Auxiliar pre examen

Repaso de contenidos escenciales

Profesor: Raimundo Undurraga
Auxiliares: Brandon Galarza, Camila Jáuregui Leonardo Meneses, Francisca Monetta, Matias Reyes, Bastian Urzua, Antonia Villegas

Comentes

1. El teorema de Gauss-Markov nos dice que siempre el estimador de Mínimos Cuadrados Ordinarios será el mejor estimador lineal insesgado

   Respuesta: Falso. Esto se cumplirá sólo en presencia de los supuestos base de MCO (exogeneidad, muestra aleatoria, no multicolinealidad, etc).

2. El método EMV sirve para estimar cualquier parámetro de cualquier distribución, ya que encuentra un máximo en la curva de verosimilitud.

   Solución:

   Falso.

   Si bien el método de Estimación de Máxima Verosimilitud (EMV) es un enfoque poderoso y ampliamente utilizado para estimar parámetros de distribuciones estadísticas, no es universalmente aplicable a todas las distribuciones y sus respectivos parámetros.

   El método EMV se basa en encontrar los valores de los parámetros que maximizan la función de verosimilitud, que representa la probabilidad de observar los datos dados ciertos valores de los parámetros. Si bien esto funciona bien en muchas situaciones y para muchas distribuciones comunes, existen casos en los que puede haber dificultades o limitaciones para aplicar este método:

   a) Condiciones de existencia de la función de verosimilitud: En algunos casos, la función de verosimilitud puede no tener un máximo bien definido o puede ser difícil de maximizar, lo que dificulta la aplicación directa del EMV.
   b) Problemas computacionales: En distribuciones complejas o con múltiples parámetros, encontrar el máximo de la función de verosimilitud puede ser computacionalmente costoso o incluso impracticable.
   c) Tamaño de la muestra: En muestras pequeñas, la estimación mediante EMV puede ser menos confiable, ya que la función de verosimilitud puede no reflejar con precisión la verdadera distribución subyacente.

Por lo tanto, aunque el método EMV es ampliamente aplicable y útil en muchas situaciones, no es necesariamente válido para todas las distribuciones y puede enfrentar limitaciones en ciertos casos.

3. Si la variable dependiente es binaria [0, 1], MCO siempre entregará predicciones dentro de ese rango.

   Respuesta: Falso, MCO asegura que el modelo será lineal, eso quiere decir que una predicción del tipo: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$ puede entregar valores mayores a 1 y menores a 0. Supongamos que tanto β_0 como β_1 son positivos. Habrá, entonces, un valor X > 0 tal que la predicción sea mayor que 1, y análogo para valores menores a 0.

4. La estimación de máxima verosimilitud siempre produce estimaciones insesgadas.

   Solución:

   La propiedad de insesgadez significa que, en promedio, el estimador produce un valor que es igual al verdadero valor del parámetro que se está estimando. Sin embargo, la máxima verosimilitud no garantiza la insesgadez en todas las situaciones. En algunos casos, las estimaciones de máxima verosimilitud pueden ser insesgadas, pero en otros casos pueden estar sesgadas.

   Aunque en muchos casos el estimador de máxima verosimilitud es insesgado, no siempre es así. En ciertos escenarios, especialmente con muestras pequeñas, puede generar estimaciones sesgadas. La afirmación entonces es FALSA.

5. Se tiene el siguiente modelo $Y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + \mu$, que se estima vía OLS. Mientras menor sea la correlación entre X_k y el resto de las variables independientes del modelo, entonces mayor será la varianza muestral del estimador de βk.

   Respuesta: Considerar que la expresión que permite conocer la desviación estándar del estimador es $SE(\hat{\beta}_k) = \frac{\sigma}{\sqrt{SST_k(1-R_k^2)}}$. Recordando también que la expresión para R_k^2 es $R_k^2 = 1 - \frac{\sum \tilde{r}_k^2}{SST_k}$, donde $\tilde{r}_k$ es el residuo obtenido al correr una regresión de X_k sobre el resto de variables independientes del modelo. En efecto, mientras mayor sea la correlación entre X_k y el resto de las variables independientes del modelo, entonces mayor será el residuo $\tilde{r}_k$. Mientras mayor sea el residuo $\tilde{r}_k$, menor es R_k^2. R_k^2 está entre 0 y 1, así mientras menor es R_k^2, menor es SE($\hat{\beta}_k$).

