COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA

ESTADÍSTICA

UNA PRUEBA DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN LOGÍSTICA

VICENTE FERNÁNDEZ GUERRERO

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE:

M A E S T R O EN C I E N C I A S

MONTECILLO, TEXCOCO, EDO. DE MÉXICO
2007

La presente tesis titulada: **UNA PRUEBA DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN LOGÍSTICA**, realizada por el alumno: **Vicente Fernández Guerrero**, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA ESTADÍSTICA

CONSEJO PARTICULAR

CONSEJERO _____
Dr. José A. Villaseñor Alva

ASESOR _____
Dr. Humberto Vaquera Huerta

ASESOR _____
Dr. Bulmaro Juárez Hernández

AGRADECIMIENTOS

A **DIOS** por que gracias al él inicié y terminé esta etapa de vida, porque siempre está a mi lado sin esperar nada a cambio, por su misericordia infinita, por apoyarme, por comprenderme, por caer y levantarse conmigo, simplemente por ser mi Dios pero sobre todo por ser mi mejor Amigo.

Al **Consejo Nacional de Ciencia y Tecnología (CONACYT)** por su gran apoyo económico que me brindó durante todos mis estudios de maestría.

Al **Colegio de Postgraduados**, por haberme brindado la oportunidad de seguir mi formación académica en sus aulas.

A los integrantes de mi Consejo Particular:

Al **Dr. José A. Villaseñor Alva** mi mas sincero agradecimiento, por su ayuda, apoyo, atinadas indicaciones y consejos ya que sin estos no hubiese sido posible la realización y conclusión de este trabajo de tesis

Al **Dr. Humberto Vaquera Huerta** y al **Dr. Bulmaro Juárez Hernández** por su valiosa participación, sus comentarios acertados, sus consejos y disponibilidad para dedicar parte de su tiempo en la revisión de este trabajo de tesis.

A la Lic. **Irán Guadalupe Pérez Velásquez** por su compañía, apoyo, comprensión, por estar siempre conmigo en las buenas y en las malas, y por enseñarme que juntos podemos llegar muy lejos.

Mi agradecimiento a los profesores, mis compañeros de clase, personal administrativo y todas aquellas personas que de alguna manera fueron coparticipes de esta tarea, a todos gracias, de manera especial a mi compañero M.C. Paulino Pérez Rodríguez.

DEDICATORIA

Dedico de manera muy especial este trabajo, al nombre sobre todo nombre, al Señor de Señores, al que fue, es y será, él verdadero autor de todo este trabajo y de mi Vida por supuesto, a Dios.

¡No a nosotros, Señor, nos des la gloria, no a nosotros,
si no a tu nombre llevado por tu amor, tu lealtad!
Sal 115:1

A mi madre Santa María de Guadalupe, por consolarme en mis momentos de angustia y por recordarme que el Señor Jesús siempre esta conmigo

A mis Padres Gregoria y Gil, que sin ser los mejores, son mi fuerza, mi apoyo y mi mayor orgullo, a ustedes dedico todos mis triunfos y derrotas.

A Irán porque siempre estuviste conmigo, apoyándome, consolándome, aconsejándome y porque gracias a todo eso, este trabajo esta terminado. Me siento muy privilegiado por Dios porque el hecho de que estemos juntos es una gran bendición, pero mas importante es el estar trabajando juntos y ser coparticipes uno del otro. Tú sabes que me siento muy orgulloso de ti, de lo que eres y de todo lo que, con muy poco, haz hecho, tú eres y serás siempre mi apoyo, mi fuerza y mi motivo para seguir adelante, gracias por estar a mi lado, por compartir este momento, por nunca negarme nada, por animarme en mis momentos de angustia pero sobre todo por ser la mujer que amo.¡Que Dios te Bendiga siempre, Flaquita!

A mis hermanos porque siempre han estado conmigo y porque cuando se los pido nunca me han negado su apoyo.

INDICE

1. INTRODUCCIÓN. 1

2. OBJETIVOS....3

3. MARCO TEORICO 4

3.1 DEFINICIÓN....4
3.2 MOMENTOS Y FUNCIÓN GENERADORA DE MOMENTOS 7
3.3 PRUEBAS DE BONDAD DE AJUSTE. 13

4. PRUEBA DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN LOGÍSTICA USANDO EL COEFICIENTE DE CORRELACIÓN 15

4.1 PLANTEAMIENTO DEL PROBLEMA....15
4.2 CONSTRUCCIÓN DE LA PRUEBA R 16
4.3 DISTRIBUCION DE R_c 17
4.4 TAMAÑO DE LA PRUEBA R....21

5. ESTUDIO DE LA POTENCIA DE LA PRUEBA R CONTRA ALGUNAS ALTERNATIVAS 25

5.1 POTENCIA DE LA PRUEBA R 25
5.1.1 PRUEBA EDF USANDO LA ESTADISTICA DE KUIPER....26
5.1.2 PRUEBA BASADA SOBRE LA REGRESIÓN Y CORRELACIÓN. 27
5.2 POTENCIA DE R CON OTRAS ALTERNATIVAS 33

6. EJEMPLOS DE APLICACIÓN DE LA PRUEBA R. 39

6.1 TIEMPOS DE VIDA 39
6.2 TIEMPO DE SUPERVIVENCIA EN RATAS. 42

7. CONCLUSIONES Y RECOMENDACIONES....45

8. BIBLIOGRAFIA. 47

ANEXO A....50

ANEXO B. 51

RESUMEN.

En el presente trabajo se propone una prueba de bondad de ajuste para la distribución logística, la cual está basada en el coeficiente de correlación muestral (R). Esta prueba R, a diferencia de otras conocidas, tiene la propiedad de ser invariante bajo transformaciones de localidad y escala, esto es, la prueba R no depende de los parámetros desconocidos. Se estudiará el tamaño de la prueba propuesta, así como la potencia; en ambos casos se hará una estimación vía simulación de Monte-Carlo. Los valores críticos de la prueba se obtienen utilizando simulación de Monte-Carlo para varios tamaños de muestras y niveles de significancia. Para el estudio de potencia de la prueba se realizara una comparación de potencias considerando diferentes alternativas, junto con otras dos pruebas, la primera es la prueba basada en regresión y correlación y la otra es la prueba basada en la Función de Distribución Empírica usando la estadística de Kuiper, para varios tamaños de muestras y niveles de significancia.

Palabras Clave: Prueba de Bondad de Ajuste, Función de Distribución Empírica, Coeficiente de correlación muestral, Simulación de Monte-Carlo.

ABSTRACT.

In this work, a goodness of fit tests for logistic distribution is proposed, which is based on the sample coefficient of correlation (R). This test R, contrary to other known tests, has the property of invariability under transformation of location and scale, this is, the test R doesn't depend on the parameters. The size of the proposed test was studied, as well as the power; in both cases it was carried out an estimate via Monte-Carlo simulation. The critical values of the test are obtained using Mount-Carlo simulation for several samples sizes and significance levels. For the study of power of the test, it was carried out a comparison of powers considering different alternatives, together with other two tests. The first one is the test based on regression and correlation and the other one is the test based on the Empirical Distribution using the statistic of Kuiper, for several samples sizes and significance levels.

Words Key: Goodness of fit stets, Function of Empiric Distribution, Coefficient of correlation sample, Simulation Monte-Carlo.

UNA PRUEBA DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN LOGÍSTICA

1. INTRODUCCIÓN.

Generalmente en cada uno de los problemas de estimación y pruebas de hipótesis se supone que las observaciones disponibles para el estadístico provienen de distribuciones cuya forma es conocida. En otras palabras, se ha supuesto que las observaciones provienen de cierta familia paramétrica de distribuciones y que se debe hacer una inferencia estadística acerca de los valores de los parámetros que definen dicha familia.

En el caso en que las observaciones disponibles para el estadístico provienen de distribuciones cuya forma exacta no es conocida, se necesitan técnicas estadísticas las cuales sean aplicables sin tener en cuanta la forma de la densidad. Estas técnicas se les conocen como técnicas no paramétricas. Como ejemplo se podría suponer simplemente que las observaciones constituyen una muestra aleatoria de una distribución continua, sin especificar la forma de esta distribución con mayor detalle y entonces investigar la posibilidad de que se trate de una distribución dada $F(x)$.

En el presente trabajo se discute de manera formal sobre de las pruebas de bondad de ajuste para la distribución logística cuya función de Distribución Acumulada esta dada por

$F(x)=\frac{1}{1+e^{-(x-\alpha)/\beta}}$ si $-\infty < x < \infty$, $-\infty < \alpha < \infty$ y $\beta > 0$.

Se investigará, en todos los casos, el siguiente juego de hipótesis.

Ho: La muestra aleatoria proviene de la distribución logística; *vs*

Ha: La muestra aleatoria no proviene de la distribución logística.

También se generan tablas de valores críticos para diferentes tamaños de muestra y valores de significancia, y la potencia de la prueba propuesta es comparada con la de otros procedimientos para mostrar su viabilidad.

Las prueba R que se estudia en este trabajo es una prueba no paramétrica, ya que, para su implementación, no es necesario estimar ningún parámetro, debido a las propiedades de la familia logística y las de la estadística R de ser invariante bajo transformaciones de localidad y escala.

2. OBJETIVOS.

- Proponer una prueba de bondad de ajuste para la función de distribución logística, utilizando el coeficiente de correlación muestral (R), que posea buena potencia.

- Realizar un estudio de comparación de potencia, por medio de simulación, de esta estadística de prueba contra otras pruebas conocidas.

3. MARCO TEORICO

3.1 DEFINICIÓN

La función de distribución logística es usada por primera vez como una herramienta para estudios demográficos por Verhulst (1838, 1845). Otros investigadores también usaron la función logística para estimar el crecimiento de las poblaciones humanas, por ejemplo Pearl and Reed (1920) y más recientemente Oliver (1982), al igual que en modelos de crecimiento en el área de biología nos lo da Pearl and Reed (1924), While Plackett (1959) uso la función logística en el análisis de datos de supervivencia. La función de crecimiento logístico fue usada por Shultz (1930) and Oliver (1964) para un modelo de datos de producción agrícola. Dos aplicaciones interesantes del modelo logístico, entre muchos otros, fueron dados por Dyke y Patterson (1952) y Grizzle (1961) en el área de la salud pública.

Esta distribución se define de manera más simple en términos de su función de distribución acumulada:

$$F_X(x) = \left[1 + \exp\left\{-(x-\alpha)/\beta\right\}\right]^{-1} \text{ con } -\infty \le x \le \infty,\ \beta > 0 \qquad (3\text{-}1)$$

Se puede ver que (3-1) define una función de distribución acumulada ya que es monótona creciente y continúa, además:

$$\lim_{x\to-\infty} F_X(x) = 0; \qquad \lim_{x\to+\infty} F_X(x) = 1$$

La correspondiente función de densidad es:

$$f_X(x) = \frac{\exp\left[-(x-\alpha)/\beta\right]}{\beta\left[1+\exp\left\{-(x-\alpha)/\beta\right\}\right]^2} \text{ ; con } \begin{array}{l} -\infty \le x \le \infty, \\ -\infty \le \alpha \le \infty, \text{ y } \beta > 0 \end{array} \qquad (3\text{-}2)$$

La función en (3-1) ha sido extensivamente usada para representar funciones de crecimientos (con *x* representando el tiempo). El uso de la función logística como una curva de crecimiento puede estar basada en la siguiente ecuación diferencial:

$$\frac{dF}{dx} = c\left[F(x) - A\right]\left[B - F(x)\right] \tag{3-3}$$

Donde *c, A* y *B* son constantes con $c > 0$, $B > A$, en palabras (3-3) puede ser interpretada como: tasa de crecimiento = [exceso inicial (asintótico) sobre el valor de A] × [deficiencia comparada (asintótica) con el valor final de B].

La solución de (3-3) nos conduce a

$$F(x) = \frac{BDe^{x/c} + A}{De^{x/c} + 1} \tag{3-4}$$

Donde D es una constante. Cuando $x \to -\infty, F(x) \to A$; Cuando $x \to \infty, F(x) \to B$ (si $D \neq 0$). La función *F(x)* representa el crecimiento de un limite inferior A a un limite superior de B. Una forma para hacer que *F(x)* sea una función de distribución apropiada, es escribir A=0, B=1; entonces la ecuación (3-4) llega a tener la siguiente forma

$$F(x) = \frac{De^{x/c}}{1 + De^{x/c}} = \left[1 + D^{-1}e^{-x/c}\right]^{-1} \tag{3-5}$$

La cual es la forma (3-1) con $c=\beta$; $D = e^{-\alpha/\beta}$

La ecuación (3-3) ha sido usada como un modelo de auto catálisis. Este es el nombre aplicado a una reacción química en al cual un catálisis M transforma a un componente G en dos componentes J y K y J mismo actúa como un catalizador para la misma reacción.

La distribución logística surge en una manera puramente estadística como el límite de la distribución del rango medio estandarizado (cuando $n \to \infty$) (definido como el promedio entre los valores más grande y más pequeño de la muestra) de una muestra aleatoria de tamaño *n*. Este resultado fue dado por Gumbel (1944). Gumbel y Keeney (1950) demostraron que la distribución logística es obtenida como la distribución límite de multiplicación apropiada de los "cocientes extremos", esto es (valor más grande)/(valor más pequeño).

Talacko (1956) demostró que la distribución logística es la distribución límite (cuando $r \to \infty$ de la variable estandarizada correspondiente a $\sum_{j=1}^{r} j^{-1} X_j$ donde las $X_j's$ son variables aleatorias independientes cada una teniendo una distribución de valores extremos tipo I.

Dubey (1969) ha demostrado que la distribución logística puede ser obtenida como una mezcla de distribuciones de valores extremos

$$P\left[X \le x|\eta\right] = 1 - \exp\left[-\eta\beta \exp\left\{-(x-\alpha)/\beta\right\}\right] \qquad (x > 0; \eta, \beta > 0)$$

Siendo η una variable aleatoria con distribución exponencial y función de densidad dado por

$$f_\eta(y) = \beta e^{-\beta y}, \qquad y > 0$$

Luego

$$P\left[X \le x\right] = 1 - \beta \int_0^\infty \exp\left[-\beta y \left\{1 + \exp(-(x-\alpha)/\beta)\right\}\right] dy$$
$$= \left[1 + \exp\left\{-(x-\alpha)/\beta\right\}\right]^{-1}$$

La cual es idéntica a (3-1).

3.2 MOMENTOS Y FUNCIÓN GENERADORA DE MOMENTOS

Haciendo la transformación $Y = (X - \alpha)/\beta$ cuando X tiene distribución logística con parámetros α y β, obtenemos, de (3-2), la función de densidad de *Y*:

$$f_Y(y) = e^{-y}(1 + e^{-y})^{-2} \tag{3-6}$$

La función de distribución acumulada de *Y* es:

$$F_Y(y) = (1 + e^{-y})^{-1} \tag{3-7}$$

Las ecuaciones (3-6) y (3-7) son las formas estándar de la distribución logística aunque no son solamente las únicas formas estándar ya la distribución logística puede también expresarse en términos de la media y desviación estándar.

