

2 ejemplares
No. 17



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE CIENCIAS

DETECCION DE OBSERVACIONES ABERRAN-
TES EN MUESTRAS DE UNA DISTRIBUCION
EXPONENCIAL

92

xcl

T E S I S

QUE PARA OBTENER EL TITULO DE
A C T U A R I O
P R E S E N T A:
LELIA MARIA MENDOZA BLANCO

MEXICO, D. F.

10100

1979



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

INDICE

INTRODUCCION

CAPITULO I.

Detección de observaciones aberrantes.

- 1.1 Historia 1
- 1.2 Ejemplos 17

CAPITULO II.

Detección de observaciones aberrantes para el caso exponencial.

- 2.1 La distribución exponencial con 2 parámetros 25
- 2.2 Historia y Propiedades 26
- 2.3 Detección de observaciones aberrantes para el caso exponencial 34

CAPITULO III.

Un enfoque general para la detección de observaciones aberrantes.

- 3.1 Estimadores insesgados de varianza mínima de funciones de distribución 51
- 3.2 Enfoque Propuesto 55
- 3.3 Aplicación al caso exponencial 59

CAPITULO IV.

Resultados.

- 4.1 Comparación de simulaciones 67
- 4.2 Resultados 82

- CONCLUSIONES 85

APENDICE.

Apéndice de Tablas 87

BIBLIOGRAFIA.

INTRODUCCION

INTRODUCCION

El tema de observaciones aberrantes ha interesado a los estadísticos desde hace bastante tiempo. Es necesario estudiar estas observaciones, que intuitivamente se dice que están "alejadas" o "distantes" del resto de las demás, ya que la presencia de ellas pueden alterar las conclusiones o los resultados encontrados. Varios problemas son de interés en este tema, en particular, la detección y el tratamiento de estas observaciones. En esta tesis se trata el problema de detección de una observación aberrante para el caso exponencial.

En el primer capítulo se presenta un bosquejo histórico de observaciones aberrantes y una introducción al problema de detección de las mismas.

En el segundo capítulo se revisan los diferentes criterios reportados en la literatura para la detección de aberrantes en muestras de una variable aleatoria exponencial.

En el tercer capítulo se discute un procedimiento propuesto por el Dr. O'Reilly para el problema de detección de una observación aberrante. Este enfoque está basado en la estimación insesgada de funciones de distribución por lo que en la primera sección se trata este tema. En la segunda sección se presenta el enfoque y en la tercera se aplica al caso exponencial.

Para el caso exponencial no se han encontrado reportes en la literatura de un estudio de las potencias de los diferentes criterios propuestos, por lo que en el cuarto capítulo se reportan y discuten los resultados obtenidos en las simulaciones realizadas para comparar las diferentes pruebas.

CAPITULO I

1. DETECCION DE OBSERVACIONES ABERRANTES

1.1. Historia

El estudio de observaciones aberrantes¹ data de 100 años atrás aproximadamente. Aún cuando no existía el desarrollo de técnicas estadísticas, ya había un interés hacia aquel tipo de datos que se decía no representaban adecuadamente a la población. Este tema ha sido de controversia entre muchos autores y ha tenido importancia en algunas ciencias como Astronomía, Geodesia, Química y Física.

Ferguson (1961)² establece lo que es una observación aberrante de la siguiente manera: "Cuando se trabaja con muestras extraídas de una cierta población, en ocasiones aparecen observaciones que no siguen el patrón común de las demás, es decir, se encuentran alejadas del resto del grupo".

Barnett y Lewis (1978) definen una observación aberrante en un conjunto de datos como "una observación o un subconjunto de ellas, que son inconsistentes con el resto del conjunto". La palabra inconsistente es a juicio subjetivo del experimentador.

1 En inglés se les llama outliers, spurious, mavericks, & rogues.

2 De Barnett y Lewis (1978)

Por ejemplo, los siguientes datos extraídos de un estudio de probabilidad de bajas temperaturas en los meses de invierno en las Islas Británicas fueron analizados por Barnett y Lewis. Estos datos consisten en una serie de temperaturas tomadas por hora en grados Fahrenheit desde la tarde del 31 de Diciembre de 1960 y temprano en la mañana del 1° de Enero de 1961 en Wick al norte de Escocia.

<u>HORA</u>	<u>TEMPERATURA</u>
5:00 P.M.	43°F
6:00 "	43°F
7:00 "	41°F
8:00 "	41°F
9:00 "	41°F
10:00 "	42°F
11:00 "	43°F
12:00 "	58°F
1:00 A.M.	58°F
2:00 "	41°F
3:00 "	41°F

Inicialmente se argumentó que los valores 58 y 58 para media noche y 1 a.m. no eran consistentes con los restantes ya que para este horario la temperatura era muy alta en Wick, sin embargo, mediante un nuevo análisis se encontró que en la media noche la Oficina Meteorológica cambió la escala de °F a 1/10 °C, por lo que los valores eran razonables y aparecieron como:

<u>HORA</u>	<u>TEMPERATURA</u>
5:00 P.M.	43
6:00 "	43
7:00 "	41
8:00 "	41
9:00 "	41
10:00 "	42
11:00 "	43
12:00 "	42
1:00 A.M.	42
2:00 "	39
3:00 "	39

Autores como Anscombe (1960) y Grubbs (1969) clasifican estas observaciones de acuerdo a sus causas de aparición en la muestra.

1. Variabilidad inherente

Este tipo de variabilidad es intrínseca y natural de los datos por lo cual no es controlable.

2. Errores de medición

Este tipo de error se comete mediante el uso de instrumentos de medición.

3. Errores de ejecución

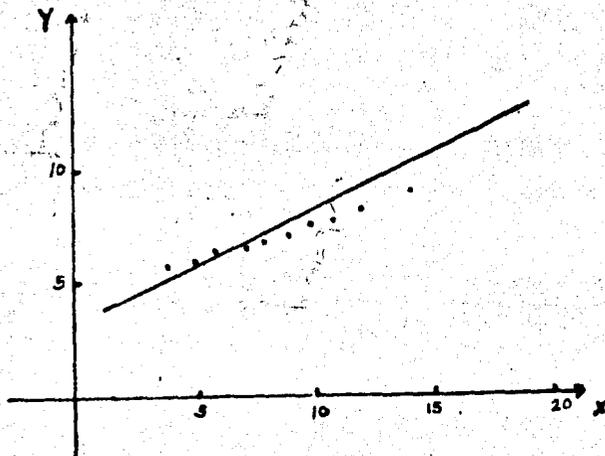
Este error se comete cuando se escogen elementos de otra población.

La importancia del estudio de las observaciones aberrantes ra-

dica en el hecho de que si no se analizan adecuadamente, las conclusiones a las que lleguemos no serán correctas y nos conducirán a resultados no confiables.

Estas observaciones pueden llegar a influir de manera determinante afectando el estudio del fenómeno o experimento en cuestión. Con el fin de ilustrar la importancia de ellas analizaremos los siguientes datos aportados por Anscombe en 1973.

<u>X</u>	<u>Y</u>
10	7.46
8	6.77
13	12.74
9	7.11
11	7.81
14	8.84
6	6.08
4	5.39
12	8.15
7	6.42
5	7.73



GRAFICA 1

Mediante el análisis de la gráfica, nos damos cuenta de la influencia que ejercen las observaciones con valores extremos sobre las restantes; en este caso hay una observación que desvía la recta del conjunto de las restantes hacia ella.

Para la solución del problema de la detección de aberrantes se han propuesto una gran variedad de criterios, desde reglas sin una base estadística formal hasta una metodología estadística sofisticada.

El primer paso que consideramos de importancia fue un trabajo sobre Geodesia publicado por Bessel (1846)¹, Director de la Escuela Alemana de Astronomía. En este trabajo se argumentaba que una observación no podía ser rechazada solo por su valor grande y que necesariamente todas las observaciones debían ser incluidas con el mismo peso a contribuir en el resultado.

Por primera vez el astrónomo Peirce (1852)² propuso un criterio para rechazar observaciones aberrantes, el cual fue reproducido 11 años después por Chauvenet (1863)². La aparición de reglas de rechazo al principio no fue muy bien recibida, pues se argumentaba que todas las observaciones deberían ser incluidas en el análisis. Después de esta literatura Wright (1884)³ propuso un procedimiento basado en el rechazo de cualquier observación que diste de la media por más de tres veces la desviación estándar y Goodwin (1913)² otro en el que se rechaza una observación extrema en una muestra de tamaño n si - - -

1 De Grubbs (1969)
2 De Barnett y Lewis (1978)
3 De Anscombe (1960)

dista de la media de las $n-1$ observaciones restantes cuatro veces la desviación promedio de estas $n-1$ observaciones.

Estos últimos procedimientos marcan un defecto que es el de escoger valores críticos constantes para sus pruebas ya que no dependen del tamaño de muestra.

Generalmente se utiliza la idea de arreglar las observaciones en orden ascendente de magnitud con el fin de ver más fácilmente qué observaciones toman valores extremos. Irwin (1925)¹ utilizó este criterio y propuso como solución, la diferencia entre la primera y la segunda observaciones más grandes para detectar si la más grande en magnitud es aberrante, como también la diferencia entre la segunda y tercera observaciones más grandes para detectar si la segunda y, por lo tanto, la primera son aberrantes, es decir,

$$T = \frac{X(n) - X(n-1)}{\sigma}$$

y

$$T = \frac{X(n-1) - X(n-2)}{\sigma} \quad (1.1.1)$$

donde σ es la desviación estándar de la población (σ conocida).

1 De De'Alba (1974)

No es recomendable el uso de esta regla en el caso de que el número de aberrantes exceda a dos ya que en estos casos tales diferencias pueden no ser significativas.

Diez años más tarde siendo el primero en proponer una regla de rechazo basada en el criterio de estandarización, Thompson -- (1935)¹ publicó un artículo donde sugirió el uso de la siguiente estadística:

$$T = \frac{x_i - \bar{x}}{s} \quad (1.1.2)$$

Donde:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

x_i observación seleccionada arbitrariamente.

La prueba propuesta por Thompson está basada en la siguiente hipótesis:

1 De De Alba (1974) y Grubbs (1950)

Ho: Todas las observaciones provienen de la misma población normal.

Esta prueba es útil en el caso de que se tenga una o más de una observación y según Thompson cuando el número de observaciones sea grande (30 ó más).

Thompson demostró que si X_i es una observación arbitraria entonces:

$$\frac{T \cdot \sqrt{n-2}}{\sqrt{n-1-T^2}} \sim t(n-2) \quad (1.1.3)$$

para la cual los valores críticos son fáciles de obtener de una Tabla t de Student con $n-2$ grados de libertad, también obtuvo una tabla de valores críticos de la estadística T .

Pearson y Chandra Sekar (1936) se basaron en los estudios realizados por Thompson sobre la desviación de una observación arbitraria con respecto a la media entre la desviación estándar y un año después publicaron un artículo donde expusieron la siguiente estadística utilizada para detectar una sola observación aberrante:

(1.1.4)

$$T_n = \frac{X(n) - \bar{X}}{S} \quad \text{Para detectar si } X(n) \text{ es aberrante}$$

$$T_1 = \frac{\bar{X} - X(1)}{S} \quad \text{Para detectar si } X(1) \text{ es aberrante} \\ (1.1.5)$$

Donde:

$$S^2 = \sum_{i=1}^n \frac{(x_i - \bar{X})^2}{n-1}$$

Las hipótesis que propusieron son:

Ho: Todas las observaciones provienen de la misma población Normal.

H1: La observación más grande (pequeña) proviene de una población con diferente media.

Los valores críticos de T fueron obtenidos sin derivar su distribución exacta.

Cabe hacer notar que la probabilidad de detectar una observación aberrante (una observación arbitraria) mediante el criterio de Thompson es diferente a la de detectar una aberrante -- (una observación particular ya sea $X(1)$ ó $X(n)$) mediante el criterio de Pearson y Chandra Sekar.

Un resultado también importante fue el propuesto por Grubbs -- (1950) que sugirió la utilización de la siguiente estadística

para probar la significancia de la observación más grande en una muestra de tamaño n de una población Normal:

$$\frac{s_n^2}{s^2} = \frac{\sum_{i=1}^{n-1} (x_{(i)} - \bar{x}_n)^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2} \quad (1.1.6)$$

Donde:

$$x_{(1)} < x_{(2)} < \dots < x_{(n)}$$

$$\bar{x}_n = \frac{1}{n-1} \sum_{i=1}^{n-1} x_{(i)}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_{(i)} = \frac{1}{n} \sum_{i=1}^n x_i$$

De manera análoga se puede probar la significancia de la observación más pequeña, es decir si la diferencia entre la observación más pequeña y las restantes es "grande", mediante la estadística:

$$\frac{s_1^2}{s^2} = \frac{\sum_{i=2}^n (x_{(i)} - \bar{x}_1)^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2} \quad (1.1.7.)$$

Donde:

$$\bar{x}_1 = \frac{1}{n-1} \sum_{i=2}^n x_{(i)}$$

Es importante notar que la estadística propuesta por Grubbs tiene una relación con la estadística T propuesta por Pearson y Chandra Sekar para el caso de detectar una observación aberrante grande.

$$\begin{aligned} \frac{s_n^2}{s^2} &= 1 - \frac{1}{n-1} \left(\frac{x_{(n)} - \bar{x}}{s} \right)^2 \\ &= 1 - \frac{1}{n-1} T_n^2 \end{aligned}$$

La distribución exacta de $\frac{s_n^2}{s^2}$ y por consiguiente de T_n fue derivada por Grubbs.

Los valores críticos para estas estadísticas se presentan en el apéndice de Tablas.

Dixon (1950) trabajó con poblaciones que siguen una distribución Normal y propuso varias estadísticas en donde puso especial cuidado de que no se vieran afectadas por la presencia de otras observaciones extremas en la muestra.

$r_{10} = \frac{X(2) - X(1)}{X(n) - X(1)}$	Para detectar si X(1) es observación aberrante.
$r'_{10} = \frac{X(n) - X(n-1)}{X(n) - X(1)}$	Para detectar si X(n) es observación aberrante.
$r_{11} = \frac{X(2) - X(1)}{X(n-1) - X(1)}$	Para detectar si X(1) es observación aberrante ignorando X(n).
$r'_{11} = \frac{X(n) - X(n-1)}{X(n) - X(2)}$	Para detectar si X(n) es observación aberrante ignorando X(1).
$r_{21} = \frac{X(3) - X(1)}{X(n-1) - X(1)}$	Para detectar si X(1) es observación aberrante ignorando X(2) y X(n).
$r'_{21} = \frac{X(n) - X(n-2)}{X(n) - X(2)}$	Para detectar si X(n) es observación aberrante ignorando X(1) y X(n-1).
$r_{22} = \frac{X(3) - X(1)}{X(n-2) - X(1)}$	Para detectar si X(1) es observación aberrante ignorando X(2), X(n-1) y X(n).
$r'_{22} = \frac{X(n) - X(n-2)}{X(n) - X(3)}$	Para detectar si X(n) es observación aberrante ignorando X(1), X(2) y X(n-1).