6. La eficiencia del estimador de máxima verosimilitud siempre es la mejor posible entre los estimadores insesgados.

   Solución:

   En general, para un tamaño de muestra grande, el estimador de máxima verosimilitud alcanza la eficiencia asintótica, lo que significa que para una gran cantidad de datos, tiende a tener la varianza más baja posible entre todos los estimadores insesgados.

   Dado lo anterior, podemos decir que el comente es verdadero.

P1.- OLS

Dado el contexto nacional, usted como estudiante está interesado en aportar en el debate de las pensiones. Para tener una opinión informada, usted se consigue los datos de la Superintendencia de Pensiones para analizar qué variables de los pensionados tiene un mayor efecto en el monto de su pensión. Para entender lo que está ocurriendo, usted propone el siguiente modelo:

$$ln(Monto\_Pension) = \beta_0 + \beta_1 Sexo + \beta_2 \text{Años}Cotizados + \beta_3 EdadInicio + \epsilon$$

Donde cada variable significa: $\ln(Monto\_Pension)$ = logaritmo natural del monto de la primera pensión en pesos considerando el aporte del pilar solidario, Sexo = el sexo del cotizante (1 es hombre y 0 es mujer), AñosCotizados = número de años cotizados, EdadInicio = edad en la que empezó a cotizar.

Los resultados que obtiene son:

Source	SS	df	MS
Model	64986.7865	3	21662.2622
Residual	139900.399	464,245	.301350363
Total	204887.186	464,248	.441331327

Number of obs	=	464,249
F(3, 464245)	=	71883.98
Prob > F	=	0.0000
R-squared	=	0.3172
Adj R-squared	=	0.3172
Root MSE	=	.54895

MontoPension	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Sex	.2686708	.0017014	157.91	0.000	.2653361	.2720054
AnosCotizados	.0286848	.0000926	309.63	0.000	.0285033	.0288664
EdadInicio	-.0077055	.000076	-101.44	0.000	-.0078544	-.0075566
_cons	16.22094	.0036753	4413.47	0.000	16.21373	16.22814

Figura 1: Resultados de la regresión

1. Testee la significancia individual de al menos 2 parámetros del modelo. Explicite el test y el método utilizado.

   Solución:

   Vamos a usar un test de significancia o también llamado test de hipótesis:

$$H_0 : \beta_i(x_i) = 0$$
$$H_1 : \beta_i(x_i) \neq 0$$

Con H_0 la hipótesis nula donde suponemos que una variable no es significativa para el modelo, H_1 la hipótesis alternativa que sostiene que sí lo es. X_i son las variables del modelo y $\beta_i(x_i)$ los parámetros asociados.

El test de signigficancia es así: P>|t|<0,05, si esta condición no se cumple; entonces se rechazal la hipótesis nula. P>|t| corresponde al P valor, que consiste en la probabilidad de que un valor estadístico calculado sea posible dada una hipótesis nula cierta, partiendo del supuesto de que la hipótesis nula es cierta y es bajo ese supuesto en el que calculamos el valor de p.

Una p $<$ 0,05 significa que la hipótesis nula es falsa y una p $>$ 0,05 que la hipótesis nula es verdadera: siempre nos movemos en el terreno de la probabilidad.

Vemos en la tabla P>|t| de las variables y vemos que todas tienen valor 0,000; por lo que como es menor que 0,05 rechazamos la hipótesis nula en que se sostiene que las variables del modelo no son significativas.

2. Para cada variable, interprete el valor obtenido. ¿Qué puede deducir al respecto?, ¿Qué variables aportan a mejorar las pensiones?

   Solución:

   - Para sexo: Ser hombre o mujer influye en las pensiones recibidas, siendo las mujeres las más perjudicadas.
   - Para años cotizados: a mayor años cotizados hay mayores pensiones, que influyen en un 2,8 % de las pensiones recibidas.
   - Para edad de inicio: a mayor edad de inicio menos pensión recibida, aunque el parámetro sea pequeño.

3. [Propuesto] Un compañero de trabajo le dice que existe alta correlación en el modelo. Identifique al menos un par de variables independientes que pudiesen estar correlacionadas. ¿Cómo afecta esto a la estimación del modelo?

