

UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE AGRONOMIA
POSTGRADO EN ESTADISTICA

**COMPARACION DE LA CALIDAD DE LOS AJUSTES OBTENIDOS
MEDIANTE LOS MODELOS DE RESPUESTA BINARIA: LOGIT VS
PROBIT, EN INDICES DE INESTABILIDAD ATMOSFERICA PARA LA
OCURRENCIA O NO DE PRECIPITACIONES.**

Autor: Daniel Miranda

Maracay, julio del 2016

Trabajo de grado presentado como requisito parcial final para optar al título de
Magister Scientiarum en Estadística.

Tutor:

Dr. Miguel Balza

Autor:

Prof. Daniel Miranda

DEDICATORIA.

Este trabajo de investigación va dedicado de manera muy especial a todas aquellas personas que, de una manera u otra, siempre han estado a mi lado apoyándome, han creído en mí y que han contribuido enormemente en todos los logros que he obtenido tanto en mi vida personal como en mi vida profesional. A mis padres Olga Margarita Cedres de Miranda y Evelio Antonio Miranda, aunque no están físicamente siempre están presente en mis recuerdos y en mi corazón. A mi hija Dayenesis Miranda Silva por ser mi motivo y mi razón de ser. A mis hermanas Olga Miranda y Adriana Miranda por mantener el apoyo irrestricto que siempre necesité.

AGRADECIMIENTOS.

Mis más sinceros agradecimientos para mi tutor académico el Dr. Miguel Balza y la MSc. Isabel Díaz por todo el apoyo prestado, por su valiosa contribución en la asesoría de este trabajo y por su excelente labor como profesores incansables. Un verdadero privilegio para mí contar con eso. Agradezco al Instituto Nacional de Meteorología e Hidrología (INAMEH) por darme el apoyo técnico y científico necesario para la ejecución del presente estudio. A mi hermano Richard Miranda por iluminar el camino de esta investigación fundamentando las ideas desde sus inicios.

INDICE DE CONTENIDOS.

DEDICATORIA -----	iii
AGRADECIMIENTOS -----	iv
INDICE DE CONTENIDOS -----	v
INDICE DE CUADROS -----	vii
RESUMEN -----	8
ABSTRACT -----	9
I. INTRODUCCION -----	10
OBJETIVOS DE LA INVESTIGACION -----	12
OBJETIVO GENERAL-----	12
OBJETIVOS ESPECIFICOS-----	12
II. MARCO TEORICO -----	13
2.1 GENERALIZACION DE LOS MODELOS-----	13
2.2 VARIABLE LATENTE-----	17
2.3 MULTICOLINEALIDAD-----	18
2.4 METODO DE MAXIMO VEROSIMIL-----	20
III. MATERIALES Y METODOS -----	22
3.1 POBLACION Y MUESTRA-----	22
3.1.1 TAMAÑOS MUESTRALES-----	23
3.1.2 VARIABLES PREDICTORAS-----	25
3.1.3 RECOLECCION DE DATOS-----	27
3.2 SELECCIÓN DE VARIABLES-----	27
3.2.1 METODO PASO ADELANTE-----	28
3.2.2 METODO PASO ATRÁS-----	28
3.2.3 METODO PASO A PASO-----	29

3.3 METODO DE MAXIMA VEROSIMILITUD-----	29
3.4 PRUEBA DE AJUSTE DE PEARSON-----	30
3.5 PRUEBA DE AJUSTE DE HOSMER-LEMESHOW-----	31
3.6 PORCENTAJES DE ACIERTOS ESTIMADOS EN EL MODELO-----	32
3.6.1 TASAS DE ACIERTOS-----	34
3.6.2 TASAS DE ERRORES-----	34
3.6.3 ESPECIFICIDAD-----	34
3.6.4 SENSIBILIDAD-----	34
3.6.5 TASA DE FALSOS CEROS-----	35
3.6.6 TASA DE FALSOS UNOS-----	35
IV. RESULTADOS Y DISCUSION-----	36
4.1 SELECCIÓN DE VARIABLES-----	36
4.2 FACTOR DE LA VARIANZA INFLADA-----	38
4.3 ESTIMACION DE LOS PARAMETROS-----	40
4.4 PRUEBAS DE AJUSTE-----	41
4.4.1 PRUEBA DE HOSMER LEMESHOW Y PRUEBA DE PEARSON-----	42
4.4.2 PORCENTAJES DE ACIERTOS-----	43
4.5 ESTUDIO DE LAS VARIABLES LATENTES-----	47
V. CONCLUSIONES-----	52
VI. RECOMENDACIONES-----	54
VII. REFERENCIAS BIBLIOGRAFICAS-----	55
VIII. BIBLIOGRAFIA ADICIONAL CONSULTADA-----	58
IX. ANEXOS-----	60
ANEXO 1. RECOLECCION DE DATOS-----	60
ANEXO 2. MUESTREO ALEATORIZADO POR EXCEL-----	77
ANEXO 3. SINTAXIS UTILIZADA EN SPSS-----	78
ANEXO 4. RESULTADOS EN MINITAB-----	80

INDICE DE CUADROS.

CUADRO 3.1.1 TAMAÑOS MUESTRALES-----	24
CUADRO 3.6 CUADRO DE PORCENTAJES-----	33
CUADRO 4.1 MATRIZ DE CORRELACIONES-----	36
CUADRO 4.1.1 VARIABLES SELECCIONADAS EN MUESTRAS ALEATORIZADAS-----	37
CUADRO 4.1.2 VARIABLES SELECCIONADAS EN MUESTRAS NO ALEATORIZADAS----	38
CUADRO 4.2.1 VALORES DE VIF PARA CADA MUESTRA ALEATORIZADA-----	39
CUADRO 4.2.2 VALORES DE VIF PARA CADA MUESTRA NO ALEATORIZADA-----	39
CUADRO 4.3.1 VALOR DE PARAMETROS PARA CADA MUESTRA ALEATORIZADA----	40
CUADRO 4.3.2 VALOR DE PARAMETROS PARA CADA MUESTRA NO ALEATORIZ.-----	41
CUADRO 4.4.1.1 PRUEBAS DE AJUSTES EN MUESTRAS ALEATORIZADAS-----	42
CUADRO 4.4.1.2 PRUEBAS DE AJUSTES EN MUESTRAS NO ALEATORIZADAS-----	42
CUADRO 4.4.2.1 PORCENTAJE DE ACIERTOS PARA LA MUESTRA COMPLETA-----	44
CUADRO 4.4.2.2 PORCENTAJE DE ACIERTOS PARA LA MUESTRA 5 ALEATORIZADA--	45
CUADRO 4.4.2.3 PORCENTAJE DE ACIERTOS PARA LA MUESTRA 6 NO ALEATOR.-----	46
CUADRO 4.5.1 CORRELACIONES ENTRE LAS VARIABLES LATENTES-----	47
CUADRO 4.5.2 ESTADISTICOS DESCRIPTIVOS-----	48
CUADRO 4.5.3 RESUMEN DEL MODELO-----	50
CUADRO 4.5.4 ANOVA-----	50
CUADRO 4.5.5 COEFICIENTES-----	51

RESUMEN.

En este trabajo de investigación se realizó un estudio comparativo de los ajustes obtenidos a través de los modelos de respuesta binaria Logit y Probit para destacar todas las características que son propia de cada uno y que proporcionan al investigador información de interés. El estudio se desarrolló utilizando datos del Instituto Nacional de Meteorología e Hidrología (INAMEH), ya que la precipitación es una variable que fácilmente puede ser recodificada, transformándola a una variable cualitativa binaria en donde 1 representa la ocurrencia de lluvias y 0 la no ocurrencia de lluvias. Se trabajó con una muestra grande que más tarde fue submuestreada en otras más pequeñas, esto para observar cómo se comportaban los modelos a medida que se aumentaba el tamaño de las muestras. Se consideraron diferentes pruebas de ajustes para la evaluación de los modelos, así como también la precisión y capacidad predictora de los mismos. Se utilizó el método de máxima verosimilitud por ser más eficiente en la estimación de los parámetros en cada una de las expresiones matemáticas usando el paquete estadístico minitab, al mismo tiempo que dicho software estadístico también seleccionaba las variables regresoras aplicando el método paso a paso. Después de realizar los análisis pertinentes y guardar toda la información requerida, se pudo observar mayor flexibilidad en el modelo Probit en la inclusión de variables predictoras, proporcionando esta particularidad mayor estabilidad estadística al modelo e incidencias positivas en el ajuste. Se encontró una relación casi perfectamente exponencial entre los Odds y la variable Latente del Probit, lo que se definió en una ecuación matemática. Como también se observaron en la mayoría de los casos estudiados una mejor sensibilidad en el Probit sobre el Logit, no así en la especificidad, en donde fue lo contrario. La escogencia de uno de los dos modelos se fundamenta principalmente en los objetivos del investigador, sobre qué información además de P01 (probabilidad de que ocurra el evento) es de su interés, teniendo presente todo lo discutido aquí, en cuanto a flexibilidad, precisión, característica teórica de los modelos y capacidad predictora.

Palabras clave: ajuste – comparación – modelos de respuesta binaria.

ABSTRACT.

In this research, it was made a comparative study about adjustments obtained through binary response models Logit and Probit to highlight all the features that are typical of each one, and provide information of interest to the researcher. The study was developed using data from the National Institute of Meteorology and Hydrology (INAMEH) because the precipitation is a variable that can be easily recorded, transforming it to a binary qualitative variable, where 1 represents the occurrence of rains and 0 the non-occurrence of rains. It was worked with a large sample that was later sub-sampled into smaller, this to see how models behaved when the sample size was increased. Testing different adjustments for evaluating models were considered, as well as their accuracy and their predictive ability. The maximum likelihood method was used by being more efficient in estimating parameters in each of the mathematical expressions using the Minitab statistical package, while this statistic software also selected the regressor variables using the stepwise method. After making the pertinent analyzes and saving all the required information, it was observed greater flexibility in the Probit model in the inclusion of predictive variables, providing this feature, greater statistical stability to the model and positive impacts on the adjustment. An almost perfect exponential relationship was found between the odds and the Probit Latent variable, which was defined in a mathematical equation. It was also observed in most of the cases studied a better sensitivity in the Probit than in the Logit, but not in the specificity, where it was the opposite. The choice of one of the two models are mainly based on researcher objectives, about what information, besides P01 (probability of the event occurring) is of interest, having in mind everything discussed here, in terms of flexibility, precision, theoretical feature of the models and predictive capacity.

Keywords: adjustment – binary response models – comparison.

I. INTRODUCCION.

Los modelos Logit y Probit son modelos que surgieron de la necesidad de buscar métodos estadísticos que, aplicados a situaciones en donde solo exigen dos tipos de respuestas puedan proveer el ajuste necesario que permitan obtener representaciones matemáticas bastantes aproximadas. Bliss (1934), fue el primero en publicar un modelo de respuesta binaria en la revista “science” en donde la inversa de la función de distribución o función asociada con la distribución normal estándar la daría a conocer como Probit. Bliss propuso transformar el porcentaje eliminado en una unidad de probabilidad que estaría linealmente relacionada con la definición moderna (él la definió arbitrariamente como igual a 0 para 0,0001 y 1 para 0,9999). Este trabajo, dió origen a la regresión logística, cuando más tarde Berkson en 1944 introdujo el término Logit como la denominación de esta transformación, para ese entonces sus trabajos popularizaron la utilización de ese modelo.

Cochran y Finney (1979), también afirman que Bliss fue uno de los pioneros en la regresión logística, ya que trabajó conjuntamente con Fisher en la creación de nuevos métodos. Tal vez el más importante de ellos fue el desarrollo, de un método iterativo para encontrar estimaciones de máxima verosimilitud en el modelo Probit.

Ambos modelos, Logit y Probit tienden a comportarse de manera muy parecida, lo que hace difícil observar las distinciones de aplicación entre ellos, debido a sus grandes similitudes. Los dos tienen la particularidad de que, la distribución de la variable dependiente, al ser binaria, no puede ser normal, como también situaciones de heterocedasticidad en donde la varianza no es constante. El cumplimiento de los supuestos de la homocedasticidad y de la normalidad no son requisitos indispensables para la aplicación de estos modelos, Alderete A. (2006).

A pesar de que ambos modelos de respuesta binaria poseen expresiones matemáticas muy particulares, estos se aplican por lo general en las mismas situaciones, lo que hace difícil determinar cuál es el más indicado en cuanto al uso respectivo de cada uno de ellos. Enchautegui (2001), menciona que estos modelos no presentan diferencias en sentido cualitativo que puedan facilitar la elección de, cual modelo utilizar, sin embargo advierte que el modelo Probit a diferencia del Logit puede determinar la existencia de una variable latente continua, la cual puede influir en gran magnitud en la respuesta estimada, proporcionando información adicional que en un momento dado puede ser de gran utilidad para el investigador.

Amemiya (1994), destaca que, los parámetros del modelo Probit al ser multiplicados por el factor 1,61 pueden dar resultados aproximados y muy cercanos al valor de los parámetros del modelo Logit, pero si la cantidad de observaciones es extremadamente grande entonces las estimaciones de ambos modelos difieren substancialmente. Este argumento es lo que lleva a pensar que la cantidad de observaciones, de muestras y variables pueden proporcionar detalles importantes de los modelos, como también puedan ayudar establecer las condiciones en que cada uno de ellos trabaja de la manera más apropiada, al mismo tiempo identificar bajo qué escenario puede ser considerada su aplicabilidad.

Existen diferentes áreas donde los modelos Logit y Probit son frecuentemente utilizados, como por ejemplo: en el área farmacéutica, específicamente en la fase experimental de los medicamentos con la finalidad de definir los efectos que estos tienen sobre los pacientes. También en la meteorología se suele recurrir a estos modelos, y hasta forma parte de la estructura lógica - algebraica de algunas redes neuronales empleadas para las investigaciones climatológicas. A los efectos de esta investigación se va a trabajar con los índices de estabilidad atmosférica para la ocurrencia o no de precipitaciones, lo que encaja perfectamente como la clase de situaciones con respuestas dicotómicas en los que estos modelos generalmente son aplicados, permitiendo así realizar los estudios correspondientes para cumplir con los objetivos aquí planteados.

En este trabajo se tratará de manejar no solo desde el punto de vista de aplicación sino que proporcionará aportes en cuanto a todos los elementos que pueden definir a cada modelo y buscar características claras, dando pasos firmes en el análisis respectivo de los datos, acercándose a toma de decisiones mucho más acertadas a sabiendas de que se ha seleccionado el modelo más indicado. Y desde el punto de vista teórico, generará discusiones y reflexiones sobre lo que hoy en día se conoce acerca de estos modelos estadísticos.

OBJETIVOS DE LA INVESTIGACION.

Objetivo general:

Comparar la calidad de los ajustes obtenidos mediante los modelos de respuestas binarias Logit vs Probit, en índices de inestabilidad atmosférica para la ocurrencia o no de precipitaciones.

Objetivos específicos:

1. Determinar las características más relevantes de los modelos Logit y Probit bajo las mismas condiciones de aplicabilidad.
2. Evaluar el comportamiento de cada uno de los modelos de respuestas dicotómicas Logit y Probit, aplicados a los índices de inestabilidad atmosférica para la ocurrencia o no de precipitaciones en diferentes tamaños de muestras.
3. Comparar ambos modelos en cuanto a la precisión y capacidad predictora, a través de pruebas de ajustes.

II. MARCO TEORICO.

2.1 Generalización de los modelos.

Perez, Kizys y Manzanedo (2002), describen los modelos Logit y Probit partiendo de una combinación de variables discretas con modelos de regresión lineal múltiple, para este caso, se puede observar que por ser Y una variable binaria (sólo podrá tomar los valores 0 y 1), donde Y es la variable respuesta del modelo planteado, siempre se cumplirá que:

$$E(Y) = y_1P(Y = y_1) + y_2P(Y = y_2) \quad (2.1.1)$$

$$E(Y) = 0P(Y = 0) + 1P(Y = 1) = P(Y = 1) \quad (2.1.2)$$

Por otra parte se pensaría en utilizar un modelo de regresión lineal múltiple para explicar el comportamiento de la variable Y :

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_kX_k + \varepsilon \quad (2.1.3)$$

Bajo el supuesto habitual de que $E(\varepsilon) = 0$, y suponiendo conocidos los valores que toman las variables explicativas o predictoras (observaciones), se tendrá que:

$$E(Y) = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_kX_k \quad (2.1.4)$$

Al igualar las ecuaciones 2.1.2, 2.1.3 y 2.1.4 se llega al resultado que le da nombre al modelo de probabilidad lineal:

$$P(Y = 1) = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_kX_k = Y - \varepsilon \quad (2.1.5)$$

En donde también se puede expresar la variable dependiente **binaria** Y como la probabilidad de “éxito” más un término de perturbación:

$$Y = P(Y = 1) + \varepsilon = E(Y) + \varepsilon \quad (2.1.6)$$

Sin embargo, se acota que este modelo inicial no será el más indicado para representar el comportamiento de variables dependientes binarias, pues presenta varios problemas:

- El término de perturbación ε no se distribuye normalmente ya que no es una variable continua sino discreta, en este sentido, no se cumple el supuesto de normalidad.
- El término de perturbación es heterocedástico por lo tanto los estimadores mínimos cuadrados no son los más eficientes.
- Al estimar los parámetros del modelo de probabilidad lineal, se estará ajustando una recta a la nube de puntos. En donde el uso de dicha recta para predecir nuevos valores de Y (2.1.6), a partir de valores dados de X_i puede proporcionar valores mayores que 1 o menores que 0 (lo cual está en contradicción con la definición de probabilidad).

Dada una variable dependiente **binaria** Y , el modelo de regresión lineal presenta serias inconsistencias. Perez, Kizys y Manzanedo (2002) explica que para evitarlas, se han desarrollado modelos no lineales, los cuales tratan de resolver los problemas anteriores.

Dichos modelos de la forma:

$$Y = f(X_1; X_2; \dots; X_k) + \varepsilon \quad (2.1.7) \quad (\text{Formulación de una regresión no lineal})$$

Donde f es una función no lineal.

a) Modelo Logit:

Se tiene que:

$$f(X_1; X_2; \dots; X_k) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}} \quad (2.1.8)$$

$$\text{Sea } P = f(X_1; X_2; \dots; X_k) \quad (2.1.9)$$

$$\text{Ln} \left(\frac{P}{1-P} \right) = \beta_0 + \sum_{i=1}^k \beta_i X_i \quad (2.1.10)$$

Si se representa

$$g(P) = \text{Ln} \left(\frac{P}{1-P} \right) \quad (2.1.11)$$

La función g se denomina **modelo Logit** la cual representa una función lineal de las variables explicativas dentro de una escala logarítmica.

b) Modelo Probit:

Ahora se tiene que:

$$f(X_1; X_2; \dots; X_k) = \int_{-\infty}^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (2.1.12)$$

Donde se desprende que si se iguala una variable P con la función:

$$P = f(X_1; X_2; \dots; X_k) \quad (2.1.13)$$

Entonces:

$$f^{-1}(P) = \beta_0 + \sum_{i=1}^k \beta_i X_i \quad (2.1.14)$$

La función $f^{-1}(P)$ se denomina **modelo Probit** y representada la inversa de la función de distribución normal acumulada.

