



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

**OBSERVACIONES DISCORDANTES EN ANALISIS
DE REGRESION; DETECCION Y TRATAMIENTO.**

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 N:

Rocio Palacios Blanco

Vicente Castañón Ibarra



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 está 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

PROLOGO.

CAPITULO I

- 1.1 INTRODUCCION GENERAL
- 1.2 LA MATRIZ SOMBRERO
- 1.3 RESIDUALES

CAPITULO II

- 2.1 METODO DE LUND
- 2.2 METODO DE ELLENBERG
- 2.3 METODO DE PRESCOTT Y COOK
- 2.4 METODO DE PRESCOTT
- 2.5 METODO DE COOK
- 2.6 METODO DE VALENCIA
- 2.7 METODO GRAFICO
- 2.8 METODO DE GENTLEMAN Y WILK
- 2.9 METODO DE ANDREWS Y PREGIBON
- 2.10 METODO DE MICKEY, DUNN Y VIRGINIA (STEPWISE)

CAPITULO III

- 3.1 OBSERVACIONES INFLUYENTES

CAPITULO IV

- 4.1 TRATAMIENTO

CAPITULO V

5.1 CUATRO EJEMPLOS

5.2 CONCLUSIONES

BIBLIOGRAFIA

PROLOGO

A través del tiempo el hombre se ha basado en observaciones para tratar de estudiar y entender algunos fenómenos, frecuentemente en el conjunto de observaciones se encuentran algunas que parecen incoherentes respecto al resto; este tipo de observaciones son llamadas discordantes y serán el objetivo principal de este estudio, enmarcadas bajo el Análisis de Regresión.

La tesis se encuentra dividida en 5 capítulos; en el primero se da una introducción general en la cual se define a una observación discordante y se ve la aparición de estas observaciones en las distintas ramas del conocimiento humano; posteriormente se realizará un breve estudio de los residuales tratando de resaltar sus principales características, muchas de las cuales serán necesarias para el desarrollo del capítulo II; por último en este capítulo se estudiarán las propiedades de la matriz llamada sombrero, cuyo estudio es fundamental en el Análisis de Regresión y en el tratamiento y detección de las observaciones discordantes.

En el capítulo II se presentan algunos métodos que han sido propuestos para detectar observaciones discordantes en el contexto de Análisis de Regresión.

En el capítulo III se hace un pequeño estudio de las observaciones influyentes y se proponen algunos métodos para detectar tales observaciones.

Una vez que las observaciones discordantes han sido de

tectadas surge la duda de ¿ Que hacer con ellas ? . Esta pregunta queda resuelta en el capítulo IV en el estudio del tratamiento de observaciones discordantes.

Finalmente en el capítulo V presentamos cuatro ejemplos a cada uno de los cuales se le aplicarán los 10 métodos propuestos en el capítulo II, para deducir después algunas conclusiones finales.

En estas condiciones acude a las casas de cambio o directamente a los productores, en demanda de ayuda y les piden - le presten determinados elementos a reserva de cubrirlos - tan pronto levanten su cosecha y la realice.

-Practicamente en este momento nacen las operaciones de -- crédito.

¿ Con que bases acudia el agricultor, siguiendo con nuestro ejemplo, ante el productor o ante la casa de cambio, - para que cualquiera de estas o el productor mismo le prestara los elementos demandados? así mismo ¿ Con qué bases - el productor prestaba dichos elementos? , estas preguntas- pueden contestarse de la siguiente forma:

- 1.- El agricultor tenía una necesidad que satisfacer.
- 2.- Ofrecía una promesa de pagar.
- 3.- El productor conocía al agricultor y consecuentemente- tenía confianza en él .

¿ Que se desprenda de lo anterior? pues es muy simple, mediante tres factores: la necesidad, la firme promesa de -- reembolsar los bienes prestados y por último, la confianza.

A través del tiempo las operaciones de crédito se han incrementado y perfeccionado. Se tienen datos en el sentido- de que el crédito nació durante la Edad Media.

" Una observación discordante es aquella que no se ajusta al patrón de las observaciones restantes. En modelos lineales es aquella, cuyo valor ajustado, difiere mucho de los demás ".
(Stefansky, 1972) [29]

" Una observación discordante, (ó conjunto de observaciones) es aquella que parece ser inconsistente con el resto del conjunto de datos ". (Grubbs, 1969) [20]

Podemos concluir que, una observación discordante es aquella que difiere substancialmente del resto de las observaciones de la muestra en la cual se encuentra, es decir, su valor es extremo en relación con los demás.

El problema de las observaciones discordantes ha sido de gran interés para científicos, estadísticos y todas aquellas personas que de alguna manera obtienen sus resultados de la interpretación de observaciones para analizar algún fenómeno de interés. Y es necesario hacer una examinación crítica de tales observaciones para poder "asegurar" que los resultados obtenidos a partir de esas observaciones van a ser representativos.

Desde los primeros seres humanos que se basaron en colecciones de datos, para tratar de explicar el mundo en que vivían, se enfrentaron al problema de tratar con observaciones discordantes, debido a que aparecían con cierta frecuencia en sus colecciones, contaminando a todo el conjunto y distorsionando, la información que podía obtenerse de él. Y ellos no tenían porque confiar en que esas observaciones pertenecían a la misma población y ser del todo dignas de confianza para poder inferir en base a los resultados obtenidos a partir de ellas; debido a que una

" Una observación discordante es aquella que no se ajusta al patrón de las observaciones restantes. En modelos lineales es aquella, cuyo valor ajustado, difiere mucho de los demás ".

(Stefansky, 1972) [29]

" Una observación discordante, (ó conjunto de observaciones) es aquella que parece ser inconsistente con el resto del conjunto de datos ".(Grubbs, 1969) [20]

Podemos concluir que, una observación discordante es aquella que difiere substancialmente del resto de las observaciones de la muestra en la cual se encuentra, es decir, su valor es extremo en relación con los demás.

El problema de las observaciones discordantes ha sido de gran interés para científicos, estadísticos y todas aquellas personas que de alguna manera obtienen sus resultados de la interpretación de observaciones para analizar algún fenómeno de interés. Y es necesario hacer una examinación crítica de tales observaciones para poder "asegurar" que los resultados obtenidos a partir de esas observaciones van a ser representativos.

Desde los primeros seres humanos que se basaron en colecciones de datos, para tratar de explicar el mundo en que vivían, se enfrentaron al problema de tratar con observaciones discordantes, debido a que aparecían con cierta frecuencia en sus colecciones, contaminando a todo el conjunto y distorsionando, la información que podía obtenerse de él. Y ellos no tenían porque confiar en que esas observaciones pertenecían a la misma población y ser del todo dignas de confianza para poder inferir en base a los resultados obtenidos a partir de ellas; debido a que una

observación suficientemente errónea puede arruinar todo el análisis que se este llevando a cabo. Debido a estos problemas y a su frecuente aparición en varios contextos van adquiriendo atención especial y se van delineando métodos para detectarlas y tratarlas.

Son varios los contextos en los cuales podemos encontrarnos con este tipo de observaciones; enseguida mencionamos algunos de ellos .

1) EN MODELOS PROBABILISTICOS :

Al enfrentarnos con el problema de determinar que función de probabilidad caracteriza el comportamiento de cierta población, lo que se hace es, en base a una muestra aleatoria de la población bajo estudio tratar de determinar la distribución de ésta. Al muestrear podemos encontrar la presencia de observaciones discordantes y esto provocaría que no se pudiera delinear con exactitud la forma de distribución de la población.

2) EN SERIES DE TIEMPO :

Con este tipo de análisis podemos estudiar fenómenos comerciales, meteorológicos, sociales etc. Y en este tipo de modelos también pueden encontrarse observaciones discordantes ocasionando los mismos problemas antes mencionados, que repercuten directamente en los resultados obtenidos por el modelo.

3) EN MODELOS DE DISEÑO DE EXPERIMENTOS Y EN GENERAL EN MODELOS LINEALES :

En este tipo de modelos podemos esperar encontrarnos con observaciones discordantes en ambas interpretaciones: Determinísticas y Probabilísticas. Las determinísticas (que vie -

nen desde la fuente de datos no estadísticos) y probabilísticos (sobre el cuestionamiento de la hipótesis de distribución o de estructuración del problema).

4) En general en todo tipo de modelos que se basen u ocupen muestras de observaciones para su desarrollo y obtención de resultados podemos encontrarnos con observaciones discordantes. Y naturalmente encontramos observaciones discordantes en modelos de Regresión Lineal General, problema, que como ya hemos mencionado, será el que se trate a través de todo el desarrollo de la presente tesis.

Antes de pasar al problema de observaciones discordantes en modelos de Regresión Lineal, es importante ver cuales pueden ser las causas de las variaciones e inconsistencias en los datos.

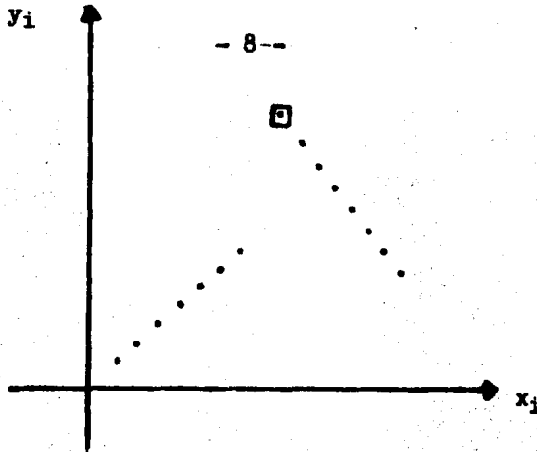
Estas variaciones se pueden deber a diferentes causas, entre las que destacan :

VARIACION INHERENTE :

Esta es la variación propia del fenómeno, es decir, es posible que una observación discordante sea una simple manifestación propia del fenómeno.

Esta variación no puede ser eliminada ya que al eliminarla estaríamos eliminando también características importantes del fenómeno bajo estudio.

Por ejemplo en el siguiente caso la observación encarrada en el cuadro podría detectarse como discordante y sin embargo es tan sólo la observación que indica que el fenómeno empieza a cambiar de comportamiento .



ERRORES DE MEDICION: r

Estos son errores que se cometen al usar los instrumentos de medición al estar recolectando las observaciones que sirven para explicar el fenómeno.

ERRORES DE EJECUCION r

Estos son errores que podemos definir, como la diferencia entre lo que se intenta hacer y lo que realmente se hace. Por ejemplo cuando observamos un elemento de una población completamente distinta a la de nuestro interés.

En muchas ocasiones cuando el valor de una observación se encuentre alejado del valor de las demás, se sospechará que es te alejamiento es causado por alguna de las variaciones antes men cionadas.

Pero surge de inmediato otra pregunta :

¿ Que hacer cuando en un modelo, muestra, etc. encontre mos la presencia de observaciones discordantes ?

Si es posible rastrear la observación podríamos detectar las causas por las que el proceso ha fallado y ésta se rechazaría,

en caso contrario, es decir, en caso de no encontrar alguna causa que explique el alejamiento del valor de esa observación, no será posible desechar fácilmente la observación y se necesitarán de ciertos criterios para hacerlo.

Una cosa importante de notar es que si la variación se justifica y se debe plenamente a un error de ejecución o de medición que no es posible rectificar se tiene entonces una buena justificación para descartar dicha observación y todo registro de ella.

Ahora bien, si se sospecha que esta observación "extraña" es producto de alguna peculiaridad del fenómeno bajo estudio (por ejemplo no normalidad o cambio de pendiente en el caso de modelos lineales) entonces es aconsejable separar la observación y estudiarla por separado y al mismo tiempo abrir una sección en el reporte de la investigación en donde se comenten este tipo de observaciones.

Lo anterior es sólo un punto de vista acerca de como tratar a las observaciones discordantes cuando éstas se encuentran presentes. Pero a través del tiempo esta disyuntiva acerca de rechazar o no rechazar ha sido motivo de gran discusión, como ejemplo de algunas opiniones acerca de esto tenemos las de :

Bessel y Baeur (1838) quienes indican que nunca se debe rechazar una observación meramente porque su naturaleza es " extraña " (valor del residual correspondiente muy grande o muy pequeño) [5] .

Legendre (1805) quién relata su experiencia sobre el juicio subjetivo en cálculos y recomienda la exclusión de estos

monetaria internacional, así como el nacimiento de un México deseoso de ocupar un primer plano en la Industrialización Latinoamericana, hicieron necesarios complementar la Legislación Bancaria de 1941. Esto ayudó a consolidar a la Banca especializada (depósito, hipoteca, fomento, uniones de Crédito Capitalización) y fue la clave de un desarrollo bancario espectacular.

Así mismo se forma también la Banca de Carácter Para-Estatal para Fomento Agrícola, Industrial, de Exportación y de Obras Públicas, en tanto que la emisión de valores Bancarios se --consolida.

Si bien por un lado las metas logradas confirmaban que el --crédito bancario era capaz de dotar a la Industria de recursos cuantiosos a mediano y largo plazo, en apoyo a su cada vez más dinámico crecimiento, por otro lado, también eran visibles ciertas resistencias o rigideces del sistema en aspectos importantes. Así por ejemplo los Bancos Financieros y --las Cédulas Hipotecarias no habían entrado al mercado propiamente dicho quedando manejados por los propios emisores además de que, dado que éstos constituían la pieza maestra de --la Banca el mercado bursátil se había mantenido en una posición auxiliar.

En cuanto a los bonos gubernamentales, tampoco alcanzaron el mercado público pues a pesar de algunos intentos de colocar-

los en los años 40, el hecho es que continuaban cautivos en el encaje de nuestro banco central.

Tampoco el encaje legal y la selectividad del crédito pudie-
ron combatir la inflación, como se pensó en un principio --
que podría lograrse mediante una banca central, tanto pro-
pulsora del financiamiento cuanto niveladora de la oferta -
monetaria.

Así, durante la década de los 60 y bajo las presiones de --
una economía cada vez más compleja, el agrupamiento Banca--
rio por instituciones cabeceras, tanto en el sector privado
como en el paraestatal, ditimo refinamiento de la estructu-
ra bancaria por especialidades que también sirviera al país
acaba por perder toda flexibilidad, hasta que en los años -
70, agotada su capacidad para seguir al país en sus necesi-
dades financieras y de manejo monetario, da paso, en 1976,-
a la Banca Múltiple. Con base en esto, el lo. de Enero de -
1979 entre en vigor, la incorporación a la Ley Bancaria de
la referida Banca Múltiple.

En la segunda mitad de la década de los ochentas, se agudi-
na la crisis económica de muchos de los países en vías de
desarrollo, debido al inesperado crecimiento demográfico, a
la carencia de adecuadas estructuras internas para el desa-
rrollo eficaz de sus actividades productivas, y a las defi-
ciencias en la administración pública y en sus sistemas de
crédito y, como consecuencia de todo ello problemas políti-
cos internos (5)

(5).- Manual del Funcionario Bancario.-Jorge Melidana Alva--
rez, Edición 1984

b) La prima a pagarse.

c) Cuanta protección nos va a brindar en caso de que ocurra el siniestro.

El primer inciso es difícil de estudiar ya que depende del tipo especial de problema que se esté estudiando, sin embargo es un hecho el peligro de las observaciones discordantes, y es por ésto, que los investigadores están interesados en estudiarlas. Este factor afecta por igual a todos los criterios y por lo tanto no se utiliza para efectuar comparaciones.

La protección y la prima (el precio de la protección) dependen tanto de la forma de detectar, como la de tratar a las observaciones discordantes, ya que existe un precio tanto para la protección en el método de detección como para la protección en el método de tratamiento. Y es por esto que se consideran en conjunto tanto la protección (por detección y tratamiento) como el precio (de detección y tratamiento) ya que no sirve tener un método óptimo de tratamiento si el método de detección es pésimo o viceversa.

Para efectos de comparación deben suponerse válidas las siguientes condiciones (aunque en la práctica no lo sean). (2)

i) Los costos de computación deben ser ignorados. Si no se considera esto, la prima tendría que incluir el costo de computación extra resultante del uso de la regla.

ii) Cualesquiera que sean las circunstancias que hayan causado que una observación sea "extraña" no afectan en forma alguna a las demás observaciones, esto es, si una observación es discordante sus compañeras no tienen porque estar con -

taminadas.

- iii) Los criterios de detección y tratamiento son objetivos, en el sentido de que no incluirán la información a priori que sobre el problema posea el investigador, esto es, los criterios no dependen de quién los aplique.

HISTORIA

A través de los años, el tema de exclusión de observaciones discordantes ha sido de gran importancia e interés en muchas ciencias como : Astronomía, Geodésica, Química, Física, Balística, etc.

Una de las primeras referencias que se tienen acerca de este tema parece haber sido hecha por Bessel, en un trabajo de Geodésica, publicado en 1838 por la dirección de la escuela alemana de astrónomos [2]. Lo que Bessel expresaba era lo siguiente :

" Tenemos pensado que sólo mediante estricta observación de esta regla se pueden eliminar arbitrariedades de nuestros resultados ". La regla a la que hacia referencia era la de " nunca excluir una observación solamente porque su residual era grande ".

El primer intento de un criterio de exclusión basado en un razonamiento probabilístico fue hecho por Peirce en 1852 y su argumento fue reproducido por Chauvenet en 1863 quien entonces dió una regla similar basada en un argumento simple.

Y debido a estas reglas, y por algunas otras sugerencias que servían de advertencia en el caso de la existencia y presencia de observaciones discordantes, se produjo una gran discusión, ya que algunos ni siquiera admitían el concepto de observaciones discordantes y la explicación que daban cuando ellas aparecían era simplemente " Hay algunas explicaciones físicas obvias de su presencia "; pero el hecho era que científicos, estadísticos y otras

personas que tuvieron que analizar datos y tomar decisiones en base a ellos se veían forzados a establecer y hacer juicios acerca de la presencia de observaciones discordantes. Y es debido a esto último que se siguieron haciendo estudios e investigaciones de ellas.

El siguiente paso significativo, después de Peirce y Chauvenet, fué hecho por Wright (1884). Tiempo después se dió una excelente discusión general, creada por la ocurrencia de este tipo de observaciones en la Astronomía. Wright sugiere entonces, que la mejor regla para detectar "lo que no vamos a observar" es excluir cualquier observación cuyo residual exceda en magnitud cinco veces la probabilidad de error (es decir 3.37 veces la desviación estándar). La razón dada para esta regla es que si la ley Gaussiana de error se cumple realmente, en promedio solo una observación en un millar será rechazada.

De 1925 en adelante, los estadísticos pusieron mucha atención en este tema, Student (1927) proporcionó una interesante relación de determinaciones analíticas químicas y propuso el uso de un criterio de rango para la detección de estas observaciones.

La exclusión de observaciones discordantes parece haber sido de particular interés para los Norteamericanos; como un ejemplo podemos mencionar a Peirce, Chauvenet y Wright.

El tema se inició en el contexto de análisis por mínimos cuadrados o en el modelo de análisis de varianza y es curioso (aparentemente) que no fué considerado por Gauss. (2)

Y he aquí que ya para nuestros días los critérios, pruebas, bases para la exclusión de observaciones discordantes, desde su detección han ido evolucionando con el paso del tiempo y adquiriendo mayor cimentación científica en cuanto a los criterios para detectarlas.

LA IMPORTANCIA DE LA MATRIZ SOMBRERO

En muchos resultados conocidos de Regresión lineal, se encuentra involucrada una matriz importante, a la cual llamaremos ' Matriz Sombrero ' .

Los elementos de la diagonal de esta matriz, nos serán muy útiles para llevar a cabo mucho más fácilmente los cálculos, de detección de observaciones discordantes. Además esta matriz sirve para determinar la influencia de un valor observado, sobre un ajustado. Por todo esto creemos necesario hablar un poco acerca de tal matriz.

En el modelo lineal general $Y = X\beta + e$, donde $Y_{(n \times 1)}$ es el vector de observaciones, $X_{(n \times p)}$ es la matriz de variables explicativas de rango p y $e_{(n \times 1)}$ es el vector de errores, al ajustar por mínimos cuadrados obtenemos que :

$$\begin{aligned}\hat{Y} &= X\hat{\beta} = X(X'X)^{-1}X'Y \\ &= HY\end{aligned}$$

en donde $H = X(X'X)^{-1}X'$ es la llamada matriz sombrero, la cuál es una proyección de y sobre \hat{y} .

Esta matriz es frecuentemente usada por el papel que juega en la determinación de la matriz de varianzas y covarianzas de \hat{y} y de $\xi = Y - \hat{Y}$, el vector de residuales

$$\text{VAR}(\hat{Y}) = \sigma^2 H$$

$$\text{VAR}(\xi) = \sigma^2 (I - H)$$

Algunas de las propiedades de la matriz H son :

H es simétrica; es decir $H = H^t$, además H es idempotente, es decir $H = H^2$.

Sea $H = (h_{ij})$, por ser H simétrica e idempotente tenemos que :

$$(1) \quad h_{ii} = \sum_{j=1}^n h_{ij}^2 = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2$$

Notemos primero que h_{ii} es una suma de elementos elevados al cuadrado, y $\therefore h_{ij} \geq 0 \quad \forall j = 1, 2, \dots, n$

por lo cual $h_{ii} \geq 0$ segundo, h_{ii} no puede ser mayor que 1, dado que si lo fuera h_{ii}^2 sería mayor que h_{ii} , y la igualdad (1) nunca podría darse $\therefore 0 \leq h_{ii} \leq 1$

Como H es idempotente entonces :

$$\begin{aligned} \text{Rango (H)} &= \text{traza (H)} = \text{tr (X(X'X)^{-1}X')} \\ &= \text{tr (X'X(X'X)^{-1})} \\ &= \text{tr (I_p)} = p \end{aligned}$$

$$\text{i.e. } \sum_{i=1}^n h_{ii} = p$$

Tomando en consideración este resultado llegamos a que el tamaño promedio de un elemento de la diagonal de la matriz sombrero es $\sum_{i=1}^n h_{ii} / n = p/n$

Desde el punto de vista estadístico, el elemento h_{ij} de H tiene la siguiente interpretación.-

Es la influencia o importancia de y_j sobre \hat{y}_i debido a que :

$$\hat{y}_i = h_{i1}y_1 + h_{i2}y_2 + \dots + h_{ij}y_j + \dots + h_{in}y_n$$

Y esto anterior es de gran importancia, ya que, conocer H nos ayudará a detectar valores de "X" que tengan una influencia grande sobre el ajuste.

En consecuencia si el diseño está bien, se esperaría que todas las y 's tuvieran la misma influencia en el ajuste, la influencia de y_i sobre el ajuste se encuentra sobre los elementos del i -ésimo renglón de la matriz (H) (o i -ésima columna debido a la simetría) esto anterior ya que analizando tenemos por ejm. que :

La influencia de y_2 sobre el ajuste la vemos reflejada como sigue :

$y_1 = h_{12}y_2 + \dots$	aquí tenemos la influencia de y_2 en y_1
$y_2 = h_{22}y_2 + \dots$	aquí tenemos la influencia de y_2 en y_2
\vdots	
\vdots	
$y_n = h_{n2}y_2 + \dots$	aquí tenemos la influencia de y_2 en y_n

Entonces nos podemos dar cuenta que dependiendo del valor de h_{ij} , $i = 1, 2, \dots, n$ $j = 1, 2, \dots, n$ (pero fija según la observación) es la influencia que y_j tiene sobre el ajuste; pero estos elementos como vimos anteriormente dependen del valor que tenga h_{11} , por lo cual es necesario tan solo fijarnos en el valor de h_{11} , (es decir ver la influencia de y_1 sobre \hat{y}_1) para saber la influencia de y_j sobre el ajuste.

Otra cosa que es importante de analizar son los casos extremos en el valor de h_{11} , puesto que ya sabemos que $0 \leq h_{11} \leq 1$ analizaremos los siguientes casos :

i) $h_{11} = 0$

ii) $h_{11} = 1$

Para el primer caso ($h_{11} = 0$) tenemos que

$$h_{11} = \sum_{j=1}^n h_{1j}^2 = 0$$

$$\Rightarrow h_{1j}^2 = 0 \Rightarrow h_{1j} = 0 \quad \forall j$$

$$Y \text{ como } y_1 = h_{11} + h_{12} y_2 + \dots + h_{1n} y_n$$

$$y_1 = 0$$

Es decir y_1 es fijo al cero por la forma del diseño, y no es afectado por y_1 o por cualquier otro y_j , es decir y_1 no tiene ninguna influencia sobre el ajuste.

Para el segundo caso ($h_{11} = 1$) tenemos que :

$$h_{11} = h_{11}^2 + \sum_{j \neq 1} h_{1j}^2 = 1 + \sum_{j \neq 1} h_{1j}^2$$

$$\Rightarrow 1 = 1 + \sum_{j \neq 1} h_{1j}^2 \quad \therefore \sum_{j \neq 1} h_{1j}^2 = 0$$

Luego como :

$$\begin{aligned} \hat{y}_1 &= \sum_{j \neq 1} h_{1j} y_j = h_{11} y_1 + \sum_{j \neq 1} h_{1j} y_j \\ &= h_{11} y_1 \\ \therefore \hat{y}_1 &= y_1 \end{aligned}$$

esto nos dice que el modelo ajusta exactamente al valor del dato.

Como mencionamos anteriormente, podríamos pensar que el tamaño promedio de h_{11} es p/n , y un posible criterio para decidir cuando h_{11} es "grande", y por lo tanto y_1 es influyente en el ajuste, es comparar h_{11} contra $2p/n$, si $h_{11} > 2p/n$, h_{11} se considerará "grande".

RESIDUALES

En el modelo de Regresión $Y = X\beta + e$ los residuales son definidos como las n diferencias $\epsilon_i = X_i - \hat{y}_i$; donde y_i es el valor observado y \hat{y}_i es el correspondiente valor ajustado. Y debido a que todos los métodos para detectar observaciones discordantes se basan en su estudio y cálculo hallamos importante hablar más acerca de ellos. La razón por la cual los métodos se basan en su estudio es debido a que por su definición son sensibles e indican la presencia de estas observaciones, ya que como la definición lo indica los residuales ϵ_i son las diferencias entre lo que se observa y lo que se predice por la ecuación de Regresión, esto es, la cantidad por la cual la ecuación de Regresión no es capaz de explicar correctamente el fenómeno. Como consecuencia de esto, podemos pensar a los ϵ_i como "errores observados si el modelo es correcto".

La importancia de los residuales radica en que mediante ellos podemos ver cuantitativamente los errores observados si el modelo es correcto.

El estudio y uso de los residuales no sólo se hace para detectar observaciones discordantes, tienen otros usos tan importantes como éste, como veremos enseguida.

Quando se esta haciendo un Análisis de Regresión se es-

tablecen los siguientes supuestos :

a) Los de tipo distribucional como son :

i) Que los errores son independientes, tienen media cero y varianza constante σ^2 .

ii) Que siguen una distribución normal .

b) Los supuestos de linealidad .

Para poder confiar en los resultados del análisis es necesario verificar tales supuestos en el fenómeno bajo estudio, porque si no verificamos, por ejemplo, los supuestos del inciso b), podría suceder como en el siguiente ejemplo [32].

Este ejemplo consta de cuatro casos y en la siguiente tabla se encuentran los datos para cada uno de ellos :

var.	caso #	$\frac{1-3}{x}$	$\frac{1}{y}$	$\frac{2}{y}$	$\frac{3}{y}$	$\frac{4}{x}$	$\frac{4}{y}$
1	10	8.04	9.14	7.46	8	6.58	
2	8	6.95	8.14	6.77	8	5.76	
3	13	7.58	8.74	12.74	8	7.71	
4	9	8.81	8.77	7.11	8	8.84	
5	11	8.33	9.26	7.81	8	8.47	
6	14	9.96	8.10	8.84	8	7.04	
7	6	7.24	6.13	6.08	8	5.25	
8	4	4.26	3.10	5.39	19	12.50	
9	12	10.84	9.13	8.15	8	5.56	
10	7	4.82	7.26	6.42	8	7.91	
11	5	5.68	4.74	5.73	8	6.89	

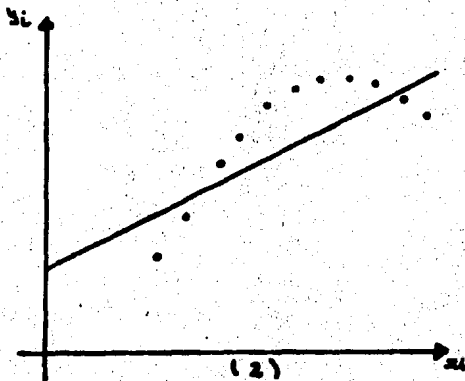
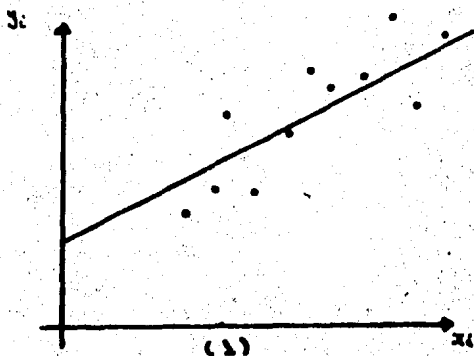
Con este ejemplo nos daremos cuenta de la necesidad de analizar siempre si los supuestos de linealidad, y en general todos los supuestos, se cumplen.

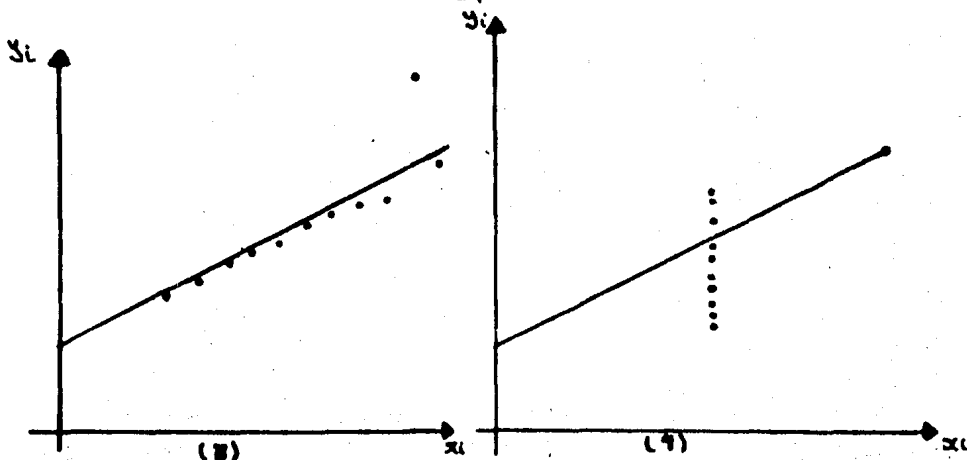
Cada conjunto consiste de 11 pares de puntos (x_i, y_i) para los cuales el modelo de Regresión Lineal simple es ajustado. Con cada conjunto de datos se obtuvo el estimador de β_0 , β_1 , σ^2 y el valor de R^2 , teniendo como resultado la misma línea ajustada para cada conjunto, con los siguientes valores :

$$\hat{\beta}_0 = 3 \quad \hat{\beta}_1 = .5 \quad \hat{\sigma}^2 = 13.75 \quad R^2 = .667$$

Ya que las estadísticas obtenidas son las mismas para todos los conjuntos de datos, uno podría concluir que el modelo de Regresión Lineal es igualmente apropiado para cada uno de ellos, esto confiando plenamente en que los supuestos de linealidad y distribucionales del modelo se cumplen para cada conjunto sin verificarlos.

Sin embargo un simple análisis gráfico de x contra y sugiere que estos resultados no son confiables y se ve el porqué en las siguientes gráficas :





En la gráfica # 1 nos podemos dar cuenta que se puede "confiar" en la línea ajustada para este primer conjunto de datos. Para el segundo conjunto de datos, la gráfica sugiere una conclusión diferente; que los resultados del análisis por Regresión Lineal son incorrectos ya que sería más apropiado, para este conjunto de datos, un ajuste cuadrático ó un ajuste polinomial. La gráfica # 3, correspondiente al tercer conjunto de datos, sugiere que lo predicho por la Regresión Lineal simple puede ser correcta para casi todos los datos, menos para uno que se encuentra muy alejado de la línea ajustada. La gráfica del conjunto final, gráfica # 4, es diferente de las otras en las cuales realmente había bastante información para hacer un juicio concerniente al modelo ajustado. En ésta el valor \hat{P}_1 está fuertemente determinado por el valor de y_0 ; si esta observación fuera omitida ni siquiera podríamos estimar a \hat{P}_1 .

Con estos ejemplos nos podemos dar cuenta de la importancia de contar con herramientas que nos permitan verificar si:

i) Existe linealidad entre las variables.

ii) Los supuestos del modelo se cumplen.

para poder confiar plenamente en los resultados obtenidos por el análisis de Regresión. Y para ayudar a la realización de estas verificaciones son de gran utilidad los residuales; así si el modelo es correcto los residuales tenderán a confirmar o no los supuestos hechos.

De esta manera cuando uno se pregunte si en realidad se cumplen los supuestos, las respuestas pueden ser :

i) Según el comportamiento de los residuales los supuestos parecen estar confirmados.

ii) Según el comportamiento de los residuales los supuestos no parecen confirmarse.

La respuesta ii) es así, ya que, no se tiene clara evidencia de que estén incorrectos dichos supuestos en el fenómeno bajo estudio.

Lo anterior implica una cuidadosa inspección de los residuales.

Hay varias formas de analizar los residuales para verificar todos los supuestos del modelo; los que aquí mencionamos son gráficos y fáciles de llevar a cabo; estos son :

i) En conjunto completo

ii) Papel normal

iii) Contra los valores ajustados \hat{y}_i

iv) Contra las variables explicatorias $X_{1i}, X_{2i}, \dots, X_{ki}$

v) En conjunto, ordenado con respecto al tiempo

En general, para las gráficas básicas, los residuales podrán ser graficados en cualquier forma que sea sensible para el problema particular bajo consideración.

i) En conjunto completo :

Este tipo de gráficas se llevan a cabo graficando por arriba de una línea recta los residuales, sin tomar en cuenta cualquier factor que haya influido en su obtención. Si el modelo ajustado es correcto, en la hipótesis distribucional, entonces los residuales mostrarán una tendencia como si provinieran de una distribución normal con media cero; (es decir se encontrarán "bien" distribuidos alrededor del cero). Pero ¿ cuándo éste tipo de gráficas van a contradecir esta idea ? Primero tenemos que notar que la media de los residuales es cero, pero esto siempre se cumple para el caso del modelo de Regresión Lineal con el parámetro β , entonces para poder juzgar como "erróneo" un modelo utilizando este tipo de gráficas es necesario tener un patrón de comparación, el patrón que nos ayuda a lo anterior consiste en lo siguiente :

Se hace uso de una tabla de desviaciones aleatorias normales; se escoge una muestra de tamaño igual al de la muestra ocupada para el modelo, cada elemento de esta muestra se grafica de la misma manera que los residuales, ésta gráfica nos exhibirá el comportamiento de anomalía para muestras de ese tamaño. Si la gráfica de residuales difiere mucho de esta última, esto quiere decir que los residuales sí se comportan como si provinieran de una distribución normal con media cero; en el caso

en que ambas gráficas se asemejen mucho será indicio de que los residuales no siguen un comportamiento de una distribución normal con media cero y entonces se podrá concluir que el modelo parece no cumplir los supuestos especificados.

Ahora bien, en situaciones donde el número de residuales es demasiado grande la gráfica en conjunto de ellos puede hacerse en forma de histogramas o en papel normal.

ii) En papel normal :

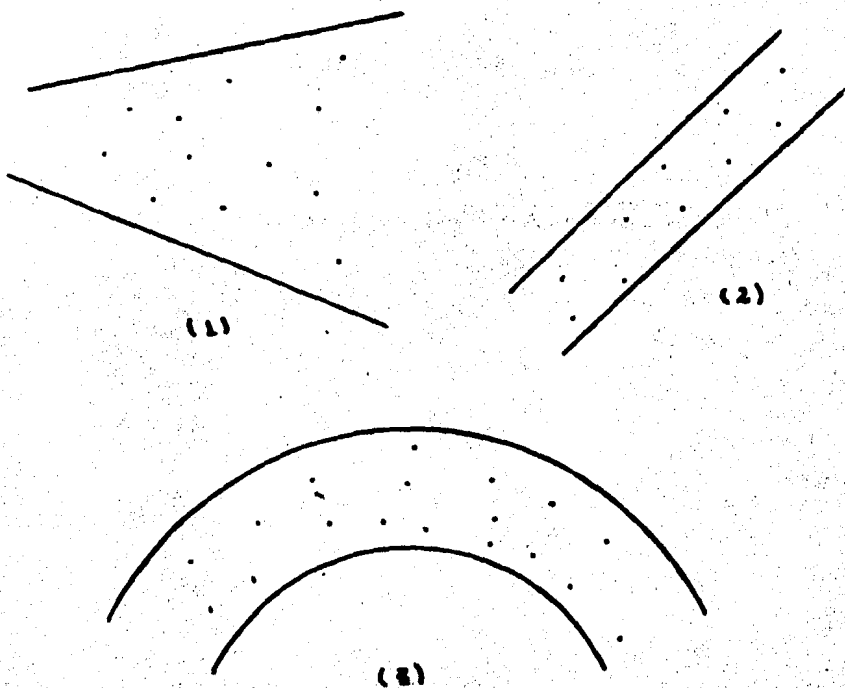
Este es un procedimiento alternativo, para probar si los residuales muestran una tendencia como si provinieran de una población normal. Estas gráficas se hacen en papel normal y se pueden hacer dos tipos de gráficas en este papel : Normal completa o medio normal, en ambas gráficas se podrá checar si existe o no normalidad de la siguiente forma: Si los puntos al graficarlos observan un comportamiento lineal entonces uno podrá decir que en efecto los residuales tienen un comportamiento "normal"; en caso contrario los residuales no se comportan como si proviniesen de una población normal.

