

# EXAMEN 2-Regresión Lineal

16/12/2020

*Dimas Ramirez Luis Daniel*

*Palma Poce Nayeli*

*Peñaloza Ponce Nayeli*

## Examen

Anexamos un documento llamado “Análisis previo para examen 2” en el cuál realizamos de manera rápida las pruebas para cada variable. No encontramos ninguna que cumpliera con todas las pruebas, así que utilizamos las “mejores” según nuestras consideraciones que fueron:

- Ser una variable estadísticamente significativa
- cumpliera con al menos una de las pruebas entre Durbin-Watson, Jarque-Bera, Breusch-Pagan

## 1.-Introducción

- a) Prueba Durbin-Watson

Esta prueba asume como hipótesis nula que la autocorrelación de los errores es 0. Buscar verificar que la covarianza entre los errores es cero, porque cuando se cumple eso se puede interpretar que las variables son aleatorias. Por lo tanto:

$$H_0 : \rho = 0 \quad , \quad H_A : \rho \neq 0$$

También es posible realizar la prueba enfocado a la hipótesis alternativa especificando que  $\rho$  sea mayor, diferente o menor que 0. Estas especificaciones se pueden hacer con el argumento *alternative* de la función.

El estadístico que utiliza esta prueba es:

$$DW = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2}$$

James Durbin era un estadista y econométrista británico, realizaba estudios matemático en la Universidad de Cambridge; estudios que fueron interrumpidos por la segunda guerra mundial. Fue presidente del Instituto Internacional de Estadística en 1985.

Geoffrey Stuart Watson fue un estadista australiano. También participó en el desarrollo de la prueba Wheeler-Watson para información cíclica.

En 1950 publicaron su primer paper introduciendo el estadístico Durbin-Watson.

- b) Prueba Jarque-Bera

Esta prueba verifica que los errores tiendan a distribuirse de manera normal, por lo tanto:

$$H_0 : \text{Los datos son normales}, H_A : \text{Los datos NO son normales}$$

Cuando una variable aleatoria se distribuye de manera normal, su tercer y cuarto momento alrededor de la media se conocen como sesgo y curtosis. Cuando el sesgo es igual a cero, quiere decir que la distribución es simétrica, mientras que la curtosis igual a cero, nos indica que la distribución es mesocúrtica o normal.

Las hipótesis, de manera más técnica, quedan definidas como:

$$H_0 : s = 0, c = 3 \quad y \quad H_A : s \neq 0 \text{ ó } c \neq 3$$

El estadístico que se utiliza es:

$$JB = T \left[ \frac{\hat{c}s^2}{6} + \frac{(\hat{c}c - 3)^2}{24} \right]$$

Carlos Jarque Uribe es economista y actuuario, ejecutivo en grandes corporativos y funcionario público mexicano. Fue presidente del INEGI de 1988-1999, Secretario de Plan Nacional de Desarrollo de 1994-2000 y Secretario de Desarrollo Social de 1999-2000.

Anil K. Bera es profesor de economía en la universidad de Illinois

- c) Prueba Breusch-Pagan

Es una prueba para explorar si se cumple el supuesto de homocedasticidad de los errores en regresión lineal. Las hipótesis quedan definidas en:

$$H_0 : \text{Los errores son homocedasticos} \quad y \quad H_A : \text{Los errores no son homocedasticos}$$

El estadístico que utiliza es:

$$\hat{u}^2 = \gamma_0 + \gamma_1 x + v$$

Esta prueba fue desarrollada en 1979 por Trevor Breusch, el cual fue un economista y profesor hasta su retiro en la Universidad Nacional Australiana. Adrian Rodney Pagan también es australiano, sin embargo el era economista y profesor en la Universidad de Sydney.

## 2.-Modelo Lineal Múltiple Dos Variables (No Dicotómicas)

- a) Propuesta con variable explicativa.
- b) Tres pronósticos del promedio actual con análisis.

```
library(lmtest)
```

```
## Warning: package 'lmtest' was built under R version 3.6.3
```

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

```
## Warning: package 'zoo' was built under R version 3.6.3
```

```
##
```

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

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
library(tseries)
```

```
## Warning: package 'tseries' was built under R version 3.6.3
```

```
## Registered S3 method overwritten by 'quantmod':
```

```
##      method      from
```

```
##      as.zoo.data.frame zoo
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.6.3
```

```
base <- read_excel("F:/Estadistica aplicada/Base de datos encuestas.xlsx")
```