   Solución:

   Años cotizados y edad de inicio podrían estar correlacionados porque a medida que la edad de inicio sea mayor, menos años de cotización significan. Esto conlleva a problemas de multicolinealidad y trae como consecuencia mayores SE($\hat{\beta}_k$) y mayores dificultades para encontrar $\hat{\beta}_k$ significativos y por ende, peores estimaciones para el modelo.

4. [Propuesto] Otro compañero le dice que hay variables relevantes que no están incorporadas al modelo. ¿Qué variable podría ser? ¿Qué implicancias tiene esto en las estimaciones que hizo? Explicite matemáticamente el efecto.

   Solución:

   Variables no incorporadas:

- escolaridad: años de escolaridad
- trabajo: tipo de trabajo que tiene una persona

Las pensiones podrían ser mejores si se hace algo respecto a sexo. Calculamos el sesgo por variable omitida:

$$\hat{\beta}_1 = \beta_1 + \beta_4 \frac{cov(Sexo, escolaridad)}{var(sexo)}$$

Con $cov(Sexo, escolaridad)$ donde suponemos que hay más mujeres con mayor años de escolaridad, por lo tanto $ov(Sexo, escolaridad) > 0$. $\beta_4 > 0$ porque a mayor hay más posibilidades de mejores empleos y con ello hay mejores sueldos. β_1 entonces estaría sobreestimado.

Vemos cómo las escolaridad con los años cotizados se relacionan:

$$\hat{\beta}_2 = \beta_2 + \beta_4 \frac{cov(AñosCotizado, escolaridad)}{var(AñosCotizado)}$$

cov$(AñosCotizado, escolaridad) < 0$, pues a mayor años de escolaridad; se tienen menos años cotizados (una persona que trabaja desde antes tiene más años de cotización), por ende $\hat{\beta}_2$ estaría subestimado.

¿Cómo se relaciona la edad de inicio con los años de escolaridad?

$$\hat{\beta}_3 = \beta_3 + \beta_4 \frac{cov(EdadInicio, escolaridad)}{var(EdadInicio)}$$

cov$(EdadInicio, escolaridad)$ es mayor que 0, pues a mayor años de escolaridad se tiene una mayor edad de inicio. Por ende $\hat{\beta}_3$ estaría sobreestimado.

Para trabajo:

$$\hat{\beta}_1 = \beta_1 + \beta_5 \frac{cov(Sexo, trabajo)}{var(Sexo)}$$

donde cov$(Sexo, trabajo)$, suponiendo que los hombres acceden a trabajos con mejores sueldos. $\hat{\beta}_1$ estaría sobreestimado.

Vemos cómo se relacionan los años cotizados con el tipo de trabajo:

$$\hat{\beta}_2 = \beta_2 + \beta_5 \frac{cov(AñosCotizados, trabajo)}{var(AñosCotizados)}$$

Aquí vemos que cov($AñosCotizados, trabajo$) $= 0$, porque no hay una relación aparente. Finalmente β_2 sería consistente.

Por último, vemos qué pasa entre la edad de inicio y el tipo de trabajo:

$$\hat{\beta}_3 = \beta_3 + \beta_5 \frac{cov(EdadInicio, trabajo)}{var(EdadInicio)}$$

tenemos que cov($EdadInicio, trabajo$) $= 0$, pues nuevamente no hay una relación aparente entre ambas variables. $\beta_3 =$ consistente.

5. Finalmente, usted quiere saber si el modelo efectivamente tiene relevancia. Para ello, fíjese en el R^2 y además realice un test de significancia global. ¿Qué concluye?

   Solución:

   Un alto R^2 implica una alta variabilidad en el modelo, en este caso $R^2 = 0{,}3172$. Planteamos el un test de hipótesis global: Prob $>$ |F| $<$ 0,05.

   Si esta condición se cumple, decimos que el modelo es significativo globalmente. En este caso Prob $>$ |F| $=$ 0,000 $<$ 0,05; por lo que el modelo sí es significativo de manera global.

$\boxed{\text{P2.-}}$ Inferencia Causal

Un grupo de estudiantes del DII deses aumentar su productividad, teorizando que el mayor problema de la baja productividad es el uso de RRSS.

