



SEGUNDO PARCIAL
29 de Octubre de 2020

Indicaciones generales

- Este es un examen **individual** con una duración de **110 minutos: de 9:00 a 10:25**. Los 5 minutos entre las 10:25 y 10:30 se destinarán a subir la información de forma correcta.
- Debe entrar al aula virtual del curso (por zoom) y encender la cámara web (puede ser mediante el celular).
- Sólo se contestarán preguntas sobre los enunciados del parcial, durante los primeros 10 minutos.
- Las respuestas deben estar totalmente justificadas. Puede determinar los cuantiles con R, pero debe explicar cómo lo hizo.
- En todos los puntos que use R, debe justificar el procedimiento tanto en papel (no el código) y debe colocar el código usado en un archivo Rmarkdown (o archivo .R) con nombre **NombreApellido.Rmd**, y posteriormente adjuntarlo en **e-aulas**.
- Puede enviar fotos rápidas de los procesos de cada punto al acabar el parcial. Posteriormente, puede enviar las imágenes escaneadas (o foto) con buena calidad de los procedimientos detallados de forma clara y ordenada. El archivo .Rmd (o .R) deben adjuntarlo al acabar el parcial.
- Cualquier incumplimiento de lo anterior conlleva la anulación del examen.

1. [25 ptos.] Sea

$$\underset{(n \times 1)}{\mathbf{Y}} = \underset{(n \times (r+1))}{\mathbf{Z}} \underset{((r+1) \times 1)}{\boldsymbol{\beta}} + \underset{(n \times 1)}{\boldsymbol{\varepsilon}}$$

donde $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ pero $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \mathbf{V}$, con $\mathbf{V}(n \times n)$ conocida y definida positiva. Para \mathbf{V} de rango completo, muestre que el estimador de mínimos cuadrados ponderado es

$$\hat{\boldsymbol{\beta}}_W = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{V}^{-1}\mathbf{Y}$$

Sugerencia: $\mathbf{V}^{-1/2}\mathbf{Y} = (\mathbf{V}^{-1/2}\mathbf{Z})\boldsymbol{\beta} + \mathbf{V}^{-1/2}\boldsymbol{\varepsilon}$ tiene la forma de regresión lineal clásica $\mathbf{Y}^* = \mathbf{Z}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$, con $E(\boldsymbol{\varepsilon}^*) = \mathbf{0}$ y $E(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*'}) = \sigma^2 \mathbf{I}$. Por lo tanto, $\hat{\boldsymbol{\beta}}_W = \hat{\boldsymbol{\beta}}^* = (\mathbf{Z}^*\mathbf{Z}^*)^{-1} \mathbf{Z}^*\mathbf{Y}^*$

2. [25 ptos.] Con los siguientes datos con un sólo predictor z y dos variables respuesta Y_1 y Y_2 :

z	-2	-1	0	1	2
y_1	5	3	4	2	1
y_2	-3	-1	-1	2	3

- a) [7.5 ptos.] Determine las estimaciones de mínimos cuadrados de los parámetros ($\boldsymbol{\beta}$) del modelo regresión múltiple multivariada $\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{E}$
- b) [7.5 ptos.] Calcule las matrices de los valores ajustados $\hat{\mathbf{Y}}$ y los residuos $\hat{\boldsymbol{\varepsilon}}$ with $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & | & \mathbf{y}_2 \end{bmatrix}$ Verifique la descomposición de suma de cuadrados y productos cruzados

$$\mathbf{Y}'\mathbf{Y} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$$

- c) [10 ptos.] Calcule un intervalo de confianza del 95% para la predicción de la respuesta Y_{02} (es decir, de y_2) correspondiente a un nuevo valor de $z_{new} = 0.5$.
Nota: Use la distribución t en vez de la distribución F, es decir, como si fuera una regresión con una variable respuesta.



Los siguientes puntos trabajaran con el dataset `razon_comprar_carro.csv`. Este dataset contiene 90 respuestas para 14 variables diferentes que los clientes consideran al comprar un carro. Las preguntas de la encuesta se enmarcaron utilizando una escala de 5 puntos, siendo 1 muy bajo y 5 muy alto.

3. [25 ptos.] **Análisis de componentes principales.**

- a) [5 ptos.] Obtenga la matriz de correlación muestral \mathbf{R} y determine sus vectores y valores propios.
- b) [5 ptos.] Determine las dos primeras componentes principales de las variables estandarizadas.
- c) [5 ptos.] Determine el porcentaje de la varianza total de las variables estandarizadas capturada por las dos componentes principales.
- d) [5 ptos.] Interprete las dos primeras componentes principales.
- e) [5 ptos.] ¿Son suficientes dos componentes principales? ¿Escogería más componentes? Si es así, ¿por qué?.

4. [25 ptos.] **Análisis de Factores.** Use las variables estandarizadas.

- a) [5 ptos.] Encuentre las estimaciones de máxima verosimilitud de \mathbf{L} y Ψ para $m = 4$ factores (sin rotación).
- b) [5 ptos.] Realice una rotación varimax de las estimaciones de máxima verosimilitud con $m = 4$.
- c) [15 ptos.] Interprete los factores rotados. ¿Qué representa cada factor?