La función generatriz de momentos de la variable aleatoria *Y* con función de densidad (3-6) es:

$$\begin{aligned} M_Y(\theta) = E(e^{\theta Y}) &= \int_{-\infty}^{\infty} e^{-(1-\theta)y}(1 + e^{-y})^{-2}\,dy \\ &= \int_0^1 \xi^{-\theta}(1 - \xi)^{\theta}\,d\xi \qquad \text{con } \xi = (\mathrm{e}^{y} + 1)^{-1} \\ &= \mathrm{B}(1\text{-}\theta, 1 + \theta) \\ &= \pi\theta \csc \pi\theta \end{aligned} \tag{3-8}$$

La función característica de *Y*, $E(e^{i\theta Y})$, está dada entonces por $i\pi\theta$ csch $\pi\theta$.

Los momentos de *Y* pueden ser determinados de (3-8) o directamente integrando (3-6). Usando el segundo método (con $r > 0$) tenemos:

$$
\begin{aligned}
E(|Y|^r) &= 2\int_0^\infty y^r e^{-y}(1+e^{-y})^{-2}\,dy \\
&= 2\int_0^\infty y^r \sum_{j=1}^{\infty}(-1)^{j-1} j\, e^{-jy}\,dy \\
&= 2\Gamma(r+1)\sum_{j=1}^{\infty}(-1)^{j-1} j^{-r} \qquad (\text{para } r>0) \\
&= 2\Gamma(r+1)(1-2^{-(r-1)})\zeta(r) \quad (\text{para } r>1)
\end{aligned}
\tag{3.9}
$$

Donde $\zeta(r) = \sum_{j=1}^{\infty} j^{-r}$ es la función zeta de Riemann.

Regresando a la forma original de la distribución (3-1), y recordando que $X = \alpha + \beta Y$, podemos ver que

$$
E(X) = \alpha \qquad\qquad Var(X) = \frac{\beta^2\pi^2}{3} \tag{3-10}
$$

La función de distribución logística acumulada expresada en términos de $E(X) = \mu,\ Var(X) = \sigma^2$, es:

$$
F(x;\mu,\sigma) = \left[1+\exp\left\{-\pi(x-\mu)/\sigma\sqrt{3}\right\}\right]^{-1} \qquad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0 \tag{3-11}
$$

Y la función de densidad es:

$$
f(x;\mu,\sigma) = \frac{\pi}{\sigma\sqrt{3}}\left[\frac{e^{-\pi(x-\mu)/\sigma\sqrt{3}}}{1+e^{-\pi(x-\mu)/\sigma\sqrt{3}}}\right]^2 \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0 \tag{3-12}
$$

La función de densidad logística estándar esta dada por:

$$
f(y) = \frac{\pi}{\sqrt{3}}\frac{e^{-\pi y/\sqrt{3}}}{\left(1+e^{-\pi y/\sqrt{3}}\right)^2} \qquad -\infty < y < \infty \tag{3-13}
$$

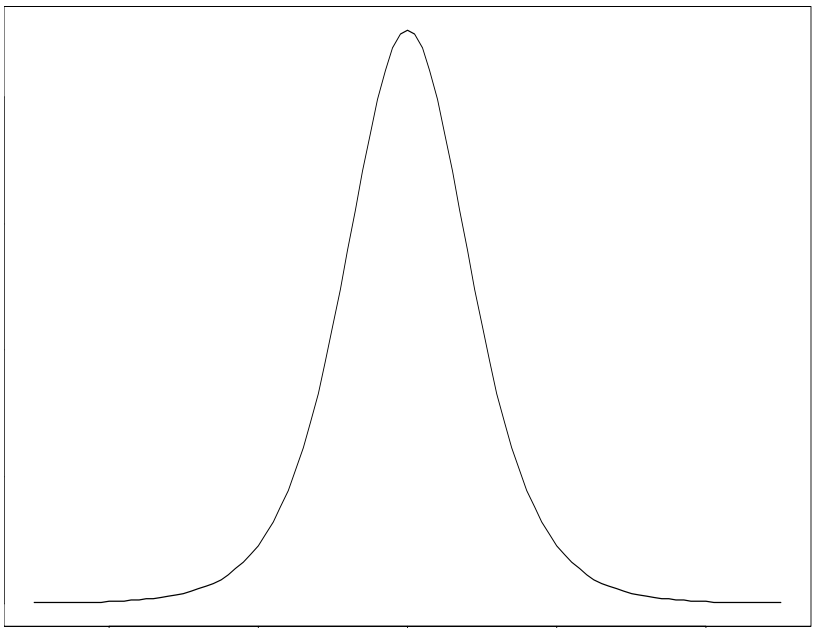
Con la variable $Y = (X - \mu)/\sigma$, y su correspondiente función de distribución acumulada:

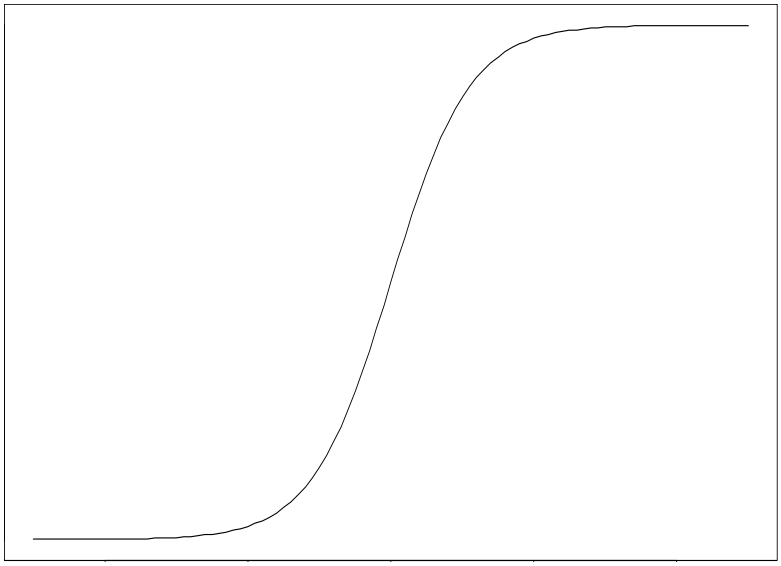
$$F(y) = \frac{1}{1 + e^{-\pi y/\sqrt{3}}} \qquad -\infty < y < \infty \tag{3-14}$$

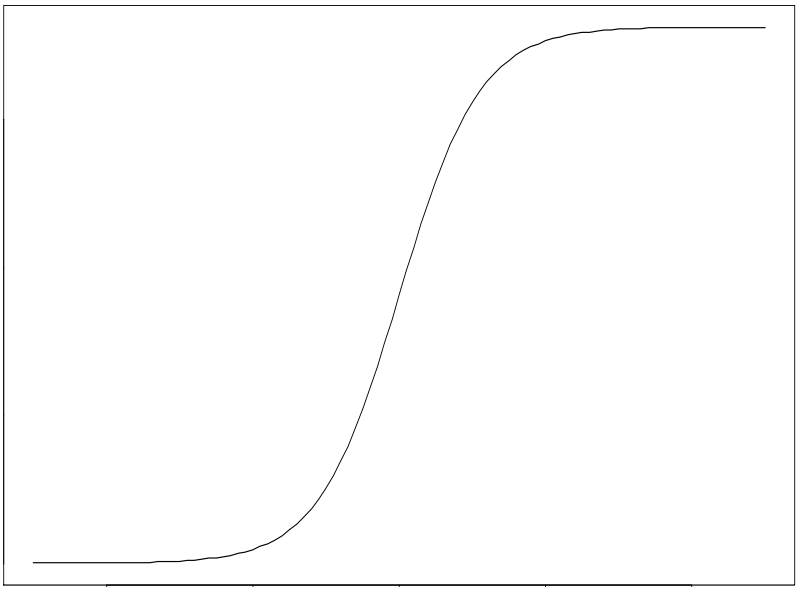
De (3-13) y (3-14) se deriva la función razón de riesgo de la logística estándar L(0,1) con variable *Y* como:

$$h(y) = \frac{f(y)}{1 - F(y)} = \frac{\pi}{\sqrt{3}\left(1 + e^{-\pi y/\sqrt{3}}\right)} \tag{3-15}$$

Las siguientes Figuras muestran las graficas de (3-13), (3-14) y (3-15), respectivamente



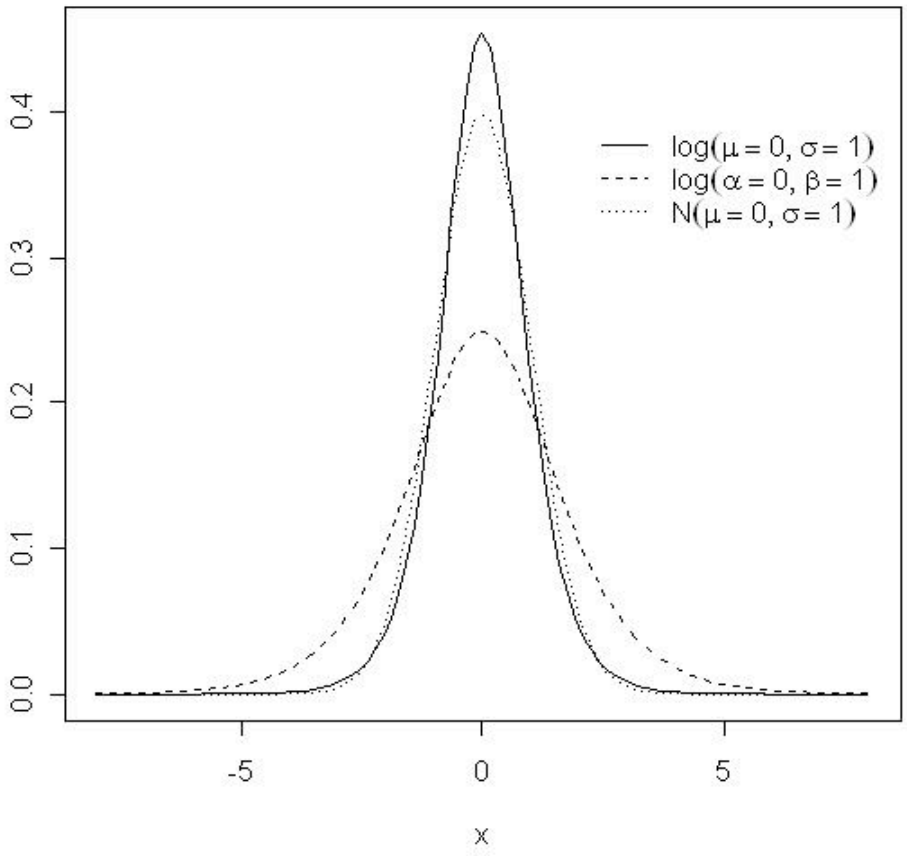




La Figura 1 muestra que la función de densidad logística es simétrica respecto a 0 también es notorio que la función razón de riesgo (Figura 3) es proporcional a la función de distribución acumulada (Figura 2).

La Figura 4 muestra que la función de densidad de la distribución logística estándar (3-13) es simétrica respecto a cero y es mas puntiaguda que la densidad normal estándar pero menos puntiaguda que la de la logística estándar usando (3-2).

Figura 4



Nótese también que las distribuciones logística estándar (3-13) y la normal estándar son muy parecidas. El hecho, de que la distribución logística tenga una forma parecida a la distribución normal, la hace aprovechable, en ocasiones conveniente, ya que el reemplazo de la normal por la logística simplifica el análisis de gran manera, sin crear grandes discrepancias en la teoría estadística. Tal sustitución debe de hacerse con cuidado y entendiendo las similitudes entre estas dos distribuciones.

Si las funciones de distribución acumuladas $G_1(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} du$ y $G_2(x) = \left[1 + \exp\left(-\pi x / \sqrt{3}\right)\right]^{-1}$ de la normal estándar y la logística estándar se comparan (ver Jonson y Kotz p. 6), las diferencias de $G_2(x) - G_1(x)$ varían de la forma en que se muestra en la Figura 5. (Ya que ambas, $G_1(x)$ y $G_2(x)$ son simétricas alrededor de $x = 0$ solo se muestran los valores para $x \geq 0$). Se puede ver que el valor máximo de $G_2(x) - G_1(x)$ esta alrededor de 0.0228 y este se alcanza cuando $x = 7$, este máximo puede ser reducido a un valor más pequeño que 0.01 cambiando la escala de x en $G_1(x)$ usando $G_1(16x/15)$ como una aproximación a $G_2(x)$. Esto, también, se muestra gráficamente en la Figura 5.

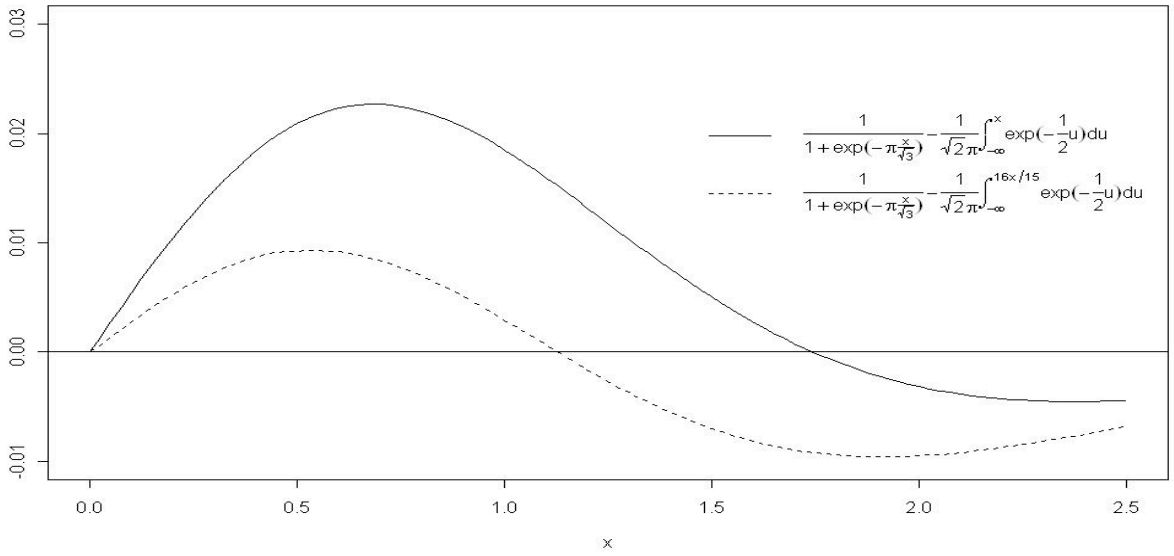


Figura 5. Comparación de la función de distribución logística y la función de distribución normal.

Usando β = 4.2, Mudholkar y George (1978) observaron que la distribución logística se aproxima a una distribución t-student con 9 grados de libertad. Un razonamiento similar ha sido dado por George y Ojo (1980) y George et al. (1986) usando la función de distribución logística generalizada aproximándose a la t-student con *v* grados de libertad.

