Pontificia Universidad Católica de Chile
Departamento de Ingeniería Industrial y de Sistemas
ICS2563 - Econometría Aplicada
Sección 1

Pauta Interrogación 1

2° semestre 2022 - Profesora Fernanda Ramirez

Nota a la corrección: Esta es una alternativa de pauta, pero no exclusiva, así podrían haber distintas formas de llegar a las mismas respuestas. Si lo explican y desarrollan de manera adecuada, pueden optar al puntaje máximo de todas formas.

Pregunta 1 (15 puntos)

1.1 (5 puntos) Elija un supuesto de la regresión lineal, explíquelo y de un ejemplo en el que este supuesto no se cumpla.

Se pueden escoger cualquiera de los siguientes supuestos:

1. Linealidad en los coeficientes.
2. Rango columna completo en la matriz de diseño.
3. Exogeneidad estricta (en los errores).
4. Varianza de los errores esférica.
5. Distribución normal en los errores.

Para el caso de la pauta, se escogerá el supuesto de rango columna completo en la matriz de diseño. Este supuesto indica que la matriz de diseño, correspondiente a la matriz de variables independientes, posee rango completo, es decir, sea $\boldsymbol{X}$ la matriz de diseño, con dimensión $(n \times p)$, el supuesto indica que Rango$(\boldsymbol{X}) = n$, con $n \geq p$.

Un caso donde el supuesto no se cumpla, es cuando se tienen más variables que observaciones en el modelo, por ejemplo, si se realiza un estudio en un curso de 5 personas, y se tienen los promedios de 7 materias distintas.

(3 puntos) Por explicar correctamente el supuesto mencionado.
(2 puntos) Por dar un ejemplo donde el supuesto no se cumpla.

1.2 (10 puntos) En la regresión lineal de y sobre X (Cuando X contiene una constante) tenemos que: (1) se puede transformar y restándole la media de y, también se puede transformar la matriz X restándole a cada columna su media, y (2) la regresión del y transformado sobre X transformado resulta en los mismos coeficientes de X que la regresión original (y sobre X), excepto por la constante. (*Hint:* en este caso, los resultados de estimar los coeficientes de la regresión original de y sobre X con constante incluida son: $b = (XM^0X)^{-1}X'M^0y$, con M^0 como la definimos en clases).

Indique:

1. ¿Se obtienen los mismos resultados si se transforma sólo y?

2. ¿Se obtienen los mismos resultados si se transforma sólo X?

Muestre el análisis que lo llevó a sus conclusiones.

Si definimos $1 = [1, 1, ..., 1]'$ como un vector de unos, de tamaño n, podemos notar que tenemos $M^0 = I_n - 1(1'1)^{-1}1'$, que es idempotente y simétrica, luego

$$b = ((X'M^{0\prime})(M^0X))^{-1}(X'M^{0\prime})(M^0y) \tag{1}$$

lo que implica que la regresión entre M^0y y M^0X produce el vector de coeficientes sin intercepto. Así:

1. Si sólo X está transformado, tenemos $b_1 = ((X'M^{0\prime})(M^0X))^{-1}(X'M^{0\prime})y$, lo que es idéntico a (1), por lo tanto se obtiene el mismo resultado. Si sólo y está transformado, tenemos $b_1 = (X'X)^{-1}X'M^0y$, lo que es muy distinto a (1), por lo tanto, no se obtiene el mismo resultado.

(2 puntos) Por notar correctamente la definición de $M^0 = I_n - 1(1'1)^{-1}1'$.
(3 puntos) Por obtener el coeficiente sólo con X transformado.
(1 punto) Por concluir que transformar sólo X implica recuperar el coeficiente.
(3 puntos) Por obtener el coeficiente sólo con y transformado.
(1 punto) Por concluir que transformar sólo y implica recuperar un coeficiente distinto.

Pregunta 2 (15 puntos)

2.1 (5 puntos) Enuncie el teorema de Gauss-Markov ¿Qué implica este teorema?

El teorema de Gauss Markov menciona que si en un modelo se cumple (1) linealidad, (2) rango completo, (3) exogeneidad estricta y (4) errores esféricos, entonces el estimador MCO es eficiente respecto a la clase de estimadores lineales insesgados.

El teorema implica que el estimador MCO es insesgado, es decir, su esperanza corresponde al coeficiente, además que tiene varianza mínima entre todos los demás estimadores lineales insesgados, luego es el más eficiente en el sentido del error cuadrático medio. En estadística, aquellos estimadores son los más codiciados, ya que generan intervalos de confianza, asumiendo un buen tratamiento de datos, con largo teóricamente pequeño.