Este tipo de gráficas cuando hay un número grande de residuales se pueden hacer graficando por partes, es decir, si tenemos por ejemplo 200 residuales a graficar, se graficarán primero el décimo residual más pequeño, enseguida el vigésimo y así hasta el 180-ésimo y entonces esta gráfica nos revelará el comportamiento en global de todos.

iii) Gráficas contra \hat{y}_i :

Este tipo de gráficas sirven para verificar la homocedasticidad de los errores y nos dan otros

indicios. Lo que se gráfica son los residuales contra sus correspondientes valores ajustados, en el eje de las abscisas va el valor de los ajustados y en el de las ordenadas se coloca el valor del residual correspondiente. Una vez obtenida la gráfica, ésta indicará error en la varianza o no (como ya habíamos mencionado) fijándonos en el efecto o patrón que muestre dicha gráfica, es decir hay que fijarse si muestra un efecto de "banda horizontal", ó que tipo de efecto nos proporciona, si se obtiene un efecto de "banda horizontal" lo que nos indicará es homocedasticidad correcta. El error en la varianza podrá ser indicado por gráficas que muestren la siguiente forma :



Este tipo de gráficas lo que indican es:

La # 1 nos da indicios acerca de la varianza, y puede sugerir que la varianza es no constante, como la supuesta. Hay necesidad de una transformación sobre las observaciones Y_i antes de efectuar un análisis de Regresión.

La # 2 nos indica error en el análisis, la desviación de la línea ajustada es sistemática, (residuales negativos corresponden a valores pequeños de \hat{Y}_i y residuales positivos a valores grandes de \hat{Y}_i). El efecto puede ser también causado por omisión errónea de β_3 en el modelo.

La última de las gráficas nos indica un modelo en el cual no hay linealidad, hay necesidad de términos extras en el modelo, (por ejemplo elevar al cuadrado o producto cruzado de algunos términos), ó también sugiere la necesidad de una transformación sobre las observaciones Y_i antes del análisis.

Uno se podría cuestionar el hecho de que grafiquemos residuales contra valores ajustados y no residuales contra valores observados, esto es debido a que estos últimos son correlacionados y los \hat{Y}_i con los Y_i no lo son. Otra forma de explicar esto anterior es la siguiente :

La pendiente de la línea por mínimos cuadrados que atraviesa a los puntos (\hat{Y}_i, Y_i) tiene un valor de $1 - R^2$, y este valor es cero solo cuando el modelo es un perfecto ajuste; y la pendiente de la línea por mínimos cuadrados que atraviesa los puntos (\hat{Y}_i, \hat{Y}_i) tiene un valor de cero siempre (Dra^{per} y Smith) (15).

La razón por la cual se prefiere que tenga la pendiente un valor cero es debido a que el modelo cumple los supuestos cuando los residuales exhiban un efecto de "banda horizontal" y sin embargo si decidiéramos graficar $\hat{\epsilon}_i$ contra \mathcal{N}_i , como la pendiente de su línea ajustada es $1 - R^2$ y solo es cero cuando el ajuste es perfecto, no tendríamos una evidencia clara de cuándo no se están violando los supuestos.

iv) Gráficas de residuales contra variables explicativas (x_{ji})

La forma de estas gráficas es la misma que la anterior, sólo lo que en lugar de poner los valores ajustados se tienen los valores de cada variable explicativa que nos interesa $x_{j1}; x_{j1}, x_{j2}, \dots, x_{jn}$, es decir se ponen los valores de la columna correspondiente a la variable de interés. Si se obtiene una impresión de banda horizontal esto resultará satisfactorio.

Las anomalías en este tipo de gráficas lo que mostrarán es : (para los mismos efectos anteriores)

- 1) Varianza no constante, necesidad de una transformación preliminar sobre los y_i 's .
 - 2) Error en cálculos, efecto lineal de x_j no tomado en cuenta.
 - 3) Necesidad de términos extras, por ejemplo, un término cuadrático en x_j en el modelo, o una transformación de los y 's.
- v) En conjunto ordenado con respecto al tiempo :

Cuando sea posible conocer el orden en que aparecen los residuales con respecto al tiempo, se puede graficar a estos en función al tiempo. De esta manera se tiene al alcance una información suficiente que proporcione un criterio para saber si el tiempo

afecta al modelo ajustado.

Si el modelo de Regresión propuesto resulta adecuado, se puede esperar que la gráfica muestre un efecto de "banda horizontal" .

Las anomalías en este tipo de gráficas mostrarán lo siguiente :

- 1) La varianza no es constante, ya que aumenta conforme avanza el tiempo, se sugiere una transformación en los Y 's .
- 2) Indica la posibilidad de incluir un término lineal en el tiempo; ó bien los cálculos han sido incorrectos.
- 3) Indica la obligación de incluir un término cuadrático o lineal en el tiempo para el modelo propuesto.

RESIDUALES Y SUS PATRONES DE VARIANZA

Como ya mencionamos un patrón "ideal" para muchas gráficas, que implica no rechazar los supuestos del modelo lineal, ocurre cuando las gráficas de residuales dan un efecto de "banda horizontal".

En los modelos de ajuste esto anterior es usualmente verdadero dentro de las limitaciones prácticas de muchas gráficas, pero no es teóricamente verdadero porque no todos los residuales tienen la misma varianza siempre.

El propósito de esta parte es en dos sentidos :

- i) Poner en evidencia que existe la probabilidad de tener, al menos, un patrón que nos indique hacer cambios en la magnitud de los residuales para tales gráficas, aún cuando los supuestos se cumplan.
- ii) Sugerir que la varianza constante de los residuales debe ser chequeada antes de descartar una observación discordante o aceptar una gráfica de residuales que ha demostrado una falla del modelo.

El estudio e inspección de las varianzas de los residuales puede poner en evidencia una no disposición de los datos para probar lo adecuado del modelo.

Como vimos anteriormente la varianza de los residuales

$$\begin{aligned} \text{Var}(\epsilon_j) &= \sigma^2 - \text{Var}(\hat{y}_j) \\ &= (1 - h_{jj}) \sigma^2 \quad j = 1, 2, \dots, n \end{aligned} \quad \dots(1)$$

Analizando la expresión anterior (1) nos podemos dar cuenta que el patrón de las varianzas de los residuales es el complemento de la de los \hat{y}_i . También, en base a que $0 \leq h_{ii} \leq 1$ de acuerdo a [22], se puede hacer notar que :

$$0 \leq (1 - h_{ii}) < 1$$

Partamos de la expresión (1) de la varianza de los residuales :

$$\text{Var}(\epsilon_i) = \sigma^2 - \text{Var}(\hat{y}_i)$$

Para el análisis de esta última expresión, tomaremos a σ^2 como constante, así que lo que queda por examinar es la $\text{Var}(\hat{y}_i)$ y para esto haremos el siguiente desarrollo para el ajuste del modelo lineal simple :

$$\text{Consideremos el modelo } y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

se puede mostrar que la ecuación ajustada va a ser:

$$\hat{y} = \bar{y} + \hat{\beta}_1 (x - \bar{x}) \quad \text{donde } \bar{x} = \frac{1}{n} \sum x_i$$

donde \bar{y} y $\hat{\beta}_1$ están sujetos a error y tendrán influencia en \hat{y} .

Queremos investigar la $\text{Var}(\hat{y}_i)$, lo primero que mostraremos es que \bar{y} y $\hat{\beta}_1$ son no correlacionados; para esto haremos lo siguiente :

Sean 'a' y 'c' dos constantes tales que :

$$\begin{aligned} a &= a_1 y_1 + a_2 y_2 + \dots + a_n y_n & \text{donde } a_1 \text{ y } c_1 \text{ son constantes.} \\ c &= c_1 y_1 + c_2 y_2 + \dots + c_n y_n \end{aligned}$$

Vamos a calcular la COV (a,c) suponiendo que y_i y y_j son no correlacionados cuando $i \neq j$ y que $\text{Var}(y_i) = \sigma^2 \forall i$

$$\Rightarrow \text{Cov}(a,c) = E(ac) - E(a)E(c)$$

donde :

$$\begin{aligned}
 E(a c) &= E[(a_1 y_1 + a_2 y_2 + \dots + a_n y_n)(c_1 y_1 + c_2 y_2 + \dots + c_n y_n)] \\
 &= E\left[\sum_{i=1}^n a_i c_i (y_i^2) + \sum_{i \neq j} (c_i a_j)(y_i y_j)\right] \\
 &= E(y_i^2) \sum_{i=1}^n a_i c_i + E(y_i y_j) \sum_{i \neq j} c_i a_j \\
 &= a_1 c_1 E(y_1^2) + a_2 c_2 E(y_2^2) + \dots + a_n c_n E(y_n^2) \\
 &\quad + \sum_{i \neq j} c_i a_j E(y_i y_j)
 \end{aligned}$$

Y debido a que y_i es no correlacionado con y_j tenemos :

$$E(a c) = \sum a_i c_i E(y_i^2) + \sum a_i c_j E^2(y_i)$$

Por otro lado

$$\begin{aligned}
 E(a) E(c) &= E[a_1 y_1 + a_2 y_2 + \dots + a_n y_n] E[c_1 y_1 + c_2 y_2 + \dots + c_n y_n] \\
 &= [a_1 E(y_1) + a_2 E(y_2) + \dots + a_n E(y_n)] [c_1 E(y_1) + c_2 E(y_2) + \dots + c_n E(y_n)] \\
 &= \sum_{i=1}^n c_i a_i E^2(y_i) + \sum a_i c_j E(y_i) E(y_j)
 \end{aligned}$$

$$E(a)E(c) = \sum_{i=1}^n a_i c_i E^2(y_i) + \sum a_i c_j E^2(y_i)$$

De donde la covarianza toma la expresión :

$$\begin{aligned} E(a_c) - E(a)E(c) &= \sum a_i c_i E(Y_i^2) + \sum a_i c_j E^2(Y_i) \\ &\quad - \sum a_i c_i E^2(Y_i) - \sum a_i c_j E^2(Y_i) \\ &= \sum a_i c_i [E(Y_i^2) - E^2(Y_i)] \end{aligned}$$

$$\text{Cov}(a, c) = \sum a_i c_i \text{Var}(Y_i)$$

Pero como $\text{Var}(Y_i) = \sigma^2 \quad \forall i = 1, \dots, n$

$$\Rightarrow \text{Cov}(a, c) = \sigma^2 \sum_{i=1}^n a_i c_i$$

Sean $a = \bar{y}$ y $c = \hat{\beta}_1$

En consecuencia se tiene:

$$\sum \frac{y_i}{n} = a_1 y_1 + a_2 y_2 + \dots + a_n y_n$$

$$\Rightarrow \frac{y_i}{n} = a_i y_i$$

$$\Rightarrow a_i = \frac{1}{n}$$

Si $c = \hat{\beta}_1$ se tiene que:

$$\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = c_1 y_1 + c_2 y_2 + \dots + c_n y_n$$

Se tendría que cumplir en general que :

$$\begin{aligned} \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})} &= \sum_{i=1}^n C_i Y_i \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) Y_i \\ &= \sum_{i=1}^n C_i Y_i \quad \Rightarrow \quad C_i = \left(\frac{X_i - \bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) \end{aligned}$$

Partiendo de que

$$\text{Cov}(a, c) = \sum_{i=1}^n a_i c_i \text{Var}(Y_i)$$

tenemos que para $a = \bar{Y}$ y $c = \hat{\beta}_1$

$$\begin{aligned} \text{Cov}(\bar{Y}, \hat{\beta}_1) &= \sum_{i=1}^n \frac{1}{n} \left(\frac{X_i - \bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) \\ &= \frac{1}{n} \left(\frac{(X_1 - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} + \frac{(X_2 - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} + \dots + \frac{(X_n - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) \end{aligned}$$

$$\begin{aligned} \text{Cov}(\bar{Y}, \hat{\beta}_1) &= \frac{1}{n} \left(\frac{X_1 + X_2 + \dots + X_n}{\sum_{i=1}^n (X_i - \bar{X})^2} - \frac{n\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) \\ &= \frac{\sum_{i=1}^n X_i}{n \sum_{i=1}^n (X_i - \bar{X})^2} - \frac{\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} = 0 \end{aligned}$$

El resultado obtenido nos indica que \hat{Y} y $\hat{\beta}_1$ son variables aleatorias no correlacionadas, \therefore la varianza de \hat{Y}_0 para un valor específico X_0 de X es :

$$\text{Var}(\hat{Y}_0) = \text{Var}(\bar{Y}) + (X_0 - \bar{X})^2 \text{Var}(\hat{\beta}_1)$$

Y como $\text{Var}(\bar{Y}) = \frac{\sigma^2}{n}$ y $\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$ tenemos lo siguiente :

$$\text{Var}(\hat{Y}_0) = \frac{\sigma^2}{n^2} + \frac{(X_0 - \bar{X})^2 \sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

tomando ahora a S^2 como un estimador de σ^2 tenemos :

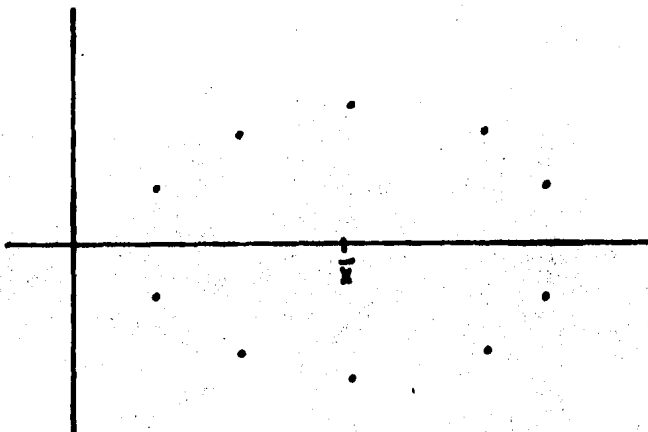
$$\widehat{\text{Var}}(\hat{Y}_0) = S^2 \left(\frac{1}{n^2} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)$$

Analizando esta última expresión nos podemos dar cuenta del comportamiento de la $\widehat{\text{Var}}(\hat{Y}_0)$, éste se puede resumir como sigue :

Alcanza su valor mínimo cuando $X_0 = \bar{X}$ y se va incrementando conforme X_0 se va alejando de \bar{X} . Como una consecuencia de este comportamiento y regresando a analizar el comportamiento de la varianza de los residuales, problema inicial, tenemos que di -

cho comportamiento es como sigue :

Para x 's cercanos a \bar{x} , los residuales, tendrán varianza mayor que para x 's lejanos a \bar{x} . Y debido a esto el patrón de los residuales al graficarlos nos presentará un efecto elipsoidal como se muestra en la siguiente figura:



La forma más alargada o achatada del elipsoide se debe a los valores de x 's usados en el análisis de Regresión y puede ser importante o no para un problema práctico.

Si al graficar la varianza de los residuales se viera que se tiene un efecto en el cual es muy alargado del centro, es decir, la varianza es demasiado grande para x 's cercanos a \bar{x} , se puede concluir que el modelo anda mal en cuanto a varianza constante se refiere, lo mismo se puede concluir para valores muy alejados de \bar{x} la varianza es, ó muy pequeña, ó demasiado grande en relación a las demás.

Si las varianzas de los residuales variáran mucho se podrá examinar $\frac{\xi_i}{(1-h_{ii})^2}$ en lugar de ξ_i en las gráficas usuales de re-

duales. Y usar en forma más correcta $\xi_i / \sqrt{1 - \lambda_{ii}}$ en lugar de ξ_i / S . Esto anterior con el fin de poder concluir algo acerca del análisis basándonos en las gráficas de residuales ya mencionadas con anterioridad.

Una variación grande de la Var (ξ_i) refleja una peculiaridad de la matriz X, llamada el espacio no homogéneo de las observaciones y tiende así a tener atención directa a las deficiencias de los datos.

Tal efecto elipsoidal en la varianza de los residuales podrá usualmente no ser perceptible en una gráfica típica de residuales.

USO DE GRAFICAS DE RESIDUALES PARCIALES

En esta parte definiremos lo que es una gráfica de residuales parciales y veremos su uso dentro del contexto de análisis de Regresión.

Estas gráficas proporcionan una gran ayuda y mejoría para los análisis por medio de Regresión, ya que, proporcionan la misma información que la dada por las gráficas usuales de residuales contra variables independientes, pero además ayudan al analista a :

- 1) Determinar la importancia (en términos del ajuste) de la i-ésima variable independiente en presencia de las otras variables independientes.
- 2) Determinar la importancia de no-linealidad en cualquiera de las variables y hacer la transformación correcta más precisamente. (Este método gráfico ayuda mucho en este sentido pues señala qué transformación y en qué variable exactamente se debe considerar, para obtener un buen análisis de Regresión).

Estableciendo una pequeña comparación entre gráficas usuales de residuales y gráficas de residuales parciales podemos decir lo siguiente :

Mientras las primeras nos señalan en donde puede haber errores por no linealidad en alguna de las variables, las segundas nos proporcionan esta misma información pero además ;

en el caso de que haya linealidad en todas las variables con respecto al ajuste nos proporcionan información acerca de qué está pasando en el ajuste (es decir, se cumplen todos los supuestos pero qué pasa con la línea ajustada) .

Estas gráficas son una representación puntual de mucha de la información disponible para un simple ajuste por mínimos cuadrados .

DEFINICION DE GRAFICAS DE RESIDUALES PARCIALES

En el Análisis de Regresión Lineal Múltiple existe un vector de observaciones Y y una matriz de observaciones de 'p' variables independientes, $X_{(n,p)}$ el estimador de β es :

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

los residuales usuales que hemos manejado son: $\epsilon = Y - X \hat{\beta}$

Para la definición de los residuales parciales consideremos lo siguiente :

Sea X_i^T la i-ésima columna de X ; entonces el vector i-ésimo de residuales parciales se define como:

$$\begin{aligned} Y_i^* &= \epsilon + X_i^T \hat{\beta}_i = Y - X \hat{\beta} + X_i^T \hat{\beta}_i \\ &= Y - \sum_{j \neq i} X_j^T \hat{\beta}_j \end{aligned}$$

Hay que hacer notar que $\hat{\beta}$ es obtenido por ajuste del modelo completo y no el estimador que podría ser obtenido ajustando el modelo sin la i-ésima variable.

Analizaremos en seguida un resultado muy importante acerca de los residuales parciales.

Consideremos la Regresión lineal simple de r_i^* sobre X_i^T . La pendiente de la línea obtenida para el ajuste por mínimos cuadrados para esta Regresión tendrá el mismo valor $\hat{\beta}_i$ que el obtenido como un estimador de β_i para el modelo de Regresión lineal múltiple involucrando las 'p' variables.

Demostración :

La regresión de r_i^* sobre X_i^T es :

$$r_i^* = \beta_i X_i^T + e_i \quad \therefore e_i = r_i^* - \beta_i X_i^T$$

Se trata de localizar $\tilde{\beta}_i$ que minimice $\|r_i^* - \tilde{\beta}_i X_i^T\|$

donde para cualquier vector 's' la norma se define como; $\|s\| = \sum z_i^2$

Por la definición de residuales parciales tenemos :

$$\|Y - \sum_{j \neq i} \hat{\beta}_j X_j^T - \tilde{\beta}_i X_i^T\| \text{----- (1)}$$

donde $\hat{\beta}_j$, es el estimador, por mínimos cuadrados, de β_j para $j \neq i$ cuando el modelo completo es ajustado.

Ahora bien $\tilde{\beta}_i$ únicamente minimiza la expresión (1) cuando $\forall b_i \neq \tilde{\beta}_i$ tenemos que se cumple lo siguiente :

$$\|Y - \sum_{j \neq i} \hat{\beta}_j X_j^T - \tilde{\beta}_i X_i^T\| \leq \|Y - \sum_{j \neq i} \hat{\beta}_j X_j^T - b_i X_i^T\|$$

pero esta última desigualdad no puede ser verdadera para $b_i \neq \hat{\beta}_i$, ya que esto implicaría que el estimador por mínimos cuadrados para el modelo completo no minimizaría la suma del cuadrado de los residuales;

$$\therefore \tilde{\beta}_i = \hat{\beta}_i$$

Este último resultado lo que nos indica es, que la pendiente de la línea ajustada por mínimos cuadrados a la i-ésima gráfica de residuales parciales es igual a $\hat{\beta}_i$, en lugar de cero como en la usual línea ajustada para la gráfica de residuales contra variables independientes. Y esto muestra al analista, como

ayudan estas gráficas de residuales parciales, a determinar la li
nealidad de cada variables (26.0).

Cuando la primera columna de la matriz X es 1, Mallows
sugiere una pequeña modificación a la expresión de los residuales
parciales :

$$V_i^* = \varepsilon + (X_i^T - \bar{X}_i) \hat{\beta}_i + \bar{Y} \quad (26.1)$$

donde \bar{X}_i y \bar{Y} son los vectores de la media para las variables X_i^T
y Y respectivamente. Esta transformación traslada la gráfica ver
ticalmente a la gráfica mencionada antes, de tal manera que la lí
nea de Regresión pasa a través del punto (\bar{X}_i, \bar{Y}) y todos los
puntos quedan dentro del rango de la variable dependiente original.

CAPITULO II

M E T O D O D E L U N D

En el ajuste del modelo lineal por mínimos cuadrados, los residuales ε_i son generalmente analizados para tratar de detectar observaciones discordantes. A menudo se transforman dichos residuales (dividiendolos entre la raíz cuadrada de la suma del cuadrado de los residuales entre $n-p$, es decir $\sqrt{\sum \varepsilon^2 / (n-p)}$) antes de iniciar el análisis de ellos. Pero debido a que la varianza de los residuales no es constante se sugiere dividirlos entre la raíz cuadrada de su varianza estimada con el fin de estandarizarlos. Obteniendo de esta manera los residuales estudentizados

$$t_i = \frac{\varepsilon_i}{[\sum \varepsilon^2 / (n-p) (1-h_{ii})]^{1/2}}$$

El presente método toma como estadística de prueba el máx $|t_i|$.

Debido a que la distribución de esta estadística es muy compleja de encontrar se hace el siguiente desarrollo para obtener cotas superiores de sus valores críticos.

Sea $\tilde{\varepsilon}_i = \varepsilon_i / (n-p)^{1/2}$, su función de densidad de probabilidad marginal esta dada por :

$$f(\tilde{\varepsilon}_i) = C (1 - \tilde{\varepsilon}_i^2)^{(n-p-1)/2} \quad \tilde{\varepsilon}_i \in [-1, 1]$$

la cual es una función student-invertida univariada; donde

$$C = \Gamma\left(\frac{n-p}{2}\right) / \pi^{1/2} \Gamma\left(\frac{n-p-1}{2}\right) \quad (16).$$

Queremos obtener una cota superior para la :

$$P[\text{máx}|t_i| > t_0]$$

pero :

$$P[\max |t_i| > t_0] = P[\max |z_i| > z_0]$$

por lo que basta con encontrar cotas superiores para

$$P[\max |z_i| > z_0]$$

lo cual resulta menos complicado.

Para encontrar esta cota ocupamos una desigualdad de Bonferroni como sigue :

$$\text{Si } \mu_i := P(|w_i| > d)$$

entonces

$$P(\max |w_i| > d) \leq \sum_{i=1}^n \mu_i$$

por lo que

$$P(\max |z_i| > z_0) \leq \sum_{i=1}^n P(|z_i| > z_0)$$

$$= 2n \int_{z_0}^{\infty} f(z) dz \dots (1)$$

[26]

De lo anterior se tiene que, podemos obtener un valor crítico $z_0 = \frac{t_0}{(n-p)^{1/2}}$, por medio de (1) ó igualando $2n \int_{z_0}^{\infty} f(z) dz = \alpha$

para poder encontrar el valor de z_0 y en consecuencia el de t_0 .

Para facilitar esto se hace lo siguiente :

$$\text{Si } 2n \int_{z_0}^{\infty} f(z) dz = \alpha$$

$$\text{entonces } \int_{z_0}^{\infty} f(z) dz = \frac{\alpha}{2n}$$

de este planteamiento se opta por hacer la siguiente transformación para obtener el valor de z_0 buscado

$$\begin{aligned} \int_{-1}^1 f(z) dz &= 1 \Rightarrow \int_0^1 f(z) dz = \frac{1}{2} \\ \Rightarrow \int_0^1 f(z) dz + \int_0^1 f(z) dz &= \frac{1}{2} \\ \Rightarrow \int_0^1 f(z) dz &= \frac{1}{2} - \int_0^1 f(z) dz \\ \Rightarrow \int_0^{\frac{1}{2}} f(z) dz &= \frac{1}{2} - \frac{1}{2n} \quad \therefore \int_0^1 f(z) dz = \frac{1}{2} \left(1 - \frac{1}{n}\right) \end{aligned}$$

Esta última expresión es más fácil de evaluar. Con la idea anterior Lund construyó tablas de cotas superiores de los valores críticos del máx $|t_1|$ para los niveles $\alpha = .10, .01, .05$ con las cuales es posible detectar una observación discordante en forma sencilla y rápida.

El presente método lo podemos resumir de la siguiente forma :

Para realizar una prueba a un nivel α :

- 1.- Calcular el máx $|t_1|$.
- 2.- Encontrar el valor $t_{\alpha, n, p, t}$ en tablas.
- 3.- Si el máx $|t_1| > t_0$ la i -ésima observación será declarada como discordante.

M E T O D O D E E L L E N B E R G

Ellenberg desarrolla un método para detectar una sola observación discordante, en el modelo de Regresión lineal general, el cual está basado en el análisis de los residuales estandarizados estimados por mínimos cuadrados; la estadística de prueba que propone es el máximo de los residuales mínimos cuadrados estandarizados, y se encuentran cotas superiores e inferiores para los puntos porcentaje de la estadística de prueba a partir de las desigualdades de Bonferroni.

Algunos métodos de detección se basan en la reducción de la suma del cuadrado de los residuales al omitir una o varias observaciones, pero con estos métodos es necesario realizar muchos ajustes, algunos otros se basan en el análisis directo de los residuales, pero en este caso puede suceder que la línea o plano ajustado pase cerca del punto discordante y en tal caso el análisis directo de los residuales no funcionaría para indicar al punto discordante; para evitar estos problemas se propone un procedimiento basado en el análisis de los residuales estandarizados (Residuales divididos entre la raíz de la varianza estimada multiplicada por la constante $(n-p)$). La expresión para estos es :

$$Z_i = \frac{\epsilon_i}{(\epsilon' \epsilon (I - A_i))^{1/2}}$$

La estadística de prueba es : $\max |Z_i|$

La regla de decisión queda definida como sigue :

Sí $\max |r_i| \geq C_\alpha$: Rechazamos la hipótesis nula de que no existen observaciones discordantes.

Sí $\max |r_i| < C_\alpha$: No rechazamos la hipótesis nula de que no existen observaciones discordantes.

para un valor crítico C_α con un nivel de significancia α .

Para el desarrollo siguiente es necesario contar con la distribución de r_i , así como la distribución bivariada de (r_i, r_j) las cuales están dadas como sigue [17] .

$$f(r_i) = \frac{\Gamma(v+1)}{\Gamma(\frac{v}{2}) \Gamma(v+\frac{1}{2})} (1 - r_i^2)^{(v-\frac{1}{2})} \quad -1 < r_i < 1, i = 1, \dots, n$$

$$f(r_i, r_j) = \frac{\gamma}{\pi(1-\rho_{ij}^2)^{\frac{v}{2}}} \left[1 - \frac{1}{(1-\rho_{ij}^2)} (r_i^2 - 2\rho_{ij}r_i r_j + r_j^2) \right]^v$$

para $r_i^2 - 2\rho_{ij}r_i r_j + r_j^2 \leq (1-\rho_{ij}^2)$

donde $v = (n-p-2)/2$

$\rho_{ij} = \frac{-h_{ij}}{(h_{ii} h_{jj})^{1/2}}$ es la correlación entre ε_i y ε_j .

La distribución exacta del $\max |r_i|$ es muy difícil de encontrar Sin embargo se pueden obtener cotas superiores e inferiores para los puntos porcentajes del $\max |r_i|$ usando la segunda desigualdad de Bonferroni [16] .

Para cualquier valor crítico C_{α} la desigualdad esta dada por :

$$(1) \quad n \Pr(|\bar{z}_i| > C_{\alpha}) - \sum_{i>j} \Pr(|\bar{z}_i| > C_{\alpha}, |\bar{z}_j| > C_{\alpha}) \\ \leq \Pr(\max |\bar{z}_i| > C_{\alpha}) \leq n \Pr(|\bar{z}_i| > C_{\alpha})$$

para evaluar la cota inferior necesitamos calcular $\binom{n}{2}$ probabilidades bivariadas QD sin embargo esto se puede reducir por un resultado dado por Srikantan (1961) [17], quien demostró que :

$$\Pr\left[(|\bar{z}_i| > C_{\alpha}), (|\bar{z}_j| > C_{\alpha}) \right] = 0 \quad \text{si} \quad C_{\alpha} \geq \left(\frac{1 + |r_{ij}|}{2} \right)^{\frac{1}{2}}$$

por lo que, si determinamos el valor crítico C_{α} para un nivel de significancia dado (α), usando la cota superior dada en (1) tal que $\Pr(|\bar{z}_i| > C_{\alpha}) = \frac{\alpha}{n}$, entonces los pares (\bar{z}_i, \bar{z}_j) para los que $C_{\alpha} \geq \left(\frac{1 + |r_{ij}|}{2} \right)^{\frac{1}{2}}$ no contribuirán para encontrar el valor de la cota inferior; con esto se puede reducir un poco el número de cálculos de probabilidades bivariadas.

Podemos darnos cuenta que, entre menor sea el nivel de significancia mayor será C_{α} por lo cual, teóricamente podemos encontrar un C_{α} apropiado tal que :

$$C_{\alpha} > \left(\frac{1 + |r_{ij}|}{2} \right)^{\frac{1}{2}} \quad \forall i > j$$

con lo cual se igualarían la cota inferior y la cota superior, y no se tendría que calcular alguna probabilidad bivariada.

Para los cálculos en donde la probabilidad bivariada sea mayor que cero, se pueden simplificar los cálculos haciendo lo siguiente :

$$\begin{aligned}
 \Pr(|z_i| > C_\alpha, |z_j| > C_\alpha) &= \Pr(z_i > C_\alpha, z_j > C_\alpha) + \\
 &\Pr(z_i > C_\alpha, z_j < -C_\alpha) + \\
 &\Pr(z_i < -C_\alpha, z_j < -C_\alpha) + \\
 &\Pr(z_i < -C_\alpha, z_j > C_\alpha) \\
 &= 2 \Pr(z_i > C_\alpha, z_j > C_\alpha) + \\
 &2 \Pr(z_i > C_\alpha, z_j < -C_\alpha)
 \end{aligned}$$

para la evaluación de estas probabilidades es necesario hacer uso de la distribución bivariada de (z_i, z_j) , así como de integración numérica, dicha integral se puede simplificar introduciendo la transformación :

$$t_i = (z_i - z_j) / (2(1 - \rho_{ij}))^{\frac{1}{2}}$$

$$t_j = (z_i + z_j) / (2(1 + \rho_{ij}))^{\frac{1}{2}}$$

con lo cual la distribución conjunta esta dada por : (17)

$$g(t_i, t_j) = \frac{v}{\pi} (1 - t_i^2 - t_j^2)^{v-1} \text{ para } t_i^2 + t_j^2 \leq 1$$

Se debe hacer notar que la prueba propuesta por Ellen - berg es equivalente a la propuesta por Richard E. Lund, la principal diferencia esta en las estadísticas que proponen para sus pruebas :

Lund propone $\max |t_1|$ donde $t_1 = z_i / (n-p)^{\frac{1}{2}}$ y para obtener la cota superior para los puntos porcentaje fija su atención en

$$\zeta_i = \frac{t_i}{(n-p)^{1/2}}$$

mientras que para Ellenberg la estadística de prueba es directamente $\max |\zeta_i|$. Por otra parte Ellenberg no solo fija su atención en la cota superior para los puntos porcentaje, sino que también calcula la inferior y para calcular estas cotas parte de ζ_i .

Demostración de la equivalencia entre ambos métodos :

- Lund propone a t_i como :

$$\zeta_i = t_i / (n-p)^{1/2} \quad \text{donde } t_i = \frac{\varepsilon_i}{[(\varepsilon^T \varepsilon / (n-p)(1-h_{ii}))^{1/2}]}$$

substituyendo el valor de t_i obtenemos :

$$\zeta_i = \frac{\frac{\varepsilon_i}{[(\varepsilon^T \varepsilon / (n-p)(1-h_{ii}))^{1/2}]}}{(n-p)^{1/2}} = \frac{\varepsilon_i}{[(\varepsilon^T \varepsilon)(1-h_{ii})]^{1/2}}$$

que son precisamente los residuales en los que Ellenberg basa su análisis.

- La función de distribución para ζ_i es idéntica en ambos casos.

Lund dice que $f(\zeta_i) = C (1 - \zeta_i^2)^{[(n-p-3)/2]}$ $\zeta_i^2 \leq 1$

en donde $C = \Gamma\left(\frac{n-p}{2}\right) / \pi^{1/2} \Gamma\left(\frac{n-p-1}{2}\right)$

substituyendo el valor de C llegamos a que :

$$f(\zeta_i) = \frac{\Gamma\left(\frac{n-p}{2}\right)}{\pi^{1/2} \Gamma\left(\frac{n-p-1}{2}\right)} (1 - \zeta_i^2)^{[(n-p-3)/2]} \quad \zeta_i^2 \leq 1$$

$$f(\xi_i) = \frac{\Gamma\left(\frac{n-p-2}{2} + \frac{2}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{n-p-2}{2} + \frac{1}{2}\right)} (1 - \xi_i^2)^{\left(\frac{n-p-2}{2} - \frac{1}{2}\right)} \quad \xi^2 \leq 1$$

$$= \frac{\Gamma(v+1)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(v+\frac{1}{2}\right)} (1 - \xi_i^2)^{\left(v - \frac{1}{2}\right)} \quad \xi^2 \leq 1$$

donde $v = (n-p-2)/2$

y esta última expresión es la distribución propuesta por Ellenberg.

- Lund se propone encontrar un valor crítico (t_0) para el $\max |t_1|$

pero para ello encuentra primero un valor crítico $\xi_0 = C_{\alpha}$
para el $\max |\xi_i|$ en donde $C_{\alpha} = \frac{t_0}{(n-p)^{1/2}}$

la hipótesis nula de que no existen observaciones discordantes para Lund, se rechaza cuando $\max |t_1| \geq t_0$

$$\iff (n-p)^{1/2} \max |t_1| \geq t_0 (n-p)^{1/2}$$

$$\iff \max |\xi_i| \geq C_{\alpha}$$

que es precisamente el criterio con el cual Ellenberg rechaza la hipótesis nula.

\therefore Las pruebas son equivalentes.

NOTA :

Para efectuar la prueba con el método de Ellenberg podemos ocupar las tablas de Lund, solo hay que dividir su valor entre $(n-p)^{1/2}$.

METODO DE PRESCOTT Y COOK

Es frecuente usar la desigualdad de primer orden de Bonferroni para obtener cotas superiores que aproximen a los valores críticos para el valor absoluto máximo del residual studentizado, para detectar una sola observación discordante en el modelo lineal general. Los métodos disponibles para medir la exactitud de esta cota requieren de integración numérica y esto hay que hacerlo en forma rutinaria. En este método se presenta una alternativa relativamente simple que hace posible medir rutinariamente la exactitud de la prueba para detectar una sola observación discordante usando la desigualdad de Bonferroni.

La estadística de prueba que se usará es el máximo de los valores absolutos de los residuales studentizados, el cual es un criterio ampliamente aceptado para detectar la presencia de dichas observaciones en modelos lineales.

Consideremos el modelo lineal general, con las suposiciones involucradas, incluyendo además la de que no existen observaciones discordantes.

Recordemos también que :

$$s^2 = (\sum \epsilon^2) / (n-p)$$

$$\text{Var}(\xi_i) = \sigma^2 (1-h_{ii})$$

la correlación entre ξ_i y ξ_j está dada por :

$$\rho_{ij} = (-h_{ij}) / [(1-h_{ij})(1-h_{jj})]^{1/2} \text{ para } i \neq j$$

la expresión para el residual studentizado es :

$$t_i = \epsilon_i / (1 - h_{ii})^{1/2} S$$

La estadística básica de prueba para detectar una sola observación discordante es el $\max |t_i|$.

Debido a que esta estadística es usualmente asociada con una alternativa, en la cual las suposiciones del modelo general son correctas excepto porque la esperanza de algún elemento de ϵ puede ser distinto de cero, se sugiere usar como estadística de prueba al $\max |V_i|$, donde:

$$\begin{aligned} V_i &= t_i / (n-p)^{1/2} \\ &= \epsilon_i / [\epsilon' \epsilon (1 - h_{ii})^{-1}]^{1/2} \end{aligned}$$

La distribución de V_i es muy difícil de encontrar, por lo que para obtener cotas superiores e inferiores para los valores críticos del $\max |V_i|$ usaremos el siguiente desarrollo:

$$\text{Sean } \alpha_i = \Pr (|V_i| > d)$$

$$\alpha_{ij} = \Pr (|V_i| > d, |V_j| > d) \quad \forall i \neq j$$

las cotas de Bonferroni de primer orden aplicadas a los eventos $\{|V_i| > d\}$ nos lleva a : (10)

$$\begin{aligned} \sum P_v (|V_i| > d) - \sum_{(i,j)} P_v (|V_i| > d, |V_j| > d) \\ \leq P_v (\max |V_i| > d) \leq \sum_i P_v (|V_i| > d) \end{aligned}$$

usando lo anterior, la desigualdad la podemos escribir como:

$$\sum_i \alpha_i - \sum_{i,j} \alpha_{ij} \leq \Pr(\max_i |r_i| < d) \leq \sum_i \alpha_i \quad (1)$$

Ya que bajo la suposición de que no hay observaciones discordantes, los $|r_i|$'s se distribuyen idénticamente, la cota superior puede ser convenientemente expresada en términos de una sola variable aleatoria F con distribución F con 1 y $(n-p-1)$ grados de libertad

$$\begin{aligned} \sum_i \alpha_i &= n \Pr(|r_i| > d) \\ &= n \Pr(r_i^2 > d^2) \\ &= n \Pr[F > d^2 (n-p-1) / (1-d^2)] \end{aligned} \quad (2)$$

con lo cual obtenemos una expresión fácilmente calculable para las cotas superiores.

