Capítulo 3

Intervalos de confianza para los coeficientes de regresión y la varianza σ^2

Conceptos importantes que debemos conocer como **probabilidad**, **distribuciones de probabilidad, errores tipo I y tipo II, nivel de significancia, potencia de una prueba estadística** e **intervalos de confianza**

Espacio muestral: El conjunto de todos los resultados posibles de un experimento aleatorio, o al azar.

Punto muestral es cada miembro de este espacio muestral

Suceso es un subconjunto del espacio muestral y pueden ser mutuamente excluyentes si la ocurrencia de uno impide la ocurrencia de otro, **exhaustivos** (colectivamente) si se agotan todos los resultados posibles de un experimento. Así, en el ejemplo de dos monedas, los sucesos *a*) dos caras, *b*) dos cruces y *c*) una cruz y una cara agotan todos los resultados posibles; por tanto, son sucesos exhaustivos (colectivamente).

Probabilidad

Sea A un suceso en un espacio muestral. Sea $P(A)$ la probabilidad del suceso A, es decir, la proporción de veces que el suceso A ocurrirá en ensayos repetidos de un experimento. En forma alterna, en un total de n posibles resultados igualmente probables de un experimento, si m de ellos son favorables a la ocurrencia del suceso A, se define la razón m/n como la **frecuencia relativa** de A. Para valores grandes de n, esta frecuencia relativa constituye una muy buena aproximación de la probabilidad de A.

Propiedades de la probabilidad

$P(A)$ es una función de valor real y tiene estas propiedades:

1. $0 \leq P(A) \leq 1$ para todo A.

2. Si A, B, C, … constituye un conjunto de sucesos exhaustivo, entonces $P(A + B + C + \cdots) = 1$, donde $A + B + C$ significa A o B o C, y así sucesivamente.

3. Si A, B, C, … son sucesos mutuamente excluyentes, entonces $P(A + B + C + \cdots) = P(A) + P(B) + P(C) + \cdots$

Variables aleatorias

Una variable cuyo valor está determinado por el resultado de un experimento al azar se denomina **variable aleatoria** (va). Las variables aleatorias se denotan usualmente por las letras mayúsculas X, Y, Z, y así sucesivamente, y los valores que toman se denotan por letras minúsculas, x, y, z, etcétera.

Una variable aleatoria puede ser **discreta** o **continua**

. Una discreta adquiere sólo un número finito (o infinito contable) de valores. Por ejemplo, al lanzar dos dados, cada uno numerado del 1 al 6, si definimos la variable aleatoria X como la suma de los números que aparecen en los dados, entonces X toma uno de los siguientes valores: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 o 12; por tanto, se trata de una variable aleatoria discreta. Una continua, por su parte, es una variable que puede tomar cualquier valor dentro de un intervalo de valores. Así, la estatura de un individuo es una variable continua —por ejemplo, en el intervalo entre 152.4 y 165.1 centímetros— y puede adquirir cualquier valor, según la precisión de la medición.

Función de densidad de probabilidad de una variable aleatoria discreta

Sea X una variable discreta que toma valores diferentes $x1, x2, \ldots xn$. Entonces, la función

$f(x) = P(X = xi)$ para $i = 1, 2, \ldots, n, \ldots$

$= 0$ para $x \neq xi$

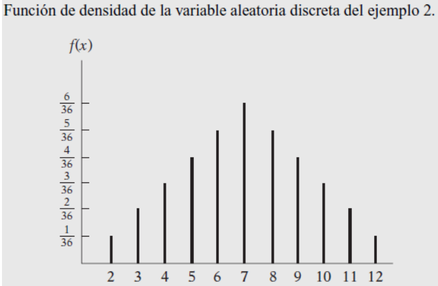
Se denomina **función de densidad de probabilidad discreta** (FDP) de X, donde $P(X = xi)$ significa la probabilidad de que la variable discreta X tome el valor de xi.

En un lanzamiento de dos dados la variable aleatoria X, la suma de los números que aparecen en dos dados, puede tomar uno de los 11 valores mostrados. La FDP de esta variable se muestra como sigue (véase también la figura A.1):

$$x = \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12$$
$$f(x) = \left(\tfrac{1}{36}\right)\left(\tfrac{2}{36}\right)\left(\tfrac{3}{36}\right)\left(\tfrac{4}{36}\right)\left(\tfrac{5}{36}\right)\left(\tfrac{6}{36}\right)\left(\tfrac{5}{36}\right)\left(\tfrac{4}{36}\right)\left(\tfrac{3}{36}\right)\left(\tfrac{2}{36}\right)\left(\tfrac{1}{36}\right)$$

Estas probabilidades se verifican fácilmente. En total, hay 36 resultados posibles, de los cuales uno es favorable al número 2, dos son favorables al número 3 (pues la suma de 3 se presenta como 1 en el primer dado y 2 en el segundo dado, o 2 en el primer dado y 1 en el segundo dado), y así sucesivamente.

Función de densidad de la variable aleatoria discreta del ejemplo 2.



Función de densidad de probabilidad de una variable aleatoria continua

Sea X una variable continua. Entonces, se dice que $f(x)$ es la FDP de X si se satisfacen las siguientes condiciones:

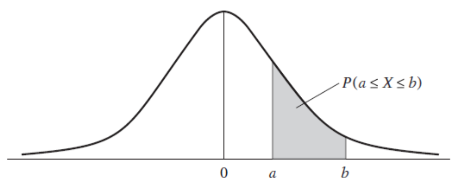
$$f(x) \geq 0$$
$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$
$$\int_{a}^{b} f(x)\,dx = P(a \leq x \leq b)$$

Donde $f(x)dx$ se conoce como el elemento probabilístico (la probabilidad asociada a un pequeño intervalo de una variable continua) y donde $P(a \leq X \leq b)$ significa la probabilidad de que X se encuentre en el intervalo a a b. Geométricamente, tenemos. Para una variable continua, en contraste con una discreta, la probabilidad de que X tome un valor específico es cero; la probabilidad para tal variable sólo se mide sobre un rango o intervalo dado, como (a, b).

Considere la siguiente función de densidad:

$$f(x) = \frac{1}{9}x^2 \qquad 0 \le x \le 3$$

Se verifica con facilidad que $f(x) \ge 0$ para toda X en el rango 0 a 3 y que $\int_0^3 \frac{1}{9}x^2 dx = 1$. (*Nota:* La integral es $(\frac{1}{27}x^3 |_0^3) = 1$.) Si deseamos evaluar la FDP anterior entre 0 y 1, obtenemos $\int_0^1 \frac{1}{9}x^2 dx = (\frac{1}{27}x^3 |_0^1) = \frac{1}{27}$; es decir, la probabilidad de que X se encuentre entre 0 y 1 es de 1/27.



La estimación y las pruebas de hipótesis constituyen las dos ramas principales de la estadística clásica. La teoría de la estimación consta de dos partes: estimación puntual y estimación por intervalos. En los dos capítulos anteriores estudiamos a fondo la estimación puntual, en donde se introdujeron los métodos MCO y MV de la estimación puntual. En este capítulo consideraremos primero la estimación por intervalos y luego trataremos el tema de las pruebas de hipótesis, muy relacionado con la estimación por intervalos.

Para poner en orden las ideas, consideremos el ejemplo de los salarios y el nivel de escolaridad. La ecuación muestra que el incremento promedio estimado del salario medio por hora relacionado con un año de aumento en la escolaridad ($\hat{\beta}_2$) es de 0.73, que constituye una cifra estimada (puntual) del valor poblacional desconocido β_2.

Debido a las fluctuaciones muéstrales, es probable que una sola estimación difiera del valor verdadero, aunque en un muestreo repetido se espera que el promedio de los valores sea igual al valor verdadero.

En estadística, la confiabilidad de un estimador puntual se mide por su error estándar. Por tanto, en lugar de depender de un solo estimador puntual, se puede construir un intervalo alrededor del estimador puntual.

Por ejemplo, dentro de dos o tres errores estándar a cada lado del estimador puntual, tal que este intervalo tenga, por ejemplo, 95% de probabilidad de incluir al verdadero valor del parámetro. Ésta es, a grandes rasgos, la idea básica de la **estimación por intervalos**.