3.3 PRUEBAS DE BONDAD DE AJUSTE.

Las pruebas estadísticas que verifican si un conjunto de datos se puede ajustar por una distribución determinada son las conocidas como pruebas de bondad de ajuste, estas pruebas miden, como el nombre lo indica, el grado de ajuste que existe entre la distribución obtenida a partir de la muestra y la distribución teórica que se supone debe seguir esa muestra.

Sí la distribución básica es conocida, se pueden realizar fácilmente las pruebas de hipótesis e intervalos de confianza basados en ésa distribución, pero en muchos casos cuando se realiza un experimento no se conoce la forma básica de la distribución y necesitamos técnicas estadísticas que sean aplicables sin tener en cuanta la forma de la densidad. Estas técnicas son conocidas como técnicas no paramétricas o métodos de distribución libre. La prueba basada en el coeficiente de correlación muestras R aquí estudiado es una técnica estadística no paramétrica ya que no depende de los parámetro por sus propiedades de invariabilidad bajo transformaciones de localidad y escala.

En todos los casos se quiere contrastar el juego de hipótesis donde la hipótesis nula afirma que la distribución desconocida es logística con fd dada por (3-2), y la hipótesis alternativa que la distribución desconocida no es logística.

Se tienen dos casos, cuando la hipótesis nula es simple (cuando los parámetros están completamente especificados) y el caso más realista donde la hipótesis nula es compuesta y los parámetros son desconocidos. En todos los casos se tiene como hipótesis alternativa que la distribución desconocida no es logística.

Aun cuando hay varias pruebas de bondad de ajuste, el hecho es que se ha prestado una atención limitada a la distribución logística. M.A. Stephens (1979) presenta un prueba basada en la Función de Distribución Empírica, pero la presentación más completa a la fecha esta dada por D'Agostino y Stephens (1986). Ellos muestran cuatro pruebas las cuales son:

- La prueba chi-cuadrada.
- La prueba basada en la función de distribución empírica (EDF tests).
- Prueba basada en espacios normalizados (para datos censurados)
- Y la prueba basada sobre regresión y correlación.

Sheng (2002) presento una metodología general para desarrollar pruebas de bondad de ajuste con la distribución libre asintótica basada en la divergencia de Kullback-Leiber. El procedimiento puede ser aplicado a una gama amplia de distribuciones entre las que se destacan las distribuciones siguientes: Normal, log-normal, logística, Gumbel y exponencial entre otras.

4. PRUEBA DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN LOGÍSTICA USANDO EL COEFICIENTE DE CORRELACIÓN

4.1 PLANTEAMIENTO DEL PROBLEMA

Supóngase que se tiene una muestra aleatoria $\{X\}_{i=1}^{n}$ de la distribución *F*, con función de densidad $f_X(x;\cdot)$ con soporte en $\Re$ y media finita. Se desea probar el siguiente juego de hipótesis:

$$H_0 : f(x;\cdot) = f_X(x;\alpha,\beta) = \frac{\exp\left[-(x-\alpha)/\beta\right]}{\beta\left[1+\exp\left\{-(x-\alpha)/\beta\right\}\right]^2} I_{(-\infty,\infty)}(x), \alpha \in \Re, \beta > 0 \qquad (4\text{-}1)$$

Con la hipótesis alternativa:

$$H_1 : f(x;\cdot) \neq f_X(x;\alpha,\beta) \qquad (4\text{-}2)$$

Hacemos notar que la hipótesis nula es compuesta, ya que los parámetros son desconocidos, la prueba rechaza cuando R calculada (R_c) es menor que R tablas (R_α), esto es:

Rechazar H_0 si $R_c < R_\alpha$ donde $\alpha = P(R_c < R_\alpha | H_0)$ la cual es una prueba de tamaño α.

4.2 CONSTRUCCIÓN DE LA PRUEBA R

Para la discriminación entre H_0 y H_1 se propone el coeficiente de correlación que se obtiene de la relación entre la muestra $x_1, x_2, ..., x_n$ y los valores $y_i, i = 1,...,n$ que se generan al evaluar en la función de distribución logística inversa los valores correspondientes de la función de distribución empírica evaluados en la muestra, por lo tanto se mide entonces el grado de asociación lineal existente entre la distribución de la muestra y la distribución logística, esto es, se espera un grado de asociación casi perfecto o sea cercano a uno. Por lo tanto la prueba rechaza la hipótesis nula cuando $R_c < R_\alpha$ donde R_c es el coeficiente de correlación y R_α es el R de tablas.

La prueba se construye de la siguiente manera:

1. Dada una realización de una muestra aleatoria de tamaño *n*.
2. Se ordenan en forma creciente las observaciones de la muestra aleatoria $x_1 < x_2 < ... < x_n$.
3. Se obtiene la función de distribución empírica (FDE). Recuérdese que la FDE esta dada por: $\hat{F_n}(x) = \frac{\#(x_j \le x)}{n}$ para una realización $x_1, x_2, ..., x_n$ de la muestra aleatoria $X_1, X_2, ..., X_n$.
4. Se evalúa la FDE en la muestra aleatoria antes obtenida, esto es se calculan $\hat{F_n}(x_1), \hat{F_n}(x_2), ..., \hat{F_n}(x_n)$.
5. Se calculan las $y_i, i = 1,...,n$ sustituyendo los $\hat{F_n}(x_i), i = 1,...,n$ en la función de distribución logística inversa esto es: $y_i = G_0^{-1}\left(\hat{F_n}(x_i)\right)$ de esta forma se obtienen las parejas (x_i, y_i), donde $G_0^{-1}(\cdot)$ es al función de distribución logística inversa en su forma estándar.
6. Por último se calcula el coeficiente de correlación muestral de la siguiente manera:

$$R_c = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})\right]^{1/2}} \qquad (4\text{-}3)$$

Recuerde que se rechaza la hipótesis nula si $R_c < R_\alpha$ para algún valor critico de R_α.

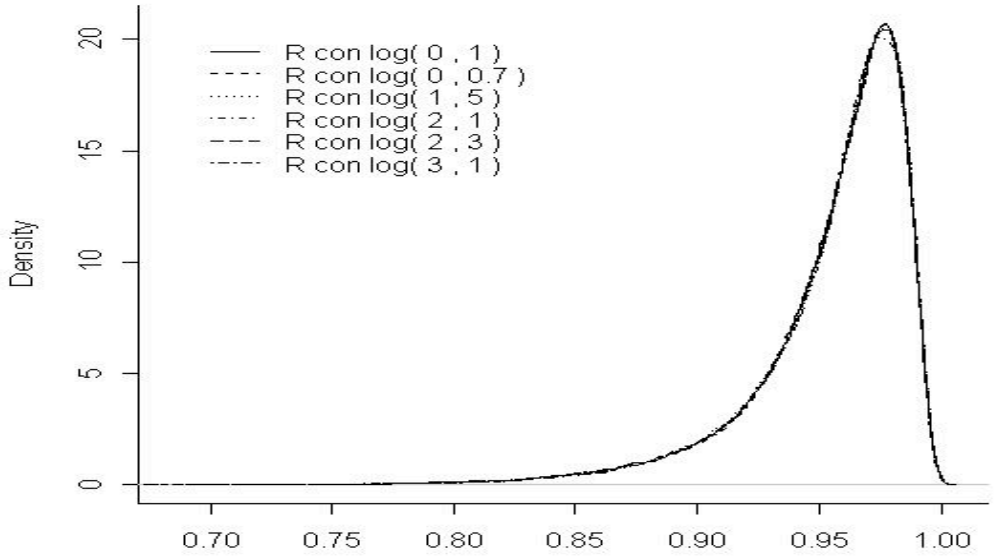
La prueba es independiente de los parámetros desconocidos ya que R_c, es invariante bajo transformaciones de localidad y escala.

4.3 DISTRIBUCION DE R_c

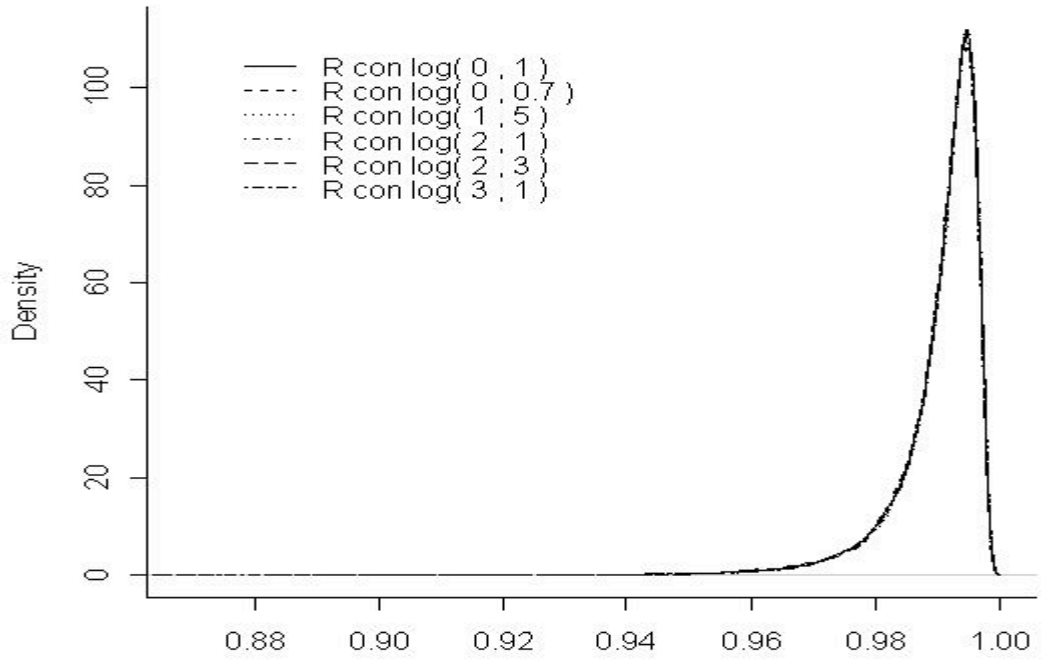
Como se pudo constatar el cálculo de R_c es relativamente fácil sin embargo no es posible obtener la distribución de R_c en forma analítica.

Para obtener la distribución empírica de R_c dada en (4-3) se realiza el siguiente algoritmo: Se generan *B* muestras aleatorias de tamaño *n* de la distribución logística (α,β) dado por (3-2) en la sección 3 y se calcula el valor de R_c esto es se obtienen *B* valores de R_c, de esta forma la distribución de R_c puede ser obtenida por la distribución empírica basada en las *B* realizaciones de R_c para *B* grande, los valores de α y β empíricos puedes ser $\alpha = 0$ y $\beta = 1$

A continuación se presenta la distribución de R_c generados por el programa 1 del anexo B para diferentes parámetros de la distribución logística con tamaño de muestra n = 10.



De igual forma se presenta la distribución de R_c con tamaño de muestra n = 100.



Las distribuciones empíricas fueron generadas con 50,000 réplicas con muestras de tamaño n=10 y n=100, para los parámetros de localidad y escala anotados en las leyendas de las gráficas. En las gráficas anteriores podemos notar que la distribución de R_c es muy parecida sin importar con que parámetros estemos trabajando, esto es importante ya que se puede concluir que la prueba de R_c no depende de los parámetros.

Para probar el juego de hipótesis planteado en (4-1) y (4-2) de este apartado, se emplea la siguiente regla de decisión:

Rechazar H_0 al nivel de significancia α si y solo si $R_c < R_\alpha$

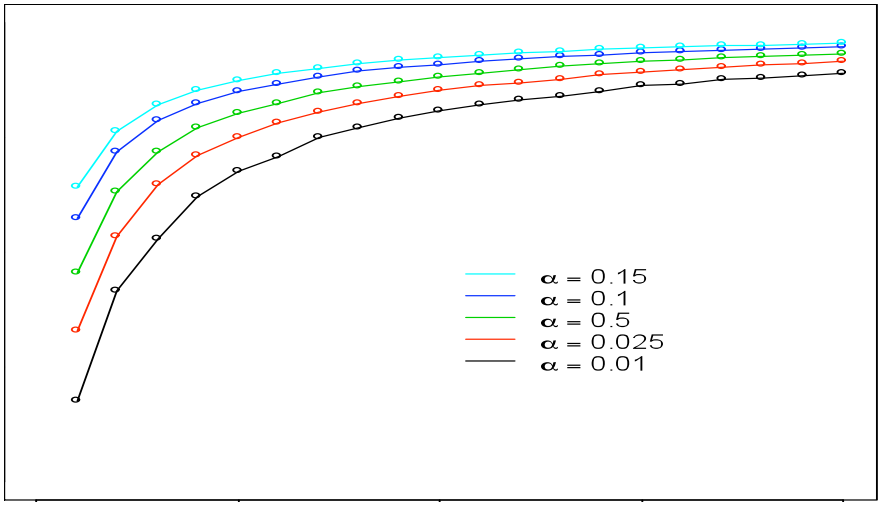
Donde el valor de R_α es tal que $\alpha = P(R_c < R_\alpha | H_0)$.

Para determinar las constantes criticas R_α de R_c, se utiliza el método de simulación empleado anteriormente para determinar la distribución empírica de R_c para niveles de significancia α = 0.01, .025, .05, .1, .15. A continuación mostramos la Tabla 1 de valores críticos de R con diferentes tamaños de muestra con 50,000 replicaciones de la distribución logística estándar. Se determinó el valor de la constante crítica R_α con el cuantil $\alpha \times 100$ de la distribución empírica de R_c, la rutina utilizada se muestra en el programa 2 del anexo B.

