Resolucion_Wooldridge

Parte 1 capitulos 1,2

Ignacio Sepulveda

2020

Capitulo 1: La naturaleza de la econometria y de los datos economicos.

Problemas.

1.1 Suponga que se le pide que realice un estudio para determinar si grupos de clase pequeños contribuyen a un mejor desempeño de los estudiantes de cuarto grado.

i) Si pudiera realizar cualquier experimento que deseara, ¿qué haría? Explique con claridad.

R: Tomaria a todos los grupos de estudiantes de cuarto año que existan los separaria en dos grupos perfectamente distinguibles donde uno corresponde a un grupo grande y otro chico, les aplicaria una prueba estadarizada para medir cual es el efecto del tamaño de la sala sobre el rendimiento, aplicado eventualmente todo los controles.

ii) Siendo más realistas, suponga que puede obtener datos observacionales de varios miles de estudiantes de cuarto grado de un determinado estado. Puede conocer el tamaño de sus grupos y las calificaciones estandarizadas obtenidas en el examen final. ¿Por qué puede esperarse una correlación negativa entre el tamaño de los grupos y las puntuaciones en el examen final?

R: Existe correlacion negativa dado que a mayor tamaño podria existir menor personalizacion, o por otro lado facilita, en general, la posibilidades de desorden o de distraccion.

iii) Una correlación negativa, ¿indicaría necesariamente que tamaños de grupo menores causan un mejor desempeño?

R: No, falta controlar por otros determinantes para asi aislar correctamente el efecto tamaño

1.2 Para justificar los programas de capacitación laboral se ha dicho que éstos mejoran la productividad de los trabajadores. Suponga que se le pide que evalúe si una mayor capacitación para el trabajo hace que los trabajadores sean más productivos. Pero, en lugar de que se le proporcionen datos sobre trabajadores individuales, se le facilitan datos de fábricas en Ohio. De cada firma se le proporcionan horas de capacitación laboral por trabajador (capacitación) y la cantidad de artículos no defectuosos producidos por hora por cada trabajador (producción).

i) Establezca cuidadosamente el experimento ceteris paribus subyacente a esta pregunta.

R: Dado que queremos evaluar el efecto de la capacitación sera necesario tener dos grupos, unos que le fue aplicado el tratamiento o capacitacion y otro que no y tener sus productividades. Para poder establecer la causalidad ademas tendremos que controlar por otros factores como la tecnologia puntual de la fabrica.

ii) ¿Parece razonable que la decisión de una empresa de capacitar a sus trabajadores sea independiente de las características de los mismos? ¿Cuáles son algunas de esas características medibles y no medibles de los trabajadores?

R: No, debiese estar condicionado a la naturaleza de su cargo y respectiva industria en que se encuentra. Es probable que el retorno marginal de la capacitacion, al menos en la mayoria de industria, disminuya a medida que aumento en jerarquía organizacional.

iii) Nombre un factor, que no sea una característica de los trabajadores, que influya en la productividad de los trabajadores.

R: La tecnologia de la fabrica.

iv) Si encontrara una correlación positiva entre producción y capacitación, ¿habría establecido de manera convincente que la capacitación para el trabajo hace que los trabajadores sean más productivos? Explique.

R: No, falta aislar la relacion, por lo que es necesario controles.

1.3 Suponga que en su universidad se le pide que encuentre una relación entre horas semanales de estudio (estudio) y horas semanales de trabajo (trabajo). ¿Tendría sentido considerar que en este problema se trata de inferir si estudio "causa" trabajo o trabajo "causa" estudio? Explique.

