Ejercicio 1

Dania Bustamante, Rosana Cardona, Jose Casanova

2024-03-27

```r
library(readxl)
attend <- read_excel("attend.xlsx")
```

Pregunta 1

Estime el modelo usando MCO e interprete el valor del parámetro β_1 hallado.

Modelo Poblacional

$\text{atndrte} = \beta_0 + \beta_1 \text{priGPA} + \beta_2 \text{ACT} + \beta_3 \text{fresh} + u$

Modelo Estimado

$\hat{\text{atndrte}} = \hat{\beta}_0 + \hat{\beta}_1 \text{priGPA} + \hat{\beta}_2 \text{ACT} + \hat{\beta}_3 \text{fresh}$

```r
rg <- lm(atndrte~priGPA+ACT+frosh,attend)
stargazer(rg,type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                               atndrte
## -----------------------------------------------
## priGPA                       18.250***
##                               (1.119)
##
## ACT                          -1.692***
##                               (0.168)
##
## frosh                        4.363***
##                               (1.365)
##
## Constant                     71.581***
##                               (4.067)
##
## -----------------------------------------------
## Observations                    680
## R2                             0.301
## Adjusted R2                    0.298
## Residual Std. Error      14.282 (df = 676)
## F Statistic           97.100*** (df = 3; 676)
```

```
## ================================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

El estimador $\hat{\beta}_1$ que acompaña a la variable *priGPA* es de 18.250. Esto significa que existe una relación positiva entre la calificación promedio acumulativa antes del semestre y el porcentaje de clases asistidas. Es decir, si priGPA aumenta en una unidad, se espera que el porcentaje de clases asistidas, en promedio, aumenta en 18.25% manteniendo todas las demás variables constantes. Por lo tanto, se puede interpretar que un anterior mejor desempeño académico se relaciona positivamente con una asistencia mayor a las clases durante el semestre.

Pregunta 2

¿Cuál es rol de las variables *ACT* y frosh?. Estime el modelo usando MCO sin incluir la variable *ACT*. ¿Cambia mucho la estimación de β_1? ¿Porqué?

La variable ACT representa la calificación de la prueba de selección universitaria. Cuando se incluye en el modelo, tiene un $\hat{\beta}_2$ de -1.692, lo que indica que un aumento en la calificación de la prueba universitaria se asocia con una disminución del 1.692% de clases asistidas, con todo lo demás constante.

La variable frosh indica si el estudiante es de primer año o no. Tiene un $\hat{\beta}_3$ de 4.363 en el anterior modelo, esto nos indica que el porcentaje de las clases asistidas aumentan, en promedio, 4.363% si el estudiante es de primer año, con todo lo demás constante.

Modelo Estimado sin la variable ACT

$\widehat{\text{atndrte}} = \hat{\beta}_0 + \hat{\beta}_1\text{priGPA} + \hat{\beta}_2\text{fresh}$

```r
rg1 <- lm(atndrte~priGPA+frosh,attend)
stargazer(rg1,type = "text")
```

```
##
## ================================================
##                          Dependent variable:
##                     ---------------------------
##                               atndrte
## ------------------------------------------------
## priGPA                       14.562***
##                               (1.134)
##
## frosh                         4.985***
##                               (1.461)
##
## Constant                     42.883***
##                               (3.110)
##
## ------------------------------------------------
## Observations                    680
## R2                             0.196
## Adjusted R2                    0.194
## Residual Std. Error      15.305 (df = 677)
## F Statistic          82.682*** (df = 2; 677)
## ================================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

$\hat{\beta}_1$ *con ATC*	$\hat{\beta}_1$ *sin ATC*
Magnitud: 18.250	Magnitud: 14.562
Signo Positivo	Signo Positivo

Al realizar la estimación sin la variable ACT, el porcentaje de las clases asistidas aumentan, en promedio, 14.56% por cada unidad más de calificación promedio acumulativa antes del semestre, con las demás variables explicativas constantes.

