

Prueba Parcial – Soluciones
Econometría I: Magíster en Economía
Universidad de Santiago de Chile
Semestre 1, 2018

Nombre de alumna/alumno y firma:	
Pregunta 1 (Máximo 12 puntos)	
Pregunta 1 (Máximo 24 puntos)	
Pregunta 2 (Máximo 24 puntos)	

Instrucciones Generales

1. Tiene 90 minutos para responder a esta prueba, dividida en 60 puntos.
2. Póngale nombre a todas las hojas.
3. Sólo el profesor puede responder dudas de enunciado, y sólo en voz alta desde el puesto.
4. Se permite el uso de calculadoras, siempre y cuando no cuenten con dispositivos de comunicación.
5. En caso de copia, se sancionará de acuerdo a lo estipulado en el programa del curso.
6. Las tablas de probabilidad necesarias están incluidas en la prueba. Si no sale la cantidad exacta de grados de libertad necesaria para un cálculo determinado, utilice la cantidad de grados de libertad más cercana.

Buena suerte!!

1. **Repaso de Álgebra Lineal** (12 puntos) Considere la cantidad $\gamma = (X'X)^{-1}X'y$ donde X es una matrix y y es un vector de columna.

a) (4 puntos) En términos generales cuáles son las dimensiones de la matrix X y el vector y . Si no es posible definir la dimensión exacta en algún caso, utilice una letra para referir a la(s) dimensión(es) relevante(s).

Solución: X es una matrix, por ende $N \times k$ (se puede utilizar cualquier dos letras en vez de N y k). y es un vector. Sabemos que y tiene que ser conformable con X' , entonces y es $N \times 1$. Aquí es fundamental que la primera dimensión (N) es lo mismo que la segunda dimensión de X' , y que la segunda dimensión de y es 1 (un vector).

b) (4 puntos) ¿En algún contexto es posible que las dos matrices X' y X no sean conformables? ¿Por qué o por qué no?

Solución: No. Supongamos que X es $N \times k$, donde N y k son cualquier dos números reales enteros positivos. En este caso X' es $k \times N$, y una matrix de $k \times N$ será conformable con una matrix de $N \times k$ para cualquier N . La clave es que la dimensión de columna de la primera matrix es igual a la dimensión de fila de la segunda matrix.

c) (4 puntos) Imagine que ahora separamos la matrix X en una serie de vectores de columna x_1, x_2, \dots, x_k , y planteamos el modelo

$$y = \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_k x_k \quad (1)$$

donde cada γ_k es un valor escalar. Para poder plantear que X es una matrix linealmente independiente, ¿cuáles restricciones tenemos que poder imponer en nuestro modelo 1?

Solución: Para el $y \in R^N$, el grupo de escalares $\gamma_1, \dots, \gamma_k$ son los únicos escalares que existen que resuelvan la ecuación 1. Es decir, necesitamos imponer unicidad (y existencia) de los escalares α .

2. **Estimación (Teoría)** [24 puntos] Se cuenta con una muestra de tamaño N extraída de una distribución normal $\mathcal{N}(\mu, \sigma^2)$. Las observaciones se denotan como y_1, y_2, \dots, y_N , o en formato vectorial simplemente como \mathbf{y} . Cada observación es independiente e idénticamente distribuida. La función de densidad de probabilidad de la distribución normal cada y_i se escribe como $f(y_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right)$. Para simplificar, denotamos el vector de parametros de interés (μ, σ^2) como θ .

a) (11 puntos) Le interesa encontrar el estimador de máxima verosimilitud $\hat{\theta}_{MV} = (\hat{\mu}, \hat{\sigma}^2)$. Escriba la función de verosimilitud para este estimador. [*Pista:* Se sugiere partir con la forma general de una función de verosimilitud, y por último sustituir para el caso específico.] Explique brevemente en palabras cómo procede en cada caso al plantear el estimador

Solución: Partimos escribiendo la fdp conjunta:

$$\begin{aligned} f(y_1, \dots, y_N|\theta) &= f(y_1|\theta) \times \dots \times f(y_N|\theta) \\ &= \prod_{i=1}^N f(y_i|\theta). \end{aligned} \quad (2)$$

Esto nos dice la probabilidad de haber observado el conjunto de datos dado los parametros θ . Queremos algo un poco distinto: la función de verosimilitud, que nos dice la probabilidad de tener parametros θ dado los datos que observamos. Escribimos la función de verosimilitud para capturar esta idea:

$$\mathcal{L}(\theta|y_1, \dots, y_n) = \prod_{i=1}^N f(y_i|\theta). \quad (3)$$

Ahora, sabemos que $f(y_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right)$, y que cada observación es independiente, y por lo tanto, podemos reescribir (3) como:

$$\mathcal{L}(\theta|y_1, \dots, y_n) = \prod_{i=1}^N \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right). \quad (4)$$

