Economía Aplicada

Regresión Lineal

Basado en Stock y Watson (cap.4-6), Wooldridge (cap.3-5)

Outline

Introducción al Modelo de Regresión Lineal Simple 1/2

- Muchas veces el análisis de economía aplicada comienza con la siguiente premisa: Y y X son dos variables, y estamos interesados en "explicar Y en función de X", o en "estudiar cómo Y varía con cambios en X".
- Para hacer eso, necesitamos considerar tres aspectos:
  1. nunca hay una relación exacta entre las variables
  2. necesitamos especificar la relación funcional entre Y y X
  3. debemos asegurarnos de capturar una relación *ceteris paribus* entre Y y X (si ese es nuestro objetivo)

Introducción al Modelo de Regresión Lineal Simple 2/2

- En el modelo de regresión lineal, la relación entre la variable a explicar (Y), y una variable explicativa (X) la expresamos de la siguiente manera:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

  donde ε representa el término de error (inobservable) y β_0 y β_1 los parámetros poblacionales (desconocidos).
- ε representa todos los factores que afectan a Y, además de X. En este modelo simple suponemos que todos esos factores son inobservables.
- Si los factores de ε se mantienen fijos, $\Delta\varepsilon = 0$, entonces X tiene un efecto lineal sobre Y: $\Delta Y = \beta_1 \Delta X$

Ejemplo

$$salario = \beta_0 + \beta_1 educ + \varepsilon$$

- Entonces el término de error (ε) puede recoger factores inobservables como la experiencia laboral, la capacidad o habilidad, otras características personales: edad, sexo, etc.
- Bajo ciertos supuestos podremos capturar el efecto, *ceteris paribus*, de la educación sobre salario.
- Si los factores de ε se mantienen fijos, $\Delta\varepsilon = 0$, entonces el efecto de un año más de educación sobre el salario está dado por β_1.

Supuestos del Modelo de Regresión Lineal

1. Linealidad en los parámetros.

   No resulta muy restrictivo, permite incorporar relaciones no lineales entre variables al introducir funciones (o transformaciones) no lineales de X y/o Y.

2. Esperanza del error, condicional en X, es cero: $E(\varepsilon|X) = 0$
   Este supuesto **CRUCIAL** implica:
   1. $E(\varepsilon) = 0$ y
   2. $C(h(X), \varepsilon) = 0$, donde $h(X)$ es cualquier función de X. En particular implica $C(X, \varepsilon) = 0$

3. Las observaciones provienen de muestras aleatorias.

   Demasiado restrictivo en algunos casos. Ejemplo: Queremos estudiar los salarios femeninos, pero solamente observamos los salarios para las mujeres que trabajan.

Media condicional nula en el ejemplo anterior

$$salario = \beta_0 + \beta_1 educ + \varepsilon$$

- Para simplificar asumamos que ε es capacidad innata.
- $E(capacidad|educ = 0)$ representa la capacidad media del grupo de personas sin educación, y $E(capacidad|educ = 12)$ la capacidad media del grupo de personas con 12 años de eduación.
- Entonces, el supuesto de esperanza condicional cero, implica que $E(capacidad|educ = 0)$ debe ser igual a $E(capacidad|educ = 12)$.
- Como no observamos la capacidad, no podemos saber si la capacidad promedio es la misma para distintos niveles educativos.
- Si, en promedio, las personas con mayor capacidad eligiesen adquirir más educación, este supuesto no sería adecuado.

Potenciales Violaciones al supuesto de Media Condicional Nula

Modelo incorrectamente especificado

- Ejemplo: *salario* no depende de manera lineal de *educ*

Variables explicativas correlacionadas con variables omitidas

- Ejemplo: no controlamos por capacidad y capacidad correlacionada con *educ*

Errores de medida en las variables explicativas

- Ejemplo: obtenemos una *proxy* para capacidad (IQ), pero que mide imperfectamente la capacidad

Variables explicativas determinadas simultáneamente con Y

- Ejemplo: El nivel educativo depende de los retornos a la educación

Interpretación de los coeficientes

- Dados los primeros dos supuestos del modelo simple, se verifica que:

$$E(Y|X) = \beta_0 + \beta_1 X$$

- En este caso, la función de esperanza condicional es lineal, y coincide con la proyección lineal de Y dado X. Entonces la pendiente β_1 tiene una **interpretación causal**.