(1.5 puntos) Por mencionar que el estimador MCO es insesgado.
(1.5 puntos) Por mencionar que el estimador MCO es eficiente (o de varianza mínima) en la clase de estimadores lineales insesgados. Puntaje parcial (1 punto) si no se menciona que es en la clase de estimadores lineales insesgados.
(2 puntos) Por explicar la implicancia estadística del teorema.

2.2 (10 puntos) Demuestre cuál es el sesgo cuando hay un error en la medición en uno de los X y cuando hay error de medición en Y. Explicite sus supuestos de ser necesario.

Para mostrar el sesgo podemos utilizar una regresión simple del tipo: $y_i = \alpha + \beta x_i + \mu_i$. Al incorporar un error de medición en x_i, del tipo $x_i' = x_i + s_i$, con s_i un error aleatorio en la medición, tenemos que el coeficiente β puede ser estimado, incorrectamente, según:

$$\begin{aligned} b &= \frac{\text{Cov}(x', y)}{\text{Var}(x')} \\ &= \frac{\text{Cov}(x + s, \alpha + x\beta + \mu_i)}{\text{Var}(x + s)} \\ &= \beta \frac{\text{Var}(x)}{\text{Var}(x) + \text{Var}(s)} \end{aligned} \tag{2}$$

donde $x = (x_1, x_2, ..., x_n)'$, $s = (s_1, s_2, ..., s_n)'$ y $y = (y_1, y_2, ..., y_n)$. Luego, si s_i y μ_i son independientes, existe efectivamente un sesgo en la estimación del coeficiente.

Por otro lado, si queremos estimar una regresión del tipo $y_i = \alpha + \beta x_i + \mu_i$, pero medimos incorrectamente la variable dependiente según $y_i' = y_i + s_i$, entonces el coeficiente estimado:

$$\begin{aligned} b &= \frac{\text{Cov}(y', x)}{\text{Var}(x)} \\ &= \frac{\text{Cov}(y + s)}{\text{Var}(x)} \\ &= \frac{\text{Cov}(\alpha + x\beta + \mu + s)}{\text{Var}(x)} \\ &= \beta \frac{\text{Var}(x)}{\text{Var}(x)} \\ &= \beta \end{aligned} \tag{3}$$

donde $x = (x_1, x_2, ..., x_n)'$, $s = (s_1, s_2, ..., s_n)'$ y $y = (y_1, y_2, ..., y_n)$. Todo lo anterior, es asumiendo que x_i y s_i, μ_i son variables independientes entre sí. Luego no existe un sesgo en la estimación de β, pero tendríamos una regresión del tipo $y_i = \alpha + \beta x_i + \mu_i + s_i$, es decir, el error tendría un término adicional. Esto puede reducir la potencia del test de significancia.

(3 puntos) Por obtener correctamente el sesgo al incluir un error en la medida de x, según (2)
(2 puntos) Por explicitar claramente que s_i y μ_i son independientes.
(3 puntos) Por obtener correctamente que no existe un sesgo en la medición del coeficiente al incluir un error en la medida de y, según (3)
(2 puntos) Por explicitar claramente que x_i y s_i, μ_i son variables independientes entre sí.

Pregunta 3 (15 puntos)

3.1 (5 puntos) Usando la definición matricial de una regresión lineal, derive el vector de coeficientes estimados usando el método de mínimos cuadrados ordinarios (MCO) (*hint:* $e = y - X\beta$).

Buscamos minimizar el cuadrado de la norma 2 del vector de errores, es decir, buscamos minimizar la siguiente función:

$$MCO(\boldsymbol{\beta}) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) = ||\boldsymbol{y} - \boldsymbol{X\beta}||_2^2 \tag{4}$$

Podemos utilizar los conocimientos en álgebra matricial:

$$\begin{aligned} MCO(\boldsymbol{\beta}) &= (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= \boldsymbol{y'y} - \boldsymbol{\beta' X' y} - \boldsymbol{y' X\beta} + \boldsymbol{\beta' X' X\beta} \\ &= \boldsymbol{y'y} - 2\boldsymbol{y' X\beta} + \boldsymbol{\beta' X' X\beta} \end{aligned}$$

Sabemos que podemos minimizar una función utilizando el gradiente e igualando a cero

$$\frac{\partial MCO(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \nabla MCO(\boldsymbol{\beta}) = -2\boldsymbol{X'y} + 2\boldsymbol{X'X\beta} \tag{5}$$

