Parcial 1

José Fernando Zea

Probabilidad y estadística aplicada

Universidad Konrad Lorenz.

6/4/2020

El parcial deberá ser presentado en un R markdown compilado en Word, cada punto deberá separarse de manera clara. Se evaluarán los procedimientos, resultados y las conclusiones y comentarios sobre cada punto los cuales deberán ser claros y con buena redacción..

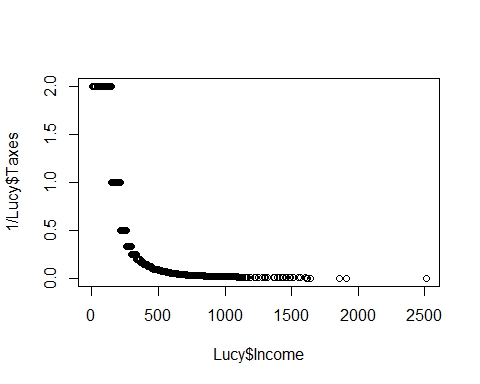
Fecha de entrega el día 16 de abril por correo electrónico (colocar asunto, parcial estadística). El parcial es de carácter individual

## Parcial (Parte teórica)

1. De las siguientes afirmaciones. ¿Cuál es verdadera?
2. El promedio es una cantidad que nunca puede ser negativa, tampoco se deja afectar por valores atípicos.
3. La mediana se calcula tomando el dato de la mitad si el número de datos es par, es una cantidad más inexacta que el promedio y su interpretación es muy complicada.
4. El promedio puede afectarse por valores atípicos, por otro lado, si existe mucha variabilidad en los datos, no es una buena medida resumen para la mayoría de los datos.
5. La mediana es robusta a datos atípicos, sin embargo, explica la variabilidad de los datos y no es una medida de tendencia central.
6. El promedio y la mediana son exactamente la misma cantidad, sólo que el promedio se calcula promediando los valores ordenados.
7. Explique en sus propias palabras que es el coeficiente de variación y como se interpreta.
8. Calcule la desviación absoluta media de los siguientes cinco datos: 2, 6, 8, 9, 14

Interprete esta medida.

1. Cuanto debe dar la correlación del siguiente conjunto de datos:



1. Cercana a cero
2. -1
3. Cercana a 1
4. Negativa
5. La siguiente tabla de contingencia corresponde al análisis de las empresas que mandan SPAM como estrategia de publicidad en cinco zonas:

table(Lucy$SPAM, Lucy$Level)

##   
## Big Medium Small  
## no 26 291 620  
## yes 57 446 956

Calcule la tabla de perfiles y analiza si son más propensas las empresas de tamaño más pequeñas de enviar SPAM.

1. Es la rentabilidad de una acción una variable:
2. Cualitativa Ordinal
3. Cualitativa nominal
4. Cuantitativa discreta
5. Cuantitativa continua.
6. No se puede cuantificar.

## Parcial práctico (valor 80%)

Use para elaborar la base titanic la base de datos Titanic:

library(readr)  
library(dplyr)  
setwd("su ruta")  
  
titanic <- read\_csv("titanic.csv")  
summary(titanic)

titanic <- titanic %>% select(PassengerId,Survived, Age,  
 Fare,  
 Pclass)  
titanic <- na.omit(titanic)  
table(titanic$Pclass, useNA = "always")

##   
## 1 2 3 <NA>   
## 186 173 355 0

summary(titanic$Age)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.42 20.12 28.00 29.70 38.00 80.00

1. Calcule las frecuencias absolutas y relativas de los sobrevivientes. Concluya
2. Calcule la edad promedio y mediana de los pasajeros del titanic. También calcule el promedio y mediana del pago del tiquete (Fare) Concluya.
3. ¿Calcule la desviación estándar y coeficiente de variación de la edad y el pago de tiquete? Existe una variabilidad alta para estas dos variables?
4. Cómo es la distribución de la edad y el pago del tiquete. Elabore un histograma y calcule la asimetría para estas dos variables. Concluya con los resultados obtenidas.
5. Existe una alta correlación entre el pago del tiquete y la edad de la persona.
6. Elabore un diagrama de caja comparando la clase del tiquete con el valor del tiquete, elabore sus conclusiones e interpretaciones del gráfico.
7. Elabore un diagrama de caja comparando el valor del tiquete con la sobrevivencia en el Titanic, elabore sus conclusiones e interpretaciones del gráfico.
8. Elabore una tabla de contingencia y un análisis de perfiles de la clase y la sobrevivencia en el Titanic ¿Sobrevivieron mas los de clase I, que puede decir de los de clase II y clase III?
9. Las pruebas saber tuvieron el año pasado en el 2019 un promedio de 300 puntos en el puntaje global, la desviación estándar fue de 20 puntos. Es razonable pensar que los datos siguen una distribución normal (el histograma es muy similar al de un conjunto de datos simulado por una normal).
   1. ¿Cuál es la probabilidad de que un estudiante seleccionado al azar saque más de 350 puntos.
   2. Calcule el percentil 25 y 75 de la distribución normal.
10. Investigue el uso de la distribución multinomial, esplique al detalle su uso, sus propiedades básicas y brinde un ejemplo de aplicación.