- La pendiente es:
$$\beta_1 = \frac{\Delta E(Y|X)}{\Delta X}$$
cuando X varía en una unidad, Y varía en promedio β_1 unidades.
- La constante β_0 se interpreta como el valor medio de Y cuando $X = 0$: $E(Y|X = 0)$. En la práctica sólo tiene interpretacin si $X = 0$ es un valor plausible para la muestra considerada.
- De todos modos, β_0 debe incluirse siempre para controlar que X e Y no tienen por qué tener media igual a 0.
- Notar que si $E(\varepsilon|X) \neq 0$ los parámetros no tienen interpretación causal.

Ejemplo

$$\hat{salario} = -0.90 + 0.54 educ$$

donde el salario por hora se mide en euros y la educación en años.

- El valor estimado de la pendiente indica que un año más de educación (cualquiera sea el nivel), incrementa el salario por hora en 54 céntimos.
- En cuanto a la constante, nuestra estimación diría que a una persona sin estudios le corresponde un salario por hora promedio de -90 céntimos. Esto se explica en este caso porque en la muestra no hay personas sin estudios, el valor mínimo de la variable educación es 8. Para una persona con 8 años de educación el salario predicho es 3.42 euros por hora.

Estimación: Principio de Analogía

- Nuestro interés es estimar los parámetros poblacionales del modelo. Para ello usamos una muestra de la población. Dada esa muestra, nuestro modelo simple se convierte en:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

  donde se verifica que $E(\varepsilon_i | X_i) = 0$, $\forall i = 1, ..., n$

- Una manera de obtener estimadores de los parámetros es utlizar el principio de analogía. La idea es aproximar las características poblacionales con sus correspondientes análogos muestrales.

Estimación: Principio de Analogía

- Bajo los supuestos del modelo, minimizando la varianza del error ε,

$$E(\varepsilon^2) = E[(Y - \beta_0 - \beta_1 X)^2]$$

obtenemos las fórmulas de los parámetros poblacionales:

$$\beta_0 = E(Y) - \beta_1 E(X)$$

$$\beta_1 = \frac{C(Y, X)}{V(X)}$$

Estimación: Principio de Analogía

- Tenemos entonces los parámetros poblacionales expresados como función de momentos poblacionales.
- Aplicando el principio de analogía, sustituyendo momentos poblacionales por muestrales en las ecuaciones obtenidas antes, obtenemos estimadores de los parámetros poblacionales β_0 y β_1.

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$
$$\hat{\beta}_1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

Estimación: MCO

- Los mismos estimadores se obtienen utilizando el criterio de minimizar la varianza del análogo muestral de los errores: los residuos.
- Los residuos son la diferencia entre el valor observado y el predicho: $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$.
- Si minimizamos la suma de los residuos al cuadrado $\sum \hat{\varepsilon}_i{}^2$, obtenemos las mismas fórmulas que utilizando el principio de analogía. Estos son los estimadores de Mínimos Cuadrados Ordinarios (MCO).

Estimación: MCO

- Las condiciones de primer orden son:

$$\sum \hat{\varepsilon}_i = 0$$
$$\sum \hat{\varepsilon}_i X_i = 0$$

- Notar que estas condiciones son el análogo muestral de

$$E(\varepsilon) = 0$$
$$C(X, \varepsilon) = 0$$

- Una vez que se obtienen los estimadores, es posible construir la línea de regresión MCO: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

Modelo de Regresión Múltiple

- Es claro que la mayoría de las relaciones económicas involucran a más de dos variables. En la práctica vamos a incorporar tantos controles como creamos necesario:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K + \varepsilon$$

- Un modelo de regresión lineal múltiple es más adecuado para análisis ceteris paribus, ya que permite controlar explícitamente por muchos de los factores que afectan simultáneamente a la variable dependiente.
- Además nos da mayor flexibilidad para modelar la relación entre Y y las X.