No tiene mucho ya que hay una limitacion practica dado que $estudio + trabajo + ocio = 24$

Ejercicios en computadora

C1.1 Para este ejercicio emplee la base de datos WAGE1.RAW.

i) Determine el nivel educativo promedio de la muestra. ¿Cuáles son los niveles de educación menor y mayor?

```r
wage1$educ %>%
  summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   12.00   12.00   12.56   14.00   18.00
```

ii) Determine el salario promedio por hora (wage) en la muestra. ¿Parece ser alto o bajo?

```r
wage1$wage %>%
  mean()
```

```
## [1] 5.896103
```

iii) Los datos de los salarios están dados en dólares de 1976. Usando el Economic Report of the President (de 2004 o posterior o el Informe de Gobierno en países de habla hispana) obtenga y dé los índices de precios al consumidor (IPC) correspondientes a 1976 y 2003.

iv) Use los valores de los IPC del inciso iii) para determinar el salario promedio por hora en dólares de 2003. ¿Parece ahora más razonable el salario promedio por hora?

v) ¿Cuántas mujeres (females) hay en la muestra? ¿Cuántos hombres?

```
wage_1=wage1
wage_1$female=ifelse(wage1$female==1,'Mujer','Hombre')
wage_1 %>% group_by(female) %>%
  summarise(Total=n())
```

```
## # A tibble: 2 x 2
##   female Total
##   <chr>  <int>
## 1 Hombre   274
## 2 Mujer    252
```

C1.2 Para responder estas preguntas emplee la base de datos BWGHT.RAW.

```
data("bwght")
```

i) ¿Cuántas mujeres hay en la muestra (male=0) y cuántas de las informantes fumaron durante un embarazo?

```
bwght %>%
  filter(male==0,cigs>0) %>%
  select(male) %>%
  summarise(Total=n())
```

```
##   Total
## 1   112
```

ii) ¿Cuál es la cantidad promedio de cigarros consumidos por día (cigs)? ¿Es el promedio, en este caso, una medida representativa de la mujer “típica”? Explique.

```
bwght$cigs %>%
  summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.000   2.087   0.000  50.000
```

No parece ser representativa outliers que fumaron hace de contrapeso de las que fumaron nada.

iii) Entre las mujeres que fumaron durante el embarazo, ¿cuál es la cantidad promedio de cigarros consumidos por día? ¿Cuál es la relación de esto con su respuesta al inciso ii) y por qué?

```
bwght %>%
  select(cigs) %>%
  filter(cigs>0) %>%
  summarise(Media=mean(cigs))
```

```
##      Media
## 1 13.66509
```

Confirma la tesis de que las que fumaban estan en el extremo de la distribuion

iv) Determine el promedio de fatheduc (años de educación del padre) en la muestra. ¿Por qué se emplean sólo 1 192 observaciones para calcular este promedio?

```
bwght$fatheduc %>% summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    1.00   12.00   12.00   13.19   16.00   18.00     196
```

Vemos que ese valor de debe a que existen 196 NA's

v) Dé el ingreso familiar promedio (famine) y su desviación estándar en dólares.

```
bwght %>% select(faminc) %>% summarise(Media=mean(faminc), SD=sd(faminc))
```

```
##      Media       SD
## 1 29.02666 18.73928
```

C1.3 Los datos de MEAP01.RAW pertenecen al estado de Michigan en el año 2001. Emplee estos datos para contestar las preguntas siguientes.

```
data("meap01")
```

i) Determine los valores mayor y menor de math4. ¿Es lógico este intervalo? Explique.

```
meap01$math4 %>% summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   61.60   76.40   71.91   87.00  100.00
```

Si dado que corresponde a un test con valores en $[0, 100]$

ii) ¿Cuántas escuelas tienen una tasa perfecta de aprobados en el examen de matemáticas? ¿A qué porcentaje del total de la muestra corresponde esta cantidad?

```
meap01 %>%
  filter(math4==100) %>%
  select(math4) %>%  summarise(Perfect_score=n(),'Percent(%)'=(Perfect_score/length(meap01$math4))*100)
```

```
##   Perfect_score Percent(%)
## 1            38   2.084476
```

iii) ¿En cuántas escuelas la tasa de aprobados en matemáticas es exactamente 50%?

```r
meap01 %>%
  filter(math4==50) %>%
  select(math4) %>%  summarise('50_score'=n())
```

```
##   50_score
## 1       17
```

iv) Compare el promedio de las tasas de aprobados en matemáticas y en lectura. ¿Cuál de estas pruebas es más difícil de aprobar?

```r
meap01 %>%
  select(math4,read4) %>%
  summarise('Mean_math4'=mean(math4,na.rm=TRUE),'Mean_read4'=mean(read4,na.rm=TRUE))
```

```
##   Mean_math4 Mean_read4
## 1     71.909   60.06188
```

v) Encuentre la correlación entre math4 y read4. ¿Qué concluye?

```r
cor(meap01$math4,meap01$read4)
```

```
## [1] 0.8427281
```