La forma gráfica que tiene cada modelo se puede observar en la figura 1:

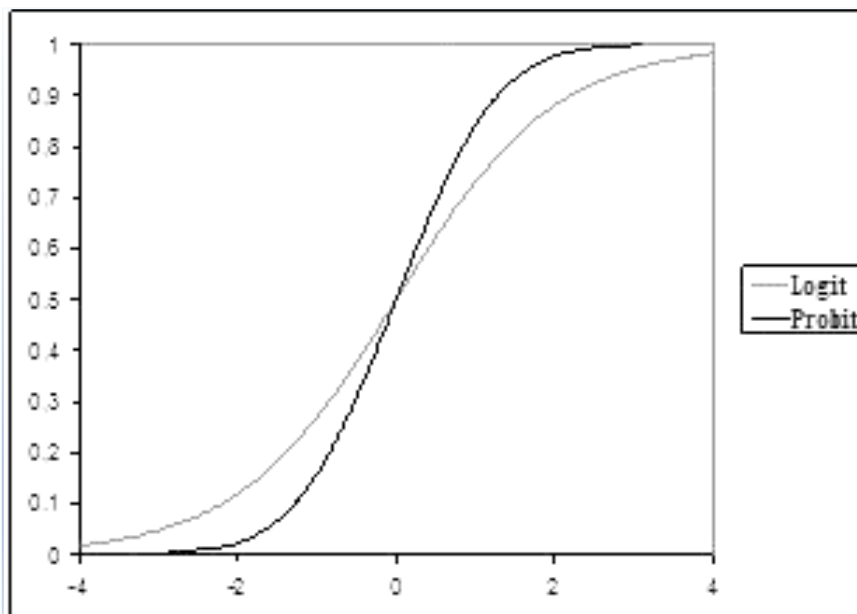


Figura 1

Fuente: página web <http://www.statisticssolutions.com/what-is-logistic-regression/>

Agresti (2007), reitera que en la práctica, el modelo Probit y los modelos de regresión logística (Logit) proporcionan ajustes similares. Sí un modelo de regresión logística se ajusta bien, entonces también lo hace el modelo Probit, y viceversa.

Walpole (2007), menciona que la distribución básica para la respuesta binaria de una regresión logística es la de Bernoulli o la Binomial. La primera la ubica en estudios observacionales donde no hay corridas repetidas en cada nivel de regresor; mientras que la segunda será el caso cuando se utilice un diseño experimental

2.2 La Variable Latente.

Moore (2000), la identifica como una variable que no se incluye entre las variables estudiadas pero que tiene una importante incidencia en la relación entre ellas.

Enchautegui, (2001) describe la variable latente de un Probit por la ecuación:

$$Y^* = \sum_{k=0}^n \beta_k X_k + \mu \quad (2.2.1)$$

Donde $X_0 = 1$ representa la constante y X_i son variables explicativas con μ como término del error. La variable Y^* no es observada; no hay datos para Y^* . Lo que se observa es una variable dicotómica cuyo valor definido es 0 ó 1 dependiendo si la variable cruza o no el límite o umbral considerado:

$$Y^* = 1 \quad \text{si} \quad Y^* = \sum_{k=0}^n \beta_k X_k + \mu > 0 \quad (2.2.2)$$

$$\mu \geq -\sum_{k=0}^n \beta_k X_k \quad (2.2.3)$$

$$Y^* = 0 \quad \text{si} \quad Y^* = \sum_{k=0}^n \beta_k X_k + \mu < 0 \quad (2.2.4)$$

$$\mu < -\sum_{k=0}^n \beta_k X_k \quad (2.2.5)$$

Agrega que el modelo Logit no necesariamente esconde una variable latente continua, La variable Y^* del modelo Logit (identificado como Odds) puede resultar de un proceso de decisión cualitativa.

Salas (1996), presenta el Odds como el cociente entre $P(X=1)$ y $P(X=0)$. Los Odds son una característica particular del modelo Logit, donde se puede determinar la relación existente de la probabilidad del éxito con respecto a la probabilidad del fracaso en la ocurrencia de un evento, dada la información de las variables predictoras.

Para Sanchez (2012), la variable latente es una variable que es observada parcialmente. En el contexto binario las variables latentes pueden aparecer de dos formas. En primera instancia puede representar un índice/propensión para la ocurrencia de un evento. En segunda instancia (contexto de elección) puede representar la diferencia, en términos de utilidad entre tomar una opción u otra.

Medina (2003), dice que la variable latente depende de un conjunto de variables explicativas que generan las alternativas que se dan en la realidad y que permiten expresar el modelo dicotómico. Cuando la variable latente supera un determinado nivel, la variable de respuesta binaria toma el valor 1, y si no lo supera toma el valor 0. Añadiendo que la distribución de ε_i determina el tipo de modelo a estimar: si se supone una función de distribución uniforme, se utiliza el Modelo Lineal de Probabilidad truncado; si se distribuye como una normal con media cero y varianza uno, el modelo generado será un Probit; mientras que si se supone que se distribuye como una curva logística, se trataría de un modelo Logit.

2.3 Multicolínealidad.

Gujarati (2010), menciona que, el término multicolínealidad se atribuye a Ragnar Frisch. Originalmente, designaba una relación lineal “perfecta” o exacta entre algunas o todas las variables explicativas de un modelo de regresión. Para la regresión con k

variables que incluye las variables explicativas X_1, X_2, \dots, X_k (donde $X_1=1$ para todas las observaciones de forma que den cabida al término del intercepto), se dice que existe una relación lineal exacta si se satisface la siguiente condición:

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k = 0 \quad (2.3.1)$$

Donde $\lambda_1, \lambda_2, \dots, \lambda_k$, son constantes tales que no todas son simultáneamente iguales a cero.

Hoy en día, sin embargo, el término multicolinealidad incluye el caso en el cual hay k variables intercorrelacionadas pero no en forma perfecta, de la siguiente manera:

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k + v_i = 0 \quad (2.3.2)$$

Donde v_i es un término de error estocástico.

Draper (1998), expone que una de las formas para determinar si existe o no correlación entre las variables predictoras que puedan incidir directamente en los estimación de los parámetros en un modelo de regresión, es por medio del factor de varianza inflada (variance inflation factors VIF) para la cual realiza un estudio de ajuste de cada una de las variables X_i con respecto a las demás covariables. Donde:

$$VIF(\beta_i) = \frac{1}{1-R_i^2} ; \text{para } i = 1 \dots k \quad (2.3.3)$$

Y en la que k es el número de variables predictoras y R_i^2 es el coeficiente de determinación de cada variable explicativa o predictora con el resto, estipulando que si VIF es mayor que 10 entonces existe un caso de correlación múltiple para la variable.

Canavos (1988) agrega que si la correlación entre las variables predictoras es pequeña entonces los problemas serán de índole menor pero si existe una correlación

fuerte entre dos o más variables entonces los resultados de la regresión serán ambiguos, especialmente con respecto a los valores de los coeficientes de regresión estimados.

Flores (2001) hace referencia a Farrar y Glauber donde definen la multicolinealidad como una condición estadística hipotética que adoptan una serie de medidas jerárquicas para cada uno de los niveles determinados en: su presencia, severidad y localización en un conjunto de datos. Señalando que su severidad más que su existencia radica precisamente en su condición estadística más que matemática.

Mendenhall (2006), menciona que bajo la condición de alta colínealidad entre las variables predictoras puede tener ciertos efectos en el análisis, ya que los coeficientes de regresión estimados tendrán errores estándar grandes, causando imprecisión en intervalos de confianza y predicción. Agregar o eliminar una variable de predicción también puede causar cambios significativos en los valores de los otros coeficientes de regresión.

2.4 Método Máximo Verosímil.

Lopez R. (2006), menciona que este método, fundamentalmente trata de encontrar el valor del parámetro para el cual sea máxima la probabilidad de obtener los valores muestrales encontrados.

Alderete (2006), reafirma que en el método de máxima verosimilitud se seleccionan las estimaciones de los parámetros que hagan posible que los resultados observados sean lo más verosímiles posible. Agregando además que no establece restricción alguna respecto a las características de las variables predictoras, éstas pueden ser nominales, ordinales o de intervalos.

Enchautegui (2001) señala que si la variable dependiente es dicótoma el valor $E(Y/X)$ no es lineal en los parámetros entonces se deben usar métodos no lineales para estimar los parámetros tal como el método de máxima verosimilitud, en vez del método de cuadrados mínimos ordinarios.

Agresti (2007) destaca el enfoque paramétrico de un modelo estadístico que asume una familia de distribuciones de probabilidad como la binomial para la variable de respuesta. Basándose, en que se puede sustituir los datos observados en la fórmula para la función de probabilidad (*fdp*) y ver cómo esa probabilidad depende del valor del parámetro desconocido. La estimación de máxima verosimilitud de un parámetro es el valor para el que la probabilidad de los datos observados toma el valor más grande, es decir, es el valor del parámetro en el que la función de probabilidad toma su máximo.

Montgomery, D., Peck, E. y Vining, G (2006), asevera que los estimadores de máxima verosimilitud tienen mejores propiedades estadísticas que los estimadores de mínimos cuadrados ya que son insesgados y también poseen estimadores consistentes, donde la consistencia es una propiedad de muestras grandes que indican que los estimadores difieren del valor verdadero del parámetro.

III. MATERIALES Y METODOS.

Este Trabajo de investigación fue desarrollado en el Instituto Nacional de Meteorología e Hidrología (INAMEH), donde se contó con la información necesaria para el logro de los objetivos inicialmente planteados. A continuación se detalla el procedimiento llevado a cabo, considerando todas las condiciones requeridas en la metodología utilizada:

3.1 Población y muestra.

Para los efectos de esta investigación fueron considerados los índices de inestabilidad atmosféricas como las variables predictoras que inciden significativamente en las precipitaciones sobre un determinado lugar, de allí la importancia de las radiosondas para el pronóstico del tiempo, Concepción (2009). Los índices considerados como las variables predictoras de los modelos son:

- Temperatura de Rocío a 850 milibares (**Td850**).
- Energía Potencial Disponible para la Convección (**CAPE**).
- Inhibición de la Convección (**CINH**).
- Índice K de Whiting (**KI**).
- Índice de Humedad (**W**).
- Índice Cross Total (**CT**).
- Índice Total de Totales (**TT**).

Y como variable dependiente:

- Precipitación (**P01**).

Concepción (2009) los define como parámetros diseñados a partir de datos de temperatura y humedad de la atmósfera, a diferentes niveles de presión, que proporcionan una medida numérica de la estabilidad estática de una capa atmosférica dada y que son usados particularmente, para evaluar la probabilidad de chaparrones y tormentas.

Se consideró la precipitación diaria como la variable de respuesta binaria definida en la ocurrencia o no de lluvia en todas sus clasificaciones, es decir, todos los valores en milímetros mayores que cero fue recodificado como 1 y el resto de los valores como 0. Al mismo tiempo se tomó una muestra total de 615 observaciones para todas las magnitudes, de un histórico de 5 años (2010 hasta el 2014) comenzando desde mayo hasta agosto, para una cantidad de 123 observaciones por año, para cada una de las variables empleadas.

3.1.1 Tamaños muestrales.

De una muestra total de 615 casos fueron validados 594 con 4752 observaciones, lo que significa 594 días en un periodo de cinco años representando una pérdida del 3,4% del tamaño inicial de la muestra. Esto debido a que la información suministrada por la estación pluviométrica utilizada (Maiquetia código 80415) no registró 21 días dentro del lapso de tiempo seleccionado. Una vez seleccionados los datos se procedió a dividir la información en 9 conjuntos de datos, comenzando con 66 casos para el primero y terminando con 594 para el último, realizando un incremento de 66 en 66. Ejecutando un muestreo intencional u opinático, esto para mantener una proporción determinada en las muestras que sea útil para el posterior análisis, además de considerarse tamaños muestrales suficientes para correr los modelos Logit y Probit. Agresti (2007) menciona que el diseño del estudio debe determinar el tamaño muestral, donde se permita los datos ofrecer la información necesaria para detectar los efectos que estos tengan dentro de la investigación. Agrega que mientras más grande la muestra

se podrán observar hasta los efectos más débiles, considerándose un $n=10$ como el mínimo de un tamaño muestral para aplicación de cualquier tipo de modelo de respuesta binaria. En el cuadro 3.1.1 se presenta el tamaño muestral creciente de cada uno de los subconjuntos de datos utilizados para cada uno de los análisis posteriores respectivos:

CUADRO 3.1.1. Tamaños muestrales.

MUESTRA	TAMAÑO	0-A	1-A	0-NA	1-NA
M1	66	31	35	27	39
M2	132	66	66	57	75
M3	198	113	85	88	110
M4	264	148	116	115	149
M5	330	175	155	156	174
M6	396	211	185	192	204
M7	462	230	232	228	234
M8	528	274	254	271	257
M9	594	308	286	308	286

En este cuadro se muestran tanto la cantidad de ceros (no precipitación) como la cantidad de unos (ocurrencia de precipitación) que existen en cada una de las muestras. Es importante mencionar que la información mantiene una secuencia temporal definida y que en la práctica bajo esta norma se maneja en los procesos de calidad y de estudio científico dentro del área climatológica. Sin embargo, las observaciones para los efectos de esta investigación se manejan de forma secuencial y de forma aleatorizada en cada una de las sub-muestras asignadas, es decir, no aleatorizada (NA) y aleatorizada (A). Canavos (1988), señala que la aleatoriedad de una muestra asegura la aplicación correcta de la probabilidad para evaluar el riesgo inherente en un proceso inductivo.

Para aleatorizar cada una de las muestras se utilizó Excel 2007 solo desde la muestra M1 hasta la muestra M8 ya que la última contenía todos los casos, al mismo

tiempo se verificó si había un numero de caso repetido y así asegurar una selección limpia de los datos.

3.1.2 Variables predictoras.

El sondeo virtual se realizó a través de la página <https://www.ready.noaa.gov/READYamet.php> colocando la estación meteorológica de Maiquetía (80415) como punto de localización, para que mediante el modelo numérico GDAS (global data assimilation system) utilizando información satelital y geoestadística proporcionara el valor de los índices de inestabilidad atmosférica diarios requeridos para la evaluación de los modelos Logit y Probit. Todas las observaciones tanto P01 como las variables predictoras fueron seleccionados desde las 12 UTC (7:30 am hora legal) hasta las 12 UTC del siguiente día, quedando este registrado en los datos. Para ilustrar mejor el formato del sondeo virtual se presenta la información del 1 de mayo del 2010 como ejemplo:

GDAS1 Sounding

```

GDAS1                Sounding                for                80415
CARACAS/S._BOLIVAR, --, VN ( Lat: 10.60 Lon: -66.98 elevation: 48 m )

File start time : 10 5 1 0 0
File ending time: 10 5 7 21 0
Chosen date in meteorological file:                10                5                1
12

YR: 2010  MON: 05  DAY: 01  HOUR: 12  AT POSITION: 294.0 101.6  LAT.:
10.60  LON.: -66.98  ELEV: 48 m
PRSS: 0.9824E+03
MSLP: 0.1014E+04
TPP6: 0.0000E+00
UMOF: 0.6562E-01
VMOF: 0.1625E-01
SHTF: -0.3136E+01
DSWF: 0.4774E+02
RH2M: 0.7655E+02
U10M: -0.1627E+01
V10M: 0.1601E+01
T02M: 0.2996E+03
TCLD: 0.4145E+02
SHGT: 0.2733E+03

```

CAPE: 0.5142E+03**
 CINH: -0.3901E+02**
 LISD: 0.2726E+03
 LIB4: -0.1394E+00
 PBLH: 0.6092E+03
 TMPS: 0.2998E+03
 CPP6: 0.1952E-05
 SOLM: 0.1369E+00
 CSNO: 0.0000E+00
 CICE: 0.0000E+00
 CFZR: 0.0000E+00
 CRAI: 0.5960E+00
 LH7F: 0.8471E+02
 LCLD: 0.4467E+01
 MCLD: 0.0000E+00
 HCLD: 0.3809E+02

PRESS HGT (MSL) TEMP DEW PT WND DIR WND SPD
 HPA M C C DEG M/S
 E = Estimated Surface Height

982.	273.	26.5	22.0	134.5	2.3
975.	343.	25.3	20.7	122.8	4.1
950.	569.	23.9	19.2	122.4	4.5
925.	803.	22.3	17.6	124.6	5.2
900.	1040.	20.8	16.6	129.5	6.2
850.	1534.	17.9**	14.4**	140.6	6.4
800.	2051.	14.5	12.0	162.2	5.0
750.	2595.	11.2	8.3	156.9	5.5
700.	3168.	8.6**	4.6**	123.1	8.1
650.	3781.	8.0	-7.2	110.4	10.1
600.	4437.	4.9	-23.5	99.8	11.5
550.	5143.	2.1	-41.0	94.2	11.0
500.	5902.	-3.9**	-31.6**	81.3	9.4
450.	6724.	-9.3	-36.9	91.0	4.3
400.	7625.	-14.8	-25.5	176.1	4.7
350.	8622.	-21.6	-42.0	111.0	4.6
300.	9738.	-30.2	-56.6	53.9	6.8
250.	11008.	-40.6	-61.5	65.7	4.0
200.	12486.	-52.6	-57.3	55.2	9.2
150.	14277.	-67.6	-68.5	48.4	11.7
100.	16632.	-81.0	-84.4	15.8	6.6
50.	20620.	-65.9	-999.0	73.0	10.6
20.	26377.	-48.6	-999.0	99.5	23.3

(**) Es la información que se utiliza para el cálculo de los índices atmosféricos.

Los índices CAPE, CINH Y Td850 se obtienen directamente del sondeo mientras que KI, W, CT y TT fueron calculados con la misma información del sondeo utilizando las ecuaciones referidas por Concepción (2009):

$$KI = (T850 - T500) + Td850 - (T700 - Td700) \quad (3.1.2.1)$$

Siendo: T_{850} temperatura en grados centígrados a 850 mb.

T_{500} temperatura a 500 mb.

Td_{850} temperatura del punto de rocío a 850mb.

T_{700} temperatura a 700 mb.

Td_{700} temperatura del punto de rocío a 700 mb.

$$W = (T_{850} - Td_{850}) + (T_{700} - Td_{700}) + (T_{500} - Td_{500})$$

(3.1.2.2)

Siendo Td_{500} temperatura del punto de rocío a 500 mb.

$$CT = (Td_{850} - T_{500}) \quad \textbf{(3.1.2.3)}$$

$$TT = T_{850} + Td_{850} - 2 * T_{500} \quad \textbf{(3.1.2.4)}$$

Donde mb es la unidad de la presión atmosférica en milibar.