Para ser más específico, supongamos que se desea encontrar qué tan "cerca" está, por ejemplo,

$\hat{\beta}_2$ de β_2. Con este fin, se trata de encontrar dos números positivos, δ y α, este último situado entre 0 y 1, de modo que la probabilidad de que el **intervalo aleatorio** ($\hat{\beta}_2 - \delta$, $\hat{\beta}_2 + \delta$) contenga al verdadero β_2 sea $1 - \alpha$. Simbólicamente

$$\Pr(\hat{\beta}_2 - \delta \le \beta_2 \le \hat{\beta}_2 + \delta) = 1 - \alpha$$

Tal intervalo, se conoce como **intervalo de confianza**

A $1 - \alpha$ se le denomina **coeficiente de confianza.**

α **se conoce como nivel de significancia** (conocida como probabilidad de cometer un error tipo I. Un error tipo I consiste en rechazar una hipótesis verdadera, mientras que el error tipo II consiste en aceptar una hipótesis falsa).

Los extremos del intervalo de confianza se conocen como **límites de confianza** (también denominados valores *críticos*), con $\hat{\beta_2} - \delta$ como *límite* de **confianza inferior** y $\hat{\beta_2} + \delta$ como *límite* de **confianza superior.**

α y $1 - \alpha$ suelen expresarse en forma porcentual

La ecuación muestra que un **estimador de intervalo**, en contraste con un estimador puntual, es un intervalo construido de manera que tenga una probabilidad específica $1 - \alpha$ de contener en sus límites al valor verdadero del parámetro.

si α = 0.05 entonces la probabilidad de que el intervalo (aleatorio) incluya al verdadero β_2 es de 95%.

Una vez que se tenga una muestra específica y se obtenga un valor numérico específico de $\hat{\beta_2}$, el intervalo deja de ser aleatorio, y queda entonces fijo. No se puede afirmar que la probabilidad de que un intervalo *fijo* dado incluya al verdadero β_2 sea $1 - \alpha$.

En esta situación, β_2 está en el intervalo fijo o fuera de él. Por consiguiente, la probabilidad será 1 o 0. Ejemplo de salarios y nivel de escolaridad, si el intervalo de confianza a 95% se obtuviera como ($0.5700 \leq \beta_2 \leq 0.8780$), **no se puede** afirmar que la probabilidad de que este intervalo incluya al verdadero β_2 sea de 95%. Esa probabilidad es 1 o 0.

¿Cómo se construyen los intervalos de confianza? De la exposición anterior se espera que si se conocen las **distribuciones muéstrales o de probabilidad** de los estimadores, se puedan hacer afirmaciones sobre intervalos de confianza

Con el supuesto de normalidad para ui, los estimadores de MCO tienden en sí mismos a una distribución normal con medias y varianzas dadas, por ejemplo, la variable Z = $\frac{\beta^\wedge 2 - \beta 2}{se(\beta^\wedge{}_2)}$ que es una variable normal estandarizada. Ahora podemos utilizar la distribución normal para hacer afirmaciones probabilísticas sobre β_2. Si se conoce la varianza poblacional (σ^2), una propiedad importante de una variable con distribución normal, con media u y una varianza σ^2, es que el área bajo la curva entre u ± σ dicha área es del 68 %; los límites entre u ± 2σ es del 95% y entre u± 3σ es de 99,7% aproximadamente.

Aunque muy pocas veces se conoce σ^2, en la práctica se utiliza el estimador insesgado $\sigma^{\wedge 2}$. Al remplazar σ por σ^, se puede reescribir la ecuación

$$t = \frac{\beta^\wedge 2 - \beta 2}{se(\beta^\wedge{}_2)} = \frac{estimador - parámetro}{error\ estándar\ del\ estimador}$$

La variable t tiene una distribución con N – 2 g de l. Por lo tanto, en lugar de utilizar la distribución normal podemos usar la distribución t para establecer intervalos de confianza para β_2

$$\Pr(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) = 1 - \alpha$$

Donde t en medio de esta doble igualdad es el valor dado por la ecuación, y $t_{\alpha/2}$ es el valor de la variable obtenida de la tabla, a menudo se denomina el valor **crítico** *t* a un nivel de significancia $\alpha/2$. Reemplazando t de la ecuación tenemos

$$\Pr\left[-t_{\alpha/2} \leq \frac{\hat{\beta}_2 - \beta_2}{\text{ee}(\hat{\beta}_2)} \leq t_{\alpha/2}\right] = 1 - \alpha$$

Pr ($\beta\hat{}_2$ - $t_{\alpha/2}$ se($\beta\hat{}_2$) ≤ β_2 ≤ $\beta\hat{}_2$ + $t_{\alpha/2}$ se($\beta\hat{}_2$)) = 1 - α

La ecuación proporciona el intervalo de confianza del 100(1 – α)% para β_2

Se puede escribir

$\beta\hat{}_2 \pm t_{\alpha/2}$ se($\beta\hat{}_2$)

Para β_1

$\beta\hat{}_1 \pm t_{\alpha/2}$ se($\beta\hat{}_1$)

Observe un rasgo importante de los intervalos de confianza dados en ambos casos *la amplitud del intervalo de confianza es proporcional al error estándar del estimador*. Es decir, entre más grande sea el error estándar, más amplio será el intervalo de confianza

De vuelta al ejemplo de regresión del salario promedio por hora (*Y*) y el nivel de escolaridad (*X*), recuerde $\hat{\beta_2}$ = 0.7240; ee ($\hat{\beta_2}$) = 0.0700. Como hay 13 observaciones, los grados de libertad (gl) son 11. Si suponemos que α = 5%, es decir, un coeficiente de confianza a 95%, entonces la tabla *t* muestra que para 11 gl el valor **crítico** $t_{\alpha/2}$ = 2.201. Al sustituir estos valores en, el intervalo de confianza para β_2 a 95% sea el siguiente:

0.5700 ≤ β_2 ≤ 0.8780

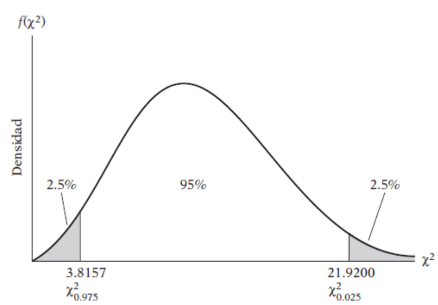
0.7240 $\pm$ 2.201(0.0700)

0.7240 ± 0.1540

La interpretación de este intervalo de confianza es: Dado el coeficiente de confianza de 95%, en 95 de cada 100 casos, los intervalos contendrán al verdadero valor de β_2. Pero, como ya advertimos, no se puede afirmar que la probabilidad de que el intervalo específico de la ecuación contenga al verdadero β_2 sea de 95%, porque este intervalo es ahora fijo y no aleatorio; por consiguiente, β_2 se encontrará o no dentro de él: la probabilidad de que el intervalo fijo específico incluya al verdadero valor de β_2 es por consiguiente 1 o 0.

Intervalo de confianza para σ^2 Como señalamos anteriormente, según el supuesto de normalidad, la variable

X^2 = (N – 2)$\frac{\sigma^{\wedge 2}}{\sigma^2}$ posee una distribución X^2 con N – 2 g de l, por lo tanto podemos utilizar la distribución X^2 para establecer intervalos de confianza para σ^2



Pr ($X^2_{1-\alpha/2} \leq X^2 \leq X^2_{\alpha/2}$) = 1 – α

En donde el valor X^2 en medio de la desigualdad está dado por la ecuación y $X^2_{1-\alpha/2}$ y $X^2_{\alpha/2}$ corresponden a los valores críticos obtenidos de la tabla para N – 2 g de l.

Reemplazando X^2 en la ecuación

Pr ($(N-2)\frac{\hat{\sigma}^2}{X^2_{\alpha/2}} \leq \sigma^2 \leq (N-2)\frac{\hat{\sigma}^2}{X^2_{1-\alpha/2}}$) = 1 – α

Continuamos con el ejemplo de salarios y nivel de escolaridad: se encontró en la tabla que para los datos se tiene $\hat{\sigma}^2$= 0.8936. Si seleccionamos α de 5%, la tabla ji cuadrada para 11 gl da los siguientes valores críticos: $\chi^2_{0.025}$ = 21.9200 y $\chi^2_{0.975}$ = 3.8157. Estos valores muestran que la probabilidad de que un valor ji cuadrada sea superior a 21.9200 es 2.5%, y la de 3.8157 es 97.5%. Por consiguiente, el intervalo entre estos dos valores es el intervalo de confianza para χ2 a 95%.

Al sustituir los datos del ejemplo, el intervalo de confianza para σ^2 a 95% es el siguiente:

0.4484 ≤ σ^2 ≤ 2.5760

La interpretación de este intervalo es la siguiente: Si establecemos límites de confianza a 95% sobre σ^2 y afirma *a priori* que entre estos límites caerá el verdadero σ^2, acertaremos, a la larga, 95% de las veces

Prueba de hipótesis

El problema de las pruebas de hipótesis estadísticas puede plantearse sencillamente de la siguiente manera: *¿es compatible o no lo es, una observación o un hallazgo dados, según algunas hipótesis planteadas?* La palabra "compatible" se utiliza aquí en el sentido de que la observación es lo "bastante" cercana al valor hipotético, de forma que no se rechaza la hipótesis planteada.