Los valores críticos para estas pruebas vienen en el apéndice de Tablas.

Ferguson (1961)¹ propuso técnicas basadas en el hecho de que las observaciones aberrantes se deben a cambios en la media (localización) y varianza (escala) de la población. Por ejemplo supongamos que se están realizando mediciones en un proceso químico determinado a una cierta temperatura y por una falla experimental cambia la temperatura y se siguen tomando mediciones, esto trae como consecuencia un cambio en la media y/o en la varianza de la población bajo estudio, por lo cual se producen observaciones aberrantes. Para esto utiliza los coeficientes de asimetría y kurtosis² muestrales como pruebas de una cola y dos colas respectivamente, o sea, el coeficiente de asimetría muestral lo utiliza para detectar observaciones aberrantes en una misma dirección ya sea con valores muy grandes o muy pequeños y el coeficiente de kurtosis muestral para detectar si varias observaciones extremas tanto muy pequeñas como muy grandes son aberrantes.

Coef. de asimetría muestral:

$$\sqrt{b_1} = \sqrt{n} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^{3/2}}$$

1 De Grubbs (1969)

2 Medida de agudeza de la curva

Coef. de kurtosis muestral:

$$b_2 = n \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2}$$

Estos coeficientes se utilizan en el caso de que la distribución de las observaciones sea Normal ya que esencialmente son pruebas para detectar fallas en la suposición de normalidad.

En caso de Normalidad sabemos que:

$$b_1 = 0$$

Los valores críticos vienen en el Apéndice de Tablas.

El procedimiento para aplicar esta prueba es el siguiente:

1. Si los valores de b_1 y b_2 exceden a los valores críticos entonces la observación más alejada de la media se rechaza.
2. El procedimiento se repite hasta que no haya más valores que se puedan catalogar como aberrantes.

En el siguiente cuadro presentamos un resumen de los principales criterios propuestos.

	ESTADISTICA	AUTOR
Estadística Unilateral.- Sirve para detectar si X(n) es aberrante.	$\frac{X(n) - X(n-1)}{S}$	Irwin (1925)
Estadística Bilateral.- Util para detectar una o más observaciones aberrantes.	$\frac{X_i - \bar{X}}{S}$	Thompson (1935)
Estadísticas Unilaterales.- Sirven para detectar si X(n) y X(1) respectivamente son aberrantes.	$\frac{X(n) - X(1)}{S}$ $\frac{X(1) - \bar{X}}{S}$	Pearson y Chandra Sekar (1936)
Estadísticas Unilaterales.- Sirven para detectar si X(n) y X(1) respectivamente son aberrantes.	$\frac{\sum_{i=1}^{n-1} (X(i) - \bar{X}_n)^2}{\sum_{i=1}^n (X(i) - \bar{X})^2}$ $\frac{\sum_{i=2}^n (X(i) - \bar{X}_1)^2}{\sum_{i=1}^n (X(i) - \bar{X})^2}$	Grubbs (1950)
Estadísticas Unilaterales.- Sirven para detectar si X(n) y X(1) respectivamente son aberrantes.	$\frac{X(n) - X(n-1)}{X(n) - X(1)}$ $\frac{X(2) - X(1)}{X(n) - X(1)}$	Dixon (1950)
Estadística Unilateral.- Sirve para detectar observaciones aberrantes en un solo sentido ya sea hacia la izquierda o hacia la derecha.	$\frac{n \sum_{i=1}^n (X_i - \bar{X})^3}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right]^{3/2}}$	Ferguson (1961)
Estadística Bilateral.- Sirve para detectar observaciones aberrantes tanto del lado derecho como izquierdo.	$\frac{n \sum_{i=1}^n (X_i - \bar{X})^4}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right]^2}$	Ferguson (1961)

1.2. Ejemplos

E.1. Con el fin de mostrar el uso de algunas de las estadísticas antes mencionadas, ejemplificaremos mediante algunos ejercicios.

Consideremos las 10 siguientes observaciones obtenidas de medir la fuerza de rompimiento de cable de cobre,

para ilustrar el uso de las estadísticas propuestas por Pearson y Chandra Sekar y Dixon.

$$X_{(1)} = 568$$

$$X_{(2)} = 570$$

$$X_{(3)} = 570$$

$$X_{(4)} = 570$$

$$X_{(5)} = 572$$

$$X_{(6)} = 572$$

$$X_{(7)} = 572$$

$$X_{(8)} = 578$$

$$X_{(9)} = 584$$

$$X_{(10)} = 596$$

Suponemos que esta muestra proviene de una población Normal.

En este conjunto de datos nos damos cuenta que la observación extrema es $X_{(10)}$ ya que está más alejada del resto en relación a la separación de las restantes.

La media es:

$$\bar{X} = \frac{1}{10} \sum_{i=1}^{10} x_i = 575.2$$

La desviación estándar muestral es:

$$\begin{aligned} s &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{X}^2 \right) \\ &= 8.7025 \end{aligned}$$

Calculamos:

$$\begin{aligned} T_{10} &= \frac{596 - 575.2}{8.7025} \\ &= 2.3901 \end{aligned}$$

De la Tabla I, encontramos que para $n = 10$ y un nivel de significancia del 5%, el valor crítico es 2.18 el cual es menor que el valor calculado de la estadística T_{10} , por esta razón consideramos a $X_{(10)}$ como observación aberrante.

Ahora utilizando la estadística de Dixon para detectar si

el valor máximo es una observación aberrante tenemos que:

$$\begin{aligned} r'_{11} &= \frac{596 - 584}{596 - 570} \\ &= 0.4615 \end{aligned}$$

Comparando este valor con el de Tabla III a un nivel de significancia de .05 encontramos que .477 es mayor, por lo tanto no consideramos a $X_{(10)}$ como aberrante.

Este ejemplo también nos sirve para ilustrar el hecho de que no siempre se llega a las mismas conclusiones al utilizar diferentes criterios. Esto se debe a que en la gran variedad de estadísticas propuestas hay unas que son más sensibles que otras. En este caso la estadística de Pearson y Chandra Sekar es más sensible para el caso de detección de una sola observación aberrante.

E.2. Un ejemplo tradicional para la ilustración de la presencia de aberrantes es el de las mediciones de los semi-diámetros verticales de Venus tomadas por Lieutenant Herndon en 1846. Estas observaciones aparecen en orden ascendente de magnitud y en desviación con respecto a la media y son:

$X_{(1)} = -1.40$	$X_{(6)} = -0.13$	$X_{(11)} = 0.20$
$X_{(2)} = -0.44$	$X_{(7)} = -0.05$	$X_{(12)} = 0.39$
$X_{(3)} = -0.30$	$X_{(8)} = 0.06$	$X_{(13)} = 0.48$
$X_{(4)} = -0.24$	$X_{(9)} = 0.10$	$X_{(14)} = 0.63$
$X_{(5)} = -0.22$	$X_{(10)} = 0.18$	$X_{(15)} = 1.01$

Suponiendo normalidad de las observaciones, consideramos que las candidatas a aberrantes son $X_{(1)}$ y $X_{(15)}$.

Utilizando la estadística de Pearson y Chandra Sekar veremos si tanto $X_{(1)}$ como $X_{(15)}$ son aberrantes:

La media de estos datos es:

$$\bar{X} = 0.018$$

La desviación estándar estimada es:

$$S = 0.5511$$

Calculamos:

$$\begin{aligned} T_1 &= \frac{0.018 - (-1.40)}{0.5511} \\ &= 2.573 \end{aligned}$$

Nuevamente de la Tabla I a un nivel de significancia del 5% y comparando con el valor crítico 2.41, resulta que -1.40 es aberrante.

Vamos a probar ahora si 1.01 es aberrante.

Si decidimos omitir $X_{(1)}$ podemos considerar las 14 observaciones restantes y volviendo a numerar los datos tenemos que:

La media es:

$$\bar{X} = 0.119$$

La desviación estándar estimada es:

$$S = 0.401$$

Por lo tanto,

$$T_{14} = \frac{0.01 - 0.119}{0.401}$$

$$= 2.2219$$

Este valor comparado al de tablas que es 2.37 a un nivel

de significancia del 5% y un tamaño de muestra $n=14$ es menor, por lo que concluimos que 1.01 no es aberrante.

Utilizando ahora el criterio de Dixon tenemos que:

$$\begin{aligned} r'_{22} &= \frac{1.01 - 0.48}{1.01 + 0.24} \\ &= 0.424 \end{aligned}$$

A un nivel del 5% encontramos en la Tabla III que para $n = 14$ el valor crítico es 0.546 y puesto que el valor calculado 0.424 es menor, decidimos que 1.01 no es aberrante.

Nuevamente con los datos del ejercicio E.2. y calculando la estadística de Grubbs tenemos que la media de las observaciones omitiendo la última es:

$$\bar{X}_{15} = -0.0529$$

y por lo tanto,

$$\frac{S^2_n}{S^2} = \frac{3.2012}{4.2828}$$

$$= 0.7475$$

De la Tabla II al nivel de significancia del 5% obtenemos el valor crítico .5559, por lo tanto la observación con valor más grande $X_{(15)}$ es aberrante.

Por último utilizando estos mismos datos ilustramos el uso de la estadística propuesta por Ferguson.

Calculando:

$$\sum_{i=1}^{15} (X_i - \bar{X})^5 = -1.6472$$

$$\left[\sum_{i=1}^{15} (X_i - \bar{X})^2 \right]^{3/2} = 8.7605$$

Por lo tanto el coeficiente de asimetría es

$$b_1 = -0.7282$$

Valor que comparado con el de la Tabla IV nos reporta a un nivel de .05 que no detectamos aberrantes en una sola dirección.

CAPITULO II

Calculando el coeficiente de kurtosis se tiene,

$$\sum_{i=1}^{15} (x_i - \bar{x})^4 = 5.279$$

$$\left(\sum_{i=1}^{15} (x_i - \bar{x})^2 \right)^2 = 18.0594$$

Por lo que el valor de esta estadística es

$$b_2 = 4.3847$$

valor que comparado con el de la Tabla IV a un nivel de .05 y un tamaño de muestra de 15, resulta ser mayor razón por la cual concluimos que -1.40 y 1.01 son aberrantes.

2. DETECCION DE OBSERVACIONES ABERRANTES PARA EL CASO EXPONENCIAL

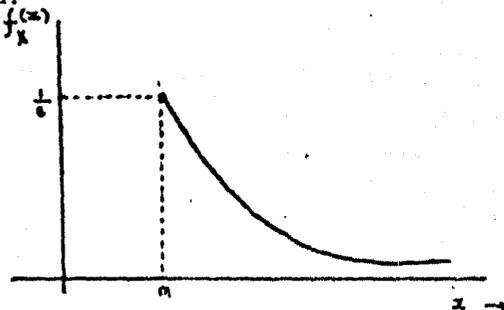
2.1. La Distribución Exponencial con dos Parámetros

A una variable aleatoria X con función de densidad de probabilidad (f.d.p.) de la siguiente forma:

$$f_X(x) = c^{-1} e^{-\frac{1}{c}(x-m)} \quad x \geq m, c > 0 \quad (2.1.1.)$$

Se le llama exponencial con dos parámetros, también se le conoce como exponencial negativa. Esta distribución es un caso particular de la distribución Gamma y cuando $m=0$ se obtiene la distribución exponencial con un parámetro.

La gráfica de esta función está representada en la siguiente figura:



$m > 0$

2.2. Historia y Propiedades¹

Durante años se han venido desarrollando técnicas para encontrar estimadores de los parámetros de la distribución exponencial, los cuales se basan en estadísticas de orden.

Entre los resultados más importantes está el de Epstein y Sobel (1954) que después de haber obtenido el estimador máximo verosímil de c (parámetro de escala) en la distribución exponencial con un parámetro publicado en 1953, ampliaron sus resultados para la distribución exponencial con dos parámetros obteniendo así los estimadores máximo verosímiles para c y m , (m parámetro de localización).

En ese mismo año Sarhan (1954) encontró los mejores estimadores lineales e insesgados (M.E.L.I.), de los dos parámetros.

Entre las aportaciones al modelo exponencial estandarizado (con parámetros $m=0$ y $c=1$) está la función de densidad de probabilidad (f.d.p.) del semi-rango y los cumulantes encontrada por Rider en 1959.

1 Las referencias de esta sección son del libro de Johnson y Kotz (1970)

Al año siguiente Epstein (1960) extendió sus resultados para la distribución exponencial con uno y dos parámetros en el caso de tener muestras censuradas a la derecha y/o a la izquierda.

La aportación que hicieron al modelo exponencial Epstein y Sobel fue una gran contribución para aplicaciones de tipo industrial dentro de la Teoría de Confiabilidad que mencionaremos más adelante.

Continuando con los estudios hechos sobre este modelo Harter (1961) propuso el mejor estimador lineal de c para la distribución exponencial con un parámetro, basado en estadísticas de orden. Sarhan (1963) generalizó estos resultados encontrando los MELI para m y c en la distribución exponencial con dos parámetros basados también en estadísticas de orden y utilizando métodos numéricos al igual que Harter.

Siddiqui (1963) presentó un método analítico para encontrar los MELI'S basado en la fórmula de Euler-Maclaurin.

La distribución exponencial tiene una gran variedad de aplicación en el campo estadístico.

En muchas ocasiones este modelo es muy útil para aproximar va-

riables relacionadas con la duración o tiempos de vida de un sistema. Una de las aplicaciones más importantes de esta variable aleatoria se encuentra en lo que denominamos Teoría de Confiabilidad de Sistemas, donde:

X Representa el tiempo de duración de vida de un sistema.

m Parámetro de escala, representa la vida mínima del sistema, y

$X_{(1)}, X_{(2)}, \dots, X_{(n)}$

Estadística de orden, representa, en una prueba de tiempos de duración de un sistema determinado, los tiempos de falla de sistemas independientes e idénticos.

Dentro de la teoría desarrollada en este tema, se encuentra una característica importante que radica en el hecho de que el tiempo de vida futuro de un sistema dado, no se ve afectado por la longevidad con la que cuenta.

