TAREA 2: MAT-269

Profesor: Ronny Vallejos

Problema 1[40 puntos]. Considere el conjunto de datos `diag2010.txt`. En este problema se analizará las variables Mate: Puntaje en la parte matemática de la PSU de los alumnos de primer año de la USM, y la variable Calculo: puntaje en la prueba de diagnóstico aplicada a los estudiantes de primer año en la USM el año 2010.

a) Haga un análisis descriptivo de ambas variables por separado.

b) ¿Es razonable asumir normalidad para cada variable? Justifique.

c) Determine si el supuesto de normalidad multivariada tiene asidero estadístico. Justifique.

d) Ajuste una distribución bivariada a los datos. Estime la media y la matriz de covarianzas.

e) Haga inferencia para el coeficiente de correlación entre las variables. Comente. ¿Existe correlación significativa entre la prueba de diagnóstico aplicada en la USM y la parte matemática de la PSU? Proponga un modelo para explicar una variable como función de la otra.

f) ¿Existen Outiers multivariados en la muestra? Justifique.

g) El instrumento consiste en 40 preguntas clasificadas en 4 categorías. Estas categorías tienen relación con los 4 cursos remediales que se ofrece a aquellos estudiantes que obtienen un puntaje menor a 620 puntos. El porcentaje de respuestas acertadas en cada categoría se encuentra en la base de datos con los nombres R1, R2, R3 y R4. ¿Es razonable asumir una distribución normal multivariada para modelar las variables R1, R2, R3 y R4?

h) Use el test T^2 de Hotelling para docimar la hipótesis de igualdad de las medias de las variables PSU y la prueba de diagnístico aplicada en la USM el año 2010 versus una hipótesis alternativa bilateral. Use $\alpha = 0.05$

Problema 2[15 puntos]. Sea $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ una muestra aleatoria desde una población $\mathcal{N}(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma}_0)$, donde μ es un vector desconocido, σ^2 es un escalar desconocido y $\boldsymbol{\Sigma}_0$ es una matriz conocida. Obtenga el estimador de máxima verosimilitud de μ y σ^2.

Problema 3[15 puntos]. Sea $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ una muestra aleatoria desde una población $N(\boldsymbol{\mu}, \boldsymbol{I}_p)$. Si $\boldsymbol{\mu}$ está definida sobre la esfera unitaria, demuestre que el estimador de máxima verosimilitud de $\boldsymbol{\mu}$ es

$$\widehat{\boldsymbol{\mu}} = \frac{\overline{\boldsymbol{X}}}{\sqrt{\overline{\boldsymbol{X}}^\top \overline{\boldsymbol{X}}}}.$$

Problema 4[15 puntos]. Suponga que un coeficiente de correlación muestral igual a 0.65 es observado en una muestra de tamaño 10. Contraste la hipótesis de independencia versus la hipótesis alternativa de correlación positiva.

Problema 5[15 puntos]. Use el test z de Fisher para contrastar la hipótesis $H_0 : \rho = 0.7$ versus $H_1 : \rho \neq 0.7$ con una significancia del 5%, con $r = 0.5$ y $N = 50$.