ &= \sum_{i=1}^a r_i \sigma_\alpha^2 - \frac{2}{n} \sum_{i=1}^a r_i E\left[\alpha_i (r_1 \alpha_1 + \dots + r_a \alpha_a)\right] + \frac{1}{n^2} \sum_{i=1}^a r_i E\left[\left(\sum_{i=1}^a r_i \alpha_i\right)^2\right] \\ &= \sigma_\alpha^2 \left(\sum_{i=1}^a r_i - \frac{\sum_{i=1}^a r_i^2}{\sum_{i=1}^a r_i}\right) \end{aligned}$$

$$\left(\frac{1}{r_i}\varepsilon_{i.}-\frac{1}{n}\varepsilon_{..}\right)^2 = \left(\frac{1}{r_i}\sum_{j=1}^{r_i}\varepsilon_{ij}\right)^2 - 2\left(\frac{1}{r_i}\sum_{j=1}^{r_i}\varepsilon_{ij}\right)\left(\frac{1}{n}\sum_{i=1}^a\sum_{j=1}^{r_i}\varepsilon_{ij}\right) + \left(\frac{1}{n}\sum_{i=1}^a\sum_{j=1}^{r_i}\varepsilon_{ij}\right)^2$$

Sustituyendo se obtiene

$$E(CM_{\text{trat}}) = \sigma^2 + \frac{\sum_{i=1}^a r_i - \left[ \sum_{i=1}^a (r_i)^2 / \sum_{i=1}^a r_i \right]}{a-1} \sigma_\alpha^2 = \sigma^2 + k\sigma_\alpha^2$$

Donde  $k = \frac{\sum_{i=1}^a r_i - \left[ \sum_{i=1}^a (r_i)^2 / \sum_{i=1}^a r_i \right]}{a-1}$

Para el Error:

$$E(CM_{\text{EE}}) = E\left(\frac{SC_{\text{EE}}}{n-a}\right) = \frac{1}{n-a} E(SC_{\text{Trat}})$$

$$E(SC_{\text{EE}}) = E\left[\sum_{i=1}^a \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i.})^2\right]$$

desarrollando esta suma y aplicando las propiedades de la esperanza y los supuestos para este modelo, se obtiene:

$$E(CM_{\text{EE}}) = \frac{1}{n-a} (n-a) \sigma^2 = \sigma^2$$

## APENDICE C

### Muestreo Aleatorio Simple

#### A. Tiempo de Desarrollo de la Enfermedad:

En las figuras 1 a 6 se muestran las plantas seleccionadas en cada variedad para el muestreo aleatorio simple, en la variable Tiempo de Desarrollo de la Enfermedad.

Parcelas donde se ubica la Variedad I														
5			7			13			23			30		
	X	X			X	X	X		X	X			X	
X	X	X	X	X							X		X	
		X	X		X			X					X	
					X				X					

Figura 1: Selección de la muestra  $n=22$  en el muestreo aleatorio simple de la variedad I (‘Fagro 122’), para evaluar el Tiempo de Desarrollo de la Enfermedad.

Parcelas donde se ubica la Variedad II														
3			12			18			19			29		
			X				X		X			X		
X		X	X	X	X				X	X			X	
X		X			X	X								X
X			X			X		X				X		

Figura 2: Selección de la muestra  $n=22$  en el muestreo aleatorio simple de la variedad II (‘Gran nain’), para evaluar el Tiempo de Desarrollo de la Enfermedad.

Parcelas donde se ubica la Variedad III														
4			8			15			21			25		
	X	X		X					X	X	X			
X		X	X	X								X	X	
			X	X		X		X		X		X		
X												X	X	X

Figura 3: Selección de la muestra  $n=22$  en el muestreo aleatorio simple de la variedad III (‘Pisang más’), para evaluar el Tiempo de Desarrollo de la Enfermedad.

Parcelas donde se ubica la Variedad IV														
6			10			17			22			27		
	X			X			X		X		X	X		
	X		X											
X			X		X	X	X			X	X	X		
X	X			X							X	X		X

Figura 4: Selección de la muestra  $n=22$  en el muestreo aleatorio simple para la variedad IV (‘Brasilero’)

Parcelas donde se ubica la Variedad V														
2			9			16			24			26		
X	X			X			X		X	X	X			X
	X					X	X			X	X	X		
X								X		X				X
			X					X			X	X		

Figura 5: Selección de la muestra  $n=22$  en el muestreo aleatorio simple de la variedad V (‘FHIA-02’), para evaluar el Tiempo de Desarrollo de la Enfermedad.