Por ejemplo: Supongamos que estamos midiendo el tiempo de vida de una bombilla. La probabilidad de que esta bombilla dure tres horas desde nueva es la misma a la de que dure también tres horas habiendo tenido un tiempo de vida anterior de 10 horas ó de un día ó de una semana.

Esto escrito formalmente se reduce a lo siguiente:

$$P_r [X \leq x + t / X > t] = P_r [X \leq x] \quad \forall x, t > 0$$

Algunas propiedades de la distribución (2.1.1.) son:

1. Momentos y función generatriz de momentos:

La función generatriz de momentos de la variable aleatoria con f.d.p. (2.1.1.) es:

$$\begin{aligned} m_X(t) &= E(e^{tX}) \\ &= (1 - ct)^{-1} e^{mt} \end{aligned} \quad (2.2.1.)$$

2. La función generatriz de cumulantes es:

$$\begin{aligned} \log_e m_X(t) &= \log_e E(e^{tX}) \\ &= mt - \log_e(1 - ct) \end{aligned} \quad (2.2.2.)$$

3. Los primeros dos cumulantes son:

$$\begin{aligned} k_1 &= E(X) = m + c \\ k_2 &= V(X) = c^2 \end{aligned} \quad (2.2.3.)$$

La mediana es $m + c \log_e 2$ y la moda es m .

4. Estimación puntual.

Sean X_1, X_2, \dots, X_n variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.), según (2.1.1.).

Los estimadores máximo verosímiles para m y c son respectivamente:

$$m = X_{(1)} = \min(X_1, \dots, X_n) \quad (2.2.4.)$$

$$c = \sum_{i=1}^n \frac{(X_i - m)}{n} = \bar{X} - X_{(1)}$$

5. La f.d.p. del estimador máximo verosímil de m es:

$$f(x) = \frac{n}{c} e^{-\frac{n}{c}(x-m)} \quad x \geq m \quad (2.2.5.)$$

La media y varianza de la variable aleatoria con f.d.p. - (2.1.1.) son $m + c$ y c^2 respectivamente.

La media de \hat{m} es $m + \frac{c}{n}$ y la varianza es $\frac{c^2}{n^2}$, ya que la densidad de $X_{(1)}$ es de la forma de (2.1.1.) sustituyendo c por $\frac{c}{n}$.

6. La f.d.p. del estimador máximo verosímil de c es:

$$f_{\bar{X}-X_{(1)}}(x) = \frac{\left(\frac{n}{c}\right)^{n-1}}{\Gamma(n-1)} x^{n-2} e^{-\frac{n}{c}x} \quad (2.2.6.)$$

la cual es una densidad gamma con parámetros $\lambda = n-1$ y $\beta = \frac{n}{c}$ por lo tanto la media de \hat{c} es $c(1 - \frac{1}{n})$ y la varianza es $c^2(\frac{1}{n} - \frac{1}{n^2})$

7. Una estadística suficiente y completa para m y c es la -
pareja:

$$(X_{(1)}, \bar{X} - X_{(1)}) \quad (2.2.7.)$$

8. La transformación de la variable aleatoria exponencial

$$U = \frac{2n}{c} (\bar{X} - m)$$

es de especial interés.

Trabajando con esta transformación tenemos que:

$$\frac{2n}{c}(\bar{X} - m) = \frac{2}{c} \sum_{i=1}^n (X_i - m),$$

y la densidad de $Z = X - m$ como sabemos es una variable aleatoria exponencial con parámetro $\frac{1}{c}$

$$f(z; c) = \begin{cases} \frac{1}{c} e^{-\frac{z}{c}} & z \geq 0 \\ 0 & \text{otro caso} \end{cases}$$

La densidad de $Z = \sum z_i$ es gamma con parámetros $\alpha = n$ y $\beta = \frac{1}{c}$.

Finalmente, mediante el cambio de variable

$$W = \frac{2}{c} Z, \text{ obtenemos:}$$

$$f_W(w) = \frac{1}{\Gamma(n)} \frac{1}{2^n} w^{n-1} e^{-\frac{w}{2}}$$

que es la densidad de una X^2 con $2n$ g.l.

Por lo tanto

$$\frac{2n}{c} (\bar{X} - m) \sim X^2_{(2n)} \quad (2.2.8.)$$

9. La siguiente transformación, también de utilidad, es:

$$V = \frac{2n}{c} (X_{(1)} - m)$$

La densidad de $Z = X_{(1)} - m$ es exponencial con parámetro $\frac{n}{c}$, lo cual es equivalente a una densidad Gamma con parámetro

tros $\alpha = 1$ y $\beta = \frac{n}{c}$.

Mediante el cambio de variable $W = \frac{2n}{c} Z$ llegamos a que:

$$f_W(w) = \frac{1}{2} e^{-\frac{w}{2}}$$

la cual es una densidad χ^2 con 2 g.l.

Por lo tanto $\frac{2n}{c} (X_{(1)} - m) \sim \chi^2_{(2)}$ (2.2.9.)

De la siguiente relación obtenemos también un resultado importante:

$$n(\bar{X} - m) = n(X_{(1)} - m) + n(\bar{X} - X_{(1)})$$

Utilizando los resultados (2.2.8.) y (2.2.9.) y debido a que la suma de χ^2 es χ^2 cuando son independientes, tenemos que:

$$\frac{2n}{c} (\bar{X} - X_{(1)}) \sim \chi^2_{(2n-2)} \quad (2.2.10.)$$

2.3. Detección de Observaciones Aberrantes para el caso

Exponencial

Hasta la fecha, la mayoría de los trabajos que se han presentado sobre observaciones aberrantes han sido desarrollados en el contexto de poblaciones Normales; sin embargo, el estudio de éstas tiene importancia práctica en otras distribuciones, en particular para la exponencial negativa.

Todos los criterios tratados en esta sección se refieren a la variable aleatoria con función de densidad de probabilidad -- (2.1.1.), y son los siguientes:

1. Laurent
2. Likes y Kabe
3. Shapiro y Wilk
4. Tiku

Criterio de Laurent

En el caso de que la distribución de las observaciones sea normal, han sido trabajadas estadísticas de la forma: $(X_1 - u)/v$ y $(X_{(i)} - u)/v$ donde X_1 es una observación arbitraria, $X_{(i)}$ la i -ésima observación ordenada, u es una estimación del parámetro de localización y v es una estimación del parámetro de es-

cala independiente del parámetro de localización. Estas estadísticas han sido utilizadas para probar normalidad de las observaciones o para probar si hay observaciones aberrantes.

Como antes habíamos mencionado¹, Thompson da la distribución de la estadística $(X_i - \bar{X})/S$ y Pearson y Chandra Sekar dan una aproximación a la distribución de la estadística $(X_{(n)} - \bar{X})/S$ para el caso normal.

Laurent (1962) presenta estadísticas de forma análoga a las anteriores para el caso exponencial:

$$\frac{X_i - X_{(1)}}{n(\bar{X} - X_{(1)})} \quad \frac{X_{(i)} - X_{(1)}}{n(\bar{X} - X_{(1)})}$$

donde $X_{(1)}$ es el estimador máximo verosímil de m

$\bar{X} - X_{(1)}$ es el estimador máximo verosímil de c

La distribución de esta estadística la encuentra a partir de la distribución condicional de $X_{(i)}$, dada la estadística suficiente y completa. Utiliza los Teoremas de Rao-Blackwell y Lehmann-Scheffé para así obtener un estimador insesgado y de varianza mínima para las funciones de distribución:

$$P \left[X_i \leq x / X_{(1)}, \bar{X} - X_{(1)} \right]$$

$$P \left[X_{(i)} \leq x / X_{(1)}, \bar{X} - X_{(1)} \right]$$

Definiendo

$$U_{(i)} = \frac{X_{(i)} - X_{(1)}}{n(\bar{X} - X_{(1)})}$$

que es independiente de $X_{(1)}, \bar{X} - X_{(1)}$ encuentra

$$P \left[U_{(i)} \leq x \right] = 1 - \sum_{r=n-i+1}^k (-1)^{r+n-i+1} \binom{n}{r} \binom{r-1}{n-1} \left(1 - \frac{r}{n}\right) (1-rx)^{n-1}$$

$$\text{para } x \geq 0, k \leq \frac{1}{x}$$

a partir de:

$$P \left[X_{(i)} \leq x / X_{(1)}, \bar{X} - X_{(1)} \right] =$$

$$= 1 - \sum_{r=n-i+1}^k (-1)^{r+n-i+1} \binom{r-1}{n-1} \binom{n}{r} \left(1 - \frac{r}{n}\right) \left(1 - \frac{x - X_{(1)}}{n(\bar{X} - X_{(1)})}\right)^{n-2}$$

$$\text{para } x \geq X_{(1)}$$

$$k \leq n(\bar{X} - X_{(1)})$$

Para el caso particular $i=n$ tenemos que:

$$P[U_{(n)} \leq x] = 1 - \sum_{r=1}^k (-1)^{r+1} \binom{n-1}{r} (1-rx)^{n-2}; \quad k \leq \frac{1}{x}$$

Por lo tanto para detectar si la observación más grande es aberrante propone:

$$U_{(n)} = \frac{X_{(n)} - X_{(1)}}{n(\bar{X} - X_{(1)})} \quad (2.3.1.)$$

y para detectar si la observación más pequeña es aberrante propone:

$$\frac{1}{nU_{(n)}} = \frac{\bar{X} - X_{(1)}}{X_{(n)} - X_{(1)}} \quad (2.3.2.)$$

Los valores críticos tanto para la distribución de $U_{(n)}$ como de $\frac{1}{nU_{(n)}}$ vienen en la Tabla VIII del Apéndice.

Ejemplo A

Los siguientes valores¹ se refieren a la duración en horas, de la vida de 10 bulbos, los cuales pertenecen a una población exponencial y están ordenados en orden ascendente de magnitud.

1 Fuente: Tabla de Números Aleatorios

$X_{(1)}$	=	3.4	$X_{(6)}$	=	78.01
$X_{(2)}$	=	21.96	$X_{(7)}$	=	79.49
$X_{(3)}$	=	51.75	$X_{(8)}$	=	90.51
$X_{(4)}$	=	55.94	$X_{(9)}$	=	232.31
$X_{(5)}$	=	61.30	$X_{(10)}$	=	576.54

En esta muestra la última observación está más alejada del resto.

Calculando el valor de la estadística para probar si $X_{(10)}$ es aberrante tenemos que:

$$\begin{aligned}\bar{X} &= \sum_{i=1}^{10} \frac{X_i}{10} \\ &= 125.12\end{aligned}$$

Por lo tanto,

$$U_{(10)} = \frac{576.54 - 3.4}{10(125.12 - 3.4)} = 0.470867$$

A un nivel de significancia de .05 y un tamaño de muestra $n=10$ se obtiene de la relación (3.3.6) el valor crítico 0.47749 ya que la Tabla VIII no contiene este valor. Por lo tanto concluimos que $X_{(10)}$ no es aberrante.

Criterio de Likes y Kabe

Otros autores que también tratan el problema de la detección - como pruebas de significancia, son Likes (1966)¹ y Kabe - - - (1970)¹.

Las estadísticas que plantean para detectar si la observación más grande o más pequeña es aberrante, son del tipo de las que propuso Dixon.

$$T_1 = \frac{X_{(2)} - X_{(1)}}{X_{(n)} - X_{(1)}} \quad (2.3.3.)$$

$$y \quad T_n = \frac{X_{(n)} - X_{(n-1)}}{X_{(n)} - X_{(1)}} \quad (2.3.4.)$$

Este par de estadísticas son vulnerables al efecto que ejercen sobre ellas los valores $X_{(n-1)}$ y $X_{(2)}$, ya que puede ser que éstas también tengan valores extremos y por lo tanto las diferencias $X_{(n)} - X_{(n-1)}$ y $X_{(2)} - X_{(1)}$ no sean realmente significativas.

Para muestras aleatorias de la distribución exponencial estas estadísticas tienen las siguientes distribuciones respectivamente:

1 De Barnett y Lewis (1977)

$$F_1(x) = (n-2) B\left(\frac{1+(n-2)x}{1-x}, n-2\right) \quad 0 \leq x \leq 1$$

y

$$F_n(x) = 1 - (n-1)(n-2) B\left(\frac{2-x}{1-x}, n-2\right) \quad 0 \leq x \leq 1$$

Los niveles de significancia tabulados para la distribución de (2.3.3.) se encuentran en Tabla VI y para la distribución - - (2.3.4.) en la Tabla VII del Apéndice.

Ejemplo

Utilizando los datos del Ejemplo A procedemos a calcular la estadística de Likes y Kabe.

$$T_n = \frac{576.54 - 232.31}{576.54 - 3.4}$$
$$= 0.6006$$

De la Tabla VII obtenemos para $n=10$ y nivel de significancia .05 el valor crítico .675 que comparado con 0.6006 es mayor, por lo tanto $X_{(10)} = 576.54$ no es aberrante.

Criterio de Shapiro y Wilk

Shapiro y Wilk (1972) desarrollan la estadística W para probar varias hipótesis simples y compuestas de exponencialidad de -

las observaciones utilizando procedimientos similares a los utilizados en trabajos anteriores, donde obtienen la estadística W para probar normalidad de las observaciones.

Derivan la estadística W del mejor estimador insesgado de c (parámetro de escala) utilizando mínimos cuadrados generalizados, mediante la siguiente técnica:

Sea $y' = (y_1, \dots, y_n)$ un vector de estadísticas de orden de la muestra, entonces

$$y_i = m + c E(x_{(i)}) \quad i = 1, 2, \dots, n$$

y utilizando mínimos cuadrados generalizados se obtiene

$$c = \frac{1' V^{-1} (1 h' - h 1') V^{-1} y}{1' V^{-1} 1 h' V^{-1} h - (1' V^{-1} h)^2}$$

donde:

$$1' = (1, 1, \dots, 1)$$

V es la matriz de varianzas y covarianzas

$$V = (v_{ij})_{n \times n}$$

$$v_{ij} = E(x_{(i)} - E(x_{(i)}))(x_{(j)} - E(x_{(j)})) \quad i, j = 1, 2, \dots, n$$

$$h_i = E(x_{(i)}) \quad i = 1, 2, \dots, n$$

$$h' = (E(x_{(1)}), E(x_{(2)}), \dots, E(x_{(n)}))$$

Trabajando con c se llega a la expresión

$$c = \frac{n(\bar{y} - y_1)}{n - 1}$$

Finalmente la estadística W que proponen depende de c y es el cociente entre el cuadrado de la combinación lineal de la estadística de orden muestral $\bar{y} - y_{(1)}$ y la suma de cuadrados de la desviación con respecto a la media:

$$\begin{aligned} W &= \frac{n(\bar{y} - y_1)^2}{(n-1) s^2} \\ &= \frac{c(\bar{y} - y_1)}{s^2} \end{aligned} \quad (2.3.5.)$$

$$\text{donde } s^2 = \sum_{i=1}^n (y_i - \bar{y})^2$$

Si bien, la hipótesis nula está bien especificada existen diferentes alternativas inherentes, es decir, cuando las conclusiones a las que nos lleva la estadística W son de rechazar la hipótesis nula puede ser por diferentes razones, una de ellas es la presencia de aberrantes. Chen (1971), en su estudio realizado sobre la potencia de la estadística W para el caso normal, encontró que esta es sensible a detectar no normalidad en

muestras normales contaminadas y que por esta razón es sensible a la presencia de aberrantes. En consecuencia Barnett y Lewis proponen la estadística W como prueba para detectar aberrantes para el caso exponencial, señalando que:

Valores significativamente altos de W indican la presencia de una o varias observaciones aberrantes grandes y/o una aberrante con valor chico.