Evaluar las cotas inferiores en (1) directamente es más complicado, dado que las α_{ij} 's tendrían que ser evaluadas mediante integración numérica, como se vio en el método desarrollado por Ellenberg; en lugar de esto, utilizaremos una aproximación para las α_{ij} , que obtendremos como sigue:

$$\begin{aligned} \alpha_{ij} &= \Pr[|r_i| > d, |r_j| > d] \\ &= \Pr[r_i > d, r_j > d] + \Pr[r_i > d, r_j < -d] \\ &+ \Pr[r_i < -d, r_j < -d] + \Pr[r_i < -d, r_j > d] \end{aligned}$$

Ya que la distribución de (T_i, T_j) es simétrica (17) tenemos que :

$$\alpha_{ij} = 2 P_r (T_i > d, T_j > d) + 2 P_r (T_i > d, -T_j > d)$$

claramente

$$P_r (T_i > d, \pm T_j > d) \leq P_r (T_i \pm T_j > 2d)$$

y entonces tenemos que

$$\alpha_{ij} \leq 2 P_r (T_i + T_j > 2d) + 2 P_r (T_i - T_j > 2d)$$

de donde

$$\alpha_{ij} \leq P_r [(T_i + T_j)^2 > 4d^2] + P_r [(T_i - T_j)^2 > 4d^2]$$

Para facilidad de escritura denotemos como:

$$\beta_{ij}^+ = P_r [(T_i + T_j)^2 > 4d^2]$$

$$\beta_{ij}^- = P_r [(T_i - T_j)^2 > 4d^2]$$

por lo que podemos escribir

$$\alpha_{ij} \leq \beta_{ij}^+ + \beta_{ij}^-$$

si hacemos uso de este resultado y de (1) obtendremos que

$$\alpha = \sum_{i,j} (\beta_{ij}^+ + \beta_{ij}^-) \leq P_r (\max |T_i| > d) \leq \alpha \dots (5)$$

donde

$$\alpha = \sum_i \alpha_i = n P_r (|T_i| > d)$$

la evaluación de la cota inferior a partir de esta última expresión es directa, una vez que se nota que :

$$(\bar{T}_i \pm \bar{T}_j)^2 (n-p-1) / [2(1 + \rho_{ij}) - (T_i + T_j)^2] \dots (4)$$

tiene una distribución F con 1 y (n-p-1) grados de libertad; este resultado fue encontrado por Milenberg (1976) (17) .

Se puede demostrar usando (4) que :

$$\beta_{ij}^- = 0 \quad \text{y} \quad \beta_{ij}^+ = 0 \quad \text{cuando} \quad 2d^2 > (1 \pm \rho_{ij}) \dots (5)$$

Sea

$$c(\pm) = \{ (i, j) / i < j, 2d^2 < (1 \pm \rho_{ij}) \}$$

usando (3), (4), (5) se puede obtener la forma final :

$$\alpha - \beta^+ - \beta^- \leq P_r (\max |T_i| > d) \leq \alpha \dots (6)$$

donde

$$\alpha = n P_r [F > d^2 (n-p-1) / (1-d^2)]$$

$$\beta^+ = \sum_{(i,j)} P_r [F > d^2 (n-p-1) / (1/2(1+\rho_{ij})-d^2)]$$

$$\beta^- = \sum_{(i,j)} P_r [F > d^2 (n-p-1) / (1/2(1-\rho_{ij})-d^2)]$$

Algunos comentarios pueden hacerse a partir de la forma final.

Primero, la cota inferior depende de la distribución

conjunta de (\bar{Y}_i, \bar{Y}_j) a través de la correlación de residuales r_{ij} , la cual depende solamente del valor de las variables explicativas y estas son conocidas.

Segundo, es inmediato de (6) que la cota superior es exacta cuando $C(+)$ y $C(-)$ son vacíos, es decir cuando :

$$1 + \max_{i < j} |r_{ij}| < 2d^2$$

Tercero, el cálculo de la cota inferior requiere del conocimiento de los r_{ij} $\forall i \neq j$ lo cual en algunos casos hace que el cálculo sea muy largo, y en tales situaciones se recomienda reemplazar a r_{ij} en $\beta^+(\beta)$ por el $\max_{C(+)} r_{ij}$ ($\min_{C(-)} r_{ij}$)

Explicación breve del método.

Calcular el valor $t_i \forall i$, encontrar el $\max |t_i|$, a partir de este calcular $\max |r_i|$, dado un α

Sí $\max |r_i| \geq d$: Rechazar la hipótesis nula de que no existen observaciones discordantes.

Sí $\max |r_i| < d$: No rechazar la hipótesis nula de que no existen observaciones discordantes.

METODO DE P. PRESCOTT.

P. Prescott desarrolla una estadística de prueba para detectar una sola observación discordante en modelos generales de Regresión lineal, haciendo uso de residuales studentizados, y sugiere que los valores críticos de esta prueba se pueden aproximar adecuadamente por cotas superiores para los valores críticos de una estadística similar, que considera residuales estandarizados pero por su desviación estándar constante.

Este procedimiento es aplicable a cualquier modelo lineal y no necesita un nuevo análisis con el valor discordante omitido.

Desarrollo.-

Tomemos en cuenta las suposiciones del modelo lineal general y recordemos que :

$$\text{Var}(\boldsymbol{\varepsilon}) = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\sigma^2$$

Si en esta expresión en lugar de σ^2 ocupamos su estimador ($S^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} / (n-p)$) tendremos :

$$\widehat{\text{Var}}(\boldsymbol{\varepsilon}) = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')S^2$$

que es la varianza estimada de los residuales.

Notemos que la desviación estándar estimada del i -ésimo residual es S_i , donde S_i^2 es el i -ésimo elemento de la diagonal de $\widehat{\text{Var}}(\boldsymbol{\varepsilon})$, es decir:

$$S_i^2 = (1 - h_{ii})S^2$$

Varios autores han examinado los residuales y en base a ellos han propuesto algunos tipos de prueba para detectar obser

vaciones discordantes. Entre estos podemos mencionar a Benhken y Draper quienes examinan el patrón de los residuales (6.1) y sugieren que si las varianzas de los residuales varían mucho, sería mejor usar ϵ_i/S en lugar de ϵ_i/S_1 como la forma de desviación normal de los residuales.

En 1973 Tietjen, Moore y Beckman propusieron un método para detectar una sola observación discordante en el modelo de regresión lineal simple ocupando como estadística de prueba a :

$$R_n = \max \left| (\epsilon_i / S_1) \right|$$

y determinaron valores críticos para esta estadística. Los resultados que ellos obtuvieron sugieren que se pueden obtener aproximaciones muy cercanas a los valores críticos exactos, suponiendo que las varianzas de los residuales son razonablemente constantes y usando el valor promedio de las varianzas en el desarrollo de los puntos porcentaje de las estadísticas de prueba. Es bueno hacer notar que Tietjen, Moore y Beckman obtuvieron puntos porcentaje para su estadística mediante un estudio de simulación a gran escala, pero solo lo hicieron para el modelo lineal simple, ya que para modelos más generales esto se complica demasiado. Con el propósito de poder salvar esta gran dificultad, así como el de obtener puntos porcentaje de R_n suficientemente exactos, Prescott propone usar una estadística muy semejante a R_n con la cual sea más fácil obtener cotas superiores para sus puntos porcentaje, y que sean muy cercanos a los puntos porcentaje de R_n .

Consideremos $R_n^* = \max \left| \epsilon_i / S \right|$ donde S^2 es el estimador de la varianza promedio de los residuales, Behnken y Draper (1972)(6.1)

mostraron que la varianza promedio de los residuales es :

$$(n-p) \bar{V}^2 / n$$

de donde podemos concluir que $\bar{S}^2 = (n-p)S^2 / n$ usando simplemente el estimador de \bar{V}^2 , con esto podemos sustituir el valor de \bar{S} en R_n^* y llegamos a :

$$R_n^* = \max |\epsilon_i| / [(n-p)S^2/n]^{\frac{1}{2}}$$

aquí podemos sustituir el valor de S^2 y obtener:

$$R_n^* = \max |\epsilon_i| / [(n-p) \frac{\sum \epsilon_i^2}{(n-p)} / n]^{\frac{1}{2}}$$

$$= \max |\epsilon_i| / [\sum_{i=1}^n \epsilon_i^2 / n]^{\frac{1}{2}}$$

$$= n^{\frac{1}{2}} \max |\epsilon_i| / [\sum_{i=1}^n \epsilon_i^2]^{\frac{1}{2}}$$

$$= n^{\frac{1}{2}} \max |z_i| \dots \dots \dots (1)$$

donde $z_i = \epsilon_i / [\sum_{i=1}^n \epsilon_i^2]^{\frac{1}{2}}$

Para obtener cotas superiores para los puntos porcentaje de R_n^* , podemos ocupar un resultado desarrollado por Stefansky (28)

quién consideró al residual máximo normado $\max |z_1|$ como una estadística de prueba para observaciones discordantes en diseño de experimentos; y encontró que $U_1 = [(n-p)F / \{n(n-p-1+F)\}]^{\frac{1}{2}}$ es una cota superior para los puntos porcentajes del $\max |z_1|$ al $100(1 - \alpha)\%$, donde F es el punto porcentaje $100(1 - \frac{\alpha}{2})$ de la distribución F con 1 y $n-p-1$ grados de libertad.

Si analizamos (1) nos daremos cuenta que en R_n^* tenemos involucrado al residual máximo normado, y de lo anterior podemos

deducir que una cota superior para los puntos porcentaje al $100(1 - \alpha)\%$ de R_n^* es :

$$C = \left[(n-p)F/(n-p-1+F) \right]^{1/2} \dots (3)$$

Para el caso de un modelo lineal general tendremos que las cotas para los puntos porcentaje de R_n^* se pueden obtener de :

$$\left[(n-2)F/(n-3+F) \right]^{1/2}$$

Dado lo anterior podemos definir la siguiente regla de decisión para detectar una sola observación discordante en el modelo lineal general :

Si $\max |t_i|/s_1 \geq C$: Rechazamos la hipótesis de que no existen observaciones discordantes en el modelo.

Si $\max |t_i|/s_1 < C$: No rechazamos la hipótesis de que no existen observaciones discordantes en el modelo.

Breve explicación sobre el uso de este método :

Calcular primero el valor de los residuales studentizados para cada uno de los residuales, es decir calcular $t_i = \hat{e}_i/s_1$

$$\forall i = 1, 2, 3, \dots, n$$

Encontrar el $\max |t_i|$ (que es la estadística de prueba).

Calcular el valor de la cota superior de R_n^* , es decir calcular $\left[(n-p)F/(n-p-1+F) \right]^{1/2}$.

Aplicar la regla de decisión para determinar si existen observaciones discordantes en el modelo.

M E T O D O D E C O O K

Como mencionamos en la introducción varios autores han propuesto varios métodos para aislar observaciones "sospechosas" que en un momento dado pudieran distorsionar la información que se va a obtener de ellas, entre los criterios propuestos para detectar tales observaciones podemos encontrar el análisis de residuales 'studentizados' (t_i) (es decir el residual dividido por su error estándar) y también se vió la necesidad de analizar la varianza estimada de los residuales [6.1] debido a que ésta contiene información más importante que la suministrada por los t_i acerca de las deficiencias de los datos.

Sin embargo pudiéramos ocupar, para detectar observaciones discordantes las dos medidas anteriores al mismo tiempo :

$$(t_i , \widehat{\text{Var}} (\epsilon_i))$$

teniendo así dos medidas asociadas con cada observación . Uno de los pasos naturales a seguir es el de examinar los efectos al quitar tal(es) observaciones, pero ahora nos enfrentamos al problema de decidir que observación quitar ya que c/u tiene dos medidas asociadas (t_i , $\widehat{\text{Var}} (\epsilon_i)$) y puede suceder que mientras una nos indique quitarla la otra nos indique lo contrario.

R. Dennis Cook propone una medida que combina ambas : (t_i , $\widehat{\text{Var}} (\epsilon_i)$) con la cual también se podrán aislar observaciones críticas y resolver el problema anterior.

Bajo las suposiciones del Modelo General de Regresión Lineal y teniendo en cuenta que :

$$\text{Var} (\hat{Y}) = X(X'X)^{-1}X' \Sigma^2 \dots(1)$$

$$\text{Var} (\hat{\epsilon}_i) = (I_n - X(X'X)^{-1}X') \Sigma^2 \dots(2)$$

son las matrices de varianza y covarianza de \hat{Y} y $\hat{\epsilon}$ respectivamente, podemos empezar el desarrollo del método de Cook.

Desarrollo :

De acuerdo con el planteamiento del problema lo que se busca es una medida que asocie las informaciones suministradas por t_i y $\hat{\epsilon}_i$ procedamos a encontrarla.

Primeramente sabemos que un elipsoide al $(1 - \alpha)$ 100% de confianza para β está dado por todos aquellos β^* que satisfagan :

$$\frac{(\beta^* - \hat{\beta})' X'X (\beta^* - \hat{\beta})}{p s^2} \leq F(p, n-p, 1-\alpha) \dots (3)$$

donde $s^2 = \epsilon' \epsilon / (n-p)$

Ahora bien, para calcular que tanta influencia tiene la i -ésima observación sobre el ajuste, resulta natural calcular el estimador mínimo cuadrático de β omitiendo dicha observación ($\hat{\beta}_{(i)}$) y una medida inmediata y que proporciona mucha información es la de la distancia de $\hat{\beta}$ a $\hat{\beta}_{(i)}$.

Una medida fácilmente interpretable de esta distancia es en el lado izquierdo de (3). Por lo cual podemos definir para la i -ésima observación lo siguiente :

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' X'X (\hat{\beta}_{(i)} - \hat{\beta})}{P S^2} \dots (4)$$

Para $i = 1, 2, \dots, n$

Esta es una medida en términos de niveles descriptivos de significancia. Por ejemplo si $D_i \approx F(p, n-p, .3)$ el eliminar la i -ésima observación moverá al estimador mínimo cuadrático de β a la orilla de una región de confianza del 70% para β basado en β .

El cálculo de D_i se puede hacer fácilmente recordando que :

$$\begin{aligned} \hat{\beta} - \hat{\beta}_{(i)} &= \frac{(X'X)^{-1} x_i' [y_i - x_i \hat{\beta}]}{1 - h_{ii}} \\ &= \frac{(X'X)^{-1} x_i' (e_i)}{1 - h_{ii}} \quad (5) \end{aligned}$$

Substituyendo (5) en (4) tenemos:

$$D_i = \frac{\left[\frac{-(X'X)^{-1} x_i' [y_i - x_i \hat{\beta}]}{1 - h_{ii}} \right] X'X \left[\frac{-(X'X)^{-1} x_i' [y_i - x_i \hat{\beta}]}{1 - h_{ii}} \right]}{P S^2}$$

$$D_i = \frac{[y_i - x_i \hat{\beta}]' \{ [X'X]^{-1} x_i \}' X'X (X'X)^{-1} x_i [y_i - x_i \hat{\beta}]}{(1 - h_{ii}) PS^2}$$

$$D_i = \frac{(y_i - \hat{\beta} x_i)' x_i \{ (X'X)^{-1} \}' x_i (y_i - x_i \hat{\beta})}{(1 - h_{ii})^2 PS^2}$$

$$D_i = \frac{(y_i - \hat{\beta} x_i)' x_i (X'X)^{-1} x_i' (y_i - x_i \hat{\beta})}{(1 - h_{ii})^2 PS^2}$$

$$D_i = \frac{(y_i - \hat{\beta} x_i)' h_{ii} (y_i - x_i \hat{\beta})}{(1 - h_{ii})^2 PS^2}$$

$$D_i = \frac{(y_i - \hat{\beta} x_i)' (y_i - x_i \hat{\beta}) h_{ii}}{(1 - h_{ii})^2 PS^2}$$

$$D_i = \frac{(y_i - \hat{\beta} x_i)^2 h_{ii}}{(1 - h_{ii})^2 PS^2}$$

$$D_i = \frac{(y_i - \hat{\beta} x_i)^2 h_{ii}}{(1 - h_{ii})^2 PS^2}$$

$$D_i = \frac{[y_i - \hat{\beta} x_i]^2}{S \sqrt{1-h_{ii}}} \frac{h_{ii}}{p(1-h_{ii})}$$

Como $y_i - \hat{\beta} x_i = \epsilon_i$ tenemos

$$D_i = \left[\frac{\epsilon_i}{S \sqrt{1-h_{ii}}} \right]^2 \frac{h_{ii}}{p(1-h_{ii})} \dots (6)$$

Notemos que ahora en D_i se encuentran tres medidas de importancia, t_i (i-ésimo residual studentizado) y además por (1) tenemos que $\text{Var}(\hat{y}_i) = x_i (X'X)^{-1} x_i' \sigma^2 = h_{ii} \sigma^2$ y por (2) tenemos $\text{Var}(\epsilon_i) = (1 - h_{ii}) \sigma^2 \therefore$ tenemos que

$$\frac{\text{Var}(\hat{y}_i)}{\text{Var}(\epsilon_i)} = \frac{h_{ii} \sigma^2}{(1-h_{ii}) \sigma^2} = \frac{h_{ii}}{(1-h_{ii})}$$

$\therefore D_i$ puede obtenerse a través de :

$$D_i = \frac{t_i^2}{p} \frac{\text{Var}(\hat{y}_i)}{\text{Var}(\epsilon_i)} \dots (7)$$

En donde t_i^2 es una medida por medio de la cual la i-ésima observación puede ser considerada como una observación discordante; mientras que la razón $\text{Var}(\hat{y}_i)/\text{Var}(\epsilon_i)$ mide la importancia que tiene la i-ésima observación en la determinación de $\hat{\beta}$. Un valor grande de esta razón indicará que la i-ésima observación tiene un gran peso en la determinación de $\hat{\beta}$. Hay que notar que

implícitamente en (7) ya tenemos involucradas las medidas que deseábamos combinar (t_i , $\widehat{\text{Var}}(\epsilon_i)$).

Cuando consideremos D_1 debemos tener cuidado de verificar que $\text{Var}(\epsilon_i) \neq 0$ ya que si $\text{Var}(\epsilon_i) = 0$ D_1 no estaría definida porque estaríamos efectuando una división entre cero, pero este es un caso extremo en el cual; debido a un lema (6):

'Si la varianza de $\epsilon_i = 0$ entonces el i-ésimo residual es cero y esto solo sucede cuando la matriz $X_{(i)}$ es de rango menor que p.

Hasta aquí hemos desarrollado la expresión para D_1 , ahora lo que haremos es explicar lo siguiente :

- 1) El porque la importancia de la medida D_1 .
- 2) El porque con D_1 podemos detectar observaciones discordantes.
- 3) Explicación breve del método.

La importancia de D_1 radica en que, como ya dijimos y podemos ver por su expresión en (7) combina dos medidas de detección de observaciones discordantes.

El porque detectamos observaciones discordantes con este método es debido a que implícitamente D_1 es una medida entre $\hat{\beta}$ y $\hat{\beta}^{(i)}$ (podemos recordar que para D_1 se partió y ocupó precisamente una medida de esta distancia en términos de niveles descriptivos de significancia). El hecho de manejar una medida de esta distancia en esos términos lo que nos quiere hacer notar es que mientras mayor sea el elipsoide de confianza en el cual encontramos $\hat{\beta}^{(i)}$ (es decir una confianza muy grande) lo que es ta indicando es que la distancia entre $\hat{\beta}$ y $\hat{\beta}^{(i)}$ es muy grande,

es decir, $\hat{\beta}$ difiere mucho de $\hat{\beta}_{(i)}$ y \therefore el eliminar la i -ésima observación cambia mucho la estimación de β esto nos está indicando que dicha observación es muy influyente para la estimación de β debido a esa anterior razón; y consecuentemente, sí al calcular D_i y buscar en tablas notamos que este valor se encuentra en unas cuyo nivel de significancia sea muy grande, lo que nos está indicando es que al eliminar la i -ésima observación nos manda al estimador de β a un elipsoide pequeño o muy pequeño (dependiendo del nivel de significancia) y esto nos indicará que la distancia entre $\hat{\beta}$ y $\hat{\beta}_{(i)}$ es muy pequeña; es decir la i -ésima observación no es muy influyente pues al quitarla no se modifica mucho la estimación del parámetro debido a que $\hat{\beta}$ difiere muy poco de $\hat{\beta}_{(i)}$.

Breve explicación de como se efectúa el método:

El primer paso consiste en determinar para cada observación su correspondiente D_i , nos fijamos en el $\max D_i$ y suponemos que $D_i \sim F(p, n-p, 1-\alpha)$ buscamos en tablas de la F con: p -no. de parámetros en el modelo, n el tamaño de la muestra

y precisamente el valor que desconocemos y que vamos a encontrar en tablas es α , esto para ver a que región de confianza (elipsoide) nos manda al estimador de β al quitar la i -ésima observación.

Dependiendo del tamaño de α (como ya explicamos antes) es que se determinará si esta i -ésima observación es, o no es influyente para la estimación de β .

METODO DE VALENCIA

Presentaremos ahora un método para detectar k observaciones discordantes que está basado en la magnitud de los residuales studentizados, obtenidos después de omitir cada una de las observaciones; aunque también toma muy en cuenta la influencia de las observaciones; es importante notar que en este método el tratamiento de las observaciones discordantes es fundamental.

Frecuentemente son usados los residuales studentizados para detectar observaciones discordantes, aunque la mayoría los usa teniendo en cuenta el conjunto completo de observaciones; en este método se define al i -ésimo residual studentizado pero cuando la i -ésima observación es omitida.

Si eliminamos a y_i y al i -ésimo renglón de la matriz X y en base a las $n-1$ observaciones restantes se efectúa el ajuste, entonces el i -ésimo residual lo calcularíamos como $y_i - \hat{y}_{(i)}$, donde $\hat{y}_{(i)}$ es la estimación de y_i en el modelo en que no se incluye a y_i ; es decir:

$$y_{(i)} = X_{(i)} \beta_{(i)}$$

además:

$$\text{Var}(\hat{y}_{(i)}) = x_{(i)} (X'_{(i)} X_{(i)})^{-1} x'_{(i)} \sigma^2$$

$$\text{y } \text{Var}(y_i - \hat{y}_{(i)}) = \left[1 + x_{(i)} (X'_{(i)} X_{(i)})^{-1} x'_{(i)} \right] \sigma^2$$

luego si denotamos como $S^2_{(i)}$ a la suma de residuales al cuadrado

entre $(n-p-1)$, obtenida al hacer el ajuste sin la i -ésima observación, entonces se puede definir :

$$z_{(i)} = \frac{y_i - \hat{y}_{(i)}}{\left[(1 + x_{(i)} (X_{(i)}^T X_{(i)})^{-1} x_{(i)}^T) \right]^{\frac{1}{2}}}$$

el cual tiene media cero y varianasa y finalmente definir al i -ésimo residual studentizado sin la i -ésima observación como :

$$t_{(i)} = \frac{z_{(i)}}{s_{(i)}} = \frac{y_i - \hat{y}_{(i)}}{\left[s_{(i)} (1 + x_{(i)} (X_{(i)}^T X_{(i)})^{-1} x_{(i)}^T) \right]^{\frac{1}{2}}}$$

se puede demostrar que $t_{(i)}$ tiene una distribución t con $n-p-1$ grados de libertad.

Aunque los residuales studentizados son de gran utilidad no es suficiente tratar de detectar a las observaciones discordantes basándose únicamente en la magnitud de estos residuales, dado que pueden existir observaciones discordantes con mucha influencia en el ajuste ocasionando con ello que los residuales studentizados de las observaciones no discordantes sean muy grandes. Es por esto importante analizar si las posibles observaciones discordantes lo son en realidad. Este análisis se puede efectuar observando el cambio que se produce en la forma de explicar a la observación por medio del ajuste cuando se está omitiendo dicha observación. Esto se puede llevar a cabo observando que tanto difiere $x_{(i)} \hat{\beta}$ de $x_{(i)} \hat{\beta}_{(i)}$; el tamaño de este cambio se puede comparar contra el error estándar de y_i , $S \sqrt{h_{ii}}$. Una mejora adicio

nal sería que una vez que se detectan a las observaciones discordantes tratarlas y volver a buscar observaciones discordantes en el nuevo conjunto de observaciones.

Podríamos ahora resumir el método; pero antes necesitamos definir a $|t_{(m)}|$ como el máximo de los residuales studentizados al haber omitido cada una de las observaciones, y y_m la observación correspondiente.

METODO.-

- 1) Para $C \in \mathbb{R}^+$ detectar como posible observación discordante a y_m si $|t_{(m)}| > C$.
- 2) Si y_m es declarada como posible observación discordante entonces calcular $x_{(m)} (\hat{\beta} - \hat{\beta}_{(-m)}) / S \sqrt{h_{mm}}$ para decidir si y_m es "peligrosa" para el ajuste.
- 3) Si $x_{(m)} (\hat{\beta} - \hat{\beta}_{(-m)})$ es significativa declarar a y_m como discordante, en caso contrario tomar como y_m a la siguiente y_1 tal que $|t_{(1)}| > C$ y ejecutar el paso 2. Esto se repite hasta encontrar una observación discordante o acabar con los $t_{(i)}$ tales que $|t_{(i)}| > C$ en cuyo caso el método acaba y se puede decir que no existen observaciones discordantes.
- 4) Una vez que y_m sea declarada como discordante se procede a tratarla.
- 5) Realizar el ajuste con el nuevo conjunto de observa-

ciones y obtener los nuevos residuales studentizados omitiendo cada una de las observaciones.

6) Ir al paso 1 .

Es importante notar que la C puede escogerse de tal manera que $\Pr(|t_{(1)}| > C) = \alpha$ para una α dada, con lo cual se obtiene que :

$$C = \left[(n-p)^2 F / n(n-p-1+F) \right]^{1/2} \quad [31]$$

donde F es el cuantil 100 $(1-\alpha/n)$ de una distribución F con 1 y $n-p-1$ grados de libertad.

Observando los siguientes resultados (Hoaglin y Welsh)

[22]

$$t_{(1)} = \epsilon_i (S_{(1)} (1-h_{ii})^{-1})^{-1}$$

$$S_{(1)}^2 = \left[(n-p)S^2 - \epsilon_i^2 / (1-h_{ii}) \right] \left[(n-p-1)^{-1} \right]$$

$$\hat{\beta} - \hat{\beta}_{(i)} = (X'X)^{-1} x'_{(1)} \epsilon_i (1-h_{ii})^{-1}$$

$$x_{(1)} (\hat{\beta} - \hat{\beta}_{(i)}) = h_{ii} \epsilon_i (1-h_{ii})^{-1}$$

nos podemos dar cuenta que solo es necesario efectuar el ajuste con todas las observaciones, obtener los residuales y conocer los elementos de la diagonal de la matriz sombrero, para poder efectuar el método.

M E T O D O G R A F I C O

En esta parte presentaremos un método gráfico para la detección de observaciones discordantes, propuesto por Atkinson.

Este método consta de dos partes, en la primera se analiza el comportamiento gráfico de los residuales studentizados sin la i -ésima observación contra la mediana de números aleatorios, en la segunda se analizará el comportamiento gráfico de la estadística de Cook modificada contra la mediana de números aleatorios también.

Es importante notar porque se usan, para este procedimiento gráfico, a los residuales studentizados sin la i -ésima observación. Esto se debe a que en ausencia de observaciones discordantes S^2 y $S_{(i)}^2$ serán estimadores insesgados de σ^2 y los dos conjuntos de residuales (aquí nos referimos a los dos conjuntos con los que mejor se puede trabajar para detectar observaciones discordantes).

1.- Residuales studentizados

$$t_i = \frac{y_i - \hat{y}_i}{s \sqrt{1 - h_{ii}}}$$

2.- Residuales studentizados sin la i -ésima observación.

$$t_{(i)} = \frac{y_i - x_{(i)} \hat{\beta}_{(i)}}{s_{(i)} \sqrt{1 + x_{(i)} (X_{(i)}^T X_{(i)})^{-1} x_{(i)}}} \left\{ \frac{1}{2} \right\}$$

serán equivalentes.

Pero en la presencia de una observación discordante, t_i no revelaría tan claramente dicha observación como $t_{(i)}$, ya que en $t_{(i)}$ tenemos impactada la presencia de esa observación al calcular $S_{(i)}$, mientras que en t_i esa observación discordante queda un tanto "disfrazada" al tener simplemente S en la cual no se esta reflejando el impacto de esa i -ésima observación.

Para la segunda parte de este método gráfico la estadística de Cook es :

$$D_i = \left[\frac{y_i - x_{ci} \hat{\beta}}{S \sqrt{1-h_{ii}}} \right]^2 \frac{h_{ii}}{p(1-h_{ii})}$$
$$= \frac{t_i^2 h_{ii}}{p(1-h_{ii})}$$

y mide el efecto de cada observación sobre el vector de parámetros estimados $\hat{\beta}$.

Esta estadística puede ser mejorada como sigue :

- 1) Usando $S_{(i)}$ en lugar de S (esto para ver claramente el impacto que tiene la i -ésima observación al ser omitida)
- 2) Usando la raíz cuadrada de D_i la cual puede ser mejor examinada sobre una gráfica medio normal y multiplicando por $\sqrt{n-p}$ para estandarizar a D_i (4).

Con lo anterior obtenemos la modificación de la estadística de Cook como :

$$T_i = \left[\left(\frac{n-p}{p} \right) \left(\frac{h_{ii}}{1-h_{ii}} \right) \right]^{\frac{1}{2}} |t_{ci}|$$

El procedimiento consiste entonces en graficar por un lado $t_{(i)}$ contra la mediana de números aleatorios como una gráfica normal y detectar como 'sospechosas' a aquellas observaciones que se encuentren muy alejadas de las demás en dicha gráfica.

En seguida hacer la gráfica normal de T_i contra la mediana de números aleatorios, y las observaciones sospechosas que vuelvan a encontrarse alejadas de las demás en esta gráfica, serán declaradas automáticamente como discordantes.

En el caso de que una observación sea detectada como sospechosa con la primera gráfica, pero con la segunda no se vea alejada gráficamente del resto, diremos que es sospechosa pero no es 'peligrosa' para el ajuste y mediante este método no podrá ser declarada como discordante.

METODO DE GENTLEMAN Y WILK

Este método se basa directamente en el uso de los residuales y requiere, para cuestiones prácticas de la economía, disponibilidad y facilidad proporcionadas por la computación, incluyendo preferentemente algunas formas de técnicas de graficación computarizadas.

Hasta ahora hemos visto métodos que sirven para la detección de una sola observación discordante, pero el hecho de que en un modelo haya 2 o más observaciones discordantes hace que estas puedan afectarse unas a otras en el análisis, de tal manera que disimulen su existencia y no se puedan detectar por métodos directos de inspección de residuales para detectar una sola observación discordante a la vez; esto sugiere la posibilidad de que 2 o más observaciones discordantes puedan ser más claramente detectadas por un método en el cual se trate de detectar a todas a un mismo tiempo. Debido a lo anterior proponemos el siguiente método para detectar 'k' observaciones discordantes en conjunto, no por separado.

DESARROLLO Y ARGUMENTOS DEL METODO

Consideremos las suposiciones del modelo lineal general

$$Y = X\beta + e$$

consideremos además una versión 'alargada' del modelo lineal general

$$Y = X\beta + S + e$$

donde $S_{(n \times k)}$ es un vector, que consiste de k valores desconocidos distintos de cero ($k < n$), estos valores estarán colocados sobre los k lugares donde se sospeche la existencia de observaciones discordantes, los lugares sobrantes tendrán ceros.

De esta manera denotaremos con k al número hipotético de posibles observaciones discordantes.

Cuando estábamos seguros de la existencia de una observación discordante lo que veníamos haciendo, generalmente, era descartarla, y nos preguntábamos si el análisis sobre el conjunto de datos completo, Y , por medio del modelo, $X\beta$, podía ser mejorado descartando la observación. Con este método haremos lo mismo sólo que ahora nos preguntaremos si dicho análisis se verá mejorado cuando hagamos la omisión de un subconjunto de k observaciones. El punto de vista de este método es considerar hasta que punto el análisis del modelo con p parámetros puede ser mejorado estadísticamente, mediante la reducción de manera selecta en el tamaño (n) de los datos.

Este método involucra la designación de un valor k , que nos indique, el número máximo hipotético de observaciones discordantes. La identificación del subconjunto 'k' más probablemente

discordante es mediante la exploración de todos los posibles subconjuntos de datos que involucren la omisión de k observaciones y en el caso de que este subconjunto no sea estadísticamente verificado, la iteración se hace usando un número hipotético de $(k - 1)$ observaciones discordantes.

Para un valor hipotético k existen $\binom{n}{k}$ particiones de los datos que involucren la omisión de k valores.

Si en efecto existen k observaciones discordantes, entonces, de las $\binom{n}{k}$ particiones

$\binom{n-k}{k}$ tendrán datos 'buenos'

y

$\binom{n}{k} - \binom{n-k}{k}$ incluirán algunos datos 'malos'

Para continuar con el desarrollo del método, para cualquier partición que involucre la omisión de k observaciones específicas se tiene asociado lo siguiente :

$Y_{(n-k) \times 1}$ denota el vector de observaciones restantes.

$Y_{(k) \times 1}$ denota el vector de observaciones omitidas.

Con lo anterior a Y lo podemos escribir como :

$$Y = \begin{pmatrix} Y_{(n-k)} \\ Y_{(k)} \end{pmatrix}$$

además las correspondientes particiones de X y e se pueden ver como :

$$X_0 \text{ } (n-k) \times p \quad e_0 \text{ } (n-k) \times 1$$

$$X_1 \text{ } (n \times p) \quad e_1 \text{ } (n \times 1)$$

En base a lo anterior tenemos que :

$$Y_0 = X_0 \beta + e_0$$

Y si denotamos a $\tilde{\beta}$ como el estimador por mínimos cuadrados basado sobre el conjunto reducido de datos tendremos que :

$$\tilde{\beta} = (X_0' X_0)^{-1} X_0' Y_0$$

Este $\tilde{\beta}$ es también el estimador por mínimos cuadrados para

$$Y = X\beta + \delta + e$$

Sea δ ordenado y escrito como $\delta' = (\delta_0', \delta_1')$ correspondiente a la partición de Y .

Podemos entonces escribir :

$$\tilde{Y}_0 = X_0 \tilde{\beta}$$

$$\tilde{e}_0 = \tilde{e}_0 = Y_0 - \tilde{Y}_0$$

$$\tilde{\delta}_0 = 0 \quad \text{el vector nulo.}$$

$$\tilde{\delta}_1 = Y_1 - X_1 \tilde{\beta}$$

donde $\tilde{\delta}_1$ es el estimador por mínimos cuadrados de δ_1 en el modelo $Y = X\beta + \delta + e$

el vector $\tilde{e}_0 = (Y - X\tilde{\beta} - \tilde{\delta})$ lo llamaremos el vector de los residuales revisados y consiste de las $(n - k)$ componentes de

$$\begin{aligned} \tilde{e}_0 & \text{ aumentado con los } k \text{ ceros de } \tilde{e}_0 = Y_0 - X_0 \tilde{\beta} - \tilde{\delta}_0 \\ & = Y_0 - X_0 \tilde{\beta} - (Y_1 - X_1 \tilde{\beta}) \\ & = 0 \end{aligned}$$

Así $\tilde{\tilde{\epsilon}}^2$ es la suma de los cuadrados de los residuales revisados que resultan después de ajustar el modelo básico después de omitir las k observaciones.

La cantidad que vamos a considerar para ver cuál de los subconjuntos de k observaciones es el más probablemente discordante, en base a lo anterior será:

$$Q_k = \tilde{\tilde{\epsilon}}^2 - \tilde{\tilde{\epsilon}}^2$$

sabemos que $\tilde{\tilde{\epsilon}}^2 \sim V^2 \chi^2(n-p)$ y $\tilde{\tilde{\epsilon}}^2 \sim V^2 \chi^2(n-k-p)$

$$\Rightarrow Q_k \sim V^2 \chi^2(k)$$

bajo los supuestos usuales.

Sin embargo el presente método involucra el cálculo de Q_k para cada una de las $\binom{n}{k}$ posibles particiones de los datos y la evaluación conjunta de $\binom{n}{k}$ estadísticas. Notemos que las Q_k son estadísticamente dependientes, y una comparación simple normalizada de este conjunto de $\binom{n}{k}$ Q_k 's con alguna suma del cuadrado de los residuales, para juzgar si alguna tiene k observaciones discordantes presentes, no puede ser hecha como una simple estadística F ni usando un "máximo F " para componentes independientes. (19).

Es por esto que la evaluación estadística para ver si el más grande de los Q_k 's identifica las observaciones discordantes requiere de comparaciones internas que tomen en cuenta la dependencia estadística y el factor de que una contribución conjunta esta hecha. Este punto de comparación estará mejor determinado en la explicación del método.

Entre las Q_k 's el $\max Q_k$ tiene el menor $\tilde{\tilde{\epsilon}}^2$ es decir con el $\max Q_k$ identificamos a las k observaciones que minimizan,

al ser omitidas, la suma del cuadrado de los residuales, por esto podemos decir que el $\max Q_{k,s}$ es la estadística que podrá identificar mejor a las k posibles observaciones discordantes.

M E T O D O L O G I A

- 1) Se elige un valor k , número máximo hipotéticamente razonable de observaciones discordantes presentes en el modelo.
- 2) Para todas las posibles $\binom{n}{k}$ particiones de los datos, se calcula el correspondiente Q_k .
- 3) La partición correspondiente al más grande Q_k , lo denotaremos por Q_k^* y este va a identificar el subconjunto k más probablemente discordante.
- 4) Para saber si el tamaño de Q_k^* , estadísticamente por medio de un juicio inferencial, es suficientemente grande en relación con los demás valores de Q_k , se hacen gráficas de probabilidad en las cuales se grafican de 50 a 100 de las más grandes $Q_{k,s}$ para los datos bajo análisis junto a valores típicos (por ejemplo las medianas de números aleatorios).
- 5) Si concluimos que Q_k^* no es estadísticamente aberrante, entonces el procedimiento se repite para el número hipotético ($k - 1$).
- 6) Si Q_k^* es encontrado como aberrante entonces el método termina proponiendo las k observaciones como discordantes.