TABLA 1 : Valores Críticos de R_c con 50,000 replicas

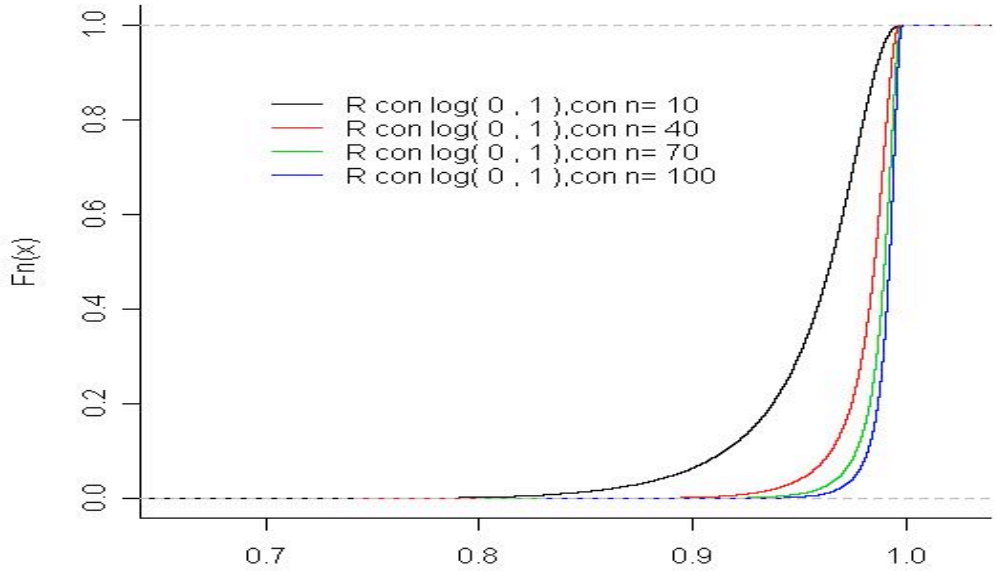
n	α				
	0.01	0.025	0.05	0.1	0.15
10	0.8352025	0.8655955	0.8911834	0.9149682	0.9283949
20	0.8832116	0.9070799	0.9261183	0.9437596	0.9527713
30	0.9057324	0.9295366	0.9440186	0.9572502	0.9642346
40	0.9243758	0.9421874	0.9541792	0.9650083	0.9707127
50	0.9351762	0.9502589	0.9605729	0.9699539	0.974965
60	0.9418916	0.9562304	0.9651204	0.9735313	0.977874
70	0.9502623	0.961226	0.9695987	0.976617	0.9804373
80	0.9541591	0.96512	0.97234	0.9789285	0.982367
90	0.9583242	0.9679767	0.9745987	0.9805384	0.9836931
100	0.961695	0.970496	0.9765398	0.9819801	0.9848631
110	0.9641202	0.9727852	0.978281	0.9832852	0.9859823
120	0.9667213	0.9739941	0.9798027	0.9845777	0.9870257
130	0.9679743	0.9757067	0.9810243	0.985481	0.9877774
140	0.9701451	0.9773414	0.9820686	0.9861958	0.9884257
150	0.9727607	0.9787826	0.9831769	0.987076	0.989093
160	0.9734928	0.9798924	0.9840169	0.987674	0.9895414
170	0.9752361	0.9807236	0.9847924	0.9882368	0.9900266
180	0.976207	0.9815637	0.9853254	0.9886908	0.9904661
190	0.9770459	0.9824545	0.9861902	0.989255	0.990898
200	0.9779043	0.9831332	0.9865273	0.9895665	0.9912121

La Gráfica 1 presenta la misma información



Gráfica 1. Valores Críticos de R_c

La Gráfica 2 muestra que conforme el tamaño de muestra aumenta la distribución empírica de R_c tiende a concentrarse en la unidad, esto implica que aumenta el valor crítico R_α, esta situación se puede ver en la tabla 1 de valores críticos al igual que en la grafica 1 de valores críticos. El programa 3 del anexo B muestra la rutina en R-2.4.1 para generar ésta gráfica.



Grafica 2. Distribución Empírica de R_c con diferentes tamaños de muestra

Por lo que si se quiere saber el valor crítico al 5% del estadístico R con un tamaño de muestra n = 150 se tiene el siguiente valor:

$$R_{n,\alpha} = R_{150\ ,\ 0.05} = 0.9831769$$

4.4 TAMAÑO DE LA PRUEBA R

En un estudio de pruebas de bondad de ajuste, es muy importante que la prueba preserve su tamaño ya que de esta forma se estará asegurando que en realidad se esta haciendo una prueba de tamaño α. Para mostrar esto se estima la probabilidad de rechazar la hipótesis nula cuando las muestras pertenecen a la distribución logística y se procede de la siguiente manera: Se genera una m.a. con distribución logística con diferentes valores en los parámetros de localidad y escala, considerando varios tamaños de muestra, esta se somete a la prueba R para tamaños de prueba α = 0.01, .05 *y* .1 el proceso se realiza con 10,000 replicas auxiliándonos del programa 4 del anexo B mediante el siguiente algoritmo.

a) Se elige el tamaño de prueba α y el tamaño de muestra n.
b) Se genera una muestra aleatoria de la distribución logística (α, β)
c) Se calcula el estadístico de prueba R_c.
d) Si $R_c < R_\alpha$ se rechaza la hipótesis nula, en cuyo caso, se agrega un número 1 a un vector que le llamaremos rechazos, en caso contrario se agrega 0.
e) Se repiten los pasos a) – d) 10,000 veces.
f) Se suma el número de rechazos y se divide entre el numero de replicas.

En otras palabras, se esta está estimando $\alpha = P(R_c < R_\alpha | H_0)$ con $\hat{\alpha} = \dfrac{No.\ de\ rechazos}{No.\ de\ replicas}$, por lo tanto se espera, por ejemplo, que para un tamaño de prueba α = 0.01 haya a lo más 100 rechazos, si α = 0.05 a lo más debe haber 500 rechazos para α = 0.1 a lo más 1000 rechazos.

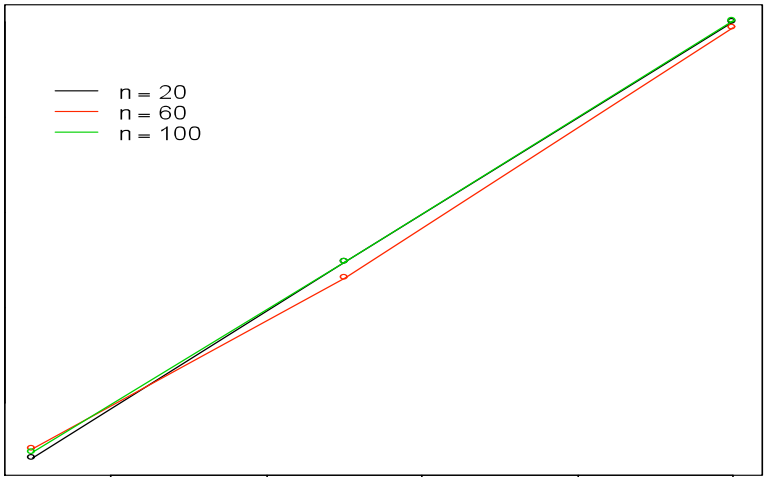
Los resultados que se obtienen son la probabilidad de cometer el error tipo I, es decir rechazar la hipótesis nula dado que es verdadera.

Los resultados obtenidos para el estudio del tamaño estimado de la prueba se muestran en la Tabla 2:

Tabla 2. Estudio del tamaño de prueba

Tamaño de muestra	α		
n=20	0.01	0.05	0.1
Logis(0, 1)	0.0085	0.0497	0.0999
Logis(1, 2)	0.0104	0.0494	0.0996
Logis(1, 3)	0.0102	0.0482	0.0979
Logis(2, 1)	0.0098	0.0498	0.0996
Logis(2, 3)	0.0105	0.0487	0.0982
Logis(3, 1)	0.0095	0.0486	0.0998
Logis(3, 2)	0.0096	0.0428	0.0967
Tamaño de muestra	α		
n=60	0.01	0.05	0.1
Logis(0, 1)	0.0102	0.0463	0.0989
Logis(1, 2)	0.01	0.0457	0.1004
Logis(1, 3)	0.0093	0.0479	0.0981
Logis(2, 1)	0.0099	0.0463	0.094
Logis(2, 3)	0.0092	0.0495	0.0988
Logis(3, 1)	0.0088	0.048	0.0976
Logis(3, 2)	0.0098	0.0498	0.097
Tamaño de muestra	α		
n=100	0.01	0.05	0.1
Logis(0, 1)	0.0095	0.0496	0.1002
Logis(1, 2)	0.0099	0.0468	0.0892
Logis(1, 3)	0.0101	0.0449	0.0967
Logis(2, 1)	0.0088	0.0464	0.0932
Logis(2, 3)	0.0098	0.0503	0.1002
Logis(3, 1)	0.0107	0.0496	0.1003
Logis(3, 2)	0.0087	0.0463	0.0975

La Grafica 3 muestra el tamaño estimado de la prueba solo para el caso de la logística estándar.



Grafica 3. Tamaño estimado de la prueba R.

Como se puede observar en este estudio el tamaño de la prueba R se preserva y esto es una cualidad muy importante de hecho más importante que su potencia, ya que siempre es mas conveniente trabajar con una prueba que preserva su tamaño a una que no lo hace, ya que la que no respeta su tamaño no esta arrojando resultados creíbles.

La Tabla 3. muestra un comparación sobre el tamaño de las pruebas de R, EK y Regresión, para el caso de las pruebas EK y regresión se procede de maneta similar que con lo realizado para estimar el tamaño de R. Los programas 5 y 6 en el anexo B muestran el procedimiento seguido.

Tabla 3. Tabla comparativa sobre el tamaño de prueba

Tamaño de muestra	α								
n=20	0.01			0.05			0.1		
	R	regresión	kuiper	R	regresión	kuiper	R	regresión	kuiper
Logis(0, 1)	0.0085	0.0109	0.0251	0.0497	0.0481	0.0637	0.0999	0.0997	0.1085
Logis(1, 2)	0.0104	0.0107	0.0207	0.0494	0.0489	0.0602	0.0996	0.0997	0.1111
Logis(1, 3)	0.0102	0.0101	0.0116	0.0482	0.0519	0.0532	0.0979	0.0951	0.0951
Logis(2, 1)	0.0098	0.0115	0.0274	0.0498	0.0497	0.0671	0.0996	0.0968	0.1138
Logis(2, 3)	0.0105	0.0101	0.011	0.0487	0.0503	0.0497	0.0982	0.0952	0.0981
Logis(3, 1)	0.0095	0.0102	0.0257	0.0486	0.0515	0.0663	0.0998	0.098	0.1131
Logis(3, 2)	0.0096	0.012	0.0186	0.0428	0.0512	0.0682	0.0967	0.101	0.1112

Tamaño de muestra	α								
n=60	0.01			0.05			0.1		
	R	regresion	kuiper	R	regresion	kuiper	R	regresion	kuiper
Logis(0, 1)	0.0102	0.0119	0.013	0.0463	0.0482	0.0545	0.0989	0.0947	0.1025
Logis(1, 2)	0.01	0.0099	0.0143	0.0457	0.0504	0.0546	0.1004	0.0955	0.1035
Logis(1, 3)	0.0093	0.0104	0.011	0.0479	0.0496	0.0557	0.0981	0.0945	0.102
Logis(2, 1)	0.0099	0.0103	0.0125	0.0463	0.0505	0.0535	0.094	0.0961	0.1062
Logis(2, 3)	0.0092	0.0094	0.0119	0.0495	0.0451	0.0565	0.0988	0.0979	0.1042
Logis(3, 1)	0.0088	0.0098	0.0129	0.048	0.0477	0.0539	0.0976	0.0939	0.1036
Logis(3, 2)	0.0098	0.0084	0.016	0.0498	0.0511	0.0536	0.097	0.1004	0.1089

Tamaño de muestra	α								
n=100	0.01			0.05			0.1		
	R	regresion	kuiper	R	regresion	kuiper	R	regresion	kuiper
Logis(0, 1)	0.0095	0.0102	0.0137	0.0496	0.0492	0.0576	0.1002	0.0985	0.1083
Logis(1, 2)	0.0099	0.0097	0.0116	0.0468	0.0496	0.0632	0.0892	0.0951	0.1129
Logis(1, 3)	0.0101	0.0102	0.0132	0.0449	0.0483	0.0614	0.0967	0.0951	0.1094
Logis(2, 1)	0.0088	0.0116	0.0111	0.0464	0.046	0.0612	0.0932	0.0926	0.1114
Logis(2, 3)	0.0098	0.0083	0.0123	0.0503	0.0503	0.0606	0.1002	0.0903	0.1118
Logis(3, 1)	0.0107	0.009	0.0127	0.0496	0.0492	0.0595	0.1003	0.0962	0.1113
Logis(3, 2)	0.0087	0.0092	0.0127	0.0463	0.045	0.057	0.0975	0.0926	0.1097

Como se puede ver la estadística EK no conserva en general su tamaño, con lo que respecta a la prueba basada sobre regresión y correlación en algunas ocasiones la prueba no respeta su tamaño pero en términos generales si lo hace. Además estos resultados dan evidencia que la estadística R no depende de los parámetros ya que la prueba preserva su tamaño cuando los parámetros α y β toman diferentes valores numéricos.

5. ESTUDIO DE LA POTENCIA DE LA PRUEBA R CONTRA ALGUNAS ALTERNATIVAS

5.1 POTENCIA DE LA PRUEBA R

En este apartado se estudia la potencia de la prueba R, la cual solo puede obtenerse cuando la hipótesis alternativa es establecida de antemano. Para poder ilustrar el comportamiento de la prueba R, se establecen diferentes distribuciones en la hipótesis alternativa H_1, dichas distribuciones que se ocuparan son: beta, normal y weibull.

La función de densidad de las distribuciones mencionadas se muestran a continuación:

La función de densidad de beta:

$$f_X(x;\alpha,\beta)=\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)},\qquad 0<x<1,\ \ \alpha>0,\beta>0 \tag{5-1}$$

La función de densidad de normal:

$$f_X(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)\right\},\qquad x\in\Re,\ \ \mu\in\Re,\sigma^2>0 \tag{5-2}$$

La función de densidad de weibull:

$$f_X(x;\alpha,\beta)=\frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}\exp\left\{-\left(\frac{x}{\beta}\right)^{\alpha}\right\},\qquad x\geq 0,\ \ \alpha>0,\beta>0 \tag{5-3}$$

De igual forma se presenta un estudio comparativo de la potencia de la prueba propuesta, la prueba de EDF (Función de Distribución Empírica) usando la estadística de Kuiper (1960) y una prueba basada en regresión y correlación.

5.1.1 PRUEBA EDF USANDO LA ESTADISTICA DE KUIPER.

Para una muestra aleatoria de observaciones $X_1, X_2, ..., X_n$ ordenada.

Primero calculamos las dos estadísticas supremas definidas como:

$$D^+ = \sup_x \left\{ \frac{i}{n} - T_i \right\},$$

$$D^+ = \sup_x \left\{ T_i - \frac{(i-1)}{n} \right\},$$

Donde $T_i = F(x_i; \hat{\mu}, \hat{\sigma})$ y $\hat{\mu}, \hat{\sigma}$ son los estimadores de máxima verosimilitud de la distribución logística definida en (3-1). También se calcula:

$$V = D^+ + D^-$$

La estadística suprema de Kuiper (1960) generalmente es usada para observaciones en círculo pero de igual forma es también usada para pruebas de bondad de ajuste para muestras de la distribución logística dada en (3-2).

Para completar la prueba, se compara la estadística de Kuiper multiplicándola por $\sqrt{n}$, con los valores que se muestran en la Tabla 4, rechazando la H_0 al nivel de significancia α deseado si la estadística excede el valor crítico.

Tabla 4. Valores críticos con varios niveles de significancia, para la estadística de Kuiper.

ESTADÍSTICA	n	NIVEL DE SIGNIFICANCIA α			
		0.10	0.05	.025	.01
$V\sqrt{n}$	5	1.170	1.246	1.299	1.373
	10	1.230	1.311	1.381	1.466
	20	1.260	1.344	1.422	1.514
	50	1.277	1.364	1.448	1.542
	∞	1.289	1.376	1.463	1.560

5.1.2 PRUEBA BASADA SOBRE LA REGRESIÓN Y CORRELACIÓN.

Sea $X_1, X_2, ..., X_n$ una muestra aleatoria, para la cual se desea probar la hipótesis compuesta dada en (4-1). Sea

$$Y_i = \frac{\sqrt{3}}{n} \ln\left(\frac{F_n(X_i)}{1 - F_n(X_i)} \right) \tag{5-4}$$

Una prueba formal se puede hacer usando el método de regresión donde el modelo de regresión es:

$$X_i = \beta_0 + \beta_1 Y_i + \varepsilon_i \tag{5-5}$$

Aquí se supone que $E(\varepsilon_i) = 0$. Mucho del trabajo para este modelo todavía necesita ser desarrollado con respecto a la distribución logística; sin embargo, Stephens en el capitulo 5 de D'Agostino y Stephens (1986), proporciona un procedimiento basado en este método.