Se ofrece a los 1.800 industriales acceder al estudio y NO usar RRSS, recibiendo como recompensa por la participación 1 décima en cualquier exámen. En la práctica, alrededor de un 10 % de los estudiantes decide participar.

Se accede a una base de datos donde se tiene la variable de productividad se mide en una escala continua de 0 a 100, y un indicador dicotómico que indica si NO usó RRSS. El modelo general que se construye es el siguiente:

$$Productividad_i = \beta_0 + \beta_1 NoUsaRRSS_i + \beta_2 VaAclases_i + \beta_3 HorasDeSueño_i + \epsilon_i$$

Los resultados de las estimaciones fueron las siguientes:

Table 1: Modelos lineales de Productividad

VARIABLES	**Modelo 1**	**Modelo 2**	**Modelo 3**	**Modelo 4**
NO usa RRSS = 1	13.98***	13.87***	8.310***	8.349***
	(1.073)	(1.095)	(1.308)	(1.315)
Va a Clases = 1		2.395		-0.012
		(0.809)		(0.809)
Horas de sueño			4.905***	5.026***
			(0.391)	(0.393)
Constante	58.64***	58.35***	49.38***	49.50***
	(0.361)	(0.620)	(1.275)	(1.374)
Observaciones	1,800	1,800	1,800	1,800
R cuadrado	0.019	0.019	0.025	0.025

Robust standard errors in parentheses
***p<0.001, **p<0.05, *p<0.1

1. Describa el mecanismo por el cual se explica que al agregar el regresor de Horas de sueño cambie el coeficiente de NoUsaRRSS, y la dirección de ese efecto. De una posible explicación de por qué esto no ocurre para el caso de la variable de Va a Clases.

Respuesta: Al agregar o quitar un regresor de tal forma que exista cambio importante en el valor de un estimador, implica que existe sesgo por variable omitida. Sabemos que este sesgo depende de: (1)la correlación de la variable omitida y la dependiente, y (2)la correlacion entre la variable omitida yel regresor estudiado. Por lo tanto, que el parámetro está siendo sobre estimado implica estos dos mecanismos:

(1) Por un lado, que horas de sueño tenga, por sí sola, un efecto directo sobre productividad (lo que es plausible)

(2) Por otro lado, la covarianza entre horas de sueño y no usar RRSS podría ser positiva, es decir, mayor horas de sueño , mayor sea la probabilidad de no usar RRRSS (lo que también es plausible, dado que tendrían más tiempo disponible que podrían usar para su descanso). Esto puede darse de manera simultánea también.

Se observa que el coeficiente asociado a NoUsarRRSS NO cambia al agregar la variable "Va a clases", lo cual puede analizarse según los mismos mecanismos anteriores: (1) Se observa que no hay significancia en el coeficiente asociado a "Va a clases", por lo que podría no ser relevante para explicar la productividad, y/o (2) No haya correlación entre ir a clases y no usar RRSS, es decir, que usar redes sociales no afecta la asistencia a las clases.

2. De forma general, describa detalladamente por qué estos estimadores no deberían tomarse como los efectos causales del experimento. Explique para este contexto específico cuál

es el problema que impide obtener una estimación limpia del impacto de RRSS sobre su productividad.

Respuesta: Los estimadores reflejan un efecto causal cuando su valor estimado mide exclusivamente el efecto del tratamiento (que es: NoUsarRRSS en este caso) y nada más. En este caso, la estimación no es limpia porque hay un problema de selección: las personas toman la decisión de no usar RRSS, es decir, se auto seleccionan. Como hay una auto-selección, no tenemos certeza de que el grupo que cumple con no usar RRSS es estadísticamente comparable con aquel que si usa RRSS. Esta diferencia se contiene en el estimador de β_1, y corresponde al sesgo de selección. Es por ello que los estimadores no pueden interpretarse como el efecto causal del estudio de "no uso de RRSS".

3. Explique formalmente, cuál es el rol del término de error en el problema que anteriormente describió y qué información podría contener ε_i

Respuesta: En línea de lo anterior, el error contiene todos los factores que no se observaron (ie: no se incluyeron en el modelo) que podrían generar diferencias en los grupos que reciben tratamiento y los que no reciben. Esto implica que el error está correlacionado con la variable NoUsaRRSS, dado que no hubo asignación aleatoria, por lo tanto: $Cov(NoUsaRRSS, \epsilon_i) \neq 0$.

