SegundoParcial

Carlos Carbone

6/14/2021

Teoria

1) Una investigadora supone el siguiente modelo:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + v_i$$

a) Que ocurre si la correlacion entre X_2 y X_3 es igual a 1 ($\rho = 1$). Comente detalladamente b) Que ocurriria si la correlacion entre X_2 y X_3 es cercana a 0.99. Comente detalladamente.

a.

Si la correlacion es 1 las variables regresoras estan correlacionadas perfectamente, es decir, existe multicolinealidad. Esto es basicamente que una de las variables es un multiplo de la otra por ejemplo, estaría sobrando una variable. Al haber multicolinealidad no voy a poder calcular los β_i con la matriz inversa puesto que su determinante va a ser igual a cero.

b.

Una correlacion alta implicaria una alta multicolinealidad si bien puedo obtener los β_i estos van a ser poco eficientes y van a dar errores altos.

2) Una investigadora supone el siguiente modelo:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + u_i$$

a) Que test utilizaria para corroborar si hay o no heterocedasticidad? Explique el procedmiento del test. Proponga los dos caminos para solucionar en caso de no existir homocedasticidad.

b) Suponga que el verdadero modelo es:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + e_i$$

donde X_3 es una variable relevante del modelo. Que ocurre?

a.

Homocedasticidad implica varianza consntate.

El test que utilizaria es el test de white. En dicho test lo que se hace es hacer una regresion tradicional con el modelo dado. Luego se obtienen los errores u_i^2 de esta primer regresion (errores al cuadrado). Luego se hace una regresión lineal multiple en funcion del cuadrado de los errores y las variables X_i, X_i^2 y $X_i * X_j$ (las cruzadas). En terminos de ecuacion queda algo asi:

$$\hat{u}_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{1i}^2 + +\alpha_4 X_{2i}^2 + +\alpha_3 X_{1i} * X_{2i} + v_i$$

Para terminar se hace una prueba de significancia global (F) con gl=2 donde la H_0: homocedasticidad y la H_a es heterocedasticidad.

Si no existe homocedasticidad tomamos dos caminos: 1) Si conocemos σ_i^2 usamos el metodo de MCG 2) Si no conocemos σ_i^2 usamos los errores robustos de white, que es construir la matriz Ω con los errores estimados $\hat{u}$.

b.

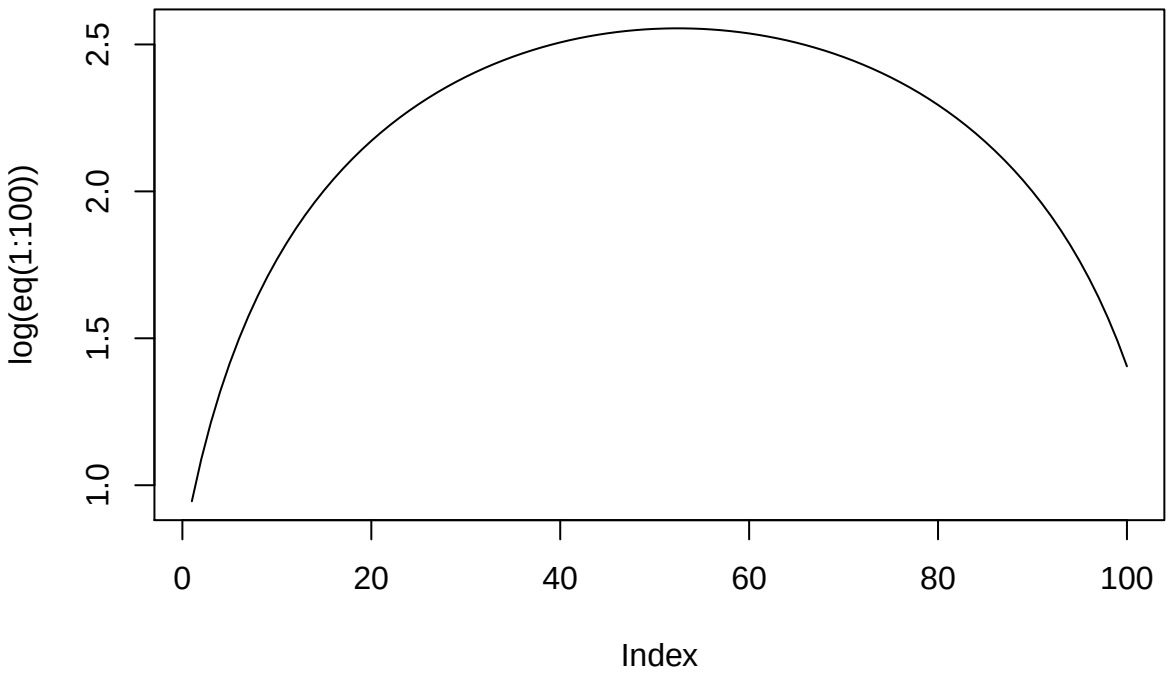
Si X_3 es una variable relvante del modelo lo que va a suceder es que mi regresion anterior va a estar llevando esa variable a u_i es decir a tomar parte del error. Se esta haciendo un subajuste del modelo. En este caso ahora los estimadores son eficientes cuando antes (con la variable omitida) eran sesgados e inconsistentes.