Supuestos 1/2

1. Linealidad en los parámetros.
2. Esperanza del Error, condicional en todas las X, es cero:

   $E(\varepsilon|X_1, X_2, ..., X_K) = 0$
   Este supuesto implica:

   1. $E(\varepsilon) = 0$ y
   2. $C(h(X_j), \varepsilon) = 0$, donde $h(X_j)$ es cualquier función de X_j. En particular, $C(X_j, \varepsilon) = 0$

Supuestos 2/2

3. Las observaciones provienen de una muestra aleatoria.
4. Necesitamos un supuesto adicional: Ausencia de multicolinealidad perfecta: $(X_1, X_2, ..., X_K)$ no forman una combinación lineal exacta.

Interpretación de los coeficientes

- La interpretación de los coeficientes es ahora de efecto parcial o ceteris paribus.
- Cuando X_j varía en una unidad, **permaneciendo el resto de las variables constantes**, Y varía en promedio en β_j unidades.

$$\beta_j = \frac{\Delta E(Y|X)}{\Delta X_j}$$

donde X representa al vector de variables explicativas: $X_1, X_2, ..., X_K$

Ejemplo

$$\hat{sal} = -2.87 + 0.60educ + 0.02exper + 0.17antig$$

donde el salario por hora está medido en euros y educación, experiencia en el mercado de trabajo (exper) y tiempo con la empresa actual (antig) se miden en años.

- El coeficiente de *educ* mide el cambio en el salario por hora si se tiene un año más de eduación, manteniendo experiencia y antigüedad fijos. En el ejemplo, un año más de educación (para cualquier nivel), incrementará el salario por hora en promedio en 60 céntimos, si *exper* y *antig* se mantienen fijos.
- Alternativamente, si comparamos dos personas con el mismo nivel de experiencia y antigüedad, el coeficiente de *educ* es la diferencia en el salario esperado cuando sus niveles de educación difieren en un año.

Otra interpretación : Regresión Múltiple

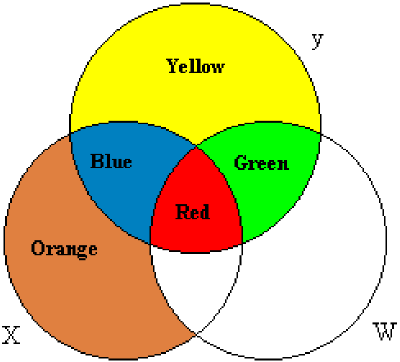
- Verdadero modelo: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K + \varepsilon$
- Dos Etapas
  1. MCO de X_1 en una constante y el resto de los regresores: computar el residuo e_1.
  2. MCO de Y en e_1.
- Los residuos de la primera etapa capturan cambios en X_1 no correlacionados con cambios en el resto de los regresores.
- La pendiente estimada en la segunda etapa es $\hat{\beta}_1$

Otra interpretación: Ejemplo

Verdadero modelo: $salario = \beta_0 + \beta_1 educ + \beta_2 IQ + \varepsilon$

- Dos Etapas
  1. MCO de *educ* en una constante e *IQ*: computar el residuo e_1.
  2. MCO de *salario* en e_1.
- Los residuos de la primera etapa capturan cambios en *educ* no correlacionados con cambios en *IQ*.
- La pendiente estimada en la segunda etapa es $\hat{\beta}_1$: el efecto "directo" de la educación en el salario

Diagramas de Venn para analizar el Modelo de Regresión Multiple[1]



- Cada círculo representa la variación de la variable correspondiente.
- En el modelo simple, MCO usa las áreas Roja y Azul para estimar β_X.
- Cuando utilizamos X y W: MCO descarta el área Roja y solamente utiliza el área Azul para estimar β_X (y la Verde para β_W).
- Idea: área Roja posee información "contaminada" (no sabemos si variación en y se debe a X o a W).
- Cuál es el efecto entonces de omitir W en la estimación?

[1] From Kennedy(2002)

Multicolinealidad

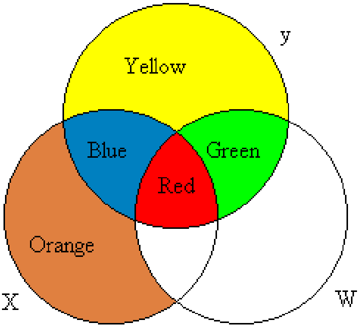


Figure 3a Modest collinearity

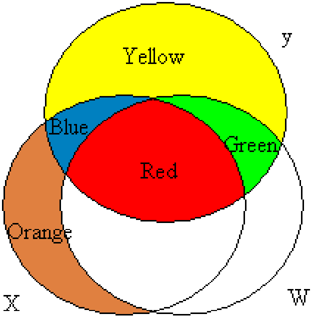


Figure 3b Considerable collinearity

En el caso de colinealidad considerable MCO sigue siendo consistente, pero al utilizar menos información (comparar las áreas azul y verde en ambas figuras), la estimación es menos precisa.