Entonces, el error podría contener variables como por ejemplo: motivación, productividad, u otros elementos que determinan la probabilidad de querer participar, y que pueden relacionarse al incentivo asociado. En otras palabras, las personas que deciden no usar RRSS pueden ser sistemáticamente distintas en no-observables que los que no se capacitan.

4. Te encomiendan rediseñar el experimento de tus colegas DII, ¿cuál sería la clave, en términos estadísticos, para poder obtener su efecto causal respecto de la productividad de los estudiantes?

Respuesta: Si no hay otro factor que se incumpla (por ejemplo los supuestos del MCO, o sesgo por omisión de variable), entonces la clave estadística es que la asignación del tratamiento sea aleatoria, o exógena (exogeno = que el parametro no tenga correlación con el error u otro regresor que no sea la variable dependiente), lo que asegure que la covarianza entre el tratamiento y los no-observables sean cero. De esta manera, la única característica que distingue al grupo de personas que no usó RRSS y los que si, sea exclusivamente esto y ninguna otra cosa. Si sólo varía el tratamiento y ninguna otra característica, entonces se tendrá el efecto causal.

$\boxed{\text{P3.-}}$ MLE - Caso continuo

Sea Y_1,..., Y_n una muestra aleatoria cuya función de densidad está dada por:

$$f(x|y,\alpha) = \begin{cases} \frac{1}{\Gamma(\alpha)\theta^\alpha} y^{\alpha-1} e^{\frac{-y}{\theta}} & si \quad y > 0 \\ \\ 0 \sim \end{cases}$$

y que cumple $\mathbb{E}(Y) = \alpha\theta$ y $\mathbb{V} = \alpha\theta^2$, con $\alpha > 0$ conocido.

1. Encuentre el estimador de máxima verosimilitud para θ.

   Solución:

   Paso 1. Planteamos la función de verosimilitud considerando α conocido:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\theta^\alpha} y^{\alpha-1} e^{-\frac{y}{\theta}} \\ &= \left[\frac{1}{\Gamma(\alpha)\theta^\alpha}\right]^n \left(\prod_{i=1}^{n} y_i\right)^{\alpha-1} e^{-\frac{\sum_{i=1}^{n} y_i}{\theta}} \end{aligned}$$

   Paso 2. Nuevamente, utilizando las propiedades de los logaritmos, la log-verosimilitud está dada por:

$$l(\theta) = -n\ln(\Gamma(\alpha)) - n\alpha\ln(\theta) + (\alpha-1)\sum_{i=1}^{n} ln(y_i) - \frac{\sum_{i=1}^{n}[y_i]}{\theta}$$

   Paso 3. Derivamos con respecto a θ e igualamos a 0:

$$\begin{aligned} \frac{\partial l(\theta)}{\partial\theta} &= \frac{-n\alpha}{\theta} + \frac{\sum_{i=1} ny_i}{\theta^2} = 0 \\ \frac{n\alpha}{\theta} &= \frac{\sum_{i=1} ny_i}{\theta^2} \\ \hat{\theta} &= \frac{1}{n\alpha}\sum_{i=1}^{n} y_i \end{aligned}$$

   Paso 4. Derivamos nuevamente con respecto a θ para comprobar que es máximo:

$$\frac{\partial^2 l(\theta)}{\partial(\theta^2)} = \frac{n\alpha\theta - 2\sum_{i=1}^{n} y_i}{\theta^3}$$

Reemplazando por el θ encontrado en Paso 3.

$$= \frac{\sum_{i=1}^{n} y_i - 2\sum_{i=1}^{n} y_i}{(\sum_{i=1}^{n} y_i)^3} \cdot n\alpha^3$$
$$= \frac{-\sum_{i=1}^{n} y_i}{(\sum_{i=1}^{n} y_i)^3} \cdot n\alpha^3$$

Como $n\alpha$ es mayor que 0, veremos que al cancelar los términos:

$$= -\frac{1}{(\sum_{i=1}^{n} y_i)^2} \cdot n\alpha^3$$

La expresión es necesariamente negativa, ya que la sumatoria será positiva al elevarla al cuadrado, $n\alpha^3$ también, y por el signo negativo al principio sabemos que es menor que 0 , por lo tanto es máximo.