3.1.3 Recolección de datos.

Ya con los datos de la estación de Maiquetía proporcionados por el Instituto Nacional de Meteorología e hidrología (INAMEH) y con los índices de inestabilidad atmosférica extraídas y calculadas a partir de los sondeos de la página antes mencionada entonces se procedió a guardarlos y a organizarlos en formato Excel 2007 (anexo 1).

3.2 Selección de Variables.

La selección de un conjunto reducido de variables explicativas o predictoras es un problema complicado, ya que si el número de variables es demasiado pequeño

entonces el modelo es poco potente y si el número de variables es muy grande entonces el modelo tendría poca utilidad. Por otro lado es necesario para evitar la multicolinealidad entre las variables predictoras, porque la misma afectaría las estimaciones, proporcionando valores para los coeficientes de regresión no idóneos para el modelo, donde además si la correlación entre dos o más variables es fuerte los resultados de la regresión serán ambiguos según Canavos G. (1988).

Como una alternativa a este tipo de problemas se encuentran diferentes metodologías para la selección de variables. Para Draper (1998) las más destacadas son:

3.2.1 Método Paso Adelante (Forward).

Es un método de selección de variables donde se comienza por un modelo que no contiene ninguna variable explicativa o predictora y se añade como primera de ellas a la que tenga un mayor coeficiente de correlación con la variable dependiente. En los pasos sucesivos se va incorporando al modelo aquella variable que presenta un mayor coeficiente de correlación parcial con la variable dependiente dadas las independientes ya incluidas en el modelo. El procedimiento se detiene cuando el incremento en el coeficiente de determinación debido a la inclusión de una nueva variable explicativa en el modelo ya no es importante, siendo esto último generalmente determinado por el investigador.

3.2.2 Método Paso Atrás (Backward).

Así como el forward es un método de selección de variables en donde el procedimiento es similar al anterior solo que en sentido contrario, es decir, se comienza por considerar incluidas en el modelo teórico a todas las variables disponibles y se van eliminando del modelo de una en una, según su capacidad explicativa. En concreto, la primera variable que se elimina es aquella que presenta un menor coeficiente de

correlación parcial con la variable dependiente y así sucesivamente hasta llegar a una situación en la que la eliminación de una variable más suponga un descenso demasiado acusado en el coeficiente de determinación.

3.2.3 Método Paso a Paso (Stepwise).

Es uno de los más empleados y consiste en una combinación de los dos anteriores. En el primer paso se procede como en el método forward pero a diferencia de éste en el que cuando una variable entra en el modelo ya no vuelve a salir, en el procedimiento stepwise es posible que la inclusión de una nueva variable haga que otra que ya estaba en el modelo resulte redundante y sea “expulsada” de él. Esta prueba se realiza de forma reiterativa hasta que todas las variables explicativas hayan sido probadas dentro del modelo. Para la presente investigación es utilizado para seleccionar las variables mediante el paquete estadístico minitab.

3.3 Método de Máxima Verosimilitud.

Mayorga J. (2004), define el método de máxima verosimilitud como una sucesión de variables aleatorias idénticamente distribuidas pero no necesariamente independientes, la función conjunta de densidad de $X_1, X_2, X_3, \dots, X_n$ se conoce con el nombre de función de verosimilitud. Donde su expresión matemática la plantea de la siguiente manera:

Sea $X_1, X_2, X_3, \dots, X_n$ una muestra aleatoria de una población con función de densidad $f_x(X, \theta)$. $\theta \in \Phi$, la función de verosimilitud de la muestra corresponde a:

$$L(\theta; X_1, X_2, X_3, \dots, X_n) = \prod_{i=1}^n f_x(X_i, \theta) \quad (3.3.1)$$

Según Salazar A. (2011), las ecuaciones de verosimilitud no son lineales en los parámetros. Para resolverlas, se propone el uso de un método iterativo como es el de “Newton- Raphson”, debido a la complejidad de los cálculos, que puede usarse para un aproximado del máximo de una función.

3.4 Prueba de ajuste de Pearson:

El test de χ^2 es una prueba de bondad de ajuste y como tal intenta determinar el ajuste entre las distribuciones de frecuencias observadas y esperadas de los diferentes niveles categoriales de una variable. De allí, se puede afirmar que esta prueba se aplica cuando la variable objeto de estudio está medida en una escala de tipo categórica o nominal. Dada esta particularidad Marques M. (2001) compara la prueba de Pearson con otras, como Kolmogorov – Smirnov, la de Anderson – Darling, etc., destacando que la primera se utiliza tanto para distribuciones continuas como para discretas (dicotómicas), mientras que las otras solo se pueden aplicar a las distribuciones continuas.

Esta prueba puede utilizarse incluso con datos medibles en una escala nominal. La hipótesis nula de la prueba Chi-cuadrado postula una distribución de probabilidad totalmente especificada como el modelo matemático de la población que ha generado la muestra.

Para realizar este contraste se disponen los datos en una tabla de frecuencias. Para cada valor o intervalo de valores se indica la frecuencia absoluta observada o empírica (O_i). A continuación, y suponiendo que la hipótesis nula es cierta, se calculan para cada valor o intervalo de valores la frecuencia absoluta que cabría esperar o frecuencia esperada ($E_i = n \cdot P_i$, donde n es el tamaño de la muestra y P_i es la probabilidad del i -ésimo valor o intervalo de valores según la hipótesis nula). El estadístico de prueba se basa en las diferencias entre la O_i y E_i y se define como:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (3.4.1)$$

Este estadístico tiene una distribución Chi-cuadrado con $n-(k+1)$ grados de libertad, donde k es el número de parámetros estimados.

Si existe concordancia perfecta entre las frecuencias observadas y las esperadas el estadístico tomará un valor igual a 0; por el contrario, si existe una gran discrepancia entre estas frecuencias el estadístico tomará un valor grande y, en consecuencia, se rechazará la hipótesis nula. Así pues, la región crítica estará situada en el extremo superior de la distribución Chi-cuadrado con $k-(r+1)$ grados de libertad.

3.5 Prueba de Ajuste Hosmer – Lemeshow.

Una prueba de la bondad del ajuste de los modelos con respuestas binarias es la que Hosmer y Lemeshow desarrollaron en 1980. Esta prueba para Méndez I. y Moreno H. (2005) consiste en agrupar las observaciones de acuerdo al orden de valores de las probabilidades estimadas P_i , es decir, estas se ordenan y se dividen en cuartiles para formar 4 grupos ó en deciles para formar diez, esta ultima la mayormente utilizada.

Hosmer, D., Lemeshow, S. y Sturdivant R. (2013) mencionan que dicho contraste consiste en realizar comparaciones entre el valor estimado y el observado por grupos. Para ello, las observaciones se dividen en J grupos (generalmente 10) aproximadamente iguales, dividiendo el recorrido de la probabilidad en deciles de riesgo (esto es, probabilidad de ocurrencia del fenómeno < 0.1 , < 0.2 , y así hasta < 1).

Cada uno de los grupos contiene n_j observaciones, y en cada uno de los J grupos se define:

- Y_j como la suma de los valores 1 en cada uno de los grupos ($Y_j = \sum Y_i$)
- \bar{O}_j como la media de los valores predichos en cada grupo.

A partir de esta información se puede construir un cuadro de contingencia (porcentajes de aciertos) a través de la que se compara tanto la distribución de ocurrencia, como la de no ocurrencia prevista por la ecuación y los valores realmente observados. El contraste se realiza comparando las frecuencias observadas y esperadas a través del cálculo del estadístico

$$HL = \sum_{j=1}^J \frac{(Y_j - n_j \bar{0}_j)^2}{n_j \bar{0}_j (1 - \bar{0}_j)} \quad (3.5.1)$$

Hosmer, D., Lemeshow, S. y Sturdivant R. (2013), demuestran que cuando el modelo es correcto el estadístico *HL* sigue una distribución chi-cuadrado con *J-2* grados de libertad, por lo que valores inferiores del estadístico calculado respecto al teórico indicarán un buen ajuste del modelo.

El uso correcto de este contraste requiere un tamaño de muestra adecuado para asegurar que cada grupo cuenta al menos con cinco observaciones. Además, el estadístico chi cuadrado es sensible al tamaño muestral, permitiendo que esta medida encuentre diferencias estadísticamente muy pequeñas cuando el tamaño muestral crece.

3.6 Porcentajes de Aciertos Estimados en el modelo.

Salas, M. (1996), hace mención de tablas de clasificación muy utilizadas en donde se pueden estudiar ciertas características que forman parte del ajuste de un modelo de respuestas que a menudo suelen ser dicotómicas o categóricas. Determinar la bondad de un modelo Logit o Probit es predecir con el modelo los valores de la variable Y_i de tal manera que $Y_i = 1$ si $M_i > c$ ó $Y_i = 0$ si $M_i < c$, considerándose M_i como el ajuste obtenido. Por lo general el valor que se le asigna a c para determinar si el valor a la predicción es igual a 1 o a 0 es de 0,5 puesto que parece lógico que la predicción sea 1 cuando el modelo dice que es más probable obtener un 1 que un 0, considerándose c como el umbral para realizar la recodificación de los ajustes obtenidos.

Sin embargo, la elección de un umbral igual a 0,5 no siempre es la mejor alternativa por lo que el valor que debe tomar depende de la distribución de datos en la muestra. Para Medina (2003), cualquier tipo de regla predictiva similar se cometerán dos errores: habrá ceros que se clasifiquen incorrectamente como unos y unos que se clasifiquen incorrectamente como ceros. Si se reduce el umbral por debajo de 0,5 aumentará el número de veces que se clasifican correctamente observaciones para las que $Y_i=1$, pero también aumentará el número de veces en que se clasifiquen observaciones como unos para las que $Y_i=0$. Cambiando el valor del umbral se reducirá siempre la probabilidad de ocurrencia de un evento y se aumentará la probabilidad de no ocurrencia del evento.

Una vez seleccionado el nivel del umbral, y dado que los valores reales de Y_i son conocidos, basta con calcular el porcentaje de aciertos para decir si la bondad del ajuste es elevada o no. A partir de este recuento se puede construir el siguiente cuadro de clasificación:

CUADRO 3.6. Cuadro de porcentajes.

CUADRO DE ACIERTOS		VALOR PRONOSTICADO	
		$M_i < c$ (0)	$M_i > c$ (1)
VALOR OBSERVADO	0	C ₁₁	C ₁₂
	1	C ₂₁	C ₂₂

Donde C₁₁; C₂₂ son todos aquellos valores pronosticados que coinciden con los valores observados (tanto en la ocurrencia o no del evento) mientras que C₁₂; C₂₁ son todos aquellos valores pronosticados que no coinciden con los valores observados, es decir, es donde falla el modelo. No obstante de aquí se pueden desprender características que puedan dar una idea del comportamiento del modelo mediante las siguientes tasas:

3.6.1) Tasas de Aciertos.

Es la relación que existe entre las predicciones correctas y el total de predicciones.

$$\frac{C_{11}+C_{22}}{C_{11}+C_{12}+C_{21}+C_{22}} \quad (3.6.1)$$

3.6.2 Tasas de Errores.

Es la relación que existe entre las predicciones incorrectas con el total de predicciones.

$$\frac{C_{12}+C_{21}}{C_{11}+C_{12}+C_{21}+C_{22}} \quad (3.6.2)$$

3.6.3 Especificidad.

Proporción entre la frecuencia de valores 0 correctos y el total de valores 0 observados.

$$\frac{C_{11}}{C_{11}+C_{12}} \quad (3.6.3)$$

3.6.4) Sensibilidad.

Es el cociente entre los valores 1 correctos y el total de valores 1 observados.

$$\frac{C_{22}}{C_{21}+C_{22}} \quad (3.6.4)$$

3.6.5) Tasa de Falsos Ceros.

Proporción entre la frecuencia de valores 0 incorrectos y el total de valores 0 observados.

$$\frac{C_{12}}{C_{11}+C_{12}} \quad (3.6.5)$$

3.6.6) Tasa de Falsos Unos.

Es el cociente entre los valores 1 incorrectos y el total de valores 1 observados.

$$\frac{C_{21}}{C_{21}+C_{22}} \quad (3.6.6)$$

IV. RESULTADO Y DISCUSION

4.1 Selección de variables.

Para saber si existen indicios de colínealidad entre las variables predictoras y que importancia tienen estas con la variable respuesta, se calcula la matriz de correlación usando SPSS:

CUADRO 4.1. Matriz de correlaciones.

		P01	Td850	CAPE	CINH	KI	W	CT	TT
P01	Correl. Pearson	1	,116**	,107**	,006	,331**	-,314**	,233**	,211**
	Sig. (bilateral)		,005	,009	,878	,000	,000	,000	,000
	N	594	594	594	594	594	594	594	594
Td850	Correl. Pearson	,116**	1	,138**	,029	,372**	-,287**	,538**	,399**
	Sig. (bilateral)	,005		,001	,475	,000	,000	,000	,000
	N	594	594	594	594	594	594	594	594
CAPE	Correl. Pearson	,107**	,138**	1	,137**	,270**	-,244**	,282**	,344**
	Sig. (bilateral)	,009	,001		,001	,000	,000	,000	,000
	N	594	594	594	594	594	594	594	594
CINH	Correl. Pearson	,006	,029	,137**	1	-,116**	,077	,020	-,167**
	Sig. (bilateral)	,878	,475	,001		,005	,061	,627	,000
	N	594	594	594	594	594	594	594	594
KI	Correl. Pearson	,331**	,372**	,270**	-,116**	1	-,820**	,720**	,702**
	Sig. (bilateral)	,000	,000	,000	,005		,000	,000	,000
	N	594	594	594	594	594	594	594	594
W	Correl. Pearson	-,314**	-,287**	-,244**	,077	-,820**	1	-,655**	-,657**
	Sig. (bilateral)	,000	,000	,000	,061	,000		,000	,000
	N	594	594	594	594	594	594	594	594
CT	Correl. Pearson	,233**	,538**	,282**	,020	,720**	-,655**	1	,917**
	Sig. (bilateral)	,000	,000	,000	,627	,000	,000		,000
	N	594	594	594	594	594	594	594	594
TT	Correl. Pearson	,211**	,399**	,344**	-,167**	,702**	-,657**	,917**	1
	Sig. (bilateral)	,000	,000	,000	,000	,000	,000	,000	
	N	594	594	594	594	594	594	594	594

** . La correlación es significativa al nivel 0,01 (bilateral).

Se puede constatar correlaciones significativas entre las variables regresoras lo que sería conveniente utilizar algún método de selección de variables, en este caso se utiliza el método paso a paso.

Hoy en día se cuenta con paquetes estadísticos que en su programación tienen integrados estos métodos, tales como SPSS, SAS y MINITAB. Este último utilizado en esta parte de la investigación. En los cuadros 4.1.1 y 4.1.2 se puede observar todas las variables predictoras que son aceptadas e incluidas dentro de la expresión matemática para cada una de las muestras, y que son marcadas con la letra x:

CUADRO 4.1.1. Resumen de variables predictoras usando el método paso a paso para cada muestra aleatorizada.

M-A	Constante		Td850		CINH		KI		W		TT	
	L	P	L	P	L	P	L	P	L	P	L	P
M-1	X	X	x	X	x	x			x	x	x	x
M-2	X	X					x	x	x	x		
M-3	X	X					x	x	x	x		
M-4	X	X					x	x				
M-5	X	X			x	x	x	x	x	x		x
M-6	X	X					x	x				
M-7	X	X					x	x	x	x		
M-8	X	X			x	x	x	x		x		
M-9	X	X					x	x	x	x		

CUADRO 4.1.2. Resumen de variables predictoras usando el método paso a paso para cada muestra secuencial o no aleatorizada.

M-S	Constante		Td850		CINH		KI		W		CT		TT	
	L	P	L	P	L	P	L	P	L	P	L	P	L	P
M-1	X	X							X	X	X	X		
M-2	X	X			X	X			X	X	X	X	X	X
M-3	X	X	X	X	X	X			X	X				
M-4	X	X			X	X			X	X				
M-5	X	X						X	X	X		X		X
M-6	X	X					X	X	X	X				X
M-7	X	X					X	X	X	X				X
M-8	X	X					X	X	X	X			X	X
M-9	X	X					X	X	X	X				

donde:

M-S: muestra secuencial o no aleatoria.

M-A: muestra aleatoria.

M-1...M-9: muestra numero 1 hasta la 9.

L: modelo Logit.

P: modelo Probit.

4.2 Factor de la Varianza Inflada (FIVó VIF).

En el capítulo II de este trabajo se mencionó que una técnica o método para detectar multicolinealidad es utilizando el estadístico VIF donde se toma en cuenta la relación lineal que tiene una variable predictora con respecto a las demás, exceptuando la variable dependiente. Esta se calcula mediante la ecuación 2.3.3:

$$VIF(\beta_i) = \frac{1}{1-R_i^2} ; \quad \text{donde } i = 1, \dots, 7 \text{ para este caso.}$$

Cuanto más se acerque R_i^2 a la unidad, es decir, cuanto mayor sea la colinealidad de la variable X_i con el resto, mayor es el valor de $VIF(\beta_i)$ y mayor es la varianza del coeficiente estimado y, tal como se ha dicho antes, si $VIF > 10$ se diagnostica multicolinealidad. Minitab realiza el cálculo de manera simultánea al correr cada

modelo (Logit y Probit). En los siguientes cuadros se resumen los valores para VIF tanto para el muestreo aleatorizado como el no aleatorizado:

CUADRO 4.2.1. Valores de VIF para cada muestra aleatorizada.

M-A	Td850		CINH		KI		W		TT	
	L	P	L	P	L	P	L	P	L	P
M-1	2,04	1,95	1,22	1,23			1,9	1,89	3,01	2,91
M-2					2,36	2,69	2,36	2,69		
M-3					2,18	2,38	2,18	2,38		
M-4					1	1				
M-5			1,02	3,78	2,78	3,05	2,75	6,38		6,31
M-6					1	1				
M-7					2,47	2,63	2,47	2,63		
M-8			1,02	1,02	1,02	2,74		2,7		
M-9					2,49	2,7	2,49	2,7		

CUADRO 4.2.2. Valores de VIF para cada muestra secuencial o no aleatorizada.

M-S	Td850		CINH		KI		W		CT		TT	
	L	P	L	P	L	P	L	P	L	P	L	P
M-1							2,49	2,32	2,49	2,32		
M-2			1,52	1,46			1,97	1,88	9,31	8,79	7,79	7,52
M-3	1,51	1,51	1,12	1,11			1,38	1,38				
M-4			1	1,01			1	1,01				
M-5						4,06	1	2,81		6,13		5,38
M-6					2,56	3,59	2,56	2,82				2,05
M-7					2,48	3,15	2,48	2,73				1,94
M-8					2,93	3,21	2,56	2,77			1,91	2
M-9					2,49	2,7	2,49	2,7				

L: Logit P: Probit.