Es recomendable que el número hipotético k sea más gran de que el verdadero número esperado de observaciones discordantes; veamos porque :

Supongamos que existen K_1 observaciones discordantes pero nosotros damos un número hipotético K donde $K < K_1$, en este caso nunca se podrán detectar a todas las observaciones discordantes ya que nunca se llegará a usar el método para tratar de detectarlas. Pero si nosotros proponemos $K > K_1$, en este caso el método nos conducirá en algún momento a probar si existen K_1 observaciones discordantes.

Uno de los principales problemas que presenta este método es que; para poder llevarlo a cabo es necesario realizar $\binom{n}{k}$ regresiones. A continuación el método es desarrollado nuevamente pero en forma tal que el problema anterior queda resuelto.

El modelo de Regresión puede ser escrito como: [14]

$$E(Y) = X\beta \quad E(Y) = E\left(\begin{matrix} Y_1 \\ Y_2 \end{matrix}\right) = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta = X\beta \dots (1)$$

donde $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ es una partición del vector Y , con Y_1 un vector de $(n - k) \times 1$ correspondiente a las $(n - k)$ observaciones no sospechosas como discordantes, mientras que Y_2 es un vector de $(k \times 1)$ correspondiente a las k observaciones sospechosas como discordantes.

En general será necesario un reacomodamiento de observaciones para poder lograr esta partición de Y .

Sabemos que si hiciéramos el ajuste por mínimos cuadrados el vector de residuales E estaría dado por $Y - X\hat{\beta}$

el cual podríamos escribirlo como :

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = (I - H) Y = \begin{pmatrix} I - H_{11} & -H_{12} \\ -H_{21} & I - H_{22} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

donde $H_{ij} = X_i (X'X)^{-1} X_j'$ es una submatriz de la matriz sombrero

$H = X(X'X)^{-1}X'$ y las dimensiones de cada submatriz son :

H_{11} es de $(n-k \times n-k)$

H_{12} es de $(n-k \times k)$

H_{21} es de $(k \times n-k)$

H_{22} es de $(k \times k)$

Si quitáramos las k observaciones sospechosas nosotros obtendríamos el modelo $E(Y_1) = X_1 \beta_1$, en donde el estimador de β_1 estaría dado por

$$\hat{\beta}_1 = (X_1' X_1)^{-1} X_1' Y_1 \dots \dots \dots (2)$$

Un modelo alternativo esta dado por: (14)

$$E \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ X_2 & I \end{pmatrix} \begin{pmatrix} \beta_1 \\ \theta \end{pmatrix}$$

en donde θ es un vector de $k \times 1$.

El estimador de β_1 estaría dado en la misma forma que (2), mientras que el de θ es : (14)

$$\hat{\theta} = (I - H_{22})^{-1} \varepsilon_2$$

Como $\epsilon_2 = -H_{21}Y_1 + (I - H_{22})Y_2$

tenemos

$$\begin{aligned}\hat{\theta} &= (I - H_{22})^{-1}(-H_{21}Y_1 + (I - H_{22})Y_2) \\ &= -(I - H_{22})^{-1}H_{21}Y_1 + Y_2 \\ &= Y_2 - (I - H_{22})^{-1}H_{21}Y_1\end{aligned}$$

Si reemplazamos Y_2 por $Y_2 - \hat{\theta}$ en (1) y realizamos nuevamente el ajuste, los nuevos residuales que obtendremos serán:

$$\tilde{\epsilon} = \begin{pmatrix} \tilde{\epsilon}_1 \\ \tilde{\epsilon}_2 \end{pmatrix} = \begin{pmatrix} I - H_{11} & -H_{12} \\ -H_{21} & I - H_{22} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 - \hat{\theta} \end{pmatrix}$$

de donde

$$\begin{aligned}\tilde{\epsilon}_1 &= (I - H_{11})Y_1 - H_{12}(Y_2 - \hat{\theta}) \\ &= IY_1 - H_{11}Y_1 - H_{12}[(I - H_{22})^{-1}H_{21}Y_1] \\ &= IY_1 - H_{11}Y_1 - H_{12}(I - H_{22})^{-1}H_{21}Y_1 \\ &= [I - H_{11} - H_{12}(I - H_{22})^{-1}H_{21}]Y_1\end{aligned}$$

$$\begin{aligned}\tilde{\epsilon}_2 &= (-H_{21})Y_1 + (I - H_{22})(Y_2 - \hat{\theta}) \\ &= -H_{21}Y_1 + (I - H_{22})[(I - H_{22})^{-1}H_{21}Y_1] \\ &= -H_{21}Y_1 + H_{21}Y_1 = 0\end{aligned}$$

Estos son los llamados residuales revisados con los cuales podemos obtener ahora que :

$$Q_k = \sum_{z}^T (I - H_{zz})^{-1} \epsilon_z$$

de esta forma para obtener las combinaciones $\binom{n}{k}$ Q_k 's es tan

solo necesario tener los residuales obtenidos al realizar el ajuste con los datos completos y la matriz sombrero.

METODO DE ANDREWS Y PREGIBON

Presentaremos aquí un método para detectar k observaciones que además de ser discordantes tengan una gran influencia sobre el estimador de los parámetros.

Definamos el operador $D_{i_1 \dots i_k}^{(k)}$ el cual borra los elementos asociados con las k observaciones (i_1, i_2, \dots, i_k) para la expresión que le sigue; por ejemplo $D_{i_1}^{(1)} Y$ es el vector de $(n-1) \times 1$ cuando Y_{i_1} ha sido borrada, $(D_{i_1}^{(2)} (X'X)^{-1})$ es la inversa de la matriz $(X'X)$ una vez que se le han borrado los renglones i y j .

La influencia de una observación puede determinarse mediante el determinante de $X'X$, es decir $|X'X|$, de forma tal que al borrar una observación que tenga una gran influencia sobre el cálculo de $|X'X|$, esta observación también tendrá una gran influencia sobre los estimadores.

De igual forma la discordancia de una observación en el ajuste puede ser detectada mediante la reducción de $\mathcal{E}'\mathcal{E}$ al borrar tal observación.

Estas dos ideas pueden combinarse, para tratar de identificar observaciones discordantes con influencia en el ajuste, calculando el cambio en $(\mathcal{E}'\mathcal{E}) |X'X|$ después de quitar una o más observaciones. Es conveniente realizar este estudio en la forma

$$|X^b X^a| = (\mathcal{E}'\mathcal{E}) |X'X|$$

donde $X^* = (X' : Y)$ es la matriz de variables independientes aumentada con el vector de variables dependientes Y entonces una medida del cambio relativo debido a la omisión de las k obser

observaciones (i_1, i_2, \dots, i_k) estará dada por

$$R_{i_1, \dots, i_k}^{(k)} = \frac{|D_{i_1, \dots, i_k}^k X^{i_1} X^{i_2} \dots X^{i_k}|}{|X^{i_1} X^{i_2} \dots X^{i_k}|}$$

luego entonces, entre más pequeño sea $R_{i_1, i_2, \dots, i_k}^{(k)}$ más discordantes e influyentes serán las k observaciones omitidas. [1]

Una forma alternativa para evaluar la estadística $R_{i_1, i_2, \dots, i_k}^{(k)}$ está dada por (14) en la cual se muestra que

$$R_{i_1, \dots, i_k}^{(k)} = [1 - (Q_k / \epsilon^2 \epsilon)] |I - H_{22}| \dots (1)$$

en donde Q_k está definida como antes (pp 82) y H_{22} es una parte de la matriz sombrero H ; con esta forma de calcular a $R_{i_1, \dots, i_k}^{(k)}$ se simplifican los cálculos.

En (1) se encuentran dos factores que podemos examinar ahora, si el primero $(1 - (Q_k / \epsilon^2 \epsilon))$ es pequeño entonces Q_k será grande y por lo tanto las k observaciones omitidas serán discordantes. El segundo factor $|I - H_{22}|$ mide la influencia que tienen las k observaciones sobre el ajuste y pequeños valores de $|I - H_{22}|$ conducirán a grandes valores de h_{11} y por lo tanto revelan puntos muy influyentes. Notemos por ejemplo que para el caso $k = 1$ $|I - H_{22}|$ se convierte en $1 - h_{11}$ con lo cual si $(1 - h_{11})$, o sea $|I - H_{22}|$ es pequeño significa que h_{11} es grande, y en consecuencia la observación i omitida tiene una gran influencia.

El método puede ser resumido como sigue :

- 1) Escoger un número k que indique el número de observaciones

discordantes e influyentes que se espera encontrar. Este número k que se escoga deberá ser el más grande hipotéticamente razonable de observaciones 'malas' que esperamos encontrar, esto por las mismas razones vistas en el Método de Gentleman y Wilk.

- 2) Calcular las $\binom{n}{k} k_{i_1, \dots, i_k}^{(k)}$'s .
- 3) Encontrar el mín $k_{i_1, \dots, i_k}^{(k)}$'s, con lo cual obtendremos las k observaciones más probablemente discordantes y a la vez influyentes en el ajuste.
- 4) Si además el mín $k_{i_1, i_2, \dots, i_k}^{(k)}$ es suficientemente diferente con respecto al resto, declarar entonces a las k observaciones como discordantes.

En caso contrario hacer $k = k - 1$ y repetir el proceso.

METODO DE MICKEY, DUNN Y VIRGINIA

En esta parte proponemos un método para detectar varias observaciones discordantes a la vez. Este método se basa en que al omitir posibles observaciones discordantes en Regresión se obtiene una gran reducción en la suma del cuadrado de los residuales. Tomando en cuenta esta idea la forma de seleccionar cuáles pueden ser observaciones discordantes se puede efectuar haciendo uso de programas de regresión por Stepwise; como hemos mencionado en varios de los métodos hasta ahora vistos, la manera de checar si en el conjunto de datos existen observaciones discordantes, es ajustar el modelo con y sin las observaciones sospechosas. El mayor problema de lo anterior radica en el hecho de ¿Qué observaciones escoger como sospechosas? .

Lo que hace el método que proponemos aquí es encontrar una observación la cual al omitirla cause la más grande reducción en la suma del cuadrado de los residuales . Habiendo encontrado esta observación, este método procede a seleccionar otra observación cuya omisión vuelva a reducir grandemente la suma del cuadrado de los residuales y así sucesivamente. Como ya mencionamos lo anterior se puede hacer por un programa de regresión por Stepwise, ya que, este tipo de regresión lo que hace es seleccionar aquellas variables que expliquen "mejor" al modelo y la manera de hacerlo es introduciendo al ajuste una a una las variables y verificando que la suma del cuadrado de los residuales se haga tan pequeña

como se pueda; las variables que expliquen "mejor" al modelo serán aquellas que produzcan la mínima suma del cuadrado de los residuales.

Teniendo estas ideas en mente procedamos a dar el procedimiento y la justificación del método.

En cualquier modelo de Regresión lineal la omisión de la i -ésima observación puede ser realizada introduciendo una variable adicional que toma el valor unitario en el i -ésimo lugar y cero en los otros lugares. Por ejemplo si a cierto modelo le añadimos las columnas $(1, 0, 0, \dots, 0)$, $(0, 0, 1, \dots, 0)$ eso significará que la primera y tercer observaciones serán omitidas.

El procedimiento consiste entonces en aumentar a la matriz usual de datos X con una matriz idéntica de $(n \times n)$ y ejecutar los cálculos con un programa de Regresión por Stepwise. Es decir la matriz X es reemplazada por la matriz aumentada $X^* = (X, I)$ de $n \times (n+p)$.

El programa de Regresión por Stepwise será hecho de tal forma que las variables correspondientes a las columnas de la matriz X siempre entrarán en los cálculos de este tipo de regresión, y mediante esto solo se escogerán a aquellas variables correspondientes a las columnas de la idéntica y serán aquellas que reduzcan lo más posible la suma del cuadrado de los residuales, y en consecuencia estas últimas nos indicarán que observaciones son sospechosas. El método podría quedar resumido de la siguiente manera :

1).- Para cada paso de stepwise declarar como sospechosa a la observación correspondiente a la columna de la matriz idéntica, que entró en la regresión.

2).- Para cada paso i de stepwise calcular :

$$F_c = \frac{SCR(i-1) - SCR(i)}{SCR(i) / (n-p-i)}$$

donde $SCR(i-1)$ = Suma de cuadrados de residuales sin las observaciones indicadas por el paso $i-1$ de stepwise.

$SCR(i)$ = Suma de cuadrados de residuales sin las observaciones indicadas por el paso i de stepwise.

3).- Para cada paso i de stepwise :

Sí $F_c > F_t$ Declarar a la observación correspondiente a la variable recién incluida por stepwise como discordante.

Sí $F_c < F_t$ La observación no es discordante.

donde F_t es el cuantil $100(1 - \alpha/n)\%$ de una distribución F con 1 y $(n-p-i)$ grados de libertad.

Es importante hacer notar que la suma del cuadrado de los residuales con k observaciones omitidas es igual a la suma del cuadrado de los residuales con k columnas de la matriz idéntica aumentadas a la matriz X .

Y además el mínimo de la suma del cuadrado de los residuales obtenido por el procedimiento de stepwise, al final de cada paso, es el mismo que el obtenido por omisión de las observaciones correspondientes en el análisis de regresión. (27)

CAPITULO III

Por lo tanto los cálculos de regresión con variables unidas es equivalente a los cálculos de regresión con observaciones omitidas.

OBSERVACIONES INFLUYENTES

Dado el hecho, de que una observación sea discordante no significa que necesariamente sea influyente en el ajuste, y debido a que muchos métodos para detectar observaciones discordantes toman muy en cuenta que también sean influyentes, decíamos hablar ahora un poco sobre observaciones influyentes, presentar algunos métodos para detectarlas y analizar un ejemplo en donde se muestre claramente que una observación discordante no siempre es influyente y viceversa.

Una observación influyente puede ser definida, como, aquella observación, tal que al quitarla, se produzca un cambio "significativo" en el ajuste.

Algunos métodos para detectar observaciones influyentes son los siguientes :

- a) Encontrar los elementos de la diagonal de la matriz H es decir encontrar h_{ii} , $i = 1, 2, \dots, n$ y compararlo contra $2p/n$, si $h_{ii} \gg 2p/n$, h_{ii} se considera "grande" y por lo tanto y_i es influyente en el ajuste. [22]
- b) Otro método para determinar la influencia de la i -ésima observación es calcular $\hat{\beta} - \hat{\beta}_{(i)}$ lo cual es equivalente a calcular $X_{(i)}\hat{\beta} - X_{(i)}\hat{\beta}_{(i)}$ si la diferencia es substancial, significa que, al excluir la i -ésima observación existe un cambio significativo en la determinación de los

estimadores y en consecuencia existirá un gran cambio en el ajuste, en tal caso diremos que la i -ésima observación es influyente.

- c) Una forma posible de determinar si $x_{ci} \hat{\beta} - x_{ci} \hat{\beta}_{(i)}$ es 'grande', es comparar tal diferencia contra el error estándar de y_i , $S\sqrt{h_{ii}}$ por lo cual si:

$$\frac{x_{ci} (\hat{\beta} - \hat{\beta}_{(i)})}{S\sqrt{h_{ii}}} > 1$$

entonces diremos que la diferencia en el ajuste es grande y por lo tanto la i -ésima observación es influyente.

- d) Con el siguiente método podremos encontrar la influencia que tienen determinadas observaciones en forma conjunta con respecto al ajuste; el método consiste en calcular

$|I - H_{22}|$, donde H_{22} es como se definió en (2.3),

$|I - H_{22}|$ nos proporciona una medida de que tan lejos se encuentran las observaciones que determinan H_{22} , del

centro de los datos restantes; entre más pequeño sea

$|I - H_{22}|$ más lejos se encontraran las observaciones y

por lo tanto más influencia tendrán en el ajuste.

Notemos que para el caso $k = 1$, $|I - H_{22}| = 1 - h_{11}$

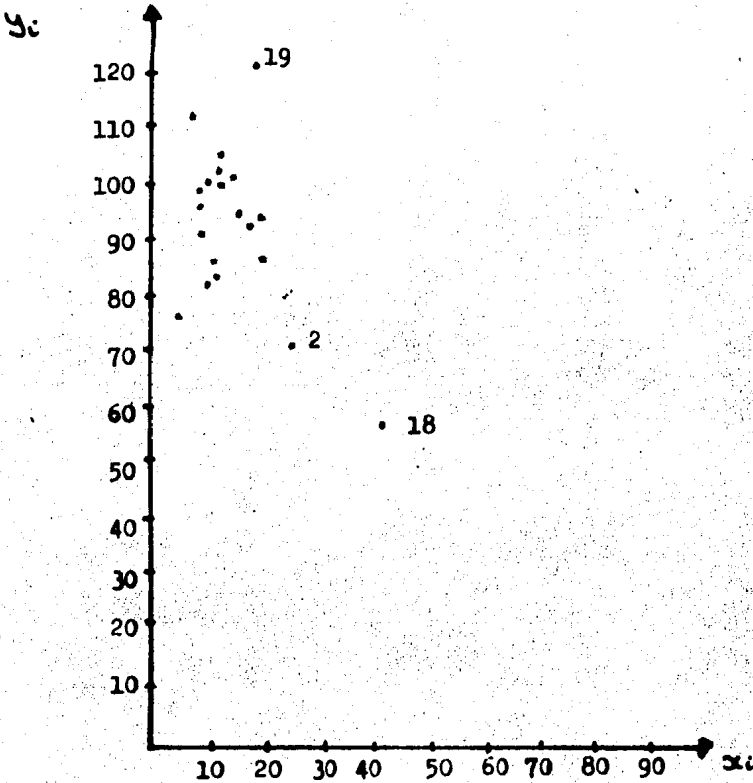
con lo cual tenemos prácticamente el caso a)

Presentaremos ahora un ejemplo de Regresión lineal simple en el cual podamos observar que las observaciones discordantes no necesariamente tienen que ser influyentes, para poder lograr tal cosa, usaremos además de los métodos propuestos aquí para detectar observaciones influyentes, las estadísticas Q_k y $R^{(k)}$ para detectar observaciones discordantes.

Analizaremos un ejemplo dado por Mickey, Dunn y Clarck, el cual consiste de 21 observaciones, las cuales aparecen a continuación :

Obs. #	x1	y	Obs. #	x1	y
1	15	95	11	7	113
2	26	71	12	9	96
3	10	83	13	10	83
4	9	91	14	11	84
5	15	102	15	11	102
6	20	87	16	10	100
7	18	93	17	12	105
8	11	100	18	42	57
9	8	104	19	17	121
10	20	94	20	11	85
			21	10	100

La gráfica que se obtiene con estos datos es :

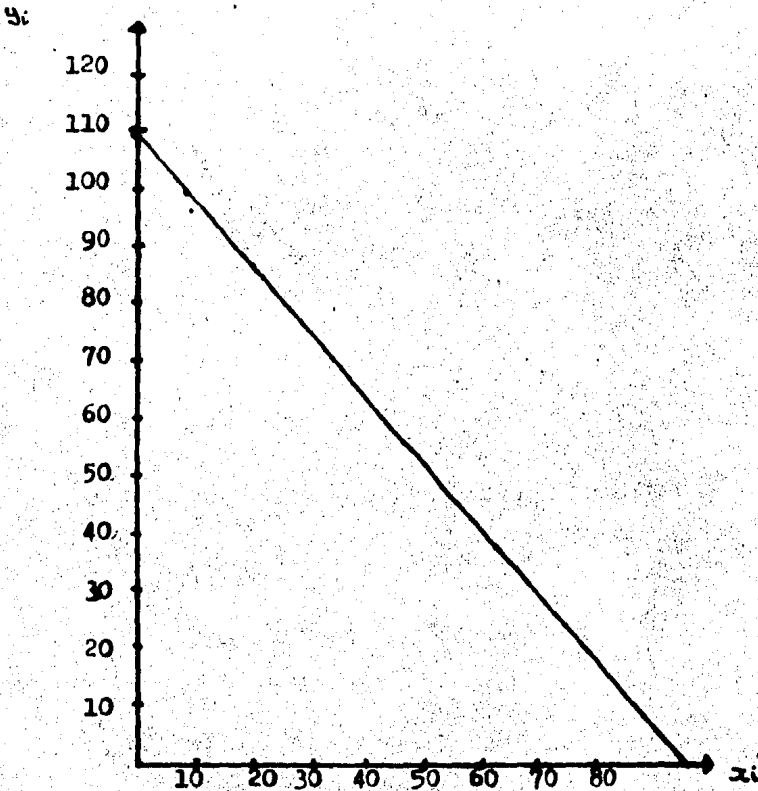


a simple vista las observaciones 19, 18 y 2 parecen ser "sospe -
chosas " .

Al hacer el ajuste para el conjunto completo de observaciones obtenemos que la línea ajustada es:

$$\hat{y} = 109.87 - 1.127 x_1$$

la cual aparece en la siguiente gráfica con los datos.



Para poder evaluar las distintas estadísticas es necesario contar con los elementos de la diagonal de la matriz H y los residuales obtenidos al realizar el ajuste con el conjunto completo de observaciones los cuales se muestran a continuación: Los elementos de la diagonal de la matriz H se encuentran en la tabla (99^{na}).

Valores de los residuales :

$\epsilon_1 = 2.03$	$\epsilon_{11} = -3.73$
$\epsilon_2 = -9.57$	$\epsilon_{12} = -15.60$
$\epsilon_3 = -15.60$	$\epsilon_{13} = -13.47$
$\epsilon_4 = -8.73$	$\epsilon_{14} = 4.52$
$\epsilon_5 = 9.03$	$\epsilon_{15} = 1.39$
$\epsilon_6 = -.33$	$\epsilon_{16} = 8.65$
$\epsilon_7 = 3.41$	$\epsilon_{17} = -5.54$
$\epsilon_8 = 2.52$	$\epsilon_{18} = 30.28$
$\epsilon_9 = 3.14$	$\epsilon_{19} = -11.47$
$\epsilon_{10} = 6.66$	$\epsilon_{20} = 1.39$
$\epsilon_{21} = 11.01$	

En la tabla siguiente se muestran los valores que tomaron las estadísticas para detectar observaciones discordantes y/o influyentes después de haber omitido cada una de las observaciones :

Observación omitida	Q_1	$R_i^{(1)}$	$x_{(i)}(\hat{\beta} - \hat{\beta}_{(i)})$	$x_{(i)}(\hat{\beta} - \hat{\beta}_{(i)})/s\sqrt{h_{ii}}$
1	4.33	.9512	.1001	.04073
2	108.37	.8072	-1.728	-.3897
3	259.80	.8324	-1.0311	-.3652
4	82.01	.8969	-.6570	-.219
5	85.66	.9176	.4453	.1813
6	.12	.9279	-.0256	-.00842
7	12.58	.9378	.2061	.0762
8	6.75	.9412	.1494	.0055
9	10.73	.9167	.2693	.08460
10	47.91	.9087	.5167	.1700
11	133.44	.8521	1.088	.3202
12	14.98	.9239	-.2807	-.0930
13	259.8	.8324	-1.031	-.3656
14	192.54	.8652	-.7990	-.2981
15	21.69	.9351	.2681	-.1000
16	2.08	.9371	.0918	.03255
17	78.94	.9155	.4744	.1837
18	88.1	.3385	10.198	1.1186
19	968.56	.5502	1.660	.642
20	139.63	.8869	-.6804	-.253
21	2.08	.9371	.0918	.03255

Los valores de las estadísticas ordenados en cuanto a influencia o discordancia son los siguientes :

h_{ii}	Obs. omitida	Q_1	Obs. omitida	R_1^1
$h_{18,18} = .648$	19	968.56	18	.3385
$h_{2,2} = .153$	3	259.8	19	.5502
$h_{11,11} = .090$	13	259.8	2	.8072
$h_{9,9} = .079$	14	192.54	3	.8324
$h_{10,10} = .072$	20	139.63	13	.8324
$h_{6,6} = .072$	11	133.44	11	.8521
$h_{12,12} = .070$	2	108.37	14	.8652
$h_{4,4} = .070$	18	88.1	20	.8869
$h_{3,3} = .062$	5	85.66	4	.8969
$h_{13,13} = .062$	4	82.01	10	.9087
$h_{16,16} = .062$	17	78.94	17	.9155
$h_{21,21} = .062$	10	47.91	9	.9167
$h_{7,7} = .057$	15	21.69	5	.9176
$h_{8,8} = .056$	12	14.98	12	.9239
$h_{14,14} = .056$	7	12.58	6	.9279
$h_{15,15} = .056$	9	10.73	15	.9351

h_{ii}	Obs. omitida	Q_i	Obs. omitida	$R_i^{(1)}$
$h_{20,20} = .056$	8	6.75	16	.9371
$h_{17,17} = .052$	1	4.33	21	.9371
$h_{19,19} = .052$	16	2.08	7	.9378
$h_{1,1} = .047$	21	2.08	8	.9412
$h_{5,5} = .047$	6	.12	1	.9512

Obs. omitida	$\chi_{(i)} (\hat{\phi} - \hat{\phi}_{(i)})$
18	10.198
2	- 1.728
19	1.66
11	1.088
13	- 1.031
3	- 1.031
14	- .799
20	- .6804
4	- .6570
10	.5107
17	.4744
5	.4453
12	- .2807
9	.2693
15	.2681
7	.2061
8	.1494
1	.1001
16	.0918
21	.0918
6	- .0256

Obs. omitida	$\chi_{(i)} (\hat{\phi} - \hat{\phi}_{(i)}) / \sqrt{s^2(1-h_i)}$
8	.0055
6	-.00842
16	.03255
21	.03255
1	.04073
7	.0762
9	.08460
12	-.0936
15	-.1000
10	.1700
5	.1813
17	.1837
4	-.219
20	-.253
14	-.2981
11	.3202
3	-.3652
13	-.3656
2	-.3897
19	.642
18	1.1186

Observemos que si $2p/n = 4/21 = .1904$ entonces bajo el criterio a) solo podríamos detectar como observaciones influyentes a las observaciones : 18 dado que $.1904 < .648 = h_{18,18}$ bajo este criterio las observaciones 2 y 11 son las que siguen en importancia de influencia.

Si observamos la estadística $x_{(i)}(\hat{\beta} - \hat{\beta}_{(i)})$ nos daremos cuenta que la observación 18 es la que tiene una gran influencia en cuanto a determinar los parámetros en el ajuste, las observaciones 2 y 19 le siguen en este sentido, aunque la 18 es mucho más influyente que las otras dos.

Con la estadística Q_1 podemos decir que la observación 19 es la más discordante en seguida están las observaciones 3 y 13 aunque no son tan discordantes como la 19.

Usando la estadística $R_i^{(1)}$ vemos que la observación 18 es la más discordante e influyente a la vez, las observaciones que le siguen son la 19 y la 2 .

CONCLUSION:

Podemos decir que la observación más discordante es la 19 mientras que la más influyente es la 18, con lo cual queda perfectamente ejemplificado que no siempre una observación discordante es influyente y viceversa.

CAPITULO IV

T R A T A M I E N T O

Hasta este momento nos hemos concretado a presentar varios métodos para la detección de observaciones discordantes; pero no hemos dicho que hacer con ellas una vez que han sido detectadas.

El propósito de esta parte de la tesis es precisamente indicar lo que se puede y debe hacer cuando una observación ha sido declarada como discordante.

Analizaremos las ventajas y desventajas de cada una de las maneras de tratar a este tipo de observaciones, para poder así concluir cual es el mejor camino de tratarlas.

Las formas de tratamiento aplicables a las observaciones discordantes se pueden agrupar en dos que son :

- 1) RECHAZO
- 2) SUSTITUCION

La primera forma consiste en que una vez que han sido detectadas las observaciones discordantes, se eliminan del ajuste tanto la observación (y_i) como el i -ésimo renglón de la matriz X . Una vez eliminadas del ajuste se reportan y estudian por separado.

Las desventajas de rechazarlas es que producen varias alteraciones en el análisis como son :

- 1.) Al eliminar una observación se corre el riesgo (dependiendo del número de observaciones restantes) de que $X_{(i)}$ sea

tal que su rango sea menor que 'p' y en consecuencia $X_{(i)}' X_{(i)}$ sea singular, lo cual implicaría no poder ajustar el modelo.

ii) En el caso de que $X_{(i)}' X_{(i)}$ sea invertible es importante analizar qué pasa con la $\text{Var}(\hat{\beta}_{(i)})$, al eliminar una observación.

La $\text{Var}(\hat{\beta}_{(i)})$ está expresada como :

$$\text{Var}(\hat{\beta}_{(i)}) = \sigma^2 (X_{(i)}' X_{(i)})^{-1}$$

y viéndola en términos de la matriz completa se tiene que:

$$\begin{aligned} \text{Var}(\hat{\beta}_{(i)}) &= \sigma^2 (X'X)^{-1} [I + x_i' x_i (X'X)^{-1} (1-h_{ii})] \\ &= \text{Var}(\hat{\beta}) [I + x_i' x_i (X'X)^{-1} (1-h_{ii})] \end{aligned}$$

Analizando esta última expresión nos podemos dar cuenta que al eliminar una observación se aumenta la varianza del estimador de parámetros $\hat{\beta}$, pero notemos además que no todas las observaciones tienen el mismo efecto sobre esta varianza en el caso de ser eliminadas, ya que, entre más influyente sea la observación eliminada, y en consecuencia mayor sea h_{ii} , mayor será el aumento en la varianza de $\hat{\beta}$ cuando se elimine la i-ésima observación correspondiente.

Ahora bien si consideramos la varianza estimada del vector de parámetros estimados $\hat{\beta}$;

$$\widehat{\text{Var}}(\hat{\beta}_{(i)}) = \hat{\sigma}^2 (X'X)^{-1} \left[I + \chi_i \chi_i' (X'X)^{-1} (1-h_{ii})^{-1} \right]$$

nos daremos cuenta de que las ventajas que se conseguirían en cuanto a la disminución de $\hat{\sigma}^2$ por haber omitido las observaciones discordantes, que implicarían esperar una disminución en $\widehat{\text{Var}}(\hat{\beta}_{(i)})$, se pueden compensar y hasta perder por el aumento debido a $(1-h_{ii})^{-1}$.

iii) Veamos ahora los efectos que provoca el omitir una observación en los residuales studentizados y en la varianza de los residuales.

$$\text{Sea } h_{k1(i)} = \chi_k' (X_{(i)}' X_{(i)})^{-1} \chi_k$$

(entrada $(k,1)$ de la matriz H al haber omitido la i -ésima observación, y $t_{j(i)}^t$ el j -ésimo residual studentizado basado sobre el conjunto de datos omitiendo la i -ésima observación .

Para analizar este punto es necesario expresar $t_{j(i)}^t$ y $h_{k1(i)}$ en términos del conjunto completo de datos, para lo cual son de gran ayuda las siguientes relaciones : (Cook, 1979)(9)

$$1) h_{k1} = h_{k1(i)} - h_{k1(i)} h_{11(i)} / (1 + h_{11(i)})$$

$$2) h_{k1} = h_{k1} + h_{ki} h_{11} / (1 - h_{11})$$

$$3) \hat{\beta} - \hat{\beta}_{(i)} = (X'X)^{-1} \chi_i \epsilon_i / (1 - h_{11})$$

$$4) (n-p) S^2 = (n-p-1) S_{(i)}^2 + \epsilon_i^2 / (1 - h_{11})$$

Si denotamos por r_{ij} la correlación entre el i -ésimo y el j -ésimo residual en el conjunto completo de datos entonces :

$$r_{ij} = \frac{\text{Cov}(\epsilon_i, \epsilon_j)}{\{\text{Var}(\epsilon_i) \text{Var}(\epsilon_j)\}^{1/2}}$$

donde $\text{Cov}(\epsilon_i, \epsilon_j)$ está dada por el elemento (i, j) de la matriz $(I-H)\sigma^2$.

Se puede obtener, partiendo de 2 que :

$$\frac{\text{Var}(\hat{y}_{j(i)})}{\text{Var}(\epsilon_{j(i)})} = \frac{h_{jj(i)}}{(1-h_{jj(i)})} = \frac{h_{jj}(1-r_{ij}^2) + r_{ij}^2}{(1-h_{jj})(1-r_{ij}^2)}$$

Analizando esta última igualdad encontramos una expresión para $h_{jj(i)}$ y podemos darnos cuenta que esta razón es grande si h_{jj} ó r_{ij} tienen un valor grande. Para $j \neq i$

Pero démonos cuenta que tener una expresión para $h_{jj(i)}$ implica tener una expresión para $\text{Var}(\hat{y}_{j(i)})$ debido a que :

$$\text{Var}(\hat{y}_{j(i)}) = h_{jj(i)} \sigma^2$$

y esto nos indica que cuando la varianza del j -ésimo valor ajustado se incrementa al omitir una observación (la i -ésima), este incremento puede deberse a una gran correlación entre el i -ésimo y el j -ésimo residual o a un gran valor de h_{jj} . Además cuando las correlaciones entre residuales son pequeñas, las varianzas de los valores ajustados no sufrirán un cambio notorio al eliminar cualquier punto.

Consecuencias de omitir una observación en los residuales studentizados.

Las expresiones (2) - (4) nos conducen a la expresión del j-ésimo residual studentizado cuando se omite la i-ésima observación y ésta es : (Cook,1979)(9)

$$t_{j(i)}^2 = (n-p-1)(t_j - r_{ij}t_i)^2 / (n-p-t_i)^2(1 - r_{ij}^2)$$

El análisis de esta expresión nos conduce a observar que si al omitir una observación se produce un incremento en $t_{j(i)}^2$ esto se deberá a un valor grande de r_{ij}^2 es decir a un valor grande de la correlación entre t_i y t_j , en conclusión esta correlación provocará un aumento en el j-ésimo residual studentizado al omitir la i-ésima observación.

Estos problemas que se generan al omitir una observación nos conducen a rechazar esta forma de tratamiento, es decir, no se debe rechazar a una observación declarada como discordante sino sustituirla.

La segunda forma de tratar a las observaciones discordantes es sustituirlas ya sea por :

- i) Nuevas observaciones.
- ii) Por nuevos valores en base a las observaciones disponibles.

La primera forma es la más aconsejable, ya que, de esta manera no se corre el riesgo de alejarse de la realidad, pero si por alguna razón no es posible obtener nuevas observaciones o no se quiere correr el riesgo de obtener otro valor discordante (cosa que puede suceder por la naturaleza de estas observaciones)

entonces es preferible optar por la segunda manera de sustitución.

Un método para poder sustituir observaciones, tomando como base a las observaciones disponibles, es el de Winsorización, sobre el cual hablaremos un poco ahora dando una visión general para después aplicar un tipo de winsorización, desarrollado primero para el caso de una sola observación discordante y posteriormente en el caso general cuando existen k observaciones discordantes.

La Regresión winsorizada es una alternativa del ajuste por mínimos cuadrados para la estimación de un modelo de regresión lineal mediante la alteración de los valores de algunas observaciones, basadas sobre la magnitud de los residuales.

La Winsorización es una técnica estadística que pretende reducir el efecto de contaminación producida por la presencia de una observación discordante, en una cierta población muestral, sobre la estimación de los parámetros, mediante la disminución del efecto de tal observación.

En seguida se explica el procedimiento general de winsorización aplicado al modelo de regresión lineal simple.

