MAT-266: Análisis de Regresión
Certamen 3. Agosto 2, 2021 **Nombre:** ____________________________
Tiempo: 70 minutos **Profesor:** Felipe Osorio

---

1. (20 pts) Considere el estimador

$$\widehat{\boldsymbol{\beta}}(d) = (\boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{I}_p)^{-1}(\boldsymbol{X}^\top \boldsymbol{Y} + d\widehat{\boldsymbol{\beta}}), \qquad d \in [0, 1),$$

donde $\widehat{\boldsymbol{\beta}}$ es el estimador mínimos cuadrados. ¿Es $\widehat{\boldsymbol{\beta}}(d)$ un estimador insesgado? Obtenga la matriz de covarianza de $\widehat{\boldsymbol{\beta}}(d)$.

2. (20 pts) Suponga el modelo $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, donde $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{Z})$ con $\boldsymbol{Z} \in \mathbb{R}^{n\times k}$ tal que $\boldsymbol{Z}^\top \boldsymbol{1} = \boldsymbol{0}$, y

$$\mathsf{var}(\epsilon_i) = \sigma^2, \qquad \mathsf{Cov}(\epsilon_i, \epsilon_j) = \sigma^2\rho,$$

suponga además que σ^2 y ρ son conocidos. Muestre que el estimador máximo verosímil de $\boldsymbol{\beta} = (\delta, \boldsymbol{\theta}^\top)^\top$ asume la forma

$$\widehat{\boldsymbol{\beta}} = (\widehat{\delta}, \widehat{\boldsymbol{\theta}}^\top)^\top, \qquad \text{con} \qquad \widehat{\delta} = \overline{Y}, \qquad \widehat{\boldsymbol{\theta}} = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\boldsymbol{Z}^\top \boldsymbol{Y}.$$

3. (30 pts) Considere $\boldsymbol{e} \sim \mathsf{N}(\boldsymbol{0}, \sigma^2(\boldsymbol{I} - \boldsymbol{H}))$ con $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ y $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ tal que $\operatorname{rg}(\boldsymbol{X}) = p$.

a) Pruebe que

$$s_{(i)}^2 = s^2\Big(\frac{n - p - r_i^2}{n - p - 1}\Big).$$

b) Obtenga la distribución marginal de e_i, $i = 1, \dots, n$, y de ahí deduzca que

$$\frac{e_i^2}{\sigma^2(1 - h_{ii})} \sim \chi^2(1).$$

c) Usando el teorema de Cochran, muestre la independencia entre $s_{(i)}^2$ y e_i^2.

4.a. (10 pts) Suponga el modelo lineal:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i, \qquad i = 1, \dots, n.$$

Sea $r = k + 1$ el número de parámetros y suponga que se desea probar que $r - p$ parámetros son cero. Considere F_p el estadístico F correspondiente. Muestre que:

$$F_p = 1 + \frac{C_p - p}{r - p}.$$

4.b. (20 pts) Considere un modelo de regresión lineal (con intercepto) $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ con $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ $\boldsymbol{\epsilon} \sim \mathsf{N}_n(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. Muestre que el estadístico F para probar la hipótesis $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$ se puede escribir como:

$$F = \frac{R^2/(p - 1)}{(1 - R^2)/(n - p)},$$

donde R^2 es el coeficiente de deteminación. De este modo, sigue que $F \sim \mathsf{F}(p - 1, n - p)$. Con lo anterior, muestre que R^2 sigue una distribución Beta con parámetros $p - 1$ y $n - p$, respectivamente. Además, obtenga $\mathsf{E}(R^2)$.