```
#summary(base)
```

```
names(base)
```

```
## [1] "MARCA TEMPORAL"
## [2] "NOMBRE DE USUARIO"
## [3] "¿CUAL ES TU EDAD?"
## [4] "¿CUAL ES TU SEXO BIOLOGICO?"
## [5] "¿QUE CARRERA ESTUDIAS?"
## [6] "¿TE GUSTA TU CARRERA?"
## [7] "¿EN DONDE HICISTE TU BACHILLERATO?"
## [8] "¿CUAL FUE TU PROMEDIO DE BACHILLERATO?"
## [9] "¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA?"
## [10] "¿CUAL ES TU AVANCE DE CREDITOS?"
## [11] "¿CUAL ES TU NIVEL DE INGLES?"
## [12] "¿CUANTAS MATERIAS INSCRIBES EN PROMEDIO AL SEMESTRE?"
## [13] "¿CUANTAS MATERIAS HAS REPROBADO?"
## [14] "¿CUANTAS HORAS AL DIA, PASAS EN PROMEDIO EN LA FACULTAD?"
## [15] "¿REALIZAS ALGUNA ACTIVIDAD EXTRACURRICULAR?"
## [16] "¿CUANTAS HORAS EN PROMEDIO, LE DEDICAS A ESA ACTIVIDAD POR SEMANA?"
## [17] "¿CUANTAS HORAS AL DIA TE TOMA TRANSPORTARTE A LA ESCUELA?"
## [18] "¿QUE TIPO DE TRANSPORTE UTILIZAS PARA IR A LA ESCUELA?"
## [19] "¿ERES FORANEO?"
## [20] "¿CON QUIEN VIVES?"
## [21] "¿CUAL ES EL GRADO MAXIMO DE ESTUDIOS DE TU MADRE?"
## [22] "¿CUAL ES EL GRADO MAXIMO DE ESTUDIOS DE TU PADRE?"
## [23] "¿CUANTAS PERSONAS HABITAN CONTIGO?"
## [24] "¿CUANTAS CON HABITACION PROPIA?"
## [25] "¿ERES RESPONSABLE DE ALGUNA MASCOTA?"
```

```
## [26] "¿CUANTAS CON ALGUNA BECA?"
## [27] "¿ESTUDIAS Y TRABAJAS?"
## [28] "¿CUANTAS HORAS A LA SEMANA TRABAJAS?"
## [29] "¿CUAL ES EL INGRESO PROMEDIO MENSUAL DE TU FAMILIA?"
## [30] "27.\t¿CUANTO GASTAS SEMANALMENTE EN COSAS RELACIONADAS CON LA ESCUELA?"
## [31] "28.\t¿TE ENCUENTRAS EN UNA RELACION CON ALGUNA PERSONA?"
## [32] "29.\t¿TIENES HIJOS?"
## [33] "30.\t¿CUANTOS HIJOS TIENES?"
## [34] "31.\t¿TE HACES CARGO AL 100% DE TUS HIJOS?"
## [35] "32.\t¿CUANTAS HORAS DIARIAS DUERMES EN PROMEDIO?"
## [36] "33.\t¿CUANTAS CON LAPTOP PROPIA?"
## [37] "¿CUANTAS VECES AL MES CONSUMES ALCOHOL?"
## [38] "¿QUE TIPO DE BEBIDA ES LA QUE MAS CONSUMES?"
```

Nuestra primer variable será “¿CUAL FUE TU PROMEDIO DE BACHILLERATO?”, y la segunda “¿CUANTAS MATERIAS HAS REPROBADO?” .

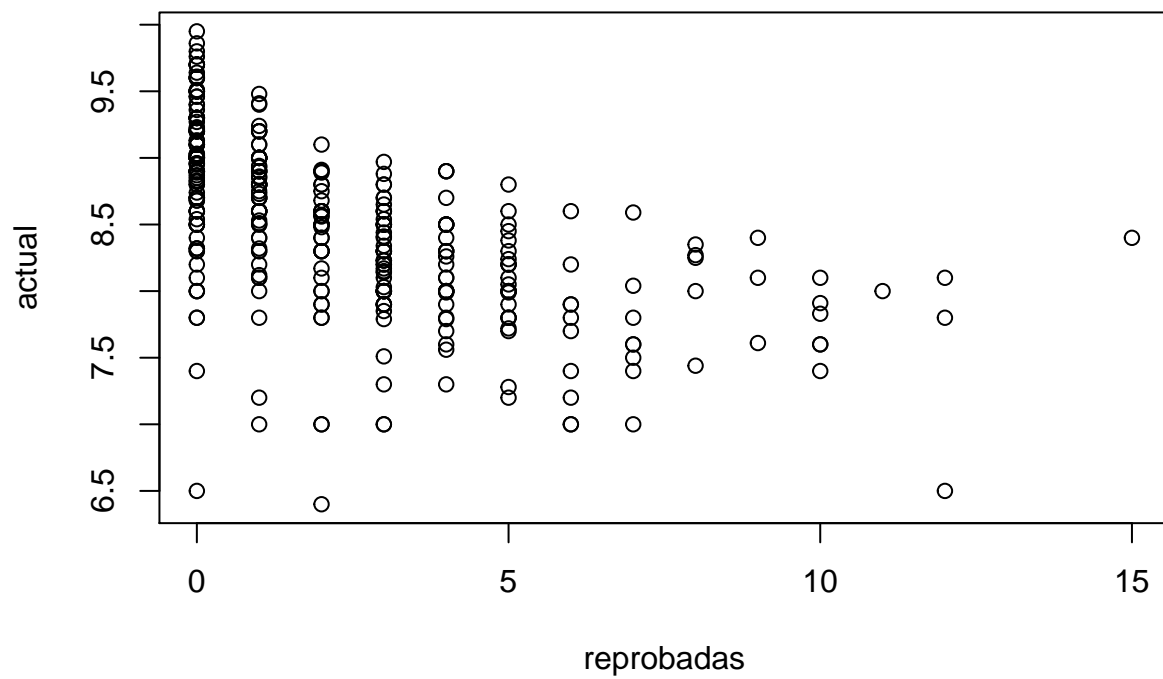
### Paso 1

