Actividad 1

Ejercicio CEF.

a) Si $e = Xu$, donde X y u son independientes $N(0,1)$. Demuestre que, condicional en X, el error tiene distribución $N(0, X^2)$
b) Lo anterior qué significa en terminos de la independencia entre e y X (vea BH, p.24-25). Relacione esto con el concepto de homocedasticidad.
c) Pruebe que: $var[Y] \geq var[Y - [E|X]] \geq var[Y - [E|X_1, X_2]]$ (vea BH 2.33)
d) Pruebe que Q_{XX} es una matriz positiva semidefinida (vea BH p.39)
e) Consider the linear projection:

$$\mathcal{P}[log(wage)|X] = 0.046experience - 0.07experience^2 + 0.11education + 2.3$$

Formally (using derivatives), what is the effect of one extra year of experience on log(wage)?

f) Consider the linear projection:

$$\mathcal{P}[log(wage)|X] = 0.046experience - 0.07experience^2 - 0.09female + 0.11education - 0.07education * female + 1.06$$

Formally (using derivatives), what is the return to one extra year of education on log(wage) for males? and for females? see BH(p. 31)

h) Draw a two-dimensional graph showing the linear projections onto education by gender.

i) Consider the linear projection:

$$\begin{aligned}\mathcal{P}[log(wage)|X] = 0.046experience - 0.07experience^2 - 0.09female + 0.05education - 0.04education * female \\ +0.05north - 0.03north * female + 0.08north * education - 0.05north * education * female + 0.98\end{aligned}$$

where *north* identifies with 1 the population living in northern Mexico and zero otherwise.

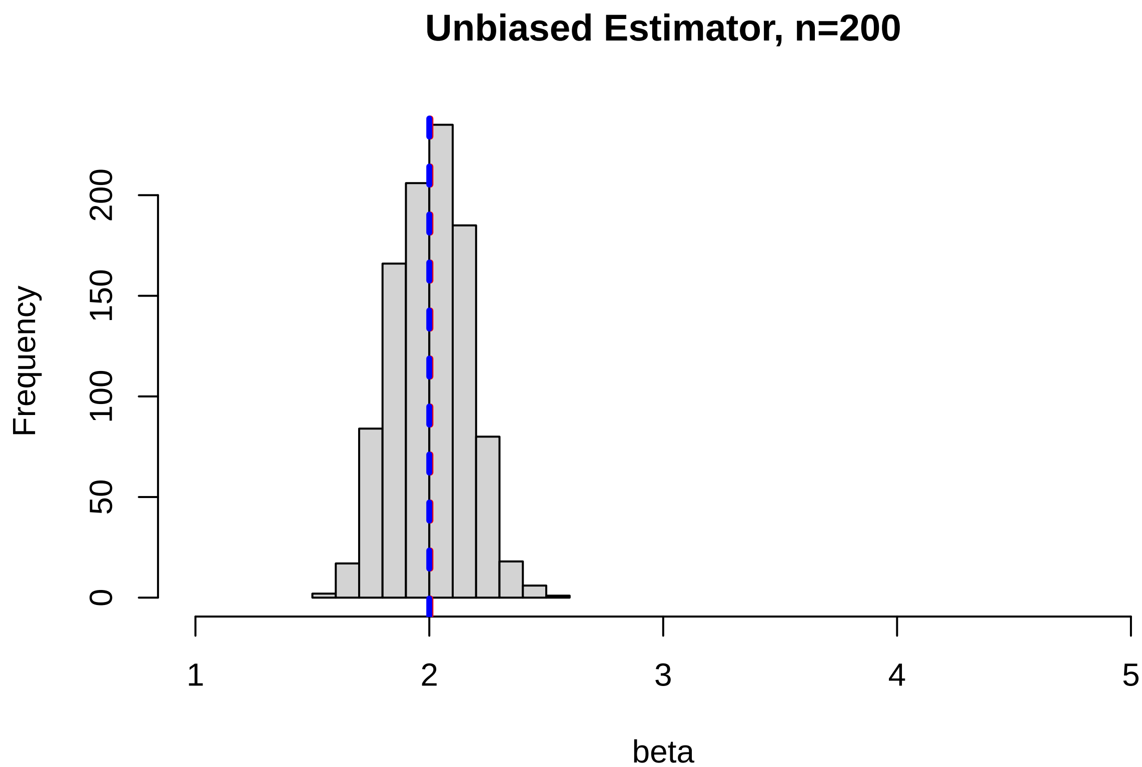
Formally (using derivatives), what is the return to one extra year of education on log(wage) for males in the north of Mexico? and for females in the south? see BH(p. 31)