Consideremos los residuales en forma ordenada es decir

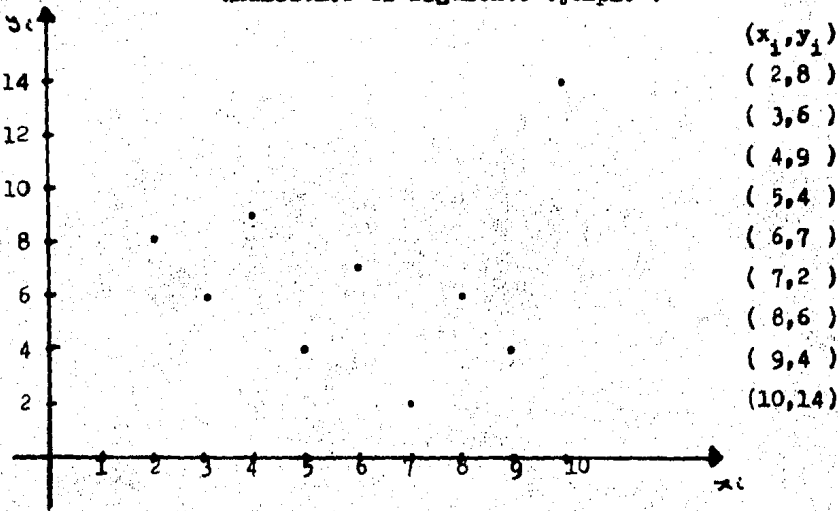
$$\varepsilon_1 \leq \varepsilon_2 \leq \varepsilon_3 \leq \dots \leq \varepsilon_n$$

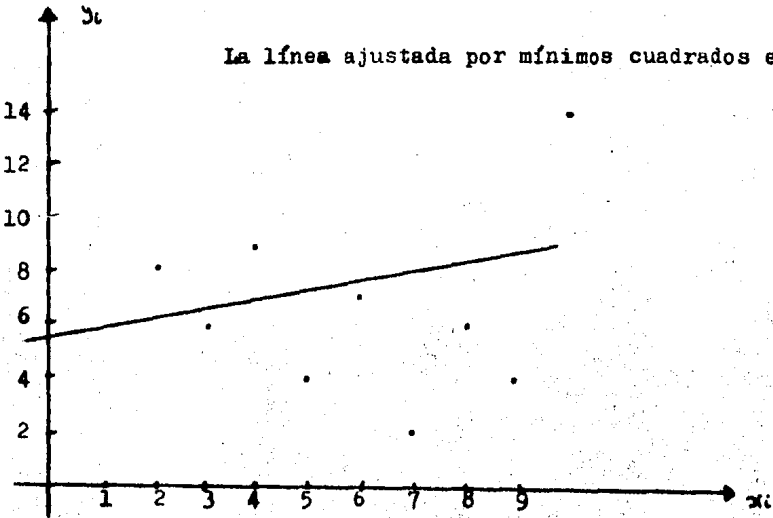
la línea de Regresión winsorizada se calcula por el método de mínimos cuadrados sobre la muestra transformada (x_i, y_i) $i = 1, \dots, n$ en donde ;

$$y_i = \hat{y}_i + \epsilon_i \quad \text{y} \quad \epsilon_i = \begin{cases} \epsilon_{g+1} & i=1, \dots, g \\ \epsilon_i & i=g+1, \dots, n-g \\ \epsilon_{n-g} & i=n-g+1, \dots, n \end{cases}$$

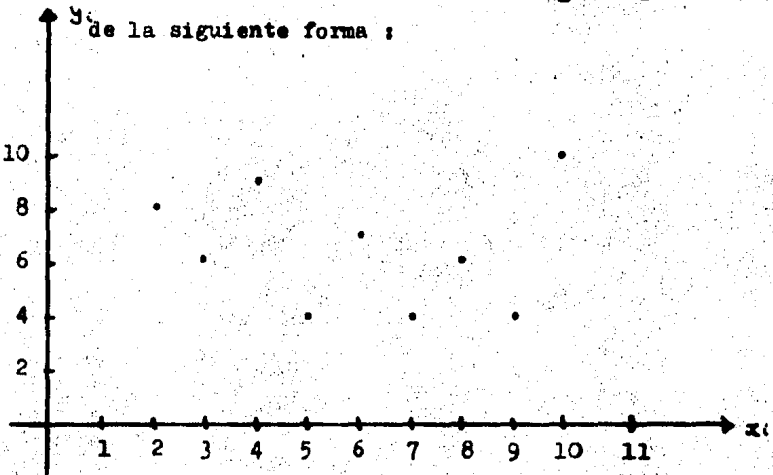
En donde n es el número de observaciones y g es el número de puntos winsorizados a cada extremo y donde \hat{y}_i es el valor ajustado por la regresión para la i -ésima observación

Analizemos el siguiente ejemplo :





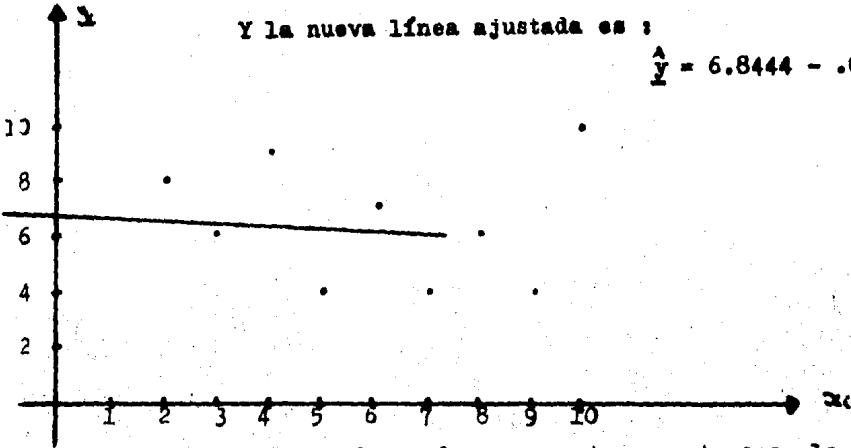
Usando winsorización con $g = 1$ las observaciones quedan



- (x_1, y_1)
- $(2, 8)$
- $(3, 6)$
- $(4, 9)$
- $(5, 4)$
- $(6, 7)$
- $(7, 4)$
- $(8, 6)$
- $(9, 4)$
- $(10, 10)$

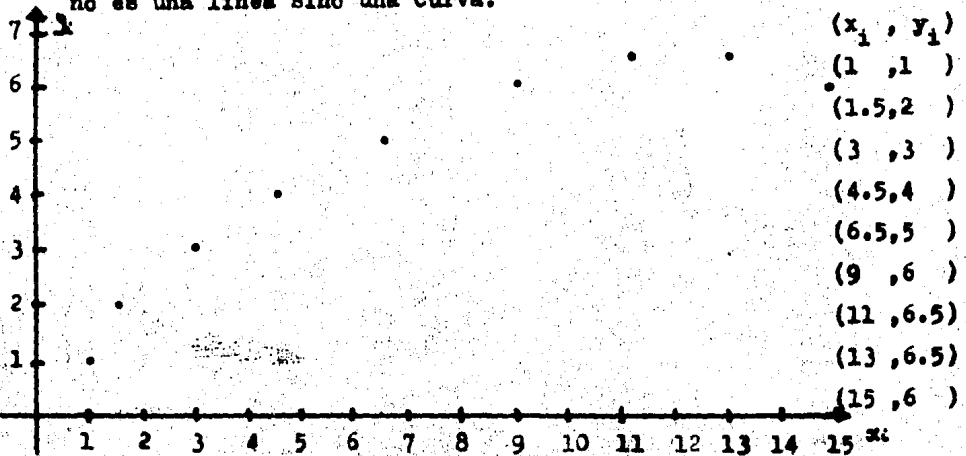
Y la nueva línea ajustada es :

$$\hat{y} = 6.8444 - .0666x$$



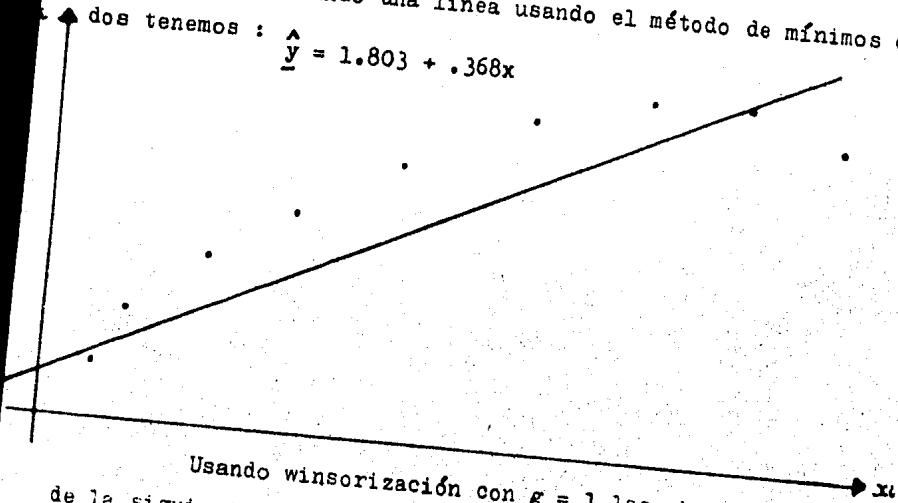
Como podemos darnos cuenta en este caso la winsorización cumple con su objetivo, de disminuir los efectos que puede tener una observación discordante en la determinación de los estimadores de los parámetros.

Vemos ahora otro ejemplo en donde el ajuste correcto no es una línea sino una curva.



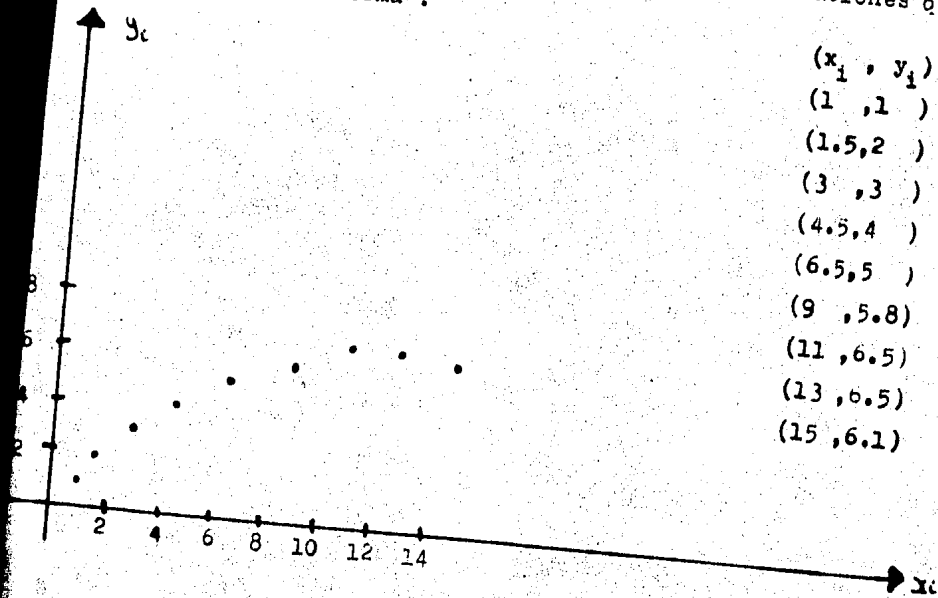
Ajustando una línea usando el método de mínimos cuadrados tenemos :

$$\hat{y} = 1.803 + .368x$$



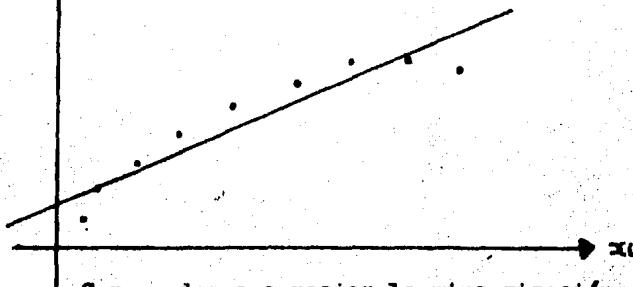
Usando winsorización con $g = 1$ las observaciones quedan de la siguiente forma :

- (x_1, y_1)
- (1, 1)
- (1.5, 2)
- (3, 3)
- (4.5, 4)
- (6.5, 5)
- (9, 5.8)
- (11, 6.5)
- (13, 6.5)
- (15, 6.1)



Ajustando la línea con los nuevos valores se obtiene :

$$\hat{y} = 1.1778 + .370 x$$



Como podemos apreciar la winsorización no tiene ningún efecto en este tipo de casos, es decir es inútil usarla en este ejemplo.

La idea fundamental del tipo de winsorización que veremos ahora es substituir el valor discordante (supongamos sin pérdida de generalidad que es y_1) por el valor $y_1^{(1)} = y_1 - N$ en donde el valor de N será estimado de tal forma que la suma de residuales al cuadrado sea mínima.

Se expresa a $y_1^{(1)}$ de esa forma ($y_1 - N$) debido a que lo que se desea es disminuir el valor discordante de y_1 en relación con los demás valores presentes en el análisis.

Si denotamos por $\epsilon_i^{(1)}$ ($i=1, 2, \dots, n$) a los residuales que se obtienen cuando y_1 ha sido substituido, entonces lo que se desea encontrar es la N tal que :

$$\sum_{i=1}^n (\epsilon_i^{(1)})^2 \text{ sea mínima.}$$

Sea $Y_{0(n \times 1)}$ el vector en el cual se substituye y_1 por $y_1 - N$, entonces $Y_0' = (y_1 - N, y_2, \dots, y_n)$ y el nuevo modelo a ajustar sería :

$$Y_0 = X \beta + M$$

donde M sería el nuevo vector de errores.

El estimador por mínimos cuadrados para β en este modelo estará dado por :

$$\hat{\beta} = (X'X)^{-1} X'Y_0$$

$$y \therefore \hat{Y}_0 = X(X'X)^{-1} X'Y_0 = HY_0$$

y el nuevo vector de residuales resultará ser :

$$\epsilon_{(n \times 1)}^{(1)} = Y_0 - \hat{Y}_0 = (I - H)Y_0$$

la i -ésima componente de este vector está dada por :

$$\epsilon_i^{(1)} = \sum_{j=1}^n a_{ij} y_j^{(1)} = \sum_{j=2}^n a_{ij} y_j^{(1)} + a_{i1} y_1^{(1)}$$

$$\text{Sea } L = \sum_{j=2}^n a_{ij} y_j^{(1)} + a_{i1} y_1^{(1)}$$

donde a_{ij} es la componente (i, j) de la matriz $(I - H)$ y $y_j^{(1)}$ es la j -ésima componente de Y_0 .

Pero como sabemos que :

$$y_1^{(1)} = y_1 - N \text{ y que } y_j^{(1)} = y_j \quad \forall j = 2, \dots, n \text{ tenemos}$$

$$\begin{aligned} \text{que } L &= \sum_{j=2}^n a_{ij} y_j + a_{i1} (y_1 - N) \\ &= \sum_{j=2}^n a_{ij} y_j - N a_{i1} \end{aligned}$$

Pero como el vector de residuales ϵ en el modelo original se expresa como :

$$\epsilon = Y - \hat{Y} = (I - H)Y$$

entonces, la i -ésima componente de ϵ , ϵ_i , es igual a $\sum_{j=1}^n a_{ij} y_j$

de donde :

$$L = \epsilon_i - N a_{i1}$$

$$y \quad \therefore \epsilon_i^{(N)} = \epsilon_i - N a_{i1}$$

dado que lo que se quiere encontrar es la N tal que minimice a

$\sum_{i=1}^n (\epsilon_i^{(N)})^2$ podemos entonces utilizar el criterio de la segunda derivada, nosotros tenemos que :

$$\begin{aligned} \sum_{i=1}^n (\epsilon_i^{(N)})^2 &= \sum_{i=1}^n (\epsilon_i - a_{i1} N)^2 \\ &= \sum_{i=1}^n (\epsilon_i^2 - 2 \epsilon_i a_{i1} N + (a_{i1} N)^2) \\ &= \sum_{i=1}^n \epsilon_i^2 - 2N \sum_{i=1}^n \epsilon_i a_{i1} + N^2 \sum_{i=1}^n a_{i1}^2 \end{aligned}$$

Derivando con respecto a N tenemos :

$$D_N \left(\sum_{i=1}^n (\epsilon_i^{(N)})^2 \right) = -2 \sum_{i=1}^n \epsilon_i a_{i1} + 2N \sum_{i=1}^n a_{i1}^2$$

igualando a cero tenemos que :

$$\begin{aligned} -2 \sum_{i=1}^n \epsilon_i a_{i1} + 2\hat{N} \sum_{i=1}^n a_{i1}^2 &= 0 \\ \Rightarrow \quad 2\hat{N} \sum_{i=1}^n a_{i1}^2 &= 2 \sum_{i=1}^n \epsilon_i a_{i1} \end{aligned}$$

$$\hat{N} = \frac{\sum_{i=1}^n \xi_i a_{i1}}{\sum_{i=1}^n a_{i1}^2}$$

si obtenemos la segunda derivada :

$$D_N^2 \left(\sum_{i=1}^n (\xi_i''')^2 \right) = 2 \sum_{i=1}^n a_{i1}^2 > 0 \therefore \text{encontramos}$$

un mínimo con $\hat{N} = \frac{\sum_{i=1}^n \xi_i a_{i1}}{\sum_{i=1}^n a_{i1}^2}$

Pero debido a que (I-H) es simétrica e idempotente teng

mos que :

$$a_{11} = \sum_{i=1}^n a_{i1} a_{1i} = \sum_{i=1}^n (a_{i1})^2$$

$$\begin{aligned} y \sum_{i=1}^n \xi_i a_{i1} &= \sum_{i=1}^n \left(\sum_{j=1}^n a_{ij} y_j \right) a_{i1} \\ &= \sum_{i=1}^n \sum_{j=1}^n (a_{ij} y_j a_{i1}) \\ &= \sum_{j=1}^n \left(\sum_{i=1}^n a_{ji} a_{i1} \right) y_j \\ &= \sum_{j=1}^n a_{j1} y_j = \sum_{j=1}^n a_{1j} y_j = \xi_1 \end{aligned}$$

$$\therefore \hat{N} = \xi_1 / a_{11}$$

Y como $\xi_i''' = \xi_i - a_{i1} \hat{N}$

$$\begin{aligned} \Rightarrow \xi_1''' &= \xi_1 - a_{11} \left(\xi_1 / a_{11} \right) \\ &= \xi_1 - \xi_1 = 0 \end{aligned}$$

y $\xi_i''' = \xi_i - a_{i1} / a_{11} \xi_1 \quad * \quad i \neq 1$

SUBSTITUCION DE K OBSERVACIONES DISCORDANTES

Veremos ahora la forma de substituir k observaciones discordantes por k valores obtenidos a partir de las observaciones disponibles.

Supongamos sin pérdida de generalidad que las k observaciones discordantes son : y_1, y_2, \dots, y_k ; entonces si :

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ \cdot \\ y_k \\ y_{k+1} \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = Y \quad \text{Se desea encontrar un nuevo vector :} \quad \begin{bmatrix} y_1^{(k)} \\ y_2^{(k)} \\ y_3^{(k)} \\ \cdot \\ \cdot \\ \cdot \\ y_k^{(k)} \\ y_{k+1}^{(k)} \\ \cdot \\ \cdot \\ y_n^{(k)} \end{bmatrix} = y^{(k)}$$

en donde $y_i^{(k)} = y_i - M_i$ para $i = 1, 2, \dots, k$ y $y_i^{(k)} = y_i$ para $i = k+1, \dots, n$.

Una forma alternativa de escribir lo anterior es la siguiente :

$$\text{Sea } Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \text{ en donde } Y_1 = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} \text{ y } Y_2 = \begin{bmatrix} y_{k+1} \\ y_{k+2} \\ \vdots \\ y_n \end{bmatrix}$$

y lo que queremos encontrar es un nuevo vector

$$Y^{(k)} = \begin{bmatrix} Y_1^{(k)} \\ Y_2^{(k)} \end{bmatrix}$$

en donde :

$$Y_1^{(k)} = Y_1 - N \quad \text{y} \quad Y_2^{(k)} = Y_2$$

en forma tal que, si el vector de residuales obtenidos con el nuevo conjunto de observaciones $Y^{(k)}$ es $\epsilon^{(k)}$, se debe cumplir que :

$$\sum_{i=1}^n (\epsilon_i^{(k)})^2 = \epsilon^{(k)'} \epsilon^{(k)} \quad \text{sea mínima}$$

sabemos que $\epsilon^{(k)} = (I-H) Y^{(k)}$

entonces $\epsilon^{(k)'} = Y^{(k)'} (I-H)$

$$\begin{aligned} \text{luego } (\epsilon^{(k)'} \epsilon^{(k)}) &= Y^{(k)'} (I-H)(I-H) Y^{(k)} \\ &= Y^{(k)'} (I-H) Y^{(k)} \end{aligned}$$

usando matrices particionadas esto lo podemos escribir en la forma siguiente :

$$\begin{aligned}
 & \begin{bmatrix} Y_1^{(k)} \\ Y_2^{(k)} \end{bmatrix}' \begin{bmatrix} I - H_{11} & -H_{12} \\ -H_{21} & I - H_{22} \end{bmatrix} \begin{bmatrix} Y_1^{(k)} \\ Y_2^{(k)} \end{bmatrix} \\
 &= (Y_1^{(k)'}, Y_2^{(k)'}) \begin{bmatrix} I - H_{11} & -H_{12} \\ -H_{21} & I - H_{22} \end{bmatrix} \begin{bmatrix} Y_1^{(k)} \\ Y_2^{(k)} \end{bmatrix} \\
 &= \left[Y_1^{(k)' (I - H_{11})} - Y_2^{(k)' H_{21}}, -Y_1^{(k)' H_{12}} + Y_2^{(k)' (I - H_{22})} \right] \begin{bmatrix} Y_1^{(k)} \\ Y_2^{(k)} \end{bmatrix} \\
 &= Y_1^{(k)' (I - H_{11})} Y_1^{(k)} - Y_2^{(k)' H_{21}} Y_1^{(k)} - Y_1^{(k)' H_{12}} Y_2^{(k)} + \\
 & \quad Y_2^{(k)' (I - H_{22})} Y_2^{(k)} \\
 &= (Y_1 - N)' (I - H_{11}) (Y_1 - N) - Y_2' H_{21} (Y_1 - N) - (Y_1 - N)' \\
 & \quad H_{12} Y_2 + Y_2' (I - H_{22}) Y_2 \\
 &= Y_1' (I - H_{11}) (Y_1 - N) - N' (I - H_{11}) (Y_1 - N) - Y_2' H_{21} Y_1 + \\
 & \quad Y_2' H_{21} N - Y_1' H_{12} Y_2 + N' H_{12} Y_2 + Y_2' (I - H_{22}) Y_2 \\
 &= Y_1' (I - H_{11}) Y_1 - Y_1' (I - H_{11}) N - N' (I - H_{11}) Y_1 \\
 & \quad + N' (I - H_{11}) N - Y_2' (H_{21}) Y_1 + Y_2' H_{21} N - Y_1' H_{12} Y_2 \\
 & \quad + N' H_{12} Y_2 + Y_2' (I - H_{22}) Y_2
 \end{aligned}$$

Si ahora usamos el criterio de la segunda derivada para obtener el mínimo tenemos que :

$$D_{\underline{N}} (\xi^{(u)}, \xi^{(u)}) = - (I - H_{11}) Y_1 - (I - H_{11}) Y_1 + 2(I - H_{11}) \hat{N} + H_{21} Y_2 + H_{12} Y_2 = 0$$

$$\Rightarrow -2(I - H_{11}) Y_1 + 2(I - H_{11}) \hat{N} + 2 H_{12} Y_2 = 0$$

$$(I - H_{11}) \hat{N} = (I - H_{11}) Y_1 - H_{12} Y_2$$

$$\hat{N} = (I - H_{11})^{-1} (I - H_{11}) Y_1 - (I - H_{11})^{-1} H_{12} Y_2$$

$$\therefore \hat{N} = Y_1 - (I - H_{11})^{-1} H_{12} Y_2$$

Se puede demostrar usando cálculo vectorial que $\xi^{(u)}, \xi^{(u)}$ alcanza un mínimo cuando:

$$\hat{N} = Y_1 - (I - H_{11})^{-1} H_{12} Y_2$$

C A P I T U L O V

C U A T R O E J E M P L O S

En este capítulo presentaremos cuatro ejemplos que serán tratados con cada método propuesto en el capítulo II. Con el fin de poder establecer comparaciones entre los métodos.

El conjunto de datos para el primer ejemplo se obtuvo de [27] , artículo escrito por Mickey, Dunn y Clark el cual consiste de 21 observaciones, las cuales aparecen a continuación :

Obs. #	x_{1i}	y_{1i}	Obs. #	x_{1i}	y_{1i}
1	15	95	11	7	113
2	26	71	12	9	96
3	10	83	13	10	83
4	9	91	14	11	84
5	15	102	15	11	102
6	20	87	16	10	100
7	18	93	17	12	105
8	11	100	18	42	57
9	8	104	19	17	121
10	20	94	20	11	85
			21	10	100

En este conjunto de datos se sabe que la observación no. 19 es discordante.

El conjunto de datos para el segundo ejemplo se obtuvo a partir del primero modificándose la observación # 10 para tener dos observaciones discordantes en el conjunto; con lo cual el conjunto de datos para el ejemplo # 2 quedó de la siguiente manera :

Obs. #	x_{1i}	y_{1i}	Obs. #	x_{1i}	y_{1i}
1	15	95	11	7	113
2	26	71	12	9	96
3	10	83	13	10	83
4	9	91	14	11	84
5	15	102	15	11	102
6	20	87	16	10	100
7	18	93	17	12	105
8	11	100	18	42	57
9	8	104	19	17	121
10	20	130	20	11	85
			21	10	100

En este conjunto de datos sabemos de antemano que las observaciones 19 y 10 son discordantes.

El conjunto de datos para el segundo ejemplo se obtuvo a partir del primero modificándose la observación # 10 para tener dos observaciones discordantes en el conjunto; con lo cual el conjunto de datos para el ejemplo # 2 quedó de la siguiente manera :

Obs. #	x_{1i}	y_{1i}	Obs. #	x_{1i}	y_{1i}
1	15	95	11	7	113
2	26	71	12	9	96
3	10	83	13	10	83
4	9	91	14	11	84
5	15	102	15	11	102
6	20	87	16	10	100
7	18	93	17	12	105
8	11	100	18	42	57
9	8	104	19	17	121
10	20	130	20	11	85
			21	10	100

En este conjunto de datos sabemos de antemano que las observaciones 19 y 10 son discordantes.

El conjunto de datos para el tercer ejemplo se obtuvo de [26]. Artículo desarrollado por Richard E. Lund, el cual consiste de 18 observaciones; las cuales aparecen a continuación :

Obs. #	y_1	x_{11}	x_{21}
1	64	0.4	53
2	60	0.4	23
3	71	3.1	19
4	61	0.6	34
5	54	4.7	24
6	77	1.7	65
7	81	9.4	44
8	93	10.1	31
9	93	11.6	29
10	51	12.6	58
11	76	10.9	37
12	96	23.1	46
13	77	23.1	50
14	93	21.6	44
15	95	23.1	56
16	54	1.9	36
17	168	26.8	58
18	99	29.9	51

En este conjunto de datos se sabe que la observación 17 es discordante.

MATRIX ELEMENTS

0.267	0.045	-0.013	0.122	0.006	0.335	0.101	-0.000	-0.070	0.167	0.035	-0.020	-0.000	-0.020	0.041	0.123	0.018	-0.068
0.045	0.189	0.159	0.136	0.170	-0.018	0.033	0.118	0.122	-0.021	0.086	0.001	-0.019	0.016	-0.047	0.122	-0.069	-0.046
-0.013	0.159	0.228	0.123	0.170	-0.093	0.046	0.138	0.152	-0.054	0.095	0.026	-0.007	0.041	-0.045	0.137	-0.061	-0.017
0.122	0.136	0.122	0.130	0.110	0.109	0.073	0.075	0.067	0.043	0.048	-0.008	-0.010	0.001	-0.013	0.121	-0.035	-0.031
0.006	0.170	0.150	0.117	0.162	-0.060	0.051	0.121	0.131	-0.027	0.036	0.032	0.011	0.044	-0.022	0.078	-0.037	0.000
0.335	-0.018	-0.050	0.107	-0.050	0.440	0.116	-0.045	-0.037	0.241	0.015	-0.032	0.015	-0.026	0.048	0.116	0.043	-0.058
0.101	0.058	0.045	0.073	0.071	0.118	0.071	0.050	0.044	0.085	0.038	0.047	0.053	0.047	0.062	0.074	0.057	0.041
-0.000	0.118	0.138	0.073	0.121	-0.045	0.030	0.103	0.113	0.001	0.083	0.065	0.050	0.071	0.026	0.059	0.024	0.057
-0.030	0.122	0.152	0.067	0.171	-0.027	0.044	0.113	0.128	-0.017	0.035	0.070	0.055	0.054	0.020	0.061	0.078	0.073
0.169	-0.021	-0.054	0.043	-0.027	0.241	0.035	0.001	-0.017	0.164	0.036	0.050	0.032	0.043	0.120	0.057	0.121	0.047
0.037	0.086	0.045	0.067	0.091	0.013	0.038	0.030	0.025	0.036	0.071	0.065	0.053	0.067	0.048	0.065	0.068	0.047
-0.000	0.001	0.025	-0.003	0.032	-0.032	0.047	0.065	0.078	0.056	0.065	0.147	0.143	0.134	0.136	-0.001	0.162	0.151
-0.000	-0.018	-0.003	-0.010	0.011	0.015	0.053	0.050	0.058	0.023	0.058	0.143	0.145	0.132	0.148	-0.000	0.174	0.150
-0.028	0.018	0.041	0.001	0.044	-0.036	0.047	0.071	0.084	0.048	0.067	0.138	0.132	0.130	0.123	0.007	0.165	0.137
0.041	-0.047	-0.045	-0.113	-0.022	0.085	0.062	0.026	0.028	0.120	0.048	0.136	0.146	0.123	0.166	-0.000	0.171	0.156
0.172	0.122	0.107	0.121	0.099	0.113	0.074	0.069	0.061	0.057	0.065	-0.001	-0.000	0.007	-0.000	0.114	-0.020	-0.077
0.018	-0.065	-0.061	-0.033	-0.035	0.063	0.037	0.024	0.029	0.121	0.048	0.162	0.174	0.145	0.191	-0.070	0.224	0.174
-0.067	-0.046	-0.013	-0.051	0.037	-0.050	0.041	0.057	0.075	0.063	0.062	0.192	0.190	0.177	0.198	-0.027	0.226	0.165

El conjunto de datos para el cuarto ejemplo se obtuvo a partir del tercero, modificándose la observación # 18 para tener dos observaciones discordantes en el conjunto; con lo cual el conjunto de datos para el ejemplo # 2 quedó de la siguiente manera :

Obs. #	y_i	x_{1i}	x_{2i}
1	64	0.4	53
2	60	0.4	23
3	71	3.1	19
4	61	0.6	34
5	54	4.7	24
6	77	1.7	65
7	81	9.4	44
8	93	10.1	31
9	93	11.6	29
10	51	12.6	58
11	76	10.9	37
12	96	23.1	46
13	77	23.1	50
14	93	21.6	44
15	95	23.1	56
16	54	1.9	36
17	168	26.8	58
18	169	29.9	65

En este conjunto de datos sabemos de antemano que las observaciones 18 y 17 son discordantes .

MATRIX EXAMPLE 4

0.271	0.050	-0.011	0.129	0.707	0.244	0.000	-0.006	-0.038	0.163	0.020	-0.042	-0.070	-0.040	0.034	0.124	-0.005	0.110
0.050	0.100	0.183	0.132	0.162	-0.005	0.061	0.115	0.119	-0.010	0.086	0.000	-0.003	0.022	-0.134	0.111	-0.004	-0.054
-0.011	0.132	0.115	0.113	0.132	-0.031	0.050	0.137	0.160	-0.043	0.097	0.039	0.012	0.052	-0.020	0.103	-0.040	-0.066
0.129	0.132	0.115	0.123	0.135	0.120	0.074	0.071	0.062	0.053	0.065	-0.011	-0.011	-0.031	-0.112	0.171	-0.035	-0.054
0.007	0.167	0.182	0.103	0.176	-0.055	0.053	0.120	0.130	-0.020	0.009	0.042	0.021	0.052	-0.110	0.065	-0.020	-0.056
0.344	-0.005	-0.091	0.120	-0.035	0.445	0.111	-0.051	-0.096	0.224	0.006	-0.062	-0.015	-0.064	0.025	0.125	0.003	0.063
0.050	0.061	0.050	0.074	0.053	0.111	0.053	0.050	0.045	0.070	0.050	0.041	0.044	0.047	0.054	0.074	0.042	0.151
-0.006	0.115	0.137	0.071	0.120	-0.051	0.050	0.104	0.114	0.101	0.032	0.071	0.070	0.077	0.032	0.060	0.032	0.111
-0.038	0.119	0.170	0.043	0.130	-0.056	0.045	0.115	0.132	-0.016	0.037	0.023	0.067	0.093	0.071	0.057	0.040	0.100
0.163	-0.010	-0.043	0.023	-0.120	0.223	0.079	0.091	-0.016	0.140	0.033	0.040	0.043	0.124	0.071	0.063	0.036	0.125
0.025	0.038	0.027	0.053	0.039	0.008	0.026	0.032	0.037	0.033	0.071	0.037	0.050	0.060	0.041	0.023	0.040	0.037
-0.045	0.000	0.030	-0.011	0.042	-0.042	0.041	0.072	0.033	0.040	0.027	0.142	0.141	0.140	0.130	-0.004	0.155	0.147
-0.070	-0.000	0.012	-0.011	0.071	-0.015	0.046	0.056	0.067	0.063	0.060	0.141	0.139	0.131	0.137	-0.003	0.162	0.151
-0.046	0.023	0.022	-0.051	0.072	-0.004	0.042	0.077	0.033	0.074	0.060	0.140	0.131	0.124	0.110	0.004	0.141	0.141
0.024	-0.004	-0.033	-0.012	-0.110	0.055	0.034	0.032	0.036	0.008	0.043	0.120	0.117	0.115	0.140	-0.001	0.170	0.002
0.121	0.110	0.107	0.121	0.035	0.125	0.074	0.060	0.057	0.060	0.043	-0.004	-0.003	0.034	-0.001	0.114	-0.001	-0.037
-0.005	-0.054	-0.040	-0.031	-0.120	0.025	0.043	0.032	0.040	0.076	0.040	0.155	0.162	0.141	0.170	-0.021	0.002	0.030
0.010	-0.004	-0.006	-0.054	-0.036	0.062	0.051	0.011	0.016	0.125	0.036	0.157	0.151	0.142	0.200	-0.037	0.036	0.037

RESULTADOS OBTENIDOS CON EL METODO DE LUND

Para el ejemplo # 1 se observa que el $\max |t_i| = 2.823$ que corresponde a la observación # 19 . Ocupando un nivel de significancia $\alpha = .1$ y utilizando las tablas de Lund para este nivel con $n = 21$ y $p = 2$ obtenemos que el valor crítico es :

$$t_{0,n,p,\alpha} = 2.64$$

Entonces debido a que :

$$\max |t_i| = 2.823 > 2.64 = t_{0,n,p,\alpha}$$

Se declara a la observación 19 como discordante

Para el ejemplo # 2 observamos que el $\max |t_i| = 2.86$ correspondiente a la observación # 10. Ocupando el mismo nivel de significancia de .1 y las tablas de Lund para $n = 21$ y $p = 2$ obtenemos que el valor crítico es :

$$t_{0,n,p,\alpha} = 2.64$$

Por lo tanto, debido a que :

$$\max |t_i| = 2.86 > 2.64 = t_{0,n,p,\alpha}$$

la observación 10 es declarada como discordante .

Con el ejemplo # 3 observamos que el $\max |t_i| = 3.224$ que corresponde a la observación # 17. Ocupando un nivel de significancia de .1 y recurriendo a las tablas de Lund con $n = 18$ y $p = 3$ se obtiene el valor crítico :

$$t_{0,n,p,\alpha} = 2.97$$

Por lo tanto debido a que :

$$\max |t_i| = 3.224 > 2.97 = t_{0,n,p,\alpha}$$

la observación 17 es declarada como discordante.

Por último para el ejemplo # 4 podemos observar que el $\max |t_i| = 2.14$ correspondiente a la observación # 17. Consultando las tablas de Lund con $\alpha = .1$, $n = 18$ y $p = 3$ se obtiene el siguiente valor crítico :

$$t_{0,n,p,\alpha} = 2.97$$

Debido a que :

$$\max |t_i| = 2.14 < 2.97 = t_{0,n,p,\alpha}$$

Concluimos que no existen observaciones discordantes para este caso.

M E T O D O D E L U N D

Tabla para el ejemplo # 1 .

i	$(1 - h_{ii})$	$\sqrt{\frac{\sum \epsilon_i^2}{n-p} (1-h_{ii})}$	ϵ_i	$ t_i $
1	.953	10.75	2.03	.188
2	.847	10.13	-9.57	-.943
3	.938	.093	-15.60	-1.46
4	.93	10.62	- 8.73	- .821
5	.953	10.75	9.03	.839
6	.928	10.61	.33	.031
7	.943	10.70	3.41	.318
8	.944	10.70	2.52	.235
9	.921	10.56	3.14	.297
10	.928	10.61	6.66	.627
11	.91	10.50	11.01	1.048
12	.93	10.62	-3.73	-.351
13	.938	10.67	-15.60	-1.460
14	.944	10.70	-13.47	-1.258
15	.944	10.70	4.52	.422
16	.938	10.66	1.39	.130
17	.948	10.72	8.65	.806
18	.352	6.53	-5.54	-.847
19	.948	10.72	30.28	2.823
20	.944	10.70	-11.47	-1.071
21	.938	16.66	1.39	1.30

Tabla para el ejemplo # 2 .

i	$(1-h_{1i})$	$\sqrt{\frac{\sum \epsilon_i^2 (1-h_{1i})}{n-p}}$	ϵ_i	$ t_i $
1	.953	14.14	.21	.015
2	.847	13.33	-13.15	.986
3	.938	14.03	-16.61	1.183
4	.93	13.97	- 9.58	.685
5	.953	14.14	7.21	.509
6	.928	13.95	- 2.94	.210
7	.943	14.07	1.12	.079
8	.944	14.07	1.35	.095
9	.921	14.49	2.44	.1616
10	.928	13.95	40.05	2.869
11	.91	13.82	10.48	.758
12	.93	13.97	-4.58	.327
13	.938	14.03	-16.61	1.18
14	.944	14.07	-14.64	1.039
15	.944	14.07	3.35	.237
16	.938	14.03	.38	.027
17	.948	14.10	7.32	.518
18	.352	8.59	-11.67	1.357
19	.948	14.10	28.15	1.99
20	.944	14.07	-12.65	.89
21	.938	14.03	.38	.027

Tabla para el ejemplo # 3 .

1	$(1-h_{ii})$	$\sqrt{\frac{\sum \epsilon_i^2 (1-h_{ii})}{n-p}}$	ϵ_i	$ t_i $
1	.7381	17.76	2.44	.137
2	.8107	18.62	1.04	.055
3	.7736	18.18	7.55	.415
4	.8705	19.29	.73	.037
5	.8379	18.92	-12.74	.673
6	.5403	15.20	12.07	.794
7	.9293	19.93	4.11	.206
8	.8974	19.59	15.99	.816
9	.872	19.31	13.47	.697
10	.8362	18.91	-32.83	1.736
11	.9293	19.93	- 2.97	.150
12	.8534	19.10	- 5.58	.292
13	.8551	19.12	-24.93	1.303
14	.8698	19.28	- 5.72	.296
15	.8336	18.88	- 7.45	.394
16	.886	19.46	- 8.77	.450
17	.7765	18.22	58.76	3.224
18	.7355	17.73	-15.18	.855

Tabla para el ejemplo # 4 .