Nuestra función de verosimilitud está dada por (4). La función de verosimilitud nos dice cuán probable es que habríamos visto los datos \mathbf{y} dado el parametro θ . Por último, para poder estimar, es más factible tomar lo-

garitmos, para evitar un valor extremadamente pequeño de la función de verosimilitud. Esto nos da la función log-verosimilitud:

$$\begin{aligned}\ell(\mu, \sigma | y_1, \dots, y_N) &= N \ln \left(\frac{1}{\sigma \sqrt{2\pi}} \right) - \sum_{i=1}^N \frac{(y_i - \mu)^2}{2\sigma^2} \\ &= -\frac{N}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^N \frac{(y_i - \mu)^2}{2\sigma^2}.\end{aligned}\quad (5)$$

y maximamos para encontrar los parametros $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$:

$$\hat{\theta}_{MV} = \arg \max_{\theta} \ell(\mu, \sigma | y_1, \dots, y_N).$$

- b) (3 puntos) ¿Cuál es la importancia del supuesto de independencia al formar el estimador de máxima verosimilitud? ¿Su respuesta anterior cambiaría si las observaciones y_1, y_2, \dots, y_N NO fueron independientes entre si?

Solución: La importancia del supuesto de independencia es que nos permite escribir la fdp conjunto como el producto simple que se presenta en la ecuación 2. Si los datos no son independientes entre si, no podemos simplificar la fdp de esta forma, y tenemos que considerar la correlación entre cada observación, y cada otra observación. Esto dificultaría significativamente la producción de un estimador de máxima verosimilitud.

- c) (4 puntos) Explique brevemente (en no más de 5 líneas) la diferencia entre un estimador de máxima verosimilitud y un estimador de método de momentos.

Solución: Cuando estimamos mediante máxima verosimilitud, se basa en el supuesto entero de las variables de interés para formar la función de (log) verosimilitud. Sin embargo, con método de momentos, simplemente aseguramos que los momentos poblacionales supuestos cumplen en la muestra.

- d) (6 puntos) Ahora, en vez de contar con una muestra de una variable para cada unidad $i = 1, \dots, N$, supongamos que se cuenta con dos variables para cada unidad: una variable x y otra y . Así, las observaciones para cada

$i = 1, \dots, N$ se denotan $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$. Y suponga que las variables para cada unidad son distribuidas según una normal bivariada, con la función de densidad de probabilidad $f(x, y)$:

$$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right]\right).$$

Cada unidad es independiente e idénticamente distribuida. ¿Cómo se podrá escribir la función de máxima verosimilitud para estimar los parametros $\hat{\theta}_{MV} = (\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\rho})$?

Solución: La clave aquí es entender que el proceso de estimación sigue siendo exactamente lo mismo que en la parte a) de la pregunta, pero ahora la función de máxima verosimilitud se basa en la normal bivariada. En este caso, escribimos la función de verosimilitud como:

$$\mathcal{L}(\theta|y_1, \dots, y_n) = \prod_{i=1}^N \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\frac{(x_i-\mu_x)^2}{\sigma_x^2} + \frac{(y_i-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y} \right]\right).$$

3. **Estimación (Aplicada)** [24 puntos] Un productor de artículos electrónicos sugiere que la vida útil de sus productos es, en promedio, 8.760 horas. Para una fiscalización del producto, se compra 100 unidades, y para cada uno se mide la vida útil. Se calcula que la vida útil es, en promedio, 6.538 horas para esta muestra, con una desviación estándar en la muestra de 3.600 horas. Suponga que la distribución de la vida útil del producto sigue una distribución normal, y que cada unidad de la muestra de 100 es *iid*.

a) (5 puntos) Calcule un estimador del intervalo de confianza de 95 % para el parámetro poblacional en base a los valores estimados con la muestra de 100 productos.

Solución: Aquí se estima tanto el promedio como la desviación estándar utilizando los datos de la muestra. En este caso, sabemos que podemos formar un estimador de intervalo como:

$$\hat{\mu} \pm t_{(N-1, \alpha/2)}(\hat{\sigma}/\sqrt{N}).$$

Aquí lo importante es poder calcular el $t_{(N-1, \alpha/2)}$, que de las tablas de probabilidad (aproximada con $N - 1 = 99$ (en vez de 100)) da un valor de 1.984. Entonces, utilizando los valores estimados:

$$6,538 \pm 1,984 \times \frac{3600}{\sqrt{100}} = 6,538 \pm 1,984 \times 360 \tag{6}$$

y el estimador de intervalo es [5.824;7.252].

b) (8 puntos) Ahora, imagine que se quiere comprobar la hipótesis nula que el promedio es inferior a 8.760 horas. Complete el test de hipótesis a un nivel de significancia de $\alpha = 0,05$.