Bondad de Ajuste

Descomposición de Varianza:

- Cada observación de Y_i se compone de dos partes: $Y_i = \hat{Y}_i + e_i$
  - valores predichos: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + ... + \hat{\beta}_k X_{ki}$
  - residuo MCO: $e_i = Y_i - \hat{Y}_i$
- Podemos definir la suma de cuadrados totales (SCT), la suma de cuadrados explicada (SCE) y la suma de cuadrados de los residuos (SCR) como:

  $SCT = \sum(Y_i - \bar{Y})^2$

  $SCE = \sum(\hat{Y}_i - \bar{Y})^2$

  $SCR = \sum(e_i)^2$

El R^2

- La variación total de Y (SCT) se divide en la variación explicada (SCE) y la no explicada (SCR).
  $SCT = SCE + SCR$
- Dividiendo entre SCT y resolviendo para SCE

$$\frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \equiv R^2$$

- El R^2 se interpreta como la proporción de la variación muestral de Y explicada por el modelo.
- Se verifica que $0 \leq R^2 \leq 1$
- Nunca decrece cuando se agregan regresores.

Modelo lineal en las variables

$$E(Y|X) = \beta_0 + \beta_1 X,$$

- Como vimos antes, en este caso,

$$\beta_1 = \frac{\Delta E(Y|X)}{\Delta X}$$

- Cuando X varía en una unidad, Y varía en promedio β_1 unidades.
- ¿Cómo se calcula la elasticidad en este caso?

$$\eta = \frac{\Delta E(Y|X)/E(Y|X)}{\Delta X/X} = \beta_1 \frac{X}{E(Y|X)}$$

- Notemos que la elasticidad depende de Y y X, y que por lo tanto no es constante.

Modelo semilogarítmico: logaritmos en la variable explicativa

$$E(Y|X) = \beta_0 + \beta_1 lnX,$$

$$\beta_1 = \frac{\Delta E(Y|X)}{\Delta lnX} \approx \frac{\Delta E(Y|X)}{\Delta X/X}$$

- Para expresar la variación de X en términos porcentuales multiplicamos y dividimos por 100:

$$\beta_1/100 \approx \frac{\Delta E(Y|X)}{100\Delta X/X}$$

- Cuando X varía en un 1%, Y varía en promedio $\beta_1/100$ unidades.

Modelo semilogarítmico: logaritmos en la variable dependiente

$$E(lnY|X) = \beta_0 + \beta_1 X,$$

$$\beta_1 = \frac{\Delta E(lnY|X)}{\Delta X} \approx \frac{\Delta E(Y|X)/E(Y|X)}{\Delta X}$$

- Para expresar la variación de Y en términos porcentuales multiplicamos por 100:

$$100\beta_1 \approx \frac{E((100 \times \Delta Y/Y)|X)}{\Delta X}$$

- Cuando X varía en una unidad, Y varía en promedio $100\beta_1$ %.

Modelo doble logarítmico

- Este modelo se suele utilizar cuando se quiere modelizar una elasticidad constante entre X e Y.

$$E(lnY|X) = \beta_0 + \beta_1 lnX,$$

$$\beta_1 = \frac{\Delta E(lnY|X)}{\Delta lnX} \approx \frac{E((\Delta Y/Y)|X)}{\Delta X/X}$$

- Cuando X varía en un 1%, Y varía en promedio β_1%: β_1 representa la elasticidad de Y con respecto a X.

Modelo con términos cuadráticos

- Se suelen utilizar términos cuadráticos para por ejemplo modelizar efectos marginales crecientes o decrecientes.
- Nótese que β_1 y β_2 no tienen interpretación por separado

$$E(Y|X) = \beta_0 + \beta_1 X + \beta_2 X^2,$$

Entonces,

$$\frac{\Delta E(Y|X)}{\Delta X} = \beta_1 + 2\beta_2 X$$

- Cuando X varía en una unidad, Y varía en promedio $\beta_1 + 2\beta_2 X$. El efecto marginal de X sobre Y dependerá del signo de β_2.