1	$(1-h_{ii})$	$\sqrt{\frac{\sum_{i=1}^n \epsilon_i (1-h_{ii})}{n-p}}$	ϵ_i	$ t_i $
1	.7289	19.55	4.056	.207
2	.8204	20.74	6.932	.334
3	.7847	20.28	12.408	.611
4	.8707	21.36	4.934	.230
5	.8436	21.03	- 9.554	.454
6	.5347	16.74	11.205	.669
7	.9321	22.10	1.648	.074
8	.8962	21.67	14.958	.689
9	.8685	21.34	11.838	.554
10	.8523	21.14	-39.194	1.853
11	.929	22.07	- 5.325	.241
12	.8514	21.13	-16.492	.780
13	.8607	21.24	-36.409	1.713
14	.8663	21.31	-15.455	.724
15	.8514	21.13	-19.784	.936
16	.8857	21.55	- 5.625	.261
17	.7969	20.44	43.929	2.14
18	.7128	19.33	35.930	1.858

T A B L A S D E L U N D

($\alpha = .10$)

n	q									
	1	2	3	4	5	6	8	10	15	25
5	1.87									
6	2.00	1.89								
7	2.10	2.02	1.90							
8	2.18	2.12	2.03	1.91						
9	2.24	2.20	2.13	2.05	1.92					
10	2.30	2.26	2.21	2.15	2.06	1.92				
12	2.39	2.37	2.33	2.29	2.24	2.17	1.93			
14	2.47	2.45	2.42	2.39	2.36	2.32	2.19	1.94		
16	2.53	2.51	2.50	2.47	2.45	2.42	2.34	2.20		
18	2.58	2.57	2.56	2.54	2.52	2.50	2.44	2.35		
20	2.63	2.62	2.61	2.59	2.58	2.56	2.52	2.46	2.11	
25	2.72	2.72	2.71	2.70	2.69	2.68	2.66	2.63	2.50	
30	2.80	2.79	2.79	2.78	2.77	2.77	2.75	2.73	2.66	2.13
35	2.86	2.85	2.85	2.85	2.84	2.84	2.82	2.81	2.77	2.55
40	2.91	2.91	2.90	2.90	2.90	2.89	2.88	2.87	2.84	2.72
45	2.95	2.95	2.95	2.95	2.94	2.94	2.93	2.93	2.90	2.82
50	2.99	2.99	2.99	2.99	2.98	2.98	2.98	2.97	2.95	2.89
60	3.06	3.06	3.05	3.05	3.05	3.05	3.05	3.04	3.03	3.00
70	3.11	3.11	3.11	3.11	3.11	3.11	3.10	3.10	3.09	3.07
80	3.16	3.16	3.16	3.15	3.15	3.15	3.15	3.15	3.14	3.12
90	3.20	3.20	3.19	3.19	3.19	3.19	3.19	3.19	3.18	3.17
100	3.23	3.23	3.23	3.23	3.23	3.23	3.23	3.22	3.22	3.21

T A B L A S D E L U N D

($\alpha = .01$)

n	q									
	1	2	3	4	5	6	8	10	15	25
5	1.98									
6	2.17	1.98								
7	2.32	2.17	1.98							
8	2.44	2.32	2.18	1.98						
9	2.54	2.44	2.33	2.18	1.99					
10	2.62	2.55	2.45	2.33	2.18	1.99				
12	2.76	2.70	2.64	2.56	2.46	2.34	1.99			
14	2.86	2.82	2.78	2.72	2.65	2.57	2.35	1.99		
16	2.95	2.92	2.88	2.84	2.79	2.78	2.58	2.35		
18	3.02	3.00	2.97	2.94	2.90	2.85	2.75	2.59		
20	3.08	3.06	3.04	3.01	2.98	2.95	2.87	2.76	2.20	
25	3.21	3.19	3.18	3.16	3.14	3.12	3.07	3.01	2.78	
30	3.30	3.29	3.28	3.26	3.25	3.24	3.21	3.17	3.04	2.21
35	3.37	3.36	3.35	3.34	3.34	3.33	3.30	3.28	3.19	2.81
40	3.43	3.42	3.42	3.41	3.40	3.40	3.38	3.36	3.30	3.05
45	3.48	3.47	3.47	3.46	3.46	3.45	3.44	3.43	3.38	3.23
50	3.52	3.52	3.51	3.51	3.51	3.50	3.49	3.48	3.45	3.34
60	3.60	3.59	3.59	3.59	3.58	3.58	3.57	3.56	3.54	3.48
70	3.65	3.65	3.65	3.65	3.64	3.64	3.64	3.63	3.61	3.57
80	3.70	3.70	3.70	3.70	3.69	3.69	3.69	3.68	3.67	3.64
90	3.74	3.74	3.74	3.74	3.74	3.74	3.73	3.73	3.72	3.70
100	3.78	3.78	3.78	3.77	3.77	3.77	3.77	3.77	3.76	3.74

T A B L A S D E L U N D

($\alpha = .05$)

n										
	1	2	3	4	5	6	8	10	15	25
5	1.92									
6	2.07	1.93								
7	2.19	2.08	1.94							
8	2.28	2.20	2.10	1.94						
9	2.35	2.29	2.21	2.10	1.95					
10	2.42	2.37	2.31	2.22	2.11	1.95				
12	2.52	2.49	2.45	2.39	2.33	2.24	1.96			
14	2.61	2.58	2.55	2.51	2.47	2.41	2.25	1.96		
16	2.68	2.66	2.63	2.60	2.57	2.53	2.43	2.26		
18	2.73	2.72	2.70	2.68	2.65	2.62	2.55	2.44		
20	2.78	2.77	2.76	2.74	2.72	2.70	2.64	2.57	2.15	
25	2.89	2.88	2.87	2.86	2.84	2.83	2.80	2.76	2.60	
30	2.96	2.96	2.95	2.94	2.93	2.93	2.90	2.88	2.79	2.17
35	3.03	3.02	3.02	3.01	3.00	3.00	2.98	2.97	2.91	2.64
40	3.08	3.08	3.07	3.07	3.06	3.06	3.05	3.03	3.00	2.84
45	3.13	3.12	3.12	3.12	3.11	3.11	3.10	3.09	3.06	2.96
50	3.17	3.16	3.16	3.16	3.15	3.15	3.14	3.14	3.11	3.04
60	3.23	3.23	3.23	3.23	3.22	3.22	3.22	3.21	3.20	3.15
70	3.29	3.29	3.28	3.28	3.28	3.28	3.27	3.27	3.26	3.23
80	3.33	3.33	3.33	3.33	3.33	3.33	3.32	3.32	3.31	3.29
90	3.37	3.37	3.37	3.37	3.37	3.37	3.36	3.36	3.36	3.34
100	3.41	3.41	3.40	3.40	3.40	3.40	3.40	3.40	3.39	3.38

RESULTADOS OBTENIDOS CON EL METODO DE ELLENBERG

En lo que respecta al ejemplo # 1, nos podemos dar cuenta en la tabla correspondiente que el $\max |r_i| = .647$. Ocupando las tabls de Lund para $n=21$, $p=2$ y $\alpha=.1$ encontramos el valor 2.64 para poder usarlo en este ejemplo es necesario dividirlo entre $(n-p)^{1/2} = 4.358$ obteniendo :

$$C_{\alpha} = .605$$

Observamos que $\max |r_i| = |r_{19}| = .647 > .605 = C_{\alpha}$ por lo tanto se rechaza la hipótesis nula de que no existen observaciones discordantes y la observación 19 es declarada discordante.

Para el ejemplo # 2 se puede observar que el $\max |r_i| = .658$ correspondiente a la observación # 10. Teniendo $n = 21$, $p = 2$ y un nivel de significancia de $\alpha = .1$ y buscando el correspondiente valor en tablas de Lund encontramos 2.64 valor que es necesario dividirlo entre $(n-p)^{1/2} = 4.358$ obtenemos :

$$C_{\alpha} = .605$$

Debido a que el $\max |r_i| = |r_{10}| = .658 > C_{\alpha} = .605$ se rechaza la hipótesis nula de que no existen observaciones discordantes y la observación #10 es declarada discordante.

El máx $|r_i|$ para el ejemplo # 3 es el correspondiente a la observación # 17 con el siguiente valor :

$$\text{máx } |r_i| = |r_{17}| = .832$$

Para el mismo nivel de significancia ocupado en los anteriores ejemplos, $p = 3$ y $n = 18$ el valor en las tablas de Lund es 2.97, y para poderlo aplicar a este ejemplo es necesario hacer : $2.97 / (18 - 3)^{1/2} = .766 = C_\alpha$

Como el máx $|r_i| = |r_{17}| = .832 > .766 = C_\alpha$ la hipótesis nula de que no existen observaciones discordantes se rechaza y la observación 17 se declara discordante.

Para el ejemplo # 4 el máx $|r_i|$ es el correspondiente a la observación # 17 con el valor de .552. El valor en las tablas de Lund, para el mismo número de observaciones que en el ejemplo # 3, el mismo número de parámetros desconocidos y el mismo nivel de significancia, es 2.97, dividiendo este valor entre $(n-p)^{1/2}$ es = .766.

Debido a que el máx $|r_i| = |r_{17}| = .552 < .766 = C_\alpha$ no rechazamos la hipótesis nula de que no existen observaciones discordantes

M E T O D O D E E L L E N B E R G

Tabla para el ejemplo # 1 .

1	$ t_i $	$ f_i $
1	.188	.0431
2	.943	.216
3	1.46	.335
4	.821	.188
5	.839	.192
6	.031	.00711
7	.318	.0729
8	.235	.0539
9	.297	.0681
10	.627	.143
11	1.048	.240
12	.351	.0805
13	1.460	.335
14	1.258	.288
15	.422	.096
16	.130	.0298
17	.806	.184
18	.847	.194
19	2.823	.647
20	1.071	.245
21	.130	.0298

Tabla para el ejemplo # 2 .

1	$ t_i $	$ r_i $
1	.015	.003
2	.986	.226
3	1.183	.271
4	.685	.157
5	.509	.116
6	.210	.048
7	.079	.018
8	.095	.021
9	.1616	.037
10	2.869	.658
11	.758	.173
12	.327	.075
13	1.18	.270
14	1.039	.238
15	.237	.0543
16	.027	.006
17	.518	.118
18	1.357	.311
19	1.99	.456
20	.89	.204
21	.027	.006

Tabla para el ejemplo # 3 .

i	$ t_i $	$ r_i $
1	.137	.035
2	.055	.014
3	.415	.107
4	.037	.009
5	.673	.173
6	.794	.205
7	.206	.053
8	.816	.210
9	.697	.179
10	1.736	.448
11	.150	.038
12	.292	.075
13	1.303	.336
14	.296	.076
15	.394	.101
16	.450	.116
17	3.224	.832
18	.855	.220

Tabla para el ejemplo # 4 .

i	$ t_i $	$ r_i $
1	.207	.053
2	.334	.086
3	.611	.157
4	.230	.059
5	.454	.117
6	.669	.172
7	.074	.191
8	.689	.177
9	.554	.143
10	1.853	.478
11	.241	.062
12	.780	.201
13	1.713	.442
14	.724	.186
15	.936	.241
16	.261	.067
17	2.14	.552
18	1.858	.479

RESULTADOS OBTENIDOS CON EL METODO DE COOK Y PRESCOTT

Este método está basado en la misma estadística de prueba que la del método de Ellenberg, es decir el máximo $|W_i|$ en este caso denotada como $\max |W_i|$.

El $\max |W_i|$ para el ejemplo # 1 corresponde a la observación no. 19. Esto anterior lo podemos verificar observando las tablas para los ejemplos con el método de Ellenberg.

Es decir, para el ejemplo 1 tenemos lo siguiente:

$\max |W_i| = |W_{19}| = .647$; teniendo un nivel de significancia $\alpha = .1$ se tiene:

$$\alpha = nP(F > d^2 (n-p-1)/(1-d^2))$$

$$\alpha = nP(F > d^2 (18)/(1-d^2))$$

\Rightarrow

$$.1 = 21 P(F > d^2(18)/(1-d^2))$$

$$.1/21 = P(F > d^2(18)/(1-d^2))$$

$$.1/21 = 1 - P(F < d^2(18)/(1-d^2))$$

$$1 - .1/21 = P(F < d^2(18)/(1-d^2))$$

$$1 - .00476 = P(F < d^2(18)/(1-d^2))$$

$$.9952238 = P(F > d^2(18)/(1-d^2)) \text{ buscando en}$$

tablas e interpolando tenemos:

$$d^2(18)/(1-d^2) = 10.3$$

$$\Rightarrow d^2/(1-d^2) = 10.3/18 = .57222$$

$$\Rightarrow d^2 = .57222 - .57222d^2$$

$$d^2 (1+.57222) = .57222$$

$$d^2 = .57222 / (1 + .57222) = .3639575$$

$$d = .603288$$

Y $\max |w_i| = .647 > d$ por lo tanto la observación 19 es declarada como discordante.

Observando la tabla del método de Ellenberg para el ejemplo # 2 nos podemos dar cuenta que el $\max |w_i| = |w_{10}| = .658$ teniendo en consideración un nivel de significancia de $\alpha = .1$ y calculando el valor de d de la misma forma que para el ejemplo # 1, tenemos lo siguiente :

$$d = .603288$$

Y debido a que $\max |w_i| > d$ rechazamos la hipótesis nula de que no existen observaciones discordantes y la observación 10 es declarada discordante.

Para el ejemplo # 3 el $\max |w_i|$ es el correspondiente a la observación 17 con un valor de $\max |w_{17}| = .832$, para $\alpha = .1$ se tiene lo siguiente :

$$.1 = 18 P(F > d^2(14)/(1-d^2))$$

$$1 - .1/18 = P(F < d^2(14)/(1-d^2))$$

$$\Rightarrow .9944 = P(F < d^2(14)/(1-d^2))$$

$$\Rightarrow d^2(14)/(1-d^2) = 11.13$$

$$\Rightarrow d^2/(1-d^2) = 11.13/14 = .795$$

$$d^2 = .795/1.795 = .442896$$

$$d = .6655043$$

Llevando a cabo la regla de decisión podemos darnos cuenta que :

$$\max |w_i| = |w_{17}| > d$$

Por lo tanto se rechaza la hipótesis nula de que no existan observaciones discordantes y la observación # 17 es declarada discordante.

Para el ejemplo # 4 tenemos que el $\max |w_i| = .552$ correspondiente a la observación #17, calculando el valor de d de la misma manera que para el ejemplo # 3 tenemos lo siguiente:

$$d = .6655043$$

Ya que :

$$d > |w_{17}|$$

No rechazamos la hipótesis nula de que no existen observaciones discordantes.

RESULTADOS OBTENIDOS CON EL METODO DE PRESCOTT

Para este método podemos ocupar las tablas de Lund ya que la estadística de prueba es la misma para ambos métodos y es el máximo de los residuales studentizados; cuando nos referimos a las tablas de Lund nos estamos refiriendo a las tablas de los ejemplos con el método de Lund.

Si observamos el método de Lund para el ejemplo # 1 observamos que el máximo residual studentizado corresponde a la observación no. 19. Es decir :

$$\max |t_i| = \max |u_i| / S_1 = |t_{19}| = 2.823$$

para un nivel de significancia de .1, $p = 2$, y $n = 21$ se tiene $F(1,18) = 10.3$ al 100 $(1 - \alpha/n)\%$ y por lo tanto :

$$C = ((n-p)F/(n-p-1+F))^{1/2} \text{ esta dada por :}$$

$$C = ((19)(10.3)/(18+10.3))^{1/2} = (195.7/28.3)^{1/2} = 2.62$$

$$\text{Debido a que } \max |t_{19}| = 2.823 > C = 2.62$$

la observación #19 se declara discordante.

Observando la tabla del método de Lund para el ejemplo # 2 nos podemos dar cuenta que el :

$$\max |t_i| = \max |t_i| / S_i = |t_{10}| = 2.86$$

Se calcula el valor de la cota superior de R_n^* es decir calculamos $C = ((n-p)F/(n-p-1+F))^{1/2}$, donde $F(1,18) = 10.3$ al 100 $(1-\alpha/n)\%$ para $p = 2$, $n = 21$ y un nivel de significancia del .1, y tenemos :

$$C = (19 (10.3)/(18+10.3))^{1/2} = 2.62$$

Ya que $|t_{10}| = 2.86 > 1.64 = C$ la observación 10 es declarada como discordante.

Para el ejemplo # 3 el $\max |t_i| = |t_{17}| = 3.224$

para un nivel de significancia de .1, $p = 3$ y $n = 18$ tenemos lo siguiente :

$$F(1,14) = 11.13 \text{ al } 100 (1-\alpha/n)\%$$

y el valor de $C = ((n-p)F/(n-p-1+F))^{1/2}$ esta dado por :

$$C = (15 (11.13) / 14+11.13)^{1/2} \\ = 2.57$$

Por lo tanto, debido a que $|t_{17}| = 3.224 > 2.57 = C$ se declara a la observación 17 como discordante.

Para el ejemplo # 4 se tiene lo siguiente :

$$\text{máx } |t_i| = 2.14$$

correspondiente a la observación no. 17.

Para $n = 18$, $p = 3$ y un nivel de significancia de .1 tenemos $F(1,14) = 11.13$ al 100 $(1 - \alpha/n) \%$, el valor de C está dado por :

$$C = ((15)(11.13)/(14+11.13))^{1/2} = 2.57$$

y debido a que $|t_{17}| = 2.14 < 2.57 = C$ no rechazamos la hipótesis nula de que no existen observaciones discordantes.

RESULTADOS OBTENIDOS CON EL METODO DE COOK

Para el ejemplo # 1 el $\max D_1$ es el correspondiente a la observación # 18 con valor de .659 el cual corresponde a un valor de $\alpha = .5$; lo cual hace que el estimador de β ($\hat{\beta}$) al eliminar del ajuste a la observación 18 esté en un elipsoide de confianza al 50% por lo tanto esta observación es la de mayor impacto en la determinación de $\hat{\beta}$.

Para el ejemplo # 2 tenemos que el $\max D_1$ es el correspondiente a la observación # 18, teniendo D_1 un valor de 1.69; buscando en las tablas de la $F(2, 19, 1 - \alpha)$ encontramos que el valor de α es .25 (aproximadamente) lo cual hace que el estimador de β ($\hat{\beta}$) al eliminar dicha observación del ajuste se encuentre en un elipsoide de confianza al 75%; por lo tanto esta observación es la de mayor impacto en la determinación de $\hat{\beta}$.

En el ejemplo # 3 podemos notar que el $\max D_1$ es el correspondiente a la observación # 17 con un valor de .987 el cual corresponde a un valor de α de .5; lo cual hace que el estimador de β ($\hat{\beta}$) al eliminar del ajuste a la observación 17 este en un elipsoide de confianza del 50% por lo cual la obs. 17 es la de mayor impacto en la determinación de $\hat{\beta}$.

Por último para el ejemplo # 4 tenemos que el máx D_i es el correspondiente a la observación # 18 con un valor de 1.6 buscando en las tablas de la F correspondiente encontramos que el valor de α es .25 (aproximadamente) lo cual hace que el estimador de β ($\hat{\beta}$) al eliminar dicha observación del ajuste se encuentre en un elipsoide de confianza al 75%; por lo tanto esta observación es la de mayor impacto en la determinación de β .

M E T O D O D E C O O K

Tabla para el ejemplo # 1 :

i	t_i^2	$h_{ii}/p(1-h_{ii})$	D_i
1	.035	.024	.0008
2	.889	.090	.080
3	2.131	.033	.070
4	.674	.037	.024
5	.703	.024	.016
6	.0009	.038	.13 ⁻⁰⁵
7	.101	.030	.003
8	.055	.029	.001
9	.088	.042	.003
10	.393	.038	.014
11	1.098	.049	.053
12	.123	.037	.004
13	2.131	.033	.070
14	1.562	.029	.045
15	.178	.029	.005
16	.016	.033	.0005
17	.649	.027	.017
18	.717	.920	.659
19	7.969	.027	.215
20	1.147	.029	.033
21	.016	.033	.0005

Tabla para el ejemplo # 2 .

1	t_1^2	$h_{11}/p(1-h_{11})$	D_1
1	.0002	.024	.5 ⁻⁰⁵
2	.972	.090	.087
3	1.399	.033	.046
4	.469	.037	.017
5	.259	.024	.006
6	.044	.038	.001
7	.006	.030	.0001
8	.009	.029	.0002
9	.026	.042	.001
10	8.231	.038	.312
11	.574	.049	.016
12	.106	.037	.003
13	1.392	.033	.045
14	1.079	.029	.031
15	.056	.029	.001
16	.0007	.033	.2 ⁻⁰⁴
17	.268	.027	.007
18	1.841	.920	1.69
19	3.96	.027	.106
20	.808	.029	.023
21	.0003	.033	.9 ⁻⁵

Tabla para el ejemplo # 3 .

1	t_1^2	$h_{11}/P(1-h_{11})$	D_1
1	.018	.118	.002
2	.003	.077	.0002
3	.172	.097	.0166
4	.001	.049	.4 ⁻⁰⁴
5	.452	.064	.028
6	.630	.283	.178
7	.042	.025	.001
8	.665	.038	.025
9	.485	.048	.023
10	3.013	.065	.195
11	.0225	.025	.0005
12	.085	.057	.004
13	1.697	.056	.075
14	.087	.049	.004
15	.155	.066	.010
16	.202	.042	.002
17	10.39	.095	.987
18	.731	.119	.086

Tabla para el ejemplo # 4 .

i	t_i^2	$h_{ii}/p(1-h_{ii})$	D_i
1	.042	.123	.005
2	.111	.072	.007
3	.373	.091	.034
4	.052	.049	.002
5	.206	.0617	.13
6	.447	.290	.129
7	.005	.0242	.0001
8	.474	.038	.018
9	.306	.050	.015
10	3.433	.0577	.198
11	.058	.0254	.001
12	.608	.0581	.035
13	2.934	.053	.155
14	.524	.051	.026
15	.876	.058	.050
16	.068	.043	.003
17	4.579	.0849	.388
18	3.452	.1343	1.600

RESULTADOS OBTENIDOS CON EL METODO DE VALENCIA

Primeramente para cada uno de los ejemplos se calcula C como :

$$C = ((n-p)^2 F/N(n-p-1+F))^{1/2}$$

Para el ejemplo # 1 C esta dada por :

$$C = ((21-2)^2 F/21(21-2-1+F))^{1/2} = 2.501$$

En la tabla para el método de Valencia del ejemplo # 1 podemos observar que el máx $|t_{(1)}|$ corresponde a la observación 19 ,ya que, $|t_{(19)}| = 3.610$.

Comparamos el valor de $|t_{(19)}|$ contra el valor de C y tenemos que : (a $|t_{(19)}|$ la vamos a denotar como $|t_m|$)

$|t_m| = 3.610 > 2.501 = C$ por lo tanto y_{19} es "sospechosa". Para saber si es discordante calculamos :

$$x_{(19)} (\hat{\beta} - \hat{\beta}_{(m)}) / S \sqrt{h_{19,19}} = .642$$

como este valor no es significativo la observación 19 no es discordante.

Como y_{19} no fué discordante nos fijamos en el siguiente valor de $|t_{(1)}|$ y este corresponde a la observación #14 a esta la denotamos como $|t_m|$ y efectuamos la comparación teniendo :

$|t_{14}| = |t_m| = 1.279 < 2.50 = C$ debido a que ya se terminó con los $|t_{(1)}| > C$ el método acaba y en este ejemplo \bar{X} observaciones discordantes.

En el ejemplo # 2, C está dada por :

$$C = ((n-p)^2 F / N(n-p-1+F))^{1/2} = 2.501$$

donde $F_{(1,18)} = 10.3$ al 100 ($1 - \alpha/n$) %

Tomamos como $|t_m|$ a $|t_{(10)}|$ y $|t_{(10)}| = 3.711 > 2.501$

por lo tanto y_{19} es "sospechosa". Para verificar si es discordante calculamos :

$$t_{(10)} (\hat{p} - \hat{p}_{(m)}) / S \sqrt{h_{10,10}} = 3.107/3.995 = .777$$

este valor no es significativo por lo tanto hacemos $|t_m|$ a $|t_{(19)}|$

$$|t_{(19)}| = 2.184 < 2.501 = C$$

Debido a que ya terminamos con la observaciones tales que $|t_{(1)}|$ son mayores que C no existen observaciones discordantes.

Para el ejemplo # 3 se tiene el siguiente valor de C :

$$C = ((15)^2 (11.13) / 18(14+11.13))^{1/2} = 2.353$$

ya que $F_{(1,14)} = 11.13$ al 100 ($1 - \alpha/n$) %

El siguiente paso consiste en hacer $|t_m|$ a $|t_{(17)}|$ debido a que $|t_{(17)}| = 5.445$ es el valor máximo encontrado en las tablas para el ejemplo # 3 con el método de Valencia. Comparamos este valor de $|t_{(17)}|$ contra C y se obtiene :

$$|t_{(17)}| = |t_m| = 5.445 > 2.353 = C$$

por lo tanto la observación 17 se declara como sospechosa. Calculamos entonces :

$$x_{(17)} (\hat{\beta} - \hat{\beta}_{(17)}) / S \sqrt{h_{17,17}} = 16.91/10.11 = 1.67$$

como este valor si es significativo la observación 17 es declarada discordante; y por lo tanto procedemos a tratarla de la siguiente manera :

$$y_{17}^m = y_{17} - [e_{17} \sqrt{1-h_{17,17}}] = 168 - 58.75/-.2225 = 92.29$$

Con este valor de y_{17} obtenemos los nuevos residuales (observar la tabla para el ejemplo # 3 con el método de Valencia con la observación discordante tratada)

Con estos nuevos residuales studentizados calculamos nuevamente G obteniendo :

$$G = 2.353 \text{ con } F_{(1,14)} = 11.13$$

En esta tabla nos podemos dar cuenta que el máx $|t_{(i)}|$ es el correspondiente a la observación 10, a este lo denotamos como $|t_m|$ y :

$$|t_m| = |t_{(10)}| = 2.772 > 2.353 = G$$

por lo tanto la observación # 10 es declarada "sospechosa" y se calcula entonces :

$$x_{(10)} (\hat{\beta} - \hat{\beta}_{(10)}) / S \sqrt{h_{10,10}} = .987$$

como este valor no es significativo la observación # 10 no es discordante.

Nos fijamos entonces en el siguiente valor de $|t_{(1)}|$ y es el correspondiente a la observación # 6 con el siguiente valor :

$$|t_{(6)}| = 1.963$$

a este lo denotamos como $|t_n|$ teniendo lo siguiente :

$$|t_n| = |t_{(6)}| = 1.963 < 2.35 = C$$

como ya terminamos con los $t_{(1)}'s > C$ ya no tenemos más observaciones discordantes.

Para el ejemplo # 4 el valor de C esta dado por :

$$C = 2.35 \text{ con una } F_{(1,14)} = 11.13 \text{ al } 100(1 - \alpha/n)\%$$

y observando la primer tabla para el ejemplo # 4 con el método de Valencia nos podemos dar cuenta que el $\max |t_{(1)}|$ corresponde a la observación # 17; efectuando la comparación de este valor de $|t_{(17)}|$ contra el valor de C tenemos :

$$|t_{(17)}| = |t_n| = 2.48 > 2.35 = C$$

por lo tanto la observación # 17 es "sospechosa" y entonces se calcula :

$$x_{(17)} (\hat{P} - \hat{P}_{us}) / S \sqrt{h_{17,17}} = 1.03$$

Debido a que esta razón es significativa la observación 17 se declara discordante y se procede a tratarla; obteniendo la nueva observación : $y_{17}^n = 112.87$

Una vez tratada la observación que fue declarada como discordante obtenemos los nuevos residuales studentizados, estos los podemos ver en la segunda tabla para el ejemplo # 4 con el método de Valencia. Volvemos a calcular el valor de C obteniendo:

$$C = 2.53 \text{ con } F_{(1,14)} = 11.3 \text{ al } 100(1 - \alpha/n)\%$$

hacemos $|t_n| = |t_{(18)}| = 4.4 > 2.35 = C$ por lo tanto la observación

18 se declara como "sospechosa" y se procede a calcular:

$$x_{(18)} (\hat{\beta} - \hat{\beta}_{(18)}) / S \sqrt{h_{18,18}} = 1.82$$

para saber si es discordante, como la razón si es significativa se declara a la observación 18 como discordante y se procede a tratarla obteniendo la nueva observación 18:

$$y_{18}^n = 101.58$$

Obtenemos de nueva cuenta a los residuales studentizados mismos que podemos ver en la tabla # 3 para el ejemplo # 4 con el método de Valencia.

Nuevamente calculamos el valor de C :

$$C = 2.35$$

Observando esta tabla observamos que el $\max |t_{(1)}|$ es el correspondiente a la observación # 10 y efectuamos la comparación entre este valor de $|t_{(10)}| = 2.85$ y el valor de C :

$$|t_{(10)}| = |t_n| = 2.85 > 2.35 = C$$

por lo tanto la observación 10 es "sospechosa"; entonces calculamos:

$$x_{(10)} (\hat{\beta} - \hat{\beta}_{(10)}) / S \sqrt{h_{10,10}} = -.947$$

como esta razón no es significativa la observación # 10 no es discordante.

Fijándonos en el siguiente valor de $t_{(1)}$ podemos observar que corresponde a la observación # 6 :

$$|t_{(6)}| = |t_n| = 1.774 < 2.35 = C$$

como ya se acabó con los $t_{(i)}$, $> C$ se acaba el método y no existen más observaciones discordantes.

M E T O D O D E V A L E N C I A

Tabla para el ejemplo # 1 .

i	$S_{(i)}$	$(1-h_{ii})^{1/2}$	ϵ_i	$t_{(i)}$
1	11.314	.976	2.03	.1838
2	11.055	.920	- 9.57	- .940
3	10.668	.969	-15.60	-1.50
4	11.121	.964	- 8.73	- .814
5	11.112	.976	9.03	.832
6	11.32	.963	- .33	- .030
7	11.29	.971	3.41	.31
8	11.30	.971	2.52	.22
9	11.29	.959	3.14	.290
10	11.20	.9630	6.66	.617
11	10.99	.953	11.01	1.05
12	11.28	.964	-3.73	- .343
13	10.66	.969	-15.60	-1.510
14	10.84	.971	-13.47	1.279
15	11.27	.971	4.52	.413
16	11.31	.968	1.39	.126
17	11.12	.973	8.65	.799
18	11.10	.593	- 5.54	- .841
19	8.62	.973	30.28	3.610
20	10.97	.971	-11.47	-1.076
21	11.31	.968	1.39	.126

Tabla para el ejemplo # 2 .

i	$S_{(i)}$	$(1-h_{ii})^{1/2}$	ϵ_i	$ t_{(i)} $
1	14.899	.976	.218	.015
2	14.503	.920	-13.148	.985
3	14.329	.969	-16.617	1.197
4	14.703	.964	- 9.583	.676
5	14.786	.976	7.217	.500
6	14.871	.963	- 2.948	.206
7	14.886	.971	1.118	.077
8	14.885	.971	1.350	.093
9	14.877	.959	2.449	.172
10	11.207	.963	40.052	3.711
11	14.662	.953	10.483	.750
12	14.847	.964	-4.583	.320
13	14.329	.969	-16.616	1.196
14	14.458	.971	-14.649	1.043
15	14.867	.971	3.350	.232
16	14.889	.968	.383	.026
17	14.783	.973	7.317	.509
18	14.14	.593	-11.678	1.392
19	13.236	.973	28.151	2.184
20	14.56	.971	-12.649	.895
21	14.889	.968	.383	.026

Tabla para el ejemplo # 3 .

i	$s_{(i)}$	$(1-h_{ii})^{1/2}$	ϵ_i	$ t_{(i)} $
1	21.39	.859	2.94	.133
2	21.40	.900	1.04	.054
3	21.28	.879	7.55	.404
4	21.40	.933	.729	.036
5	21.08	.915	-12.74	.660
6	20.95	.735	12.07	.784
7	21.37	.964	4.11	.199
8	20.93	.947	15.98	.806
9	21.05	.934	13.47	.685
10	19.17	.914	-32.83	1.873
11	21.38	.964	- 2.96	.139
12	21.34	.924	- 5.58	.283
13	20.18	.925	-24.93	1.335
14	21.34	.933	- 5.72	.287
15	21.29	.913	- 7.45	.383
16	21.26	.941	- 8.77	.438
17	12.25	.881	58.76	5.445
18	20.89	.857	-15.18	.848

Tabla para el ejemplo 3 con la obs. 17 tratada

i	$S_{(i)}$	$(1-h_{ii})^{1/2}$	ϵ_i	$ t_{(i)} $
1	12.23	.859	2.92	.278
2	12.19	.900	-4.60	.419
3	12.24	.879	2.49	.231
4	12.24	.933	-2.56	.224
5	11.34	.915	-15.96	1.538
6	10.86	.735	15.67	1.963
7	12.09	.964	7.39	.634
8	11.29	.947	16.98	1.588
9	11.51	.934	14.83	1.379
10	9.87	.914	-25.04	2.772
11	12.26	.996	-0.34	.0274
12	12.17	.924	5.17	.459
13	11.65	.925	-13.36	1.239
14	12.21	.9333	3.84	.337
15	12.16	.913	5.34	.481
16	11.86	.941	-10.99	.985
17	12.25	.881	-1.85	.171
18	12.26	.857	.087	.008

Tabla para el ejemplo # 4 .

i	$S_{(i)}$	$(1-h_{ii})^{1/2}$	ξ_i	$ t_{(i)} $
1	23.67	.854	4.05	.2
2	23.62	.906	6.93	.32
3	23.41	.886	12.40	.59
4	23.67	.933	4.93	.22
5	23.55	.918	-9.55	.44
6	23.35	.731	11.20	.656
7	23.71	.654	1.65	.106
8	23.33	.946	14.95	.67
9	23.47	.931	11.84	.54
10	20.85	.923	-39.19	2.03
11	23.66	.964	- 5.32	.23
12	23.23	.923	-16.49	.77
13	21.30	.927	-36.41	1.84
14	23.29	.930	-15.45	.71
15	23.02	.923	-19.78	.93
16	23.65	.941	- 5.62	.25
17	19.81	.892	43.93	2.48
18	20.89	.844	35.93	2.03

Tabla para el ejemplo # 4 con la obs. 17 tratada

i	$s_{(i)}$	$(1-h_{ii})^{1/2}$	ϵ_i	$ t_{(i)} $
1	19.79	.854	3.26	.193
2	19.78	.906	3.70	.206
3	19.58	.886	9.94	.573
4	19.80	.933	2.67	.144
5	19.56	.918	-10.96	.610
6	19.33	.731	11.90	.842
7	19.79	.654	3.77	.291
8	19.27	.946	16.30	.894
9	19.43	.931	13.63	.753
10	17.12	.923	-34.63	2.192
11	19.79	.964	- 3.16	.165
12	19.66	.923	- 8.65	.476
13	18.09	.927	-28.24	1.683
14	19.67	.930	- 8.38	.458
15	19.55	.923	-11.13	.616
16	19.71	.941	- 7.19	.387
17	19.81	.892	- 0.89	.05
18	12.94	.844	48.06	4.4

Tabla para el ejemplo # 4 con la observación 18 tratada

i	$S_{(i)}$	$(1-h_{11})^{1/2}$	ϵ_i	$ t_{(i)} $
1	12.91	.854	3.21	.291
2	12.92	.906	-2.97	.253
3	12.90	.886	3.80	.332
4	12.94	.933	-1.46	.121
5	12.17	.918	-15.16	1.352
6	11.70	.731	15.18	1.774
7	12.83	.654	6.43	.776
8	12.09	.946	16.46	1.433
9	12.30	.931	14.13	1.233
10	10.31	.923	-27.21	2.85
11	12.95	.964	-1.26	.101
12	12.94	.923	1.60	.134
13	11.99	.927	-17.10	1.538
14	12.95	.930	.64	.053
15	12.95	.923	1.32	.110
16	12.62	.941	-10.21	.859
17	12.27	.892	13.95	1.274
18	12.94	.844	-1.35	.123

RESULTADOS OBTENIDOS CON EL METODO GRAFICO

Análisis para el ejemplo # 1 :

Una vez que hemos obtenido los valores de $t_{(i)}$ y T_i los graficamos y procedemos a analizar las graficas obtenidas.

Al hacer el análisis de la gráfica de los valores de $t_{(i)}$ nos podemos dar cuenta de que las observaciones que se encuentran muy alejadas de las demás son únicamente :

La observación 19 por lo tanto se detecta a la observación 19 como "sospechosa". Para determinar si esta observación es discordante se analiza la gráfica de los valores de T_i y podemos observar en esta segunda gráfica que la observación que se encuentra muy alejada del resto es la observación 18; por lo tanto la observación 18 se declara peligrosa, y aún más no se detecta a ninguna observación como discordante.

Análisis para el ejemplo # 2.

Al analizar las gráficas de los valores de $t_{(i)}$ y T_i para el segundo ejemplo, nos pudimos dar cuenta que las observaciones que se encuentran muy alejadas del resto, en la gráfica de los valores de $t_{(i)}$ son las observaciones 10 y 19 por lo que dichas observaciones son declaradas "sospe

chosas". Al hacer el análisis de la gráfica de $T_{1.2}$, para determinar si las anteriores observaciones son discordantes, se puede observar que las únicas observaciones que se encuentran alejadas del resto en esta gráfica son las observaciones 18 y 10.

Debido a que la observación # 10 en ambas gráficas se encuentra alejada del resto se declara como discordante.

Análisis para el ejemplo # 3:

Al observar las gráficas de $t_{(1)}$ y T_1 nos podemos dar cuenta que las observaciones que se encuentran alejadas del resto en la gráfica de los valores de $t_{(1)}$ son las observaciones #13, #10, #17 ; por lo cual se declaran a estas observaciones como sospechosas.

En la gráfica de los valores de T_1 podemos observar que las observaciones que se encuentran alejadas del resto son las observaciones #6, #10, y #17.

Ya que las observaciones 10 y 17 se encuentran alejadas del resto de las observaciones en ambas gráficas se declaran observaciones discordantes.

Análisis para el ejemplo # 4 :

Al hacer el análisis de las gráficas, para este último ejemplo, nos podemos dar cuenta que las observaciones "sospechosas" son las observaciones 13, 10, 18 y 17 debido a que estas se encuentran alejadas del resto en la gráfica de los valores de $t_{(1)}$.

Y las observaciones que se declaran peligrosas son las siguientes :

La 17 y la 18

Por lo tanto las observaciones que se declaran discordantes en este ejemplo son las observaciones 17 y 18, debido a que ambas se encuentran alejadas de las demás observaciones en las dos gráficas.