En ambos cuadros (4.2.1 y 4.2.2) muestran valores VIF por debajo de 10 lo que indica que no existe multicolinealidad, aunque el valor VIF de la muestra 2 de la no aleatorizada se puede observar que está cercano al límite en la variable CT, sin embargo. no altera significativamente la estimación de los parámetros.

4.3 Estimación de los Parámetros.

Al usar Minitab, el método de estimación de máxima verosimilitud se corre automáticamente y de forma simultánea con la selección de variables. En los siguientes cuadros se pueden verificar un resumen de los parámetros para cada variable y cada muestra para su posterior estudio:

CUADRO 4.3.1. Valores de β_i para cada muestra aleatorizada.

	L	P	L	P	L	P	L	P	L	P	L	P
	β_0	β_0	β_1	β_1	β_3	β_3	β_4	β_4	β_5	β_5	β_7	β_7
M-1	19,94	12,3	0,497	0,302	-0,023	-0,0146			-0,101	-0,0636	-0,58	-0,359
M-2	-2,07	-1,17					0,0961	0,056	-0,0452	-0,0278		
M-3	-1,87	-1,15					0,0799	0,0489	-0,0398	-0,0238		
M-4	-4,475	-2,52					0,1381	0,0779				
M-5	-0,58	3,7			0,00971	0,024	0,0543	-0,031	-0,0455	0,218		-0,187*
M-6	-5,037	-2,99					0,1595	0,0949				
M-7	-2,34	1,359					0,0935	0,0553	-0,0261	-0,0165		
M-8	-3,914	-1,54			0,00685	0,00416	0,1315	0,0604		-0,011*		
M-9	-2,48	-1,44					0,0947	0,0556	-0,0241	-0,0147		

CUADRO 4.3.2. Valores de β_i para cada muestra secuencial o no aleatorizada.

	L	P	L	P	L	P	L	P	L	P	L	P	L	P
	β_0	β_0	β_1	β_1	β_3	β_3	β_4	β_4	β_5	β_5	β_6	β_6	β_7	β_7
M-1	13,04	7,59							-0,142	-0,082	-0,533	-0,309		
M-2	2,2	1,24			0,025	0,015			-0,129	-0,078	-1,02	-0,613	0,48	0,29
M-3	5,71	3,53	-0,24	-0,15	0,021	0,013			-0,092	-0,056				
M-4	1,64	1,02			0,009	0,006			-0,060	-0,037				
M-5	1,24	3,78						0,037*	-0,061	-0,028		0,131*		-0,16*
M-6	-1,56	1,65					0,0742	0,062	-0,038	-0,024				-0,07*
M-7	-1,13	1,69					0,0637	0,051	-0,042	-0,028				-0,06*
M-8	2,18	1,4					0,1018	0,061	-0,035	-0,021			-0,109	-0,067
M-9	-2,48	-1,44					0,0947	0,056	-0,024	-0,015				

L: Logit P: Probit.

En ambos cuadros se pueden observar cada uno de los valores de los parámetros de las variables seleccionadas por el método paso a paso y en donde se verifica los valores que pueden ser significativos para la ocurrencia de las precipitaciones. No obstante, los valores con (*) señala los parámetros que aparecen en el modelo Probit pero que no están en el modelo Logit, tanto en las muestras aleatorizadas como en las no aleatorizadas. Se puede observar que de 17 muestras diferentes en 5 el Probit añade una o tres variables regresoras más que el Logit, esto revela que existe una pequeña tendencia por parte del modelo Probit incluir al menos una variable más en su expresión matemática que el modelo Logit y en un caso hasta tres (muestra 5 no aleatorizada) lo que determinaría mayor flexibilidad del Probit en la inclusión de predictores sobre el Logit.

4.4 Pruebas de Ajuste.

4.4.1 Prueba Hosmer – Lemeshow y Prueba de Pearson.

Para la verificación del ajuste del modelo con los datos observados se consideraron las pruebas de Hosmer – Lemeshow y Pearson. Esto para revisar el ajuste desde diferentes perspectivas en cada una de las muestras aleatorizadas y no aleatorizadas, utilizando Minitab para luego resumir los resultados en los cuadros 4.4.1.1 y 4.4.1.2.

CUADRO 4.4.1.1. Pruebas de ajustes para cada modelo en muestras aleatorizadas.

MA	LOGIT						PROBIT					
	G L	H.L.	P-VA	G L	PEAR	P-VA	G L	H.L.	P-VA	G L	PEAR	P-VA
M_1	8	9,73	0,284	61	61,72	0,450	8	9,86	0,275	61	61,46	0,460
M_2	8	9,41	0,284	129	131,27	0,428	8	9,57	0,296	129	131,17	0,430
M_3	8	14,7	0,065	195	196,15	0,463	8	14,79	0,063	195	195,89	0,469
M_4	8	6,09	0,637	262	288,21	0,128	8	6,8	0,558	262	287,78	0,131
M_5	8	17,12	0,029*	326	340,08	0,284	8	11,87	0,157	326	344,50	0,219
M_6	8	6,04	0,643	394	398,84	0,423	8	6,22	0,622	394	400,38	0,401
M_7	8	7,61	0,472	459	461,38	0,460	8	7,94	0,440	459	461,04	0,464
M_8	8	11,48	0,176	525	537,71	0,341	8	8,4	0,395	525	536,98	0,338
M_9	8	17,13	0,029*	591	606,19	0,324	8	16,85	0,03*	591	606,98	0,316

CUADRO 4.4.1.2. Pruebas de ajustes para cada modelo en muestras no aleatorizadas.

MS	LOGIT						PROBIT					
	G L	H.L.	P-VA	G L	PEAR	P-VA	G L	H.L.	P-VA	G L	PEAR	P-VA
M_1	8	6,46	0,596	63	65,31	0,396	8	6,41	0,602	63	64,64	0,419
M_2	8	4,58	0,801	127	129,18	0,429	8	4,74	0,785	127	128,69	0,441
M_3	8	9,25	0,322	194	191,72	0,533	8	9,34	0,314	194	191,79	0,531
M_4	8	16,89	0,031*	261	262,07	0,470	8	18,03	0,02*	261	262,06	0,470
M_5	8	20,97	0,007*	328	335,46	0,376	8	21,6	0,01*	328	331,27	0,393
M_6	8	16,54	0,035*	393	402,23	0,363	8	12,86	0,117	393	400,72	0,370
M_7	8	8,88	0,353	459	466,74	0,391	8	9,27	0,320	459	466,98	0,376
M_8	8	15,84	0,045*	524	536,54	0,343	8	17,06	0,03*	524	536,22	0,346
M_9	8	17,13	0,029*	591	606,19	0,324	8	16,85	0,03*	591	606,98	0,316

Donde GL es grado de libertad; H.L. es la prueba de Hosmer – Lemeshow; Pear es la prueba de Pearson; P-VA es el P-Valor; MA son las muestras aleatorizadas y MS son las muestras no aleatorizadas.

Los valores con (*) indican que el modelo no tiene buen ajuste ya que se rechaza la hipótesis nula con una significancia de 0,05, según la prueba H.L. Esto dá pie para el análisis comparativo, ya que se puede verificar que en dos de las siete muestras en rojo utilizando el modelo Logit (con muestras aleatorizadas y no aleatorizadas), el modelo Probit se separa arrojando que en dos de ellas se determina un buen ajuste, según la prueba de Hosmer-Lemeshow (ver muestra 5 de la aleatorizada y muestra 6 de la no aleatorizada).

Se puede comprobar que bajo el criterio de la prueba de Pearson no se descartó ninguno de los modelos en sus diferentes expresiones. En tal sentido, con la información de la prueba de bondad de ajuste Hosmer-Lemeshow estaría constituyendo un segundo diagnostico desde otra perspectiva, que ayuda establecer las condiciones favorables para cada uno de los modelos (Logit y Probit).

4.4.2 Porcentaje de Aciertos o Tablas de clasificación.

Para realizar el estudio respectivo de los aciertos y desaciertos se utiliza los cuadros de clasificación o de contingencia únicamente para la muestra total y para las que se observan ciertas diferencias (muestra 5 aleatorizada y 6 no aleatorizada).

CUADRO 4.4.2.1. Porcentajes de aciertos de los modelos Logit y Probit para toda la muestra y bajo el método de selección de variables Paso a Paso.

PROBIT		PRONOSTICADO		
OBSERVADO	P01	0	1	%
	0	185	123	60,06
	1	79	207	72,38
				66,22

LOGIT		PRONOSTICADO		
OBSERVADO	P01	0	1	%
	0	185	123	60,06
	1	80	206	72,03
				66,05

El 60,06% corresponde a la especificidad y el 72,38% a la sensibilidad del Probit, mientras que el 66,22% es el promedio entre ambos porcentajes. En el mismo orden para el Logit, 60,06% en especificidad, 72,03% en sensibilidad y 66,05% como porcentaje general. De igual forma, haciendo uso de las fórmulas correspondientes especificadas en el capítulo anterior para cada tasa, se tienen los siguientes resultados:

Probit:

Tasa de aciertos: 65,99%

Tasa de errores: 34,01%

Tasa de falsos ceros: 39,93%

Tasa de falsos unos: 27,62%

Logit:

Tasa de aciertos: 65,82%

Tasa de errores: 34,17%

Tasa de falsos ceros: 39,93%

Tasa de falsos unos: 27,97%

Esta clase de evaluación es comúnmente realizada para obtener diagnósticos de aciertos para modelos de respuestas binarias con un umbral de 0,5. Aquí se puede apreciar que bajo esta condición no existe diferencia en cuanto a la especificidad entre ambos modelos, existiendo ligeramente una mayor sensibilidad en el Probit que en el

Logit. Se muestra una pequeña diferencia entre los modelos a favor del Probit en cuanto al pronóstico de la ocurrencia o no de precipitaciones.

CUADRO 4.4.2.2. Porcentajes de aciertos de los modelos Logit y Probit para la muestra 5 aleatorizada y bajo el método de selección de variables paso a paso.

PROBIT		PRONOSTICADO		
OBSERVADO	P01	0	1	%
	0	117	58	66,86
	1	51	104	67,10
				66,98
LOGIT		PRONOSTICADO		
OBSERVADO	P01	0	1	%
	0	118	57	67,43
	1	54	101	65,16
				66,29

Probit:

Tasa de aciertos: 66,96%

Tasa de errores: 33,03%

Tasa de falsos ceros: 33,14%

Tasa de falsos unos: 32,90%

Logit:

Tasa de aciertos: 66,36%

Tasa de errores: 33,63%

Tasa de falsos ceros: 32,57%

Tasa de falsos unos: 34,83%

En esta parte del diagnóstico aunque, el modelo Logit se muestra ligeramente con mayor especificidad que el Probit en los ceros, este sigue manteniendo más aciertos con la ocurrencia de lluvias y una condición favorable para el pronóstico de forma general.

CUADRO 4.4.2.3. Porcentajes de aciertos de los modelos Logit y Probit para la muestra 6 no aleatorizada y bajo el método de selección de variables paso a paso.

PROBIT		PRONOSTICADO		
OBSERVADO	P01	0	1	%
	0	101	91	52,60
	1	50	154	75,49
				64,05

LOGIT		PRONOSTICADO		
OBSERVADO	P01	0	1	%
	0	103	89	53,65
	1	51	153	75,00
				64,32

Probit:

Tasa de aciertos: 64,39%

Tasa de errores: 35,60%

Tasa de falsos ceros: 47,39%

Tasa de falsos unos: 24,50%

Logit:

Tasa de aciertos: 64,64%

Tasa de errores: 35,35%

Tasa de falsos ceros: 46,35%

Tasa de falsos unos: 25%

Para la muestra 6 no aleatorizada se observa mayor especificidad en el Logit y como en todas los cuadros anteriores mayor sensibilidad en el Probit, aunque de manera general en este caso, el modelo Logit muestra un diagnostico levemente favorable sobre el otro modelo. Vale la pena recordar que se está trabajando con el umbral de 0,5 y que al realizar ínfimas variaciones del mismo podría alterar de forma significativa dicha tendencia, todo depende de la concentración de los unos (ocurrencia de precipitación). Aunado a esto en el cuadro 4.4.1.2 se verifica que según el estadístico de prueba HL

con una significancia de 0,05 señala que el modelo Logit no muestra un buen ajuste al conjunto de datos observados.

4.5 Estudio de las Variables Latentes.

Cada modelo posee una variable que, aunque no forma parte de las variables observadas directamente, proveen cierta información que puede ser de interés para el investigador, estas son los Odds en el modelo Logit que muestra propensión de las probabilidades y la variable latente en el modelo Probit que representa la diferencia en términos de utilidad (capítulo II). A los efectos de este trabajo se procede a calcular ambas variables y a verificar si existe algún tipo de relación entre ellas. Se calculó con los parámetros estimados para las variables seleccionadas por el método paso a paso (ODDS y Latente_P), (Anexos 1).

Se verifica si existe una correlación bivariada entre ambas, también se realiza un estudio descriptivo para cada una:

CUADRO 4.5.1. Correlaciones entre ODDS y Latente_P.

		Correlaciones	
		ODDS	Latente_P
ODDS	Correlación de Pearson	1	,908**
	Sig. (bilateral)		,000
	N	594	594
Latente_P	Correlación de Pearson	,908**	1
	Sig. (bilateral)	,000	
	N	594	594

** . La correlación es significativa al nivel 0,01 (bilateral).

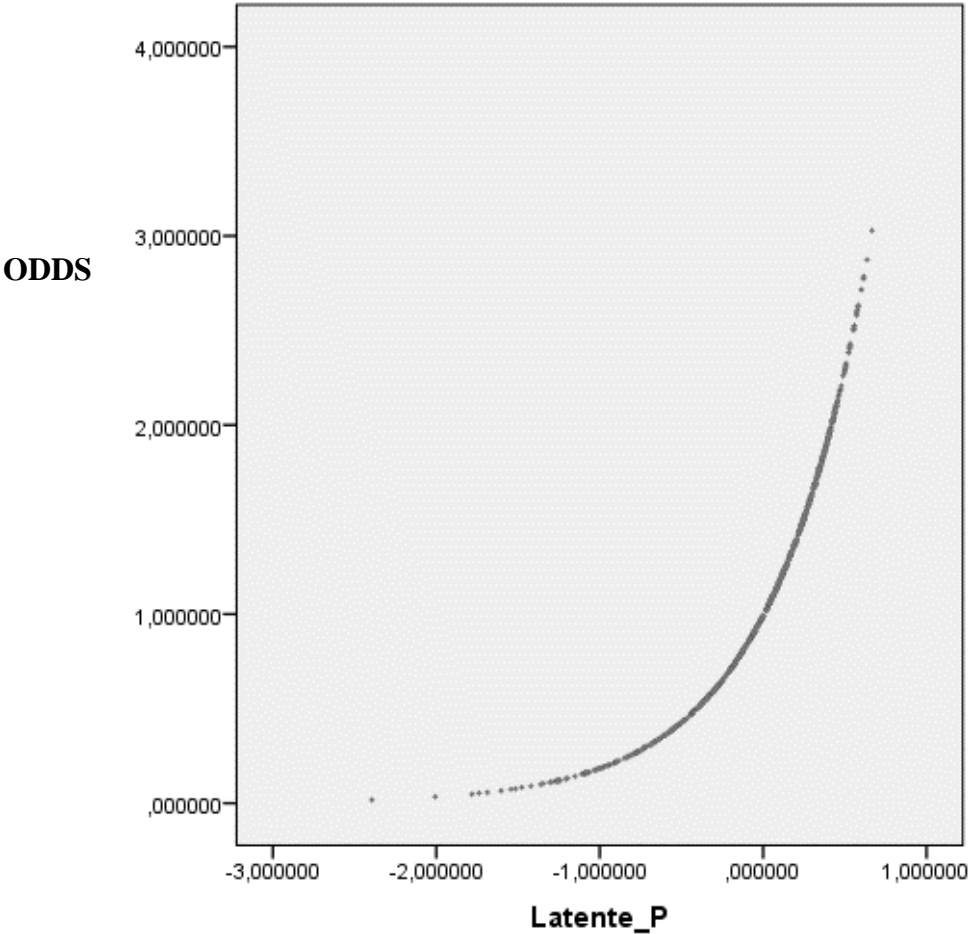
Aquí se puede ver que existe una correlación de Pearson bastante significativa entre ODDS y Latente_P.

CUADRO 4.5.2. Descriptivos de ODDS y Latente_P.

Descriptivos			Estadístico	Error típ.
ODDS	Media		1,13754688	,026981979
	Intervalo de confianza para la media al 95%	Límite inferior	1,08455502	
		Límite superior	1,19053875	
	Media recortada al 5%		1,11835138	
	Mediana		1,11306952	
	Varianza		,432	
	Desv. Típ.		,657607899	
	Mínimo		,017507	
	Máximo		3,028629	
	Rango		3,011122	
	Amplitud intercuartil		1,065418	
	Asimetría		,273	,100
	Curtosis		-,715	,200
	Latente_P	Media		-,06674055
Intervalo de confianza para la media al 95%		Límite inferior	-,10640870	
		Límite superior	-,02707241	
Media recortada al 5%			-,02840929	
Mediana			,07124000	
Varianza			,242	
Desv. Típ.			,492265841	
Mínimo			-2,393176	
Máximo			,666272	
Rango			3,059448	
Amplitud intercuartil			,615916	
Asimetría			-1,247	,100
Curtosis			1,546	,200

Como se puede observar, no existen valores negativos en los ODDS, no así en la Latente_P, aunque los rangos son aproximados en los demás estadísticos conservan cierto margen de diferencia.

Ya obtenidos los valores de las variables latentes de cada modelo para cada una de las observaciones y tomando en cuenta la correlación que existe entre ambas se procede a graficar los pares ordenados (ODDS, Latente_P) para estudiar su forma y el tipo de relación que existe entre ellas:



En esta grafica se puede apreciar que existe una relación exponencial entre las variables, en donde se observa que la mayor probabilidad de que ocurra una precipitación se encuentra en $\text{Latente_P} > 0$ y $\text{ODDS} > 1$.

Usando el mismo paquete estadístico SPSS se puede realizar una estimación curvilínea tratando de ajustar los datos a un modelo exponencial definido para este caso como:

$$Y = \beta_0 e^{\beta_1 X} \text{ ó } \ln(Y) = \ln(\beta_0) + \beta_1 X \quad (4.5.1)$$

Esto debido a la relación observada en la gráfica anterior. Dicha estimación revela las siguientes características:

CUADRO 4.5.3. Ajuste del modelo.

Resumen del modelo			
R	R cuadrado	R cuadrado corregida	Error típico de la estimación
1,000	1,000	1,000	,004

La variable independiente es Latente_P.

CUADRO 4.5.4. Anavar.

ANOVA					
	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	406,697	1	406,697	20253771,915	,000
Residual	,012	592	,000		
Total	406,709	593			

La variable independiente es Latente_P.

CUADRO 4.5.5. Coeficientes.