En el lenguaje de estadística, la hipótesis planteada se conoce como hipótesis nula, y se denota con el símbolo H_0. La hipótesis nula suele probarse frente a una hipótesis alternativa (también conocida como hipótesis mantenida) denotada con H_1, que puede plantear, por ejemplo, que el verdadero β2 es diferente a la unidad. La hipótesis alternativa puede ser simple o compuesta.
Por ejemplo, H_1: β_2 = 1.5 es una hipótesis simple, pero H_1: $\beta_2 \neq 1.5$ es una hipótesis compuesta

*Hay dos métodos mutuamente complementarios para diseñar tales reglas: el **intervalo de confianza** y la **prueba de significancia***

Estos dos enfoques plantean que la variable o estimador en consideración sigue alguna distribución de probabilidad y que la prueba de hipótesis establece afirmaciones sobre el valor del parámetro de tal distribución. La mayoría de las hipótesis estadísticas se hace afirmaciones sobre uno o más valores de los parámetros de algunas distribuciones de probabilidad supuestas, como la normal, F, t o χ2.

Pruebas de hipótesis: método del intervalo de confianza

Para ilustrar el enfoque del intervalo de confianza, una vez más nos referiremos al ejemplo de salarios y nivel de escolaridad. Por los resultados de la regresión obtenidos en la ecuación (3.6.1), sabemos que el coeficiente de pendiente es 0.7240. Supongamos que se postula que

H_0: $\beta_2 = 0.5$

H_1: $\beta_2 \neq 0.5$

es decir, el verdadero coeficiente de la pendiente es 0.5 según la hipótesis nula, pero menor o mayor que 0.5 según la hipótesis alternativa. La hipótesis nula es una hipótesis simple, mientras

La hipótesis nula es una hipótesis simple, mientras que la alterna es compuesta. Esto es lo que se conoce como hipótesis de dos colas. Es común que una hipótesis alterna de dos colas refleje el hecho que no tenemos una fuerte base teórica o a priori de hacia dónde se debe mover la hipótesis.

¿Es el $\hat{\beta}_2$ observado compatible con H_0? Para responder, consultemos el intervalo de confianza. Sabemos que, a la larga, los intervalos como (0.5700, 0.8780 que contendrán el verdadero valor de β_2 con una probabilidad del 95%. En consecuencia, si β_2 bajo la H_0 cae dentro del intervalo de confianza de 100(1 – α) por ciento podemos aceptar la hipótesis nula y si no la rechazamos.

Regla de decisión Construya un intervalo de confianza para β_2 a $100(1-\alpha)\%$. Si el β_2 en H_0 se encuentra dentro de este intervalo de confianza, no rechace H_0, pero si está fuera del intervalo, rechace H_0.

Por lo que rechazamos la hipótesis nula.

En estadística, cuando se rechaza la hipótesis nula, se dice que el hallazgo es **estadísticamente significativo**. Por otra parte, cuando no se hace, se dice que el resultado no es **estadísticamente significativo**.

En este caso decimos que nuestra estimación de 0,7240 es estadísticamente significativa, es decir, significativamente diferente en términos estadísticos del valor hipotético de 0,5.

Prueba unilateral o de una cola

Algunas veces tenemos una expectativa *a priori* o teórica sólida (o existen expectativas basadas en algún trabajo empírico previo) de que la hipótesis alternativa es unilateral o unidireccional, en lugar de ser bilateral o de dos colas, como acabamos de analizar. Así, para el ejemplo de los salarios y el nivel de escolaridad, se puede postular que

H_0: $\beta_2 \leq 0.5$ y H_1: $\beta_2 > 0.5$

Quizá la teoría económica o el trabajo empírico previo indiquen que la pendiente es mayor que 0.5.

Prueba de Hipótesis

La Prueba de significancia

Una prueba de significancia es un procedimiento mediante el cual se utilizan los resultados de la muestra para verificar la veracidad o falsedad de una hipótesis. La idea fundamental es utilizar un estadístico de prueba (estimador) y la distribución muestral de dicho estadístico bajo la hipótesis nula. La decisión de aceptar o rechazar H_0 se toma sobre la base del valor del estadístico obtenido a partir de los datos.

Si el valor verdadero de β_2 se especifica en la hipótesis nula, el valor de t puede calcularse a partir de la muestra disponible, sirviendo como estadístico de prueba. Y como este estadístico posee la distribución t, es posible hacer intervalo de confianza

Pr (- $t_{\alpha/2}$ ≤ $\frac{\hat{\beta}_2 - \beta^*_2}{se(\hat{\beta}_2)}$ ≤ $t_{\alpha/2}$) = 1 – α

Donde β_2^* es el valor de β_2 bajo la H_0 ; y - $t_{\alpha/2}$ y $t_{\alpha/2}$ son los valores de t que se obtiene de la tabla para un nivel de significancia de α/2 y N – 2 g de l

Reordenando tenemos

Pr (β_2^* - $t_{\alpha/2}$se($\hat{\beta}_2$) ≤ $\hat{\beta}_2$ ≤ β_2^* + $t_{\alpha/2}$se($\hat{\beta}_2$)) = 1 – α

Esta expresión proporciona el intervalo en el que $\hat{\beta}_2$ se encontrará con una probabilidad de (1 – α)% dado que β_2 = β_2^*.

El intervalo de confianza 100(1 – α) se conoce como región de aceptación (de la H_0), mientras que la región fuera del intervalo se denomina región de rechazo (de H_0) o región crítica, los extremos del intervalo se llaman valores críticos

En el procedimiento de intervalos de confianza tratamos de establecer los límites dentro de los cuales se puede encontrar el β_2 verdadero pero desconocido, mientras en el enfoque de la prueba de significancia se plantea un valor hipotético para β_2 y se busca comprobar si el $\hat{\beta}_2$ calculado se encuentra dentro de ciertos límites razonables (de confianza) con respecto al valor de la hipótesis.

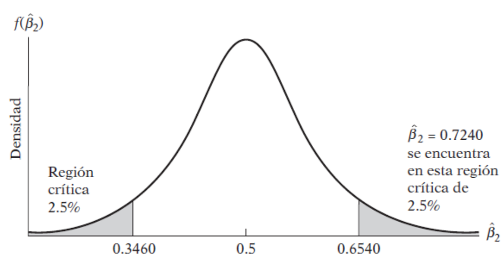
Ej. el ejemplo de los salarios y el nivel de escolaridad. Sabemos que $\hat{\beta}_2$ = 0.7240, ee ($\hat{\beta}_2$) = 0.0700 y gl = 11. Si suponemos α = 5%, $t_{\alpha/2}$ = 2.201.

Si suponemos que H_0: β_2 =β_2^* = 0.5 y H_1: $\beta_2 \neq 0.5$

Pr (0.3460 ≤ $\hat{\beta}_2$ ≤ 0.6540)

valor que con claridad se encuentra en la región crítica de la figura 5.4. La conclusión se mai tiene; es decir, rechazamos H_0.

FIGURA 5.3
Intervalo de confianza a 95% para $\hat{\beta}_2$ según la hipótesis de que $\beta_2 = 0.5$.



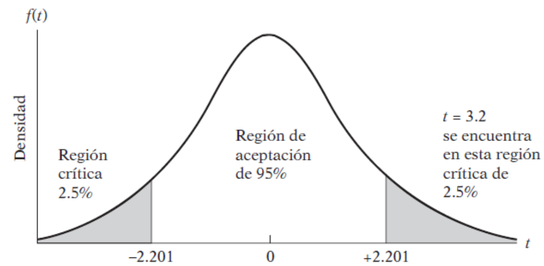
En la práctica, se puede calcular el valor de t y observar si se encuentra dentro de los valores de t críticos o fuera de ellos ej.

t = $\frac{\hat{\beta}_2 - \beta^\circ_2}{se(\hat{\beta}_2)}$

t = $\frac{0{,}724 - 0{,}5}{0{,}0700}$ = 3,2

Como vemos este valor se encuentra en la zona crítica. Se dice que un estadístico es estadísticamente significativo si el valor del estadístico de prueba se encuentra en la región crítica. En esta caso t es significativo y rechazamos la H_0

FIGURA 5.4
Intervalo de confianza a 95% para t(11 gl).



