**Quizzes Eco Computacional**

Semana 3 – High Dimensional Inference

Objetivos de aprendizaje:

* Inferencia clásica: Parámetro vs estimador; Sesgo, ECM, Normalidad, Pruebas de hipótesis, pruebas múltiples,
* Big data inference: FDP, FDR

Preguntas:

1. ¿Por qué los estimadores son variables aleatorias?
   1. Porque varía dependiendo de cuantas variables use para estimar el parámetro
   2. Porque varía dependiendo de cuantas observaciones use en mi base de datos
   3. No son variables aleatorias, son un número fijo (i.e beta = -0.5)
   4. **Porque varían dependiendo de la muestra de la que viene mi base de datos**
2. Conforme sube el tamaño de mi muestra, mi estimador:
   1. **Tiene mayor probabilidad de ser igual al parámetro, menor varianza y se distribuye normal alrededor del parámetro**
   2. Tiene mayor probabilidad de ser igual al parámetro, mayor varianza y se distribuye normal alrededor del parámetro
   3. Tiene mayor probabilidad de ser igual al parámetro, menor varianza y no se distribuye normal alrededor del parámetro
   4. Tiene menor probabilidad de ser igual al parámetro, menor varianza y se distribuye normal alrededor del parámetro
3. Intuitivamente, el estadístico F calcula:
   1. **La proporción de varianza total de la variable y que es explicada por la regresión vs la varianza no explicada por la regresión, ajustando por cuantas variables incluí en la regresión**
   2. La proporción de varianza total de y que es explicada por la regresión vs la varianza total de la variable y, ajustando por cuantas variables incluí en la regresión
   3. La proporción de varianza total de y que no es explicada por la regresión vs la varianza explicada por la regresión, ajustando por cuantas variables incluí en la regresión
4. En Big data, ¿por qué las pruebas de hipótesis simples y múltiples arrojan muchos falsos positivos?
   1. Porque los estadísticos T y F son crecientes al número de observaciones en una regresión; eso hace que salgan muchas pruebas significativas
   2. Porque los estadísticos T y F son crecientes al número de variables incluidas en una regresión; eso hace que salgan muchas pruebas significativas
   3. Porque los estadísticos T y F son decrecientes al número de variables incluidas en una regresión; eso hace que salgan muchas pruebas significativas
   4. Porque los estadísticos T y F son crecientes al número de observaciones en una regresión; eso hace que salgan muchas pruebas significativas
5. Una variable incluida en una regresión de 130 variables tiene un p-value de 0.03. Este es el vigésima variable con menor p-value de las 130. Es un falso positivo para un FDR(0.05)?
   1. Si
   2. **No**

k=20, q=0.05, p=0.03

kq/N = 20\*0.04/0.05 = 0.04

Semana 4 – Regresión

Objetivos de aprendizaje:

* Aprendizaje supervisado y no supervisado, in and out of sample, paramétrico y no paramétrico
* Gauss-Markov, omitted variable bias, selección de variables
* Regresión para predicción

Preguntas:

1. Supervised vs un supervised
2. In and out of sample (overfitting)
3. Agrego una variable no relevante. Que pasa con beta
4. Calcula la matriz de varianza covarianza (X’X) para la siguiente base de datos:
   1. Y=(-3, 0, 2, 1.5), X1=(1,2,3,4), X2= (-5, 5, 0, 3)

X’ = , X = , X’X= , X’Y =

1. f