Que es positiva y alta.

vi) La variable exppp es gasto por alumno. Determine el promedio y la desviación estándar de exppp. ¿Parece haber una gran variación en el gasto por alumno?

```r
meap01 %>% select(exppp) %>% summarise('mean'=mean(exppp,na.rm=TRUE), 'sd'=sd(exppp,na.rm=TRUE))
```

```
##       mean      sd
## 1 5194.865 1091.89
```

vii) Suponga que la escuela A gasta 6 000 dólares por alumno y la escuela B gasta 5 000 dólares por alumno. Dé el porcentaje en el que el gasto de la escuela A supera al gasto de la escuela B. Compare este porcentaje con $100 \cdot [\log(6\,000) - \log(5\,500)]$, que es la diferencia porcentual aproximada basada en la diferencia de los logaritmos naturales. (Veáse la sección A.4 del apéndice A.)

```r
dif_1=100*(1-(5000/6000))
dif_2=100*(log(6000)-log(5000))
dif_1-dif_2
```

```
## [1] -1.565489
```

C1.4 La base de datos de JTRAIN2.RAW proviene de un experimento de capacitación para el trabajo realizado para hombres con bajos ingresos durante 1976-1977; véase Lalonde (1986).

```
data("jtrain2")
```

i) Emplee la variable indicadora train para determinar la proporción de hombres a los que se les dio capacitación para el trabajo.

```
jtrain2 %>%
  select(train) %>%    summarise('Prop_capacitados(%)'=sum(train)/length(train)*100)
```

```
##   Prop_capacitados(%)
## 1            41.57303
```

ii) La variable re78 es ingresos desde 1978, dados en dólares de 1982. Determine el promedio de re78 para la muestra de hombres a los que se les dio capacitación laboral y para la muestra de hombres a los que no se les dio. ¿Es esta diferencia económicamente grande?

```
jtrain2_1=jtrain2
jtrain2_1$train=ifelse(jtrain2$train==1,'Capacitado','No_Capacitado')
jtrain2_1 %>%
  group_by(train) %>%
  summarise('Mean_78'=mean(re78))
```

```
## # A tibble: 2 x 2
##   train         Mean_78
##   <chr>           <dbl>
## 1 Capacitado       6.35
## 2 No_Capacitado    4.55
```

Es una diferencia bastante razonable.

iii) La variable unem78 indica si un hombre estuvo desempleado o no en 1978. ¿Qué proporción de los hombres a los que se les dio capacitación para el trabajo están desempleados? ¿Y de aquellos a los que no se les dio capacitación laboral? Comente la diferencia.

```
jtrain2 %>%
  filter(train==1) %>%
  select(unem78) %>% summarise('Prop_desempleados(%)'=sum(unem78)*100/length(unem78))
```

```
##   Prop_desempleados(%)
## 1             24.32432
```

iv) Con base en los incisos ii) y iii), ¿parece haber sido efectivo el programa de capacitación laboral? ¿Qué haría que nuestra conclusión fuera más convincente?

Parece haber un mayor sueldo promedio en los que reciben la capacitacion ademas de haber llegado a un grupo de desempleados. Esto sumado a que podrian existir otras variables que influencien tal relacion.

Una conclusion mas convincente seria tomar demostrar que:

$E(Salario|Capacitado) > E(Salario|NoCapacitado)$ y asumiendo que hemos puesto todos los controles.

Capitilo 2: El modelo de regresión simple.