Coeficientes					
	Coeficientes no estandarizados		Coeficientes estandarizados	T	Sig.
	B	Error típico	Beta		
Latente_P	1,682	,000	1,000	4500,419	,000
(Constante)	,990	,000		5389,504	,000

La variable dependiente es $\ln(\text{ODDS})$.

De los cuadros 4.5.3, 4.5.4 y 4.5.5 se puede verificar que los Odds (ODDS) se pueden estimar casi perfectamente mediante la variable latente del modelo Probit, mostrando un excelente ajuste al utilizar el modelo:

$$ODDS = 0.99e^{1.682Latente_P} \quad (4.5.2)$$

Esta ecuación, particularmente para este caso, permitiría establecer un puente de comunicación a las características particulares de cada modelo, tendencias o diferencias en términos de aplicabilidad, que para este estudio es la ocurrencia de precipitaciones (P01) con solo utilizar uno de los dos modelos. Sería importante realizar estudios posteriores en donde se pueda verificar si la relación existente es constantemente exponencial y si se podría establecer una aproximación muy cercana en cualquier tipo de situación en donde la variable respuesta sea binaria, buscando una generalización de la misma con factores o parámetros fijos, haciendo de esta ecuación una herramienta útil para la investigación.

V. CONCLUSIONES.

Para el primer objetivo específico se señala que las características teóricas ó experimental de los modelos son muy parecidas, salvo a la información que puede proporcionar cada uno a través de sus variables latentes, las cuales, aun siendo no observables de manera directa ofrecen al investigador diferentes perspectivas. Tanto el Odds del Logit como la variable Latente del Probit son características muy particulares, pero el estudio realizado arroja que existe una relación exponencial casi perfecta entre ellas, mediante una ecuación matemática que las aproxima (ecuación 4.5.2). Dicha expresión podría en algunos casos resolver la disyuntiva en la elegibilidad de los modelos, ya que, desde la aplicación de uno también se pueden obtener aportes complementarios del otro. Esta se convertiría en una importante herramienta al servir de puente entre ambas alternativas.

Para el cumplimiento del objetivo dos se puede decir que los modelos de respuesta binaria estudiados en esta investigación (Logit y Probit) mostraron un comportamiento similar bajo las mismas condiciones de aplicabilidad hasta determinado tamaño muestral donde se evidencia que:

- a) Los valores de una de las pruebas de ajuste (Hosmer – Lemeshow) aplicadas a ambos modelos se distancian a partir de 198 observaciones, es decir, desde la muestra 4 en ambos tipos de muestreo aleatorizado y no aleatorizado, lo que hace pensar que, en cualquier situación donde se tenga una muestra con un tamaño similar o menor, la utilización de cualquiera de los dos modelos es indiferente.
- b) El modelo Probit se muestra más flexible en la inclusión de variables predictoras bajo el criterio del método de selección de variable paso a paso, lo que le da mayor estabilidad estadística en la estimación de P_0 y un buen

ajuste a los datos observados según la prueba de Hosmer – Lemeshow, ya que, en dos de las muestras donde el Probit fue aprobado bajo este criterio se incluyó una variable más que en el modelo Logit.

Para el tercer objetivo específico se puede decir que, de manera general ambos modelos se ajustan muy bien a los datos observados, pero antes vale la pena reseñar que, de manera empírica existen variables en la cual su relación física inciden significativamente en la ocurrencia de un evento, pero estadísticamente no, de allí la necesidad de encontrar un método de selección de variables que resuelva la multicolinealidad y mantenga una correlación tanto física como estadística importante entre la variable respuesta con las regresoras. Frente a esta situación el modelo probit podría ser una opción, ya que, al mostrar mayor flexibilidad que el modelo Logit en la inclusión de variables predictoras también muestra una capacidad predictora más estable, al menos bajo los criterios explicados en el párrafo anterior.

Por otro lado, en los porcentajes de aciertos se verifica en la comparación hecha en tres casos, que el modelo Probit muestra mayor sensibilidad que el Logit aunque con una mínima diferencia y una tasa de falsos 1 menor, lo que le da mayor precisión en el pronóstico de la ocurrencia de precipitaciones para este caso, mientras que el Logit mantiene una ligera diferencia a su favor en cuanto a especificidad en dos de los tres casos e igualando en el tercero, como también una tasa de falsos 0 menor, lo que le da mayor precisión en el pronóstico de la no ocurrencia de precipitaciones.

La escogencia de uno de los dos modelos se debe fundamentar principalmente en los objetivos del investigador sobre qué información además de P01 (probabilidad de que ocurra el evento) es de su interés, teniendo presente todo lo discutido aquí, en cuanto a flexibilidad, precisión, característica teórica de los modelos y capacidad predictora.

VI. RECOMENDACIONES.

1. Resultaría de gran importancia realizar un estudio comparativo de ambos modelos bajo condiciones de multicolinealidad, para así verificar la robustez de los mismos en alguna de sus características explicadas en este trabajo de investigación.

2. Realizar estudios acerca de las variables latentes explicadas aquí con el propósito de verificar y generalizar la relación exponencial entre los Odds y la variable latente del Probit representada en la ecuación 4.5.2, para así definir un puente de conexión entre ambos modelos.

3. Ampliar el estudio realizado en este trabajo de investigación pero utilizando un método de selección de variables distinto al utilizado aquí (método de selección de variables paso a paso) y tomando en cuenta otros modelos como el Gompit.

4. Conviene una incorporación más frecuente del modelo Probit en el área meteorológica, ya que se demostró que posee todas las condiciones para realizar pronósticos efectivos del tiempo.

5. Se recomienda realizar un trabajo de investigación comparativo entre los modelos aquí discutidos partiendo de un caso donde tanto la variable respuesta como las variables predictoras sean binarias y plantearse los mismos objetivos aquí desarrollados.

VII. REFERENCIAS BIBLIOGRAFICAS.

- AGRESTI, A. (2007). An Introduction to Categorical Data Analysis. University of Florida. Second Edition. Gainesville, EEUU. 372 P.
- AGRESTI, A. (2002). Categorical Data Analysis. University of Florida. Second Edition. Gainesville, EEUU. 710 P.
- ALDERETE, A. M. (2006), Art. Fundamentos del Análisis de Regresión Logística en la Investigación Psicológica. Universidad Nacional de Córdoba. ISSN 1667-4545. Córdoba – Argentina. 52-67 PP.
- AMEMIYA, T. (1994). Qualitative Response Models. Part IV. Studies in Econometric Theory. The Collected Essays of Takeshi Amemiya, Edward Elgar. USA.
- BLISS, C. (1934). The Method of Probits. Science. Science, Volume 79, Issue 2037. pp. 38-39.
- CANAVOS, G. (1988). Probabilidad y Estadística, Aplicaciones y Métodos. McGraw-Hill. Estado de México – México. P 520.
- COCHRAN, W. y FINNEY, D. (1979). Artículo: Chester Ittner Bliss. IBS Vol. 35, número 4. PP. 715 – 717.
- CONCEPCION, M. (2009). Aplicación de Índices de Estabilidad Atmosférica al Pronóstico Meteorológico de la Región Centro Norte Costera de Venezuela. Tesis de grado para optar al título de Ingeniero Hidrometeorologista. U.C.V. Caracas, Venezuela.

DRAPER, N. AND SMITH, H. (1998). Applied Regression Analysis. Third Edition.
John Wiley. New York - EEUU.

ENCHAUTEGUI, M. E. (2001). Módulo de estudio sobre Modelos Probit y Logit.
Departamento de Economía, Universidad de Puerto Rico, Recinto de Río
Piedras. 16 P.

FLORES, J. (2001). Press-Ridge en la Estimación de un Modelo de Regresión Lineal
Múltiple en Presencia de Multicolínealidad Severa. Trabajo de grado para optar
al título de Magister Scientiarum en Estadística. U.C.V. Maracay, Venezuela.

GUJARATI, D. PORTER, D. (2010). Econometría. McGrawHill, Quinta Edición.
México. 946 P.

HOSMER, D., LEMESHOW, S. y STURDIVANT, R. (2013). Applied Logistic
Regression, Third Edition. New Jersey - EEUU. 157-160.

LOPEZ, R. (2006). Cálculo de Probabilidades e inferencia estadística con Tópicos de
Econometría. Universidad Católica Andrés Bello. Cuarta Edición. Caracas
Venezuela. 601 P.

MARQUES, M. (2001). Estadística Básica: Un Enfoque No Paramétrico. UNAM.
México. Zaragoza. 31-32.

MAYORGA, J. (2004). Inferencia Estadística. Primera Edición. Universidad Nacional
De Colombia. Bogotá. 68-70.

MEDINA, E. (2003). Modelos de Elección Discreta. Universidad Autónoma de
Madrid. Madrid – España. 18 – 20.

- MENDENHALL, W., BEAVER, R. y BEAVER, B. (2006). Introducción a la Probabilidad y a la Estadística. Décima tercera Edición. Editorial CLE. México. 780 P.
- MENDEZ, I. y MORENO, H. (2005). Razones de Momios y Cocientes de Probabilidades, Conceptos y Estimación con Modelos Logísticos. Universidad Nacional Autónoma de México. Ediciones UNAM. P – 37.
- MONTGOMERY, D., PECK, E. y VINING, G. (2006). Introducción al Análisis de Regresión Lineal. Editorial Continental. Tercera edición. México. 567 P.
- MOORE, D. (2000). The Basic Practice of Statistics. Purdue University. Second Edition. Lafayette – USA. 845 P.
- PÉREZ, A. KIZYS, R. y MANZANEDO, L. (2002). Regresión Logística Binaria. Universidad Oberta de Catalunya, Barcelona - España 5–16 PP.
- SALAS, M. (1996). La Regresión Logística. Una Aplicación a la Demanda de Estudios Universitarios. Universidad de Granada. Granada – España. 213-215.
- SALAZAR A. (2011). Modelos de Respuestas Discretas en R y Aplicación con datos Reales. Universidad de Granada. Granada – España. 12 – 28 PP.
- SANCHEZ, O. (2012). Estructuras con variables latentes como alternativa para modelar la autoselección. Tesis para optar al título de Maestro en Ciencias con Especialidad en Probabilidad y Estadística. CIMAT. Guanajuato, Mexico. 80 P.
- WALPOLE, R., MYERS, R., MYERS, S. y YE, K. (2007). Probabilidad y Estadística para Ingeniería y Ciencias. Octava edición. Pearson educación. Mexico. 840 P.

VIII. BIBLIOGRAFIA ADICIONAL CONSULTADA.

- ALDRICH, J. AND NELSON, F. (1984). Linear Probability, Logit and Probit Models Vol. 45. SAGE Publications. California – EEUU. 95 P.
- ALVAREZ R. (1995). Estadística Multivariante y No paramétrica con SPSS Aplicación a las ciencias de la salud. Ediciones Díaz de Santos. España. Madrid. 305-306 PP.
- ARNAU, J. (1996). Métodos y Técnicas Avanzadas de Análisis de Datos en Ciencias del Comportamiento. Ediciones de la Universidad de Barcelona. Barcelona – España. 101 – 112.
- DEL BARRIO, T., LOPEZ, M. y SURINACH, J. (2002). Modelo de Regresión Lineal Múltiple. Universidad Abierta de Cataluña (UOC), editorial UOC. Barcelona – España. 21 – 23.
- DIAZ, J. Y BENITEZ, F. (1998). Introducción a los Métodos Numéricos para la Resolución de Ecuaciones. Cuarta edición. Universidad de Cádiz. 23-26.
- GARCÍA V., ALVARADO J. y JIMÉNEZ A. (2000). Art. La Predicción del Rendimiento Académico: Regresión Lineal Versus Regresión Logística. Universidad Complutense de Madrid. Madrid – España. Vol.12 PP. 248 -252.
- LIAO T. F. (1994). Interpreting Probability Models Logit, Probit and Other Generalized Linear Models, Issue 101. SAGE Publications. California – EEUU. 88 P.
- MENDENHALL, W., SCHEAFFER R. y WACKERLY, D. (2010). Estadística Matemática con Aplicaciones. Séptima Edición. Editorial CLE.

México. 911 P.

RODRIGUEZ, J y FIUZA, M. (2000). Artículo: La Regresión Logística: Una Herramienta Versátil. *Nefrología*, vol. XX, número 6. Canaria – España. 405 – 500.

ROJO, J. M. (2007). *Regresión Lineal Múltiple*. Instituto de Economía y Geografía. Madrid – España. P - 6.

ROJO, J. M. (2007). *Regresión con variable dependiente cualitativa*. Instituto de Economía y Geografía. Madrid – España. P - 6.

WILKS, D. S. (2006). *Statistical Methods in the Atmospheric Sciences, Volume 91 in the International Geophysics Series. Second Edition*. California – EEUU. 201-215 PP.

IX. ANEXOS.

ANEXO 1. Recolección de datos.

M-9	P01	Td850	CAPE	CINH	KI	W	CT	TT	ODDS	Latente_P
1	0	14,4	514,2	-39,01	32,2	35,2	18,3	40,1	0,76	-0,17
2	0	10,9	791,7	-12,62	26	49,8	15,8	38,7	0,3	-0,73
3	0	12	883,2	-28,75	16,4	55,5	17,3	40,2	0,1	-1,34
4	0	18,2	246,5	-59,37	7,2	65,7	13,4	35,5	0,03	-2,01
5	0	18,7	348,5	-20,22	12,1	66,1	14,1	37,1	0,05	-1,74
6	0	18,2	662,4	-37,7	19,7	51	17,6	40,9	0,16	-1,09
7	1	18,2	1141	-16,16	34,3	16,5	17,8	40,7	1,45	0,23
8	1	13,6	844,3	-47,17	32	17,6	18	41,2	1,13	0,08
9	1	15,5	1087	-15,87	37,7	14,7	20,7	44,7	2,09	0,44
10	0	13,4	709,6	-20,54	22	31,6	18	41	0,31	-0,68
11	1	10	508,7	-52,04	27,4	25,7	14,6	37,5	0,6	-0,29
12	1	13,9	549,8	-9,295	32,7	8,6	18,6	41	1,51	0,25
13	1	14,2	668,5	-9,42	33,2	7,9	19,1	41,5	1,61	0,29
14	1	15,9	945,5	-68,89	36,9	3,9	20,9	43,5	2,51	0,56
15	0	14,8	1377	0	30,4	11,8	19,5	42,3	1,12	0,08
16	0	14,6	717,3	-56,47	25,6	17,2	19	41,7	0,62	-0,27
17	1	15	1093	-12,29	34,5	9,8	20	43,6	1,73	0,33
18	1	14,9	1560	-19,26	37	8	20,4	45,1	2,3	0,5
19	1	16,9	1376	-9,847	37,4	7,4	21,9	45,3	2,42	0,53
20	1	14,9	1215	-9,945	30,6	13,6	18,5	40,6	1,09	0,06
21	0	15,4	1114	-10,01	35,3	17,6	19,5	42	1,55	0,26
22	1	12,5	1054	-41,07	34,8	17,3	17,2	41,4	1,49	0,24
23	0	15,9	1431	-9,684	36,9	14,7	20,9	44,5	1,94	0,4
24	0	14,8	1783	-28,44	36,5	16,9	19,3	43,3	1,77	0,34
25	1	14,8	1421	-23,85	33,4	9,5	19,9	43,9	1,57	0,28
26	0	14,5	1284	-28,86	36,1	8,3	18,6	42,2	2,09	0,45
27	1	15,2	1080	-32,54	36,2	6,5	19,6	42,7	2,21	0,48
28	1	14,9	1560	-19,26	37	8	20,4	45,1	2,3	0,5
29	1	15,9	499,8	-28,71	37	4,3	20,7	43,5	2,51	0,55
30	1	14,9	1143	-0,313	36,5	6,4	19,2	41,4	2,28	0,5
31	1	15,7	1579	-9,454	35,6	11,9	20,2	43,3	1,83	0,37

32	1	15,5	1841	-9,582	34,2	25,1	19,3	41,7	1,17	0,09
33	0	13,6	1811	-14,41	31,6	15,7	19,2	43,3	1,14	0,09
34	1	11,4	1360	-10,07	21,3	42,7	17,2	41,3	0,22	-0,88
35	0	12,6	1401	-15,48	25,9	17,7	18,9	43,1	0,64	-0,26
36	1	14,5	1459	-15,57	35,1	12,6	20,3	44,3	1,72	0,33
37	1	15,9	1207	-18,79	36,4	11,9	21,5	44,7	1,97	0,41
38	0	13,9	636,4	-59,91	35,8	10,6	19,5	43,4	1,92	0,4
39	0	15,7	1767	-9,669	33,6	12,3	21,5	45,4	1,5	0,25
40	1	13,5	1832	-6,256	32,3	14,3	18,6	42,2	1,26	0,15
41	1	14,8	922,2	-6,563	34,2	9,8	20,2	42,9	1,69	0,32
42	1	15,4	903,7	-0,195	37,1	3,5	20,9	43,2	2,58	0,57
43	1	14,5	1277	-6,592	35,7	10	19,4	41,5	1,93	0,4
44	1	15,2	843,3	-28,21	38	2,6	21,5	44,9	2,87	0,64
45	1	12,5	1214	-22,05	34,7	14,7	18,4	42	1,57	0,27
46	1	9,3	847,2	-23,76	17,2	39,6	13,8	36,3	0,16	-1,07
47	1	10,1	801,1	-28,22	24,1	30,9	15,2	38,6	0,39	-0,55
48	1	14,5	1444	-0,126	33,4	13,3	19,3	42,1	1,44	0,22
49	0	13,6	1173	-16,12	33,4	26,3	18,2	41,1	1,05	0,03
50	1	12,7	1454	-11,08	27,5	29,7	17,2	40,4	0,55	-0,35
51	1	14,7	1711	0	32,4	12,5	20,1	43,9	1,33	0,18
52	1	14	863,4	-15,98	34,2	8,1	19	42	1,76	0,34
53	1	15,2	380	-42,09	35,1	5,8	20,4	43	2,02	0,43
54	1	15	774,1	-44,52	38,3	5,1	21,7	46,3	2,78	0,62
55	0	14,6	1233	-9,476	34,9	8	19,3	42,1	1,88	0,38
56	0	15,5	1623	-9,626	35,7	13,7	20,2	43,1	1,77	0,34
57	0	15,7	1478	-28,79	36,9	13,6	20,2	43,3	1,99	0,41
58	0	14,9	1111	-41	32,8	20,6	19,8	43	1,14	0,08
59	0	14,3	1133	-19,13	29,9	18,9	19,8	43,2	0,9	-0,05
60	0	9,3	987,3	-31,67	24,3	34,6	14,3	37,4	0,36	-0,6
61	0	15	738,4	-0,124	29,7	36,6	19,7	41,1	0,58	-0,33
62	1	12,3	1648	-3,213	32,8	19	18,1	41,7	1,18	0,11
63	1	13,7	1113	-0,13	36,7	4,8	19,1	42,3	2,41	0,53
64	0	14,8	1502	0	29,5	12	20,4	43,4	1,02	0,02
65	0	11,7	1097	-18,76	25,9	22	17,4	40,9	0,57	-0,32
66	1	13,6	721,8	-37,66	31,5	17	19,3	42,7	1,1	0,06
67	0	13,6	1305	-15,9	36,3	7,7	19,6	43,4	2,16	0,47
68	1	14,9	1765	0	34,9	13	19,9	42,8	1,67	0,31