```
actual <- base$`¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA?`
reprobadas <- base$`¿CUANTAS MATERIAS HAS REPROBADO?`
prepa <- base$`¿CUAL FUE TU PROMEDIO DE BACHILLERATO?`

lm1<-lm(actual~reprobadas+prepa)
#hist(base$`¿CUANTAS MATERIAS HAS REPROBADO?`)
lm1
```

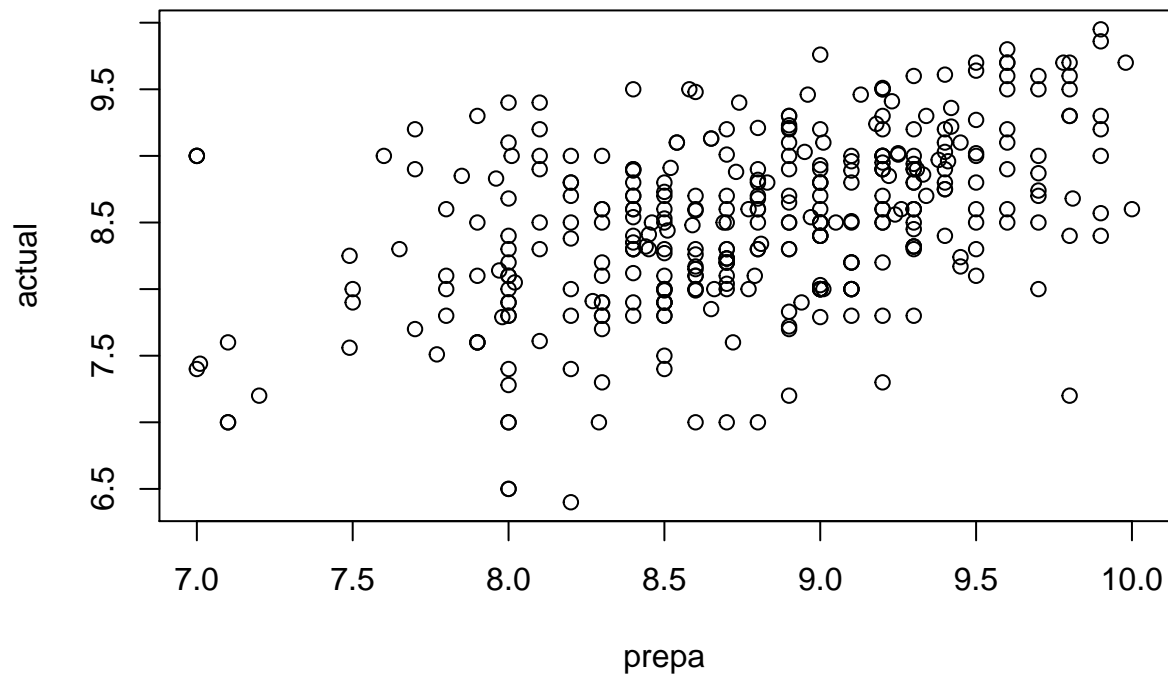
```
##
## Call:
## lm(formula = actual ~ reprobadas + prepa)
##
## Coefficients:
## (Intercept)    reprobadas        prepa
##      6.2261      -0.1171       0.2907
```

```
plot(actual~reprobadas+prepa)
```



```
abline(lm1)
```

```
## Warning in abline(lm1): only using the first two of 3 regression coefficients
```



El modelo quedaría descrito como

$$Y_i = \beta_1 + \beta_2 X_1 + \beta_3 X_2$$

Por lo tanto:

$$P_A = 6.2261 - 0.1171(\text{MateriasReprobadas}) + 0.2907(\text{Promedio Bachillerato})$$