2. Calcule $\mathbb{E}(\hat{\theta})$.

   Ahora se calcula la $\mathbb{E}(\hat{\theta})$. Como n y α son constantes, salen del paréntesis.

$$\mathbb{E}(\hat{\theta}) = \frac{1}{n\alpha}\mathbb{E}(\sum_{i=1}^{n} y_i)$$

   Como Y_i es una muestra aleatoria i.i.d. podemos calcular la sumatoria de las esperanzas de cada y_i

$$\mathbb{E}(\hat{\theta}) = \frac{1}{n\alpha}\mathbb{E}(y_i) = \frac{n\alpha\theta}{n\alpha} = \theta$$

Ejercicios propuestos

MLE - Caso discreto

1. Suponga que tiene dos monedas en su bolsillo. Una de ellas es equilibrada, mientras que la otra no lo es. La probabilidad de observar cara con la moneda equilibrada es 0,5 mientras que con la no equilibrada la probabilidad de observar cara es 0,8. Identificando cara y sello con 1 y 0, respectivamente. Suponga ahora que saca una de las monedas de su bolsillo y deberá determinar si se trata de la moneda equilibrada o de la otra, pudiendo tirarla muchas veces en forma independiente. Suponga que en n = 100 lanzamientos observa la muestra. Plantee la función de verosimilitud para ambas monedas.

$$\underbrace{1,\dots,1}_{\text{12 veces}}\underbrace{0,\dots,0}_{\text{5veces}}\underbrace{1,\dots,1}_{\text{23 veces}}\underbrace{0,\dots,0}_{\text{8veces}}\underbrace{1,\dots,1}_{\text{15 veces}}\underbrace{0,\dots,0}_{\text{3 veces}}\underbrace{1,\dots,1}_{\text{11 veces}}\underbrace{0,\dots,0}_{\text{4 veces}}\underbrace{1,\dots,1}_{\text{13 veces}}\underbrace{0,\dots,0}_{\text{6 veces}}$$

Solución

Para responder a esta pregunta, vamos a calcular L(0,8) y L(1/2), la probabilidad de observar la muestra con cada una de las posibles monedas.

$$L(0,8) = \underbrace{0,8\dots0,8}_{\text{12 veces}}\underbrace{0,2\dots0,2}_{\text{5 veces}}\underbrace{0,8\dots0,8}_{\text{23 veces}}\underbrace{0,2\dots0,2}_{\text{8 veces}}\underbrace{0,8\dots0,8}_{\text{15 veces}}\underbrace{0,2\dots0,2}_{\text{3veces}}\underbrace{0,8\dots0,8}_{\text{11 veces}}\underbrace{0,2\dots0,2}_{\text{4veces}}\underbrace{0,8\dots0,8}_{\text{13 veces}}\underbrace{0,2\dots0,2}_{\text{6 veces}}$$

$$L(0,8) = (0,8)^{74}(0,2)^{26}$$

$$L(0,8) = 4,15231285 \cdot 10^{-26}$$

Siendo 74 el n umero de caras observadas en las N = 100 repeticiones. Análogamente, se tiene que:

$$L(1/2) = (1/2)^{74}(1/2)^{26}$$
$$L(1/2) = 7,8886091 \cdot 10^{-31}$$

La propuesta de máxima verosimilitud consiste en pensar que la moneda que estamos utilizando es aquella que más probabilidad le asigna a la muestra observada. Como L(0,8) > L(1/2) se concluye que se está utilizando la moneda no equilibrada.

2. Suponga ahora que desconoce por completo qué tipo de monedas tiene en su bolsillo. Siendo que en la muestra de n = 100 lanzamientos observó 74 caras. Calcule $\hat{\theta}$ el estimador de máxima verosimilitud para la obtención de caras y siga todos los pasos.