A medida en que el valor de β_2 estimado se aleje del valor hipotético de β_2, el $|t|$ (es decir, el valor absoluto de *t; nota*: *t* puede ser positivo o negativo) será cada vez mayor. *Por consiguiente, un valor "grande" de $|t|$ será evidencia en contra de la hipótesis nula*.

Siempre se puede utilizar la tabla *t* para determinar si un valor *t* particular es grande o pequeño; la respuesta, como sabemos, depende de los grados de libertad igual que de la probabilidad del error tipo I que estemos dispuestos a aceptar.

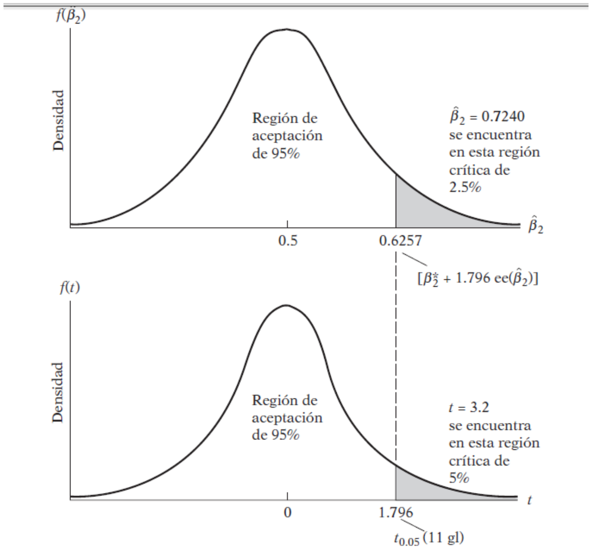
Para una cola supóngase ahora que la experiencia o la teoría nos sugieren que el valor esperado de la PMC es mayor que 0,5. En este caso tenemos

H_0 =, $\beta_2 \leq 0{,}5$

H_1 =, $\beta_2 > 0{,}5$

Para probar esta hipótesis utilizamos una prueba de una cola (derecha), el límite de confianza superior o valor critico es t_α = $t_{0,05}$

La utilización de una prueba de significancia de una o dos colas dependerá de la forma como se formule la hipótesis alternativa, la cual, a su vez, puede depender de algunas consideraciones *a priori* o de experiencia empírica previa.



Prueba de significancia para σ^2 : la prueba X^2

En la metodología de la prueba de significancia consideramos la siguiente variable

X^2 = (N - 2) $\frac{\hat{\sigma}^2}{\sigma^2}$

La cual sigue una distribución X^2 con N – 2 g de l. Si tenemos que $\hat{\sigma}^2$ = 42,1591 y los g de l = 8. Podemos plantear que H_0 = σ^2 = 85 versus H_1 =$\sigma^2 \neq$ 85.

De la ecuación calculamos el estadístico que nos da 3,97. Si suponemos que α = 95%, tenemos los valores críticos de la tabla para X^2 son 2,1797 y 17.5346. Como el valor calculado cae dentro de los limites aceptamos la H_0 .

Regla Práctica del “2-t” y la hipótesis nula o cero “t”

Si el número de grados de libertad es de 20 o más y si α, el nivel de significancia se fija en 0,05, entonces la hipótesis nula β_2 = 0 se puede rechazar a favor de la hipótesis alterna de que $\beta_2 \neq$ 0, si el valor t calculado a partir de la ecuación $\frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$ es mayor a 2 en valores absolutos.

Rechazamos la H_0: β_2 = 0 cuando

t = $\frac{\hat{\beta}_2}{se(\hat{\beta}_2)} > t_{\alpha/2}$ cuando $\hat{\beta}_2 > 0$

t = $\frac{\hat{\beta}_2}{se(\hat{\beta}_2)} < -t_{\alpha/2}$ cuando $\hat{\beta}_2 < 0$

Cuando decidimos “aceptar” la hipótesis nula, todo lo que se afirma es que, con base en la evidencia dada por la muestra, no existe razón para rechazarla; no se sostiene que la hipótesis nula sea verdadera con absoluta certeza

Formación de las hipótesis nula y alternativa

¿Cómo se formulan estas hipótesis? No existen reglas específicas. Muy a menudo, el fenómeno en estudio sugerirá la forma de las hipótesis nula y alternativa

El caso de la demanda de dinero. Que es una función del ingreso. Estudios anteriores indican que la elasticidad ingreso de la demanda de dinero suele ubicarse en un rango de 0.7 a 1.3. Por consiguiente, un nuevo estudio de la demanda de dinero, si se postula que el coeficiente β_2 de la elasticidad ingreso es 1, la hipótesis alternativa puede ser que $\beta_2 \neq 1$, una hipótesis alternativa bilateral.

Es en extremo importante que el investigador plantee estas hipótesis antes de la investigación empírica

Selección del nivel de significancia α

La única forma de decidir sobre esta compensación es encontrar los costos relativos de ambos tipos de error Desde luego, el problema es que pocas veces se conocen los costos de cometer los dos tipos de error. Por tanto, los econometristas tienen por costumbre fijar el valor de α en niveles de 1, 5 o 10% como máximo, y escogen un estadístico de prueba que haga que la probabilidad de cometer un error tipo II sea lo más pequeña posible.

Como uno menos la probabilidad de cometer un error tipo II se conoce como la potencia de la prueba, el problema relacionado con la selección del valor apropiado de α se evita al emplear lo que se conoce como valor p del estadístico de prueba.

Se denomina valor p (es decir, valor de probabilidad), también conocido como nivel observado o exacto de significancia, o probabilidad exacta de cometer un error tipo I. Más técnicamente, el valor p se define como nivel de significancia más bajo al cual puede rechazarse una hipótesis nula.

En la mayor parte de los análisis económicos aplicados, la hipótesis nula postulada hace las veces de comodín, y el objetivo del trabajo empírico es, rechazar la hipótesis nula.

Análisis de Regresión y Análisis de Varianza (Prueba F)

Veremos el análisis de regresión desde el punto de vista del análisis de varianza, anteriormente desarrollamos la siguiente igualdad STC = SCE + SCR, lo que indica es que la suma total de cuadrados = suma de cuadrados explicados + suma de los cuadrados de los residuos. Desde el punto de vista de la regresión el estudio de los componentes de la STC se conoce como análisis de varianza (ADV).

Asociados a cualquier suma de cuadrados existen gl correspondientes, los cuales están en función con el número de observaciones independientes en los cuales dicha suma se basa.

La distribución *F* con 1 gl en el numerador y ($n - 2$)gl en el denominador.

La STC tiene N – 1 g de l, porque perdemos un grado de libertad en el cálculo de la media muestral $\bar{Y}$.

La SCR tiene N – 2 g de l (valido caso dos variables y β_1 presente)

La SCE tiene 1 g de l (valido caso dos variables, es una función de $\hat{\beta}_2$ solamente)

$$F = \frac{\text{SCM de SCE}}{\text{SCM de SCR}}$$

F constituye un estadístico que sirva para probar la hipótesis nula de que el verdadero $\beta_2 = 0$ .Para hacer esto se debe calcular el valor de F y compararlo con el valor critico F que se obtiene de las tablas para el nivel de significancia que se escoja.

En el ejemplo el valor calculado de F = 202,87. Si $\alpha = 5\%$, el valor critico de F para 1 y 8 g de l es 5,32. En este caso el valor calculado de F es estadísticamente significativo y, por lo tanto se rechaza la hipótesis nula de que el ingreso no tiene influencia en los gastos de consumo.

El valor t de la tabla corresponde a un valor F con 1 g de l en el numerador y k g de l en el denominador.

Para el ejemplo si suponemos que H_0 =, $\beta_2 = 0$ se puede verificar el valor de F, usando el valor t estimado que tiene 8 g de l .

$(t)^2 = (14{,}26)^2 = F$.

Tanto las pruebas t y F son dos maneras alternas y complementarias para evaluar la H_0 de que $\beta_2 = 0$, para el modelo "predecir" o "pronosticar" el salario promedio futuro *Y* correspondiente a algún nivel dado de escolaridad *X* de dos variables no es necesario recurrir a F. Pero para regresión múltiple tiene diferentes aplicaciones.

Aplicación del análisis de regresión: problema de predicción

$\hat{Y_i} = -0.0144 + 0.7240X_i$

Donde $\hat{Y_i}$ es el estimador del verdadero $E(Y_i)$ correspondiente a *X* dada.

"predecir" o "pronosticar" el salario promedio futuro *Y* correspondiente a algún nivel dado de escolaridad *X.*

Hay dos clases de predicciones: **predicción media** y **predicción individual**.

