

Actividad 2

Ejercicio 1. Propiedades Finitas.

- a) Suopnga que $E[e|X] = 0$ y $var[e|X] = \Omega$ pruebe que:

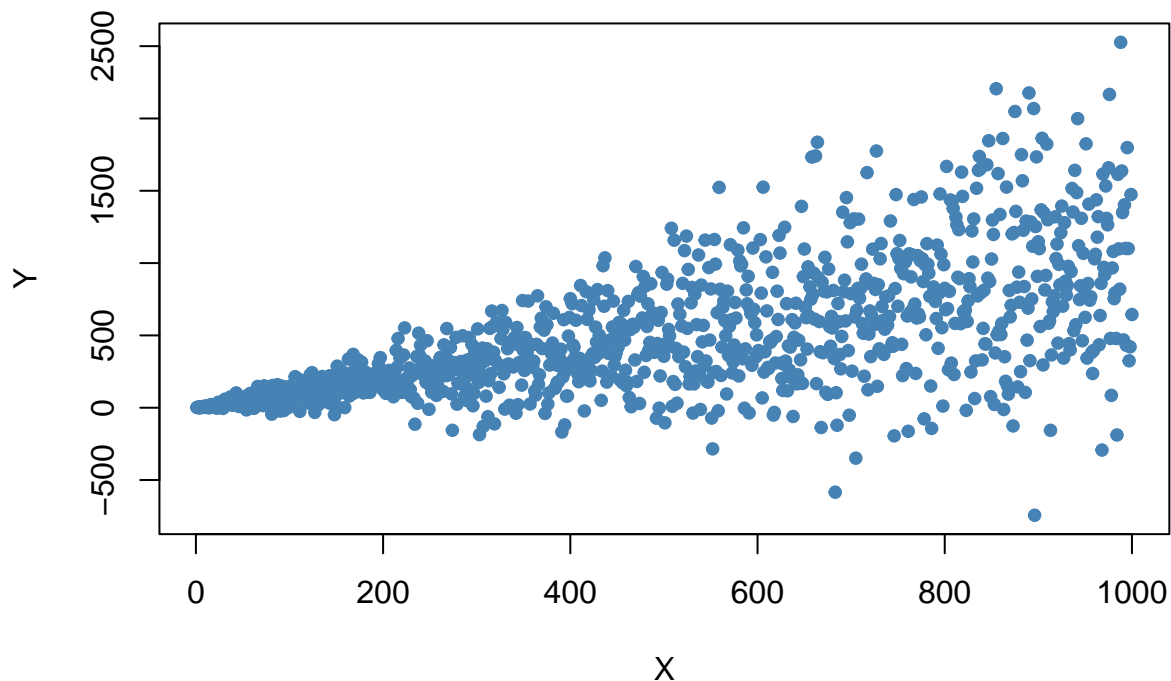
$$E[\hat{\beta}|X] = \beta;$$

$$var[\hat{\beta}|X] = (X'X)^{-1}(X'\Omega X)(X'X)^{-1}$$

- b) The parameter β is esimated by OLS $\hat{\beta} = (X'X)^{-1}X'Y$ and GLS $\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$. Let $\hat{e} = Y - X\hat{\beta}$ and $\tilde{e} = Y - X\tilde{\beta}$ denote the residuals. How would you compute \hat{R}^2 and \tilde{R}^2 . If the error e_i is truly heteroskedastic will \hat{R}^2 or \tilde{R}^2 be smaller?
- c) Weighted Least Squares is a special case of GLS. Where $var(e_i|X) = \sigma^2 h(X)$ for $h(X) > 0$. So the functional form of heteroskedasticity is known. Show that $var(e_i h_i^{-1/2}) = \sigma^2$. Is WLS efficient?
- d) Take the linear homoskedastic CEF $Y^* = X'\beta + e$; $E[e|X] = 0$; $E[e^2|X] = \sigma^2$. Suppose that Y^* is measured with error. Instead, we observe $Y = Y^* + u$ where u is a measurement error, e and u are independent and $E[u|X] = 0$; $E[u^2|X] = \sigma_u^2(X)$. Describe the effect if any of this measurement error on variance calculation for $\hat{\beta}$.
- e) Muestre que $E[\hat{V}_{\hat{\beta}}^{HC0}|X]$ se encuentra sesgado bajo homoscedasticidad (vea BH 113). Junto con esto, muestre que $HC2$, bajo homoscedasticidad, es insesgado.
- f) Siga el ejemplo en R abajo. Obtenga la proyección lineal de Y en X utilizando OLS. Encuentre los errores estándar robustos (hint: lmtest package) y muestre que, empíricamente, $HC1 < HC2 < HC3$