Primero definiremos a $F_n(X_i) = \frac{i}{n+1}$, lo sustituimos en la definición (5-4) y realizamos la regresión de X_i sobre Y_i usando mínimos cuadrados ordinarios para estimar β_0 y β_1 en (5-5). Entonces calculamos:

- SSR = Suma de cuadrados de la Regresión $= \sum_{i=1}^{n} \left(\hat{X}_i - \overline{X} \right)^2$.
- SST = Suma de cuadrados del total = $\sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2$.

Aquí $\hat{X}_i$ es el valor predicho de X_i usando los estimadores de la regresión β_0 y β_1 en (5-5). Sea $R^2(X,Y) = SSR/SST$ (Nota, $R^2(X,Y) = r^2(X,Y)$ es el coeficiente de correlación de Pearson entre X_i y Y_i.). Sea $Z(X,Y) = n\left[1 - R^2(X,Y)\right]$. Sí el ajuste a una línea recta es bueno, entonces $R^2(X,Y)$ es grande y cercano a 1. (sí es exacto entonces $R^2(X,Y)$ es 1.) Por lo tanto si $Z(X,Y)$ es

grande, entonces $R^2(X,Y)$ es pequeña y 4-1 debería ser rechazada. Esto es sí $Z(X,Y)$ excede el valor tabular (ver Tabla 5) entonces la hipótesis nula dada en 4-1 es rechazada.

Tabla 5. Valores críticos con varios niveles de significancia, de regresión y correlación.

n	NIVEL DE SIGNIFICANCIA α				
	0.15	0.10	0.05	0.025	.01
10	1.29	1.51	1.93	2.31	2.84
20	1.84	2.19	2.78	3.42	4.20
40	2.46	2.94	3.76	4.64	5.94
60	2.88	3.40	4.38	5.37	6.99
80	3.15	3.74	4.83	5.93	7.73
100	3.35	3.99	5.16	6.34	8.26

Definición 1. La función de potencia de una prueba ϕ se define como una función:

$$\beta_\phi = \Omega \rightarrow [0,1] \text{ tal que}$$

$$\beta_\phi(\underline{\theta}) = P_{\underline{\theta}}\left(\phi(x) = 1\right)$$

$$= P(\text{Rechazar } H_0 \text{ usando } \phi \mid \underline{\theta}) = \text{P(error tipo I).}$$

Donde $\underline{\theta}$ es el vector de parámetros de la distribución.

Además $P(\text{Error tipo II usando } \phi) = 1 - \beta_\phi(\underline{\theta})$.

El juego de hipótesis que se emplea para la simulación es:

$$H_0 = \text{La muestra aleatoria viene del modelo logistico}$$
$$\text{vs.} \qquad (5\text{-}6)$$
$$H_1 = \text{La muestra aleatoria no viene del modelo logistico}$$

El programa 4 del anexo B fue usado para generar 10,000 muestras de tamaño *n* para los parámetros dados de cada distribución y calcular la estadística R y posteriormente decidir si se rechaza o no se rechaza la hipótesis nula planteada en (5-6). La potencia de la prueba se estimó al dividir el número de veces que se rechaza la hipótesis nula entre el número de muestras de tamaño *n* generadas para los parámetros de la distribución dada.

Para el caso de la estadística de Kuiper (EK) y la prueba basada en regresión y correlación se procedió de manera similar, los programas 5 y 6 en el anexo B muestran el algoritmo usado para estimar la potencia de estas pruebas.

Las siguientes tablas muestran los resultados obtenidos para las diferentes distribuciones mostradas en (5-1), (5-2), (5-3).

Tabla 6. Estimación de la potencia mediante simulación para las pruebas R, EK y regresión (re) usando como alternativa la distribución beta.

α	Parámetro (α,β)	*n*														
		10			20			60			80			100		
		pruebaR	*re*	*EK*	*pruebaR*	*re*	*EK*	*pruebaR*	*re*	*EK*	*pruebaR*	*re*	*EK*	*pruebaR*	*re*	*EK*
0.01	(1,1)	0.0011	0.0009	+	0	0	+	0.003	0.0013	+	0.038	0.0143	+	0.1832	0.0823	+
	(2,2)	0.0019	0.0008	+	0	0.0001	+	0	0	+	0.0001	0	+	0.0003	0.0001	+
	(3,3)	0.0029	0.0012	+	0.0002	0	+	0.0001	0	+	0	0	+	0.0001	0	+
	(4,4)	0.0012	0.0018	+	0	0	+	0	0	+	0	0	+	0	0	+
	(6,6)	0.0024	0.0021	+	0.0002	0.0002	+	0	0	+	0	0	+	0	0	+
0.05	(1,1)	0.0162	0.0156	+	0.009	0.0107	+	0.2672	0.1963	+	0.6469	0.5009	+	0.8884	0.7928	+
	(2,2)	0.0151	0.0105	+	0.002	0.0014	+	0.0099	0.0033	+	0.0451	0.0119	+	0.1152	0.0337	+
	(3,3)	0.0153	0.0111	+	0.0042	0.002	+	0.0038	0.0007	+	0.0101	0.0017	+	0.0237	0.0048	+
	(4,4)	0.017	0.012	+	0.0048	0.0026	+	0.0017	0.0003	+	0.0056	0.0008	+	0.0111	0.0014	+
	(6,6)	0.0194	0.0152	+	0.0056	0.0039	+	0.0018	0.0004	+	0.0031	0.0003	+	0.0049	0.0007	+
0.1	(1,1)	0.0504	0.1515	+	0.0562	0.0504	+	0.6226	0.5277	+	**0.9107**	**0.8322**	+	**0.9845**	**0.9659**	+
	(2,2)	0.042	0.0956	+	0.0181	0.0111	+	0.0823	0.0327	+	0.2039	0.0902	+	0.3693	0.1909	+
	(3,3)	0.0428	0.1042	+	0.0195	0.0015	+	0.0322	0.0098	+	0.0721	0.0203	+	0.1335	0.0439	+
	(4,4)	0.0444	0.1032	+	0.0219	0.0121	+	0.0252	0.0062	+	0.041	0.0105	+	0.0735	0.0174	+
	(6,6)	0.0507	0.1096	+	0.0235	0.0132	+	0.0167	0.004	+	0.029	0.0044	+	0.0398	0.0094	+

Aquí el símbolo + indica que la estadística no arrojo resultado ya que no se cumplen los condiciones en los estimadores de los parámetros (α,β) de máxima verosimilitud en el que β>0 o simplemente la función de verosimilitud no converge. Como puede notarse la estadística R posee una buena potencia con respecto a las otras pruebas de hecho con la estadística EK no se tiene resultados. También puede notarse que la potencia va aumentando conforme aumenta el tamaño de la muestra.

Tabla 7. Estimación de la potencia mediante simulación en R para las pruebas R, EK y regresión (re) usando como alternativa la distribución Normal.

α	Parámetros (μ,σ)	*n*														
		10			20			60			80			100		
		pruebaR	*re*	*EK*	*pruebaR*	*re*	*EK*	*pruebaR*	*re*	*EK*	*pruebaR*	*re*	*EK*	*pruebaR*	*re*	*EK*
0.01	(0,1)	0.0036	0.0024	0.0102	0.0009	0.0007	0.0106	0.0002	0	0.0199	0	0	0.0254	0	0	0.0324
	(1,1)	0.0037	0.0034	0.0104	0.0011	0.0008	0.0115	0	0	0.0182	0	0	0.0224	0	0	0.0267
	(2,2)	0.0042	0.0031	0.0151	0.0011	0.0005	0.0161	0.0001	0	0.0186	0.0001	0.0001	0.0236	0.0002	0	0.0276
	(4,4)	0.0032	0.0032	0.3314	0.0005	0.0007	0.0119	0.0001	0	0.0186	0	0	0.0221	0	0	0.0287
	(5,5)	0.004	0.0036	0.2504	0.0013	0.0006	0.011	0.0002	0	0.0154	0.0001	0.0001	0.0213	0	0	0.0301
0.05	(0,1)	0.0282	0.0223	0.0477	0.0131	0.0102	0.0557	0.0039	0.0021	0.081	0.003	0.0007	0.0858	0.0019	0.0002	0.1059
	(1,1)	0.0266	0.0249	0.049	0.0132	0.0099	0.0571	0.0047	0.0008	0.079	0.003	0.0009	0.0896	0.0024	0.001	0.1101
	(2,2)	0.0277	0.0245	0.0533	0.0128	0.0115	0.0622	0.004	0.0016	0.076	0.0037	0.0008	0.0921	0.0031	0.0008	0.1115
	(4,4)	0.0271	0.0232	0.457	0.012	0.0124	0.0573	0.0027	0.0016	0.074	0.0027	0.0014	0.0892	0.0027	0.0003	0.1111
	(5,5)	0.0274	0.022	0.3527	0.0118	0.0107	0.0558	0.0037	0.0026	0.0784	0.0034	0.001	0.0912	0.0024	0.0004	0.1068
0.1	(0,1)	0.0611	0.1323	0.0976	0.0431	0.0292	0.1039	0.0199	0.0083	0.1382	0.0229	0.0047	0.1594	0.0217	0.0063	0.1914
	(1,1)	0.0674	0.1394	0.0967	0.0428	0.0315	0.104	0.0199	0.0085	0.1396	0.0196	0.0058	0.1635	0.0196	0.0053	0.1777
	(2,2)	0.0633	0.1352	0.1035	0.0417	0.0323	0.1053	0.0191	0.0087	0.1413	0.0218	0.0049	0.1594	0.0191	0.0047	0.183
	(4,4)	0.064	0.1307	0.5203	0.0416	0.0322	0.104	0.0201	0.0082	0.1346	0.0193	0.0055	0.1669	0.0214	0.0047	0.1854
	(5,5)	0.0621	0.132	0.4158	0.0407	0.0316	0.1038	0.0206	0.0078	0.135	0.0212	0.0069	0.1605	0.0181	0.0057	0.1883

Aquí podemos notar, que la prueba R posee mejor potencia con respecto a la prueba usando regresión, y aunque la potencia de la prueba de EK es mayor que las otras dos recuerde que la estadística EK no preserva su tamaño y esto es más importante que su potencia ya que con la estadística EK no se esta presentando una prueba de tamaño α.

Tabla 8. Estimación de la potencia mediante simulación en R para las pruebas R, EK y regresión (re) usando como alternativa la distribución Weibull.

α	Parámetros (α,β)	*N*														
		10			20			60			80			100		
		pruebaR	*re*	*EK*	*pruebaR*	*re*	*EK*	*pruebaR*	*re*	*EK*	*pruebaR*	*re*	*EK*	*pruebaR*	*re*	*EK*
0.01	(1,1/ Γ(1+1/1))	0.0316	0.117	0.2205	0.0836	0.2806	0.5248	0.7226	0.8667	0.9946	0.9513	0.9738	0.9997	0.9957	0.9969	1
	(2,1/ Γ(1+1/2))	0.0016	0.007	+	0.0004	0.0071	+	0.0007	0.0057	+	0.0018	0.007	+	0.005	0.0134	+
	(4,1/ Γ (1+1/4))	0.0037	0.0018	+	0.0004	0.0002	+	0.0001	0	+	0	0	+	0.0001	0	+
	(6,1/ Γ (1+1/6))	0.0062	0.0032	+	0.0024	0.0017	+	0.0006	0.0002	+	0.0016	0.0002	+	0.0015	0.0003	+
	(8,1/ Γ (1+1/8))	0.0122	0.0032	+	0.007	0.006	+	0.0086	0.0024	+	0.0113	0.0038	+	0.0169	0.0035	+
0.05	(1,1/ Γ(1+1/1))	0.1332	0.2942	0.4243	0.3255	0.5828	0.752	0.9847	0.9927	0.9995	0.9995	0.9993	1	1	0.9999	1
	(2,1/ Γ(1+1/2))	0.0153	0.0397	+	0.0074	0.041	+	0.0335	0.0828	+	0.0866	0.1258	+	0.1624	0.1949	+
	(4,1/ Γ (1+1/4))	0.0228	0.017	+	0.0101	0.0063	+	0.0033	0.0005	+	0.0043	0.0004	+	0.0056	0.0005	+
	(6,1/ Γ (1+1/6))	0.0438	0.0263	+	0.0388	0.0216	+	0.0362	0.0104	+	0.0521	0.0102	+	0.067	0.0136	+
	(8,1/ Γ (1+1/8))	0.0557	0.0373	+	0.0601	0.0381	+	0.1048	0.0424	+	0.1452	0.0497	+	0.1952	0.0653	+
0.1	(1,1/ Γ(1+1/1))	0.2256	0.5957	0.5361	0.5247	0.7355	0.8494	0.9976	0.999	0.9999	0.9998	1	1	1	1	1
	(2,1/ Γ(1+1/2))	0.0467	0.1855	+	0.0318	0.0957	+	0.1424	0.2116	+	0.281	0.3132	+	0.4298	0.4558	+
	(4,1/ Γ (1+1/4))	0.0622	0.1237	+	0.0359	0.0189	+	0.0225	0.0052	+	0.0321	0.0056	+	0.0398	0.0066	+
	(6,1/ Γ (1+1/6))	0.0904	0.1531	+	0.0911	0.045	+	0.1212	0.0455	+	0.1585	0.0529	+	0.2072	0.0651	+
	(8,1/ Γ (1+1/8))	0.1125	0.1751	+	0.1394	0.0794	+	0.2465	0.1122	+	0.33	0.1485	+	0.403	0.1948	+

Como podemos ver la potencia de R va aumentando conforme crece el parámetro α y cuando el parámetro α es 1 o 2 las potencias son muy similares por lo que podemos decir que con la distribución gamma la estadística R posee buena potencia.

5.2 POTENCIA DE R CON OTRAS ALTERNATIVAS

Ahora presentamos la potencia de la prueba R con algunas otras alternativas.

a) Cuando la alternativa es la distribución Cauchy, donde la fdp de dicha distribución esta dada por:

$$f_X(x;\alpha,\beta)=\frac{1}{\beta\pi\left[1+\left(\frac{x-\alpha}{\beta}\right)^2\right]} \qquad x\in\Re,\alpha\in\Re,\beta>0$$

α es el parámetro de localización y β es el parámetro de escala.

b) Cuando la alternativa es la distribución F, donde la fdp de dicha distribución esta dada por:

$$f_X(x;m,n)=\frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\cdot\frac{x^{\frac{m-2}{2}}}{\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} \qquad x>0$$

con *m* y *n* grados de libertad.

c) Cuando la alternativa es la distribución Gamma, donde la fdp de dicha distribución esta dada por:

$$f_X(x;\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right) \qquad x>0,\alpha>0,\beta>0$$

con α parámetro de forma y β el parámetro de escala.

d) Cuando la alternativa es la distribución t-student, dicha distribución esta dada por:

$$f_X(x;n)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}\left[1+\left(\frac{x^2}{n}\right)\right]^{\frac{n+1}{2}}} \qquad x\in\Re,n=1,2,3,...$$

Tabla 9. Estimación de la potencia mediante simulación de la prueba R, usando como alternativa la distribución Cauchy.