Podemos utilizar (5) para obtener un candidato a mínimo según

$$\begin{aligned} -2\boldsymbol{X'y} + 2\boldsymbol{X'X\beta} &= 0 \\ \boldsymbol{X'X\beta} &= \boldsymbol{X'y} \end{aligned} \tag{6}$$

De esta manera, las ecuaciones (6) se conocen como ecuaciones normales. Verificamos que tenemos un mínimo según el hessiano:

$$\frac{\partial^2 MCO(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = 2\boldsymbol{X'X} \tag{7}$$

Bajo el supuesto que las columnas de la matriz de diseño son independientes, es decir, que la matriz es de rango columna completo, tenemos que la matriz hessiana presentada en (7) es definida positiva, luego tratamos con un mínimo. Además, esto garantiza la existencia de la inversa de $\boldsymbol{X'X}$, luego podemos resolver las ecuaciones normales en (6) para $\boldsymbol{\beta}$:

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

Luego se tiene el resultado que se buscaba.

(1 punto) Por determinar que la ecuación a minimizar es (4)
(3 puntos) Por obtener correctamente el gradiente de la función en (4), i.e., por obtener la ecuación (6)
(1 punto) Por concluir matemáticamente que se trata de un mínimo, i.e. ecuación en (7)

3.2 (5 puntos) Demuestre por qué es necesario incluir un vector de 1's en X para que la suma de los errores sea cero.

Denotemos $\boldsymbol{X}_1 = \boldsymbol{1}_n = [1\ 1\ ...\ 1]'$ como el vector de 1's de dimensión n. Luego,

$$\boldsymbol{X}_1'\boldsymbol{e} = \boldsymbol{1}_n'\boldsymbol{e} = \sum_{i=1}^{n} e_i \tag{8}$$

Como los residuos son ortogonales a la matriz de diseño, tenemos que $\boldsymbol{X}_1'\boldsymbol{e} = 0$, luego $\sum_{i=1}^{n} e_i = 0$.

(2.5 puntos) Por plantear la ecuación (8)
(2.5 puntos) Por utilizar la propiedad de que la matriz de diseño es ortogonal a los residuos para concluir correctamente lo pedido

3.3 (5 puntos) La regresión lineal puede entenderse como descomponer y en una proyección de y sobre x y sobre un vector de residuos y. Derive qué matrices M_1 (proyección) y M_2 (aniquiladora) permiten descomponer y de la manera descrita.

Utilizando el estimador MCO tenemos $\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}$, luego considerando la regresión $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, tenemos que

$$\begin{aligned}\boldsymbol{\epsilon} &= \boldsymbol{y} - \boldsymbol{X\beta} \\ &= \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'y} \\ &= \boldsymbol{y}(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X})\end{aligned}$$

Por lo tanto, definiendo $M_1 = \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X}$ y $M_2 = \boldsymbol{I}_n - M_1$, tenemos

$$\boldsymbol{y} = M_1\boldsymbol{y} + M_2\boldsymbol{y}$$

Luego que M_1 es una matriz de proyección y M_2 es la matriz aniquiladora.

(2 puntos) Por derivar M_1 y M_2
(3 puntos) Por concluir correctamente que M_1 es una matriz de proyección y M_2 es la matriz aniquiladora.

Pregunta 4 (15 puntos)

4.1 (5 puntos) Usando la notación matricial de la regresión lineal, derive el sesgo de la variable omitida, donde Z es la variable que se está omitiendo de la regresión poblacional:

$$Y = X\beta + Z\gamma + \epsilon$$

Nota: Durante la prueba se aclaró que podía no utilizarse la notación matricial, por lo que el resultado también sería válido sin matrices.

Si corremos Y sin incluir la variable relevante Z, tenemos que el estimador MCO b, insesgado, es:

$$b = \beta + (X'X)^{-1}X'(X\gamma) + (X'X)^{-1}X'\epsilon \tag{9}$$

y podemos conocer el sesgo calculando su esperanza:

$$\mathbb{E}[b \mid X, Z] = \beta + (X'X)^{-1}X'(X\gamma) \tag{10}$$

Por lo tanto, el sumando $(X'X)^{-1}X'(X\gamma)$ equivale al sesgo por omisión de variable relevante.

(2 puntos) Por reconocer que el coeficiente tiene una estimación puntual según (9)
(3 puntos) Por derivar correctamente el sesgo por omisión de variable relevante, segun la ecuación (10).

4.2 (5 puntos) Suponga que está estudiando cómo perciben el tiempo de espera los usuarios del transporte público.