Predicción media

Suponga que $X_0 = 20$ y deseamos predecir $E(Y|X_0 = 20)$. Ahora, puede demostrarse que la regresión histórica proporciona la estimación puntual de esta predicción media de la siguiente forma:

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$$

$$\hat{Y}_0 = -0.0144 + 0.7240\,(20)$$
$$\hat{Y}_0 = 14.4656$$

Donde $\hat{Y}_0$ = estimador de $E(Y|X_0)$. Puede comprobarse que este predictor puntual es el mejor estimador lineal e insesgado (MELI).

Como $\hat{Y}_0$ es un estimador, es probable que éste sea diferente de su verdadero valor. La diferencia entre los dos valores dará alguna idea del error de predicción o pronóstico. Para evaluar este error es necesario encontrar la distribución muestral de $\hat{Y}_0$

$\hat{Y}_0$ está normalmente distribuida con media ($\beta_1 + \beta_2 X_0$) y una varianza dada por la siguiente

$$\text{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right]$$

Al reemplazar la σ^2 desconocida por su estimador insesgado $\sigma^{\wedge 2}$, vemos que la variable

$$t = \frac{\hat{Y}_0 - (\beta_1 + \beta_2 X_0)}{\text{ee}(\hat{Y}_0)}$$

Sigue una distribución t con $n - 2$ gl. La distribución t sirve por consiguiente para construir intervalos de confianza para el verdadero $E(Y_0 \mid X_0)$ y pruebas de hipótesis acerca de tal valor de la manera usual, a saber

$$\Pr[\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\alpha/2}\,\text{ee}(\hat{Y}_0) \leq \beta_1 + \beta_2 X_0 \leq \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\alpha/2}\,\text{ee}(\hat{Y}_0)] = 1 - \alpha$$

Donde ee ($\hat{Y}_0$) se obtiene de la siguiente manera

$$\begin{aligned} \text{var}(\hat{Y}_0) &= 0.8936 \left[\frac{1}{13} + \frac{(20-12)^2}{182} \right] \\ &= 0.3826 \end{aligned}$$

y

$$\text{ee}(\hat{Y}_0) = 0.6185$$

Por tanto, el intervalo de confianza a 95% para el verdadero $E(Y \mid X_0) = \beta_1 + \beta_2 X_0$ está dado por

$$14.4656 - 2.201(.6185) \leq E(Y_0 \mid X = 20) \leq 14.4656 = 2.20(0.6185)$$

$$13,1043 \leq E(Y \mid X = 20) \leq 15,8260$$

Predicción individual

Lo que interesa es predecir un valor individual Y, Y_0 correspondiente a un valor dado de X, digamos, X_0, entonces, el mejor estimador lineal insesgado de Y_0, pero su varianza es la siguiente.

$$\text{var}(Y_0 - \hat{Y}_0) = E[Y_0 - \hat{Y}_0]^2 = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right] \qquad (5.10.6)$$

Además, se demuestra que Y_0 también sigue una distribución normal con media y varianza dadas por (5.10.1) y (5.10.6), respectivamente. Al sustituir $\hat{\sigma}^2$ por la desconocida σ^2, se colige que

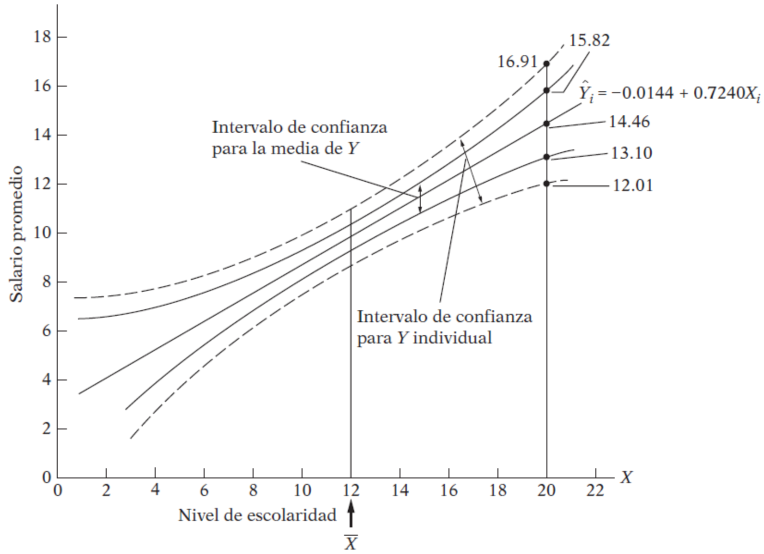
$$t = \frac{Y_0 - \hat{Y}_0}{\text{ee}(Y_0 - \hat{Y}_0)}$$

También sigue una distribución t. Por consiguiente, la distribución t sirve para inferir sobre la verdadera Y_0. Al continuar con el ejemplo, vemos que la predicción puntual de Y_0 es 14.4656, igual a $\hat{Y}_0$, y su varianza es 1.2357

$12.0190 \leq (Y_0/X_0=20) \leq 16,9122$

Vemos que el intervalo de confianza para la Y_0 individual es más amplio que el intervalo para el valor medio de Y_0.

Se debe tener cuidado al “extrapolar” la línea de regresión histórica para predecir $E(Y \mid X_0)$ o Y_0 asociada a una X_0 dada muy alejada de la media muestral $\bar{X}$.



Informe de resultados del análisis de regresión

$\hat{Y}_i = -0.0144 + 0.7240X_i$

ee = (0.9317) (0.0700) $r^2 = 0.9065$

t = (−0.0154) (10.3428) gl = 11

p = (0.987) (0.000) $F_{1,11} = 108.30$

Evaluación de los resultados del análisis de regresión

Primero, ¿están los signos de los coeficientes estimados de acuerdo con las expectativas teóricas o previas? *A priori*, β_2 en el ejemplo de los salarios y el nivel de escolaridad debe ser positivo. En el presente ejemplo, lo es. Segundo, si la teoría sostiene que la relación no debe ser sólo positiva sino también estadísticamente significativa, ¿es el caso en la presente aplicación? Como analizamos, el coeficiente del nivel de escolaridad no sólo es positivo, sino también estadísticamente significativo, es decir, diferente de cero; el valor *p* del valor *t* estimado es muy pequeño. Valen los mismos comentarios para el coeficiente del intercepto. Tercero, ¿qué tan bien explica el modelo de regresión la variación en el ejemplo? Se puede responder con r^2. En nuestro ejemplo, r^2 es de alrededor de 0.90, un valor muy alto si consideramos que r^2puede ser máximo 1.

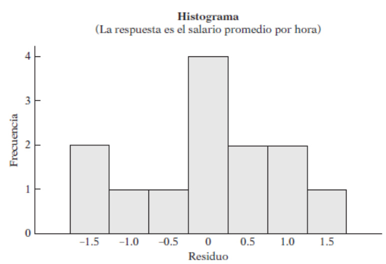
Así, parece muy bueno el modelo escogido para explicar el comportamiento de los salarios promedio. Pero antes de comprometerse con él, sería interesante averiguar si satisface los supuestos del MCRLN. No veremos ahora los diversos supuestos, pues la simplicidad del modelo es clara. Sólo hay un supuesto que se puede verificar, a saber, el de normalidad del término de perturbación, u_i. Recuerde que las pruebas *t* y *F* requieren que el término de error siga una distribución normal. De lo contrario, el procedimiento de prueba no será válido en muestras pequeñas, o finitas.

Pruebas de normalidad

Aunque se han estudiado diversas pruebas de normalidad en la teoría, sólo consideraremos tres: 1) histograma de residuos, 2) gráfica de probabilidad normal (GPN) y 3) prueba **Jarque-Bera**.

Histograma de residuos

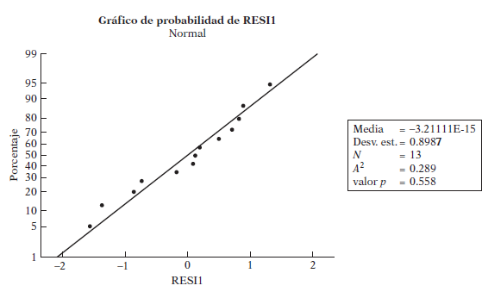
Este diagrama muestra que los residuos no tienen distribución normal perfecta; para una variable distribuida normalmente, la asimetría (una medida de la simetría) debe ser cero, y la cuûrtosis (que mide si la distribución normal es alta o baja), 3. Siempre es aconsejable trazar el histograma de los residuos de cualquier regresión como método aproximado y rápido para probar el supuesto de normalidad.



