

# Segundo Examen Parcial

## Análisis Estadístico Multivariado

ITESO

Departamento de Matemáticas y Física

Nombre:

Fecha:

### 1. Regresión Lineal Simple y Múltiple

Elija una base de datos para proponer un modelo de regresión simple y un modelo de regresión múltiple y conteste las siguientes preguntas:

- a) ¿Qué supuestos debe cumplir un modelo que describa la relación lineal entre dos variables? Describa en qué consiste cada uno de ellos.
- b) Con base a los datos que eligió, escriba un enunciado planteando el objetivo o el problema a resolver.
- c) Obtenga el modelo de regresión simple y escriba su ecuación. De una descripción de cómo estos parámetros impactan sobre la variable dependiente.
- d) ¿Por qué es importante separar los datos en 80% para entrenamiento y 20% para prueba?
- e) De una interpretación de los resultados obtenidos (Summary)
- f) ¿Cuál sería el valor de  $T_{\text{tablas}}$  con el que contrastaría el valor de  $T_{\text{calculada}}$  si se tuviera un nivel de significancia del 0.05?
- g) Indique si el modelo lineal se ajusta a los datos basado en las predicciones obtenidas. Justifique su respuesta.
- h) Si no existiera normalidad y homocedasticidad, ¿qué puede concluir de los resultados del análisis? ¿Qué solución propone ante la falta de normalidad y heterocedasticidad?
- i) ¿Qué son los “outliers”? ¿Cómo influyen en el análisis de regresión? ¿qué solución propone ante la presencia de estos valores?
- j) Escriba un enunciado planteando el problema a resolver.
- k) Obtenga el modelo de regresión lineal múltiple y escriba su ecuación. De una descripción de cómo estos parámetros impactan sobre la variable dependiente.
- l) Interprete los resultados obtenidos (Summary)
- m) ¿Qué modelo de regresión es mejor, el simple o el múltiple? Justifique su respuesta.

## 2. Análisis de la Varianza

Elija una base de datos para realizar un ANOVA de dos factores y responda las siguientes preguntas:

- a) ¿Qué supuestos se deben cumplir para realizar el ANOVA?
- b) ¿Cuál es el objetivo del ANOVA?
- c) Indique las variables que va a utilizar y especifique quién es la variable dependiente y las variables independientes
- d) Escriba un enunciado planteando el problema a resolver. (Plantear la  $H_0$  y  $H_1$  del problema a resolver).
- e) ¿Cuántos niveles tiene cada factor que eligió?
- f) Obtenga el modelo del ANOVA y de una interpretación de los resultados.
- g) ¿Cuál sería el valor de  $F_{\text{tablas}}$  con el que contrastaría el valor de  $F_{\text{calculada}}$  si se tuviera un nivel de significancia del 0.05?
- h) ¿Para que sirven las pruebas Post-Hoc en el ANOVA?
- i) De ser necesario, incluya la prueba de Tukey y de la interpretación de los resultados.
- j) Conclusiones generales de análisis de la varianza que acaba de realizar.

## 3. Análisis de Componentes Principales

Elija una base de datos para realizar el análisis de componentes principales y responda las siguientes preguntas:

- a) ¿Qué supuestos se deben de cumplir para realizar un Análisis de Componentes Principales?
- b) ¿Cuál es el objetivo de realizar el análisis de componentes principales?
- c) ¿Qué es lo que sucede si las variables que se están utilizando para realizar el PCA no están correlacionadas? Justifique su respuesta.
- d) ¿Bajo qué circunstancias se recomienda hacer una estandarización de los datos?
- e) ¿Qué representan los vectores propios de la matriz de varianzas y covarianzas en el análisis de componentes principales?
- f) ¿Qué representan los valores propios de la matriz de varianzas y covarianzas en el análisis de componentes principales?
- g) Obtenga la matriz de vectores propios y de una interpretación de los resultados.
- h) En el análisis de componentes principales que está realizando ¿Cuál es número óptimo de componentes principales? Justifique su respuesta
- i) ¿Cuáles son las variables que más influyen en los primeros dos componentes principales?
- j) Conclusiones generales del análisis de componentes principales que acaba de realizar.