M E T O D O G R A F I C O

Tabla para el ejemplo # 1 .

i	$ t_{(i)} $	$\left[\frac{n-p}{p} h_{ii} / (1-h_{ii}) \right]^{1/2}$	T_i
1	.183	.684	.125
2	.940	1.309	1.230
3	1.50	.791	1.865
4	.814	.838	.682
5	.832	.675	.561
6	.030	.849	.025
7	.310	.754	.233
8	.220	.742	.163
9	.290	.893	.259
10	.617	.722	.445
11	1.05	.964	1.012
12	.343	.838	.287
13	1.510	.791	1.194
14	1.279	.742	.949
15	.413	.742	.306
16	.126	.627	.079
17	.799	.716	.572
18	.841	4.180	3.515
19	3.610	.513	1.851
20	1.076	.742	.798
21	1.260	.627	.995

Tabla para el ejemplo # 2 .

i	$ t_{ii} $	$\left[\frac{n-p}{p} h_{ii} / (1-h_{ii}) \right]^{1/2}$	r_i
1	.015	.684	.010
2	.985	1.309	1.289
3	1.197	.791	.947
4	.676	.838	.566
5	.500	.675	.337
6	.206	.849	.175
7	.077	.754	.058
8	.093	.742	.069
9	.172	.893	.159
10	3.711	.722	2.679
11	.750	.964	.723
12	.320	.838	.268
13	1.196	.791	.946
14	1.043	.742	.811
15	.232	.742	.172
16	.026	.627	.016
17	.509	.716	.364
18	1.392	4.180	5.818
19	2.184	.513	1.203
20	.895	.742	.664
21	.026	.627	.016

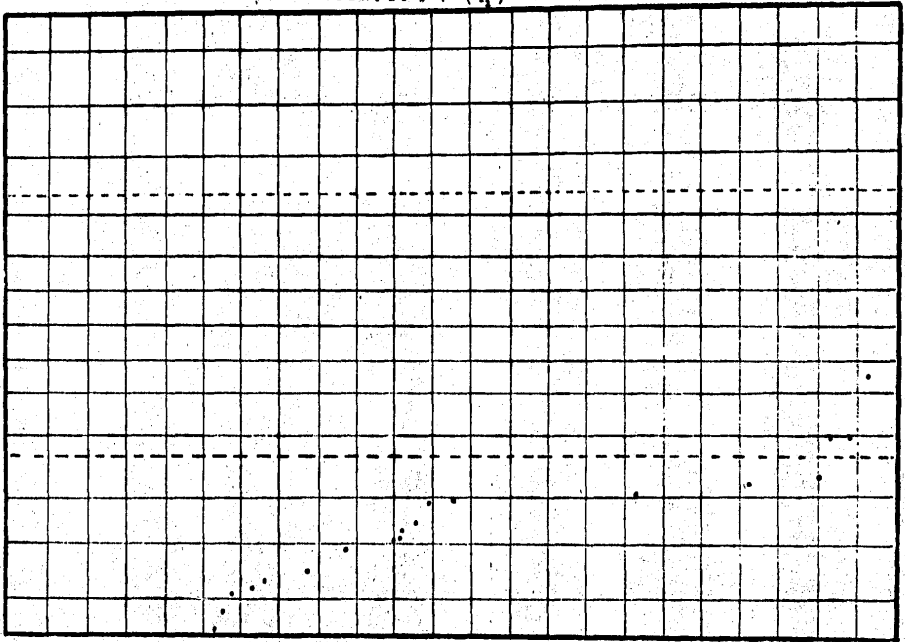
Tabla para el ejemplo # 3 .

i	$ t_{ii} $	$\left[\frac{n-p}{p} h_{ii} / (1-h_{ii}) \right]^{1/2}$	T_i
1	.133	1.330	.176
2	.054	1.075	.058
3	.404	1.206	.487
4	.036	.857	.031
5	.660	.979	.646
6	.784	2.060	1.615
7	.199	.612	.121
8	.806	.755	.608
9	.685	.848	.580
10	1.873	.987	1.848
11	.139	.612	.085
12	.283	.924	.261
13	1.335	.916	1.222
14	.287	.857	.245
15	.383	.994	.380
16	.438	.793	.347
17	5.445	1.193	6.495
18	.848	1.336	1.133

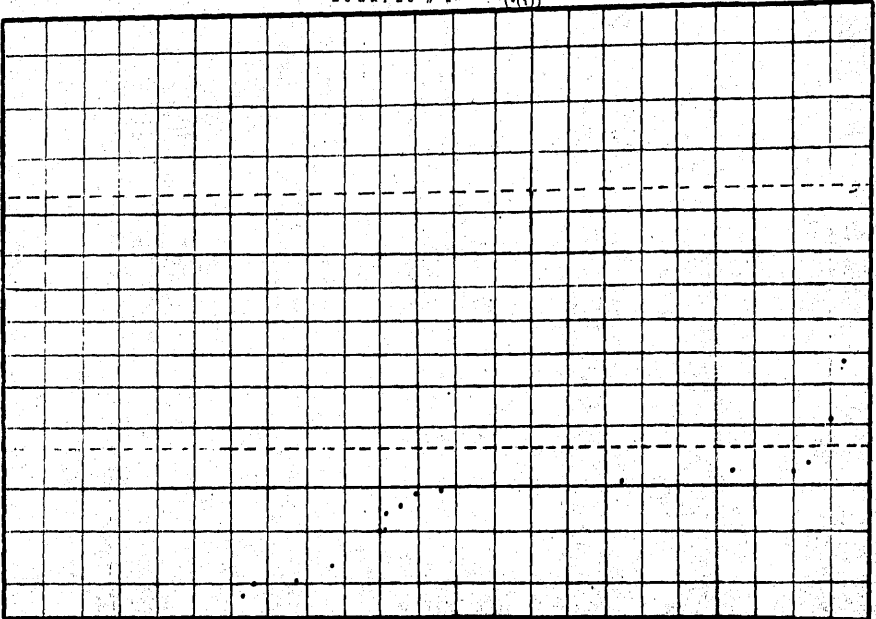
Tabla para el ejemplo # 4 .

1	$ t_{11} $	$\left[\frac{n-p}{p} h_{11} / (1-h_{11}) \right]^{1/2}$	T_1
1	.2	1.368	.271
2	.32	1.039	.332
3	.59	1.168	.689
4	.22	.857	.188
5	.44	.962	.423
6	.656	2.08	1.364
7	.106	.602	.064
8	.67	.755	.505
9	.54	.866	.467
10	2.033	.930	1.88
11	.23	.617	.142
12	.77	.933	.718
13	1.84	.894	1.644
14	.71	.874	.620
15	.93	.932	.866
16	.25	.803	.200
17	2.48	1.128	2.79
18	2.03	1.419	2.88

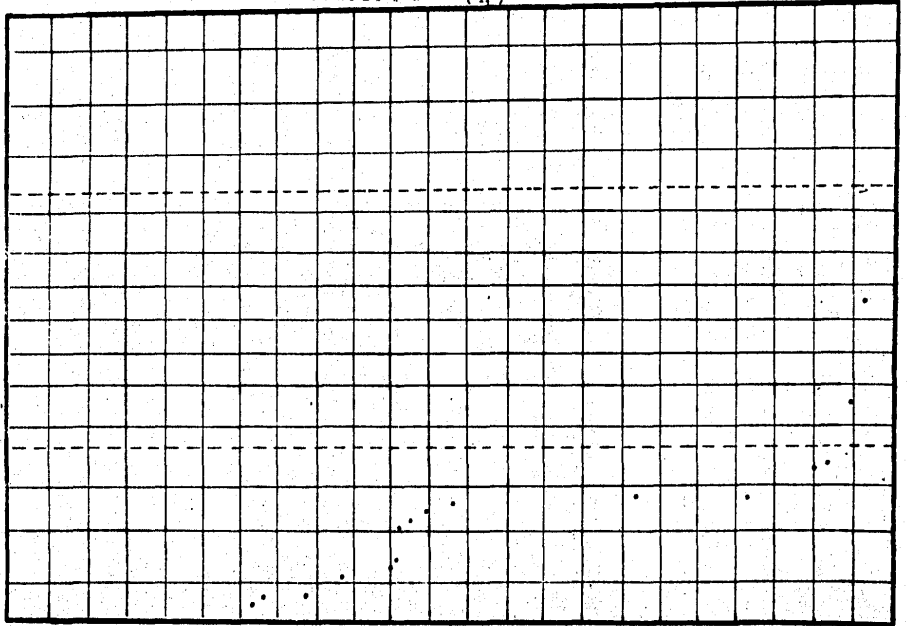
EJEMPLO # 1 (2.)



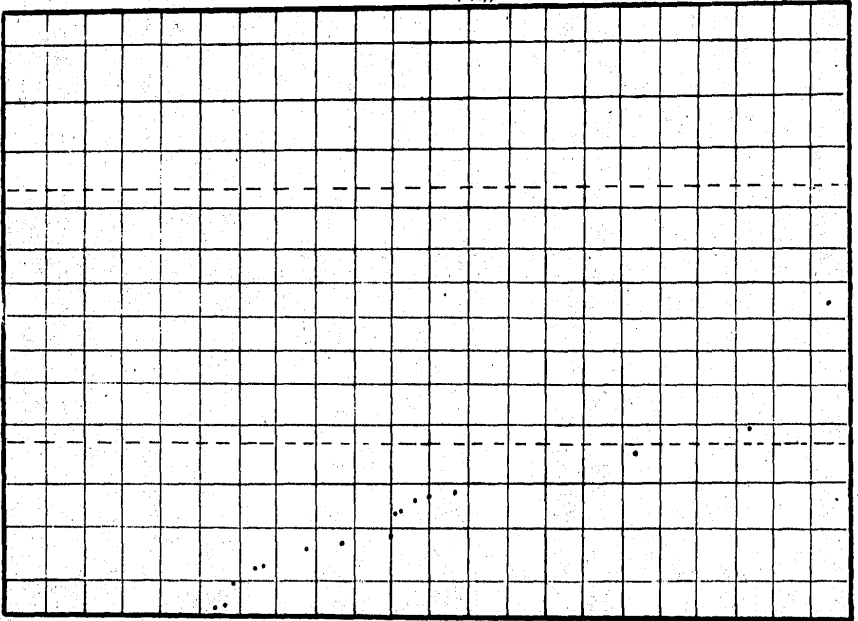
EJEMPLO # 2 (4(1))



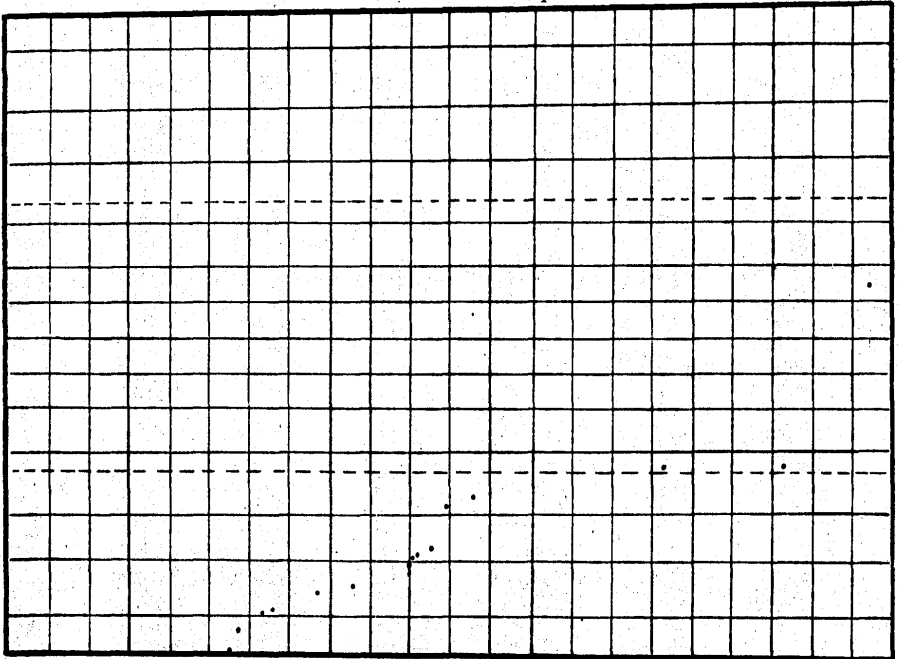
EJEMPLO # 2 (T₁)



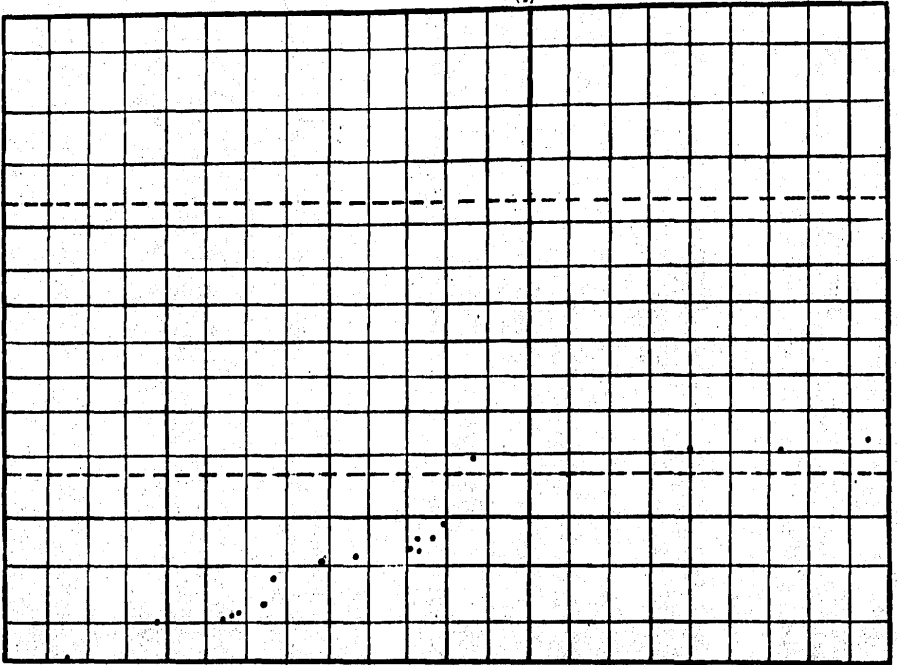
EJEMPLO # 3 (4)



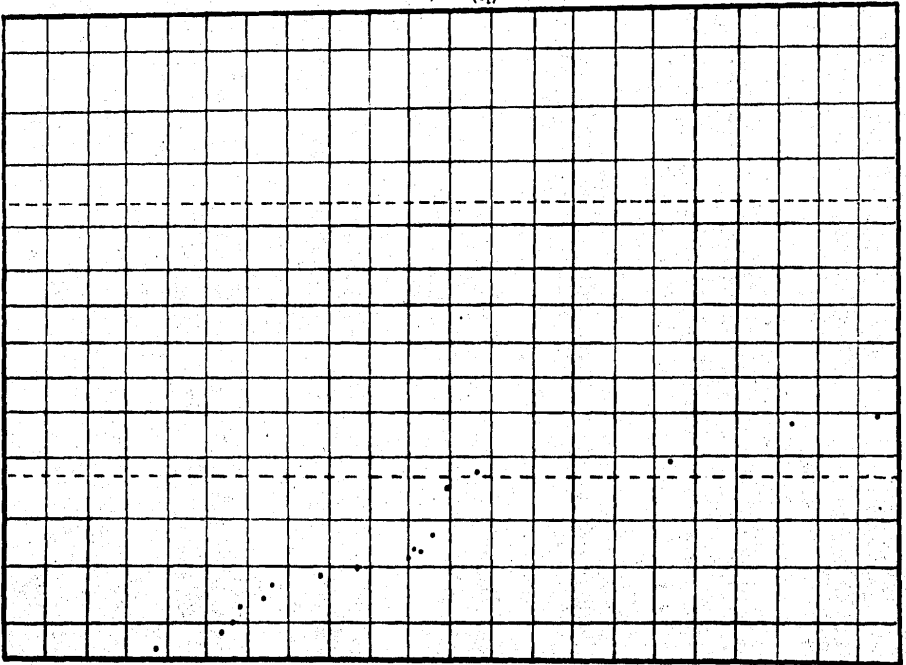
EJEMPLO # 3 T_1



EJEMPLO # 4 (t₁)



EJEMPLO # 4 (π_t)



RESULTADOS PARA EL METODO DE GENTLEMAN Y WILK

Observando la tabla para el método de Gentleman y Wilk para el primer ejemplo y suponiendo inicialmente la existencia de dos observaciones discordantes; es decir suponiendo $k = 2$, nos podemos dar cuenta que el subconjunto de 2 observaciones más probablemente discordante es :

$$Q_2^{**} = 1189.32$$

Partición que corresponde a la pareja (19,3).

Para saber si esta $Q_2^{**} = 1189.32$ es suficientemente grande, estadísticamente hablando, analizamos la gráfica para este ejemplo con $k = 2$, y podemos observar que dicha gráfica no revela que $Q_2^{**} = 1189.32$ sea suficientemente grande en relación con los demás, y estadísticamente hablando, aún mas no revela la existencia de algún par de observaciones discordantes, lo que si se puede apreciar en esta gráfica es que las parejas más posiblemente discordantes son aquellas que tienen involucrada la observación # 19.

Debido a que esta primer gráfica no revela la existencia de algún par de observaciones discordantes suponemos la existencia de una sola observación discordante haciendo $k = 1$ y nos podemos dar cuenta que el subconjunto de 1 observación más probablemente discordante es :

$$Q_1^{**} = 968.56$$

Partición correspondiente a la observación 19.

Para estar seguros de que $Q_1^{**} = 968.56$ es suficientemente grande estadísticamente analizamos la gráfica correspondiente,

y podemos observar que ésta nos revela claramente que la observación 19 es la discordante ya que podemos apreciar que su correspondiente Q_1^m es suficientemente grande en relación con las demás.

Por lo tanto se declara a la observación 19 discordante.

Para el ejemplo # 2 nos podemos dar cuenta, suponiendo primeramente que existen dos observaciones discordantes, que el subconjunto de dos observaciones más probablemente discordantes es el correspondiente a la pareja (10,19), debido a que su correspondiente Q_2 es el máximo de los $Q_{2,g}$. Por lo tanto :

$$Q_2^* = 2729.75$$

Para determinar si este subconjunto de dos observaciones es suficientemente grande en base a un juicio inferencial analizamos la gráfica correspondiente y podemos observar que Q_2^* es suficientemente grande en relación con los demás, estadísticamente hablando, por lo tanto se declara a la pareja (10,19) como discordante.

Analizando los resultados y la gráfica para el ejemplo 3, tenemos primeramente que para $k = 2$ el máximo de las $Q_{2,g}$ es el siguiente :

$$Q_2^* = 5078.61$$

correspondiente a las observaciones (10,17). Para poder inferir si este subconjunto de observaciones es suficientemente aberrante para declararlo como discordante se analiza la gráfica para ver si la correspondiente Q_2^* es suficientemente grande en rela -

ción con los demás y para ello se analiza la gráfica correspondiente y observamos que, aún cuando los pares de observaciones más probablemente discordantes son aquellos que tienen involucrada a la observación 17, no se puede inferir a partir de ella que una de dichas parejas sea discordante ya que no se presenta un cambio brusco entre ellas.

De acuerdo con el método suponemos ahora la existencia de una sola observación discordante y bajo esta suposición hallamos que el máximo de los $Q_{1,s}$ es :

$$Q_1^{\max} = 4312.65$$

correspondiente a la observación # 17 . Analizando la gráfica correspondiente encontramos que esta es suficientemente grande en relación con las demás por lo que se declara a la observación 17 discordante.

Y por último para el ejemplo # 4 tenemos, que bajo la suposición de 2 observaciones discordantes, el máximo de las $Q_{2,s}$ es la correspondiente al par (17,18) con el valor siguiente:

$$Q_2^{\max} = 5770.20$$

Analizando la gráfica correspondiente para poder decidir si este par de observaciones es discordante observamos que su correspondiente Q_2^{\max} es suficientemente grande por lo que se declara al par (17,18) discordante.

METODO DE GENTLEMAN Y WILK

Tabla para el ejemplo # 1 . (sin 1 observación)

Observación omitida	$\sum_{i=1}^n \epsilon_{ii}^2$	Q_k	$\sum_{i=1}^n \epsilon_{ii}^2$
6	2308.46	0.12	2308.58
16	2306.50	2.08	
21	2306.50	2.08	
1	2304.25	4.33	
8	2301.83	6.75	
9	2297.85	10.73	
7	2296.22	12.36	
12	2293.60	14.98	
15	2286.89	21.69	
10	2260.67	47.91	
17	2229.64	78.94	
4	2226.57	82.01	
5	2222.92	85.66	
18	2220.48	88.10	
2	2200.21	108.37	
11	2175.14	133.44	
20	2168.90	139.68	
14	2116.04	192.54	
13	2048.78	259.80	
3	2048.78	259.80	
19	1340.02	968.56	

Tabla para el ejemplo # 1. (sin 2 observaciones)

Observación omitida	$\sum_{i=1}^n \epsilon_{ii}^2$	Q_k	$\sum_{i=1}^n \epsilon_i^2$
3, 19	1119.26	1189.32	2308.58
13, 19	1119.26	1189.32	
11, 19	1180.17	1128.41	
14, 19	1182.51	1126.07	
5, 19	1221.80	1086.78	
20, 19	1230.00	1078.58	
17, 19	1233.99	1074.59	
10, 19	1260.66	1074.92	
2, 19	1276.95	1031.63	
4, 19	1278.35	1030.23	
15, 19	1304.02	1004.56	
7, 19	1311.49	997.09	
9, 19	1320.44	988.14	
8, 19	1324.51	984.07	
18, 19	1325.83	982.75	
1, 19	1326.41	982.17	
19, 21	1332.63	975.95	
16, 19	1332.63	975.95	
12, 19	1332.93	975.65	
6, 19	1337.38	971.20	
3, 13	1751.65	556.93	
3, 14	1826.08	482.50	
13, 14	1826.08	482.50	
2, 18	1867.00	441.58	
13, 20	1883.41	425.17	
3, 20	1883.41	425.17	
2, 3	1937.62	370.96	
2, 13	1937.62	370.96	
3, 11	1942.57	366.01	
11, 13	1942.57	366.01	
4, 13	1944.19	364.39	

Tabla para el ejemplo # 1. (sin dos observaciones)

Observación omitida	$\sum_{i=1}^n \epsilon_{it}^2$	Q_k	$\sum_{i=1}^n \epsilon_{it}^2$
3, 4	1944.19	364.39	2308.58
14, 20	1955.43	353.15	
3, 5	1976.71	331.87	
5, 13	1976.71	331.87	
13, 18	1983.88	324.70	
3, 18	1983.88	324.70	
3, 17	1985.69	322.89	
13, 17	1985.69	322.89	
2, 14	2002.22	306.36	
11, 14	2004.31	304.27	
3, 10	2007.32	301.26	
10, 13	2007.32	301.26	
4, 14	2016.10	292.48	
3, 12	2023.50	285.08	
12, 13	2023.50	285.08	
11, 18	2031.88	276.70	
13, 15	2035.48	273.10	

Tabla para el ejemplo # 2. (sin 2 observaciones)

Observación omitida	$\sum_{i=1}^n \epsilon_{ii}^2$	Q_k	$\sum_{i=1}^n \epsilon_i^2$
10, 19	1260.66	2729.75	3990.41
10, 13	2007.32	1983.09	
10, 3	2007.32	1983.09	
10, 14	2074.53	1915.88	
10, 11	2124.60	1865.81	
10, 20	2126.51	1863.90	
10, 2	2166.68	1823.73	
10, 5	2167.73	1822.68	
10, 17	2176.67	1813.74	
10, 4	2181.77	1808.64	
10, 18	2201.95	1788.46	
10, 15	2236.65	1753.76	
10, 7	2244.69	1745.72	
10, 12	2247.02	1743.39	
10, 9	2248.97	1741.44	
10, 8	2252.60	1737.81	
10, 1	2254.63	1735.78	
10, 16	2257.94	1732.47	
10, 21	2257.94	1732.47	
10, 6	2260.63	1729.78	
2, 18	2672.76	1317.65	
13, 19	2897.70	1092.71	
3, 19	2897.70	1092.71	
18, 19	2935.95	1054.46	
14, 19	2961.62	1028.79	
2, 19	3009.08	981.33	
11, 19	3009.33	981.08	
20, 19	3014.47	975.94	
5, 19	3074.34	916.07	
17, 19	3075.60	914.81	
4, 19	3075.69	914.72	

Tabla para el ejemplo # 2. (sin 2 observaciones)

Observación omitida	$\sum_{i=1}^n \epsilon_{(i)}^2$	Q_k	$\sum_{i=1}^n \epsilon_i^2$
9, 20	3118.05	872.36	3990.41
15, 19	3131.47	858.94	
12, 19	3140.33	850.08	
9, 19	3140.41	850.00	
7, 19	3145.43	844.98	
8, 19	3146.58	843.83	
1, 19	3150.59	839.82	
16, 19	3151.03	839.38	
19, 21	3151.03	839.38	
19, 6	3152.00	838.41	
3, 18	3357.11	633.30	
13, 18	3357.11	633.30	
3, 13	3358.85	631.56	
11, 18	3364.55	625.86	
14, 18	3397.59	592.82	

Tabla para el ejemplo # 3. (sin 1 observación)

Observación omitida	$\sum_{i=1}^n \epsilon_{i0}^2$	Q_k	$\sum_{i=1}^n \epsilon_i^2$
4	6413.33	0.61	6413.94
2	6412.60	1.34	
1	6405.86	8.08	
11	6404.54	9.40	
7	6395.87	18.07	
12	6378.13	35.81	
14	6376.91	37.03	
15	6348.98	64.96	
3	6340.26	73.68	
16	6327.20	86.74	
5	6220.58	193.36	
9	6207.00	206.94	
6	6144.20	269.74	
8	6130.77	283.17	
18	6109.97	303.97	
13	5702.23	711.71	
10	5142.61	1271.33	
17	2101.29	4312.65	

Tabla para el ejemplo # 3. (sin 2 observaciones)

Observación omitida	$\sum_{i=1}^n \epsilon_{i1}$	Q_k	$\sum_{i=1}^n \epsilon_i^2$
10, 17	1335.33	5078.61	6413.94
6, 17	1651.04	4762.90	
8, 17	1783.02	4630.92	
5, 17	1800.77	4613.17	
9, 17	1851.91	4562.03	
13, 17	1878.33	4535.61	
16, 17	1966.68	4447.26	
7, 17	2044.32	4369.62	
15, 17	2071.35	4342.59	
12, 17	2073.56	4340.38	
2, 17	2076.90	4337.04	
14, 17	2086.69	4327.25	
1, 17	2089.85	4324.09	
3, 17	2092.19	4321.75	
4, 17	2094.29	4319.65	
18, 17	2101.08	4312.86	
11, 17	2101.10	4312.84	
10, 13	4272.77	2141.17	
10, 18	4757.64	1656.30	
10, 8	4849.34	1564.60	
10, 9	4904.69	1509.25	
10, 5	4984.91	1429.03	
10, 15	4986.89	1427.05	
10, 16	5010.53	1403.41	
10, 3	5018.29	1395.65	
10, 12	5083.46	1330.48	
10, 14	5086.71	1327.23	
10, 1	5121.08	1292.86	
10, 6	5123.54	1290.40	
10, 11	5125.29	1288.65	
10, 2	5137.95	1275.99	

Tabla para el ejemplo # 3. (sin 2 observaciones)

Observación omitida	$\sum_{i=1}^n \epsilon_i^2$	Q_k	$\sum_{i=1}^n \epsilon_i^2$
10, 7	5141.16	1272.78	6413.94
10, 4	5141.47	1272.47	
13, 18	5150.70	1263.24	
13, 6	5434.90	979.04	
13, 8	5457.47	956.47	
13, 5	5505.32	908.62	
13, 9	5534.95	878.99	
13, 15	5553.58	860.36	
13, 12	5603.73	810.21	
13, 14	5608.69	805.25	
13, 16	5620.12	793.82	
13, 3	5624.20	789.74	
13, 11	5682.31	731.63	
13, 1	5691.98	721.96	

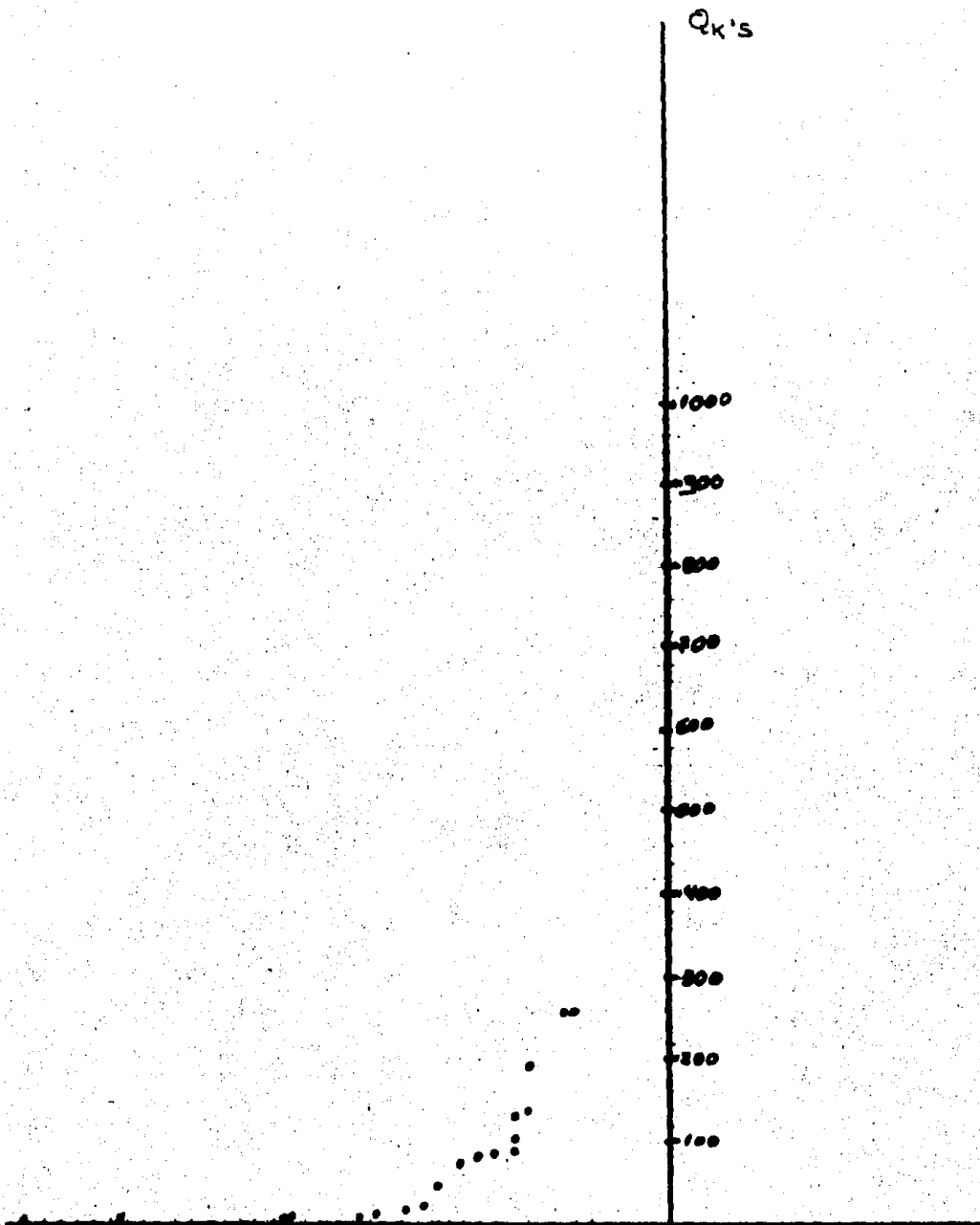
Tabla para el ejemplo # 4. (sin 2 observaciones)

Observación omitida	$\sum_{i=1}^n \epsilon_{ui}^2$	Q_k	$\sum_{i=1}^n \epsilon_i^2$
17, 18	2101.08	5770.20	7871.28
17, 10	4083.01	3788.27	
13, 10	4359.14	3512.14	
17, 13	4543.14	3328.14	
18, 10	4757.64	3113.64	
18, 13	5150.70	2720.58	
17, 8	5202.61	2668.67	
17, 6	5233.53	2637.75	
17, 9	5285.13	2586.15	
17, 15	5344.35	2526.93	
17, 5	5355.94	2515.34	
17, 3	5369.87	2501.41	
17, 12	5404.78	2466.50	
17, 14	5412.50	2458.78	
10, 15	5432.01	2439.27	
17, 16	5439.60	2431.68	
17, 2	5480.13	2391.15	
17, 7	5482.65	2388.63	
17, 1	5482.79	2388.49	
17, 11	5486.48	2384.80	
17, 4	5488.98	2382.30	
13, 15	5617.19	2254.09	
10, 12	5718.84	2152.44	
18, 6	5763.50	2107.78	
10, 14	5775.22	2096.06	
13, 12	5783.50	2087.78	
10, 3	5820.15	2051.13	
10, 8	5831.49	2039.79	
18, 8	5857.46	2013.82	
13, 14	5864.24	2007.04	

Tabla para el ejemplo # 4. (sin 2 observaciones)

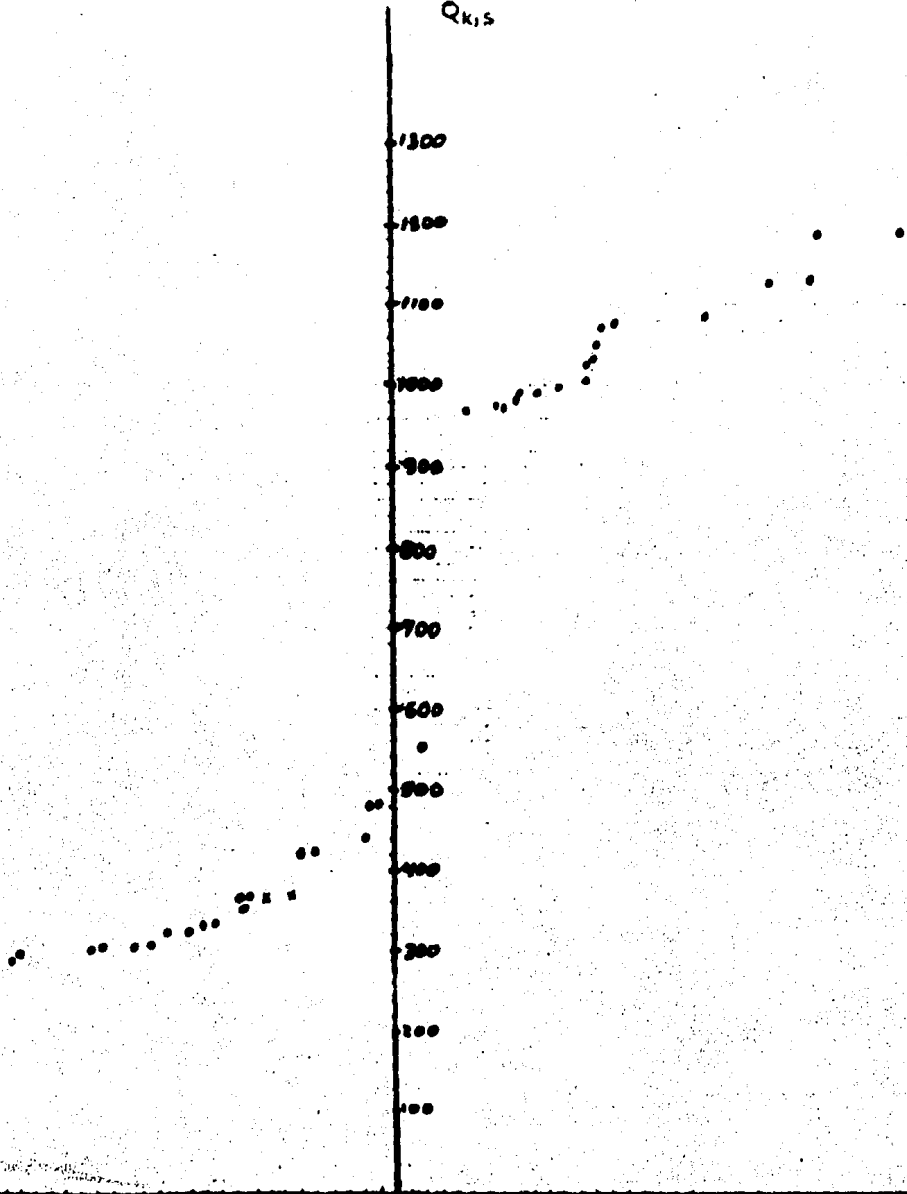
Observación omitida	$\sum_{i=1}^n \epsilon_{ui}^2$	Q_k	$\sum_{i=1}^n \epsilon_i^2$
9, 10	5898.18	1973.10	7871.28
18, 5	5920.48	1950.80	
9, 18	5939.26	1932.02	
15, 18	5969.55	1901.73	
5, 10	6001.78	1869.50	
10, 16	6011.55	1859.73	
12, 18	6011.69	1859.59	
14, 18	6017.20	1854.08	
10, 2	6019.40	1851.88	
18, 3	6028.56	1842.72	
16, 18	6040.76	1830.52	
10, 11	6041.84	1829.44	
6, 13	6070.29	1800.99	
10, 1	6073.03	1798.25	
10, 4	6079.23	1792.05	