j) This is a piece of code simulating an unbiased estimator β_1 and β_2.

```r
repet <- 1000
n <- 200
beta <- NULL

for (i in 1:repet){
  x <- rnorm(n)
  x2 <- rnorm(n)
  u <- rnorm(n,0,1)
  y=2+2*x+2*x2+u
  beta[i] <- lm(y~x)$coef[2] #
}
hist(beta, main="Unbiased Estimator, n=200", xlim = c(1,5) )
```

```r
abline(v = mean(beta), col="red", lwd=3, lty=2,)
abline(v = 2, col="blue", lwd=3, lty=2)
```



- Modify the code so that it shows a **biased** estimator of β_1.
- While keeping this last modification make another (different) modification to the code so that you can get rid of the bias (see BH 2.24)

k) Give a concrete example of a linear projection on which we most likely suffer from OVB.

Ejercicio LS.

a) Use la base de datos de sueldos provista por Wooldridge:

```r
library(wooldridge)
data("wage1")
```

- Construya los vectores X con educación, experiencia y experiencia al cuadrado, el vector Y es logaritmo del sueldo. Obtenga:

$$\hat{\beta} = \hat{Q}_{XX}^{-1}\hat{Q}_{XY}$$

- Muestre explícitamente el vector $\sum_{i=1}^{n} X_i Y_i$ y la matriz $(\sum_{i=1}^{n} X_i X_i')^{-1}$
- Interprete β_1 y el efecto de la experiencia en sueldos.
- ¿Son los *residuos* ortogonales a la educación? ¿Son los errores ortogonales a la educación? Explique.
- Demuestre *formalmente* si existe sesgo del estimador $\hat{\beta}_1$.
- Obtenga ESS y TSS así como R^2 e interpretela.

b) Muestre que si $X = [X_1, X_2]$ entonces $PX_1 = X_1 y MX_1 = 0$

c) En R-Studio genere dos parámetros x y $residuo$ (n=100, mean=0 y sd=5) donde $y < -15 + (7 * x) + residuo$

- Realice una regresión de x en y.
- Realice un scatterplot de la regresión estimada que muestre cada observación Y_i y la línea LS.
- Obtenga ESS y TSS así como R^2.
- Demuestre, modificando los datos creados en b), que si RSS aumenta R^2 disminuye. Discuta si $\hat{\beta}$ es insesgada.
- Cree una variable x_2 y agreguela a la regresión estimada en b) Qué sucede con ESS, TSS y R^2
- Modifique su código para sesgar $\hat{\beta}_1$ considerablemente y estime de nuevo la misma regresión ¿Cambia esto R^2? Discuta.

e) Considere dos regresiones por LS:

$$Y = X_1\tilde{\beta} + \tilde{e}$$

$$Y = X_1\hat{\beta} + X_2\hat{\beta}_2 + \tilde{e}$$

Donde R_1^2 y R_2^2 corresponden a cada regresión respectivamente. Muestre que $R_1^2 \geq R_2^2$ ¿cuál es el caso en que serían iguales?

f) A dummy variable takes on only the values 0 and 1. It is used for categorical variables. Let D_1 and D_2 be vectors of 1's and 0's, with the $i_t h$ element of D_1 equaling 1 and that of D_2 equaling 0 if the person is a man, and the reverse if the person is a woman. Suppose that there are n_1 men and n_2 women in the sample.

Consider fitting the following equations by OLS:

$$Y = \mu + D_1\alpha_1 + D_1\alpha_2 + e$$

$$Y = D_1\alpha_1 + D_1\alpha_2 + e$$

$$Y = \mu + D_1\phi_1 + e$$

- Can all three equations be estimated by OLS? Explain if not.
- In the OLS regression $Y = D_1\hat{\gamma}_1 + D_2\hat{\gamma}_2 + \hat{u}$ show that $\hat{\gamma}_1$ is the sample mean of the dependent variable among the men of the sample ($\bar{Y}_1$), and that $D_2\hat{\gamma}_2$ is the sample mean among the women (Y_2).