Ejemplo

El archivo gasoline.gdt contiene información sobre el mercado de gasolina de EEUU para los años 1960-1995. La variable *pcgas* es el consumo per capita de gasolina e *inc* es el ingreso disponible per capita medidos en dólares. En cada punto, estime el modelo, interprete β_1 y calcule la elasticidad ingreso del gasto en gasolina.

1. $pcgas = \beta_0 + \beta_1 inc + \varepsilon$
2. $pcgas = \beta_0 + \beta_1 log(inc) + \varepsilon$
3. $log(pcgas) = \beta_0 + \beta_1 log(inc) + \varepsilon$
4. $log(pcgas) = \beta_0 + \beta_1 log(inc) + \beta_2 log(P_g) + \varepsilon$, donde P_g es el precio de la gasolina. Calcule la elasticidad precio.

Propiedades de los estimadores MCO

- Insesgadez (bajo linealidad y esperanza condicional del error cero)
  - $E\left(\widehat{\beta}_0\right) = \beta_0$
  - $E\left(\widehat{\beta}_1\right) = \beta_1$
  - Recordemos que la insesgadez es una característica de la distribución de los estimadores, lo que no nos asegura nada sobre la estimación puntual que obtengamos en una muestra determinada. Confiamos que si la muestra es "típica", entonces nuestros estimadores deberían estar "cerca" del valor poblacional.

Muestras Grandes

- nos centramos en las propiedades de los estimadores a medida que el tamaño de la muestra n se hace cada vez más grande:

1. ¿Cuán lejos está $\hat{\beta}_j$ del verdadero parámetro β_j cuando $n \to \infty$?
2. ¿Cómo es la distribución de $\hat{\beta}_j$ cuando $n \to \infty$?

Ley de Grandes Números (1/2)

- esencialmente establece que el promedio de los resultados obtenidos de un número grande de experimentos (iid) se vuelve más cercano al valor esperado cuantas más pruebas se realicen
- la LGN es importante porque "garantiza" resultados de muestras grandes estables para eventos aleatorios
- no significa que si aparece un cierto valor, inmediatamente será compensado por los siguientes (falacia del jugador)

Ley de Grandes Números (2/2)

Teorema

Para cualquier variable aleatoria y con valor esperado μ, definiendo a $\bar{y}_n$ como el promedio de la muestra de tamaño n.
Entonces

$$plim(\bar{y}_n) = \mu$$

$$plim(\hat{cov}(y,x)) = cov(y,x)$$

$$plim(\hat{var}(x)) = var(x)$$

Convergencia en Probabilidad

Definición

A medida que el tamaño de la muestra crece, cualquier distancia positiva entre $\hat{\beta}_j$ y β_j se hace arbitrariamente improbable.

- β_j es el límite en probabilidad de $\hat{\beta}_j$
- $\hat{\beta}_j$ converge en probabilidad a β_j
- $plim(\hat{\beta}_j) = \beta_j$

Consistencia MCO

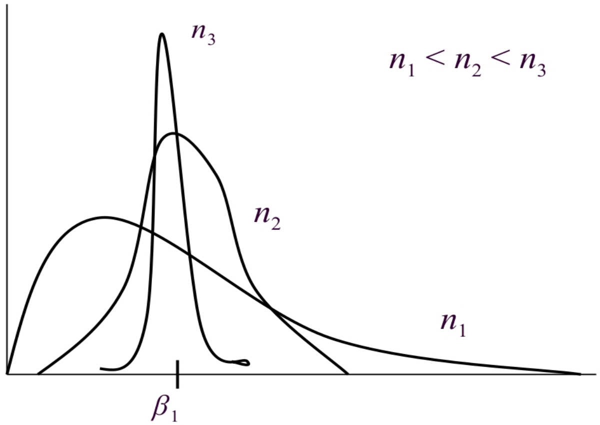
Teorema

Bajo los supuestos del modelo de regresión, los estimadores MCO son consistentes

Ejemplo: $wages = \beta_0 + \beta_1 educ + u$ con $cov(educ, u) = 0$

- $\hat{\beta}_1 = \beta_1 + \frac{\hat{cov}(educ_i, u_i)}{\hat{var}(educ_i)}$
- $plim\left(\hat{\beta}_1\right) = plim(\beta_1) + \frac{plim(\hat{cov}(educ_i, u_i))}{plim(\hat{var}(educ_i))} = \beta_1 + \frac{cov(educ, u)}{var(educ)}$
- dado que $cov(educ, u) = 0 \Rightarrow plim\left(\hat{\beta}_1\right) = \beta_1$