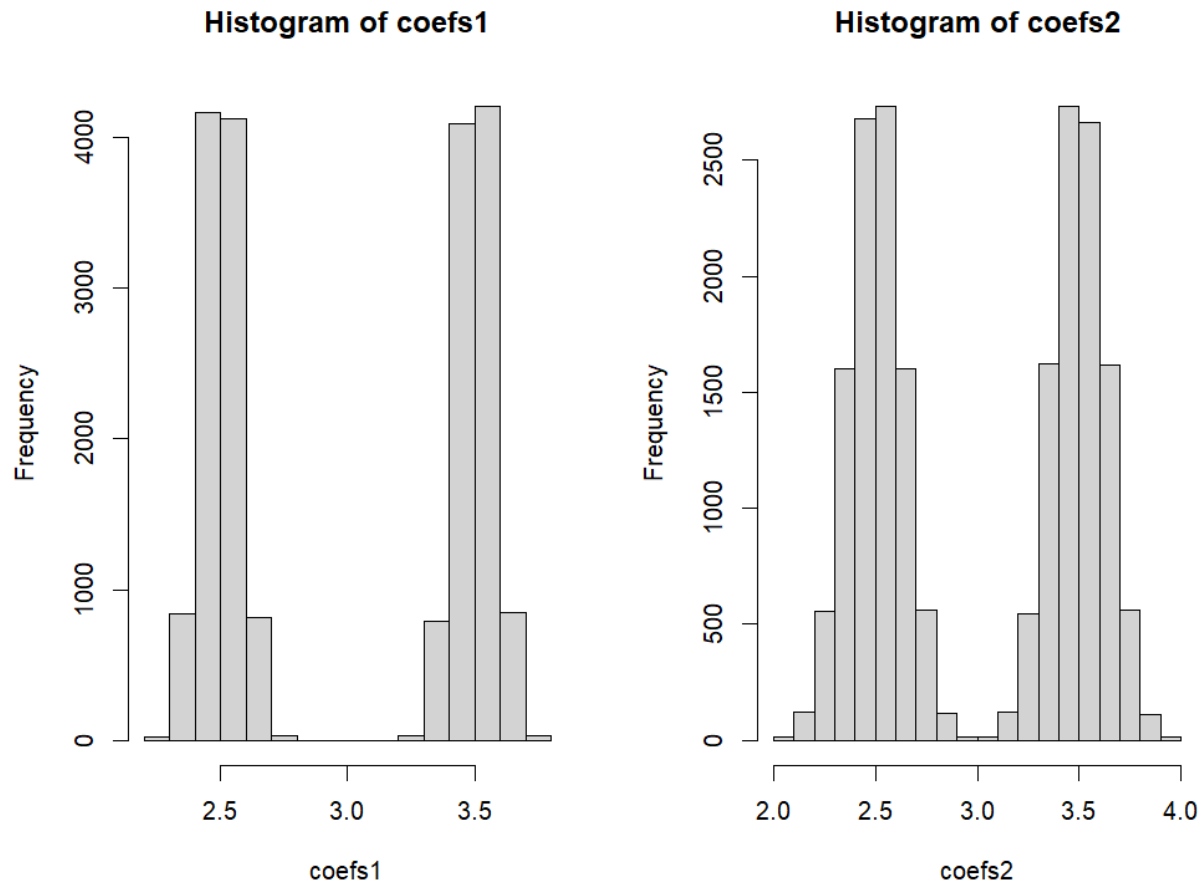
```
# generate heteroskedastic data
X <- 1:1000
Y <- rnorm(n = 1000, mean = X, sd = 0.6 * X)

# plot the data
plot(x = X, y = Y,
     pch = 19,
     col = "steelblue",
     cex = 0.8)
```