3

El logaritmo del ingreso mensual viene dado por

$$ln(w_i) = 2.17 + 0.408 * Ed - 0.0038896 * Ed^2 + u_i$$

```
eq<-function(x){2.17+0.408*x-0.0038896*x^2}
plot(log(eq(1:100)),type='l')
```




```
deriv(~2.17+0.408*x-0.0038896*x^2,"x")
```

```
## expression({
##     .value <- 2.17 + 0.408 * x - 0.0038896 * x^2
##     .grad <- array(0, c(length(.value), 1L), list(NULL, c("x")))
##     .grad[, "x"] <- 0.408 - 0.0038896 * (2 * x)
##     attr(.value, "gradient") <- .grad
##     .value
## })
```

```
polyroot(c(0.408,-0.0038896*2))
```

```
## [1] 52.44755+0i
```

Derivando e igualando a cero obtenemos el punto donde se hace máximo.

```
deriv(~2.17+0.408*x-0.0038896*x^2,"x")
```

```
## expression({
##     .value <- 2.17 + 0.408 * x - 0.0038896 * x^2
##     .grad <- array(0, c(length(.value), 1L), list(NULL, c("x")))
##     .grad[, "x"] <- 0.408 - 0.0038896 * (2 * x)
##     attr(.value, "gradient") <- .grad
##     .value
## })
```

```
polyroot(c(0.408,-0.0038896*2))
```

```
## [1] 52.44755+0i
```

Aca se puede ver que el maximo se alcanza aproximadamente a los 52 años. Eso se da porque al aumentar la edad en un año el salario se incrementa en 40.8% y se decrementa en 0.48896% por cada cuadrado de la edad. En principio parece algo grande el valor en que se incrementa por cada año, pero estoy asumiendo que es un modelo log-lin.

Practica

Punto 1

Suponga que esta interesado en estimar:

$$Y_i = \beta_0 + \beta_1 * X_i + \epsilon_i$$

a)

Estimamos la regresion a traves de MCO

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
x= c(1,1,2,2,3,3,4,4,5,5)
y= c(10,20,10,40,20,60,20,80,30,100)
datos_punto1<-data.frame(x,y)
regression_mco<-lm(y~.,data=datos_punto1)

stargazer(regression_mco,type="text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                   y
## -----------------------------------------------
## x                               12.500*
##                                 (5.898)
##
## Constant                         1.500
##                                (19.560)
##
## -----------------------------------------------
```

```
## Observations                    10
## R2                             0.360
## Adjusted R2                    0.280
## Residual Std. Error       26.375 (df = 8)
## F Statistic             4.492* (df = 1; 8)
## ================================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

```
summary(regression_mco)
```

```
##
## Call:
## lm(formula = y ~ ., data = datos_punto1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -34.00 -18.38   1.00  19.12  36.00
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.500     19.560   0.077   0.9408
## x             12.500      5.898   2.120   0.0669 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.37 on 8 degrees of freedom
## Multiple R-squared:  0.3596, Adjusted R-squared:  0.2796
## F-statistic: 4.492 on 1 and 8 DF,  p-value: 0.06688
```

El coeficiente β_1 es significativa al 10% y no sigfnificativa al 5% y al 1%. La constante no es significativa El modelo global es significativo al 10% solamente. La variacion en una unidad de X implica una variacion de 12.5 unidades en y.

b)

Hacemos el test de breuch Pagan para ver si hay Heterocedasticidad.