Problemas

2.1 Sea niños la cantidad de hijos que ha tenido una mujer, y educ los años de educación que tiene esta mujer. Un modelo sencillo para relacionar fertilidad con años de educación es

$niños = \beta_0 + \beta_1 * Educ + u$, donde u es no observado.

i) ¿Qué tipo de factores son los contenidos en u? ¿Es posible que estos factores estén correlacionados con el nivel de educación?

R: Factores contenidos en u son en general sociodemograficos como estructura familiar, estado civil, etc. Y es probable que esten correlacionados con el nivel de educacion ya que tambien constituyen parte de sus determinantes.

ii) ¿Es posible que con un análisis de regresión simple se halle el efecto ceteris paribus de educación sobre fertilidad? Explique.

No, dado que podrian existir no observables que efectivamente esten covariando con educacion y produzcan problemas en el estimador. Ejemplo de esto seria estado civil esperamos que condional en que es casada tiene en promedio un esperanza de niñoz mas alta que las que no, entonces aplicar estos controles permitiria mejorar en cuanto a propiedades del estimador.

2.2 En el modelo de regresión lineal simple $y = \beta_0 + \beta_1 x + u$, suponga que $E(u) \neq 0$. Sea $\alpha_0 = E(u)$, muestre que siempre es posible reescribir el modelo con la misma pendiente, pero con otro intercepto y otro error, de manera que el nuevo error tenga valor esperado cero.

- $E(u) = \alpha_0$
- $E(y - \beta_0 - \beta_1 x) = \alpha_0$
- $E(y) - \beta_0 - \beta_1 E(x) = \alpha_0$, dado que $E(xu) = 0$
- $E(y) - \gamma_0 - \beta_1 E(x) = 0$, donde $\beta_0 + \alpha_0 = \gamma_0$
- $E(y - \gamma_0 - \beta_1 x) = 0$
- $E(u^*) = 0$

2.3 En la tabla(pag.61) siguiente se presentan las puntuaciones obtenidas en el examen de ingreso a la universidad en Estados Unidos, ACT (American College Test), y en el GPA (promedio escolar) por ocho estudiantes universitarios. El GPA está medido en una escala de cuatro puntos y se ha redondeado a un dígito después del punto decimal.

```
library(tidyverse)
datos=data.frame(Estudiante=seq(1,8,1),
                 GPA=c(2.8,3.4,3,3.5,3.6,3,2.7,3.7),
                 ACT=c(21,24,26,27,29,25,25,30))
datos %>% knitr::kable()
```

Estudiante	GPA	ACT
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25
7	2.7	25
8	3.7	30

i) Estime la relación entre GPA y ACT empleando MCO; es decir, obtenga las estimaciones para la pendiente y para el intercepto en la ecuación

$\widehat{GAT} = \hat{\beta}_0 + \hat{\beta}_1 ACT$

Comente la dirección de la relación ¿tiene, en este caso, el intercepto una interpretación útil? Explique, ¿qué tanto más alto será el GPA predicho si ACT aumenta cinco puntos?

R: Ocupando una estimacion por medio del metodos de momentos,

Si estimamos por Metodo de momentos.

- $E(u) = 0 \rightarrow E(GAT - \beta_0 - \beta_1 ACT) = 0$
- $E(ACTu) = 0 \rightarrow E(ACT(GAT - \beta_0 - \beta_1 ACT)) = 0$

Por lo tanto sus contraparte muestrales estan dadas por:

- $n^{-1} \sum_{i=1}^{n} (GAT_i - \hat{\beta}_0 - \hat{\beta}_1 * ACT_i) = 0$
- $n^{-1} \sum_{i=1}^{n} ACT_i (GAT_i - \hat{\beta}_0 - \hat{\beta}_1 * ACT_i) = 0$
- $\hat{\beta}_0 = \overline{GAT} - \hat{\beta}_1 \overline{ACT}$
- $n^{-1} (\sum_{i=1}^{n} GAT_i ACT_i - \hat{\beta}_0 \sum_{i=1}^{n} ACT_i - \hat{\beta}_1 \sum_{i=1}^{n} ACT_i^2) = 0$