Gráfica de probabilidad normal

Papel de probabilidad normal, especialmente diseñado para gráficas. Sobre el eje horizontal, o eje *X*, se grafican los valores de la variable de interés (por ejemplo, los residuos de MCO, *ûi*), y sobre el eje vertical, o eje *Y*, el valor esperado de esta variable si estuviera normalmente distribuida.

FIGURA 5.8
Residuos de la regresión de salarios y nivel de escolaridad.



Prueba de normalidad de Jarque-Bera (JB)

La prueba de normalidad JB es una prueba *asintótica*, o de muestras grandes. También se basa en los residuos de MCO. Esta prueba calcula primero la **asimetría** y la **curtosis** de los residuos de MCO, con el siguiente estadístico de prueba: donde *n* = tamaño de la muestra, *S* = coeficiente de asimetría y *K* = coeficiente de curtosis.

Para una variable normalmente distribuida, *S* = 0 y *K* = 3. Por tanto, la prueba de normalidad JB constituye una prueba de la hipótesis conjunta de que *S* y *K* son 0 y 3, respectivamente. En este caso, se espera que el valor del estadístico JB sea igual a cero. De acuerdo con la hipótesis nula, la cual afirma que los residuos están normalmente distribuidos, Jarque y Bera mostraron que *asintóticamente (es decir, en muestras grandes) el estadístico JB) sigue la distribución ji cuadrada, con 2 gl*. Si el valor *p* calculado del estadístico JB es lo bastante bajo en una aplicación, lo cual sucederá si el valor del estadístico difiere en gran medida de cero, se puede rechazar la hipótesis de que los residuos están normalmente distribuidos. Pero si el valor *p* es razonablemente alto, lo cual sucede cuando el valor del estadístico está cerca de cero, NO rechazamos la suposición de normalidad.

Naturaleza de la multicolinealidad

El término *multicolinealidad* se atribuye a Ragnar Frisch, este tiene que ver con el supuesto 8 del *modelo clásico de regresión lineal* (MCRL) plantea que no existe **multicolinealidad** entre las regresoras incluidas en el modelo de regresión.

En estricto sentido, la *multicolinealidad* se refiere a la existencia de más de una relación lineal exacta, y *colinealidad*, a la existencia de una sola relación lineal. Pero esta distinción pocas veces se mantiene en la práctica, y se hace entonces referencia a multicolinealidad en ambos casos

El término multicolinealidad incluye el caso de multicolinealidad perfecta, y también el caso en el cual hay *X* variables intercorrelacionadas pero no en forma perfecta, de la siguiente manera.

$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k = 0$

$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_2 X_k + v_i = 0$

Si la multicolinealidad es perfecta en el sentido, los coeficientes de regresión de las variables *X* son indeterminados, y sus errores estándar, infinitos. Si la multicolinealidad es menos que perfecta, los coeficientes de regresión, aunque sean determinados, poseen grandes errores estándar (en relación con los coeficientes mismos), lo cual significa que los coeficientes no pueden ser estimados con gran precisión o exactitud.

Existen diversas fuentes de multicolinealidad. Como afirman Montgomery y Peck, la multicolinealidad puede deberse a los siguientes factores:

1. El método de recolección de información. Por ejemplo, la obtención de muestras en un intervalo limitado de valores tomados por las regresoras en la población.

2. *Restricciones en el modelo o en la población objeto de muestreo*. Por ejemplo, en la regresión del consumo de electricidad sobre el ingreso (X_2) y el tamaño de las viviendas (X_3) hay una restricción física en la población, pues las familias con ingresos más altos suelen habitar viviendas más grandes que las familias con ingresos más bajos.

3. *Especificación del modelo.* Por ejemplo, la adición de términos polinomiales a un modelo de regresión, en especial cuando el rango de la variable *X* es pequeño.

4. *Un modelo sobre determinado*. Esto sucede cuando el modelo tiene más variables explicativas que el número de observaciones. Esto puede suceder en investigación médica, donde en ocasiones hay un número reducido de pacientes sobre quienes se reúne información respecto de un gran número de variables.

Otra razón para la multicolinealidad, sobre todo en los datos de series de tiempo, puede ser que las regresoras del modelo compartan una *tendencia común*; es decir, que todas aumenten o disminuyan a lo largo del tiempo. Por tanto, en la regresión del gasto de consumo sobre el ingreso, la riqueza y la población, las regresoras ingreso, riqueza y población tal vez todas crezcan con el tiempo a una tasa aproximadamente igual, con lo cual se presentaría la colinealidad entre dichas variables de regresión, en especial cuando el rango de la variable X es pequeño.

Consecuencias prácticas de la multicolinealidad

1. Aunque los estimadores de MCO son MELI, presentan varianzas y covarianzas grandes que dificultan la estimación precisa.

2. Debido a la consecuencia 1, los intervalos de confianza tienden a ser mucho más amplios, lo cual propicia una aceptación más fácil de la "hipótesis nula cero" (es decir, que el verdadero coeficiente poblacional es cero).

3. También debido a la consecuencia 1, la razón t de uno o más coeficientes tiende a ser estadísticamente no significativa.

4. Aunque la razón t de uno o más coeficientes sea estadísticamente no significativa, R^2, la medida global de bondad de ajuste, puede ser muy alta.

5. Los estimadores de MCO y sus errores estándar son sensibles a pequeños cambios en los datos.

Detección de la multicolinealidad

1. La multicolinealidad es una cuestión de grado y no de clase. La distinción importante no es entre presencia o ausencia de multicolinealidad, sino entre sus diferentes grados.

2. Como la multicolinealidad se refiere a la condición de las variables explicativas que son no estocásticas por supuestos, es una característica de la muestra y no de la población. Por consiguiente, no es necesario "llevar a cabo pruebas sobre multicolinealidad", pero, si se desea, es posible medir su grado en cualquier muestra determinada.

1. **Una R^2 elevada pero pocas razones t significativas**

La prueba F, en la mayoría de los casos, rechazará la hipótesis de que los coeficientes parciales de pendiente son simultáneamente iguales a cero, pero las pruebas t individuales no son significativas.

2. **Altas correlaciones entre parejas de regresoras**

Las correlaciones de orden cero elevadas son una condición suficiente pero no necesaria para la existencia de multicolinealidad, debido a que puede existir a pesar de que las correlaciones de orden cero o correlaciones simples sean comparativamente bajas (es decir, inferiores a 0.50).

3. **Examen de las correlaciones parciales.**

4. **Regresiones auxiliares**

5. **Valores propios e índice de condición.**

7. **Diagrama de dispersión**

Medidas correctivas

No hacer nada

Reglas prácticas

1. **Información *a priori***

2. **Combinación de información de corte transversal y de series de tiempo**

$\ln Y_t = \beta_1 + \beta_2 \ln P_t + \beta_3 \ln I_t + u_t$

$Y^*_t = \beta_1 + \beta_2 \ln P_t + u_t$

Donde $Y^* = \ln Y - \beta_3 \ln I$, es decir, Y^* representa ese valor de Y después de eliminarle el efecto del ingreso.

Aunque es una técnica atractiva, la mezcla de datos de series de tiempo y de corte transversal de esta forma puede crear problemas de interpretación porque se supone implícitamente que la elasticidad ingreso estimada a partir de datos de corte transversal es igual a la que se habría obtenido a partir de un análisis puro de series de tiempo.

3. **Eliminación de una(s) variable(s) y el sesgo de especificación**

Sin embargo, al eliminar una variable del modelo se puede incurrir en un **sesgo de especificación** o **error de especificación**

4. Transformación de variables

5. Datos nuevos o adicionales.

6. Reducción de la colinealidad en las regresiones polinomiales

Video

Heteroscedasticidad

Un supuesto importante del modelo clásico de regresión lineal es que las perturbaciones *ui* que aparecen en la función de regresión poblacional son homoscedásticas; es decir, que todas tienen la misma varianza.

La varianza condicional de *Yi* (la cual es igual a la de *ui*), condicional a las *Xi* dadas, permanece igual sin importar los valores que tome la variable *X*.

Hay diversas razones por las cuales las varianzas de *ui* pueden ser variables

1) Los *modelos de aprendizaje de los errores; 2) A medida que aumentan los ingresos, la gente posee más ingreso discrecional; 3) A medida que mejoran las técnicas de recolección de datos es probable que* σ_i^2 *se reduzca; 4)* La heteroscedasticidad también surge por la presencia de datos atípicos o aberrantes; 5) Otra fuente de heteroscedasticidad surge de la violación del supuesto 9 del MCRL, que establece que el modelo de regresión está correctamente especificado; 6) la asimetría en la distribución de una o más regresoras

incluidas en el modelo; 7) Otra fuente de heteroscedasticidad: como señala David Hendry, es debido a 1) la incorrecta transformación de los datos (por ejemplo, las transformaciones de razón o de primeras diferencias) y 2) una forma funcional incorrecta (por ejemplo, modelos lineales frente a modelos log-lineales).