Valores significativamente bajos de W indican la presencia de una o varias aberrantes chicas.

Esta estadística es de 2 colas, es decir, podemos detectar observaciones aberrantes tanto de un lado como de otro, sin embargo, no hay manera de saber cual o cuales fueron las observaciones detectadas, sólo nos dice cuando la muestra fue contaminada con observaciones no pertenecientes a dicha población, las cuales se consideran como aberrantes.

Algunas de las propiedades de esta estadística son:

1. Es invariante a cambios en la escala y localización
2. El valor máximo que toma es 1
3. El valor mínimo que toma es $(n-1)^{-2}$
4. La distribución de W bajo H_0 es independiente del mínimo

$X_{(1)}$ y la media \bar{X} y sólo depende del tamaño de muestra n .

Los valores críticos de W vienen dados en la Tabla IX del Apéndice.

Ejemplo

Utilizando los datos del Ejemplo A tenemos que:

$$W = \frac{10(125.12 - 3.4)^2}{9(260468.78)}$$
$$= 0.0632$$

De la Tabla IX observamos que para un nivel de 5% repartido entre las 2 colas, .2879 es mayor que .0632 por lo que no rechazamos la hipótesis de exponencialidad de las observaciones, razón por la cual podemos pensar que $X_{(20)}$ no es aberrante.

Criterio de Tiku

Tiku (1975) trata el problema de la detección de observaciones aberrantes planteando la siguiente hipótesis:

H_0 : Todas las observaciones de una muestra aleatoria de tamaño n provienen de la misma población exponencial,

es decir, la muestra no contiene aberrantes.

H_a : Existen observaciones aberrantes, es decir, observaciones muy grandes o muy pequeñas comparadas con el resto.

La estadística que propone es el cociente entre el estimador máximo verosímil o máximo verosímil modificado de c , calculado de la muestra censurada¹, y el máximo verosímil de c calculado a partir de toda la muestra.

Formalmente,

$$T = \frac{C_m}{c} h$$

Donde

$$h = \frac{E(c)}{E(C_m)}$$

C_m es el estimador máximo verosímil modificado de c y

c es el estimador máximo verosímil de C

1 Muestra censurada es aquella que se obtiene de omitir las observaciones con valores extremos.

Para el caso de la detección de la observación aberrante con el valor más pequeño, la estadística se reduce a la siguiente expresión:

$$T = \frac{(n-1) \sum_{i=2}^n (X_{(i)} - X_{(2)})}{n \sum_{i=1}^{(n-2)} (X_i - X_{(1)})} \quad (2.3.6.)$$

y para el caso de detectar la de valor más grande:

$$T = \frac{(n-1) \left(\sum_{i=1}^{n-1} X_{(i)} + X_{(n-1)} - n X_{(1)} \right)}{n \sum_{i=1}^{(n-2)} (X_i - X_{(1)})} \quad (2.3.7.)$$

La distribución de $\frac{n-2}{n-1} T$ es una distribución Beta con parámetros $p = n-2$ y $q = 1$.

Si el valor del parámetro m es conocido, solo basta reemplazarlo por $X_{(1)}$ y la distribución seguirá siendo Beta pero con parámetros $p = n-1$ y $q = 1$.

Para valores pequeños de T rechazamos la hipótesis nula, es decir, si T es menor que el valor de la Tabla entonces rechazamos H_0 .

Los valores críticos para algunos tamaños de muestra y nivel de significancia .01 y .05 vienen dados en la Tabla V del Apéndice, sin embargo, estos son fácilmente obtenibles ya que si $\frac{n-2}{n-1} T$ tiene una distribución Beta con parámetros $p = n-2$ y $q = 1$, su densidad es

$$f(t; p, q) = \frac{\Gamma(n-2+1)}{\Gamma(n-2)\Gamma(1)} t^{n-2-1} (1-t)^0 \\ = (n-2) t^{n-3}$$

y su distribución es:

$$F(t; p, q) = \int_0^x (n-2) t^{n-3} dt \\ = x^{n-2}$$

Por lo tanto, si queremos los cuantiles que nos dejen hacia la izquierda una área de tamaño α , es decir,

$$P[T \leq x] = \alpha$$

los obtendremos evaluando la siguiente expresión para determinados niveles de significancia y tamaños de muestra:

$$x = e^{\frac{\ln \alpha}{n-2}}$$

Ejemplo

Finalmente con los datos que hemos utilizado para ilustrar el uso de las estadísticas, calculamos el valor de la estadística para detectar si $X_{(10)}$ es aberrante y tenemos que:

$$\left(\sum_{i=1}^9 X_{(i)} + X_{(9)} - 10 X_{(1)} \right) = 1449.52$$

$$\sum_{i=1}^{10} (X_i - X_{(1)}) = 1217.21$$

Por lo que,

$$T = 1.19085$$

valor que, comparado con el valor crítico 0.68765 a un nivel de .05 resulta ser mayor, por lo tanto, $X_{(10)} = 576.54$ no es aberrante.

El siguiente cuadro es un resumen de las estadísticas utilizadas en la detección de aberrantes para el caso exponencial con 2 parámetros, clasificados por autor.

AUTOR	REFERENCIA	DESCRIPCION DE LA PRUEBA	ESTADISTICA
LAURENT	2.3.6.	Prueba para detectar si $X_{(1)}$ es aberrante	$\frac{\bar{X} - X_{(1)}}{X_{(n)} - X_{(1)}}$
	2.3.7.	Prueba para detectar si $X_{(n)}$ es aberrante	$\frac{X_{(n)} - X_{(1)}}{n(\bar{X} - X_{(1)})}$
LIKES Y KARE	2.3.3.	Prueba para detectar si $X_{(1)}$ es aberrante	$\frac{X_{(2)} - X_{(1)}}{X_{(n)} - X_{(1)}}$
	2.3.4.	Prueba para detectar si $X_{(n)}$ es aberrante	$\frac{X_{(n)} - X_{(n-1)}}{X_{(n)} - X_{(1)}}$
SHAPIRO Y WILK *	2.3.5.	Prueba para detectar si $X_{(1)}$ y/o $X_{(n)}$ es aberrante	$\frac{n(\bar{X} - X_{(1)})^2}{(n-1) \sum_{i=1}^n (X_i - \bar{X})^2}$
TIKU	2.3.1.	Prueba para detectar si $X_{(1)}$ es aberrante	$\frac{\frac{1}{n-1} \sum_{i=2}^n (X_i - X_{(1)})}{\frac{1}{n} \sum_{i=1}^n (X_i - X_{(1)})}$
	2.3.2.	Prueba para detectar si $X_{(n)}$ es aberrante	$\frac{\frac{1}{n-1} \sum_{i=1}^{n-1} (X_i - X_{(1)})}{\frac{1}{n} \sum_{i=1}^n (X_i - X_{(1)})}$

* Las estadísticas que proponen sólo detectan si existen aberrantes en la muestra, más no las identifican.

CAPITULO III

UN ENFOQUE GENERAL PARA LA DETECCION

DE OBSERVACIONES ABERRANTES

En este capítulo tratamos un enfoque general para la construcción de pruebas de detección de observaciones aberrantes. Utilizando este enfoque se han reconstruido, para diferentes distribuciones, criterios que han probado ser los más potentes.

Este enfoque se basa en los estimadores de la función de distribución, por lo que en la primera sección se discute un método para estimar insesgadamente la función de distribución utilizando los Teoremas de Rao-Blackwell y Lehmann-Scheffé. Estos estimadores se han utilizado principalmente en Teoría de Confiabilidad para distribuciones de tiempos de falla de un sistema y en estimación de la distribución de una observación futura en Regresión¹.

En la segunda sección tratamos los pasos a seguir para la construcción de este enfoque y finalmente en la tercera obtenemos el estimador insesgado y de varianza mínima para el caso de una variable aleatoria exponencial con 2 parámetros y se construye el criterio para este caso.

1 O'Reilly (1975)

3.1. Estimadores Insesgados de Varianza Mínima de Funciones de distribución

La estimación de la función de distribución la determinamos, - buscando dentro de la clase de estimadores insesgados aquél - que tenga varianza mínima utilizando los siguientes dos Teoremas:

Teorema de Rao-Blackwell

Sean X_1, X_2, \dots, X_n una muestra aleatoria de la densidad $f(x; \theta)$ y sea $S_1 = S_1(X_1, \dots, X_n), \dots, S_k = S_k(X_1, \dots, X_n)$ un conjunto de estadísticas conjuntamente suficientes. Sea la estadística $T = t(X_1, \dots, X_n)$ un estimador insesgado de $\gamma(\theta)$. Definiendo T' por:

$$T' = E [T/S_1, \dots, S_k] \quad \text{entonces}$$

1. T' es una estadística y es función de las estadísticas suficientes S_1, \dots, S_k

$$T' = t'(S_1, \dots, S_k)$$

2. $E_{\theta} [T'] = \gamma(\theta)$. T' es un estimador insesgado de $\gamma(\theta)$.

$$3. \text{Var}_\theta(T') \leq \text{Var}_\theta(T) \quad \forall \theta$$

Teorema de Lehmann-Scheffé

Sea X_1, \dots, X_n una muestra aleatoria de una densidad $f(x; \theta)$. Si $S = s(X_1, \dots, X_n)$ es una estadística suficiente y completa y si $T^* = t^*(s)$, una función de S , es un estimador insesgado de $\mathcal{T}(\theta)$, entonces T^* es el único estimador insesgado de varianza mínima (EIVM) de $\mathcal{T}(\theta)$.

Por lo tanto si existe una estadística suficiente y completa, entonces existe el EIVM de $F(x; \theta)$, el cual obtenemos del cálculo de la esperanza condicional del estimador insesgado dada la estadística suficiente y completa.

Un estimador insesgado de $F(x; \theta)$ es la función indicadora:

$$I_{[X \leq x]} = \begin{cases} 1 & \text{Si } X \leq x \\ 0 & \text{En otro caso} \end{cases}$$

ya que

$$\begin{aligned} E[I_{[X \leq x]}] &= P[X \leq x] \\ &= F(x; \theta) \end{aligned}$$

Por lo que para obtener el EIVM de $F(x; \theta)$ utilizando el Teorema de Rao-Blackwell y siendo T la estadística suficiente y completa, tenemos:

$$\begin{aligned} E[I_{[X \leq x]} / T] &= P[X \leq x / T] \\ &= \widehat{P[X \leq x]} \\ &= \tilde{F}_X(x; \theta) \quad \forall x \end{aligned}$$

Por lo tanto $E[I_{[X \leq x]} / T]$ es el estimador insesgado y de varianza mínima de $F(x; \theta)$ que denotaremos por $\tilde{F}_X(x; \theta)$.

Como ejemplo obtendremos la función estimadora de la función de distribución del siguiente modelo:

Sean X_1, X_2, \dots, X_n v.a.i.i.d. según una población Normal con media μ y varianza σ^2 conocida. La estadística suficiente y completa para μ es

$$T = \sum_{i=1}^n X_i$$

$$\tilde{F}(X_1; \mu) = P(X_1 \leq x/T) = \sum_{i=1}^n X_i$$

$$= \int_{-\infty}^x \frac{f_{X_1/T}(x_1, t)}{f_T(t)} dx_1$$

donde:

$$f(x_1, t) = \left(\frac{1}{\sqrt{2\pi}} \right)^2 \frac{1}{\sqrt{n-1}\sigma^2} e^{-\left\{ \frac{1}{2\sigma^2} \frac{1}{n-1} (n(x_1-\mu)^2 + (t-n\mu)^2 - 2(x_1-\mu)(t-n\mu)) \right\}}$$

$$f(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n}\sigma} e^{-\left\{ \frac{1}{2\sigma^2} \frac{1}{n} (t-n\mu)^2 \right\}}$$

y efectuando el cociente tenemos que:

$$\tilde{F}(x_1; \mu) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \frac{\sqrt{n}}{\sqrt{n-1}} \frac{1}{\sigma} e^{-\left\{ \frac{1}{2\sigma^2} \frac{n}{n-1} (x_1 - \bar{X})^2 \right\}} dx_1$$

Por lo tanto $\tilde{F}(X_1; \mu)$ es una distribución Normal con media \bar{X} y varianza $\frac{n-1}{n} \sigma^2$.

Este método para estimar la función de distribución no es el único. Existen otros que también pueden ser aprovechados como son los propuestos por: Sathe, Y.S. y Varde, S.D. (1969); Tate (1959), Abbey, S.L. y David, H.T. (1970) y para el caso multivariado: Ghurye y Olkin (1969).

3.2. Enfoque Propuesto

Un camino razonable a seguir para detectar si existen observaciones aberrantes es el de observar el comportamiento de cada X_1 en relación a su función de distribución.

Sean X_1, X_2, \dots, X_n variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) según $F(x; \theta)$ $\theta \in \Theta$.

Deseamos verificar estadísticamente las hipótesis:

H_0 : Todas las observaciones provienen del esquema hipotetizado.

H_a : Alguna X_1 pertenece a otra población.

El procedimiento que a continuación detallaremos se resume en los siguientes tres pasos:

1. Dado el esquema de inferencia se obtiene la estadística suficiente minimal, si es completa es mejor ya que se puede obtener el único EIVM, y se identifica la estimadora de la función de distribución $\tilde{F}(x; \theta)$.