Solución

Paso 1. Planteamos la función de verosimilitud:

$$L(\theta) = (\theta)^{74}(1-\theta)^{26}$$

Paso 2. Aplicando logaritmo, la log-verosimilitud está dada por:

$$l(\theta) = 74\ln(\theta) + 26\ln(1-\theta)$$

Paso 3. Derivamos con respecto a θ e igualamos a 0:

$$\begin{aligned}
\frac{\partial l(\theta)}{\partial \theta} &= \frac{74}{\theta} - \frac{26}{1-\theta} = 0 \\
\frac{74}{\theta} &= \frac{26}{1-\theta} \\
74 - 74\theta &= 26\theta \\
74 &= 100\theta \\
\hat{\theta} &= 0,74
\end{aligned}$$

Paso 4. Verificamos que sea máximo viendo la segunda derivada:

$$\frac{\partial^2 l(\theta)}{\partial \theta^2} = \frac{-74}{\theta^2} - \frac{26}{(1-\theta)^2}$$

Como θ^2 y $(1-\theta)^2$ son siempre mayores que 0. Para cualquier valor de θ esta expresión es negativa, por lo tanto es máximo.

Más comentes

1. Sean dos muestras de variables aleatorias de tamaño n: A = $(A_1,\ A_2,\ ..,\ A_n)$ y B = $(B_1,\ B_2,\ ..,\ B_n)$, ambas independientes e idénticamente distribuidas en su respectivo conjunto. Si se asume que tanto las variables de A y B distribuyen Exponencial de parámetro λ, el cual es desconocido pero se puede estimar por el método de máxima verosimilitud. Se aplica este procedimiento sobre cada muestra de manera separada, es decir, se maximiza $L(\lambda|\text{A})$ y $L(\lambda|\text{B})$. Y resulta que usando los datos de A se tiene un EMV equivalente λ_A=1 y si se usan las variables B se tiene que el EMV es igual a λ_B=0.5. ¿Cómo explica esta diferencia desde el punto de vista del EMV?

   Respuesta: Si no hay otro factor que se incumpla (por ejemplo los supuestos del MCO, o sesgo por omisión de variable), entonces la clave estadística es que la asignación del tratamiento sea aleatoria, o exógena (exogeno = que el parametro no tenga correlación con el error u otro regresor que no sea la variable dependiente), lo que asegure que la covarianza entre el tratamiento y los

Recordar que el Estimador Máximo Verosímil es el resultado de maximizar la función de verosimilitud L(θ|X), es decir, el parámetro θ es el óptimo que maximiza la probabilidad de ocurrencia o verosimilitud de los datos X que se observan. Por lo tanto, λA es aquel parámetro que maximiza L(λ|A), es decir, la probabilidad de que los datos A ocurran. Del mismo modo que λB maximiza la probabilidad de ocurrencia de los datos B mediante la función objetivo L(λ|B). Es por este motivo que $\lambda_A \neq \lambda_B$, ya que A y B pueden ser muestras distintas.

2. El teorema de Gauss-Markov indica que bajo los supuestos de MCO, entonces el estimador OLS es, dentro de todos los estimadores, aquel con mínima varianza.

   Respuesta: Falso, el teorema sólo aplica sobre estimadores lineales e insesgados. Recordar que la receta"del teorema exige estas tres condiciones para establecer que el estimador obtenido (los beta de la regresión) es el mejor.

3. El método EMV sirve para estimar cualquier parámetro de cualquier distribución, ya que encuentra un máximo en la curva de verosimilitud.

   Solución:

   Si bien es cierto que la estimación por máxima verosimilitud puede ser aplicada en múltiples distribuciones para obtener los valores de los parámetros de interés, es importante recordar que este método considera un problema de optimización, el cual tiene por objetivo maximizar L(θ|X). Por lo tanto, solo se encuentra un θ que maximiza la verosimilitud si es que L(θ|X) es una función cóncava. Lo cual se puede verificar calculando su segunda derivada y corroborando si es mayor a cero.

4. El modelo Logit se utiliza porque no podemos estimar decisiones discretas usando el modelo de regresión lineal.

   Solución:

   Falso. El modelo de regresión lineal permite modelos del tipo $Y = \beta^T X$ con Y una variable binaria. Sin embargo, el problema que ocurre con esto es que las probabilidades asociadas a los valores que puede tomar Y pueden estar fuera del rango [0,1]. Por lo tanto, es conveniente usar modelos de elección discreta en estos casos como Logit o Probit, ya que se encargan de que las probabilidades se calculen correctamente, es decir, que sus valores se mantengan en el intervalo [0,1].