Interpretación Gráfica de Consistencia



Un ejemplo de inconsistencia

Modelo Verdadero: $wages = \beta_0 + \beta_1 educ + \beta_2 IQ + v$

- $cov(educ, v) = cov(IQ, v) = 0$
- $cov(educ, IQ) \neq 0$, $\beta_2 \neq 0$

- Ecuación estimada por MCO: $wages = \hat{\gamma}_0 + \hat{\gamma}_1 educ + \hat{u}_{educ}$

$\hat{\gamma}_1 = \hat{\beta}_1 + \hat{\beta}_2 \frac{\hat{cov}(educ, IQ)}{\hat{var}(educ)} \Rightarrow plim(\hat{\gamma}_1) = \beta_1 + \beta_2 \frac{cov(educ, IQ)}{var(educ)}$

- $plim(\hat{\gamma}_1) \neq \beta_1$ si
  - inteligencia es relevante: $\beta_2 \neq 0$
  - educación correlacionada con inteligencia: $cov(educ, IQ) \neq 0$

Teorema Central del Límite

Teorema

Para cualquier variable aleatoria y con valor esperado μ y varianza σ^2 se define el promedio de una muestra de tamaño n como $\overline{y}_n$. Entonces

$$n^{\frac{1}{2}} \frac{\overline{y}_n - \mu}{\sigma} \rightarrow N(0,1) \, \text{cuando} \, n \rightarrow \infty$$

- lo destacable del TCL es que la distribución de la variable aleatoria y es irrelevante

Normalidad Asintótica de los Estimadores MCO

Bajo los supuestos del Modelo de Regresión

$$n^{\frac{1}{2}} \frac{\hat{\beta}_j - \beta_j}{a_j} \to N(0,1) \text{ cuando } n \to \infty$$

$$\text{donde } a_j^2 = Asy.Var\left(n^{\frac{1}{2}} \hat{\beta}_j\right)$$

Errores Estándar (Robustos) 1/2

podemos estimar a_j de manera consistente: $se\left(n^{\frac{1}{2}}\hat{\beta}_j\right)$

- A medida que n aumenta, los estimadores MCO, convenientemente transformados, se acercan tanto como queramos a una distribución normal.

$$n^{\frac{1}{2}}\hat{\beta}_j \approx N\left(\beta_j, a_j^2\right)$$

- podemos aproximar la desviación típica de $\hat{\beta}_j$, $sd\left(\hat{\beta}_j\right) \approx \frac{a_j}{n^{1/2}} \approx se\left(\hat{\beta}_j\right)$

Errores Estándar (Robustos) 2/2

- estos errores estándar son “robustos a la heteroscedasticidad”
- un detalle importante: en la mayoría de los paquetes econométricos, los “errores estándar” reportados no son $se\left(\hat{\beta}_j\right)$ sino otros válidos solamente bajo supuestos muy restrictivos (homoscedasticidad condicional).
- Homoscedasticidad Condicional : $V(\varepsilon|X) = \sigma^2$
- En la ecuación salarial, podemos interpretarlo como que la capacidad tiene la misma dispersión para diferentes niveles educativos. O también que la variabilidad salarial no depende del nivel educativo. Por ejemplo, $V(wage|educ = 0) = V(wage|educ = 12)$.
- ¿Qué sucede si las personas con bajo nivel educativo tienen pocas oportunidades y deben trabajar al salario mínimo?