α	Parametros (α,β)	*n*				
		10	20	60	80	100
		pruebaR	*pruebaR*	*pruebaR*	*pruebaR*	*pruebaR*
0.01	(0,1)	0.2902	0.5385	**0.9427**	**0.9801**	**0.9963**
	(1,1)	0.2865	0.5461	**0.9377**	**0.9844**	**0.9963**
	(2,2)	0.2888	0.5521	**0.9396**	**0.9865**	**0.996**
	(4,4)	0.2809	0.547	**0.9424**	**0.9824**	**0.9961**
	(5,5)	0.2887	0.5513	**0.9375**	**0.9829**	**0.9962**
0.05	(0,1)	0.4325	0.6896	**0.9826**	**0.9962**	**0.9988**
	(1,1)	0.4277	0.6939	**0.9816**	**0.9958**	**0.9994**
	(2,2)	0.4236	0.6888	**0.9815**	**0.9962**	**0.9989**
	(4,4)	0.4312	0.6953	**0.9797**	**0.9962**	**0.9995**
	(5,5)	0.4291	0.6895	**0.9805**	**0.9956**	**0.999**
0.1	(0,1)	0.5133	0.7745	**0.9916**	**0.9984**	**0.9998**
	(1,1)	0.5132	0.7608	**0.9893**	**0.9985**	**0.9996**
	(2,2)	0.5072	0.7705	**0.9892**	**0.9983**	**1**
	(4,4)	0.5048	0.7687	**0.9907**	**0.997**	**0.9994**
	(5,5)	0.5082	0.7637	**0.9888**	**0.998**	**0.9996**

Puede notarse que R tiene buena potencia con esta distribución de cola pesada. Además note que con $n = 60$ ya se tiene buena potencia.

Tabla 10. Estimación de la potencia por simulación de la prueba R, usando como alternativa la distribución F.

α	Parametros (a,b)	*n*				
		10	20	60	80	100
		pruebaR	*pruebaR*	*pruebaR*	*pruebaR*	*pruebaR*
0.01	(2,2)	0.2919	0.7597	**1**	**1**	**1**
	(3,3)	0.1571	0.5215	**0.9978**	**1**	**1**
	(5,5)	0.0653	0.2621	**0.9435**	**0.9957**	**0.9998**
	(10,10)	0.0177	0.0691	0.5241	0.7907	**0.9273**
	(20,20)	0.0055	0.0107	0.115	0.2542	0.4063

0.05	(2,2)	0.5076	**0.923**	**1**	**1**	**1**
	(3,3)	0.3507	0.779	**1**	**1**	**1**
	(5,5)	0.1918	0.5266	**0.9968**	**0.9999**	**1**
	(10,10)	0.0812	0.2254	0.8693	**0.9723**	**0.9959**
	(20,20)	0.0414	0.0734	0.4395	0.6623	0.8363
0.1	(2,2)	0.6435	**0.9661**	**1**	**1**	**1**
	(3,3)	0.472	0.8813	**1**	**1**	**1**
	(5,5)	0.3045	0.6876	**0.9995**	**0.9999**	**1**
	(10,10)	0.1578	0.368	**0.9495**	**0.9945**	**0.9996**
	(20,20)	0.0776	0.1597	0.6584	0.8522	**0.9414**

En esta tabla se puede, notar que la potencia aumenta conforme aumenta el tamaño de muestra, en forma general para la distribución F se tiene una buena potencia.

Tabla 11. Estimación de la potencia mediante simulación de la prueba R, usando como alternativa la distribución Gamma.

α	Parametros (a,b)	*n*				
		10	20	60	80	100
		pruebaR	*pruebaR*	*pruebaR*	*pruebaR*	*pruebaR*
0.01	(1/2,2)	0.132	0.4058	**0.9987**	**1**	**1**
	(1,2)	0.0304	0.0866	0.734	**0.9535**	**0.9969**
	(3,2)	0.0048	0.0053	0.0362	0.0983	0.2162
	(4,2)	0.003	0.0022	0.0126	0.0409	0.0806
	(6,2)	0.0022	0.0015	0.0034	0.0088	0.0172
0.05	(1/2,2)	0.3447	0.7676	**1**	**1**	**1**
	(1,2)	0.1334	0.3243	**0.9809**	**0.9997**	**1**
	(3,2)	0.0282	0.0465	0.3006	0.5388	0.7375
	(4,2)	0.025	0.0283	0.1666	0.3195	0.4867
	(6,2)	0.0194	0.0157	0.0684	0.133	0.2211
0.1	(1/2,2)	0.5073	0.8993	**1**	**1**	**1**
	(1,2)	0.2242	0.5269	**0.998**	**1**	**1**
	(3,2)	0.0726	0.1171	0.5421	0.7741	**0.9076**
	(4,2)	0.0605	0.0741	0.3592	0.5821	0.7515
	(6,2)	0.0546	0.052	0.1795	0.3231	0.4627

Como se puede ver la potencia va disminuyendo conforme la distribución se va pareciendo a la logística pero también notemos que la potencia aumenta conforme el tamaño de muestra aumenta.

Tabla 12. Estimación de la potencia por simulación de la prueba R, usando como alternativa la distribución t-student.

α	Parametro (λ)	*N*				
		10	20	60	80	100
		pruebaR	*pruebaR*	*pruebaR*	*pruebaR*	*pruebaR*
0.01	0.1	**0.9721**	**1**	**1**	**1**	**1**
	0.5	0.5991	**0.8967**	**1**	**1**	**1**
	1	0.2851	0.543	**0.9373**	**0.9829**	**0.9955**
	1.5	0.1553	0.3153	0.6952	0.8215	**0.8995**
	2	0.0977	0.1923	0.4677	0.5831	0.6732
0.05	0.1	**0.9918**	**1**	**1**	**1**	**1**
	0.5	0.7285	**0.9494**	**1**	**1**	**1**
	1	0.4281	0.6938	**0.9803**	**0.9958**	**0.999**
	1.5	0.2737	0.4614	0.8452	**0.9206**	**0.9635**
	2	0.1924	0.3366	0.6404	0.7578	**0.8295**
0.1	0.1	**0.9959**	**1**	**1**	**1**	**1**
	0.5	0.7917	**0.9715**	**1**	**1**	**1**
	1	0.5106	0.7697	**0.9912**	**0.9976**	**0.9991**
	1.5	0.3549	0.5584	0.8947	**0.9543**	**0.9789**
	2	0.2679	0.4157	0.7357	0.821	0.8852

En términos generales la prueba R posee buena potencia con la alternativa t-student.

La figura 6 muestra un estudio comparativo de las potencias estimadas mediante simulación Monte-Carlo para algunas alternativas que se consideraron en este trabajo, para varios tamaños de muestra y nivel de significancia $\alpha = 0.05$.

Figura 6. Estudio comparativo.

Potencias estimadas para α = 0.05 contra algunas alternativas con n=10

		Alternative	R	regresion	EDF
Distribuciones Simetricas	Colas Ligeras	logistica(0,0.7)	0.0479	0.0511	0.0529
		Normal Estandar	0.0282	0.0223	0.0477
		t(12)	0.0379	0.0404	0.0524
	Colas Pesadas	t(4)	0.0846	0.1001	0.0849
		Cauchy Estandar	0.4325	0.5519	0.5482
Distribuciones Asimetricas	Colas Ligeras	Gamma(1,1)	0.1206	0.3029	0.4106
		Weibull(2,1/Γ(1+1/2))	0.0153	0.0397	+
	Colas Pesadas	Weibull(0.5,1/Γ(1+1/0.5))	0.576	0.7956	0.9104
		Log-Normal(-0.2, √0.4)	0.0703	0.2485	0.5289
		gumbel(0,1)	0.0256	0.0960	0.0920

Potencias estimadas para $\alpha = 0.05$ contra algunas alternativas con n=20

		Alternative	R	regresión	EDF
Distribuciones Simetricas	Colas Ligeras	logistica(0,0.7)	0.0474	0.0479	0.0551
		Normal Estandar	0.0131	0.0102	0.0557
		t(12)	0.0314	0.0316	0.0511
	Colas Pesadas	t(4)	0.1195	0.1492	0.0991
		Cauchy Estandar	0.6896	0.8177	0.7653
Distribuciones Asimetricas	Colas Ligeras	Gamma(1,1)	0.3308	0.5862	0.7505
		Weibull(2,1/Γ(1+1/2))	0.0074	0.0410	+
	Colas Pesadas	Weibull(0.5,1/Γ(1+1/0.5))	0.952	0.9851	0.9970
		Log-Normal(-0.2, √0.4)	0.1894	0.4632	0.8173
		gumbel(0,1)	0.0336	0.1495	0.1301

Potencias estimadas para $\alpha = 0.05$ contra algunas alternativas con n=60

		Alternative	R	regresión	EDF
Distribuciones Simetricas	Colas Ligeras	logistica(0,0.7)	0.0459	0.0535	0.0567
		Normal Estandar	0.0039	0.0021	0.0810
		t(12)	0.0288	0.0250	0.0557
	Colas Pesadas	t(4)	0.2113	0.2705	0.1165
		Cauchy Estandar	0.9826	0.9928	0.9906
Distribuciones Asimetricas	Colas Ligeras	Gamma(1,1)	0.9811	0.9924	0.9997
		Weibull(2,1/Γ(1+1/2))	0.0335	0.0828	+
	Colas Pesadas	Weibull(0.5,1/Γ(1+1/0.5))	1.0000	1.0000	1.0000
		Log-Normal(-0.2, √0.4)	0.8297	0.9307	0.9991
		gumbel(0,1)	0.1697	0.3826	0.2986

Potencias estimadas para $\alpha = 0.05$ contra algunas alternativas con n=80

		Alternative	R	regresión	EDF
Distribuciones Simetricas	Colas Ligeras	logistica(0,0.7)	0.0462	0.0495	0.0614
		Normal Estandar	0.0030	0.0007	0.0858
		t(12)	0.0252	0.0239	0.0673
	Colas Pesadas	t(4)	0.2519	0.3187	0.1319
		Cauchy Estandar	0.9962	0.9989	0.9971
Distribuciones Asimetricas	Colas Ligeras	Gamma(1,1)	0.9995	0.9996	1.0000
		Weibull(2,1/Γ(1+1/2))	0.0866	0.1258	+
	Colas Pesadas	Weibull(0.5,1/Γ(1+1/0.5))	1.0000	1.0000	1.0000
		Log-Normal(-0.2, √0.4)	0.9658	0.9833	0.5289
		gumbel(0,1)	0.3114	0.5056	0.3937

Potencias estimadas para $\alpha = 0.05$ contra algunas alternativas con n=100

		Alternative	R	regresión	EDF
Distribuciones Simetricas	Colas Ligeras	logistica(0,0.7)	0.0478	0.0514	0.0625
		Normal Estandar	0.0019	0.0002	0.1059
		t(12)	0.0268	0.0216	0.0650
	Colas Pesadas	t(4)	0.2810	0.3568	0.1553
		Cauchy Estandar	0.9988	0.9998	0.9994
Distribuciones Asimetricas	Colas Ligeras	Gamma(1,1)	1.0000	1.0000	1.0000
		Weibull(2,1/Γ(1+1/2))	0.1624	0.1949	+
	Colas Pesadas	Weibull(0.5,1/Γ(1+1/0.5))	1.0000	1.0000	1.0000
		Log-Normal(-0.2, √0.4)	0.9941	0.9975	1.0000
		gumbel(0,1)	0.4655	0.6291	0.4963

Se puede ver de la Figura 6 que la prueba R tal pareciera que no posee buena potencia, pero nótese que la prueba R respeta en todos los casos su tamaño de muestra, mientras que la prueba basada en regresión y correlación lo hace solo en algunos casos y en los casos en donde ésta respeta su tamaño las diferencias entre la potencia de R y regresión no son muy grandes de hecho para un tamaño de muestra $n \geq 60$ las potencias de las tres pruebas son ya muy similares, teniendo en cuenta que la estadística de Kuiper no respeta en ningún caso su tamaño.

De este estudio podemos decir que es muy recomendable usar la estadística R sobre las otras dos ya que, como se dijo en la sección 4.4, es mucho mas importante que una estadística de prueba respete su tamaño que su propia potencia, porque de esta manera se esta asegurando que la prueba es en verdad de de tamaño α.

6. EJEMPLOS DE APLICACIÓN DE LA PRUEBA R.

En esta sección se mostraran algunos ejemplos en los que se ilustra los pasos a seguir para aplicar la prueba de bondad de ajuste descrita en la sección 4.2.

6.1 Tiempos de Vida

Davis (1952) muestra datos de tiempos de vida medido en horas de 417 lámparas incandescentes de 40 watts. Enseguida se muestran las primeras 100 observaciones.

lámpara	vida en horas	lámpara	vida en horas	lámpara	vida en horas	lámpara	vida en horas
1	1067	26	938	51	830	76	1250
2	919	27	970	52	1063	77	1203
3	1196	28	1237	53	930	78	1078
4	785	29	956	54	807	79	890
5	1126	30	1102	55	954	80	1303
6	936	31	1022	56	1063	81	1011
7	918	32	978	57	1002	82	1102
8	1156	33	832	58	909	83	996
9	920	34	1009	59	1077	84	780
10	948	35	1157	60	1021	85	900
11	855	36	1151	61	1062	86	1106
12	1092	37	1009	62	1157	87	704
13	1162	38	765	63	999	88	621
14	1170	39	958	64	932	89	854
15	929	40	1311	65	1035	90	1178
16	950	41	1037	66	944	91	1138
17	905	42	702	67	1049	92	951
18	972	43	521	68	940	93	902
19	1035	44	933	69	1122	94	923
20	1045	45	928	70	1115	95	1333
21	1157	46	1153	71	833	96	811
22	1195	47	946	72	1320	97	1217
23	1195	48	858	73	901	98	1085
24	1340	49	1071	74	1324	99	986
25	1122	50	1069	75	818	100	958

Primero veremos si los datos muestran correlación entre ellos, para eso mostramos la función de autocorrelación en la siguiente gráfica.