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
bptest(regression_mco)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  regression_mco
## BP = 9.4697, df = 1, p-value = 0.002089
```

La H_0 es que hay Homocedasticidad como el p-value es significativo al %5 entonces RHN y por lo tanto hay heterocedasticidad, es decir, la media de los errores no es constante.

C)

Estimamos el la varianza a través de los errores robustos de white.

Desvio estandar de los coeficientes a traves de MCO

```
coeftest(regression_mco)
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.5000    19.5600  0.0767  0.94076
## x            12.5000     5.8976  2.1195  0.06688 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Desvio estandar de los coeficientes a traves de white

```
library(car)
```

```
## Loading required package: carData
```

```
coeftest(regression_mco,vcov=hccm)
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.5000    16.5107  0.0908   0.9298
## x            12.5000     7.6758  1.6285   0.1421
```

Aca podemos ver que el interceptor y la pendiente dan igual y los desvios estandar son bastante similare: las varianzas son dichos valore elevados al cuadrado. varianza $x = 7.6758^2$ varianza intercepto $= 16.5107^2$

d)

Suponiendo que:

$$\sigma_i^2 = \sigma^2 * X_i^3$$

dividimos todo por $\sigma_i = \sqrt{(X_i^3)}$ lo que nos deja la siguiente ecuacion:

$$\frac{Y}{\sqrt{(X_i^3)}} = \frac{\beta_0}{\sqrt{(X_i^3)}} + \beta_1 * \frac{\sqrt{(X_i)}}{X_i} + \frac{u_i}{\sqrt{(X_i^3)}}$$

Hacemos la nueva regresion PO EL ORIGEN

```
datos_punto1$nuevaY=datos_punto1$y/sqrt(datos_punto1$x^3)
# datos_punto1$nuevaX=sqrt(datos_punto1$x)/datos_punto1$x
datos_punto1$nuevaX=datos_punto1$x/sqrt(datos_punto1$x^3)
datos_punto1$unoSobreRaizCubo=1/sqrt(datos_punto1$x^3)

regresion_ajustada<-lm(nuevaY~0+nuevaX+unoSobreRaizCubo,data = datos_punto1)

summary(regresion_ajustada)
```

```
##
## Call:
## lm(formula = nuevaY ~ 0 + nuevaX + unoSobreRaizCubo, data = datos_punto1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8851 -3.7163  0.2719  4.0468  5.1257
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## nuevaX             11.771      4.921   2.392   0.0437 *
## unoSobreRaizCubo    3.103      6.829   0.454   0.6616
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.804 on 8 degrees of freedom
## Multiple R-squared:  0.8248, Adjusted R-squared:  0.781
## F-statistic: 18.83 on 2 and 8 DF,  p-value: 0.0009416
```

En terminos del problema no tiene interceptores. Una de de las dos variables es significativa al 5% y no al 1% y la otra variable no es significativa. Ademas el modelo global es es signficativo asi como tambien la R^2 ajustada explica el 78.1% de los resultados.

Punto 2

a) Que significa que se usaran errores corregidos por White significa que existía heterocedasticidad. Eso significa que la varianza la calculamos como

$$var(\hat{\beta}) = (X^T X)^{-1} * X^T * \hat{u}_i \hat{u}_i^t * X(X^T X)^{-1}$$

b) Cual es la categoria base para el modelo c?

Categoria base para el modelo C es mujer sin estudios.

c) Nos quedamos con el modelo que mejor R^2 ajustado tenga, mas vairables significativs, es decir el que tiene 0.743 y ademas este tiene el AIC mas bajo -0..5663. Esto es el modelo C.

d) interpretamos los coeficientes.

El modelo no tiene constante, eso significa que si no hay estudios, la edad es cero no tiene experiencia y es mujer no percibe ingreso alguno. Ahora por ser varon gana 2.1% mas que por ser mujer, por cada año de experiencia gana 1.2% mas, por cada nivel edudcativo, universitario, secundario, primario se gana 39%,18.1% y 7,9% mas respectivamente.

Solo son significativos edad y experiencia y al 5% y al 10% respectivamente. el 79,1% (R^2 ajustado)de los salarios se explican por estas variables.