Parcelas donde se ubica la Variedad VI														
1			11			14			20			28		
		X			X	X	X	X		X	X		X	
X				X				X		X				X
X	X		X										X	
		X			X				X	X		X		

Figura 6: Selección de la muestra  $n=22$  en el muestreo aleatorio simple de la variedad VI (‘Cien Bta-03’), para evaluar el Tiempo de Desarrollo de la Enfermedad.

### B. Número de Frutos

En las figuras 7 a 12 se muestran las plantas seleccionadas en cada variedad para el muestreo aleatorio simple, en la variable Número de Frutos.

Parcelas donde se ubica la Variedad I														
5			7			13			23			30		
				X								X		
		X										X		
X			X	X		X								
											X	X		

Figura 7: Selección de la muestra  $n=10$  en el muestreo aleatorio simple de la variedad I (‘Fagro 122’), para evaluar el Número de Frutos

Parcelas donde se ubica la Variedad II														
3			12			18			19			29		
X	X						X							
	X					X						X		
X							X				X		X	
			X					X			X	X		

Figura 8: Selección de la muestra  $n=14$  en el muestreo aleatorio simple DE la variedad II (‘Gran nain’), para evaluar el Número de Frutos

Parcelas donde se ubica la Variedad III														
4			8			15			21			25		
	X											X		
	X					X						X		X
			X	X								X		

Figura 9: Selección de la muestra  $n=9$  en el muestreo aleatorio simple de la variedad III (‘Pisang más’), para evaluar el Número de Frutos

Parcelas donde se ubica la Variedad IV														
6			10			17			22			27		
	X	X	X	X					X	X	X			
X		X	X	X					X				X	
			X	X		X		X		X		X		
X				X	X							X	X	X

Figura 10: Selección de la muestra  $n=25$  en el muestreo aleatorio simple de la variedad IV (‘Brasileiro’), para evaluar el Número de Frutos

Parcelas donde se ubica la Variedad V														
2			9			16			24			26		
			X			X								
X					X					X	X			X
		X									X	X		X
			X						X		X			

Figura 11: Selección de la muestra  $n=14$  en el muestreo aleatorio simple de la variedad V (‘FHIA-02’), para evaluar el Número de Frutos

Parcelas donde se ubica la Variedad VI														
1			11			14			20			28		
										X	X			
		X	X	X		X	X			X	X			
						X		X						
	X		X						X	X				

Figura 12: Selección de la muestra  $n=15$  en el muestreo aleatorio simple de la variedad VI (‘Cien Bta-03’), para evaluar el Número de Frutos

### C. Número de Frutos, considerando la mayor varianza

En las figuras 13 a 18 se muestran las plantas seleccionadas en cada variedad para el muestreo aleatorio simple, en la variable Número de Frutos para el mayor tamaño de muestra.

Parcelas donde se ubica la Variedad I														
5			7			13			23			30		
	X			X			X		X			X		X
	X	X	X				X					X	X	X
X	X		X	X		X	X				X	X		
				X							X	X		X

Figura 13: Selección de la muestra  $n=25$  en el muestreo aleatorio simple de la variedad I (‘Fagro 122’), para evaluar el Número de Frutos con  $n$  mayor

Parcelas donde se ubica la Variedad II														
3			12			18			19			29		
X	X			X			X		X		X			X
X	X					X	X			X	X	X	X	
X							X			X	X		X	X
			X					X			X	X		

Figura 14: Selección de la muestra  $n=25$  en el muestreo aleatorio simple DE la variedad II (‘Gran nain’), para evaluar el Número de Frutos con  $n$  mayor



Parcelas donde se ubica la Variedad III														
4			8			15			21			25		
	X	X		X					X	X	X	X		X
X	X		X			X						X		X
			X	X		X				X		X		
				X	X			X				X	X	X

Figura 15: Selección de la muestra  $n=25$  en el muestreo aleatorio simple de la variedad III (‘Pisang más’), para evaluar el Número de Frutos con  $n$  mayor

Parcelas donde se ubica la Variedad IV														
6			10			17			22			27		
	X	X	X	X					X	X	X			
X		X	X	X					X				X	
			X	X		X		X		X		X		
X				X	X							X	X	X

Figura 16: Selección de la muestra  $n=25$  en el muestreo aleatorio simple de la variedad IV (‘Brasileiro’), para evaluar el Número de Frutos con  $n$  mayor