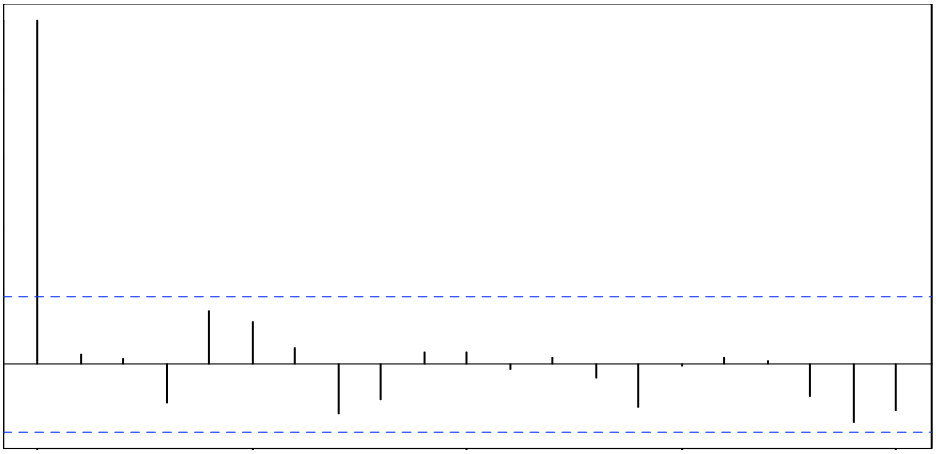


Figura 7. Gráfica de la función de autocorrelación de los datos de tiempo de vida

Nótese en la Figura 7 que los datos no muestran correlación por lo que podemos seguir con nuestro estudio.

Una manera informal de investigar si los datos se ajustan a una distribución logística es usar el método de descrito por Balakrishnan, N. (1991) en el cap. 13.3.

Primero haremos una gráfica conocida como gráfica de probabilidad logística en donde se graficaran

$$x \text{ vs } y = \frac{\sqrt{3}}{\pi}\log\left(\frac{F_n(x)}{1-F_n(x)}\right). \qquad (6\text{-}1)$$

$$\text{donde } F_n(x) = \frac{i-0.5}{n}, \text{ con } i = \#(X_j \le x)$$

D'Agostino en el capitulo 2 de D'Agostino y Stephens(1986), demostraron que (6-1) puede ser usada para muestras de tamaño $n \ge 10$.

En segundo lugar calculas los estimadores de mínimos cuadrados y sustituimos en el siguiente modelo de regresión $x = \widehat{\mu} + \widehat{\sigma} y$.

Si la distribución desconocida es logística entonces la gráfica de probabilidades logística se ajustara a línea dada por el modelo de regresión del segundo paso.

La Figura 8 muestra la grafica de probabilidades logística y la recta $x = \widehat{\mu} + \widehat{\sigma} y$.

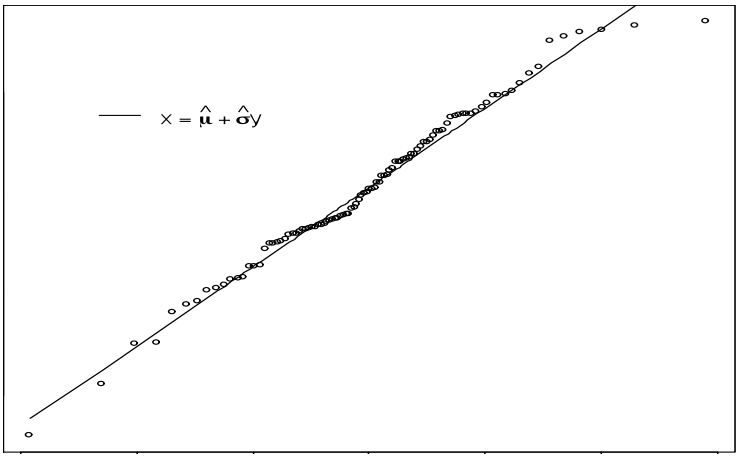


Figura 8. Gráfica de probabilidades logística de los datos de tiempo de vida con la línea recta obtenida usando los estimadores de mínimos cuadrados $\left(\hat{\mu} = 1012.0, \hat{\sigma} = 156.9\right)$.

Se puede notar de la Figura 8 parece que existe una fuerte relación lineal entre *x e y*, a excepción de las últimas observaciones, sin embargo podemos decir que tal pareciera que los datos provienen de una distribución logística, en seguida se prueba si los datos vienen de una distribución logística utilizando la prueba de bondad de ajuste descrita en la sección 4.2 , el juego de hipótesis de interés es el planteado en (4-1) y (4-2).

El tamaño de muestra es $n = 100$, para el nivel de significancia $\alpha = 0.05$, de la tabla 1 se toma $R_{100,0.05} = 0.97653980$, solo resta calcular el valor de R_c obteniéndose $R_c = 0.99356$, como $0.99356 > 0.97653980$ no se rechaza la H_0 y se concluye que los datos se ajustan bien a un modelo logístico.

6.2 Tiempo de supervivencia en ratas.

Los datos siguientes son tiempos de supervivencia en semanas para 20 ratas hembras que fueron sometidas a niveles altos de radiación. Estos datos fueron originalmente dados por Furth et al (1959) y también son discutidos por Lawless (1982) p. 206. Los tiempos son:

rata	tiempo de vida (en semanas)
1	152
2	152
3	115
4	109
5	137
6	88
7	94
8	77
9	160
10	165
11	125
12	40
13	128
14	123
15	136
16	101
17	62
18	153
19	83
20	69

Empezamos por investigar si los datos presentan correlación entre ellos por lo que nos auxiliamos de la gráfica de función de Autocorrelación la cual se muestra en la Figura 9.

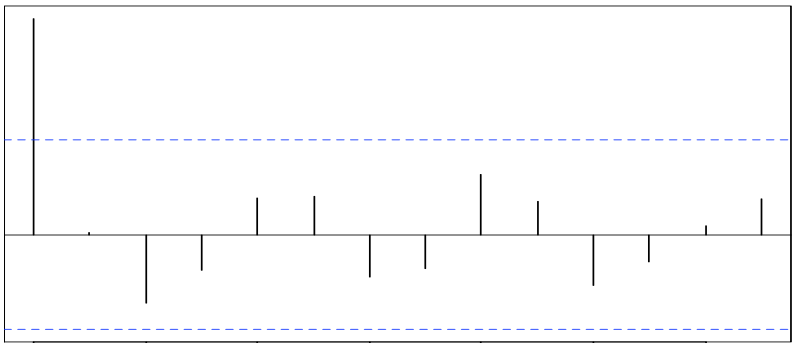


Figura 9. Gráfica de la función de autocorrelación para los datos supervivencia en ratas.

La gráfica de auntocorrelación muestra que no hay correlación en los datos (la mayor esta cerca de .25) por lo que podemos seguir con el análisis.

La Figura 10 nos muestra la gráfica de probabilidades Logística paro los datos de supervivencia en ratas.

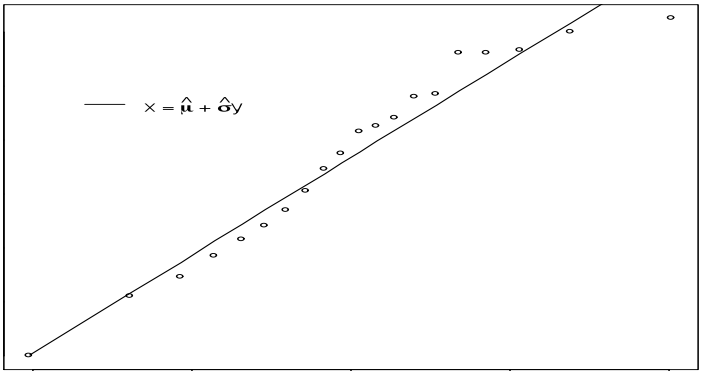


Figura 10. Gráfica de probabilidades logística de los datos de supervivencia en ratas con la línea recta obtenida usando los estimadores de mínimos cuadrados $\left(\hat{\mu} = 113.45, \hat{\sigma} = 36.13\right)$.

Podemos ver de la Figura 10 que los datos parecen ajustarse a la línea recta, esto nos permite suponer que existe una relación lineal entre *x e y* tal pareciera que los datos provienen de una

distribución logística, en seguida se prueba si los datos vienen de una distribución logística utilizando prueba de bondad de ajuste descrita en la sección 4.2 , el juego de hipótesis de interés es el planteado en (4-1) y (4-2).

El tamaño de muestra es $n = 20$, para el nivel de significancia $\alpha = 0.05$, de la tabla 1 se toma $R_{20,0.05} = 0.9261183$, solo resta calcular el valor de R_c obteniéndose $R_c = 0.9866977$, como $0.9866977 > 0.9261183$ no se rechaza la H_0 y se concluye que los datos se ajustan bien a un modelo logístico.

7. CONCLUSIONES Y RECOMENDACIONES.

a. La estadística R es una estadística fácil de construir a diferencia de las otras estadísticas por lo que su cálculo es de mayor rapidez.

b. La potencia de la prueba de la estadística R es superior con respecto a las estadísticas EK y la basada en la regresión y correlación, para las alternativas Beta, Normal y Weibull.

c. Aunque para las otras alternativas estudiadas la prueba R no fue superior, se ve que tiene una buena potencia y es muy similar a las otras estadísticas estudiadas, además recuérdese que las estadísticas de Kuiper no respeta su tamaño al igual que la basada en regresión y correlación.

d. La estadística R siempre respeta su tamaño, esto asegura que la prueba trabaja adecuadamente, rechazando cuando debe de rechazar y no rechazando cuando no debe de rechazar.

e. Ya que la estadística R no depende de parámetros se puede recomendar su uso para el análisis de pruebas de bondad de ajuste, recordemos que no es complicada de implementar como en el caso de la estadística de Kuiper, en la cual, si hay que estimar los parámetros.

f. El Hecho de que la prueba R respete su tamaño es un indicador de que su uso es de mejor opción que las otras pruebas, ya que ésta característica es un mejor indicador que la potencia.

g. Es recomendable, en estudios posteriores, observar el comportamiento de la estadística de Kuiper usando los estimadores de momentos, ya que con los estimadores de máxima verosimilitud en varias ocasiones no se muestra resultado.

h. Es recomendable, en estudios posteriores, observar el comportamiento de la estadística R, para cuando los casos en donde la muestra en estudio presenta correlación, ya que aquí solo se tomo en cuenta muestras independientes.

i. Es también recomendable, para estudios posteriores, observar el comportamiento de la prueba R. para datos censurados ya que en muchos ejercicios de aplicación, como por ejemplo en análisis de supervivencia, los datos muestran algún tipo de censura.

8. BIBLIOGRAFIA.

Balakrishnan, N. (1991), Handbook of the logistic distribution, edit, Dekker, vol 123.

D' Agostino And Stephens (1986) Goodness-of-Fit Techniques, Marcel Dekker, New York.

Davis D.J. (1952), An Análisis of some failure data*, J. Amer. Stat. Assoc*., 47, 133-150.

Dubey, S.D. (1969). A new derivation of the logistic distribution , *Naval Research Logistics Quarterly 16,* 37-40.

Dyke, G. V. and Patterson, H. D. (1952). Analysis of factorial arrangements when the data are proportions. *Biometrics, 8,* 1-12.

Goerge, E.O., El-Saidi, M., and Singh, K. (1986). A generalized logistic approximation of the Student *t* distribution. *Commun. Statist.-Simul. Comput., 15,* 1199-1208.

Govind S. Mudholkar and Olusegun Goerge, (1978) A remark on shape of logistic distribution, *Biometrika,* 65, 3, pp. 667-668.

Grizzle, J.E. (1961) A new method of testing hypotheses and estimating parameters for the logistic model. *Biometrics , 17,* 375-385.

Gumbel, E. J. (1944). Ranges and midranges, *Annals of mathematical Statistics*, 15, 414-422.

Gumbel, E. J. And Keeney, R.D. (1950). The extremal quotient, *Annals of mathematical Statistics*, 21, 523-538.

Hegazy Y. A. S. and Green J. R. (1975), Some New Goodness-of-fit using order statistics, department of computational and statistical, University of Liverpool.

J. Aitchison and S. M. Shen, (1980), Logistic- Normal distribution: son properties and uses, *biometrika,* 67-2, 261-272.

Jhonson, N.L and Kotz, S. (1970). Distribution in Statistics: Continuous Univariate Distributión 2, Vol. 2. Wiley, New York.

Kuiper, N.H.(1960). Tests concerning random points on a circle, *Proc. Koninkl. Neder. Akad. van. Wetenschappen. A, 63,* 38-47.

Lawless, J.F. (1982), Statistical models & Methods for lifetime data. Edit. Wiley-interscience. p.206.

Meade, Nigel (1988), A modified logistic model applied to human populations, , *J. R. Statist. Soc*. A 151, part 3 pp. 491-498.

Mudholkar, G.S. and George, E.O. (1978) A remark on the shape of the logistic distribution. *Biometrics., 65,* 667-668.

Oliver, F.G. (1964). Methods of estimating the logistic growth function. *Appl. Statist., 13,* 57-66.

Oliver, F.G. (1982). Notes on logistic curve for human populations, *J. R. Statist*. Soc. 36, pp. 1103-115.

Pearl, R. and Read, L. J. (1920).On the rate of growth of the population of the United States since 1970 and its mathematical representation. *Proc. Nat. Acad. Sci., 6,* 275-288.

Pearl, R. and Read, L. J. (1924). *Studies in Human Biology.* Williams And Wilkins, Baltimore.

Plackett, R.L. (1959). The analysis of Life test data. *Technometrics, 1,* 9-19.

R Development Core Team (2006). R: A lenguaje and Enviroment for Statistical computing, R Foundation for Statistical Computing, Viena, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org.

Sheng S. K. (2002), Goodness-of-fit Test based on Kullback-Leibler Discrimination information, *IEEE Transactions on Information Teory,* 48, 1103-1117.

Stephens, M. A. (1979), Tests of fit for the logistic distribution based on the empirical distribution function. *Biometrica* vol. 66. No. 3, pp. 591-596.

Talacko, J. (1956) Perks' distributions and their role in the theory of Wiener's stochastic variables, *Trabajos de estadística, 7,* 159-174.

Verhulst, P. J. (1838). Notice sur lois que la population suit dans sons accroissement . *Corr. Math. Pys., 10,* 113-121.

Verhulst, P. J. (1845). Recherches mathématiques sur la loi d'accroissement de la population. *Acad. Bruxelles, 18,* 1-38.

ANEXO A

Propiedades de la distribución Logística.

- Función de Densidad:

$$f_X(x) = \frac{\exp\left[-(x-\alpha)/\beta\right]}{\beta\left[1+\exp\left\{-(x-\alpha)/\beta\right\}\right]^2} \qquad -\infty < x < \infty, -\infty < \alpha < \infty, \beta > 0$$

$$= (4\beta)^{-1}\operatorname{sech}^2\left\{\frac{1}{2}(x-\alpha)/\beta\right\}$$

- Función de Distribución:

$$F_X(x) = \frac{1}{\left[1+\exp\left\{-(x-\alpha)/\beta\right\}\right]}$$

$$= \frac{1}{2}\left[1+\tanh\left\{\frac{1}{2}(x-\alpha)/\beta\right\}\right] \qquad \text{con } \beta > 0$$

- Función Generatriz de momentos:

$$m_X(t) = e^{\alpha t}\pi\beta\csc(\pi\beta t)$$

- Media:

$$E(X) = \alpha$$

- Varianza:

$$Var(X) = \frac{\beta^2\pi^2}{3}$$

ANEXO B.