- g) Utilice la base “cps09mar.txt” y siga el código provisto en R en BH p.91. Obtenga los errores HC0 a HC3.
- Modifique h_{ii} para que tenga un valor cercano a .4 y obtenga de nuevo HC0 a HC3 ¿a qué conclusión llega?
- h) Obtenga en forma matricial (similar a como hizo en g) los errores estándar en cluster utilizando la base de datos “DDK2011.txt”.
- i) Carga el conjunto de datos ‘PublicSchools’ desde la paquetería ‘sandwich’. Ajusta un modelo de regresión lineal simple usando OLS donde ‘Expenditure’ es la variable dependiente e ‘Income’ la variable independiente.
- Realice un diagnóstico de la regresión (pueden usar funciones como `plot()`).
 - Interprete los resultados de sus diagnósticos ¿Hay algún estado que influya desproporcionadamente en el modelo? ¿Cómo podría modificar el modelo para manejar estas influencias?
- j) Carga el conjunto de datos ‘Journals’ de Stock y Watson (2007), que contiene información sobre suscripciones a revistas de economía en bibliotecas de EE. UU. desde la paquetería ‘AER’.
- Genere un nuevo **data frame** que contenga solo las variables de interés: número de suscripciones (subs), precio de suscripción (price) y el precio por cita calculado (citeprice).
 - Transforme las variables subs y citeprice usando el logaritmo natural para su análisis.
 - Ajuste un modelo de regresión lineal usando OLS donde el logaritmo del número de suscripciones se explique por el logaritmo del precio por cita.
 - Realice un diagnóstico del modelo de regresión.
 - Interprete los resultados de los diagnósticos.
 - Reporte los errores estándar correctos para los coeficientes estimados.

- k) En el contexto de una encuesta nacional sobre la satisfacción con los servicios de salud pública, se ha utilizado un diseño de muestreo por conglomerados. Los investigadores seleccionaron inicialmente ciertas áreas geográficas (unidades primarias de muestreo) y luego realizaron una selección aleatoria de hogares dentro de estas áreas para las entrevistas. Discuta por qué es importante utilizar errores estándar agrupados en la estimación de los coeficientes de un modelo de regresión que analice los determinantes de la satisfacción con los servicios de salud.
- l) Con base en el ejemplo abajo, ejecute dos simulaciones con $n=500$ y $10,000$ repeticiones que reproduzcan, en la medida de lo posible, los siguientes gráficos:



Example: Here we create a matrix X with two vectors (regressors X_1 and X_2) with median 50 and 100 each, and $cov(X_1, X_2) = 0.25$

```
library(MASS)
X <- rmvnorm(n, c(50, 100), sigma = cbind(c(10, 2.5), c(2.5, 10)))
summary(X)
plot(X[,1], X[,2])
```

Explique en qué factor aumenta la varianza de acuerdo con los parametros establecidos n , σ^2 y ρ siguiendo la definición de $var[\hat{\beta}|X]$ en BH p.121.

Ejercicio 2. MLE.

- a) Show that plugging in the estimators β_{mle}^2 and $\hat{\sigma}_{mle}^2$ into:

$$l_n(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - X_i' \beta)^2$$

We get the maximized log likelihood function (see BH p.142-143). How does it work as a measure of fit?

- b) Suponga que $y_i | \mathbf{x}_i \sim \mathcal{N}(m(\mathbf{x}_i, \beta_0), \sigma_0^2)$, donde $m(\mathbf{x}, \beta)$ es una función del vector de variables explicativas \mathbf{x} y del vector de parámetros β de dimensión $(k \times 1)$. Entonces, $E(y_i | \mathbf{x}_i) = m(\mathbf{x}_i, \beta_0)$ y $V(y_i | \mathbf{x}_i) = \sigma_0^2$.
 - Escriba la función de log verosimilitud condicional para la observación i . Muestre que el estimador de máxima verosimilitud $\hat{\beta}$ resuelve el problema de minimización $\min_{\beta} \sum_i (y_i - m(\mathbf{x}_i, \beta))^2$.
- c) Sea $\theta \equiv (\beta' \sigma^2)'$ un vector de parámetros de dimensión $(k+1) \times 1$. Encuentre el vector score para la observación i . Muestre que $E(s_i(\theta_0) | \mathbf{x}_i) = \mathbf{0}$.
 - Usando las condiciones de primer orden, encuentre $\hat{\sigma}^2$ en términos de $\hat{\beta}$.
 - Encuentre la matriz hessiana de la función de log verosimilitud con respecto a θ .
- d) Show that, under $\epsilon N(0, \sigma^2)$ MLE estimators of β and σ^2 are equivalent to OLS.
- e) In the normal regression model show that the leave-one out prediction errors \tilde{e}_i and the standardized residuals \bar{e}_i are independent of $\hat{\beta}$, conditional on \mathbf{X} . Hint: Use (3.45) and (4.24) in Hansen.
- f) Ejecute el siguiente código en R que sirve como un ejemplo para obtener SE robustos a heterocedasticidad y los t-test de hipótesis correspondientes.