Parcelas donde se ubica la Variedad V														
2			9			16			24			26		
			X			X	X	X		X		X	X	
X				X	X		X			X	X			X
		X									X	X	X	X
			X		X		X		X		X	X		

Figura 17: Selección de la muestra  $n=25$  en el muestreo aleatorio simple de la variedad V (‘FHIA-02’), para evaluar el Número de Frutos con  $n$  mayor

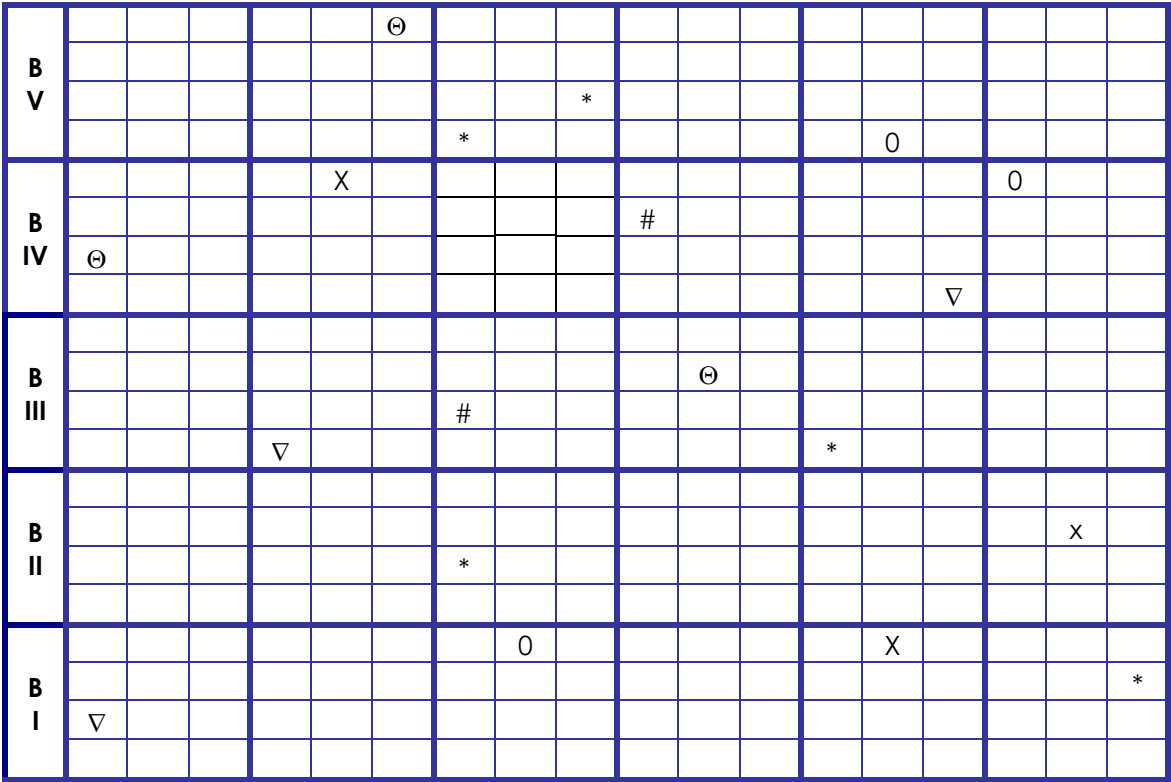
Parcelas donde se ubica la Variedad VI														
1			11			14			20			28		
	X			X						X	X			
		X	X	X		X	X			X	X		X	
			X	X		X		X				X		
X	X		X						X	X		X	X	X

**Figura 18: Selección de la muestra  $n=25$  en el muestreo aleatorio simple de la variedad VI (‘Cien Bta-03’), para evaluar el Número de Frutos con  $n$  mayor**



**B. Número de Frutos**

En la figura 2 se muestran las plantas seleccionadas en cada variedad para el muestreo estratificado aleatorio, en la variable Número de Frutos.



**Figura 2: Plano en campo del experimento en Musáceas, de las plantas seleccionadas en cada parcela, en el Muestreo Estratificado Aleatorio, para evaluar el Número de Frutos. Donde; X representan a la variedad I, los 0 a la variedad II, # a la variedad III, \* a la variedad IV, Θ a la variedad V y ∇ a la variedad VI.**

**C. Número de Frutos, considerando la mayor varianza**



## APENDICE E

### Muestreo Estratificado Doble

#### A. Tiempo de Desarrollo de la Enfermedad:

En la figura 1 se muestran las plantas seleccionadas en cada variedad para el muestreo estratificado doble, en la variable Tiempo de Desarrollo de la Enfermedad.