Es importante notar que si X_1, X_2, \dots, X_n son idénticamente distribuidas, se obtendrán para cada X_i las funciones estimadoras $\tilde{F}(X_i; \theta)$.

2. Se calcula el intervalo de longitud mínima $\tilde{I}_{1-\frac{\alpha}{n}}$ de probabilidad condicional $1-\frac{\alpha}{n}$ de $\tilde{F}(x; \theta)$, $\theta \in \Theta$.
3. Se compara cada X_i con el intervalo $\tilde{I}_{1-\frac{\alpha}{n}}$ y por consecuencia decimos que X_i es aberrante sí y sólo si $X_i \notin \tilde{I}_{1-\frac{\alpha}{n}}$.

Justificación y comentarios al procedimiento

Como un primer paso está el determinar la función estimadora de $F(x; \theta)$ procedimiento que ya analizamos en la primera sección.

Una vez identificada la función estimadora de $F(x; \theta)$, construimos un intervalo de longitud mínima de probabilidad $1-\frac{\alpha}{n}$ y contra el cual compararemos todas las observaciones. En caso de que las variables no sean idénticamente distribuidas, se construyen los intervalos $\tilde{I}_{1-\frac{\alpha}{n}}$, X_i para cada $\tilde{F}(x_i; \theta)$. La razón por la que la probabilidad del intervalo es $1-\frac{\alpha}{n}$ la explicamos a continuación.

Como $\tilde{F}(x; \theta)$ es una distribución condicionada a T y es a par---

tir de ella que se construye el intervalo $\tilde{I}_{1-\frac{\alpha}{n}}$, se tiene -
que:

$$P\left[X_i \in \tilde{I}_{1-\frac{\alpha}{n}} / T\right] = P\left[X_i \in \tilde{I}_{1-\frac{\alpha}{n}}\right] = 1 - \frac{\alpha}{n}$$

Ahora definimos

$$A = \bigcap_{i=1}^n (X_i \in \tilde{I}_{1-\frac{\alpha}{n}})$$

es decir, que todas las observaciones caen en $\tilde{I}_{1-\frac{\alpha}{n}}$, por lo tanto

$$A^c = \bigcup_{i=1}^n (X_i \notin \tilde{I}_{1-\frac{\alpha}{n}})$$

equivale a decir que alguna X_i no está en dicho intervalo por lo que consideramos a A^c como una región crítica para probar - las hipótesis establecidas anteriormente con una probabilidad de error tipo I de a lo más α ya que

$$\begin{aligned} P(A^c) &= P\left(\bigcup_{i=1}^n (X_i \notin \tilde{I}_{1-\frac{\alpha}{n}})\right) \\ &\leq \sum_{i=1}^n P(X_i \notin \tilde{I}_{1-\frac{\alpha}{n}}) \\ &= \alpha \end{aligned}$$

A manera de validación del procedimiento se obtiene:

$$P(A^c) \geq \sum_{i=1}^n P(X_i \notin \tilde{I}_{1-\frac{\alpha}{n}}) - \sum_{i \neq j} P(X_i \notin \tilde{I}_{1-\frac{\alpha}{n}}) \cap (X_j \notin \tilde{I}_{1-\frac{\alpha}{n}})$$

con lo cual obtenemos la cota inferior y podemos ver que tan buena es la aproximación.

Siguiendo el ejemplo tratado en la sección anterior, obtenemos intervalos de la forma:

por lo que,
$$\left\{ \frac{|X_i - \bar{X}|}{\sigma} \leq c \right\}$$

$$\begin{aligned} P(A^c) &= P\left(\bigcup_{i=1}^n (X_i \notin \tilde{I}_{1-\frac{\alpha}{n}})\right) \\ &= P\left(\bigcup_{i=1}^n \left(\frac{|X_i - \bar{X}|}{\sigma} \leq c\right)^c\right) \\ &= P\left(\bigcup_{i=1}^n \left(\frac{|X_i - \bar{X}|}{\sigma} > c\right)\right) \\ &= P\left(\max_i \left(\frac{|X_i - \bar{X}|}{\sigma} > c\right)\right) \end{aligned}$$

En este caso las tablas para obtener el valor c son conocidas.

3.3. Aplicación al caso exponencial

La construcción de este enfoque para el caso exponencial está basada en la estimación insesgada de funciones de distribución como antes ya habíamos mencionado, la cual encontraremos a continuación.

Sean X_1, X_2, \dots, X_n v.a.i.i.d. según una población exponencial con parámetros m y c , cuya f.d.p. es:

$$f_X(x; m, c) = \begin{cases} c^{-1} e^{-\frac{1}{c}(x-m)} & x \geq m \\ 0 & \text{otro caso} \end{cases}$$

Dada la estadística suficiente y completa $T = (X_{(1)}, \bar{X} - X_{(1)})$ tenemos que:

$$\begin{aligned} \tilde{F}(X_n; m, c) &= P(X_n \leq x / T) \\ &= P(X_n \leq x / X_{(1)}, \bar{X} - X_{(1)}) \\ &= \int \frac{f(X_n, X_{(1)}, \bar{X} - X_{(1)})}{f(X_{(1)}, \bar{X} - X_{(1)})} dx_n \end{aligned}$$

Calculando la función de densidad conjunta $f(X_{(1)}, y)$ donde $y = \bar{X} - X_{(1)}$ obtenemos:

$$f(x_{(1)}, y) = f(x_{(1)}) f(y) \\ = \frac{\left(\frac{n}{c}\right)^n e^{-\frac{n}{c}[(x_{(1)}-m)+y]} y^{n-2}}{\Gamma(n-1)} \quad (3.3.1.)$$

basándonos en la independencia de $X_{(1)}$ y $\bar{X} - X_{(1)}$ y en los resultados (2.1.6.) y (2.1.7.) que son las densidades de \hat{m} y \hat{c} .

La función de densidad conjunta de $X_{(1)}$, $\bar{X} - X_{(1)}$ la obtenemos mediante el siguiente procedimiento:

Sea $X = (X_1, X_2, \dots, X_n)$ una muestra de tamaño n de una variable aleatoria exponencial negativa y sea $X^* = (X_1, X_2, \dots, X_{n-1})$ una muestra de X de tamaño $n-1$.

Calculamos la densidad $f(X_n, X_{(1)}^*, \bar{X}^* - X_{(1)}^*)$ donde X_n , $X_{(1)}^*$ y $\bar{X}^* - X_{(1)}^*$ son independientes.

$$f(x_n, x_{(1)}^*, \bar{x}^* - x_{(1)}^*) = \frac{\frac{1}{c} e^{-\frac{1}{c}(x_n - m)}}{\Gamma(n-2)} \cdot \left(\frac{n-1}{c}\right)^{n-1} y^{n-3} e^{-\frac{n-1}{c}(x_{(1)}^* - m + y)}$$

donde $Y^* = \bar{X}^* - X_{(1)}^*$.

Para obtener la densidad $f(X_n, X_{(1)}, \bar{X} - X_{(1)})$ realizamos el siguiente cambio de variable:

$$Y^* = \frac{1}{n-1} \sum_{i=1}^{n-1} X_i - X_{(1)}^* \\ = \frac{n}{n-1} Y + \frac{1}{n-1} (X_{(1)} - X_n)$$

donde $Y = \bar{X} - X_{(1)}$.

Hay dos casos que considerar de acuerdo a si $X_{(1)}^*$ es mayor o igual que $X_{(1)}$.

Para el caso $X_{(1)}^* > X_{(1)}$ se tiene que $X_{(1)} = X_n$ y por lo tanto

$$f(X_n / X_{(1)}, Y) = \frac{1}{n}$$

Para el caso $X_{(1)}^* = X_{(1)}$ se tiene que $X_{(1)} = X_{(1)}^* < X_n$ y -- calculando el Jacobiano para este cambio de variable tenemos:

$$|J| = \begin{vmatrix} \frac{n}{n-1} & \frac{1}{n-1} & -\frac{1}{n-1} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

$$= \frac{n}{n-1}$$

Por lo tanto

$$f(X_n, X_{(1)}, Y) = \frac{\frac{1}{c} e^{-\frac{1}{c}(X_n - m)}}{\Gamma(n-2)} \cdot \left(\frac{n-1}{c}\right)^{n-1} \cdot \left[\frac{n}{n-1} Y + \frac{1}{n-1} (X_n - Y_n)\right]^{n-3}$$

$$e^{-\frac{n-1}{c} \left[(X_{(1)} - m) + \frac{n}{n-1} Y + \frac{1}{n-1} (X_{(1)} - Y_n) \right]} \cdot \frac{n}{n-1} \quad (3.3.2.)$$

Efectuando el cociente de densidades para obtener la función de densidad condicional $f(X_n / X_{(1)}, Y)$ se llega a:

$$f(X_n | X_{(1)}, Y) = \begin{cases} \frac{1}{n} & \text{si } X_n = X_{(1)} \\ (n-2) \left(1 - \frac{1}{n}\right) \left[1 - \frac{X_n - X_{(1)}}{nY}\right]^{n-3} \cdot \frac{1}{nY} & \text{si } X_{(1)} \leq X_n \leq X_{(1)} + nY \\ 0 & \text{otro caso} \end{cases} \quad (3.3.3)$$

De aquí que,

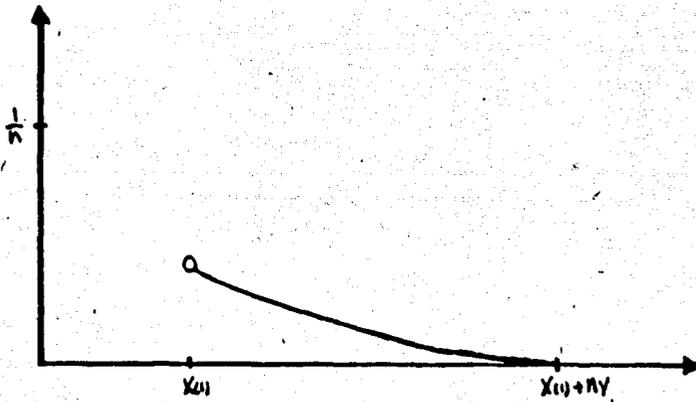
$$\begin{aligned} P[X \geq x | T] &= \int_x^{X_{(1)} + nY} f(x | X_{(1)}, Y - X_{(1)}) dx \\ &= \left(1 - \frac{1}{n}\right) \left[1 - \frac{x - X_{(1)}}{n(Y - X_{(1)})}\right]^{n-2} \Bigg|_x^{X_{(1)} + nY} \\ &= \left[1 - \frac{1}{n}\right] \left[1 - \frac{x - X_{(1)}}{n(Y - X_{(1)})}\right]^{n-2} \end{aligned}$$

y finalmente

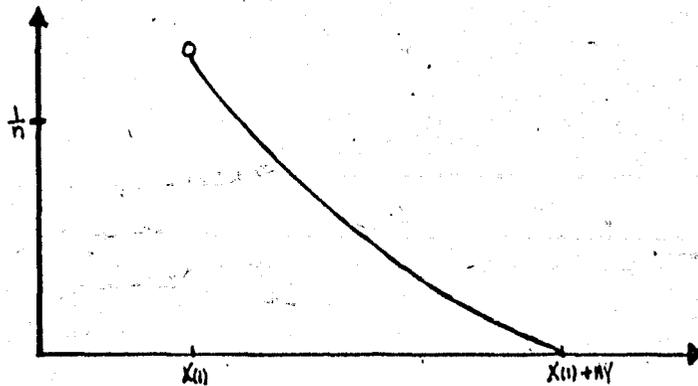
$$\tilde{F}(x) = P[X \leq x | T] = \begin{cases} 0 & \text{si } x < X_{(1)} \\ 1 - \left[1 - \frac{1}{n}\right] \left[1 - \frac{x - X_{(1)}}{n(Y - X_{(1)})}\right]^{n-2} & \text{si } X_{(1)} \leq x \leq X_{(1)} + nY \\ 1 & \text{si } x > X_{(1)} + nY \end{cases} \quad (3.3.4.)$$

Gráficas de la densidad condicional (3.3.3.) para algunos valores de n , Y y $X_{(1)}$ están representadas por las siguientes figuras:

$n = 5$
 $Y = 5$
 $X_{(1)} = 2$



$n = 10$
 $Y = 5$
 $X_{(1)} = 2$



Esta función de densidad es decreciente en el intervalo
 $(X_{(1)}, X_{(1)} + nY]$

Podemos encontrar los intervalos de longitud mínima a partir de (3.3.4.)

$$P(X \leq x/T) = 1 - \left(1 - \frac{x - X_{(1)}}{nY}\right)^{n-1} = 1 - \frac{x - X_{(1)}}{nY}$$

y son de la forma:

$$\left\{ X: \frac{|X - X_{(1)}|}{n(\bar{X} - X_{(1)})} \leq c \right\} \quad (3.3.5.)$$

por lo tanto,

$$\begin{aligned} P(A^c) &= P\left(\bigcup_{i=1}^n \left| \frac{X_i - X_{(1)}}{n(\bar{X} - X_{(1)})} \right| > c\right) \\ &= P\left(\max_i \left| \frac{X_i - X_{(1)}}{n(\bar{X} - X_{(1)})} \right| > c\right) \\ &= P\left(\frac{X_{(n)} - X_{(1)}}{n(\bar{X} - X_{(1)})} > c\right) \end{aligned}$$

Es importante notar que el criterio conduce a identificar a $X_{(n)}$ como la posible observación aberrante, lo cual intuitivamente era de esperarse. Explícitamente podríamos tratar de obtener la distribución de:

$$U_{(n)} = \frac{X_{(n)} - X_{(1)}}{n(\bar{X} - X_{(1)})}$$

que es la estadística de Laurent. En nuestro procedimiento, lo que hacemos es observar si $X_i \in \tilde{I}_1 - \frac{\alpha}{n}$, y no necesitamos encontrar la distribución de $U_{(n)}$, sino que utilizamos una aproximación a ésta.

Para encontrar el valor de c de la expresión (3.3.5.) evaluamos la siguiente probabilidad:

$$p(X \leq x / X_{(1)}, \bar{X} - X_{(1)}) = 1 - \frac{\alpha}{n}$$

Por lo tanto

$$\left(1 - \frac{1}{n}\right) \left[1 - \frac{X - X_{(1)}}{n(\bar{X} - X_{(1)})}\right]^{n-2} = \frac{\alpha}{n}$$

de donde,

$$\frac{X - X_{(1)}}{n(\bar{X} - X_{(1)})} = 1 - \exp \left[\frac{1}{n-2} \ln \frac{\alpha}{n-1} \right] \quad (3.3.6.)$$

Es importante notar que los valores críticos para la distribución de $U_{(n)} = \frac{X_{(n)} - X_{(1)}}{n(\bar{X} - X_{(1)})}$ de la propuesta de Laurent dados en la tabla VIII del Apéndice se pueden obtener de la relación (3.3.6.).