69	1	15,2	1455	0	34,8	8,3	20,7	43,8	1,85	0,37
70	1	14,6	910,5	-9,351	35	7,4	20,4	44,1	1,93	0,4
71	1	13,8	1259	-9,176	33,6	11,1	19,2	42,3	1,54	0,27
72	0	13	1647	-15,56	27,8	31,3	18,7	42,7	0,55	-0,35
73	0	14,7	1427	-6,4	31	20,7	19,7	42,4	0,96	-0,02
74	0	14,8	1041	0	25,5	28,3	20,5	43,4	0,47	-0,44
75	1	12,9	994,5	-12,52	31,6	12,3	18,1	41,1	1,24	0,14
76	1	14,9	1466	0	32,5	11,7	20,2	43,4	1,37	0,2
77	1	14,1	1166	-9,612	36,6	6,6	20,3	44,1	2,29	0,5
78	0	14,3	912	-22,98	36,6	4,9	20,1	43,4	2,38	0,52
79	1	15	801,9	-19,15	37,3	5,2	21	44,2	2,53	0,56
80	0	14,5	1073	-25,7	31,1	16,6	20,1	42,8	1,07	0,05
81	0	14,8	868,6	-19,25	34,3	16,9	19,9	42,3	1,43	0,22
82	0	14,5	981,5	-25,58	33,3	19,3	20,5	44,2	1,23	0,13
83	1	15,3	1283	-25,4	37,8	8,8	21,2	44,8	2,43	0,53
84	1	13,5	1548	-0,124	31,9	13,8	18,9	41,7	1,23	0,13
85	1	13	1526	-6,248	26,7	30,3	18,1	41,1	0,51	-0,4
86	0	14,2	1312	0	31	26,3	19,2	41,7	0,84	-0,1
87	1	12,9	1336	0	24,3	24,6	18,6	41,9	0,46	-0,45
88	0	12,6	413,3	-15,74	29,1	23	17,7	40,1	0,76	-0,16
89	1	13,3	503,7	-34,37	35	9,4	18,5	41,6	1,84	0,37
90	0	13	715,9	-47,3	31,6	31,5	18	41,4	0,78	-0,15
91	1	14,3	485,1	-47,26	34	9,2	19,4	42,3	1,68	0,32
92	1	15,3	1152	-25,09	35,7	10,5	21,2	45,3	1,91	0,39
93	1	12,3	867,1	-78,4	34,5	14,9	18,4	43,4	1,53	0,26
94	1	13,5	593,1	-41,09	35,5	7,4	19,3	43,3	2,02	0,43
95	1	14,4	198,1	-28,54	35,2	5,9	19,7	42,5	2,04	0,43
96	0	14,5	645,5	-36,74	35,8	9,7	20,5	44,1	1,97	0,41
97	0	13,5	217,3	-57,79	33,9	11,4	18,9	42,2	1,58	0,28
98	0	14,2	536,7	-19,9	33,8	12,4	19,5	42,7	1,53	0,26
99	1	13,9	403,5	-19,71	32,2	17,8	19	41,7	1,15	0,09
100	0	14,5	386,7	-4,896	32,2	26,4	18,7	40,1	0,94	-0,04
101	0	14	714,6	-25,02	34,3	17	19,5	43,1	1,43	0,22
102	0	12,9	711,8	-83,46	33,2	13,1	18,5	42,7	1,42	0,21
103	0	14,4	620,4	-45,18	35,2	10,3	20,2	44,3	1,83	0,37
104	1	11,5	508	-38,91	32,5	19,3	16,8	40,8	1,14	0,08
105	1	13,5	1147	-50,94	34,2	11,3	19,2	43,3	1,63	0,3

106	1	14,7	1031	-70,88	39	10,8	21,6	46,9	2,59	0,57
107	0	13,5	1005	-47,17	35,4	11,4	20,5	45,6	1,82	0,36
108	1	13,3	577,4	-25,15	33,2	25,4	19,4	43,6	1,05	0,03
109	1	12	965,6	-28,35	32,7	15	18,5	42,7	1,29	0,16
110	0	13,1	986,9	-32,93	29,1	23,4	19,3	43,6	0,75	-0,17
111	1	13,5	759,5	-53,31	30,3	26,4	19,2	42,8	0,78	-0,14
112	1	11,9	1009	-21,95	31,5	12	18,4	42,7	1,24	0,14
113	0	14,2	839,9	-63,17	34,7	14,8	19,9	43,8	1,57	0,27
114	1	13,9	797	-50,62	33,6	19,8	19,8	43,9	1,25	0,14
115	1	12,6	985,1	-56,7	34,6	14,5	18,7	43,4	1,56	0,27
116	1	14,9	764,4	-37,95	36,7	10,5	20,5	44,2	2,1	0,45
117	0	14,3	537	-22,46	35,7	11,7	19,9	43,5	1,86	0,37
118	0	12,7	637	-108	33,6	13,6	18,6	43,7	1,45	0,23
119	1	13,1	1236	-47,94	35,1	14,4	18,8	43,5	1,64	0,3
120	0	13,5	770,8	-62,72	32,4	20,4	20,2	45,5	1,1	0,06
121	1	10,9	830,2	-82,18	29,9	27,3	16,3	41	0,74	-0,18
122	0	13,3	859,1	-44,22	33,8	13,3	18,9	43,1	1,49	0,24
123	1	14,1	880,8	-44,16	35,5	12,7	19,9	44,3	1,78	0,35
124	1	14,4	1019	-18,82	36,6	6,5	20,3	44,3	2,29	0,5
125	0	13,9	1140	-25,28	35	8,6	19,2	43	1,87	0,38
126	1	18,5	785,5	-34,75	35	8,1	18,9	42,8	1,9	0,39
127	1	17,9	592,9	-28,32	34,2	8,4	19,7	42,6	1,74	0,34
128	1	17,8	501,8	-20,48	31,5	11,1	18,6	41,1	1,27	0,15
129	0	18,2	964	-25,22	32,6	14,1	19,4	41,8	1,31	0,17
130	0	13,9	628,5	-62,72	34,1	14,9	18,6	41,6	1,48	0,24
131	0	13,1	966,3	-37,63	23,9	30,3	17,9	41,3	0,39	-0,56
132	0	12,5	1329	-31,62	33,3	14,6	17,4	41,1	1,38	0,2
133	1	14,9	1264	-42,75	32,3	11,1	20	43,6	1,37	0,19
134	0	13,3	572,5	-47,71	34,6	8,2	18,7	42,4	1,82	0,36
135	1	11,5	822,5	-31,71	29	15,3	16,8	40,9	0,9	-0,05
136	1	13,4	137,5	-66,27	35,6	7,3	18,9	42,6	2,04	0,43
137	1	14,3	371,6	-31,65	35,5	6,5	19,8	42,7	2,07	0,44
138	0	12,6	330	-31,68	34,3	7,9	18	41	1,78	0,35
139	1	15,3	354,8	-25,47	37,6	3,4	21,2	44,1	2,71	0,6
140	1	13,5	326,9	-45,54	35,5	8,4	19,4	43,1	1,97	0,41
141	1	13,5	673,7	-15,68	32	9,7	19,6	43,2	1,37	0,2
142	1	16	902,1	-17,34	37,9	3,7	21,7	44,7	2,77	0,61

143	1	15,1	1078	-12,7	37,4	7,6	21,3	45,2	2,41	0,53
144	0	12,1	1303	-50,56	31,8	12,6	18	42,5	1,26	0,14
145	1	13	1292	-23,87	31,1	17	18,9	43,3	1,06	0,04
146	0	12,5	663	-34,5	31,1	17,8	18,1	41,9	1,04	0,03
147	0	14,8	913,9	-31,68	32	25,2	19,7	43	0,94	-0,03
148	0	16	1143	-19,14	38	11,4	21,7	45,6	2,33	0,51
149	0	14,6	1122	-34,69	35,4	27,7	19,4	42,6	1,23	0,12
150	0	12,4	678,3	-19,26	28,6	22,8	18	41,7	0,73	-0,18
151	1	14,4	571,1	-25,28	27,9	28,5	18,6	40,5	0,59	-0,31
152	1	16,2	1197	-9,728	37,8	11,8	20,9	43,5	2,26	0,49
153	1	14,4	969,9	-50,37	34,3	8,6	19,4	43,1	1,75	0,34
154	0	14,4	718,1	-15,68	32,9	18,1	19,1	42	1,22	0,12
155	1	9,7	789,9	-40,77	13	36,9	14,4	37,8	0,12	-1,26
156	1	18	1342	-22,14	36,6	6,2	20,4	43,9	2,31	0,5
157	0	17,8	923,8	-24,38	36,5	6,1	20,1	43,7	2,29	0,5
158	0	18,8	345,9	-49,15	33,4	12	18,1	42,7	1,48	0,24
159	0	19,6	1383	-78,72	27	22,3	15,7	41	0,63	-0,27
160	0	15,5	1078	-33,18	33,4	17	21	44,7	1,31	0,17
161	1	14,2	1990	-44,22	34,3	15,9	19,9	44,9	1,47	0,23
162	0	13,9	2038	-37,95	32,5	23,7	19,2	43,6	1,03	0,02
163	0	12,7	1687	-18,82	25	38	17,5	41,2	0,36	-0,61
164	0	8,5	743,7	-72,32	20,7	48,4	13,3	36,9	0,19	-1
165	0	14,2	882,9	-15,68	30,4	30,7	19,1	42,1	0,71	-0,2
166	1	15	1053	-31,55	32,1	24,5	20,2	43,3	0,97	-0,01
167	1	14,5	1031	-16	36,5	10,5	20,6	44,8	2,06	0,44
168	1	13,5	1088	-25,28	34,6	8,8	19,3	42,5	1,79	0,36
169	1	10,1	921,1	-31,55	23,6	33,5	15,7	39,6	0,35	-0,62
170	1	10,9	858,7	-25,09	26,6	21,7	16,2	39	0,62	-0,28
171	1	15,1	1024	-14,56	37,1	8,2	21,3	45,6	2,31	0,5
172	1	15,2	682,1	-19,01	33,2	11,2	21	44,4	1,48	0,24
173	1	12,8	919,7	-40,77	25,3	21	18,9	42,8	0,55	-0,34
174	1	15,1	939,6	-12,54	35,5	8,9	20,5	43,7	1,95	0,4
175	1	13,8	780,2	-25,41	35,3	8,4	19,4	42,6	1,94	0,4
176	1	10,3	771,6	-34,5	23,1	22	16,1	39,6	0,44	-0,48
177	0	13,9	620,7	-22,14	35,4	6,9	20,4	44,2	2,03	0,43
178	1	13,6	1223	-9,6	33,3	9,2	20,5	44,6	1,57	0,28
179	0	8,9	854,8	-91,07	25,1	23,7	15,7	40,9	0,51	-0,39

180	1	11,9	780,4	-43,9	27,9	32,8	17,5	41,1	0,53	-0,37
181	0	10,4	962,3	-68,99	30	24,8	16,6	41,1	0,79	-0,14
182	0	14,3	1110	-37,63	37,3	7,6	20,9	45,8	2,38	0,52
183	1	13,9	1351	-28,22	31,2	14,3	20	44,2	1,14	0,09
184	0	14,8	1092	-37,95	36,6	10	21,3	45,8	2,11	0,45
185	1	13,9	672	-60	35,2	15,5	19,6	43,7	1,62	0,29
186	0	17,6	553	-28,61	33,9	26,9	20	43,1	1,09	0,05
187	0	17,9	474,4	-34,5	21,6	26,9	16,5	40	0,34	-0,63
188	1	17,2	201,3	-19,14	32,4	15,3	18,7	41,3	1,25	0,14
189	0	12,7	706,4	-72,45	31,1	20,8	19,4	43,7	0,96	-0,02
190	1	14,6	793,1	-66,24	31,2	13,2	20,8	45,1	1,17	0,1
191	0	12,5	1213	-31,36	32,2	21,2	18,9	43,6	1,06	0,04
192	0	9,3	338,8	-47,23	23,2	34,5	15,4	39,1	0,33	-0,66
193	1	13,8	1032	-43,9	33	27,4	19,4	42,3	0,98	-0,01
194	1	13,9	755,1	-31,74	36,1	9,4	20,8	45,1	2,04	0,43
195	0	12,9	508,7	-65,98	33,7	16,4	19,6	44,2	1,37	0,19
196	0	10,7	1038	-68,99	25,1	23,8	18,1	43,7	0,51	-0,39
197	0	11,1	652	-53,5	29,6	17,3	17,9	42,3	0,91	-0,05
198	0	13,1	773,5	-103,8	32,2	25,1	18,9	42,7	0,97	-0,02
199	1	14,6	897,4	-45,98	36,6	8,4	21	45,4	2,19	0,47
200	1	13,1	442,2	-64,61	33,2	21,9	18,7	42,8	1,15	0,08
201	1	12,7	865	-76,99	33,5	15,4	19,1	43,9	1,38	0,2
202	0	11,6	757,8	-47,04	26,1	30,5	17,3	41,1	0,48	-0,44
203	0	12,1	223,2	-23,71	28,5	30,6	17,6	40,4	0,6	-0,3
204	1	14	623,1	-50,62	32,3	21,1	19,8	43,3	1,07	0,05
205	1	13,3	1009	-47,68	35,2	8,8	19,9	44,6	1,9	0,39
206	1	13,8	386,7	-79,23	32,7	11,9	20,1	43,9	1,39	0,2
207	0	12,9	568,1	-55,2	33	19,5	18,7	42,3	1,19	0,11
208	1	13,2	929	-60,22	35,1	11,5	19,6	44,1	1,76	0,34
209	0	13,7	842,4	-34,69	33,8	13,6	19,5	43,2	1,48	0,24
210	0	12,7	520,4	-59,58	31,6	13,2	19	43,1	1,21	0,12
211	0	10,7	631,9	-53,31	27,9	25,8	16,7	40,9	0,63	-0,27
212	0	12,8	278,6	-47,36	33,2	14,6	18,9	42,9	1,37	0,19
213	0	10	1003	-6,272	17,4	35,6	15,9	39,7	0,18	-1
214	1	11,8	485,9	-60,54	31,7	15,4	17,5	41,4	1,16	0,1
215	1	12,7	480,4	-99,07	34,4	10,8	18,6	43	1,68	0,31
216	1	18,7	1255	-46,05	35,2	27,5	19,6	43,8	1,21	0,11

217	1	18,3	768,2	-63,17	30,4	20,9	20	44	0,9	-0,06
218	1	17,9	566	-41,09	34,1	13,4	18,9	42,1	1,53	0,26
219	0	18	992,6	-41,41	35,9	18,6	20,3	44,7	1,6	0,28
220	0	12,8	667,4	-38,08	34	23,9	18,8	42,9	1,18	0,1
221	1	12,5	975,2	-56,9	29,5	27,3	18,4	42,7	0,71	-0,2
222	0	13,6	1057	-47,81	36	14,8	20,3	45,3	1,77	0,34
223	0	13,1	924,1	-72,58	32,7	31,7	19	43,1	0,86	-0,09
224	0	13,1	1148	-34,5	32,6	24,2	18,7	42,3	1,02	0,02
225	0	10,5	833	-53,31	27,2	34,4	16,5	40,7	0,48	-0,43
226	1	12,7	438,3	-19,14	33,6	20,8	19,1	43,2	1,22	0,12
227	1	13,5	1208	-43,9	34,6	23	20,1	45,3	1,27	0,15
228	1	10,8	536,1	-28,22	26,7	39,6	16,2	39,9	0,4	-0,54
229	1	10,8	1480	-40,77	31,7	20,9	17,5	43,1	1,02	0,02
230	1	12,7	1407	-40,77	33,7	17,6	19	43,8	1,33	0,18
231	1	11,5	909,9	-75,27	33	20,2	17,8	43	1,17	0,1
232	1	11,6	1564	-47,04	27,2	26,1	17,8	43,3	0,59	-0,31
233	0	11,8	1241	-34,62	28,4	25,9	17,4	41,6	0,66	-0,24
234	0	13,2	1245	-18,98	29,5	26	18,2	41,7	0,73	-0,18
235	1	14,3	1166	-27,71	35,7	13,2	19,6	43,1	1,79	0,35
236	0	13,5	1115	-32,77	35,5	11,9	19,4	43,7	1,81	0,36
237	1	12,7	629,6	-50,82	33,1	12,2	18,9	42,9	1,43	0,22
238	0	11,8	1476	-49,31	34	11,6	18,2	43,6	1,58	0,28
239	0	13,7	840,4	-54,91	35,8	20,3	20,8	46	1,52	0,25
240	0	13,3	1524	-39,36	36	11,5	20,5	46,2	1,92	0,39
241	1	13,6	1368	-25,25	34,4	21,1	19,8	43,9	1,31	0,16
242	1	13,8	886,9	-52	35,3	12,5	20,3	44,9	1,75	0,34
243	1	13,8	1463	-41,28	36	15	19,2	43,1	1,76	0,34
244	1	13,5	355,6	-29,79	24,9	45,1	19,1	42,6	0,3	-0,72
245	1	15	1176	-20,64	36,6	10,8	21	44,8	2,07	0,44
246	1	14,9	1392	-28,42	35,5	7,3	20,8	44,5	2,03	0,43
247	1	18	1068	-26,75	32,1	10,6	19	42,6	1,36	0,19
248	1	18	667,1	-50,5	37,7	5,5	20,9	45	2,61	0,58
249	1	17,4	793,5	-22,3	37,3	4,3	21	44	2,58	0,57
250	1	17,6	726,4	-26,66	37,3	5,6	21	44	2,5	0,55
251	1	14,1	787,4	-36,45	35,6	7,9	19,2	42,1	2,02	0,42
252	1	15,3	1166	-34,69	37,7	5,2	20,9	44,7	2,62	0,58
253	1	14,5	849,8	-30,05	35,9	7	20,1	43,5	2,12	0,45