El problema de heteroscedasticidad es quizá más común en la información de corte transversal que en la de series de tiempo**.** En las series de tiempo, por el contrario, las variables tienden a ser de órdenes de magnitud similares.

Estimación por MCO en presencia de heteroscedasticidad

¿Seguirá siendo MELI aunque sólo eliminemos el supuesto de homocedasticidad y lo reemplacemos por el de heteroscedasticidad?

La varianza de *ui*, homocedástica o heteroscedástica, no desempeña papel alguno en la determinación de la propiedad de insesgamiento. El β_2 estimado converge a su valor verdadero

En el supuesto de que β_2 continúe siendo lineal, insesgado y consistente, ¿es "eficiente" o "el mejor"? Es decir, ¿tendrá varianza mínima en la clase de los estimadores lineales e insesgados?

β_2 deja de ser el mejor y ya no es el de varianza mínima.

El método de mínimos cuadrados generalizados (MCG)

El método de MCO usual, no aprovecha la "información" contenida en la variabilidad desigual de la variable dependiente *Y*, como sucede con la compensación salarial de los empleados este método asigna igual peso o importancia a cada observación. Pero existe un método de estimación, conocido como **mínimos cuadrados generalizados** (**MCG**), que toma en cuenta esa información explícitamente y, por consiguiente, es capaz de producir estimadores que son MELI. Para ver cómo se hace, considere el modelo ya familiar con dos variables

$Y_i = \beta_1 + \beta_2 X_i + u_i$

Se transforma en un modelo donde se utiliza las varianzas heteroscedástica para obtener

$Y^*_i = \beta^*_1 X^*_{0i} + \beta^*_2 X^*_i + u^*$

Los β^*_1 *y* β^*_2 *estimados son ahora MELI y no los estimadores de MCO,* β_1 *y* β_2*.*

Este procedimiento de transformar las variables originales de forma que las variables transformadas satisfagan los supuestos del modelo clásico y de aplicar luego MCO a ellos se conoce como método de mínimos cuadrados generalizados (MCG).

Consecuencias de utilizar MCO en presencia de heteroscedasticidad

Los intervalos de confianza basados en estos últimos serán innecesariamente grandes. Como resultado, es probable que las pruebas *t* y *F* den resultados imprecisos.

En resumen, si insistimos en los procedimientos de prueba usuales a pesar de la presencia de heteroscedasticidad, las conclusiones o inferencias que obtengamos pueden ser muy equivocadas.

Detección de la heteroscedasticidad

En la mayoría de las investigaciones econométricas, la heteroscedasticidad puede ser un asunto de intuición, de conjeturas refinadas, de un trabajo basado en experiencia empírica previa o de pura especulación.

Métodos informales

Naturaleza del problema: en la información de corte transversal que comprende unidades heterogéneas, la heteroscedasticidad puede ser la regla y no la excepción.

Método gráfico: llevar a cabo un análisis de regresión con el supuesto de que no hay heteroscedasticidad y luego hacer un examen post mortem de los residuos elevados al cuadrado, u^2_i *, para ver si exhiben algún patrón sistemático.*

Métodos formale

Prueba de Park Como σ_i^2 *por lo general no se conoce, Park sugiere utilizar* u_i^2 *como aproximación y correr la siguiente regresión:*

$ln u^2 = ln\ \sigma^2 + \beta\ ln X_i + v_i$

$\alpha + \beta \ln X_i + v_i$

Prueba de Goldfeld-Quandt

Este popular método es aplicable si se supone que la varianza heteroscedástica, σ_i^2 , está relacionada positivamente con *una* de las variables explicativas en el modelo de regresión. Por simplicidad, considere el modelo usual con dos variables:

$Yi = \beta_1 + \beta_2 X_i + u_i$

Suponga que σ_i^2 está relacionado positivamente con X_i, en la forma

$\sigma_i^2 = \sigma^2 X_2$

Donde σ^2 es una constante

Prueba Breusch-Pagan-Godfrey

El éxito de la prueba de Goldfeld-Quandt depende no sólo del valor de *c* (el número de observaciones centrales que se van a omitir), sino también de la identifi cación de la variable *X* correcta que servirá de referencia para ordenar las observaciones. Esta limitación de la prueba se evita si consideramos la prueba Breusch-Pagan-Godfrey (BPG).

Prueba general de heteroscedasticidad de White

A diferencia de la prueba de Goldfeld-Quandt, que requiere reordenar las observaciones respecto de la variable *X* que supuestamente ocasiona la heteroscedasticidad, o de la prueba BGP, sensible al supuesto de normalidad, la prueba general de heteroscedasticidad propuesta por White no se apoya en el supuesto de normalidad y es fácil aplicarla.

La **prueba de Koenker-Basset (KB)**. Al igual que las pruebas Park, Breusch-Pagan-Godfrey y la de White, la prueba KB se basa en los residuos al cuadrado, u_i^2, pero en vez de hacer la regresión sobre una o más regresoras, se efectúa la regresión de los residuos al cuadrado sobre los valores estimados de la regresora al cuadrado.

Medidas correctivas

La heteroscedasticidad no destruye las propiedades de insesgamiento y consistencia de los estimadores de MCO; sin embargo, éstos ya no son eficientes, ni siquiera asintóticamente.

Cuando se conoce $\sigma^2{}_i$: método de los mínimos cuadrados ponderados

Cuando no se conoce σ^2

Varianzas y errores estándar consistentes con heteroscedasticidad de White, los errores estándar de White corregidos mediante heteroscedasticidad también se conocen como errores estándar robustos.

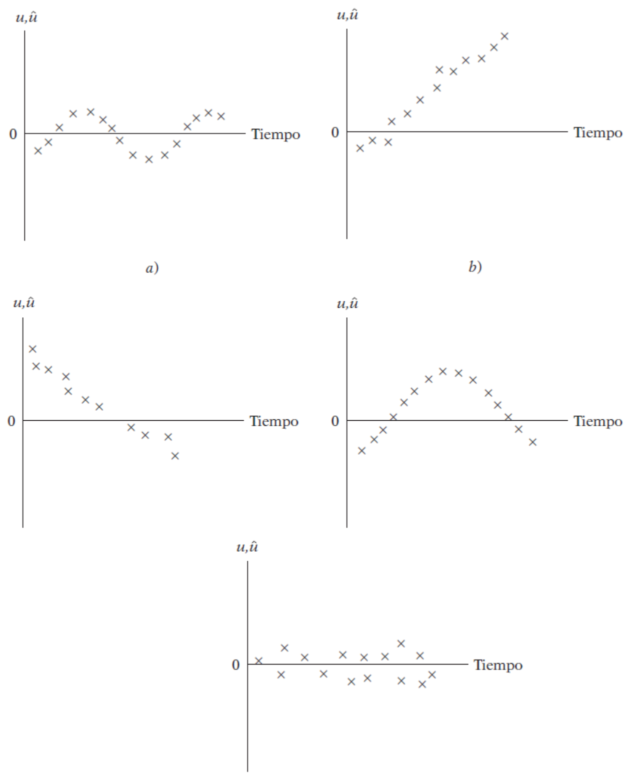
Una desventaja del procedimiento de White, además de ser de muestras grandes, es que los estimadores obtenidos por este medio pueden no ser tan eficientes como los obtenidos por métodos que transforman la información para reflejar tipos específicos de heteroscedasticidad.
Consideraremos ahora diversos supuestos sobre el patrón de heteroscedasticidad.