Reemplazando

- $n^{-1} (\sum_{i=1}^{n} GAT_i ACT_i - (\overline{GAT} - \hat{\beta}_1 \overline{ACT}) \sum_{i=1}^{n} ACT_i - \hat{\beta}_1 \sum_{i=1}^{n} ACT_i^2) = 0$
- $n^{-1} (\sum_{i=1}^{n} GAT_i ACT_i - \overline{GAT} \sum_{i=1}^{n} ACT_i + \hat{\beta}_1 \overline{ACT} \sum_{i=1}^{n} ACT_i - \hat{\beta}_1 \sum_{i=1}^{n} ACT_i^2) = 0$
- $\sum_{i=1}^{n} GAT_i ACT_i - \overline{GAT} \sum_{i=1}^{n} ACT_i = \hat{\beta}_1 (\sum_{i=1}^{n} ACT_i^2 - \overline{ACT} \sum_{i=1}^{n} ACT_i)$
- $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} ACT_i (GAT_i - \overline{GAT})}{\sum_{i=1}^{n} ACT_i (ACT_i - \overline{ACT})}$

Por propiedad de la sumatoria sabemos que eso es equivalente a:

- $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (ACT_i - \overline{ACT})(GAT_i - \overline{GAT})}{\sum_{i=1}^{n} (ACT_i - \overline{ACT})^2}$

Calculamos la suma de productos cruzados y la suma de productos totales de x.

```r
datos_1=datos %>% select(GPA,ACT) %>% mutate(dif_gpa=GPA-mean(GPA),dif_act=ACT-mean(ACT))
SPC=sum(datos_1$dif_gpa*datos_1$dif_act)
SPT=sum(datos_1$dif_act**2)
```

Entonces nuestro betas seran

```r
beta_1=SPC/SPT
beta_0=mean(datos$GPA)-beta_1*mean(datos$ACT)
paste('beta_0 es', round(beta_0,4),'y beta_1 es' ,round(beta_1,4))
```

```
## [1] "beta_0 es 0.5681 y beta_1 es 0.1022"
```

```r
lm(GPA~ACT,datos)
```

```
##
## Call:
## lm(formula = GPA ~ ACT, data = datos)
##
## Coefficients:
## (Intercept)          ACT
##      0.5681       0.1022
```

Vemos que existe una relacion positiva entre GPA y ACT dado que nuestro $\beta_1 > 0$. Parece que β_0 carece de un sentido practico, pero eventualmente podria decirnos cuantos es el minimo puntaje que en promedio obtienen cuando su ACT es 0.

El aumento en 5 puntos de ACT, segun la regresion anterior:

```r
paste('En promedio 5 puntos mas en la ACT son',  round(beta_1*5,3), 'puntos mas en la GPA')
```

```
## [1] "En promedio 5 puntos mas en la ACT son 0.511 puntos mas en la GPA"
```

ii) Calcule los valores ajustados y los residuales para cada observación y verifique que los residuales (aproximadamente) sumen cero.

Valor ajustado:

$$\hat{y_i} = \hat{\beta_0} + \hat{\beta_1} ACT_i$$

Residuales:

$$y_i - \hat{y_i} = \hat{u_i}$$

y queremos que $\sum_{i=1}^{n} \hat{u_i} \approx 0$

R:

```r
datos_2=datos %>% mutate(fitted=beta_0+beta_1*ACT,residuals=fitted-GPA)
datos_2 %>% knitr::kable()
```

Estudiante	GPA	ACT	fitted	residuals
1	2.8	21	2.714286	-0.0857143
2	3.4	24	3.020879	-0.3791209
3	3.0	26	3.225275	0.2252747
4	3.5	27	3.327472	-0.1725275
5	3.6	29	3.531868	-0.0681319
6	3.0	25	3.123077	0.1230769
7	2.7	25	3.123077	0.4230769
8	3.7	30	3.634066	-0.0659341

```
paste('Es la suma de los residuales aproximadomente 0?',round(datos_2$residuals %>% sum(),15)==0)
```

```
## [1] "Es la suma de los residuales aproximadomente 0? TRUE"
```

Vemos que es practicamente igual a 0.

iii) ¿Cuál es el valor que se predice para el GPA si $ACT = 20$?