Ejemplo #1 sin una observación

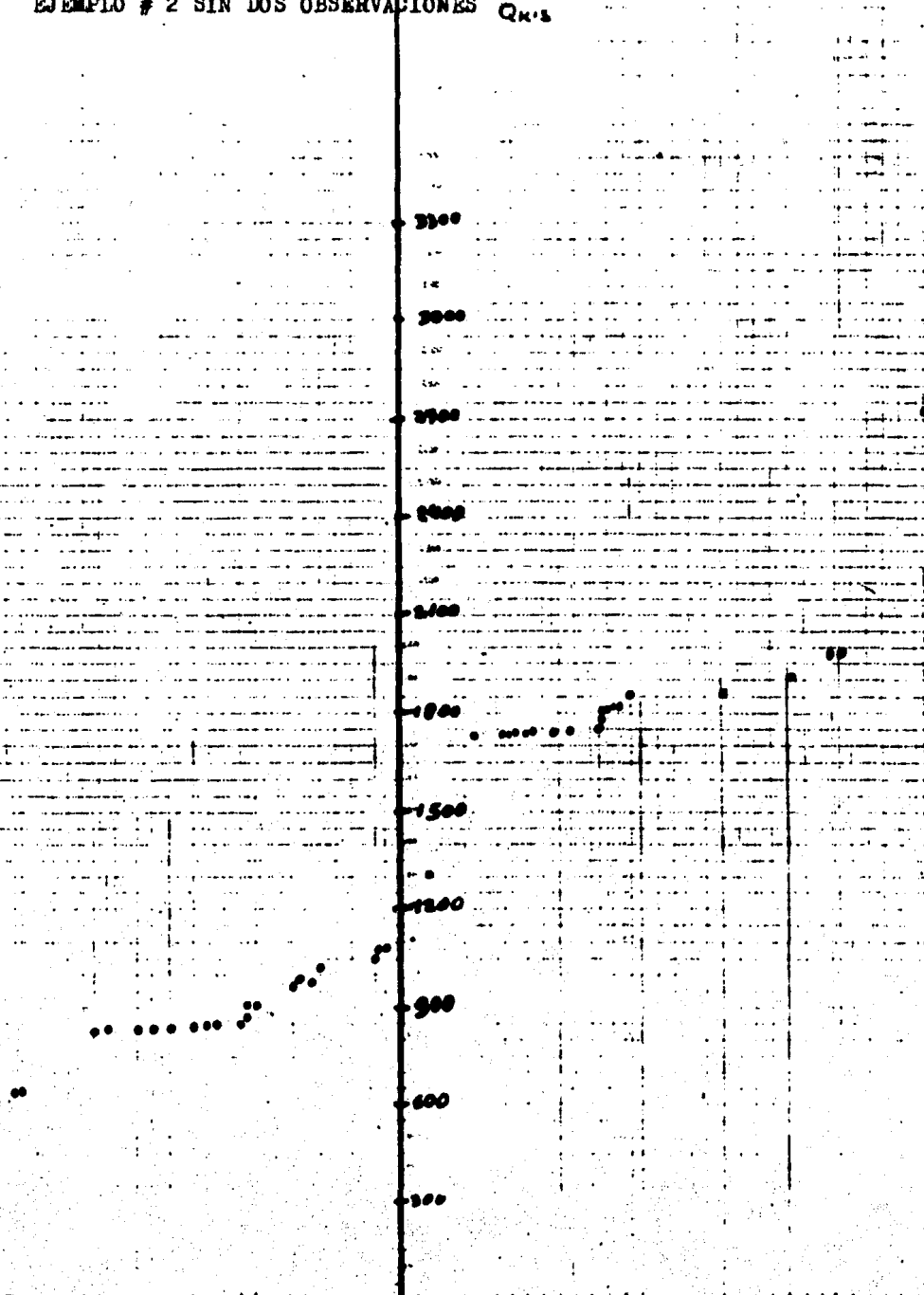


EJEMPLO # 1 SIN DOS OBSERVACIONES

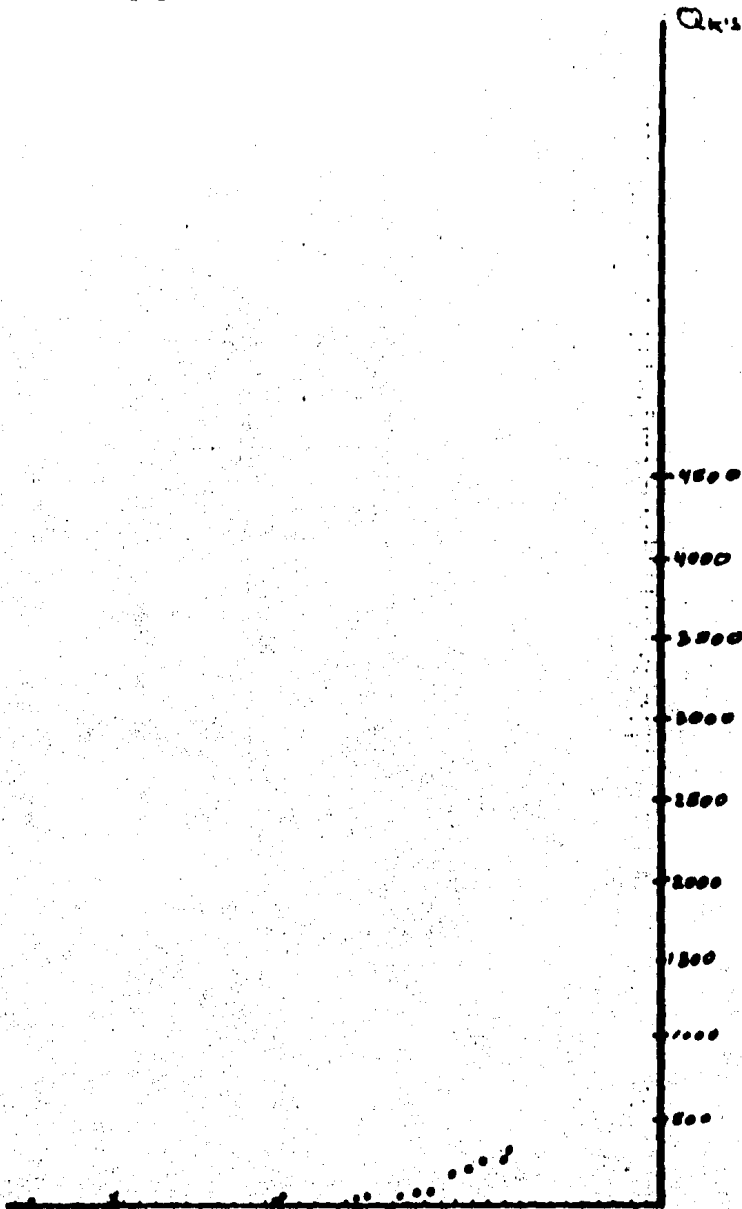
$Q_{k,5}$



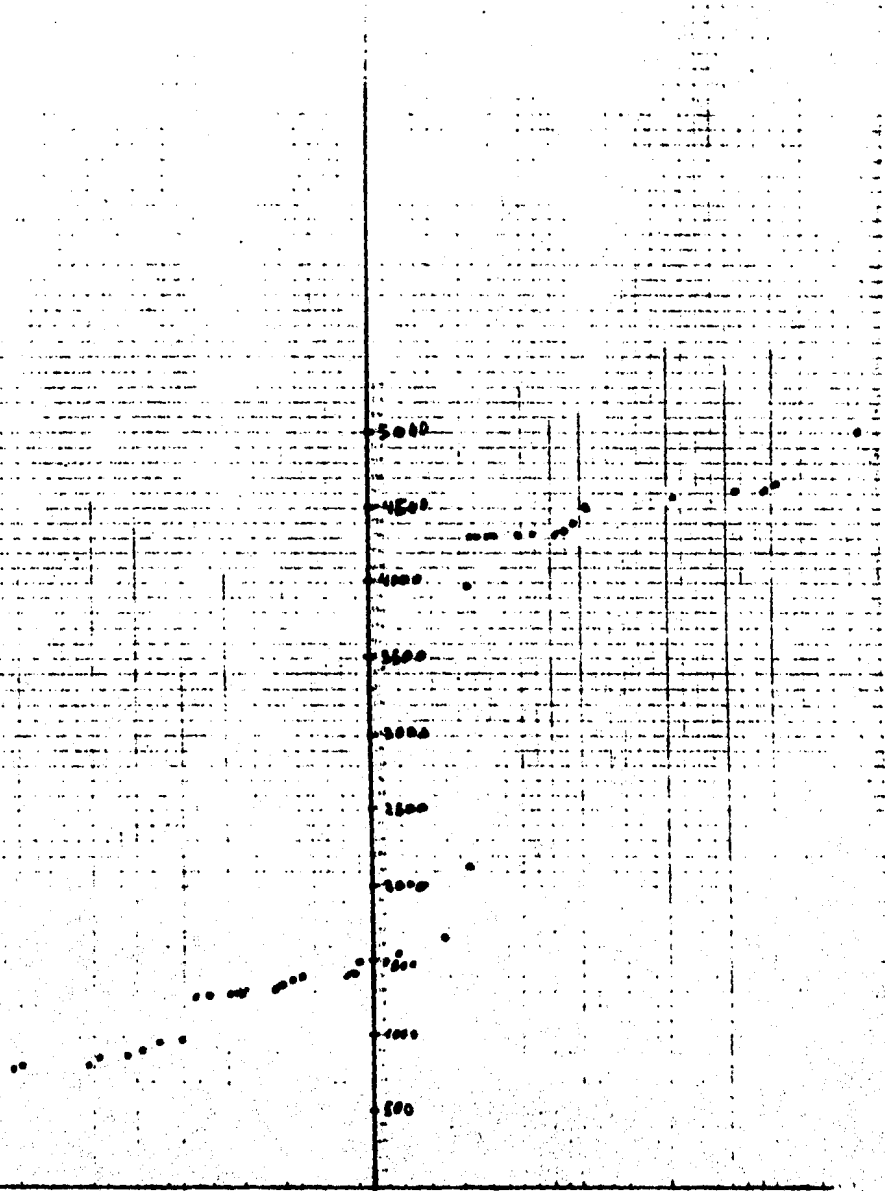
EJEMPLO # 2 SIN DOS OBSERVACIONES Q_K's



EJEMPLO # 3 SIN UNA OBSERVACION

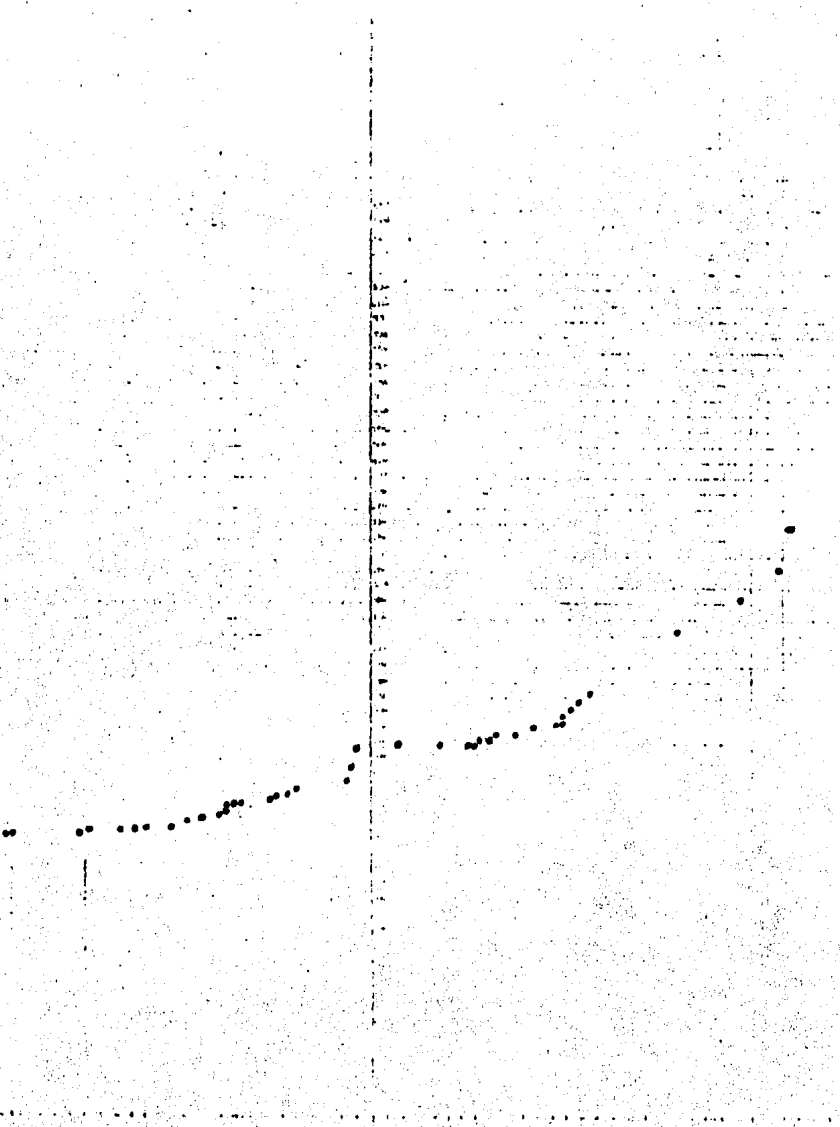


EJEMPLO # 3 SIN LAS OBSERVACIONES



EJEMPLO # 4 SIN DOS OBSERVACIONES

Q_k 's



RESULTADOS OBTENIDOS CON EL METODO DE ANDREWS Y PREGIBON

Para el ejemplo # 1, suponemos inicialmente que existen dos observaciones discordantes, es decir suponemos $k = 2$. Calculamos los valores de R_2 .

Observando la tabla para el ejemplo # 1 sin dos observaciones, para este método, nos podemos dar cuenta que el mín R_2 es el correspondiente a las observaciones (18,19), pero su valor que es $R_2 = .1364$ no difiere mucho del siguiente valor que corresponde a las observaciones (18, 2) que es de $R_2 = .1662$; por lo tanto el par de observaciones (19,18) no es discordante.

Siguiendo el método suponemos ahora que existe una sola observación discordante, es decir $k=1$, y observando la tabla para el ejemplo # 1 sin una observación nos podemos dar cuenta que el mín R_1 es el correspondiente a la observación 18 con el siguiente valor :

$$R_1 = 0.3366$$

debido a que este valor si difiere mucho de los siguientes valores la observación 18 se declara como discordante.

Para el ejemplo # 2, se supone inicialmente la existencia de dos observaciones discordantes, $k = 2$; analizando la tabla sin dos observaciones para el ejemplo # 2 podemos notar que el $\min R_2 = .1376$ correspondiente a las observaciones (18, 2) no difiere mucho del resto de los $R_{2,s}$, por lo tanto no declaramos a las observaciones (18, 2) como discordantes y suponemos entonces que solo existe 1 observación discordante; observando la tabla correspondiente al ejemplo # 2 sin una observación podemos notar que el $\min R_1$ es el correspondiente a la observación 18; $\min R_1 = .3157$ y éste valor si difiere substancialmente del resto de los $R_{1,s}$, por lo tanto se declara a la observación 18 discordante.

Para el ejemplo # 3 teniendo $k = 2$, el $\min R_2 = .1068$ correspondiente al par de observaciones (17, 6), pero como este valor no difiere mucho de los restantes $R_{2,s}$, la pareja (17, 6) no se declara discordante. Suponemos en seguida la existencia de una sola observación discordante en el conjunto de observaciones para el ejemplo # 3 y en la tabla correspondiente podemos notar que el $\min R_1 = .2542$ correspondiente a la observación # 17 y analizando si este valor difiere mucho de los restantes $R_{1,s}$, podemos observar que este valor si difiere mucho de los demás, por lo tanto la observación 17 se declara discordante.

En la tabla para el ejemplo # 4 sin dos observaciones el $\min R_2 = .1366$ valor que corresponde a las observaciones (17,18) este valor al compararlo con los restantes $R_{2,s}$ nos podemos dar cuenta que difiere mucho, por lo tanto el par de observaciones (17, 18) se declara discordante.

METODO DE ANDREWS Y PREGIBON

Tabla para el ejemplo # 1 . (sin una observación)

Observación omitida	$\sum_{i=1}^n \epsilon_{(i)}^2$	Q_k	$ I-H_{22} $	R_k
18	2220.48	88.10	0.35	0.3366
19	1340.02	968.56	0.947	0.5497
2	2200.21	108.37	0.846	0.8063
13	2048.78	259.80	0.937	0.8316
3	2048.78	259.80	0.937	0.8316
11	2175.14	133.44	0.909	0.8565
14	2116.04	192.54	0.943	0.8644
20	2168.90	139.68	0.943	0.8859
4	2226.57	82.01	0.93	0.8970
10	2260.67	47.91	0.928	0.9087
17	2229.64	78.94	0.948	0.9156
9	2297.85	10.73	0.92	0.9157
5	2222.92	85.66	0.952	0.9167
12	2293.60	14.98	0.93	0.9240
6	2308.46	0.12	0.928	0.9280
15	2286.89	21.69	0.943	0.9341
21	2306.50	2.08	0.937	0.9362
16	2306.50	2.08	0.937	0.9362
7	2296.22	12.36	0.942	0.9370
8	2301.83	6.75	0.943	0.9402
1	2304.25	4.33	0.952	0.9502

Tabla para el ejemplo # 1. (sin dos observaciones)

Observación omitida	$\sum_{i=1}^n \epsilon_{i1}^2$	Q_k	$ I - H_{22} $	R_k
18, 19	1325.83	982.75	.3206	0.1364
18, 2	1867.00	441.58	.2054	0.1662
18, 11	2031.88	276.70	.3051	0.2686
18, 13	1983.88	324.70	.3256	0.2798
18, 3	1983.88	324.70	.3256	0.2798
18, 10	2201.95	106.63	.2959	0.2822
18, 6	2209.49	99.09	.2959	0.2832
18, 14	2039.33	269.25	.3293	0.2909
18, 20	2090.57	218.01	.3293	0.2982
18, 9	2196.82	111.76	.3136	0.2984
18, 4	2157.09	151.49	.3206	0.2996
18, 7	2218.30	90.28	.313	0.3007
18, 5	2151.39	157.19	.3294	0.3070
18, 12	2212.99	95.59	.3206	0.3073
18, 17	2140.24	168.34	.3317	0.3076
18, 16	2215.45	93.13	.3256	0.3125
18, 21	2215.45	93.13	.3256	0.3125
18, 15	2194.53	114.05	.3293	0.3131
18, 8	2211.28	97.30	.3293	0.3154
18, 1	2219.29	89.29	.3294	0.3166
19, 13	1119.26	1189.32	.8857	0.4294
19, 3	1119.26	1189.32	.8858	0.4295
19, 11	1180.17	1128.41	.8597	0.4395
19, 2	1276.95	1031.63	.7961	0.4404
19, 14	1182.51	1126.07	.8914	0.4566
19, 20	1230.00	1078.58	.8914	0.4749
19, 5	1221.80	1086.78	.8991	0.4759
19, 10	1260.66	1047.92	.8753	0.4780
19, 17	1233.99	1047.59	.8959	0.4789
19, 4	1278.35	1030.23	.8794	0.4870
19, 9	1320.44	988.14	.8700	0.4977

Tabla para el ejemplo # 2. (sin una observación)

Observación omitida	$\sum \epsilon_k^2$	Q_k	$ I - H_{22} $	R_k
18	3598.95	391.46	.35	0.3157
10	2260.67	1729.74	.928	0.5257
19	3153.53	836.88	.947	0.7484
2	3785.97	204.44	.846	0.8027
3	3695.80	294.61	.937	0.8678
13	3695.80	294.61	.937	0.8678
11	3869.55	120.86	.909	0.8815
14	3762.91	227.50	.943	0.8892
20	3820.79	169.62	.943	0.9029
4	3891.60	98.81	.93	0.9070
9	3983.89	6.52	.92	0.9185
12	3967.81	22.60	.93	0.9247
6	3981.04	9.37	.928	0.9258
17	3933.93	56.48	.948	0.9346
16	3990.26	0.15	.937	0.9370
21	3990.26	0.15	.937	0.9370
5	3935.70	54.71	.952	0.9389
15	3978.52	11.89	.943	0.9402
7	3989.09	1.32	.942	0.9417
8	3988.48	1.93	.943	0.9425
1	3990.36	0.05	.952	0.9520

Tabla para el ejemplo # 2. (sin dos observaciones)

Observación omitida	$\sum_{i=1}^n \epsilon_{i1}^2$	Q_k	$ I - H_{22} $	R_k
18, 2	2672.76	1317.65	.2054	0.1376
10, 18	2201.95	1788.46	.2959	0.1633
18, 19	2935.95	1054.46	.3206	0.2359
18, 11	3364.55	625.86	.3051	0.2573
18, 6	3510.01	480.40	.2959	0.2603
18, 13	3357.11	633.30	.3256	0.2740
18, 3	3357.11	633.30	.3256	0.2740
10, 19	1260.66	2729.75	.8753	0.2765
18, 9	3564.80	425.61	.3136	0.2802
18, 4	3397.59	592.82	.3293	0.2804
18, 7	3588.00	402.41	.313	0.2814
18, 4	3541.76	448.65	.3206	0.2846
18, 20	3451.84	538.57	.3293	0.2849
18, 12	3593.49	396.92	.3206	0.2887
18, 21	3594.66	395.75	.3256	0.2934
18, 16	3594.66	395.75	.3256	0.2934
18, 17	3540.14	450.27	.3317	0.2943
18, 5	3570.58	419.83	.3294	0.2948
18, 15	3579.92	410.49	.3293	0.2954
18, 8	3593.65	396.76	.3293	0.2966
10, 2	2166.68	1823.73	.7760	0.4214
10, 13	2007.32	1983.09	.8687	0.4370
10, 3	2007.32	1983.09	.8687	0.4370
10, 11	2124.60	1865.81	.8433	0.4490
10, 14	2074.53	1915.88	.8740	0.4544
10, 20	2126.51	1863.90	.8740	0.4658
10, 4	2181.77	1808.64	.8624	0.4716
10, 5	2167.73	1822.68	.8809	0.4786
10, 17	2176.67	1813.74	.8783	0.4791
10, 9	2248.97	1741.44	.8534	0.4810
10, 6	2260.63	1729.78	.856	0.4849

Tabla para el ejemplo # 3. (sin una observación)

Observación omitida	$\sum_{i=1}^n \epsilon_{ii}^2$	Q_k	$ I - H_{22} $	R_k
17	2101.29	4312.65	.776	0.2542
6	6144.20	269.74	.54	0.5173
10	5142.61	1271.33	.836	0.6703
18	6109.97	303.97	.735	0.7002
1	6405.86	8.08	.738	0.7371
13	5702.23	711.71	.855	0.7601
3	6340.26	73.68	.774	0.7651
2	6412.60	1.34	.811	0.8108
5	6220.58	193.36	.838	0.8127
15	6348.98	64.96	.834	0.8256
12	6378.13	35.81	.853	0.8482
14	6376.91	37.03	.87	0.8650
4	6413.33	0.61	.87	0.8699
16	6327.20	86.74	.886	0.8740
9	6207.00	206.94	.92	0.8903
8	6130.77	283.17	.943	0.9014
7	6395.87	18.07	.929	0.9264
11	6404.54	9.40	.929	0.9276

Tabla para el ejemplo # 3. (sin dos observaciones)

Observación omitida	$\sum_{i=1}^n \epsilon_{ii}^2$	Q_k	$ I - H_{22} $	R_k
17, 6	1651.04	4762.90	.4150	0.1068
17, 10	1335.33	5078.61	.6340	0.1320
17, 18	2101.08	4312.86	.5192	0.1701
17, 5	1800.77	4613.17	.6490	0.1822
17, 13	1878.33	4535.61	.6332	0.1854
17, 1	2089.85	4324.09	.5723	0.1865
17, 8	1783.02	4630.92	.6954	0.1933
17, 3	2092.19	4321.75	.5969	0.1947
17, 9	1851.91	4562.03	.6758	0.1951
17, 15	2071.35	4342.59	.6107	0.1972
17, 2	2076.90	4337.04	.6245	0.2022
17, 12	2073.56	4340.38	.6356	0.2055
17, 16	1966.68	4447.26	.6871	0.2107
17, 14	2086.69	4327.25	.6540	0.2128
17, 4	2094.29	4319.65	.6738	0.2200
17, 7	2044.32	4369.62	.7176	0.2287
17, 11	2101.10	4312.84	.7186	0.2354
1, 6	5957.81	456.13	.2862	0.2659
10, 6	5123.54	1290.40	.3933	0.3142
18, 6	5763.50	650.44	.3935	0.3536
13, 6	5434.90	979.04	.4614	0.3910
5, 6	5904.09	509.85	.4489	0.4132
2, 6	6143.69	270.25	.4376	0.4192
15, 6	6103.97	309.97	.4429	0.4216
3, 6	6106.86	307.08	.4501	0.4286
9, 6	5994.07	419.87	.4633	0.4330
12, 6	6094.05	319.89	.4595	0.4367
4, 6	6132.39	281.55	.4579	0.4378
16, 6	6099.76	314.18	.4649	0.4422
8, 6	5898.65	515.29	.4823	0.4436
14, 6	6091.38	322.56	.4685	0.4449

Tabla para el ejemplo # 4. (sin dos observaciones)

Observación omitida	$\sum_{i=1}^n \epsilon_{ii}^2$	Q_k	$ I - H_{22} $	R_k
17, 18	2101.08	5770.20	.5116	0.1366
16, 1	7389.50	481.78	.2716	0.2550
18, 6	5763.50	2107.78	.3776	0.2765
17, 6	5233.53	2637.75	.4257	0.2831
10, 6	6084.82	1786.46	.4038	0.3122
10, 17	4083.01	3788.27	.6698	0.3475
13, 6	6070.29	1800.99	.4604	0.3551
10, 18	4757.64	3113.64	.5918	0.3577
13, 18	5150.70	2720.58	.5811	0.3803
13, 17	4543.14	3328.14	.6599	0.3809
3, 6	7494.44	376.84	.4116	0.3920
10, 13	4359.14	3512.14	.7240	0.4010
1, 18	6087.81	1783.47	.5196	0.4019
1, 17	5482.79	2388.49	.5809	0.4047
6, 15	7224.00	647.28	.4522	0.4151
6, 12	7256.64	614.64	.4514	0.4162
2, 6	7580.64	290.64	.4386	0.4225
3, 18	6028.56	1842.72	.5523	0.4230
3, 17	5369.87	2501.41	.6240	0.4257
6, 14	7303.51	567.77	.4592	0.4261
5, 6	7499.05	372.23	.4485	0.4273
4, 6	7571.85	299.43	.4515	0.4344
6, 9	7526.22	345.06	.4551	0.4352
15, 17	5344.35	2526.93	.6486	0.4404
12, 18	6011.69	1859.59	.5788	0.4421
6, 16	7626.27	245.01	.4583	0.4441
2, 18	6104.61	1766.67	.5758	0.4466
12, 17	5404.78	2466.50	.6542	0.4492
6, 8	7424.26	447.02	.4767	0.4497
5, 18	5920.48	1950.80	.5986	0.4503

RESULTADOS PARA EL METODO DE STEPWISE

Para los cuatro ejemplos se usó un nivel de significancia $\alpha = .1$.

Para el ejemplo # 1 tenemos que para la observación 19:

$$F_{\text{calc}} = 13.01 > F_{\text{tablas}} = 10.3$$

Para el resto de las observaciones omitidas sucede que:

$$F_{\text{calc}} < F_{\text{tablas}}$$

por lo que solo la observación 19 es declarada discordante.

Para el ejemplo # 2 tenemos que :

$$F_{\text{calc}} = 13.48 > 10.45 = F_{\text{tablas}}$$

$$F_{\text{calc}} = 13.77 > 10.3 = F_{\text{tablas}}$$

para las observaciones 19 y 10 respectivamente.

Para el resto de las observaciones :

$$F_{\text{calc}} < F_{\text{tablas}}$$

de lo anterior concluimos que las observaciones 19 y 10 son declaradas discordantes.

Con el ejemplo # 3 sucede que solo para la observación 17 se tiene :

$$F_{\text{calc}} = 28.73 > 11.13 = F_{\text{tablas}}$$

Y para el resto de las observaciones :

$$F_{\text{calc}} < F_{\text{tablas}}$$

por lo que en este caso solo la observación 17 es declarada discordante.

Y por último para el ejemplo # 4, solo para la observación 18 tenemos que :

$$F_{\text{calc}} = 21.01 > 11.46 = F_{\text{tablas}}$$

Y para el resto de las observaciones :

$$F_{\text{calc}} < F_{\text{tablas}}$$

Por lo cual concluimos que solo la observación 18 es discordante.

S T E P W I S E

Tabla para el ejemplo # 1 .

Suma del cuadrado de los residuales	Observación omitida.	Grados de libertad.	F tablas	F calculada.
2308.58	NINGUNA	19		
1340.02	19	18	10.3	13.01
1119.26	19,3	17	10.45	3.35
865.9	19,3,13	16	10.62	4.68
650.1	19,3,13,14	15	10.8	4.98
462.1	19,3,13,14,20	14	11.13	5.70

$$F_{\text{calc}} = \frac{(2308.58 - 1340.02)}{1340.02/18} = \frac{968.56}{74.44} = 13.01$$

$$F_{\text{calc}} = \frac{(1340.02 - 1119.26)}{1119.26/17} = \frac{220.76}{65.83} = 3.35$$

$$F_{\text{calc}} = \frac{(1119.26 - 865.9)}{865.9/16} = \frac{253.36}{54.11} = 4.68$$

$$F_{\text{calc}} = \frac{(865.9 - 650.1)}{650.1/15} = \frac{215.8}{43.34} = 4.98$$

$$F_{\text{calc}} = \frac{(650.1 - 462.1)}{462.1/14} = \frac{188}{33} = 5.70$$

Tabla para el ejemplo # 2 .

Suma del cuadrado de los residuales	Observación omitida.	Grados de libertad.	F tablas	F calculada.
3990.41	NINGUNA	19		
2260.67	10	18	10.3	13.77
1260.66	10,19	17	10.45	13.48
1048.29	10,19,3	16	10.62	3.24
803.96	10,19,3,13	15	10.8	4.55
597.47	10,19,3,13,14	14	11.13	4.83
418.11	10,19,3,13,14 20	13	11.46	5.57
264.86	10,19,3,13,14 20,4	12	11.8	6.94

$$F_{\text{calc}} = \frac{(3990.41 - 2260.67)}{2260.67/18} = \frac{1729.74}{125.59} = 13.77$$

$$F_{\text{calc}} = \frac{(2260.67 - 1260.66)}{1260.66/17} = \frac{1000.01}{74.156} = 13.48$$

$$F_{\text{calc}} = \frac{(1260.66 - 1048.29)}{1048.29/16} = \frac{212.37}{65.518} = 3.24$$

$$F_{\text{calc}} = \frac{(1048 - 803.96)}{803.96/15} = \frac{244.04}{53.59} = 4.55$$

$$F_{\text{calc}} = \frac{(803.96 - 597.47)}{597.47/14} = \frac{206.49}{42.67} = 4.83$$

$$F_{\text{calc}} = \frac{(597.47 - 418.11)}{418.11/13} = \frac{179.36}{32.16} = 5.57$$

$$F_{\text{calc}} = \frac{(418.11 - 264.86)}{264.86/12} = \frac{153.25}{22.07} = 6.94$$

Tabla para el ejemplo # 3 .

Suma del cuadrado de los residuales	Observación omitida.	Grados de libertad.	F tablas	F calculada.
6413.94	NINGUNA	15	10.8	
2101.29	17	14	11.13	28.73
1335.33	10,17	13	11.46	7.45
1010.65	10,17,8	12	11.8	3.85
654.85	10,17,8,9	11	12.3	5.97
404.65	10,17,8,9,13	10	12.8	6.18
242.16	10,17,8,9,13 3	9	13.6	6.03

$$F_{\text{calc}} = \frac{(6413.94 - 2101.29)}{2101.29/14} = \frac{4312.65}{150.09} = 28.73$$

$$F_{\text{calc}} = \frac{(2101.29 - 1335.33)}{1335.33/13} = \frac{765.96}{102.71} = 7.45$$

$$F_{\text{calc}} = \frac{(1335.33 - 1010.65)}{1010.65/12} = \frac{324.68}{84.2208} = 3.85$$

$$F_{\text{calc}} = \frac{(1010.65 - 654.85)}{654.85/11} = \frac{355.8}{59.53} = 5.97$$

$$F_{\text{calc}} = \frac{(654.85 - 404.65)}{404.65/10} = \frac{250.2}{40.46} = 6.18$$

$$F_{\text{calc}} = \frac{(404.65 - 242.16)}{242.16/9} = \frac{162.49}{26.9} = 6.03$$

Tabla para el ejemplo # 4 .

Suma del cuadrado de los residuales	Observación omitida.	Grados de libertad.	F tablas	F calculada.
7871.28	NINGUNA	15	10.8	
5497.51	17	14	11.13	6.04
2101.08	17,18	13	11.46	21.01
1325.47	17,18,10	12	11.8	7.02
944.79	17,18,10,13	11	12.3	4.43
670.21	17,18,10,13 8	10	12.8	4.09
378.07	17,18,10,13 8,9	9	13.6	6.95

$$F_{\text{calc}} = \frac{(7871.28 - 5497.51)}{5497.51/14} = \frac{2373.77}{392.67} = 6.04$$

$$F_{\text{calc}} = \frac{(5497.51 - 2101.08)}{2101.08/13} = \frac{3396.43}{161.62} = 21.01$$

$$F_{\text{calc}} = \frac{(2101.08 - 1325.47)}{1325.47/12} = \frac{775.61}{110.45} = 7.02$$

$$F_{\text{calc}} = \frac{(1325.47 - 944.79)}{944.79/11} = \frac{380.68}{85.89} = 4.43$$

$$F_{\text{calc}} = \frac{(944.79 - 670.21)}{670.21/10} = \frac{274.58}{67.02} = 4.09$$

$$F_{\text{calc}} = \frac{(670.21 - 378.07)}{378.07/9} = \frac{292.14}{42.01} = 6.95$$

C O N C L U S I O N E S

En esta parte estableceremos una serie de comparaciones entre los 10 métodos propuestos, en base a los resultados obtenidos para cada uno de los ejemplos.

En la tabla (I) podemos observar los métodos y los resultados obtenidos con ellos. Podemos darnos cuenta que los métodos que solo toman en cuenta la discordancia sin poner énfasis en la influencia para detectar una sola observación discordante, detectan exactamente las mismas observaciones en los cuatro ejemplos, algunos comentarios se podrían hacer al respecto:

- Para los ejemplos 1 y 3 los métodos detectan correctamente y esto es debido a que sólo existe una observación discordante.
 - En el ejemplo 2 los métodos detectan solo una observación discordante.
 - En el ejemplo 4 los métodos no son capaces de detectar observaciones discordantes, esto es debido a que al existir más de una observación discordante y tratar de detectar solamente una, entre ellas pueden disimular su existencia teniéndose como resultado que los métodos para detectar una sola observación discordante no detecten nada.
- Debido a esto es recomendable que, cuando se sospeche la existencia de más de una observación discordante se empleen

métodos para detectar más de una observación discordante a la vez.

- Entre las ventajas de estos métodos se encuentran la sencillez y rapidez para llevarlos a cabo.

Cuando se sospeche la existencia de una sola observación discordante se recomienda usar el método de LUND por ser el más sencillo.

Entre los métodos para detectar una sola observación discordante se encuentra el de COOK el cual es distinto a los anteriores debido a que toma en cuenta fundamentalmente la influencia de las observaciones, además de que el método está basado en encontrar un nivel descriptivo de significancia. Comparando en base a los resultados obtenidos podemos darnos cuenta de lo siguiente :

Para el ejemplo 1 detectó a la observación 18 que es muy influyente aún cuando no es la más discordante; para el ejemplo 2 ocurre lo mismo. Para el ejemplo 3 si detecta correctamente debido a que la observación 17 si es discordante pero además muy influyente y por último para el ejemplo 4 detecta a la observación 18 que es discordante pero muy influyente.

De los métodos para detectar más de una observación discordante a la vez, podemos decir lo siguiente :

- No coinciden en los resultados obtenidos.

- El método de Andrews y Pregibon toma en cuenta la influencia obteniendo que para el ejemplo 1 detecta a la observación más influyente; lo mismo ocurre para el ejemplo 2, en el

cual solo detecta una observación. En los ejemplos 3 y 4 detecta correctamente a las observaciones.

Entre las desventajas de este método tenemos :

- i) Es subjetivo.
- ii) Se requieren de muchos cálculos para poder llevarlo a cabo.

-- Consideramos que el método de VALENCIA es muy exigente y debido a esto en muchos casos no detecta; como en los ejemplos 1 y 2; aunque en los ejemplos 3 y 4 si detecto bien debido a la influencia de las observaciones detectadas.

La principal desventaja de este método es la gran cantidad de cálculos necesarios para llevarlo a cabo.

-- Con el método GRAFICO no se puede esperar que siempre se tengan buenos resultados debido a la subjetividad para interpretar las gráficas.

-- Con el método de GENTLEMAN y WILK, a pesar de tener cierto grado de subjetividad, detectó correctamente en los cuatro ejemplos.

La principal desventaja de este método es la gran cantidad cálculos y la dificultad de la graficación, que son necesarios.

-- El método de STEPWISE sólo detecto de manera incompleta en el ejemplo 4 y esto podría deberse a que después de detectar la primera observación discordante la suma del cuadrado de residuales no sufre una reducción sensible al omitir

cualquier otra observación.

Entre los métodos para detectar varias observaciones discordantes a la vez se recomienda usar el de STEPWISE debido a su gran rapidez y sencillez en comparación con los otros.

TABLA (I)

METODOS	EJEM. 1	EJEM. 2	EJEM. 3	EJEM. 4
LUND	19	10	17	No detecta
ELLENBERG	19	10	17	No detecta
COOK Y PRESCOTT	19	10	17	No detecta
PRESCOTT	19	10	17	No detecta
ANDREWS Y PREGI- BON	18	18	17	(17,18)
VALENCIA	No detecta	No detecta	17	(17,18)
GRAFICO	No detecta	10	(10,17)	(17,18)
GENTLEMAN. Y WILK	19	(10,19)	17	(17,18)
STEPWISE	19	(10,19)	17	18
COOK	18	18	17	18

TABLA DE RESULTADOS

B I B L I O G R A F I A

- 1) .- Andrews, D.F. & Pregibon, D. (1978) : "Finding the outliers that matter". J.R. Statist. Soc. B 40, 85-93
- 2) .- Anscombe, F.J. (1960) : "Rejection of outliers". Technometrics, 2, p. 123-148.
- 3) .- Anscombe, F.J. and Tukey, J.W. (1963) : "The examination and Analysis of Residuals". Technometrics, 5. p. 141-160.
- 4) .- Atkinson, A.C. : "Two graphical displays for outlying and influential observations in Regression". Biometrika (1981), 68, 1. p. 13-20.
- 5) .- Barnett, V. (1978) : "The study of outliers : Purpose and Model". Appl. Statist. 27, 3, p. 242-250.
- 5.1) .- Barnett, V. and Lewis, T. (1978) : "Outliers in Statistical Data". Chichester : Wiley.
- 6) .- Beckman, R.J. and Trusell, H.J. (1974) : "The distribution of an arbitrary studentized residual and the effects of updating in multiple regression". J. Amer. Statist. Ass 69, p. 199-201.
- 6.1) .- Behnken, D.W. and Draper, N.R. (1972) : "Residuals and their variance patterns". Technometrics, 14, p. 102-111.
- 7) .- Collett, D. and Lewis, T. (1976) : "The subjective nature of outlier rejection procedures". App. Stat. 25, p. 228-237.
- (8) .- Cook, R.D. (1977) : "Detection of influential observations in linear regression". Technometrics, 19,15-8
- (9) .- Cook, R.D. (1979) : "Influential observations in linear regression". J. Am. Statist. Assoc., 74, p. 169-74.