```
summary(lm1)
```

```
##
## Call:
## lm(formula = actual ~ reprobadas + prepa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.05192 -0.29909  0.05328  0.33602  1.48804
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.22605    0.39950  15.585 < 2e-16 ***
## reprobadas   -0.11708    0.01045 -11.201 < 2e-16 ***
## prepa         0.29073    0.04436   6.555 1.92e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.4878 on 363 degrees of freedom
## Multiple R-squared:  0.4174, Adjusted R-squared:  0.4142
## F-statistic: 130 on 2 and 363 DF, p-value: < 2.2e-16
```

Observamos que para nuestra variable de *Materias Reprobadas* :

$$Pr(> |t|) = < 2e - 16$$

Y para *Promedio Bachillerato*:

$$Pr(> |t|) = 1.92e - 10$$

Por lo tanto, hasta el momento las dos variables son estadísticamente significativas para nuestro modelo.

## Paso 2

- Durbin-Watson test

```
dwtest(lm1)
```

```
##
## Durbin-Watson test
##
## data: lm1
## DW = 1.7247, p-value = 0.003933
## alternative hypothesis: true autocorrelation is greater than 0
```

La prueba Durbin Watson nos indica que  $p - value = 0.003933$ . Como  $p \leq \alpha$  existe suficiente evidencia para rechazar  $H_0 : \rho = 0$ . Por lo tanto no son variables aleatorias y hay algo que las relaciona.

## Paso 3

- Jarque-Bera test

```
jarque.bera.test(lm1$residuals)
```

```
##
## Jarque Bera Test
##
## data: lm1$residuals
## X-squared = 68.826, df = 2, p-value = 1.11e-15
```

La prueba Jarque-Bera nos indica que  $p - value = 1.11e - 15$ . Como  $p \leq \alpha$  existe suficiente evidencia para rechazar  $H_0 : Los errores son normales$

## Paso 4

-Breusch-Paga test

```
bptest(lm1)
```

```
##
## studentized Breusch-Pagan test
##
## data:  lm1
## BP = 5.8766, df = 2, p-value = 0.05296
```

La prueba Breusch-Pagan nos indica que  $p - value = 0.05296$ . Como  $p \geq \alpha$  no existe suficiente evidencia para rechazar que  $H_0$ : Los errores son homocedásticos.

## Paso 5

- Verificar que todas las  $\beta$ 's son iguales a cero.

```
summary(lm1)
```

```
##
## Call:
## lm(formula = actual ~ reprobadas + prepa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.05192 -0.29909  0.05328  0.33602  1.48804
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.22605     0.39950  15.585 < 2e-16 ***
## reprobadas   -0.11708     0.01045  -11.201 < 2e-16 ***
## prepa         0.29073     0.04436   6.555 1.92e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4878 on 363 degrees of freedom
## Multiple R-squared:  0.4174, Adjusted R-squared:  0.4142
## F-statistic: 130 on 2 and 363 DF, p-value: < 2.2e-16
```

$p - value : < 2.2e - 16$ . Como  $p \leq \alpha$  nos indica que hay alguna  $\beta \neq 0$  por lo tanto existe suficiente evidencia para rechazar  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

## Paso 6

Verificamos la información con base en la función summary en la que planteamos

$$H_0 : \beta = k \text{ y } H_A : \beta \neq k$$

Observando la columna  $Pr(> |t|)$  concluimos que no existe suficiente evidencia para rechazar la hipótesis nula y que la variable “Materias reprobadas” y “Promedio Bachillerato” son estadísticamente significativas.

## Paso 7

- Coeficiente de determinación ajustada

Adjusted R-squared: 0.4142

El 41.42% de la varación en el promedio actual se debe a un cambio en las variables de materias reprobadas y promedio de bachillerato. El otro 58.58% se debe a variables aleatorias.

## Paso 8



- Intervalos de confianza

```
confint(lm1)
```

```
##              2.5 %      97.5 %
## (Intercept)  5.4404330  7.0116717
## reprobadas  -0.1376386 -0.0965279
## prepa        0.2035055  0.3779602
```

Para las Materias Reprobadas existe un 95% de probabilidades que el coeficiente poblacional (real) se encuentre entre -0.14 y -0.097. Para el promedio de bachillerato, existe un 95% de probabilidades de que el coeficiente poblacional se encuentre entre 0.2 y 0.37.

## Paso 9

- Pronósticos

1.-Un alumno de la FI tiene 2 materias reprobadas, hasta el día de hoy que se encuentra en quinto semestre y su promedio de bachillerato fue de 9.9. Por lo tanto su promedio actual, según nuestro modelo sería