R:

```
paste('Predice que en promedio el puntaje de la GPA cuando el de la ACT=20 es de:', round(beta_0+beta_1
```

```
## [1] "Predice que en promedio el puntaje de la GPA cuando el de la ACT=20 es de: 2.612"
```

iv) ¿Qué tanto de la variación en el GPA de estos ocho estudiantes es explicada por el ACT? Explique.

Para eso tendriamos que calcular el R^2,

$R^2 = SEC/STC$

$SEC = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$

$STC = \sum_{i=1}^{n}(y_i - \overline{y})^2$

```
datos_3=datos_2 %>% mutate(dif_sct=(GPA-mean(GPA))**2,dif_sec=(fitted-mean(GPA))**2)
datos_3 %>% select(residuals,dif_sct,dif_sec) %>% summarise(SCR=sum(residuals^2),SCT=sum(dif_sct),SEC=s
```

SCR	SCT	SEC	R2
0.4347253	1.02875	0.5940247	0.5774238

Explica el 0.58% de la variacion

2.4 La base de datos BWGHT.RAW contiene cifras sobre los hijos nacidos de mujeres en Estados Unidos. Las dos variables de interés son la variable independiente, peso en onzas del niño al nacer (bwght) y la variable explicativa, cantidad promedio diaria de cigarros consumidos por la madre durante el embarazo (cigs). La siguiente ecuación de regresión simple se estimó con datos de n=1,388 nacimientos:

$$\widehat{bwght} = 119.77 - 0.514cigs$$

i) ¿Cuál es el peso al nacer que se predice si cigs=0? ¿Y cuando cigs=20 (un paquete por día)? Analice la diferencia.

R:

```r
paste('Si es cigs=0 entonces bwght es',119.77 - 0.514*0 )
```

```
## [1] "Si es cigs=0 entonces bwght es 119.77"
```

```r
paste('Si es cigs=20 entonces bwght es',119.77 - 0.514*20 )
```

```
## [1] "Si es cigs=20 entonces bwght es 109.49"
```

La evidencia reafirma la crencia que se podria tener a priori sobre la relacion negativa entre la cantidad de cigarros consumidos y el peso del niño.

ii) ¿Capta esta ecuación de regresión simple una relación causal entre el peso del niño al nacer y el hábito de fumar de la madre? Explique.

R: No, dado que podria haber otros factores que se correlacionen con fumar que tambien esten afectando el peso del bebe como alimentacion o todo lo que tiene que ver con salud en general.

iii) Para que el peso al nacer predicho sea de 125 onzas, ¿cuál tiene que ser el valor de cigs? Explique.

R:

```r
paste('Tendria que fumar:',round((-125+119.77)/0.514,0), 'cigarrillos')
```

```
## [1] "Tendria que fumar: -10 cigarrillos"
```

No tiene mucho sentido practico.

iv) La proporción de mujeres en la muestra que no fumaron durante el embarazo es aproximadamente 0.85. ¿Ayuda esto a entender sus hallazgos del inciso iii)?

R: Si la muestra es representativa de la poblacion en general no debiese haber un problema en principio con esa proporcion.

La conclusion que podemos sacar de eso, es que el promedio de ese 85% corresponde a 119.77 onzas. Dado que es un estimador tiene varianza por lo que perfectamente podriamos predecir valores negativos para los cigarros.

2.5 En la función lineal de consumo $\widehat{consline} = \hat{\beta}_0 + \hat{\beta}_1 inc$

la propensión marginal a consumir estimada (PMgC) del ingreso no es más que la pendiente, β_1 , mientras que la propensión media a consumir(PMeC) es $\widehat{consline}/sinc = \hat{\beta}_0/inc + \hat{\beta}_1$ .Usando las observaciones sobre ingreso anual y consumo de 100 familias (ambos medidos en dólares), se obtiene la ecuación siguiente:

$$\widehat{consline}/sinc = -124.84 + 0.853inc$$

$$n = 100, R^2 = 0.692$$

i) Interprete el intercepto en esta ecuación y analice su signo y su magnitud.