El contraste t

- Usando el TCL (y una ley de grandes números) se puede demostrar que

$$t = \frac{\hat{\beta}_j - \beta_j}{se\left(\hat{\beta}_j\right)} \rightarrow N(0,1) \text{ cuando } n \rightarrow \infty$$

- este resultado puede ser utilizado entre otras cosas para contrastar si un coeficiente es significativamente distinto de 0

El contraste de significatividad (1/3)

Ejemplo: $wages = \beta_0 + \beta_1 educ + \beta_2 IQ + u$

- queremos comprobar que la capacidad afecta significativamente a los salarios, ceteris paribus
- definimos la hipótesis que queremos rechazar (será la hipótesis nula)
  - hipótesis nula: $H_0 : \beta_2 = 0$ (IQ no afecta a los salarios)
  - hipótesis alternativa: $H_1 : \beta_2 \neq 0$ (IQ afecta a los salarios)

El contraste de significatividad (2/3)

- bajo la hipótesis nula, $t = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} \approx N(0,1)$

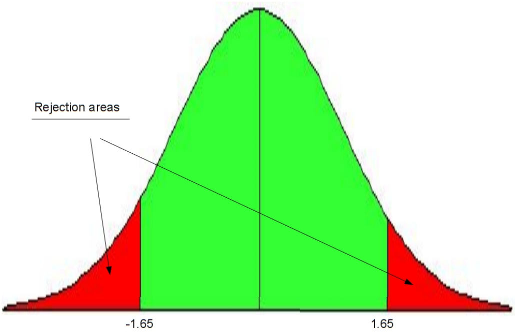
para rechazar la nula debemos encontrar suficiente evidencia en su contra

- suficiente evidencia sería que t es muy grande en valor absoluto: controlamos por la probabilidad de rechazar erróneamente (es decir, controlamos el nivel de significatividad)
- fijamos el nivel de significación al, digamos, 10%, y definimos el valor crítico como

$$c \text{ tal que } \Pr(|t| \geq c) = 0.10 \Rightarrow c \simeq 1.65$$

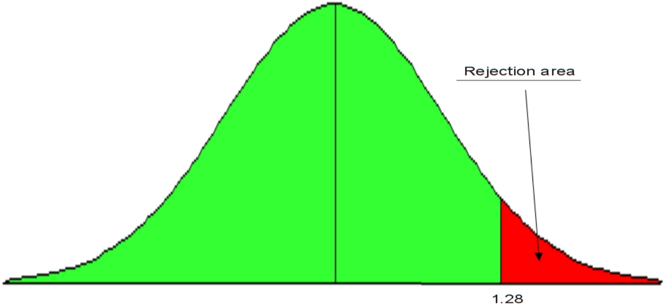
El contraste de significatividad (3/3)

Cuando contrastamos $H_0 : \beta_j = 0$ vs $H_1 : \beta_j > 0$ para un nivel de significatividad de 10%, el valor crítico es 1.65, como muestra la figura:



para un nivel de significatividad del 5%: si el test es mayor que 1.96, rechazamos H_0 :

Si fijamos el nivel de significación en 10%, el valor crítico para un contraste a una cola, es 1.28. Si el estadístico t es mayor al valor crítico, rechazamos la hipótesis nula a favor de que el parámetro es mayor que el postulado.



Potencia del contraste *t*

- controlamos la probabilidad de rechazar erróneamente: error del tipo I
- en cualquier contraste, podemos incurrir también en error de tipo II: no-rechazamos-pero-deberíamos-rechazar
- la probabilidad de rechazar cuando debemos rechazar se conoce como potencia del contraste: depende de la alternativa
- en la práctica, si rechazamos está bien, si no, quizás el test no tiene potencia...

valores-*p*

- una alternativa es preguntar: ¿cuál es el menor nivel de significatividad para el que el contraste se rechaza con el valor del test obtenido?
- para responder a esta pregunta, primero calculamos el estadístico t
- después miramos la masa de probabilidad, bajo la nula, de obtener un estadístico mayor
- este es el valor-p
- la mayoría de los programas econométricos computan el valor-p para contrastes de dos colas
- Si p es menor al nivel de significatividad deseado, rechazamos H_0 :.

Contraste de una combinación lineal

- Dada la siguiente hipótesis nula $H_0 : \sum a_j \beta_j = a^0$, definimos el siguiente estadístico de contraste:

$$t = \frac{\sum a_j \hat{\beta}_j - a^0}{s_{\sum a_j \hat{\beta}_j}}$$

- Podemos demostrar que $t \stackrel{\cdot}{\sim} N(0,1)$.
- Notar que este contraste requiere conocer o estimar la varianza de la combinación lineal de los estimadores. Esto implica la necesidad de conocer o estimar covarianzas entre estimadores.