254	1	12,8	647,5	-48,96	33,1	10,3	18,9	43,1	1,5	0,25
255	1	13,5	355,6	-29,79	24,9	45,1	19,1	42,6	0,3	-0,72
256	0	10,3	733	-18,82	22,7	20,6	17,5	41,5	0,44	-0,48
257	0	10,8	665,5	-90,95	25,4	19,1	17,7	42,3	0,59	-0,31
258	0	11,6	638,3	-87,81	28,2	22,8	18,7	43,6	0,7	-0,21
259	0	12,1	801,8	-50,18	32,9	23,3	18,4	43	1,08	0,05
260	1	13	1292	-31,36	29,6	33,8	18,3	42,2	0,61	-0,29
261	0	13,1	640,8	-46,11	35,3	16,1	19,2	42,7	1,61	0,29
262	0	11,3	555,8	-44,03	27,1	22	17,8	41,6	0,64	-0,26
263	1	10,4	673,4	-72,64	24,8	39,7	16,6	41,1	0,34	-0,64
264	0	14,2	786,7	-15,87	35,7	16,2	19,9	43,3	1,67	0,31
265	0	13,3	1023	-3,136	31,4	14,5	19,4	43,1	1,16	0,09
266	0	13,9	1201	-12,54	35,9	14,3	21,2	46,6	1,78	0,35
267	0	14,1	840,4	-15,87	35,9	10,3	20,5	45	1,96	0,41
268	0	14,5	1015	-21,95	35	25,4	20,5	44,4	1,25	0,13
269	0	14,1	1229	-15,81	27,8	36,6	20	43,9	0,48	-0,43
270	0	12,9	852,7	-12,54	24,2	46,3	17,3	39,7	0,27	-0,78
271	1	13,6	707,8	-12,54	32,2	31,2	18,3	41	0,83	-0,11
272	1	13,8	864,4	-23,62	36,3	8,4	20,4	45,4	2,13	0,46
273	1	14,3	936,7	-18,91	35	10,1	20,3	44,4	1,81	0,36
274	0	13,5	847,4	-18,98	34,3	10	19,7	43,6	1,69	0,32
275	0	8,3	1651	-14,24	23,6	20,8	14,5	38,6	0,47	-0,43
276	0	8,8	982,4	-25,09	4,5	40,2	15	38,1	0,05	-1,78
277	0	17,2	718,8	-53,31	13,8	39,3	17	40,7	0,12	-1,25
278	0	18,1	1413	-15,68	30,8	19,5	19,4	43,3	0,97	-0,01
279	1	17,6	1125	-6,272	29	26,3	19	42,2	0,69	-0,21
280	0	13,6	1086	-7,84	28,2	20,5	19,4	43	0,74	-0,17
281	1	12,6	728,7	-12,54	29,4	23,3	18,6	41,8	0,77	-0,15
282	1	14,9	290,9	-10,98	32,7	22,8	20,9	44	1,07	0,04
283	1	15,2	1012	-9,408	36	14,5	21	44,7	1,79	0,35
284	1	14,9	700,4	-15,68	34,5	10,4	21	44,4	1,71	0,33
285	1	13,8	612	-21,95	35,5	7,5	19,9	43,6	2,02	0,42
286	1	12,8	573,3	-25,15	33,9	7,6	18,9	42,4	1,73	0,33
287	0	11,5	718,1	-12,54	27,3	18	17,8	41,9	0,72	-0,19
288	1	14	837,8	-19,01	35,1	8,9	20,8	45,2	1,88	0,38
289	1	14,1	1204	-28,42	36	9	20,8	45,6	2,04	0,43
290	0	9,3	799,9	-18,82	23,9	26,6	14,8	38,3	0,42	-0,5

291	0	12,2	643,1	-65,92	16,3	36,9	18,5	42,7	0,16	-1,08
292	0	14,7	1223	-12,54	28,4	31,9	20,4	43,4	0,57	-0,33
293	0	12,9	1303	-21,95	27,4	21,4	21,1	47,7	0,67	-0,23
294	1	13,7	940	-6,272	31,7	25,5	21	45,6	0,91	-0,05
295	0	13,6	1005	-25,09	35,7	9,2	21,1	46,1	1,97	0,41
296	0	12,9	1215	-15,68	33,4	10,5	20,5	45,1	1,54	0,26
297	0	12	912,1	-34,5	26,2	42,5	19,2	43,5	0,36	-0,61
298	1	13,5	623,6	-15,81	33,9	11,2	21	45	1,58	0,28
299	1	12,5	401,9	-15,68	25	49,1	18,5	41,1	0,27	-0,77
300	0	11,6	433	-9,408	26,1	39,1	17,5	40,1	0,39	-0,56
301	0	12,3	591,2	0	27,4	34,3	19,2	42,9	0,49	-0,42
302	0	13,1	587	-56,58	35	10	20,6	45,5	1,81	0,36
303	1	12,5	469	-22,14	33,3	11,4	19,4	43,6	1,49	0,24
304	0	11,5	579,8	-12,54	28,2	20,8	17,3	40,4	0,73	-0,18
305	0	17,8	844	-31,36	36,2	15,6	21,1	45,3	1,77	0,34
306	0	17,3	547,8	-12,54	33,4	17,7	19	42,4	1,29	0,16
307	0	17,7	568,1	-15,68	22,6	33,7	15,6	39,1	0,32	-0,68
308	0	13,2	436	-9,408	27,8	25,7	18,9	41,6	0,63	-0,27
309	0	13,6	723,7	-20,38	33,2	23,6	19,6	43	1,1	0,06
310	1	12,6	753,7	-12,54	19	59,6	17,4	39	0,12	-1,26
311	0	12,4	519,9	-31,36	27,9	24,7	17,4	39,9	0,65	-0,25
312	0	13,4	537,2	-12,54	29,2	23,7	18,9	41,8	0,75	-0,16
313	0	12,8	734,6	-17,44	34,1	10,9	19,1	42,8	1,63	0,3
314	0	13,6	848,6	-6,272	28,7	21	19,6	43,4	0,76	-0,15
315	0	12,1	567,5	-50,18	25,8	25,3	16,9	39,5	0,52	-0,38
316	1	14,3	737,5	-28,35	35,5	11,3	20,6	44,6	1,84	0,37
317	1	11,9	592,2	-12,54	25,6	27,8	17,9	41,1	0,48	-0,42
318	0	13,3	428,5	0	30,2	25	18,7	40,9	0,8	-0,13
319	1	12,4	549,2	-6,272	27,1	28,6	17,8	40,5	0,55	-0,35
320	0	12,3	370,3	-9,536	30,9	23,7	18,2	41,1	0,88	-0,07
321	1	13,5	271,1	-28,67	34,1	14	20,4	44,3	1,51	0,25
322	0	12,8	669,9	-78,98	28,4	18,1	19,6	44,5	0,8	-0,13
323	1	11,8	678,9	-50,18	26,6	25,8	18,4	42,9	0,56	-0,34
324	1	13,6	731,8	-28,22	35,3	13,9	19,4	43,3	1,7	0,32
325	1	15,1	1091	-16	36,3	14,9	21,4	45,1	1,82	0,36
326	0	12	715	-6,272	16,6	46,2	16,9	39,3	0,13	-1,2
327	0	9,2	201,1	0	14,2	43,9	13,9	35,6	0,11	-1,3

328	1	14	278,9	-4,704	31,4	26,1	18,6	40,1	0,87	-0,08
329	0	14,1	675,8	-10,98	30,1	28,9	19,1	41,9	0,72	-0,19
330	0	12,8	680,8	0	26,1	37,6	17,6	40,3	0,4	-0,54
331	1	14,2	764	-16	33,6	20,7	20,1	43,7	1,23	0,12
332	0	14	448,1	-29,31	32,9	18,4	19,3	42,7	1,21	0,12
333	0	13,2	605,4	-6,272	31,2	24,9	18,8	41,7	0,88	-0,07
334	0	14,1	358,5	-12,54	27,5	33,8	19,3	42,2	0,5	-0,41
335	1	17,1	659,2	-9,6	34,1	14	20,7	44,1	1,51	0,25
336	0	17,7	663,2	-21,95	25,9	17	21	45	0,65	-0,25
337	1	17,2	339,5	-9,408	32	23,2	18,4	41,1	0,99	0
338	0	17,3	409,7	-6,784	35,6	9,3	20,2	42,9	1,95	0,4
339	1	13,5	829,3	-25,41	35,8	8,9	19,4	43,3	2,01	0,42
340	1	13,7	431,9	-44,35	36	10,7	19,8	43,3	1,96	0,41
341	1	13,7	123,6	-56,64	35	14,5	19,5	43	1,62	0,29
342	1	12,1	439,5	-18,82	20,6	41	16,6	38,5	0,22	-0,9
343	1	14,3	271,9	-28,22	35,1	8,7	20,9	44,7	1,89	0,38
344	1	13,7	657,9	-25,09	32,1	15,8	19,8	43,6	1,2	0,11
345	0	13,6	1147	-22,14	35,4	12,4	20,9	45,7	1,77	0,35
346	0	13,2	1159	-31,36	31,5	30,4	19,2	43	0,79	-0,14
347	1	12,4	711,5	-25,09	24,2	32	17,9	41,2	0,38	-0,56
348	1	14,3	910	-22,27	36,8	12,4	21	45,1	2,03	0,42
349	0	10	433,8	-98,3	34,1	16,2	17,1	42,9	1,43	0,22
350	0	5,9	630	0	17,7	43,2	11,7	35,6	0,16	-1,09
351	0	11,4	201,2	-9,408	18,1	45,8	16	37,6	0,15	-1,11
352	1	12	532	-40,9	28,1	19,2	18,4	42,6	0,75	-0,16
353	1	12,6	583,7	-20,99	30,5	21,4	18,6	42,7	0,9	-0,06
354	1	14,6	488,7	-39,68	36,7	9,6	20,7	44,4	2,15	0,46
355	1	13	472,1	-51,84	34	11,8	19,5	44,1	1,58	0,28
356	0	13,1	722,8	-76,99	36,1	12,7	20,1	45,5	1,88	0,38
357	1	13,1	792,4	-9,408	34,5	9,7	20	43,6	1,74	0,34
358	0	12,8	825,3	-25,09	32,5	21,8	18,9	42,5	1,08	0,05
359	0	12,8	631	-37,76	32,4	18,8	18,8	42,6	1,14	0,09
360	1	13,1	729,8	-31,55	31,2	28,2	18,9	42,5	0,81	-0,12
361	1	13,2	557,9	-28,48	32,7	29,4	19	42,1	0,91	-0,05
362	1	14,6	887,2	-11,07	36,1	13,3	20,7	44,2	1,86	0,37
363	0	15	691,2	-22,18	32,9	19	19,9	42,5	1,19	0,11
364	1	13,6	749,8	-20,54	34,5	11,1	20,1	44,9	1,68	0,32

365	1	18,6	461	-58,46	32,5	13,7	19,1	43,5	1,31	0,17
366	1	18,2	557,4	-41,63	37,8	5,5	20,9	45,1	2,63	0,58
367	1	17,5	497,7	-27,01	37,4	3,9	20,7	43,8	2,63	0,58
368	0	17,8	1077	-6,272	21,4	23,8	17,8	41	0,36	-0,6
369	0	12,3	501	-31,36	30,5	26,3	18	41,6	0,8	-0,13
370	0	14,1	852,7	-4,704	18,2	37,6	19	41,1	0,19	-0,98
371	0	14,2	280,8	-7,84	27,4	31,5	18,8	40,7	0,52	-0,38
372	0	10,4	312,8	-9,408	14,9	47,1	16,1	40,1	0,11	-1,3
373	0	12,8	509,4	0	20,7	36,5	19,2	43,4	0,25	-0,83
374	0	10,8	458,2	-18,82	18,5	41,7	16,1	39,6	0,18	-1,02
375	0	13,2	491,8	-6,272	19,5	46,2	18,7	42,3	0,17	-1,03
376	0	13,6	463	-15,78	25	37,3	19,1	42,7	0,36	-0,6
377	0	14,4	449,4	0	27,9	33,5	19,2	41,3	0,52	-0,38
378	0	11,4	360,1	0	18,2	27,7	16,9	40	0,24	-0,83
379	0	14,3	385,5	-3,136	33,7	13,6	19,6	42,1	1,47	0,23
380	0	12,4	417,4	-23,52	29,2	16,8	18	41,7	0,89	-0,06
381	0	15,3	1041	-12,54	35	10,9	21,1	45	1,77	0,35
382	0	10,2	546,3	-3,136	29,9	13,6	15,9	39,4	1,02	0,02
383	0	12,3	448,8	-4,768	26,6	16,6	17,7	40,7	0,7	-0,2
384	0	11,1	511,5	-4,704	29,8	19,7	16,7	40,4	0,88	-0,07
385	0	13,7	575,2	-26,66	28,8	27,2	19,7	43,6	0,66	-0,24
386	0	12,4	1037	-21,95	21,6	25,8	18,6	43	0,35	-0,62
387	1	11,9	828,4	-6,272	31,3	14	18,2	42,7	1,16	0,1
388	1	13,9	1347	0	35,9	8,1	20,1	44,1	2,06	0,44
389	0	11,8	814,6	-25,09	25,8	18,5	17,8	41,8	0,62	-0,28
390	1	12,9	880,1	-6,272	31,1	15,3	19,1	43,3	1,1	0,06
391	1	14	738,6	-3,136	33,2	15,1	20,1	43,8	1,35	0,18
392	0	11,8	514	-12,54	30,8	14,3	18,1	42,2	1,1	0,06
393	1	10,2	434,2	0	22,4	30,1	16,2	39,5	0,34	-0,64
394	1	14,9	1086	-12,54	36,8	13,8	21,4	45,8	1,96	0,4
395	1	18	1101	-28,42	33,7	13,2	20,2	44,9	1,48	0,24
396	0	17,8	963,3	-21,95	25,7	18,6	19	43,5	0,61	-0,28
397	0	16,9	891,6	-6,272	27,8	20	19,5	42,9	0,72	-0,19
398	0	16,9	765,2	-34,5	31,4	12,9	20,9	44,5	1,2	0,12
399	0	13,2	715,5	-17,25	24,6	30,4	18,6	41,7	0,41	-0,52
400	1	13,5	686,8	-25,09	31,7	18,3	19,2	42,8	1,08	0,05
401	0	13,8	1180	-15,81	33	14	19,6	43,4	1,36	0,19

402	0	10,4	422,3	-6,272	23,5	22,9	15,9	39,1	0,45	-0,47
403	1	14,2	631,8	-6,272	25,5	24,1	20	42,5	0,52	-0,38
404	0	13,3	971,4	-15,68	34,5	8,1	18,8	41,9	1,81	0,36
405	1	13,6	906,8	-19,14	35,6	10,6	19,2	42,8	1,89	0,38
406	1	12,9	398,1	-12,54	14	39,6	18,2	40,4	0,12	-1,24
407	1	12,7	351,4	-6,272	27,6	20,5	18	40,4	0,7	-0,21
408	1	14,2	784,2	-12,54	33,9	16,4	20,7	45,1	1,4	0,2
409	1	12,3	477,3	-33,41	33	13,2	19	43,8	1,39	0,2
410	1	13,6	732,5	-28,22	34,1	15,1	20,7	44,9	1,47	0,23
411	0	11,3	639,5	-18,82	14,5	29,3	17,7	41,3	0,16	-1,06
412	0	11,1	258,8	-18,82	14,2	38	16,9	39,9	0,13	-1,21
413	0	14,6	408,2	-18,82	32,9	35,1	19,6	41,9	0,81	-0,13
414	1	11	449,9	0	25,8	34	16	38,9	0,42	-0,51
415	1	10,8	735,6	-25,09	28	20	16,9	41,1	0,73	-0,18
416	0	13,6	663,4	-67,42	32,4	15	19,8	44,3	1,25	0,14
417	0	13,3	1174	-15,94	24,2	29,5	19	42,8	0,41	-0,53
418	0	13,2	686,8	-15,81	31,2	26,2	19,2	42,9	0,85	-0,09
419	1	11,6	750,5	-18,82	31,5	29,1	18,1	42,4	0,82	-0,12
420	1	13,7	548,5	-31,49	31,2	15,4	20,2	44,3	1,11	0,07
421	1	11,9	932,5	-9,408	19	26,6	18,5	42,5	0,27	-0,77
422	0	7,3	611,7	-18,82	18,4	40,6	13,7	38,2	0,18	-1,01
423	0	11,7	218,7	-23,68	31,7	46,4	16,4	38,3	0,55	-0,36
424	0	16,9	655,5	-21,95	29,3	35,4	18,5	41,1	0,57	-0,33
425	0	16,9	512,5	-18,82	19,9	52,5	14	36,5	0,16	-1,11
426	1	17	418,2	-9,536	26,4	25,2	17,6	40,9	0,56	-0,34
427	1	17,9	1128	-22,14	35,3	10	20,7	45,2	1,86	0,38
428	0	14	923,3	-28,22	34,9	22,8	20,5	45	1,32	0,17
429	1	13,7	829,6	-12,54	35,3	23,8	20,8	45,4	1,34	0,17
430	0	14,4	1267	-28,54	36,4	7,8	21,7	47	2,18	0,47
431	0	10,8	785,3	-28,22	16,9	33,2	17,3	41,4	0,19	-0,99
432	0	12	582,9	-42,34	19,2	44,6	18,2	41,6	0,18	-1,03
433	1	13,2	532,7	-15,68	28,4	34,6	18	39,8	0,54	-0,37
434	0	13,2	557,6	-10,98	33	30,6	18,2	40,4	0,91	-0,05
435	1	13,4	781,2	-37,63	34,2	22,7	18,7	41,6	1,24	0,13
436	0	12	1052	-15,68	12,6	46,3	18	41,5	0,09	-1,42
437	0	14	1037	-12,54	31,3	22,9	20,4	44,4	0,93	-0,04
438	1	12,1	937,7	-28,22	30,4	26,9	17,6	41,4	0,78	-0,14