Es negativo y pequeño en comparacion a la magnitud del ingreso, que al estar medido anualmente debiese en promedio ser un significativamente mas alto que los 124 usd.

Desde una interpretacion economica segun la funcion keynesiana del consumo correspoderia al consumo autonomo, osea ese que no depende del ingreso. No tiene mucho sentido, puedo consumir 0 pero no negativo.

ii) ¿Cuál es el consumo que se predice si el ingreso familiar es de $30,000?

```
paste('Nuestra prediccion es:', -124.84+0.853*30000)
```

```
## [1] "Nuestra prediccion es: 25465.16"
```

iii) Con inc en el eje x, trace una gráfica la PMgC estimada y de la PMeC estimada.

2.6 Usando los datos de Kiel y McClain (1995) sobre las casas vendidas en 1988 en Andover, Massachusetts, en la ecuación siguiente se relaciona el precio de las casas (price) con la distancia a un incinerador de basura construido recientemente (dist):

$$\widehat{log(price)} = 9.4 + 0.312 log(dist)$$
$$n = 135, R^2 = 0.162$$

i) Interprete el coeficiente de log(dist). ¿Es el signo de esta estimación el que se esperaba?

Es la elasticidad precio-distancia, y si es el signo esperado dado que mientras mas lejos de la basura mayor seria el precio.

ii) ¿Considera que la regresión simple proporciona un estimador insesgado de la elasticidad ceteris paribus de price (precio) respecto a dist? (Reflexione en la decisión de la ciudad sobre dónde colocar el incinerador.)

No, hay caracteristicas endogenas a las casas que determinar su precio. Por lo tanto, la ciudad tenderia a irse lo mas alejada del incinerador.

iii) ¿Qué otros factores relacionados con una casa afectan su precio? ¿Pueden estos factores estar correlacionados con la distancia al incinerador?

Podria ser de menos probabilidad que haya un edificio, casas de lujos. Si la cantidad de ventanas es determinante en el precio de una vivienda y asumimos que mas cerca de incinerador menos ventanas.

2.7 Considere la funcion de ahorro.

$sav = \beta_0 + \beta_1 inc + u, u = \sqrt{inc} * e$, donde e es una varible aleatoria con $E(e) = 0$ y

$$Var(e) = {\sigma^2}_e$$

. Suponga que e es independiente de inc.

i) Muestre que $E(u|inc) = 0$ de manera que el supuesto clave de media condicional cero (Supuesto RLS.4) se satisface.

R:

- $E(e) = 0$ Multiplico por la $E(\sqrt{inc})$,
- $E(e)E(\sqrt{inc})$=0 Por independencia,
- $E(\sqrt{inc} * e|inc) = 0$
- $E(u|inc) = 0$

ii) Muestre que $Var(u|inc) = \sigma^2{}_e inc$, de manera que se viola el supuesto RLS.5 de homocedasticidad. En particular, la varianza de sav (ahorro) aumenta con inc.

- $Var(e) = Var(e|inc) = \sigma^2{}_e$
- $Var(\frac{u}{\sqrt{inc}}|inc) = \sigma^2{}_e$
- $\frac{Var(u|inc)}{\sqrt{inc}^2} = \sigma^2{}_e$
- $Var(u|inc) = \sigma^2{}_e inc$

iii) Proporcione un análisis para sostener el supuesto de que la varianza de los ahorros aumenta con el ingreso familiar.

A medida que aumenta el ingreso mi capacidad de ahorro se vuelve muchos mas grande dado que es mas facil cubrir mis necesidades basicas.