Solución: Es importante aquí darse cuenta que el test de hipótesis es un test de una sola cola. Comprobamos la nula $H_0 : \mu \leq 8760$ contra la alternativa $H_1 : \mu > 8760$. Bajo la nula sabemos que $t = \frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{N}}$ se distribuye según una distribución t de Student con 99 grados de libertad. A un nivel de significancia de $\alpha = 0,05$ entonces rechazamos la nula si $t > t_{critico}$,

donde aproximamos el $t_{critico}$ con 99 grados de libertad con el valor para 100 grados de libertad: $t_{critico} = 1,66$. Aquí es importante notar que rechazamos la nula cuando t es más grande que el valor crítico, pero no cuando t es inferior que $-t_{critico}$, ya que valores bajos de $\hat{\mu}$, (y por ende valores bajos de t), son evidencia a favor de la nula. Entonces, calculamos que $t = (6538 - 8760)/360 = -6,17$, y dado que $-6,17 < 1,66$, no rechazamos la nula.

- c) (5 puntos) Explique en palabras cómo se podría comprobar esta misma hipótesis utilizando un test de razón de verosimilitudes. Asegure de indicar la distribución del estadístico de prueba bajo la nula, y el procedimiento general a seguir.

Solución: Aquí vamos a querer estimar dos estimadores de máxima verosimilitud para el vector $\theta = (\mu, \sigma)$. Uno, $\hat{\theta}_{MV}$, donde no imponemos ninguna restricción en los parametros μ y σ , y otro, $\hat{\theta}_r$, donde imponemos que $\mu = 8760$. Si el valor de la función de log-verosimilitud en el modelo restringido $\ell(\hat{\theta}_r)$ es muy inferior al valor de la función de log-verosimilitud en el modelo no restringido, $\ell(\hat{\theta}_{MV})$, esto sugiere que la restricción no es muy realista, y que hay evidencia en contra de la nula que $\mu = 8760$. Específicamente, sabemos que la cantidad $2(\ell(\hat{\theta}_{MV}) - \ell(\hat{\theta}_r))$ sigue una distribución χ^2 , con un grado de libertad. [NOTA DE CORRECCIÓN: No es necesario utilizar la misma notación].

- d) (6 puntos) Un representante de la empresa sugiere que el promedio estimado de 6.538 horas es demasiado bajo, y sugiere estimar el promedio considerando el promedio en los 10 mejores productos, y los 90 peores productos como distintas sub-muestras. Plantea que si se hace así, el promedio de los 10 mejores productos es $\mu_h = 12,280$ y el promedio de los 90 peores productos es $\mu_l = 5900$. Y argumenta que el promedio simple de estos dos valores es 9.090, un valor que supera el promedio sugerido de 8.760. Demuestre que este argumento es inválido. Calcule el promedio real de la vida útil utilizando la ley de esperanzas iteratadas.

Solución: Hay varias maneras de responder a pregunta. Uno es simplemente notando que, sin otras restricciones, el estimador que la persona sugiere pa-

ra el promedio es un estimador *sesgado* del promedio poblacional, mientras el estimador del promedio simple no está sesgado. Para ver esto, notamos que:

$$E[\hat{\mu}] = \sum_{i=1}^N E[Y_i/N] = \sum_{i=1}^N \mu/N = N\mu/N = \mu,$$

y por ende $\hat{\mu}$ está insesgado, pero

$$E[0,9 \times \hat{\mu}_l + 0,1 \times \hat{\mu}_h] = 0,9E[\hat{\mu}_l] + 0,1E[\hat{\mu}_h]$$

que por lo general no está igual a μ . También se puede responder notando que el estimador que sugiere es completamente arbitrario, y consiste en dar una ponderación 9 veces más grande a los productos que funcionan bien que los productos que funcionan mal. Para calcular el promedio utilizando la ley de esperanzas iteradas, hacemos:

$$\begin{aligned} E[\hat{\mu}] &= \sum_{i \in (l,h)} E[\hat{\mu} | \text{producto} = i] \cdot Pr(\text{producto} = i) \\ &= 0,1 \times 12280 + 0,9 \times 5900 \\ &= 6538. \end{aligned}$$

Tab. 2: Valores Críticos de la Distribución t de Student

k	75.0 %	80.0 %	87.5 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
1	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307
45	0.680	0.850	1.165	1.301	1.679	2.014	2.412	2.690	3.281
50	0.679	0.849	1.164	1.299	1.676	2.009	2.403	2.678	3.261
55	0.679	0.848	1.163	1.297	1.673	2.004	2.396	2.668	3.245
60	0.679	0.848	1.162	1.296	1.671	2.000	2.390	2.660	3.232
100	0.677	0.845	1.157	1.290	1.660	1.984	2.364	2.626	3.174
∞	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090