<b>B V</b>			#		⊖	⊖	*						0				X	
	#		#		⊖					▽	▽	▽			0		X	X
	#					⊖	*	*		▽			0		0	X		X
			#		⊖		*		*			▽	0					
<b>B IV</b>		⊖	⊖		X			*					▽				0	
	⊖					X		*	*	#		#		▽			0	0
	⊖			X					*	#	#	#	▽		▽			0
	⊖			X	X			*					▽				0	
<b>B III</b>		X	X		▽							⊖						
	X				▽	▽	#	#		⊖			*	*	*			
	X		X	▽	▽		#				⊖	⊖			*	0	0	0
							#	#				⊖	*			0		0
<b>B II</b>		0		▽	▽					⊖			#	#			X	
		0	0		▽	▽	*		*						#	X	X	X
			0			▽	*				⊖	⊖	#					
		0					*	*		⊖		⊖			#			X
<b>B I</b>		▽			⊖		0						X					*
		▽	▽	⊖				0	0	#	#	#						*
	▽		▽				0						X		X	*	*	*
				⊖	⊖	⊖			0		#	#	X	X				

Figura 1: Plano en campo del experimento en Musáceas, de las plantas seleccionadas en cada parcela, en el Muestreo Estratificado Doble, para evaluar el Tiempo de Desarrollo de la Enfermedad. Donde las X representan a la variedad I, los 0 a la variedad II, # a la variedad III, \* a la variedad IV, ⊖ a la variedad V y ▽ a la variedad VI.

## B. Número de Frutos

En la figura 2 se muestran las plantas seleccionadas en cada variedad para el muestreo estratificado aleatorio, en la variable Número de Frutos.

<b>B V</b>								*					0					X
						Θ			*	▽	▽	▽			0			
	#	#							*			▽					X	
				Θ	Θ													
<b>B IV</b>			Θ							#					▽			
		Θ				X	*							▽		0		
	Θ						*											
		Θ				X	*			#				▽		0		0
<b>B III</b>													*	*				
			X		▽	▽				Θ		Θ		*	*	0	0	
			X		▽		#		#				*		*			
														*				0
<b>B II</b>																		
			0	▽		▽	*	*		Θ	Θ	Θ	#					
									*							X		
		0				▽	*	*	*			Θ						
<b>B I</b>			▽				0	0									*	
								0	0		#						*	*
		▽		Θ									X		X	*		*
											#			X	*			

**Figura 2: Plano en campo del experimento en Musáceas, de las plantas seleccionadas en cada parcela, del muestreo estratificado doble, para evaluar el Número de frutos (dedos). X representan a la variedad I, los 0 a la variedad II, # a la variedad III, \* a la variedad IV, Θ a la variedad V y ▽ a la variedad VI.**

### C. Número de Frutos, considerando la mayor varianza

En la figura 3 se muestran las plantas seleccionadas en cada variedad para el muestreo estratificado aleatorio, en la variable Número de Frutos, con el tamaño de muestra mayor.

<b>B V</b>		#	#					*			▽	▽	0	0				X
		#				⊖			*	▽	▽	▽			0	X		
	#	#		⊖					*			▽		0		X	X	
		#		⊖	⊖	⊖				▽								
<b>B IV</b>	⊖	⊖	⊖		X					#				▽	▽		0	
		⊖	⊖			X	*			#				▽		0	0	
	⊖				X		*				#					0		
		⊖				X	*			#	#		▽	▽		0		0
<b>B III</b>							#				⊖		*	*				
			X		▽	▽	#		#	⊖		⊖		*	*	0	0	
	X	X	X		▽		#		#	⊖			*		*		0	
			X		▽	▽		#						*			0	0
<b>B II</b>		0									⊖	⊖						X
			0	▽	▽	▽	*	*		⊖	⊖	⊖	#					
									*	⊖				#		X		
		0		▽		▽	*	*	*			⊖	#			X	X	
<b>B I</b>			▽			⊖	0	0	0	#				X			*	
								0	0	#	#		X	X	X		*	*
		▽	▽	⊖			0			#			X	X	X	*		*
									0			#			X	*		

**Figura 3: Plano en campo del experimento en Musáceas, de las plantas seleccionadas en cada parcela, del muestreo estratificado doble, para evaluar el Número de frutos (dedos) con n mayor. X representan a la variedad I, los 0 a la variedad II, # a la variedad III, \* a la variedad IV, ⊖ a la variedad V y ▽ a la variedad VI.**