Autocorrelación

El término **autocorrelación** se define como la "correlación entre miembros de series de observaciones ordenadas en el tiempo o en el espacio. El modelo clásico supone que el término de perturbación relacionado con una observación cualquiera no recibe influencia del término de perturbación relacionado con cualquier otra observación
En los estudios transversales, a menudo los datos se recopilan con base en una muestra aleatoria de unidades transversales; como familias o empresas, de modo que no existe razón previa para creer que el término de error que correspondiente a una familia o a una empresa esté correlacionado con el término de error de otra familia o empresa. Si por casualidad se observa dicha correlación en unidades transversales, se conoce como **autocorrelación espacial.**
No esperaremos que el efecto de un incremento en el ingreso de una familia sobre su gasto de consumo incida en el gasto de consumo de otra.
La situación es muy distinta si trabajamos con datos de series de tiempo, pues las observaciones en estos datos siguen un ordenamiento natural respecto del tiempo, de modo que es muy posible que las observaciones sucesivas muestren intercorrelaciones, sobre todo si el intervalo entre observaciones sucesivas es muy corto, como un día, una semana o un mes, en lugar de un año.
Los índices bursátiles, como el Dow Jones o el S&P 500 en días sucesivos, no es raro que descubra que dichos índices aumentan o disminuyen durante varios días sucesivos. Obvio, en esta clase de situaciones se viola el supuesto del MCRL en cuanto a que **no existe autocorrelación, ni correlación serial** en los términos de error.
Si tratamos con información trimestral de series de tiempo, que implica una regresión de la producción sobre los insumos trabajo y capital, y si, por ejemplo, hay una huelga laboral que afecta la producción en un trimestre, no hay razón para pensar que esta interrupción afectará la producción del trimestre siguiente

En presencia tanto de autocorrelación como de heteroscedasticidad, los estimadores de MCO usuales, a pesar de ser lineales, insesgados y tener distribución asintóticamente normal (es decir, en muestras grandes), dejan de tener varianza mínima entre todos los estimadores lineales insesgados. En resumen, no son eficientes en relación con los demás estimadores lineales e insesgados. Dicho de otro modo, es posible que no sean los mejores estimadores lineales insesgados (MELI). Como resultado, las pruebas usuales *t*, *F* y $\chi 2$ pueden no ser válidas.

FIGURA 12.1
Patrones de autocorrelación y no autocorrelación.



¿Por qué ocurre la correlación serial?

Inercia: Una característica relevante de la mayoría de las series de tiempo económicas es la inercia o pasividad.
Sesgo de especificación: caso de variables excluidas:
Sesgo de especificación: forma funcional incorrecta
Fenómeno de la telaraña
La oferta de muchos productos agrícolas refleja el llamado fenómeno de la telaraña, en donde la oferta reacciona al precio con un rezago de un periodo debido a que la instrumentación de las decisiones de oferta tarda algún tiempo

$\text{Oferta}_t = \beta_1 + \beta 2 P_{t-1} + u_t$

Rezagos

En una regresión de series de tiempo del gasto de consumo sobre el ingreso no es extraño encontrar que el gasto de consumo en el periodo actual dependa, entre otras cosas, del gasto de consumo del periodo anterior

$\text{Consumo}_t = \beta_1 + \beta_2\, \text{ingreso}_t + \beta_3\, \text{consumo}_{t-1} + u_t$

“Manipulación” de datos

En el análisis empírico con frecuencia se “manipulan” los datos simples. Por ejemplo, en las regresiones de series de tiempo con datos trimestrales, por lo general estos datos provienen de datos mensuales a los que se agregan simplemente las observaciones de tres meses y se divide la suma entre 3.

La **interpolación** o **extrapolación**

El Censo de Población se realiza cada 10 años en Estados Unidos, y los dos últimos se efectuaron en 1990 y 2000. Ahora bien, si necesitamos datos para algún año comprendido en el periodo intercensal, la práctica común consiste en interpolar con base en algunos supuestos *adhoc*.
Transformación de datos
Lo importante del ejemplo anterior es que a veces la autocorrelación puede inducirse como resultado de transformar el modelo original

$Y_t = \beta_1 + \beta_2 X_t + u_t$

$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1}$

No estacionariedad

Mencionamos en el capítulo 1 que, al trabajar con datos de series de tiempo, quizá habría que averiguar si una determinada serie de tiempo es estacionaria
Una serie de tiempo es estacionaria, de manera informal, si sus características (por ejemplo, media, varianza y covarianza) son invariantes respecto del tiempo; es decir, no cambian en relación con el tiempo. Si no es así, tenemos una serie de tiempo no estacionaria.

Detección de la autocorrelación

Método gráfico

Recuerde que el supuesto de no autocorrelación del modelo clásico se relaciona con las perturbaciones poblacionales u_t, las cuales no pueden observarse directamente. En su lugar disponemos de valores sustitutos, los residuos $\hat{u}_i$, a partir del procedimiento usual MCO. Aunque las $\hat{u}_i$ no son lo mismo que las u_t, con mucha frecuencia un examen visual de las $\hat{u}$ da algunas claves sobre la posible presencia de autocorrelación en las u.

Prueba de “las rachas

Prueba *d* de Durbin-Watson

Es simplemente la razón de la suma de las diferencias al cuadrado de residuos sucesivos sobre la SCR. En el numerador del estadístico d, el número de observaciones es $n - 1$ porque se pierde una observación al obtener las diferencias consecutivas.
Una gran ventaja del estadístico d es que se basa en los residuos estimados, que se calculan de manera rutinaria en los análisis de regresión.

Supuestos en los cuales se basa

1. El modelo de regresión incluye el término del intercepto. Si dicho término no está presente, como en la regresión a través del origen, es esencial efectuar de nuevo la regresión con dicho término para obtener la SCR
2. Las variables explicativas, X, son no estocásticas, es decir, son fijas en muestreo repetido.

3. Las perturbaciones *ut* se generan mediante el esquema autorregresivo de primer orden:
$u_t = \rho u_{t-1} + \varepsilon_t$. Por tanto, no se pueden utilizar para detectar esquemas autorregresivos de orden superior.
4. Se supone que el término de error *ut* está normalmente distribuido.
5. El modelo de regresión no incluye valor(es) rezagado(s) de la variable dependiente como una variable explicativa. Por tanto, la prueba es inaplicable a modelos del siguiente tipo:
$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + \gamma Y_{t-1} + ut$ donde Y_{t-1} es el valor de Y rezagada un periodo. Tales modelos se conocen como **modelos autorregresivos**
6. No hay observaciones faltantes en los datos. Por tanto, en la regresión de salarios-productividad de 1960 a 2005, si por alguna razón faltaran observaciones, por ejemplo, de 1978 y 1982, el estadístico *d* no permitiría la ausencia de tales observaciones.
Es evidente de la ecuación que si $\hat{\rho} = 0$, $d = 2$; es decir, si no hay correlación serial (de primer orden), esperamos que *d* esté alrededor de 2. *Por consiguiente, como regla práctica, si en una aplicación vemos que* d *es igual a 2, podemos suponer que no hay autocorrelación de primer orden, positiva o negativa*. Si $\hat{\rho} = +1$, indica una correlación positiva perfecta en los residuos, $d \approx 0$. Por consiguiente, entre más cercano esté *d* a 0, mayor será la evidencia de correlación serial positiva.
Entre más se acerque *d* a 4, mayor será la evidencia de correlación serial negativa.

Debe añadirse que un estadístico *d* significativo no necesariamente indica autocorrelación. Más bien, puede indicar una omisión de las variables pertinentes en el modelo.

La prueba de Breusch-Godfrey (BF)

Medidas correctivas
1. Trate de averiguar si se trata de **autocorrelación pura** y no el resultado de una mala especificación del modelo a veces se observan patrones en los residuos porque el modelo está mal especificado —es decir, se excluyeron variables importantes— o porque su forma funcional no es correcta.

2. Si se trata de autocorrelación pura, se puede utilizar una transformación apropiada del modelo original de manera que en el modelo transformado no se presente el problema de la autocorrelación (pura). Como en la heteroscedasticidad, habrá que emplear algún **método generalizado de mínimos cuadrados (MCG**).

El remedio depende de la naturaleza de la interdependencia entre las perturbaciones *ut*. Pero como las *ut* no son observables, la práctica común es suponer que algún mecanismo las generó.
5. El mecanismo más común es el esquema autorregresivo de primer orden de Markov, que supone que la perturbación en el tiempo actual está linealmente relacionada con el término de perturbación en el tiempo anterior, el coeficiente de autocorrelación ρ que da el grado de interdependencia. Este mecanismo se conoce como esquema AR(1).
6. Si el esquema AR(1) es válido y se conoce el coeficiente de autocorrelación, el problema de correlación serial se resuelve fácilmente mediante la transformación de los datos según el procedimiento de diferencias generalizado. El esquema AR(1) se generaliza sin dificultad a un esquema AR(p). También se puede suponer un mecanismo de promedios móviles (PM) o una mezcla de los esquemas AR y PM, conocido como ARMA..
7. Aunque utilicemos un esquema AR(1), el coeficiente de autocorrelación ρ no se conoce *a priori*. Consideramos diversos métodos para estimar ρ, como el *d* de Durbin-Watson, el *d* modificado de Theil Nagar, el procedimiento de dos etapas de Cochrane-Orcutt (C-O), el procedimiento iterativo C-O y el método de dos etapas de Durbin. En muestras grandes, estos métodos suelen producir estimaciones similares de ρ, aunque en muestras pequeñas tienen un desempeño diferente. En la práctica, el método iterativo C-O ha cobrado gran popularidad.