Ejemplo: $wages = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 IQ + u$

- queremos investigar si un año más de educación tiene el mismo efecto en el salario que un año más de experiencia.
- La hipótesis adecuada es: $H_0 : \beta_1 = \beta_2$ y $H_1 : \beta_1 \neq \beta_2$
- O equivalentemente $H_0 : \beta_1 - \beta_2 = 0$ y $H_1 : \beta_1 - \beta_2 \neq 0$
- Entonces $t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{s_{\hat{\beta}_1 - \hat{\beta}_2}} \approx N(0, 1)$
- Pero para computar $s_{\hat{\beta}_1 - \hat{\beta}_2}$ necesitamos conocer o estimar la covarianza entre los estimadores.

- Alternativamente podríamos definir un nuevo parametro, $\theta = \beta_1 - \beta_2$. Entonces reescribimos el modelo:
  $wages = \beta_0 + \theta\, educ + \beta_2(educ + exper) + \beta_3 IQ + u$
- Si estimamos este nuevo modelo, basta con el estadístico t para la significatividad de θ
- Por supuesto que los paquetes econométricos nos permiten contrastar este tipo de hipótesis sin necesidad de reescribir el modelo.

Contrastes sobre varias restricciones lineales 1/3

- Un contraste de varias restricciones lineales es de la forma:

$$\beta_j = \beta_k = \beta_l = 0$$

- Cuando queremos contrastar varias restricciones lineales necesitamos distinguir el modelo no restringido del restringido.
- El modelo no restringido es el modelo sobre el que deseamos contrastar la hipótesis, el modelo original. El modelo restringido es el modelo una vez impuesta la hipótesis nula.

Contrastes sobre varias restricciones lineales 2/3

- Para cada uno de los modelos se computan la suma de cuadrados de los residuos y con estos valores se define el siguiente estadístico de contraste:

$$W^0 = n\frac{SRR - SRS}{SRS} \stackrel{\cdot}{\sim} \chi^2_q$$

  donde SRR representa la suma de cuadrados de los residuos en el modelo restringido, SRS la suma de cuadrados de los residuos en el modelo no restringido y q el número de restricciones incluidas en la hipótesis nula.

- Un estadístico equivalente es el estadístico F:

$$F = n\frac{SRR - SRS}{qSRS}$$

$$qF \stackrel{\cdot}{\sim} \chi^2_q$$

Contrastes sobre varias restricciones lineales 3/3

- Notar que W^0 (o F) nunca es negativo, puesto que $SRR \geq SRS$, ya que el modelo restringido no puede tener residuos menores que el modelo sin restringir.
- Alternativamente, cuando la variable explicativa en los modelos restringido y no restringido es la misma, podemos expresar el estadístico de contraste en términos de los R^2 de ambos modelos:

$$W^0 = n\frac{R_S^2 - R_R^2}{1 - R_S^2} \overset{\cdot}{\sim} \chi_q^2$$

donde R_R^2 representa el R^2 del modelo restringido y R_S^2 el R^2 del modelo sin restringir.

Ejemplo: $salario = \beta_0 + \beta_1 ESec + \beta_2 EUniv + \beta_3 IQ + u$

- queremos investigar si un año más de educación (secundaria o universitaria) tiene efecto en el salario o no.
- La hipótesis adecuada es: $H_0 : \beta_1 = \beta_2 = 0$ y $H_1 : \beta_1 \neq 0, \beta_2 \neq 0$
- para este contraste podemos utilizar el estadístico W^0 (o F)
- Modelo Restringido: $salario = \beta_0 + \beta_3 IQ + u$ con $q = 2$
- Por supuesto que los paquetes econométricos nos permiten contrastar este tipo de hipótesis sin necesidad de estimar ambos modelos.

Significatividad Conjunta

- El ejemplo más utilizado de estos contrastes es el llamado contraste de significatividad conjunta. Partiendo del siguiente modelo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K + \varepsilon$$

- La hipótesis nula, de que ninguna variable es relevante, incluye K restricciones lineales:

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_K = 0$$

- El modelo restringido por lo tanto es $Y = \beta_0 + \varepsilon$, y el estadístico correspondiente es:

$$W^0 = n \frac{R_S^2}{1 - R_S^2} \overset{\cdot}{\sim} \chi_K^2$$