Por lo tanto esta aproximación reproduce la tabla de Laurent y no es necesario usar la distribución de la estadística $U_{(n)}$.

Con este procedimiento solo podemos detectar observaciones aberrantes grandes. Si la observación aberrante proviene de una variable exponencial con parámetros c y $m + \delta c$ ($\delta > 0$) el criterio no la detecta. Esto es un limitante al procedimiento propuesto y por lo tanto restringe la aplicabilidad del mismo. Sin embargo es posible modificarlo analizando los intervalos de longitud mínima. Pensamos que la información para poder decidir si $X_{(1)}$ es aberrante, básicamente está contenida en la distribución condicional por lo siguiente:

Notemos que

$$\frac{X_{(2)} - X_{(1)}}{n (\bar{X} - X_{(1)})}$$

es una medida de la discrepancia entre $X_{(2)}$ y $X_{(1)}$. La función de distribución condicional evaluada en $x_{(2)}$

$P. (x_{(1)} < X \leq x_{(2)} / T)$ nos da información acerca de esta discrepancia. Este punto no lo tratamos en esta tesis.

En el caso en que el parámetro m sea conocido se puede ver fácilmente que el procedimiento propuesto conduce a rechazar H_0 si

$$\frac{X_{(n)}}{\sum_{i=1}^n X_i} > c$$

y aquí no se presentan los problemas mencionados en el párrafo anterior. La estadística antes mencionada es la que más se ha utilizado cuando el parámetro m es conocido.

CAPITULO IV

4. RESULTADOS

4.1. Comparación de Simulaciones

Con el fin de hacer comparaciones entre los criterios expuestos en el Capítulo anterior, realizamos simulaciones de tal manera que los resultados pudieran ser examinados entre sí.

Para la simulación se generaron muestras de la variable aleatoria exponencial utilizando el siguiente Teorema¹:

"Sea X es una variable aleatoria con función de distribución $F_X(x)$ y $Y = h(X)$.

La función de distribución de Y está dada por

$$G_Y(y) = \int_{h^{-1}(T_y)} dF_X(x)$$

donde $T_y = \{ Y/Y \leq y \}$."

Corolario:

Sea X una variable aleatoria con función de distribución $F_X(x)$.

Sea $h(X) = F_X(X) = y$.

1 Harris (1966)

La función de distribución de Y está dada por

$$G_Y(y) = F_X(F_X^{-1}(y)) = y \quad 0 \leq y \leq 1$$

es decir, Y tiene una distribución uniforme en el intervalo (0, 1).

Aplicando estos resultados para el caso exponencial tenemos:

$$\begin{aligned} G_Y(y) &= P(Y \leq y) = F_X(F_X^{-1}(y)) \quad 0 \leq y \leq 1 \\ &= P[X \leq F_X^{-1}(y)] \\ &= P[X \leq -c \ln(1 - y) + m] \end{aligned}$$

De donde,

$$F_X^{-1}(y) = -c \ln(1 - y) + m \quad (4.1.1.)$$

tiene una distribución exponencial negativa. De esta manera, conociendo el valor de los parámetros m y c y generando valores para una población uniforme en (0, 1), se pueden obtener muestras pertenecientes a una población exponencial negativa.

Para la generación de aberrantes en la simulación, consideremos 2 tipos de contaminación, uno con cambios en el parámetro

de localización y otro con cambios en el parámetro de escala.

Esquema 1

X_1, X_2, \dots, X_{n-1} provienen de una población con parámetro de localización m y parámetro de escala c .

X_n proviene de una población con parámetro de localización $m + \delta c$ y mismo parámetro de escala c , con $\delta \neq 0$.

Esquema 2

X_1, X_2, \dots, X_{n-1} provienen de una población con parámetro de localización m y parámetro de escala c .

X_n proviene de una población con parámetro de escala δc ($\delta > 1$) y mismo parámetro de localización m .

Para encontrar el número apropiado del total de simulaciones o repeticiones del proceso queremos que bajo la hipótesis nula - el verdadero valor del nivel de significancia α sea aproximadamente igual al nivel de significancia estimado $\hat{\alpha}$, para esto controlamos el rango de variación de $\hat{\alpha}$ con respecto a α dando una precisión deseada d y una probabilidad alta γ a este evento, es decir,

$$\Pr \left[\left| \hat{\alpha} - \alpha \right| \leq d \right] = \gamma \quad (4.1.2.)$$

Donde

- $\hat{\alpha}$ nivel de significancia estimado
- α nivel de significancia real
- d precisión deseada

El nivel de significancia estimada lo obtuvimos de la siguiente consideración:

Sea X_i es una variable aleatoria tal que

$$X_i = \begin{cases} 0 & \text{si no se rechaza la hipótesis en la } i\text{-ésima} \\ & \text{realización de la prueba bajo interés} \\ 1 & \text{si se rechaza la hipótesis} \end{cases}$$

Ya que X es una variable aleatoria Bernoulli con parámetro $p = \alpha$, el estimador de α está dado por :

$$\hat{\alpha} = \sum_{i=1}^R \frac{X_i}{R}$$

donde R es el número de repeticiones.

Por lo tanto,

$$P_1 \left[\left| \hat{\alpha} - \alpha \right| \leq d \right] = P_1 \left[\left| \sum \frac{X_i}{R} - \alpha \right| \leq d \right]$$

$$= \Pr \left[\left| \sum X_i - Rd \right| \leq Rd \right]$$

y utilizando la aproximación a una distribución normal (0, 1)

$$\begin{aligned} \Pr \left[\left| \sum X_i - Rd \right| \leq Rd \right] &= \Pr \left[\left| \frac{\sum X_i - Rd}{\sqrt{Rd(1-d)}} \right| \leq \frac{Rd}{\sqrt{Rd(1-d)}} \right] \\ &= \Pr \left[|z| \leq \frac{Rd}{\sqrt{Rd(1-d)}} \right] \end{aligned}$$

Con una precisión $d = .005$, una probabilidad $\gamma = .95$ y $d = .05$ tenemos:

$$P \left[|z| \leq \frac{R(.005)}{\sqrt{R(.05)(.95)}} \right] = .95$$

De aquí que

$$\frac{R(.005)}{\sqrt{R(.05)(.95)}} = 1.96$$

Por lo tanto el tamaño total de repeticiones del proceso es $R = 7299$. Para la realización de las simulaciones tomamos $R = 8000$.

Las simulaciones realizadas, están enfocadas principalmente a medir la potencia de las pruebas, pero como se dijo anterior--

mente, también se midieron otros conceptos.

Sea T la estadística de interés con t_α el valor crítico de T a un nivel de significancia α .

H_0 la hipótesis nula y

H_a la hipótesis alternativa, de la cual consideramos los siguientes tipos:

1. H_a : $X_{(1)}$ no pertenece a la población bajo estudio
2. H_a : $X_{(n)}$ no pertenece a la población bajo estudio
3. Alguna X_i no pertenece al esquema propuesto.

Los conceptos que se midieron en la simulación son:

1. FUNCION POTENCIA

Probabilidad de rechazar H_0 dado que es falsa

$$P_1 = \Pr \left[T > t_\alpha / H_a \right] \quad (4.1.3.)$$

2. NIVEL DE SIGNIFICANCIA

El nivel de significancia lo obtuvimos como control con el objeto de revisar el comportamiento de la -

prueba cuando las muestras no contenían aberrantes.

Probabilidad de rechazar H_0 dado que es cierta

$$P_2 = \Pr \left[T > t_d / H_0 \right] \quad (4.1.4.)$$

3. PROBABILIDAD DE RECHAZAR H_0 Y DETECTAR LA OBSERVA---
CION CONTAMINANTE

Este caso es diferente a la función potencia pues se puede rechazar H_0 dado que es falsa y no detectar la observación que proviene de otra población.

La medición de estos conceptos se obtuvo mediante la ayuda de los siguientes 3 contadores :

1. El primero controló el número de veces que la prueba bajo interés detectó correctamente una aberrante dado que la muestra si contenía.
2. El segundo contador llevó cuenta del número de veces que la prueba utilizada detectó incorrectamente una aberrante dado que la muestra sí contenía.
3. Finalmente, el último que se utilizó para contar el número de veces que la prueba bajo interés no detectó la abe-

rrante dado que la muestra sí contenía.

La potencia de la prueba la medimos a partir de la suma del - contenido de los 2 primeros contadores entre el total de repeticiones, el nivel de significancia a partir de la suma de estos 2 contadores también pero cuando no generamos aberrantes, y el error tipo II a partir del tercer contador entre el total de repeticiones.

A continuación anexamos las Tablas y Gráficas de la potencia - de la prueba para todos los criterios.

POTENCIA DE LA PRUEBA

Esquema 1

Muestras generadas con parámetros $m=0$ y $c=1$

Observación aberrante generada con parámetro de localización $m=\int c$ y parámetro de escala $c=1$

N=5

	<u>TIKU</u>	<u>LIKES</u>	<u>O' REILLY</u>	<u>SHAPIRO</u>
0	.04838	.05050	.04913	.04988
1	.03025	.02962	.03175	.04775
2	.06275	.06100	.05988	.04488
3	.11538	.11238	.11663	.06625
4	.17988	.17650	.18100	.10813
5	.25475	.24388	.25988	.15275

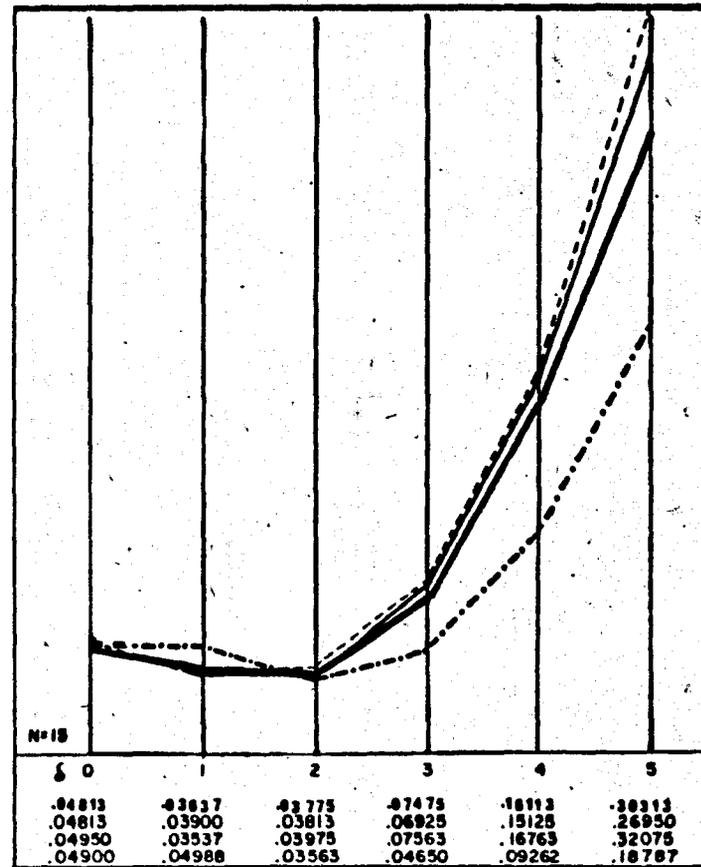
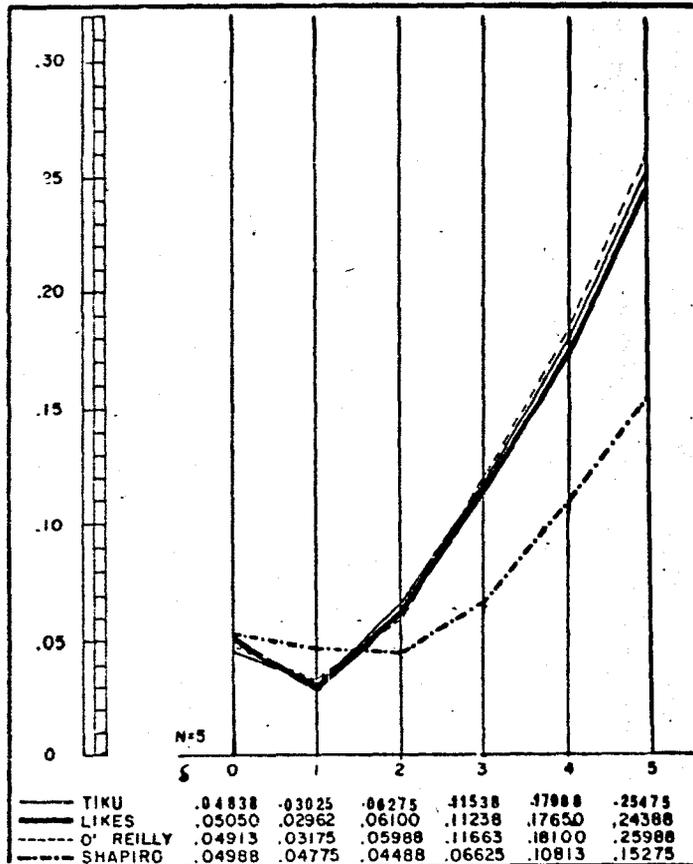
N=15

	<u>TIKU</u>	<u>LIKES</u>	<u>O' REILLY</u>	<u>SHAPIRO</u>
0	.04813	.04813	.04950	.04900
1	.03637	.03900	.03537	.04988
2	.03775	.03813	.03975	.03563
3	.07475	.06925	.07563	.04650
4	.16113	.15125	.16763	.09262
5	.30313	.26950	.32075	.18787

POTENCIA DE LA PRUEBA

ESQUEMA I

MUESTRAS GENERADAS CON PARAMETROS $m=0$ Y $c=1$
 OBSERVACION ABERRANTE GENERADA CON PARAMETRO DE
 LOCALIZACION $m + \delta c$ Y PARAMETRO DE ESCALA $c=1$



POTENCIA DE LA PRUEBA

Esquema 1

Muestras generadas con parámetros $m=0$ y $c=1$
observación aberrante generada con parámetro de localización
 $m= c$ y parámetro de escala $c=1$.