```
mr <- 2
prom_prepa <- 9.9
prom_actual <- 6.22605-(0.11708*mr)+(0.29073*prom_prepa)
prom_actual
```

```
## [1] 8.870117
```

Por lo tanto, su promedio actual será 8.870117.

2.-Un segundo alumno tiene 4 materias reprobadas y se encuentra en noveno semestre, su promedio en bachillerato fue de 7.6. Por lo tanto su promedio actual, según nuestro modelo sería

```
mr2 <- 4
prom_prepa2 <- 7.6
prom_actual2 <- 6.22605-(0.11708*mr2)+(0.29073*prom_prepa2)
prom_actual2
```

```
## [1] 7.967278
```

Por lo tanto, su promedio actual será 7.967278.

3.-Un tercer alumno que apenas va en primer semestre quiere saber cual sería su promedio si para noveno semestre tiene ocho materias reprobadas, su promedio en bachillerato fue de 8. Por lo tanto su promedio sería, según nuestro modelo:

```
mr3 <- 8
prom_prepa3 <- 8
prom_actual3 <- 6.22605-(0.11708*mr3)+(0.29073*prom_prepa3)
prom_actual3
```

```
## [1] 7.61525
```

Por lo tanto, su promedio será 7.61525.

En este caso, sí se esperaría que el hecho de tener varias materias reprobadas fuera prueba de un deficiente aprovechamiento académico que se ve reflejado en el promedio. Nuestro coeficiente de materias reprobadas es negativo por lo tanto es un buen indicador de nuestro planteamiento previo. Notamos que cuando cuando tenemos reprobada una materia, nuestro promedio se verá disminuido en 0.11708 con respecto al promedio original.

## Modelo Lineal Múltiple dos variables (Dicotómicas)

- a) Propuesta con variable explicativa.
- b) Tres pronósticos del promedio actual con análisis.

Nuestra primer variable será “¿CUAL FUE TU PROMEDIO DE BACHILLERATO?”, y la segunda “¿Cuentas con alguna Beca?” .

### Paso 1

```
beca <- base$`¿CUANTAS CON ALGUNA BECA?`
beca[beca=="SI"]<-1
beca[beca=="NO"]<-0
becaa <- as.numeric(beca)

lm2 <- lm(actual~prepa+as.factor(beca))
lm2
```

```
##
## Call:
## lm(formula = actual ~ prepa + as.factor(beca))
##
## Coefficients:
##      (Intercept)          prepa  as.factor(beca)1
##           4.6003           0.4257           0.3748
```

```
summary(lm2)
```

```
##
## Call:
## lm(formula = actual ~ prepa + as.factor(beca))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.72101 -0.31797  0.03715  0.36868  1.39435
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.60029    0.39861  11.541 < 2e-16 ***
## prepa          0.42567    0.04575   9.304 < 2e-16 ***
## as.factor(beca)1 0.37483    0.05693   6.584 1.61e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5348 on 363 degrees of freedom
## Multiple R-squared:  0.2997, Adjusted R-squared:  0.2958
## F-statistic: 77.67 on 2 and 363 DF, p-value: < 2.2e-16
```

El modelo quedaría descrito como

$$Y_i = \beta_1 + \beta_2 X_1 + \beta_3 X_2$$

Por lo tanto:

$$P_A = 4.6 + 0.42567(\textit{Promedio Bachillerat}) + 0.37483(\textit{Tiene Beca})$$

Observamos que para nuestra variable de *Promedio Bachillerato* :

$$Pr(> |t|) = < 2e - 16$$

Y para *Beca*:

$$Pr(> |t|) = 1.61e - 10$$

Por lo tanto, hasta el momento las dos variables son estadísticamente significativas para nuestro modelo.

## Paso 2

- Durbin-Watson test

```
dwtest(lm2)
```

```
##  
## Durbin-Watson test  
##  
## data: lm2  
## DW = 1.742, p-value = 0.006466  
## alternative hypothesis: true autocorrelation is greater than 0
```

La prueba Durbin Watson nos indica que  $p - value = 0.006466$ . Como  $p \leq \alpha$  existe suficiente evidencia para rechazar  $H_0 : \rho = 0$ . Por lo tanto no son variables aleatorias y hay algo que las relaciona.

## Paso 3

- Jarque-Bera test

```
jarque.bera.test(lm2$residuals)
```

```
##  
## Jarque Bera Test  
##  
## data: lm2$residuals  
## X-squared = 16.409, df = 2, p-value = 0.0002734
```

La prueba Jarque-Bera nos indica que  $p - value = 0.0002734$ . Como  $p \leq \alpha$  existe suficiente evidencia para rechazar  $H_0 : \textit{Los errores son normales}$

## Paso 4

-Breusch-Paga test

```
bptest(lm2)
```

```
##
## studentized Breusch-Pagan test
##
## data:  lm2
## BP = 8.9175, df = 2, p-value = 0.01158
```

La prueba Breusch-Pagan nos indica que  $p - value = 0.01158$ . Como  $p \leq \alpha$  existe suficiente evidencia para rechazar que  $H_0$ : Los errores son homocedásticos.

## Paso 5

- Verificar que todas las  $\beta$ 's son iguales a cero.