Programa 1. Rutina en R 2.4.1 para obtener la Distribución empírica de R.

```
#función que calcula la estadistica R.
Rstat<-function(x,n)
{
  x<-sort(x)
  Fn<-ecdf(x);
  y<-qlogis(Fn(x[1:n]),0,1,TRUE)
  return(cor(x[1:n-1],y[1:n-1]))
}
#Función que calcula la Distribución de R,B indica el numero de replicas a
generar.
disR<-function(n=10,B=50000,params)
{
   R<-rep(0,B)
   i<-1
   while(i<=B)
   {
     x<-rlogis(n,params[1],params[2])
     {
       R[i]<-Rstat(x,n)
       i<-i+1
     }
   }
   return(R)
}
grafica<-function(tmuestras,B=1000,params)
{
   cuantas<-length(tmuestras)

plot(density(disR(tmuestras[1],B,c(params[1],params[2]))),main="",xlab="",bty="l"
,lty=1)
   legpos<-locator(1)
   legend(legpos,paste("R con log(",params[1],",",params[2],")"),bty="n",lty=1)
   for(i in 2:cuantas)
   {
     lines(density(disR(tmuestras[i],B,c(params[2*i-
1],params[2*i]))),bty="l",lty=i)
     legend(legpos$x,legpos$y-(i-1)*5,paste("R con log(",params[2*i-
1],",",params[2*i],")"),bty="n",lty=i)
   }
}

grafica(rep(100,6),50000,c(0,1,0,0.7,1,5,2,1,2,3,3,1)).
```

Programa 2. Rutina en R 2.4.1 para determinar las constantes críticas de R.

```
#funcion para clacular la estadística R.
log<-function(x,n)
{
   x <- rlogis(n); xord <- sort(x);x;xord
   Fn <- ecdf(x);
   Fn(xord[1:n])
   y <- qlogis(Fn(xord[1:n]),0,1,TRUE);y
   cor(xord[1:n-1],y[1:n-1])
}
#Función para calcular los puntos críticos para diversos tamaños de muestra(n)
#el parametro B indica el numero de muestras a generar, el parámetro
#alpha es el nivel de significancia de la prueba
Ralpha<-function(n=50,B=50000, alpha=c(0.01,0.05,0.10,0.15))
{
      aleatorios<-rep(0,B)
      i<-1
      while(i<=B)
      {
        aleatoriosr<-rlogis(n,0,1)
        aleatorios[i]<-log(aleatoriosr,n);
        i<-i+1;
      }
      return(quantile(aleatorios,alpha))
}
Ralpha(n=10,B=50000, alpha=c(0.01,0.05,0.10,0.15))
```

Programa 3. Rutina en R 2.4.1 para obtener la Función de Distribución Empírica FDE de R con diferentes tamaños de muestra.

```
#funcion para clacular la estadística R.
Rstat<-function(x,n)
{
  x<-sort(x)
  Fn<-ecdf(x);
  y<-qlogis(Fn(x[1:n]),0,1,TRUE)
  return(cor(x[1:n-1],y[1:n-1]))
}

#Función que calcula la Distribución de R,B indica el numero de replicas a
generar.
disR<-function(n=10,B=1000,params)
{
   R<-rep(0,B)
   i<-1
   while(i<=B)
   {
     x<-rlogis(n,params[1],params[2])
     {
       R[i]<-Rstat(x,n)
       i<-i+1
     }
   }
   return(R)
```

```
}

grafica<-function(tmuestras,B=1000,params)
{
   cuantas<-length(tmuestras)

plot(ecdf(disR(tmuestras[1],B,c(params[1],params[2]))),verticals=T,do.p=F,main=""
,xlab="",bty="l",lty=1,col.h=1,col.v=1)
   legpos<-locator(1)
   legend(legpos,paste("R con log(",params[1],",",params[2],"),con
n=",tmuestras[1],""),bty="n",lty=1,col=1)
   for(i in 2:cuantas)
   {

lines(ecdf(disR(tmuestras[i],B,c(params[1],params[2]))),verticals=T,do.p=F,bty="l
",lty=1,col.h=i,col.v=i)
     legend(legpos$x,legpos$y-(i-1)*.05,paste("R con
log(",params[1],",",params[2],"),con n=",tmuestras[i],""),bty="n",lty=1,col=i)
   }
}

grafica(seq(from = 10, to = 100, by =30),50000,c(0,1))
```

Programa 4. Rutina en R 2.4.1 para calcular el tamaño de la prueba y la potencia de R.

```
#Función para calcular la estadística R
#x es el vector de datos
Rstat<-function(x,n)
{
       x <- sort(x)
       Fn <- ecdf(x)
       y <- qlogis(Fn(x[1:n]),0,1,TRUE)
       cor(x[1:n-1],y[1:n-1])

}
CriticoR<-function(n,alpha)
{
    #La tabla 2 de puntos críticos
    #columna 1(alpha=0.01), columna 2(alpha=0.05), columna 3(alpha=0.10)
    Criticos<-matrix(nrow=5, ncol=3)
    Criticos[1,]<-c(0.8352025,0.8911834,0.9149682) #Tamaño de muestra 10
    Criticos[2,]<-c(0.8832116,0.9261183,0.9437596) #Tamaño de muestra 20
    Criticos[3,]<-c(0.9418916,0.9651204,0.9735313) #Tamaño de muestra 60
    Criticos[4,]<-c(0.9541591,0.9723400,0.9789285) #Tamaño de muestra 80
    Criticos[5,]<-c(0.9616950,0.9765398,0.9819801) #Tamaño de muestra 100
    i<-c(10,20,60,80,100)        #indice de fila
    j<-c(0.01,0.05,0.10)            #indice de columna

    {
      return(Criticos[match(n,i),match(alpha,j)])
    }

}

#Función que indica si se rechaza  o no se rechaza la Hipotesis nula
#Ho:F(x) = Logistica, Ha:F(x)<> Logistica
#Si se rechaza regresa un 1, caso contrario un 0
```

```
#alpha es el nivel de significancia de la prueba
#n, es el tamaño de la muestra

Phix<-function(R,n,alpha)
{
    pcritico<-CriticoR(n,alpha)
    if(R<pcritico) return(1)
    else return(0)
}

PotR<-function(n=80,alpha=0.05,B=1000,alternativa="logistica",params=c(0,1))
{
   rechazos<-rep(0,B)               #Crea un vector para contar el numero de
veces que se rechaza Ho
   i<-1
   while(i<=B)
   {
     x<-switch(alternativa,logistica=rlogis(n,params[1],params[2]),
                           tstudent=rt(n,params[1]),
                           gamma=rgamma(n,shape=params[1],scale=params[2]),
                           F=rf(n,params[1],params[2]),
                           weibull=rweibull(n,params[1],params[2]),
                           cauchy=rcauchy(n,params[1],params[2]),
                           normal=rnorm(n,params[1],params[2]),
                           lognormal=rlnorm(n,params[1],params[2]))
     R<-Rstat(x,n)

     rechazos[i]<-Phix(R,n,alpha)
     i<-i+1
   }
   return(sum(rechazos)/B)  #Regresa la potencia de la prueba o el tamaño según
sea el caso.
}
PotR(10,0.01,10000,alternativa="logistica",params=c(2,1/2))# calcula tamaño
PotR(10,0.01,10000,alternativa="tstudent",params=c(2,1/4))# calcula potencia

Para calcular el tamaño de la prueba solo la leyenda alternativa se cambia el
nombre a las distribuciones dadas para x.
```

Programa 5. Rutina en R 2.4.1 para calcular el tamaño de la prueba de la estadística de Kuiper (EK) y la potencia

```
#Funcion que expresa el log-verosimilitud
fvlogistica<-function(parametros,x)
{
      alfa<-parametros[1]
      beta<-parametros[2]
      n<-length(x)
      y<-1+exp(-(x-alfa)/beta)
      sum((x-alfa)/beta)+n*log(beta)+2*sum(log(y))
}
#función que calcula los estimadores de maxima verosimilitud.
emvlogistica<-function(x)
{
  estimadores<-c(NaN,NaN)
  alfa.i<-mean(x)
  beta.i<-sqrt(var(x))
```

```
  semillas<-c(alfa.i,beta.i)
  emv<-nlminb(semillas,fvlogistica,x=x)
if(emv$convergence==0)
  {
    estimadores<-emv$par
  }
  return(estimadores)
}

#Función para calcular la estadística V
#x es el vector de datos
#máximo verosimiles de los parámetros de localidad y escala.
Vstat<-function(x)
{
    x<-sort(x)                              #Ordena el vector de datos en forma
ascendente
    emv<-emvlogistica(x)
    if(!is.na(emv[1]) && !is.na(emv[2]) &&  emv[2]>0)
    {
      z<-plogis(x,emv[1],emv[2])                  #Evalua la función de
distribucion Gumbel
      Dmas<-max(c(1:length(x))/length(x)-z)        #Calcula D+
      Dmenos<-max(z-c(0:(length(x)-1))/length(x))  #Calcula D-
      return(Dmas+Dmenos)
    }
    else return(-1)
}

CriticoKuiper<-function(n,alpha)
{
    #La tabla 2 de puntos críticos de Chandra
    #columna 1(alpha=0.10), columna 2(alpha=0.05), columna 3(alpha=0.025),
columna 4(alpha=0.01)
    Criticos<-matrix(nrow=4, ncol=4)
    Criticos[1,]<-c(1.230,1.311,1.381,1.466) #Tamaño de muestra 10
    Criticos[2,]<-c(1.260,1.344,1.422,1.514) #Tamaño de muestra 20
    Criticos[3,]<-c(1.277,1.364,1.448,1.542) #Tamaño de muestra 50
    Criticos[4,]<-c(1.289,1.376,1.463,1.560) #Tamaño de muestra infinito
    i<-c(10,20,50,NA)                        #indice de fila
    j<-c(0.10,0.05,0.025,0.01)               #indice de columna
    if(n<=50)
    {
      return(Criticos[match(n,i),match(alpha,j)])
    }
    else
    {
      return(Criticos[4,match(alpha,j)])
    }
}

#Función que indica si se rechaza  o no se rechaza la Hipotesis nula
#Ho:F(x) = Gumbel, Ha:F(x)<> Gumbel
#Si se rechaza regresa un 1, caso contrario un 0
#V es el valor de la estadística V, alpha es el nivel de significancia de la
prueba
#n, es el tamaño de la muestra
Phix<-function(V,n,alpha)
{
    pcritico<-CriticoKuiper(n,alpha)
```

```
    if(sqrt(n)*V>pcritico) return(1)

 else return(0)
}

PotKuiper<-
function(n=150,alpha=0.05,B=10000,alternativa="logistica",params=c(0,1))
{
   rechazos<-rep(0,B)               #Crea un vector para contar el numero de
veces que se rechaza Ho
   i<-1
   while(i<=B)
   {
     x<-switch(alternativa,logistica=rlogis(n,params[1],params[2]),
                           tstudent=rt(n,params[1]),
                           beta=rbeta(n,params[1],params[2]),
                           gamma=rgamma(n,shape=params[1],scale=params[2]),
                           F=rf(n,params[1],params[2]),
                           weibull=rweibull(n,params[1],params[2]),
                           cauchy=rcauchy(n,params[1],params[2]),
                           normal=rnorm(n,params[1],params[2]),
                           lognormal=rlnorm(n,params[1],params[2]))
     V<-Vstat(x)
     if(V!=-1)
     {
       rechazos[i]<-Phix(V,n,alpha)
       i<-i+1
     }
   }
   return(sum(rechazos)/B)  #Regresa la potencia de la prueba o el tamaño
}
PotKuiper(10,0.01,10000,alternativa="logistica",params=c(1/2,2))#calcula el
tamaño
PotKuiper(10,0.01,10000,alternativa="tstudent",params=c(1,2))# calcula la
potencia

Para calcular el tamaño de la prueba solo la leyenda alternativa se cambia el
nombre a las distribuciones dadas para x.
```

Programa 6. Rutina en R 2.4.1 para calcular el tamaño de la prueba y la potencia del método basado sobre regresión y correlación.

```
#Función para calcular la estadística Z
#x es el vector de datos
Zstat<-function(x,n)
{
        x<-sort(x)
        t<-c(1:n)
        F<-t/(n+1)
        Y<-(sqrt(3)/pi)*log(F/(1-F))
        y<-lm(x~Y)
        p<-predict(y)
        SSR<-sum((p-mean(x))^2)
        SST<-sum((x-mean(x))^2)
        RC<-SSR/SST
        return(n*(1-RC))
}
```

```
Criticoregresion<-function(n,alpha)
{
    #La tabla 2 de puntos críticos
    #columna 1(alpha=0.10), columna 2(alpha=0.05), columna 3(alpha=0.01)
    Criticos<-matrix(nrow=5, ncol=3)
    Criticos[1,]<-c(1.51,1.93,2.84) #Tamaño de muestra 10
    Criticos[2,]<-c(2.19,2.78,4.20) #Tamaño de muestra 20
    Criticos[3,]<-c(3.40,4.38,6.99) #Tamaño de muestra 60
    Criticos[4,]<-c(3.74,4.83,7.73) #Tamaño de muestra 80
    Criticos[5,]<-c(3.99,5.16,8.26) #Tamaño de muestra 100
    i<-c(10,20,60,80,100)        #indice de fila
    j<-c(0.10,0.05,0.01)            #indice de columna

    {
      return(Criticos[match(n,i),match(alpha,j)])
    }

}

#Función que indica si se rechaza  o no se rechaza la Hipotesis nula
#Ho:F(x) = Logistica, Ha:F(x)<> Logistica
#Si se rechaza regresa un 1, caso contrario un 0
#alpha es el nivel de significancia de la prueba
#n, es el tamaño de la muestra
Phix<-function(Z,n,alpha)
{
    pcritico<-Criticoregresion(n,alpha)
    if(Z>pcritico) return(1)
    else return(0)
}

Potregresion<-
function(n=100,alpha=0.05,B=1000,alternativa="logistica",params=c(0,1))
{
   rechazos<-rep(0,B)               #Crea un vector para contar el numero de
veces que se rechaza Ho
   i<-1
   while(i<=B)
   {
     x<-switch(alternativa,logistica=rlogis(n,params[1],params[2]),
                           tstudent=rt(n,params[1]),
                           beta=rbeta(n,params[1],params[2]),
                           gamma=rgamma(n,shape=params[1],scale=params[2]),
                           F=rf(n,params[1],params[2]),
                           weibull=rweibull(n,params[1],params[2]),
                           cauchy=rcauchy(n,params[1],params[2]),
                           normal=rnorm(n,params[1],params[2]),
                           lognormal=rlnorm(n,params[1],params[2]))
     Z<-Zstat(x,n)

     rechazos[i]<-Phix(Z,n,alpha)
     i<-i+1
   }
   return(sum(rechazos)/B)  #Regresa la potencia de la prueba o el tamaño segun
sea el caso
}
Potregresion(10,0.01,10000,alternativa="logistica",params=c(3,1/2))#calcula el
tamaño
```

```
Potregresion(10,0.01,10000,alternativa="tstudent",params=c(1,2))#    calcula    la
potencia
```

Para calcular el tamaño de la prueba solo la leyenda alternativa se cambia el nombre a las distribuciones dadas para x.