5. Si se desea diseñar un modelo con OLS para evaluar el efecto de un tratamiento, ¿Qué elementos estadísticos son clave para obtener el efecto causal de este?

   Respuesta: Hay tres claves estadísticas. La primera es el cumplimiento de los supuestos del MCO. Lo segundo es que no exista sesgo por omisión de variable, por lo cual se deben encontrar todas las variables que sean relevantes para explicar a la dependiente (y), esto debe cumplirse ya que de esta forma las estimaciones de los coeficientes (beta) serán precisas.

La última y quizás la más relevante, es que la asignación del tratamiento sea aleatoria o exógena, esto significa que asegure que la covarianza entre el tratamiento (regresor binario elegido como tal) y los no-observables sea cero. De esta manera, lo único que distingue al grupo de personas que recibió el tratamiento y los que no lo recibieron, sea el hecho de recibir sea esta asignación.

Con todos estos elementos, la diferencia entre ambos grupos corresponderá al efecto causal del tratamiento.

6. Se estima un modelo Probit para encontrar la probabilidad de que un individuo n escoja usar transporte público durante la pandemia ($Y_n = 1$) versus usar transporte particular ($Y_n = 0$). La función de probabilidad estimada tiene la siguiente forma: P(Y_n) = -0.07 - 2.798 · Wagen + ... Este modelo nos indica que si aumenta el salario de una persona en una unidad entonces disminuye en un 2.7 % la probabilidad de que utilice el transporte público.

   Solución:

   Falso. Ya que el modelo Probit propone el modelo Y = $\phi(\beta^T X)$ con $\phi()$ la función de densidad acumulada de la distribución normal estándar. Por ende, la interpretación de los parámetros no es tan directa como ocurre con el modelo MCO.

Modelos de elección discreta

Una cadena de supermercados dispone de un panel de clientes que registran rutinariamente sus compras en la tienda. En lo que sigue estudiaremos el comportamiento de compra en la categoría ketchup donde hay dos marcas principales: Heinz, Kroger y Hunts. El panel posee información del comportamiento de 447 hogares que realizan 4.887 compras en esta categoría durante 138 semanas de actividad. Más específicamente, el panel describe para cada ocasión de compra la marca elegida, los precios actividad promocional de cada una de las marcas e información demográfica de los panelistas.

Teniendo en cuenta los datos, responda:

1. Formule un modelo de elección discreta que permita obtener la probabilidad de que una persona elija la marca Kroger bajo los atributos disponibles en los datos.

   Solución:

   $$P(Y_{Kroger} = 1) = F(\beta\theta + \beta_{weight}Weight + \beta_{pHeinz}PHeinz + ...)$$

   Con Y_{Kroger} una variable binaria tal que si Y_Kroger=1 el cliente elige comprar Kroger, mientras que si Y_{Kroger}=0, el cliente opta por otra marca distinta.

Y F() es una función de probabilidad que toma $\beta^T X$ y retorna la probabilidad asociada a que un cliente opte por la marca Kroger.

2. Exprese las probabilidades de acuerdo los modelos Logit y Probit.

Solución:

Logit

$$P(Y_{\text{Kroger}} = 1 \mid X) = \frac{e^{\beta_0 + \beta_{\text{weight}} \cdot \text{Weight} + \beta_{\text{pHeinz}} \cdot \text{PHeinz} + \ldots}}{1 + e^{\beta_0 + \beta_{\text{weight}} \cdot \text{Weight} + \beta_{\text{pHeinz}} \cdot \text{PHeinz} + \ldots}}$$

$$P(Y_{\text{Kroger}} = 0 \mid X) = 1 - P(Y_{\text{Kroger}} = 1 \mid X) = \frac{1}{1 + e^{\beta_0 + \beta_{\text{weight}} \cdot \text{Weight} + \beta_{\text{pHeinz}} \cdot \text{PHeinz} + \ldots}}$$

Probit

$$P(Y_{Kroger} = 1|X) = \phi(beta_0 + \beta_{weight} Weight + \beta_{pHeinz} PHeinz + ...)$$
$$P(Y_{Kroger} = 0|X) = 1 - \phi(beta_0 + \beta_{weight} Weight + \beta_{pHeinz} PHeinz + ...)$$