```
summary(lm2)
```

```
##
## Call:
## lm(formula = actual ~ prepa + as.factor(beca))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.72101 -0.31797  0.03715  0.36868  1.39435
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.60029    0.39861   11.541 < 2e-16 ***
## prepa          0.42567    0.04575    9.304 < 2e-16 ***
## as.factor(beca)1 0.37483    0.05693    6.584 1.61e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5348 on 363 degrees of freedom
## Multiple R-squared:  0.2997, Adjusted R-squared:  0.2958
## F-statistic: 77.67 on 2 and 363 DF, p-value: < 2.2e-16
```

$p - value : < 2.2e - 16$ . Como  $p \leq \alpha$  nos indica que hay alguna  $\beta \neq 0$  por lo tanto existe suficiente evidencia para rechazar  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

## Paso 6

Verificamos la información con base en la función summary en la que planteamos

$$H_0 : \beta = k \text{ y } H_A : \beta \neq k$$

Observando la columna  $Pr(> |t|)$  concluimos que no existe suficiente evidencia para rechazar la hipótesis nula y que la variable “Tiene Beca” y “Promedio Bachillerato” son estadísticamente significativas.

## Paso 7

- Coeficiente de determinación ajustada

Adjusted R-squared: 0.2958

El 29.58% de la varación en el promedio actual se debe a un cambio en las variables de promedio de bachillerato y si tiene Beca. El otro 70.42% se debe a variables aleatorias.

## Paso 8

- Intervalos de confianza

```
confint(lm2)
```

```
##                2.5 %    97.5 %
## (Intercept)    3.8164057 5.3841732
## prepa          0.3356943 0.5156448
## as.factor(beca)1 0.2628707 0.4867899
```

Para el promedio de bachillerato, existe un 95% de probabilidades de que el coeficiente poblacional se encuentre entre 0.3356 y 0.5157. Mientras que para la variable de si tiene beca o no, existe un 95% de probabilidades de que el coeficiente poblacional se encuentre entre 0.26 y 0.486.

## Paso 9

- Pronósticos

1.-Un alumno de segundo semestre al cual no le fue otorgado la beca, quieres saber su promedio si su promedio de bachillerato fue de 7.7 2.-Un alumno de nuevo ingreso quiere saber cual sera su promedio en la carrera si le llegan a otorgar la beca. Su promedio de bachillerato fue de 9.3 3.-Un alumno de nuevo ingreso quiere saber cual sera su promedio en caso de que le asignen una beca. SU promedio de bachillerato fue de 8.8.

```
pb <- c(7.7, 9.3, 8.8)
be <- c(0,1,1)
prono1 <- 4.60029+(0.42567*pb)+(0.37483*be)
prono11 <- round(prono1,2)
```

Los promedios serán: 7.88, 8.93, 8.72

De este modelo podriamos afirmar que el tener una beca influiara en nuestro promedio actual. Es decir; si tengo una beca voy a tener un mejor promedio y viceversa. Sin embargo, dejando un poco el análisis estadístico, creemos que no es cierto. Probablemente tenerla sí ayude en algo, pero el hecho de no tenerla no debería disminuir (en la mayoría de los casos) tu promedio actual en 0.37483.

Utilizando las pruebas estadísticas también podriamos afirmar que aunque es una variable estadísticamente significativa, los errores no son aleatorios, normales,ni homocedasticos.

## Modelo Lineal Múltiple de Más de Dos Variables

- a) Propuesta con variable explicativa.
- b) Tres pronósticos del promedio actual con análisis.

Nuestra primer variable será “¿CUAL FUE TU PROMEDIO DE BACHILLERATO?”, la segunda “¿Cuántas materias inscribes en promedio al semestre?” y la tercera “¿Estudias y trabajas?”

### Paso 1

```
materias <- base$`¿CUANTAS MATERIAS INSCRIBES EN PROMEDIO AL SEMESTRE?`

trabaja <- base$`¿ESTUDIAS Y TRABAJAS?`
trabaja[trabaja=="SI"] <- 1
trabaja[trabaja=="NO"] <- 0
trabaja <- as.numeric(trabaja)

lm3 <- lm(actual~prepa+materias+as.factor(trabaja))
lm3
```

```
##
## Call:
## lm(formula = actual ~ prepa + materias + as.factor(trabaja))
##
## Coefficients:
##          (Intercept)              prepa              materias
##             4.4125             0.4483             0.0328
## as.factor(trabaja)1
##             -0.1773
```