2.8 Considere el modelo estandar regresion simple $y = \beta_0 + \beta_1 x + u$ bajo los supuestos de Gauss-markov. Los estimadores usuales de MCO de $\hat{\beta_0}$ y $\hat{\beta_1}$ son insesgados para sus respectivos parámetros poblacionales. Sea $\tilde{\beta_1}$ el estimador de β_1 obtenido cuando el intercepto es 0.

i) Determine $E(\tilde{\beta_1})$ en términos de x_i,β_0 y β_1 Verifique que $\tilde{\beta_1}$ es insesgado respecto a β_1 cuando el intercepto poblacional es cero. ¿Hay otros casos en los que $\tilde{\beta_1}$ sea insesgado?

R:

- $E(\tilde{\beta_1}) = E(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2})$
- $E(\tilde{\beta_1}) = \frac{E(\sum_{i=1}^n x_i y_i)}{E(\sum_{i=1}^n x_i^2)}$
- $E(\tilde{\beta_1}) = \frac{\sum_{i=1}^n E(x_i y_i)}{\sum_{i=1}^n E(x_i^2)}$
- $E(\tilde{\beta_1}) = \frac{\sum_{i=1}^n E(x_i(\beta_1 x_i + u_i))}{\sum_{i=1}^n E(x_i^2)}$
- $E(\tilde{\beta_1}) = \frac{\beta_1 \sum_{i=1}^n E(x_i^2))}{\sum_{i=1}^n E(x_i^2)}$
- $E(\tilde{\beta_1}) = \beta_1$

Si hay otro caso cuando el estimador es insesgado independiente del valor de los β_0 y es cuando $\overline{x} = 0$.

ii) Determine la varianza de $\tilde{\beta_1}$

- $Var(\tilde{\beta_1}) = E(\tilde{\beta_1} - E(\tilde{\beta_1}))^2$

- $Var(\tilde{\beta}_1) = E(\tilde{\beta}_1 - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2})^2$
- $Var(\tilde{\beta}_1) = E(\frac{\sum_{i=1}^{n} \beta_1 x_i^2 + u_i x_i}{\sum_{i=1}^{n} x_i^2} - E(\frac{\sum_{i=1}^{n} \beta_1 x_i^2 + u_i x_i}{\sum_{i=1}^{n} x_i^2}))^2$
- $Var(\tilde{\beta}_1) = E(\beta_1 + \frac{\sum_{i=1}^{n} u_i x_i}{\sum_{i=1}^{n} x_i^2} - \beta_1)^2$
- $Var(\tilde{\beta}_1) = E(\frac{\sum_{i=1}^{n} u_i x_i}{\sum_{i=1}^{n} x_i^2})^2$
- $Var(\tilde{\beta}_1) = (\frac{\sum_{i=1}^{n} E(u_i^2)E(x_i^2)}{\sum_{i=1}^{n} E(x_i^2)^2})$
- $Var(\tilde{\beta}_1) = (\frac{\sum_{i=1}^{n} E(u_i^2)}{\sum_{i=1}^{n} E(x_i^2)})$
- $Var(\tilde{\beta}_1) = \frac{\sigma^2}{ST_i}$, donde $ST_i = \sum_{i=1}^{n} x_i^2$

Estimado seria$= \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i}{n-1}$

iii) Muestra que $Var(\tilde{\beta}_1) \preceq Var(\hat{\beta}_1)$.

- $\frac{\sum_{i=1}^{n} e_i^2}{(n-1)\sum_{i=1}^{n}(x_i^2)} \preceq \frac{\sum_{i=1}^{n} e_i^2}{(n-2)\sum_{i=1}^{n}(x_i - \overline{x})^2}$
- $(n-1)\sum_{i=1}^{n}(x_i^2) \succeq (n-2)\sum_{i=1}^{n}(x_i - \overline{x})^2$
- $\frac{\sum_{i=1}^{n}(x_i^2)}{\sum_{i=1}^{n}(x_i - \overline{x})} \succeq \frac{n-2}{n-1}$

Donde,

- $\sum_{i=1}^{n}(x_i^2) \succeq \sum_{i=1}^{n}(x_i - \overline{x})$, $\forall i \in 1, ..., n$
- $(n-1) > (n-2)$, $\forall n \in 1, ..., N)$

Queda demostrado.