N=5

	<u>TIKU</u>	<u>LIKES</u>	<u>SHAPIRO</u>
-5	.97100	.96138	.87775
-4	.91900	.90263	.77063
-3	.81000	.78088	.60388
-2	.56563	.53863	.36763
-1	.21463	.21050	.12925
0	.05225	.05150	.04988

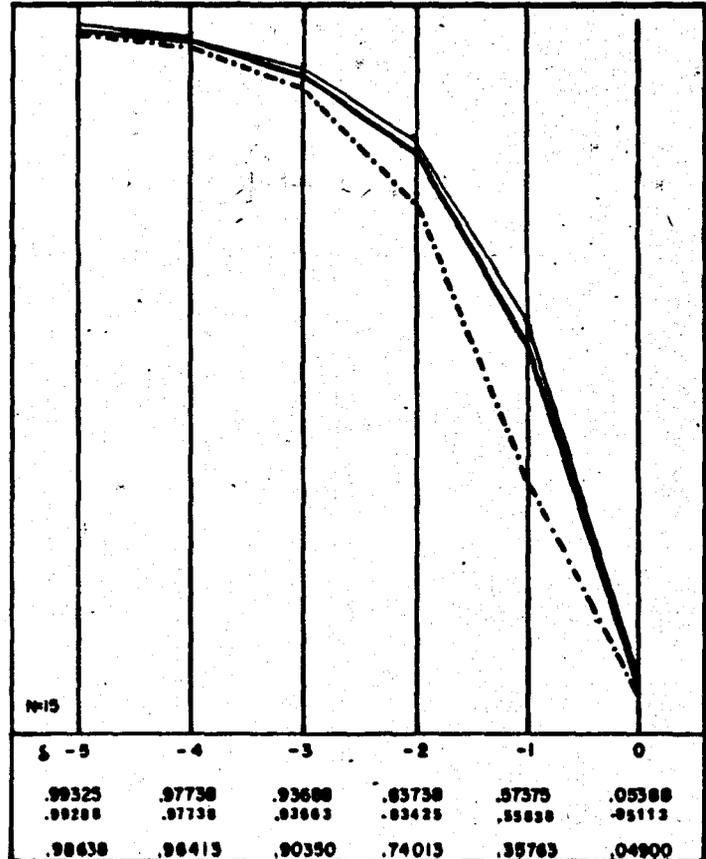
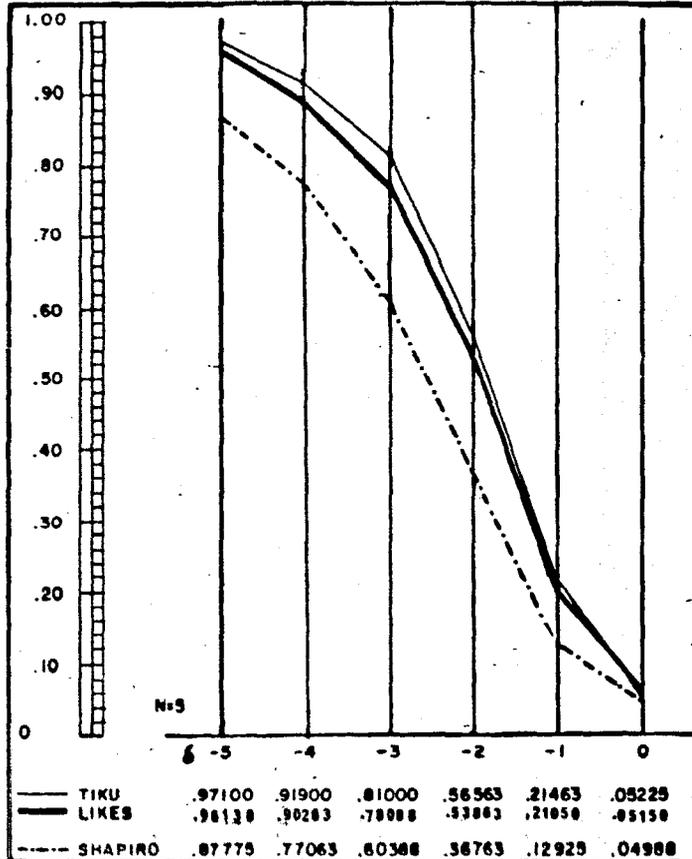
N=15

	<u>TIKU</u>	<u>LIKES</u>	<u>SHAPIRO</u>
-5	.99325	.99288	.98638
-4	.97738	.97738	.96413
-3	.93688	.93663	.90350
-2	.83738	.83425	.74013
-1	.57375	.55838	.35763
0	.05388	.05113	.04900

POTENCIA DE LA PRUEBA

ESQUEMA I

MUESTRAS GENERADAS CON PARAMETROS $m=0$ Y $c=1$
 OBSERVACION ABERRANTE GENERADA CON PARAMETRO DE
 LOCALIZACION $m + bc$ Y PARAMETRO DE ESCALA $c=1$



POTENCIA DE LA PRUEBA

Esquema 2

Muestras generadas con parámetros $m=0$ y $c=1$

Observación aberrante generada con parámetro de localización $m=0$ y parámetro de escala δ .

N=5

δ	<u>TIKU</u>	<u>LIKES</u>	<u>O' REILLY</u>	<u>SHAPIRO</u>
1	.04838	.05050	.04913	.04988
2	.07138	.07013	.07163	.06563
3	.11838	.11475	.11838	.09488
4	.17175	.16512	.17525	.12950
5	.22250	.21387	.22525	.17275
6	.27112	.25837	.27388	.20900
7	.31275	.29962	.31488	.24825
8	.34625	.33663	.34988	.28250
9	.38100	.37088	.38200	.31438
10	.41488	.40313	.41800	.34388
11	.44850	.43363	.45150	.37063

POTENCIA DE LA PRUEBA

Esquema 2

Muestras generadas con parámetros $m=0$, $c=1$

Observación aberrante generada con parámetro de localización $-m=0$ y parámetro de escala $c\delta$.

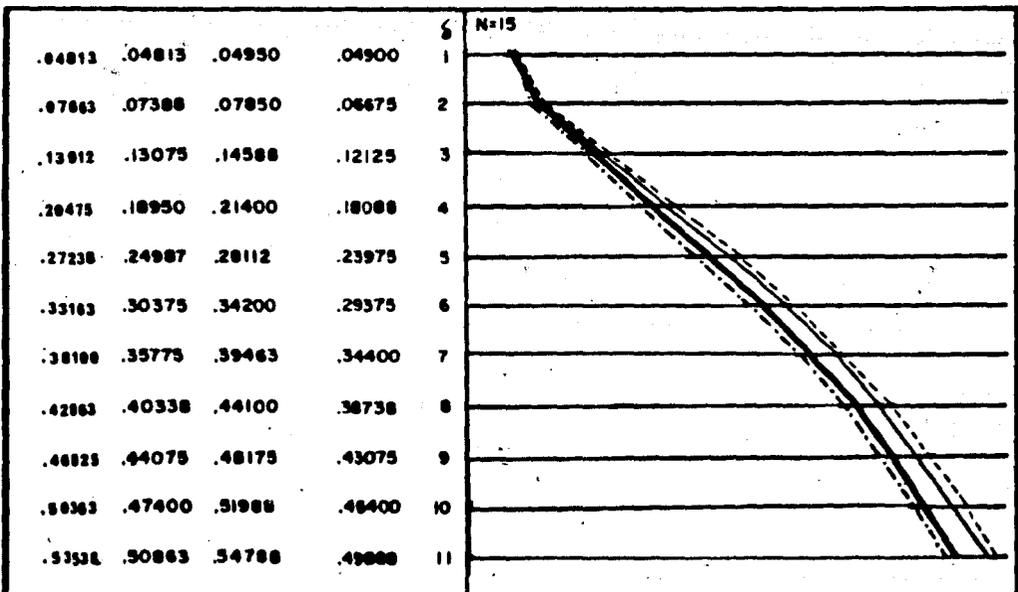
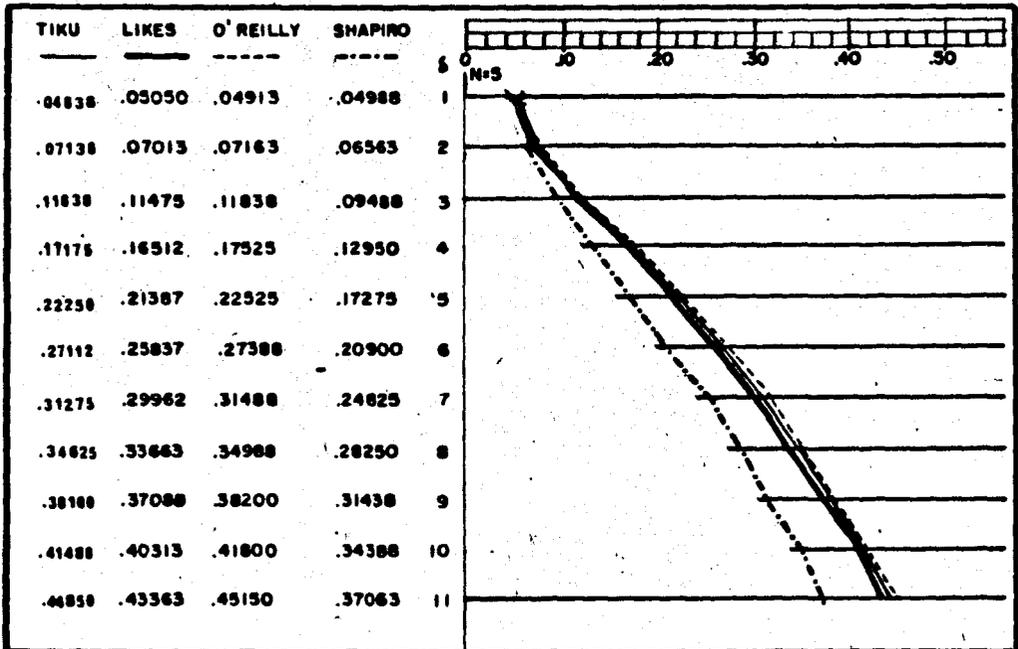
N=15

δ	<u>TIKU</u>	<u>LIKES</u>	<u>O' REILLY</u>	<u>SHAPIRO</u>
1	.04813	.04813	.04950	.04900
2	.07663	.07388	.07850	.06675
3	.13912	.13075	.14588	.12125
4	.20475	.18950	.21400	.18088
5	.27238	.24987	.28112	.23975
6	.33163	.30375	.34200	.29375
7	.38100	.35775	.39463	.34400
8	.42863	.40338	.44100	.38738
9	.46825	.44075	.48175	.43075
10	.50363	.47400	.51988	.46400
11	.53538	.50863	.54788	.49888

POTENCIA DE LA PRUEBA

ESQUEMA 2

MUESTRAS GENERADAS CON PARAMETROS $m=0$ Y $\sigma=1$
 OBSERVACION ABERRANTE GENERADA CON PARAMETRO DE LOCALIZACION $m=0$ Y PARAMETRO DE ESCALA $c=5$



4.2 Resultados

Para la detección de observaciones aberrantes grandes en muestras generadas con el esquema 1 (modificación del parámetro de localización), obtuvimos los siguientes resultados:

Para el criterio de Shapiro encontramos que el nivel de significancia estimado para muestras de tamaño $N=5$ es $\hat{\alpha} = 0.04988$ y para muestras de tamaño $N=15$ es $\hat{\alpha} = 0.049$ cuya diferencia con el valor real de $\alpha = .05$ es pequeña. Estadísticamente esto significa que la prueba de hipótesis $H_0: \alpha = .05$ no se rechaza.

Este criterio no resulta ser tan potente como los criterios de Likes, Tiku y el de O'Reilly.

En el criterio de Likes encontramos que para los diferentes tamaños de muestra 5 y 15 el valor del nivel de significancia estimado es muy parecido al real y su potencia para un valor de $\alpha = 5$ y $N = 15$ es 0.2695.

El criterio de Tiku demostró tener para muestras de tamaño 5 y 15 un nivel de $\hat{\alpha} = 0.04838$ y $\hat{\alpha} = 0.04813$

respectivamente, por lo que la región crítica podemos decir que es de tamaño α al igual que en el criterio de Likes.

Finalmente el criterio de O'Reilly demostró ser el mejor que los anteriores criterios en este caso. Para éste, el valor estimado de α es $\hat{\alpha} = 0.0495$ y la potencia para $\delta = 5$ y $N = 15$ es 0.32075. Este criterio también demuestra tener una región crítica de tamaño α .

Para la detección de aberrantes chicas en muestras generadas con el esquema 1 obtuvimos los siguientes resultados:

El criterio de Shapiro mejora para la detección de aberrantes chicas. El nivel de significancia para muestras de tamaño $N = 15$ es $\hat{\alpha} = .049$ que estadísticamente es muy parecido a α , y la potencia para $\delta = -5$ y $N = 15$ es 0.98638 que casi es la potencia máxima.

Para el criterio de Likes encontramos que el nivel de significancia estimado para el caso de detección de aberrantes chicas, en muestras de tamaño $N = 15$ es $\hat{\alpha} = 0.05113$ y la potencia para $\delta = -5$ y $N = 15$ es 0.99288 que al igual que la potencia de Shapiro es bastante alta.

El criterio de Tiku resulta con un nivel de significancia estimado $\hat{\alpha} = 0.05388$ para $N = 15$ que también es muy parecido a α ..

Para la detección de observaciones aberrantes grandes en muestras generadas con el esquema 2 (modificación del parámetro de escala), llegamos a los siguientes resultados:

Tanto el criterio de Shapiro, Likes, Tiku y O'Reilly tienen un nivel de significancia estimado que estadísticamente es muy cercano a α , por lo que podemos decir que estas pruebas tienen una región crítica de tamaño α .

Como observamos en las gráficas del esquema 2 en orden de menor a mayor potencia, los criterios siguen la siguiente secuencia: primero el de Shapiro, en segundo lugar el de Likes, en tercero el de Tiku y finalmente el de O'Reilly.

Estos criterios reportan una potencia para $d = 11$ y $N = 15$ de 0.49888 el de Shapiro, 0.50863 el de Likes, 0.44850 el de Tiku y 0.54788 el de O'Reilly.

CONCLUSIONES

CONCLUSIONES

El criterio de Laurent para detectar observaciones aberrantes chicas demuestra que el valor del nivel de significancia estimado es $\hat{\alpha} = .50013$, por lo que concluimos que esta prueba no tiene una región crítica de tamaño α , mas aún, - la potencia de la prueba no aumenta y por lo tanto este criterio se descarta para la detección de aberrantes chicas. Para la detección de aberrantes grandes, como hemos visto - coincide con el criterio de O'Reilly.