```
summary(lm3)
```

```
##
## Call:
## lm(formula = actual ~ prepa + materias + as.factor(trabaja))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85224 -0.34215  0.01293  0.31279  1.38192
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.41254    0.41888  10.534 < 2e-16 ***
## prepa          0.44826    0.04749   9.439 < 2e-16 ***
## materias       0.03280    0.01203   2.725  0.00674 **
## as.factor(trabaja)1 -0.17730    0.07272  -2.438  0.01524 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5558 on 362 degrees of freedom
## Multiple R-squared:  0.2456, Adjusted R-squared:  0.2394
## F-statistic: 39.29 on 3 and 362 DF, p-value: < 2.2e-16
```

Observamos que para nuestra variable de *Materias inscritas al semestre* :

$$Pr(> |t|) = 0.00674$$

Para *Promedio Bachillerato*:

$$Pr(> |t|) = 2e - 16$$

Y para *Trabaja o no*:

$$Pr(> |t|) = 0.01524$$

Por lo tanto, hasta el momento las tres variables son estadísticamente significativas para nuestro modelo. Sin embargo, del análisis de la función `summary` podemos notar que las variables de las materias inscritas y el trabajo, son significativas en menor medida que el promedio de bachillerato.

## Paso 2

- Durbin-Watson test

```
dwtest(lm3)
```

```
##
## Durbin-Watson test
##
## data: lm3
## DW = 1.7709, p-value = 0.01373
## alternative hypothesis: true autocorrelation is greater than 0
```

La prueba Durbin Watson nos indica que  $p - value = 0.01373$ . Como  $p \leq \alpha$  existe suficiente evidencia para rechazar  $H_0 : \rho = 0$ . Por lo tanto no son variables aleatorias y hay algo que las relaciona.

## Paso 3

- Jarque-Bera test

```
jarque.bera.test(lm3$residuals)
```

```
##
## Jarque Bera Test
##
## data: lm3$residuals
## X-squared = 7.0553, df = 2, p-value = 0.02937
```

La prueba Jarque-Bera nos indica que  $p - value = 0.02937$ . Como  $p \leq \alpha$  existe suficiente evidencia para rechazar  $H_0 : \text{Los errores son normales}$

## Paso 4

-Breusch-Paga test



```
bptest(lm3)
```

```
##
## studentized Breusch-Pagan test
##
## data:  lm3
## BP = 9.0038, df = 3, p-value = 0.02924
```

La prueba Breusch-Pagan nos indica que  $p - value = 0.02924$ . Como  $p \leq \alpha$  existe suficiente evidencia para rechazar que  $H_0$ : Los errores son homocedásticos

### Paso 5

- Verificar que todas las  $\beta$ 's son iguales a cero.

```
summary(lm3)
```

```
##
## Call:
## lm(formula = actual ~ prepa + materias + as.factor(trabaja))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85224 -0.34215  0.01293  0.31279  1.38192
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.41254     0.41888  10.534 < 2e-16 ***
## prepa             0.44826     0.04749   9.439 < 2e-16 ***
## materias          0.03280     0.01203   2.725  0.00674 **
## as.factor(trabaja)1 -0.17730     0.07272  -2.438  0.01524 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5558 on 362 degrees of freedom
## Multiple R-squared:  0.2456, Adjusted R-squared:  0.2394
## F-statistic: 39.29 on 3 and 362 DF,  p-value: < 2.2e-16
```

$p - value : < 2.2e - 16$ . Como  $p \leq \alpha$  nos indica que hay alguna  $\beta \neq 0$  por lo tanto existe suficiente evidencia para rechazar  $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

### Paso 6

Verificamos la información con base en la función summary en la que planteamos

$$H_0 : \beta = k \text{ y } H_A : \beta \neq k$$

Observando la columna  $Pr(> |t|)$  concluimos que no existe suficiente evidencia para rechazar la hipótesis nula y que la variable “Promedio Bachillerato”, “Materias promedio al semestre” y “Trabaja y Estudia” son estadísticamente significativas.

### Paso 7

- Coeficiente de determinación ajustada

Adjusted R-squared: 0.2394

El 23.94% de la varación en el promedio actual se debe a un cambio en las variables de promedio de bachillerato, catidad de materias al semestre y si trabaja y estudia. El otro 76.06% se debe a variables aleatorias.

## Paso 8

- Intervalos de confianza

