Material para las clases teóricas virtuales

Clase 5: El modelo lineal general en notación matricial. (Link a lista de reproducción)

1. Presentación en pdf para imprimir disponible en este link

2. Videos (son 5 videos en total):
   - **Video 1 – El modelo en notación matricial**
     Se presenta la formulación matricial del modelo lineal general o modelo con K variables.

     Preguntas guía para el video 1:

     a) En la página 3 de las filminas dice que el modelo puede escribirse como un sistema de n ecuaciones lineales ¿lineales en qué?

     b) Escribir los vectores y matrices de la página 4 para el caso en que K=3 y n=4.

     c) Desarrollar la operación $X\beta + u$. Verificar que el resultado es un vector 4x1, donde cada elemento de ese vector corresponde al modelo lineal para cada una de las observaciones.

   - **Video 2 – Repaso de álgebra matricial**
     Se repasan algunas definiciones y resultados de álgebra matricial y de reglas de derivación de expresiones que involucran matrices y vectores.

     Preguntas guía para el video 2:

     a) Argumentar por qué una matriz de dimensión $n \times K$, con $n \neq K$, no puede tener inversa.

     b) En nuestro caso, siempre $n > K$ (esto no lo habíamos dicho antes). Entonces, del argumento en (a) se desprende que X no tiene inversa. ¿Cuál es el mayor rango que puede tener la matriz X?

     c) ¿Qué es lo que garantiza que X tenga rango columna completo? ¿Qué implica esto acerca de la relación entre las variables explicativas?

     d) Escribir la matriz X para el caso en que $K = 3$ y $n = 4$. Obtener la matriz $(X'X)$ y prestar atención a su composición (es decir, cómo se obtienen los distintos elementos de la matriz). ¿Qué rango tiene esta matriz y bajo qué supuesto? ¿Tiene inversa?

   - **Video 3 – Obtención de los MCO con matrices**
     Se resuelve el problema de minimización de SRC y se obtiene el vector de estimadores MCO usando notación matricial. También se repasan algunas propiedades algebraicas de los estimadores MCO en notación matricial.

     Preguntas guía para el video 3:

     a) A partir del minuto 1:07 del video 3, la profesora dice algo así: “$X\beta$ es el modelo de regresión, es la parte lineal, observable, del modelo”. ¿Cuál es el error?

b) Verificar que entienden todos los pasos hasta la obtención de $\hat{\beta} = (X'X)^{-1}X'Y$. Recodar: es útil ir chequeando la dimensión de las matrices mientras hacen las demostraciones; sirve como forma de detectar errores.
c) Escribir la matriz X y el vector Y para $K = 2$ y n genérico (es decir, no fijar un tamaño muestral particular). Para ese caso, operar para obtener el siguiente producto matricial: $(X'X)^{-1}X'Y$. Verificar que el resultado obtenido es un vector 2x1 que contiene las fórmulas para los estimadores MCO de la ordenada al origen y la pendiente del modelo con 2 variables que vimos en la clase 2 (ecuaciones 3 y 4 de las filminas de la clase 2).
d) Demostrar las propiedades algebraicas 2 y 3 (páginas 20 y 21 de las filminas). Mostrar que, tal como dice en la página 20, la propiedad 2 implica que la suma de los errores de estimación es cero y que la covarianza entre cada uno de los regresores y los errores de estimación es cero. ¿Qué pasaría si el modelo no incluyera una ordenada al origen?

- **Video 4 – R cuadrado y R cuadrado ajustado**
  Se presenta la notación matricial del R cuadrado. Se discute que el R cuadrado no decrece con el agregado de variables explicativas y se propone el R cuadrado ajustado como forma de resolver el trade-off entre la mejora del ajuste y la pérdida de eficiencia ante el agregado de variables explicativas.

  Preguntas guía para el video 4:
  a) Mostrar que la STC puede escribirse como $Y'Y - n(\bar{Y})^2$.
  b) ¿Por qué decimos que no tiene sentido basarnos en el R cuadrado para comparar la bondad del ajuste de modelos que difieren en el número de variables explicativas?
  c) Quiero saber qué porcentaje de la variabilidad de Y es explicada por mi modelo. ¿Qué uso, el R cuadrado o el R cuadrado ajustado?
  d) Decir si puede darse lo siguiente y por qué: (i) se agrega una nueva variable al modelo y el R cuadrado no aumenta, (ii) se agrega una nueva variable que no resulta estadísticamente significativa y el R cuadrado ajustado aumenta.

- **Video 5 – Supuestos clásicos en notación matricial**
  Se presentan los supuestos clásicos en notación matricial. Se introduce el concepto de matriz de varianzas y covarianzas de un vector aleatorio. Se prueban las propiedades estadísticas del vector de estimadores MCO (insesgadez y la expresión de la matriz de varianzas y covarianzas).

  Preguntas guía para el video 5:
  a) El segundo supuesto clásico en notación matricial es $VarCov\ (u) = \sigma^2 I_n$. Escribir esa matriz y mostrar que resume los supuestos de homocedasticidad y no correlación serial.
  b) ¿Bajo qué supuestos podemos escribir $VarCov\ (u) = E(uu')$?
  c) Repetir la demostración de insesgadez del vector de estimadores MCO (página 32 de las filminas), justificando detalladamente cada uno de los

pasos. ¿Cuál es el supuesto clave en esta demostración? ¿Qué pasa con la propiedad de insesgadez si $VarCov\ (u) \neq \sigma^2 I_n$?

d) Repetir la demostración de la obtención de la matriz de varianzas y covarianzas del vector de estimadores MCO (página 33 de las filminas), justificando detalladamente cada uno de los pasos. ¿cuál es el supuesto clave en esta demostración?

e) Obtener la matriz $VarCov\ (\hat{\beta}) = \sigma^2 (X'X)^{-1}$ para el caso en que $K = 2$. Chequear que el elemento en la fila 2 y columna 2 es justamente la varianza del estimador de la pendiente de un modelo con 2 variables, cuya fórmula ya la habíamos visto en la clase 3 (ver página 34 de las filminas de la clase 3).