Ejemplo:

```
#Environment:
install.packages("AER")
library(AER)
data("CPSSWEducation")
attach(CPSSWEducation)
?CPSSWEducation

reg <- lm(earnings ~ education)
summary(reg)

# plot observations and add the regression line
plot(education,
     earnings,
     ylim = c(0, 150))

abline(labor_model,
       col = "steelblue",
       lwd = 2)

# compute homoskedastic-robust standard errors.
t <- linearHypothesis(reg, "X = 0")$'Pr(>F)'[2] < 0.05

# compute heteroskedasticity-robust (HC1) standard errors
t.rob <- linearHypothesis(reg, "X = 0", white.adjust = "hc1")$'Pr(>F)'[2] < 0.05

# show both t-tests, where 1="true" meaning Ho is "true" at the 95% level.
round(cbind(t = mean(t), t.rob = mean(t.rob)), 3)

# Same for the varcovar matrix
vcov <- vcovHC(labor_model, type = "HC1")
```

```
vcov

# compute the square root of the diagonal elements in vcov
robust_se <- sqrt(diag(vcov))
robust_se

# we use `coeftest()` on our robust model and show the assumed-homoskedastic model:
coeftest(labor_model, vcov. = vcov)
summary(labor_model)
```

- Genere datos heteroscedasticos para dos variables X e Y (como en 1.f) los cuales cumplen con la ecuación poblacional $Y = \alpha + \beta X + e$, donde $\beta = 1$
- Realice un scatterplot de estas variables con una linea ajustada mostrando su relación.
- Ejecute 10,000 regresiones, cada una con nuevas muestras aleatorias provenientes de la misma distribución establecida en (1.f).
- Para cada $i=1 \dots 10,000$ ejecute los test-t “HCO” y “HC2” y almacénalos en un vector “t” y “t.rob”
- Compute el porcentaje de rechazos de la hipótesis nula $\beta = 1$
- ¿Cuál es la conclusión de este resultado?

g) MLE is the technique that helps us determine the parameters of the distribution that best describe the given data. Imagine that we have a sample that was drawn from a normal distribution with mean $\mu = 5$ and variance $\sigma^2 = 100$. The objective is to estimate these parameters with MLE.

The normal log-likelihood function is given by:

$$l = -\frac{1}{2}n\ln(2\pi) - \frac{1}{2}n\ln(\sigma^2) - \frac{1}{2\sigma^2} \sum (y_i - \mu)^2$$

*Note that minimizing a negative likelihood function is the same as maximizing the likelihood function.

```
#we define:

X <- rnorm(n = 1000, mean = 5, sd = 10)
df <- data.frame(X)

#We program the log-likelihood function in R:

normal.lik1 <- function(theta,y){
  mu<-theta[1]
  sigma2<-theta[2]
  n<-nrow(y)
  logl<- -.5*n*log(2*pi)-.5*n*log(sigma2)-(1/(2*sigma2))*sum((y-mu)^2)
  return(-logl)
}

#Here theta is a vector containing the two parameters of interest
#(i.e. theta[1] is equal to mu). The remainder sets n, and the log-likelihood function.

# we use optim(starting values, log-likelihood, data) with starting values 0 and 1.
optim(c(0,1), normal.lik1, y=df)

#We can ask for the method-of-moments-mean directly:
mean(X)
var(X)
```

- Now, estimate the MLE parameters β and σ^2 with $Y = 5 + 2X + e$.

Bis) Lets try to estimate “manually” a poisson regression using MLE (hint: Google the log-likelihood function of a conditional Poisson Distribution):

```
# Parameters of the Poisson distribution
lambda <- 2 # Mean parameter of the Poisson distribution

# Generate Poisson-distributed data
doctor_visits <- rpois(n = 2950, lambda = lambda)
hist(doctor_visits)

education <- data.frame(CPSSWEducation$education)

nll <- function(theta0,theta1) {
  x <- education
  y <- doctor_visits
  mu = exp(theta0 + x*theta1)
  -sum(y*(log(mu)) - mu) #Why is this?
}

#optimize using the stats4 library
stats4::mle(minuslogl = nll, start = list(theta0 = 2, theta1 = 0))
```

- Estimate the same model using the GLM library, and compare the results with the “manual” optimization.
- Estimate the same model using OLS and make a histogram of the residuals Think: Do they distribute as expected for OLS estimation to be equivalent to MLE?