439	1	12	815,1	-12,54	26,4	26,8	17,9	41	0,53	-0,37
440	1	11,8	989	-3,136	29	17,3	18,6	42,5	0,86	-0,08
441	0	9,3	475,9	0	19,5	30,5	15,2	38,6	0,25	-0,8
442	0	14,4	848,1	-18,82	33	11,2	21,5	45,9	1,46	0,23
443	0	12,2	975,9	-28,22	22,2	42,4	18,5	42,5	0,25	-0,83
444	0	12,4	911,2	-18,82	21,8	39,8	18,6	42,5	0,25	-0,81
445	1	13,5	664,2	0	27,7	24,2	18,8	40,9	0,64	-0,26
446	0	13,3	385,8	-19,97	34,3	10,1	19,6	42,9	1,69	0,32
447	0	12,1	58,75	-90,82	31,3	18,8	17,9	40,8	1,03	0,02
448	1	12,3	391,5	-29,82	33,2	20,5	18,4	42,3	1,19	0,11
449	0	13	598	-37,63	28,9	26,5	18,7	41,5	0,68	-0,22
450	0	12	628,2	-25,22	27,6	30,6	18,2	42,1	0,55	-0,35
451	1	14	687,2	-21,95	35,3	19,2	21	45,2	1,49	0,24
452	0	11,2	739,8	-37,63	29,7	18,2	18,3	43,2	0,9	-0,06
453	1	13,8	816	-37,63	30,5	18	21	46,2	0,97	-0,01
454	0	17,5	1097	-6,272	18,8	34,1	20,9	45,1	0,22	-0,9
455	0	18	593,3	-18,82	16,6	34,7	16,3	40,4	0,17	-1,03
456	0	17,7	414,5	-9,408	20,9	34,5	16,2	38,9	0,26	-0,78
457	0	18,2	747,3	-15,68	31,6	32,3	19,5	42,8	0,77	-0,16
458	1	11,8	985,3	-22,08	29,4	20,8	18,1	42,6	0,82	-0,11
459	1	13	1067	-15,68	31,7	16,3	18,8	42,8	1,14	0,08
460	1	12,3	751,6	-34,5	33,9	14,7	19,1	44	1,46	0,23
461	1	13,2	607,9	-78,4	35,8	8,5	19,6	44,7	2,02	0,43
462	1	15,1	664,4	-25,28	38,5	2,4	21,5	45	3,03	0,67
463	1	6,9	626,7	-9,408	9,1	51,2	12,1	36,1	0,06	-1,69
464	1	14,2	886	-25,09	35,1	21,4	20,3	44,3	1,39	0,2
465	0	10,2	335,5	-6,272	28,4	19,5	15,5	38,2	0,77	-0,15
466	0	12,8	513,1	-34,5	28,8	40	18,8	43,3	0,49	-0,43
467	1	13,9	690,6	-12,83	31,5	24,1	19,2	42,1	0,93	-0,04
468	1	14,5	739,3	-9,408	34,2	10	20,7	44,3	1,68	0,32
469	1	14,1	516,7	-38,59	36,7	11	20,5	44,8	2,08	0,44
470	0	13,4	706,7	-9,408	31,2	15	19,5	43	1,12	0,07
471	0	10,8	950	-56,45	19,4	37,5	16,7	40,4	0,21	-0,91
472	1	11,4	569,6	-9,408	19,8	43	16,3	38,4	0,19	-0,97
473	1	14,5	766,1	-51,81	36	17,4	20,5	44,5	1,67	0,31
474	0	13,4	1068	-23,52	30,5	21,1	20	44,7	0,9	-0,05
475	1	14,3	827,8	-41,89	36,6	18,3	20,1	43,9	1,72	0,33

476	1	13,1	591,6	-48,32	35,9	16,7	19,8	44,4	1,68	0,31
477	1	12,6	379,4	-38,98	33,1	13,9	18,6	42,2	1,38	0,2
478	0	13,8	256,8	-51,26	32,9	9,6	20,3	43,5	1,5	0,25
479	0	11,5	229,2	-53,79	34,5	13,1	18,4	43,5	1,6	0,29
480	0	11,8	211,3	-50,5	20,3	26	17,8	41,8	0,31	-0,69
481	0	10,7	314,9	-17,25	18,8	24,6	16,6	40	0,27	-0,76
482	0	12,5	271,2	-39,2	22	22,9	18,9	42,7	0,39	-0,55
483	0	18,1	708,5	-20,45	33,1	18,9	21,1	46,5	1,22	0,12
484	0	19,3	198,5	-91,39	25,2	34,8	13,6	40	0,39	-0,55
485	1	19,2	368,1	-94,46	25	33,9	18,2	43,6	0,39	-0,55
486	1	18,7	765,4	-20,51	28,1	41,5	17,8	41,3	0,44	-0,49
487	1	14,3	719,6	-3,136	35,4	9,5	20	43,3	1,9	0,39
488	1	9,9	211,3	-17,31	31,3	15	17	41,5	1,13	0,08
489	0	9,8	116,7	-31,36	31,9	13,3	17	41	1,25	0,14
490	0	10,7	12,54	-22,14	31,2	17,6	18,1	42,5	1,05	0,04
491	0	13,3	431,1	-31,36	31,3	17,6	20	44,1	1,06	0,04
492	0	13,6	235,4	-97,28	32,2	21,5	19,3	43	1,05	0,03
493	0	14,9	221,9	-72,9	18,9	26,3	20	42,4	0,27	-0,78
494	0	14,6	336,4	-6,272	30,2	22,9	20,1	43,2	0,84	-0,1
495	0	14,7	430,6	-6,272	32,2	29,3	19,6	42,2	0,87	-0,08
496	0	13,4	271,5	-34,56	28,6	29,6	18,9	42,7	0,62	-0,28
497	0	13,2	279,9	-37,63	21,8	42,3	18,8	42,7	0,24	-0,85
498	0	11,4	235,5	-18,94	25	41,7	16,3	39,8	0,33	-0,66
499	0	8,5	474,4	-23,65	25,3	43,3	14	38,9	0,32	-0,67
500	0	13,3	720,5	-4,704	26,3	28,3	19,3	42,9	0,51	-0,39
501	0	13,3	235,8	-56,45	22,2	39,7	19	41,9	0,26	-0,79
502	0	10	451,7	-45,54	-0,8	61,8	16	39,8	0,02	-2,39
503	0	12	420,6	-32,93	12,7	55	18,3	42,4	0,07	-1,54
504	1	14,9	790,5	-6,368	36,9	17,1	20,1	43	1,83	0,36
505	1	9,5	617,1	-12,54	20,4	23,8	15,4	39,4	0,33	-0,66
506	0	13,8	713,2	-9,408	32,8	10,1	19,7	43	1,47	0,24
507	1	11,7	238,2	-20,38	30,1	13,1	17,9	41,8	1,06	0,04
508	0	13,2	239,5	-9,408	28,2	17,5	18,4	41	0,79	-0,13
509	0	13,5	333,4	-32,93	22,9	33,7	17,9	39,9	0,33	-0,66
510	0	13,5	700,5	-21,95	30,5	18	18	40,7	0,97	-0,01
511	0	11,1	670,7	-7,904	10	42,9	16	38,9	0,08	-1,51
512	0	7,4	425,5	-9,856	15,6	61,5	11,9	34,6	0,08	-1,48

513	0	17,5	504,6	-28,74	20,6	48,1	16,2	38,1	0,18	-1
514	0	17,3	363,4	-78,47	9,3	46,3	18,4	39,9	0,07	-1,6
515	0	18	461,4	-40,77	25,7	35,6	16	38	0,4	-0,53
516	1	18	583,5	-43,9	33,9	27,4	19	41,5	1,07	0,04
517	0	15,2	944,4	-6,272	28,2	21	20,8	43,9	0,73	-0,18
518	0	14,7	796,5	-6,272	33,1	14,3	19,9	42,8	1,36	0,19
519	0	13	747,3	-6,272	32,9	18,1	18,5	42,1	1,22	0,12
520	1	13,6	704	-6,272	23,5	33,3	19,3	42,9	0,35	-0,62
521	0	13,6	714,4	0	30,5	15,5	19,3	42,4	1,04	0,03
522	1	15,1	900,7	0	35	18,5	20,3	43,2	1,47	0,23
523	0	13,3	545,9	-10,98	30,6	18,4	19,2	42,7	0,97	-0,01
524	1	13,8	620,5	-18,82	31,7	22,2	19,8	43,5	0,99	0
525	1	12	461,4	-40,77	25,7	35,6	16	38	0,4	-0,53
526	0	11,5	934,1	-3,136	21,1	21,1	17,4	40,4	0,37	-0,58
527	0	12,8	376,4	0	19,4	26	18,5	41	0,28	-0,74
528	1	12,4	452,9	-14,11	26,8	26,1	18	41,4	0,56	-0,33
529	1	15,1	716,8	-19,07	35,6	14,5	20,9	44,4	1,72	0,33
530	0	14	487,8	-25,22	29	21	19,5	42,3	0,79	-0,14
531	1	12,6	476	0	29,5	24	18	41,1	0,77	-0,15
532	1	12,1	434,9	0	16,9	30,9	18,1	40,8	0,2	-0,95
533	0	12	933	-10,98	34	13,4	18	42	1,52	0,25
534	0	12,6	564,2	-21,95	27,8	17,2	18,8	43	0,77	-0,15
535	0	13,6	517,6	-21,95	33,6	11,9	19,4	42,9	1,51	0,25
536	0	14,7	836,9	-25,09	34,8	12,2	21,6	46,1	1,68	0,32
537	0	8,9	688,6	0	11,2	37	14,7	37,7	0,1	-1,36
538	0	12,8	156,3	-9,536	23,8	26,9	18,8	41,4	0,42	-0,51
539	0	16,9	592,9	-6,272	15,1	44,2	16,7	39,2	0,12	-1,25
540	0	16,7	125,9	-4,704	15,8	40,1	13,3	34,7	0,14	-1,15
541	0	18	553,7	-15,68	32,5	28,6	19,5	43,1	0,91	-0,05
542	0	17,4	905,2	-12,54	17,2	30,5	17,6	41,7	0,2	-0,93
543	0	9,1	438,6	-18,82	12,1	34,2	14,9	38	0,12	-1,27
544	0	10,8	319,5	-6,272	26,1	24,4	16,4	39	0,55	-0,35
545	1	13,4	654,5	-31,36	31,1	19,7	19,4	42,6	0,99	0
546	1	13,9	970,1	-4,704	32,3	10,5	20,3	43,7	1,39	0,2
547	1	13,8	656,9	-6,272	32,3	11,1	20,3	43,3	1,37	0,19
548	0	12	548,5	-10,98	25,7	19,3	17,7	40,8	0,6	-0,29
549	0	13,7	522,5	-6,272	30,5	17,5	19,6	42,5	0,99	0

550	1	14	618,7	-37,63	32,6	16,2	19,8	43,3	1,24	0,14
551	0	14,6	848,2	-39,36	34,6	10,1	21,2	45,9	1,74	0,34
552	0	14,4	1320	-9,408	29,4	25,9	21	44,8	0,73	-0,19
553	1	13,5	737,8	-15,68	31,2	23,6	19,2	42,1	0,91	-0,05
554	1	15,2	709,6	-18,82	34,6	17,7	21,1	44,7	1,45	0,22
555	0	8,5	825,9	-12,54	24,9	29,2	15	39,5	0,44	-0,48
556	0	12,2	355,1	0	20,2	40	17,9	40,2	0,22	-0,9
557	1	14	415,7	-14,11	32,4	16,7	19	41,2	1,2	0,12
558	0	11,3	588,7	-9,408	20,5	31,6	16,7	40,2	0,27	-0,76
559	1	11,5	582,9	-6,272	30,3	25,1	17,9	42	0,81	-0,12
560	1	13,8	353,8	-25,09	35	8,8	20,7	45,1	1,86	0,38
561	0	13,5	762,7	-42,46	34,4	9,4	20,5	44,4	1,74	0,34
562	0	11,1	831,9	-37,63	29,1	16,4	17,7	42,2	0,89	-0,06
563	0	13,6	718,6	-26,66	30,2	17,7	19,7	43,5	0,95	-0,02
564	1	13,6	823,5	-28,22	29,6	22,3	20,3	44,5	0,81	-0,12
565	0	12,8	550,8	-25,34	31,2	15,1	19,2	43,2	1,12	0,07
566	0	12,3	784	-23,52	24,1	42,8	17,5	40,5	0,29	-0,73
567	1	12,8	499,1	-47,36	33,3	21,4	19,1	43,1	1,17	0,1
568	0	10	686,8	-44,16	28,3	17,7	17,5	42,7	0,8	-0,13
569	0	17,9	469,9	-9,408	23,5	37,3	16,5	39,3	0,32	-0,68
570	1	18	680,4	-47,49	33,5	15,7	19,8	44,1	1,37	0,19
571	1	17	511,1	-22,08	32,8	15,5	20,4	43,9	1,29	0,16
572	1	14	782,2	-28,32	34,6	13,8	20,4	44,2	1,59	0,28
573	0	10,9	481,9	-9,408	19,6	50,2	16,6	39,7	0,16	-1,09
574	1	13,7	510,2	-9,408	32,9	17,7	19,7	43,2	1,23	0,13
575	0	13,6	1010	-37,95	35,5	14,8	20,7	46	1,69	0,32
576	1	14,7	814,1	-37,63	35,8	20,9	21,2	45,7	1,5	0,24
577	1	13,1	705,3	-31,36	32	25,4	20	44,4	0,94	-0,03
578	1	14,4	713,3	-51,74	32	19,1	20,2	43,7	1,09	0,06
579	1	14,5	521,2	-28,35	34,9	13,7	20,5	43,7	1,64	0,3
580	0	14	1014	-21,95	31,8	26,5	20	43,9	0,9	-0,06
581	0	14,8	706,8	-18,82	29,5	19	21,4	45,1	0,87	-0,08
582	0	12	577	-32,93	20,5	44	17,8	40,7	0,2	-0,95
583	1	12,6	542,7	-9,504	25,7	25,9	18,5	41	0,51	-0,39
584	0	13,7	286,7	-12,61	29,2	17	19,3	41,8	0,88	-0,07
585	0	14	527,9	-9,728	32,3	18,2	19	41,5	1,15	0,09
586	1	12,6	375,3	-58,62	35,4	9	18,9	43,5	1,93	0,4

587	0	13,7	971	-25,09	34	11,8	19,7	43,7	1,58	0,28
588	1	14	1128	-36,13	29,2	19,3	20,9	46,2	0,84	-0,1
589	0	14,4	1384	-25,22	36,5	17,5	21,2	46,3	1,74	0,33
590	1	15,5	872,8	-7,904	35,8	7	21,8	45,6	2,1	0,45
591	0	12,6	867,3	-20,38	33,2	10,6	19,6	44,2	1,5	0,25
592	1	9,6	839	-18,94	19,1	48,2	15,8	40	0,16	-1,09
593	1	11,6	942,6	-59,58	34,6	20,2	19	44,8	1,36	0,19
594	1	12,8	750,6	-14,18	33,2	22,6	19,3	43,8	1,13	0,07

ANEXO 2. Muestreo aleatorizado.

- Se parte de la muestra completa en Excel aleatorizando N°.
- Luego se verifico que no existieran valores repetidos.
- Se procedió a dar la ruta para que se ubicara toda la data correspondiente según el caso seleccionado.

=BUSCAR(AAi;\$N\$3:\$N\$596;\$O\$3:\$O\$596) en P01

=BUSCAR(AAi;\$N\$3:\$N\$596;\$P\$3:\$P\$596) en Td850

=BUSCAR(AAi;\$N\$3:\$N\$596;\$Q\$3:\$Q\$596) en CAPE

=BUSCAR(AAi;\$N\$3:\$N\$596;\$R\$3:\$R\$596) en CINH

=BUSCAR(AAi;\$N\$3:\$N\$596;\$S\$3:\$S\$596) en KI

=BUSCAR(AAi;\$N\$3:\$N\$596;\$T\$3:\$T\$596) en W

=BUSCAR(AAi;\$N\$3:\$N\$596;\$U\$3:\$U\$596) en CT

=BUSCAR(AAi;\$N\$3:\$N\$596;\$V\$3:\$V\$596) en TT

=BUSCAR(AAi;\$N\$3:\$N\$596;\$W\$3:\$W\$596) en DIA

=BUSCAR(AAi;\$N\$3:\$N\$596;\$X\$3:\$X\$596) en MES

=BUSCAR(AAi;\$N\$3:\$N\$596;\$Y\$3:\$Y\$596) en AÑO

donde $i = 1, \dots, 596$

ANEXO 3. Sintaxis utilizada en SPSS.

GET DATA /TYPE=XLSX

/FILE='C:\Users\dmiranda\Documents\Muestra Completa.xlsx'

/SHEET=name 'Hoja1'

/CELLRANGE=full

/READNAMES=on

/ASSUMEDSTRWIDTH=32767.

DATASET NAME Conjunto_de_datos1 WINDOW=FRONT.

COMPUTE ODDS=EXP(-2.48+0.0947*KI-0.0241*W).

EXECUTE.

COMPUTE Latente_P=-1.439+0.0556*KI-0.01472*W.

EXECUTE.

CORRELATIONS

/VARIABLES=ODDS Latente_P

/PRINT=TWOTAIL NOSIG

/MISSING=PAIRWISE.

CORRELATIONS

/VARIABLES=P01 Td850 CAPE CINH KI W CT TT

/PRINT=TWOTAIL NOSIG

/MISSING=PAIRWISE.

* Estimación curvilínea.

TSET NEWVAR=NONE.

CURVEFIT

/VARIABLES=ODDS WITH Latente_P

/CONSTANT

/MODEL=EXPONENTIAL

/PRINT ANOVA

/PLOT FIT.

GRAPH

/SCATTERPLOT(BIVAR)=Latente_P WITH ODDS BY P01

/MISSING=LISTWISE.

ANEXO 4. Resultados en minitab.

Método

Función de enlace **Logit**
Filas utilizadas 594

Selección de términos escalonada

α a entrar = 0,15. α a retirar = 0,15

Información de respuesta

Variable	Valor	Conteo	
P01	1	286	(Evento)
	0	308	
	Total	594	

Resumen del modelo

R-cuad. de desviación	R-cuad. (ajust) de desviación	AIC
9,20%	8,95%	753,00

Coefficientes

Término	Coef	EE del coef.	VIF
Constante	-2,48	1,09	
KI	0,0947	0,0278	2,49
W	-0,0241	0,0139	2,49

Ecuación de regresión

$$P(1) = \exp(Y') / (1 + \exp(Y'))$$

$$Y' = -2,48 + 0,0947 \text{ KI} - 0,0241 \text{ W}$$

Pruebas de bondad del ajuste

Prueba	GL	Chi-cuadrada	Valor p
Desviación	591	747,00	0,000
Pearson	591	606,19	0,324
Hosmer-Lemeshow	8	17,13	0,029

Tabla de desviaciones

Fuente	GL	Desv. ajust.	Media ajust.	Chi-cuadrada	Valor p
Regresión	2	75,647	37,823	75,65	0,000
KI	1	12,435	12,435	12,44	0,000
W	1	3,040	3,040	3,04	0,081
Error	591	746,997	1,264		
Total	593	822,644			

Método

Función de enlace **Normit**
 Filas utilizadas 594

Selección de términos escalonada

α a entrar = 0,15. α a retirar = 0,15

La selección escalonada se detuvo porque el paso 2 y el paso 6 incluyen términos idénticos.

Información de respuesta

Variable	Valor	Conteo	
P01	1	286	(Evento)
	0	308	
	Total	594	

Resumen del modelo

R-cuad. de desviación	R-cuad. (ajust) de desviación	AIC
9,09%	8,84%	753,90

Coefficientes

Término	Coef	EE del coef.	VIF
Constante	-1,439	0,645	
KI	0,0556	0,0163	2,70
W	-0,01472	0,00833	2,70

Ecuación de regresión

$$P(1) = \Phi(Y')$$

$$Y' = -1,439 + 0,0556 \text{ KI} - 0,01472 \text{ W}$$

Φ = CDF de la distribución normal estándar

Pruebas de bondad del ajuste

Prueba	GL	Chi-cuadrada	Valor p
Desviación	591	747,90	0,000
Pearson	591	606,98	0,316
Hosmer-Lemeshow	8	16,85	0,032

Tabla de desviaciones

Fuente	GL	Desv. ajust.	Media ajust.	Chi-cuadrada	Valor p
Regresión	2	74,747	37,373	74,75	0,000
KI	1	12,097	12,097	12,10	0,001
W	1	3,134	3,134	3,13	0,077
Error	591	747,897	1,265		
Total	593	822,644			