```
confint(lm3)
```

##	2.5 %	97.5 %
## (Intercept)	3.588809038	5.23627879
## prepa	0.354869997	0.54164587
## materias	0.009129949	0.05646214
## as.factor(trabaja)1	-0.320305933	-0.03429694

Para el promedio de bachillerato, existe un 95% de probabilidades de que el coeficiente poblacional se encuentre entre 0.35y 0.55. Mientras que para la variable de materias inscritas , existe un 95% de probabilidades de que el coeficiente poblacional se encuentre entre 0.009 y 0.056. Por último, para la variable de si trabaja y estudia, existe un 95% de probabilidades de que el coeficiente poblacional se encuentre entre -0.32 y -0.034

## Paso 9

- Pronósticos

1.-Un alumno que estudia y trabaja, mete en promedio 3 materias por semestre, quiere saber su promedio actual si su promedio en bachillerato fue de 9.5

```
tra <- 1
mat <- 3
pr <- 9.5
pronostico <- 4.41254+(0.44826*pr)+(0.03280*mat)-(0.17730*tra)
pronosticor <- round(pronostico, 2)
```

El promedio de este alumno será de 8.59

2.-Un alumno que solamente estudia, mete en promedio 7 materias por semestre, quiere saber su promedio actaul si su promedio fue de 8.5

```
tra2 <- 0
mat2 <- 7
pr2 <- 8.5
pronostico2 <- 4.41254+(0.44826*pr)+(0.03280*mat2)-(0.17730*tra2)
pronostico2r <- round(pronostico2, 2)
```

El promedio de este alumno será de 8.9

3.-Un alumno que trabaja y estudia mete en promedio 4 materias por semestre, quiere saber su promedio actual si su promedio en bachillerato fue de 8.5

```
tra3 <- 1
mat3 <- 4
pr3 <- 8.5
pronostico3 <- 4.41254+(0.44826*pr3)+(0.03280*mat3)-(0.17730*tra3)
pronostico3r <- round(pronostico3, 2)
```

El promedio de este alumno será de 8.18

4.-Un alumno que estudia y trabaja mete en promedio 7 materias por semestre, quiere saber su promedio actual si su promedio en bachillerato fue de 9

```
tra4 <- 1
mat4 <- 7
pr4 <- 9
pronostico4 <- 4.41254+(0.44826*pr4)+(0.03280*mat4)-(0.17730*tra4)
pronostico4r <- round(pronostico4, 2)
```

El promedio de este alumno será de 8.5

Para este modelo esperaríamos que la variable afirmativa en el caso de que estudie y trabaje, se vea reflejada de manera negativa en el desempeño académico lo cual es cierto, sin embargo tenemos que recordar que aunque es una variable estadísticamente significativa, no es suficientemente representativa. De igual manera pasa con las materias que se inscriben por semestre. Tal vez en un modelo con solamente dos variables, tendría sentido, sin embargo para este modelo podemos hacer el análisis de que se esperaría que si trabaja y estudia, y además mete 7 materias (son las máximas posibles en un semestre dentro de la FI) su promedio se vería muy afectado a pesar de haber tenido un buen promedio en el bachillerato. En el pronóstico 4 observamos que su promedio solamente se vio disminuido en 0.5 lo cual se ve un poco alejado de la realidad promedio pero no suena descabellado del todo.

## Referencias

- **Durbin Watson Test.** Recuperado el 15/12/2020 de: <http://math.furman.edu/~dcs/courses/math47/R/library/lmtest/html/dwtest.html>
- **Durbin Watson Statistic.** Recuperado el 15/12/2020 de: <https://corporatefinanceinstitute.com/resources/knowledge/other/durbin-watson-statistic/>
- **Overview: James Durbin.** Recuperado el 15/12/2020 de: <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095736356>
- **Overview: Geoffrey Stuart Watson.** Recuperado el 15/12/2020 de: <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803121318164>
- **Pruebas de normalidad.** Recuperado el 15/12/2020 de: [https://rpubs.com/luisxsuper/normalidad\\_test](https://rpubs.com/luisxsuper/normalidad_test)
- **Prueba de normalidad: Jarque Bera & Shapiro Wilks en RStudio** Recuperado el 15/12/2020 de: <https://www.youtube.com/watch?v=SFSg0EzzTUE>
- **Carlos Jarque.** Recuperado el 15/12/2020: [https://es.wikipedia.org/wiki/Carlos\\_Jarque](https://es.wikipedia.org/wiki/Carlos_Jarque)
- **Trevor S. Breusch.** Recuperado el 15/12/2020: [https://en.wikipedia.org/wiki/Trevor\\_S.\\_Breusch](https://en.wikipedia.org/wiki/Trevor_S._Breusch)
- **Adrian Pagan.** Recuperado el 15/12/2020: [https://en.wikipedia.org/wiki/Adrian\\_Pagan](https://en.wikipedia.org/wiki/Adrian_Pagan)