El criterio de Shapiro es la única prueba bilateral propuesta y si la hipótesis alternativa es que las aberrantes provienen del esquema 1, es decir, con cambios en el parámetro de localización m por $m + \delta c$ con $\delta \neq 0$, este criterio es bueno a pesar de que no tiene tanta potencia como otros.

El criterio de O'Reilly está implicando una hipótesis alternativa que descarta la posibilidad de detección de aberrantes chicas como ya se vió en el capítulo anterior.

Para la detección de aberrantes grandes y si la hipótesis alternativa está especificada es decir, las aberrantes son generadas por cambios en el parámetro de localización m por $m + \delta c$ ($\delta > 0$) o por cambios en el parámetro de escala c por δc ($\delta > 1$) es recomendable utilizar el criterio de O'Reilly.

Para la detección de aberrantes chicas donde éstas son generadas por cambios en el parámetro de localización m por $m + \delta c$ ($\delta < 0$) es recomendable el criterio de Tiku.

El criterio de Shapiro también se recomienda su uso para detectar aberrantes tanto grandes como chicas, al igual que - el de Likes.

Finalmente el único criterio que descartamos es el de
Laurent para la detección de aberrantes chicas.

APENDICE

APENDICE DE TABLAS

TABLA I	Tabla de Valores Críticos para la Prueba de - Pearson y Chandra Sekar en Muestras Normales.
TABLA II	Tabla de Valores Críticos para la Prueba de - Grubbs en Muestras Normales.
TABLA III	Tabla de Valores Críticos para la Prueba de - Dixon en Muestras Normales.
TABLA IV	Tabla de Valores Críticos para la Prueba de - Ferguson en Muestras Normales.
TABLA V	Tabla de Valores Críticos para la Prueba de - Tiku en Muestras Exponenciales.
TABLA VI	Tabla de Valores Críticos para la Prueba de - Dixon. Detección de Aberrantes Chicas en Muestras Exponenciales.
TABLA VII	Tabla de Valores Críticos para la Prueba de - Dixon. Detección de Aberrantes en Muestras Exponenciales.
TABLA VIII	Tabla de Valores Críticos para la Prueba de - Laurent para Muestras Exponenciales.
TABLA IX	Tabla de Valores Críticos para la Prueba de - Shapiro y Wilk para Muestras Exponenciales.

T A B L A I
VALORES CRITICOS PARA LA ESTADISTICA
DE PEARSON Y CHANDRA SEKAR ¹

n	Niveles de Significancia		
	5%	2.5%	1%
3	1.15	1.15	1.15
4	1.46	1.48	1.49
5	1.67	1.71	1.75
6	1.82	1.89	1.94
7	1.94	2.02	2.10
8	2.03	2.13	2.22
9	2.11	2.21	2.32
10	2.18	2.29	2.41
11	2.23	2.36	2.48
12	2.29	2.41	2.55
13	2.33	2.46	2.61
14	2.37	2.51	2.66
15	2.41	2.55	2.71
16	2.44	2.59	2.75
17	2.47	2.62	2.79
18	2.50	2.65	2.82
19	2.53	2.68	2.85
20	2.56	2.71	2.88
21	2.58	2.73	2.91
22	2.60	2.76	2.94
23	2.62	2.78	2.96
24	2.64	2.80	2.99
25	2.66	2.82	3.01
30	2.75	2.91	
35	2.82	2.98	
40	2.87	3.04	
45	2.92	3.09	
50	2.96	3.13	
60	3.03	3.20	
70	3.09	3.26	
80	3.14	3.31	
90	3.18	3.35	
100	3.21	3.38	

$$T_n = \frac{X_{(n)} - \bar{X}}{S}$$

$$T_1 = \frac{\bar{X} - X_{(1)}}{S}$$

¹ Grubbs (1969)

T A B L A II
VALORES CRITICOS PARA LA ESTADISTICA
DE GRUBBS¹

Niveles de Significancia				
n	1%	2.5%	5%	10%
3	.0001	.0007	.0027	.0109
4	.0100	.0248	.0494	.0975
5	.0442	.0808	.1270	.1984
6	.0928	.1453	.2032	.2826
7	.1447	.2066	.2696	.3503
8	.1948	.2616	.3261	.4050
9	.2411	.3101	.3742	.4502
10	.2831	.3526	.4154	.4881
11	.3211	.3901	.4511	.5204
12	.3554	.4232	.4822	.5483
13	.3864	.4528	.5097	.5727
14	.4145	.4792	.5340	.5942
15	.4401	.5030	.5559	.6134
16	.4634	.5246	.5755	.6306
17	.4848	.5442	.5933	.6461
18	.5044	.5621	.6095	.6601
19	.5225	.5785	.6243	.6730
20	.5393	.5937	.6379	.6848
21	.5548	.6076	.6504	.6958
22	.5692	.6206	.6621	.7058
23	.5827	.6327	.6728	.7151
24	.5953	.6439	.6829	.7238
25	.6071	.6544	.6923	.7319

$$\frac{s_n^2}{s^2} = \frac{\sum_{i=1}^{n-1} (x_{(i)} - \bar{x}_n)^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2}$$

$$\frac{s_1^2}{s^2} = \frac{\sum_{i=2}^n (x_{(i)} - \bar{x}_1)^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2}$$

1 Grubbs (1950)

TABLA III
VALORES CRITICOS PARA LAS ESTADISTICAS
DE DIXON¹

n	criterio	Niveles de significancia		
		10%	5%	1%
3	$r_{10} = \frac{X_2 - X_1}{X_n - X_1}$.886	.941	.988
4	$r_{10} = \frac{X_n - X_1}{X_n - X_{n-1}}$.679	.765	.889
5	$r_{10} = \frac{X_n - X_{n-1}}{X_n - X_1}$.557	.642	.780
6	$r_{10} = \frac{X_n - X_{n-1}}{X_n - X_1}$.482	.560	.698
7	$r_{10} = \frac{X_n - X_1}{X_n - X_1}$.434	.507	.736
8	$r_{11} = \frac{X_2 - X_1}{X_n - X_1}$.479	.554	.683
9	$r_{11} = \frac{X_2 - X_1}{X_n - X_1}$.441	.512	.635
10	$r_{11} = \frac{X_{n-1} - X_1}{X_n - X_{n-1}}$.409	.477	.597
	$r_{11} = \frac{X_n - X_{n-1}}{X_n - X_1}$			
	$r_{11} = \frac{X_n - X_2}{X_n - X_1}$			
11	$r_{21} = \frac{X_3 - X_1}{X_n - X_1}$.517	.576	.679
12	$r_{21} = \frac{X_{n-1} - X_1}{X_n - X_1}$.490	.546	.642
13	$r_{21} = \frac{X_n - X_{n-2}}{X_n - X_1}$.467	.521	.615
	$r_{21} = \frac{X_n - X_2}{X_n - X_1}$			
	$r_{21} = \frac{X_3 - X_1}{X_n - X_1}$			
14	$r_{22} = \frac{X_{n-2} - X_1}{X_n - X_1}$.492	.546	.641
15	$r_{22} = \frac{X_{n-2} - X_1}{X_n - X_1}$.472	.525	.616
16	$r_{22} = \frac{X_{n-2} - X_1}{X_n - X_1}$.454	.507	.595
17	$r_{22} = \frac{X_{n-2} - X_1}{X_n - X_1}$.438	.490	.577
18	$r_{22} = \frac{X_{n-2} - X_1}{X_n - X_1}$.424	.475	.561

n	critério	10%	5%	1%
	$r_{22} = \frac{X_n - X_{n-2}}{X_n - X_3}$			
19		.412	.462	.574
20		.401	.450	.535
21		.391	.440	.524
22		.382	.430	.514
23		.374	.421	.505
24		.367	.413	.497
25		.360	.406	.489

T A B L A I V
VALORES CRITICOS PARA LAS ESTADISTICAS
DE FERGUSON¹

Niveles de Significancia Para $\sqrt{b_1}$

	n									
	5	10	15	20	25	30	35	40	50	60
1%	1.34	1.31	1.20	1.11	1.06	.98	.92	.87	.79	.72
5%	1.05	.92	.84	.79	.71	.66	.62	.59	.53	.49

Niveles de Significancia Para b_2

	n							
	5	10	15	20	25	50	75	100
1%	3.11	4.83	5.08	5.23	5.00	4.88	4.59	4.39
5%	2.89	3.85	4.07	4.15	4.00	3.99	3.87	3.77

$$\sqrt{b_1} = \sqrt{n} \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right]^{3/2}}$$

$$b_2 = n \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right]^2}$$

1 Grubbs (1969)

T A B L A V
VALORES CRITICOS PARA LA ESTADISTICA
DE TIKU

Niveles de Significancia		
n	1%	5%
5	.2154436	.3684031
10	.5623413	.687656
15	.7017038	.7941833
20	.7742637	.8466824

$$T_1 = \frac{\frac{1}{n-1} \sum_{i=2}^n (X_i - X_{(1)})}{\frac{1}{n} \sum_{i=1}^n (X_i - X_{(1)})}$$

$$T_n = \frac{\frac{1}{n-1} \sum_{i=1}^{n-1} (X_i - X_{(1)})}{\frac{1}{n} \sum_{i=1}^n (X_i - X_{(1)})}$$

T A B L A VI
VALORES CRITICOS PARA LA ESTADISTICA
DE DIXON ¹

Prueba para una Observación Aberrante
Chica en una Muestra Exponencial

Niveles de Significancia		
n	5%	1%
3	0.905	0.980
4	0.618	0.808
5	0.429	0.618
6	0.316	0.479
7	0.246	0.381
8	0.198	0.312
9	0.165	0.262
10	0.140	0.224
11	0.121	0.194
12	0.106	0.171
13	0.094	0.152
14	0.085	0.136
15	0.077	0.124
16	0.070	0.113
17	0.064	0.103
18	0.059	0.095
19	0.055	0.088
20	0.051	0.082

$$T_1 = \frac{X_{(2)} - X_{(1)}}{X_{(n)} - X_{(1)}}$$

1 Barnett y Lewis (1977)

T A B L A VII
VALORES CRITICOS PARA LA ESTADISTICA
DE DIXON ¹

Prueba para una Observación Aberrante
Grande en una Muestra Exponencial

Niveles de Significancia		
n	5%	1%
3	0.974	0.995
4	0.894	0.957
5	0.830	0.912
6	0.782	0.875
7	0.746	0.845
8	0.717	0.821
9	0.694	0.800
10	0.675	0.783
11	0.658	0.768
12	0.644	0.755
13	0.631	0.743
14	0.620	0.733
15	0.610	0.724
16	0.601	0.715
17	0.593	0.707
18	0.586	0.700
19	0.579	0.694
20	0.573	0.687
21	0.567	0.682

¹ Barnett y Lewis (1977)

TABLA VIII
VALORES CRITICOS PARA LA ESTADISTICA DE
LAURENT¹

Niveles de significancia

n	3	4	5	6	7	8	9
.90	.95000	.81743	.70760	.62894	.55907	.50741	
.99	.99500	.91226	.80428	.78853	.72170	.66410	.61517

¹ Laurent (1962)

T A B L A IX
VALORES CRITICOS PARA LA ESTADISTICA
DE SHAPIRO Y WILK¹

Niveles de Significancia				
n	1%	5%	95%	99%
3	0.254	0.270	0.993	0.9997
4	0.130	0.160	0.858	0.968
5	0.0905	0.119	0.668	0.860
6	0.0665	0.0956	0.509	0.678
7	0.0591	0.0810	0.416	0.571
8	0.0512	0.0710	0.350	0.485
9	0.0442	0.0633	0.300	0.401
10	0.0404	0.0568	0.253	0.339
12	0.0358	0.0494	0.202	0.272
14	0.0317	0.0428	0.165	0.213
16	0.0280	0.0374	0.136	0.177
18	0.0250	0.0332	0.116	0.148
20	0.0227	0.0302	0.100	0.129
30	0.0164	0.0213	0.0593	0.0719
40	0.0131	0.0164	0.0414	0.0499
50	0.0111	0.0137	0.0317	0.0360
60	0.0095	0.0117	0.0252	0.0291
70	0.0084	0.0103	0.0209	0.0241
80	0.0075	0.0091	0.0177	0.0205
90	0.0069	0.0082	0.0156	0.0176
100	0.0063	0.0074	0.0139	0.0153

¹ Barnett y Lewis (1977)
Una tabla más completa está dada por Shapiro (1972)

BIBLIOGRAFIA

BIBLIOGRAFIA

- (1925) Irwin, J.O. "On a criterion for the rejection of outlying observations". *Biometrika* 17, 238-250.
- (1936) Pearson, E.S. y Chandra Sekar, C. "The efficiency of statistical tools and a criterion for the rejection of outlying observations". *Biometrika* 28, 308-320.
- (1950) Grubbs, F.E. "Sample criteria for testing -- outlying observations". *Ann. Math. Statist.* 21, 27-58.
- (1950) Dixon, W.J. "Analysis of extreme values". *Ann. Math. Statist.* 21, 488-506.
- (1960) Anscombe, F.J. "Rejection of outliers". *Technometrics* Vol. 2 , No. 2.
- (1962) Laurent, A.G. "Conditional distribution of order statistics and distribution of the reduced i th order statistic of the exponential model". *Ann. Math. Statist.* 34, 652-657.
- (1965) Basu, A.P. "On some tests of hypotheses relating to the exponential distribution when outliers are present". *American Statistical Association Journal*, Junio 1965, 548-558.

- (1966) Harris, B. "Theory of probability".
Addison-Wesley.
- (1969) Grubbs, F.E. "Procedures for detecting out-
lying observations in samples".
Technometrics Vol. 2, No. 1, 1-21.
- (1970) Johnson, N.L. and Kotz, S. "Continuous Uni-
variate distributions 1".
Houghton Mifflin Company, 1970.
- (1971) Chen, E.H. "The power of the Shapiro-Wilk
W-test for normality on samples from conta-
minated distributions". J. Amer. Statist.
Ass., 66, 760-762.
- (1972) Shapiro, S.S. and Wilk, M.B. "An analysis
of variance test for the exponential distri-
bution (Complete samples)".
Technometrics Vol. 14, No. 2, 355-370.
- (1975) Tiku, M.L. "A new statistic for testing sus-
pected outliers". Technometrics Vol. 4,
No. 8, 737-752.
- (1975) O'Reilly T. F. "Algunos usos de estimación
de distribuciones en Regresión".
Comunicaciones Técnicas. Serie B, Vol. 6,
No. 90. IIMAS. UNAM.
- (1978) Barnett, V. and Lewis, T. "Outliers in Sta-
tistical data".
John Wiley & Sons, 1978.