Asuma que la relación real entre el tiempo de espera percibido, si el usuario está en horarios punta y la regularidad del servicio de transporte está dado por:

$$\text{espera recibida} = \alpha_1 + \beta_1(\text{horario punta}) + \gamma_1(\text{regularidad servicio}) + \epsilon$$

Pero usted sólo puede estimar la siguiente regresión:

$$\text{espera recibida} = \alpha_2 + \beta_2(\text{horario punta}) + \eta$$

¿Usted cree que el coeficiente β_2 está sesgado? De ser así, describa las características de este sesgo. Si hace supuestos, explicítelos y explique por qué son razonables.

En clases, vimos que existe un sesgo por omisión de variable, y en tal caso estimaríamos

$$\beta_2 = \beta_1 + \gamma_1 \frac{\text{Cov(horario punta, regularidad servicio)}}{\text{Var(horario punta)}} \tag{11}$$

Por lo tanto, podemos discutir las características del sesgo:

1. Si $\gamma_1 = 0$, entonces no hay seso por omisión de variable. Esto sucede si la regularidad en el servicio efectivamente no tiene un efecto sobre la espera recibida. Podemos pensar diferentes situaciones donde un servicio puede ser interrumpido, luego la espera percibida aumenta, es decir, existe una relación entre las variables.
2. Si $\frac{\text{Cov(horario punta, regularidad servicio)}}{\text{Var(horario punta)}} \geq 0$ entonces existe un sesgo positivo, en otro caso sería negativo. Para que exista sesgo a su vez debe existir dependencia entre las variables horario punta y regularidad del servicio. Podemos notar que en el día a día, se intenta que en el horario punta funcione el servicio, pero ¿la falla de un servicio está asociada al horario punta? En caso que si, existiría sesgo. Si no es así, no existiría sesgo.

(2 puntos) Por explicitar que β_2 podría estar sesgado, según la ecuación (11)
(1.5 puntos) Por explicar correctamente que podría existir un sesgo de acuerdo a γ_1.
(1.5 puntos) Por explicar correctamente que podría existir un sesgo de acuerdo a la covarianza entre horario punta y regularidad del servicio.

4.3 (5 puntos) Observe los resultados de una regresión lineal entre el logaritmo natural del ingreso, años de educación, educación de la madre y número de hijos.

	(1) Log(salario)
Años de educación	0.118
	(0.0156307)
Número de hijos	-0.04
	(0.025)
Años de educación de la madre	-0.0203
	(0.0107632)
Constante	-0.0453
	(0.1920918)
N	428

Errores estándar entre paréntesis
$^{+}p < 0.10$, $^{*}p < 0.05$
Valor crítico del t-test al 5% de confianza para esta regresión: 1.96

Responda:

1. Interprete el coeficiente de número de hijos en el ingreso.

   Como es una log-regresión, entonces tener un hijo más implica una disminución promedio en un 4% en los salarios.

   (1 punto) Por interpretar correctamente el valor del coeficiente de acuerdo al contexto. No asignar puntaje si no se adecúa al contexto del problema.

2. Defina una hipótesis nula sobre el coeficiente de número de hijos y calcule el t-test de ese coeficiente.

   Se puede definir, para algún $\gamma \in \mathbb{R}$, el siguiente test de hipótesis:

   $$H_0 : \beta_{\text{número de hijos}} = \gamma \qquad v/s \qquad H_a : \beta_{\text{número de hijos}} \neq \gamma$$

   De esta manera, se puede obtener el test-t según $t = \dfrac{\beta_{\text{número de hijos}} - \gamma}{SE(\beta_{\text{número de hijos}})}$, con $SE(\beta_{\text{número de hijos}})$ el error estándar del coeficiente. Así, utilizando $\gamma = 0$ obtenemos $t = 1.6$

   (1 punto) Por plantear alguna hipótesis nula del coeficiente del número de hijos.
   (1 punto) Por calcular correctamente el t-test asociado a la hipótesis definida anteriormente.

3. Basada/o en la información de la tabla, ¿qué concluiría usted de la influencia del número de hijos en el ingreso?

   Podemos notar que el valor crítico para el t-test está dado por la tabla, equivalente a 1.96, luego como la hipótesis del t-test es de doble cola, tenemos que no rechazamos H_0 si $-1.96 < t < 1.96$ para el t definido en el ítem anterior. Para $\gamma = 0$, tenemos que $t = 1.6$, luego no rechazamos H_0, es decir, no existe evidencia estadística suficiente para decir que el coeficiente asociado al número de hijos pueda ser distinto de 0. En conclusión, utilizando el modelo descrito no podemos concluir que el número de hijos tenga un efecto lineal en el logaritmo de los ingresos.

   (2 puntos) Por concluir que el coeficiente asociado al número de hijos no es significativo.