Se observa que el valor estimado de $\hat{\beta}_1$ disminuye de 18.250 a 14.560 al eliminar ACT del modelo. Esto indica que parte del efecto de la calificacion de la prueba de selección universitaria está siendo capturado por la calificacion promedio acumulativa antes del semestre. Es decir, una parte de la explicación que la variable ACT tenía en la variable atndrte está siendo dada por la variable priGPA. Además, tambien se concluye que al omitir una variable significativa se está cometiendo un error de sesgo y el $\hat{\beta}_1$ cambiará su coeficiente.

Pregunta 3

¿Cómo se interpreta el valor del parámetro β_3?. Estime ahora el siguiente modelo:

Modelo Poblacional

$\text{atndrte} = \beta_0 + \beta_1 \text{priGPA} + \beta_2 \text{ACT} + \beta_3 \text{frosh} + \beta_4 \text{priGPA} \times \text{frosh} + u$

Modelo Estimado

$\hat{\text{atndrte}} = \hat{\beta}_0 + \hat{\beta}_1 \text{priGPA} + \hat{\beta}_2 \text{ACT} + \hat{\beta}_3 \text{frosh} + \hat{\beta}_4 \text{priGPA} \times \text{frosh}$

```r
rg2 <- lm(atndrte~priGPA+ACT+frosh+priGPA:frosh,attend)
stargazer(rg2,type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                               atndrte
## -----------------------------------------------
## priGPA                       18.806***
##                               (1.257)
##
## ACT                          -1.713***
##                               (0.169)
##
## frosh                         10.542
##                               (6.493)
##
## priGPA:frosh                  -2.622
##                               (2.694)
##
## Constant                     70.572***
##                               (4.198)
##
```

```
## -----------------------------------------------
## Observations                    680
## R2                            0.302
## Adjusted R2                   0.298
## Residual Std. Error      14.283 (df = 675)
## F Statistic           73.056*** (df = 4; 675)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

$\hat{\beta}_3$ nos indica que, si el estudiante es de primer año, el porcentaje de clases asistidas aumentan en promedio en 10.542%, manteniendo constantes todas las demás variables en el modelo. Esto sugiere que los estudiantes de primer año tienden a asistir a un mayor porcentaje de clases en comparación con los estudiantes de otros años.

¿Cuál es la interpretación del parámetro β_4?.

El parámetro $\hat{\beta}_4$ representa el efecto de la interacción entre las variables *priGPA* y frosh. Nos muestra cómo cambia el efecto de *priGPA* en *atndrte* dependiendo de si el estudiante es de primer año (frosh) o no.

El valor negativo de $\hat{\beta}_4$ nos dice que hay una relación negativa entre la interacción de priGPA y frosh y el porcentaje de clases asistidas (atndrte). Lo que indica que la relación entre *priGPA* y *atndrte* es menor para los estudiantes de primer año en comparación con los estudiantes de otros años, después de haber considerado la interacción. Es decir, si la calificacion promedio acumulada antes del semestre, para los estudiantes de primer año aumenta en una unidad, en promedio, la asistencia a clases disminuye en un 2,625%, en comparacion a los estudiantes de otros años, manteniendo las demás variables constantes.

Sin embargo, dado que el p-valor asociado al coeficiente $\hat{\beta}_4$ no es significativo, no se tiene suficiente evidencia estadistica para afirmar que la interacción entre priGPA y frosh tiene un efecto significativo en el porcentaje de clases asistidas.

Pregunta 4

Realice un test estadístico para probar la hipótesis nula que $\beta_1 = 12$. ¿Es plausible este valor en la población? Explique cómo realizó el test.

Para probar la hipótesis nula de que $\beta_1 = 12$, debemos realizar un *test de hipótesis utilizando el estadístico t.*

El modelo de regresión utilizado para realizar la inferencia es: $\widehat{\text{atndrte}} = \hat{\beta}_0 + \hat{\beta}_1\text{priGPA} + \hat{\beta}_2\text{ACT} + \hat{\beta}_3\text{fresh} + \hat{\beta}_4\text{priGPA} \times \text{frosh}$

- Hipotesis nula: $H_0 : \beta_1 = 12$
- Hipotesis alterna: $H_1 : \beta_1 \neq 12$

  Para calcular el estadistico t: $t = \frac{\hat{\beta}_1 - \beta_0}{\text{error estándar de } \hat{\beta}_1}$

```r
t_calculado <- (18.806-12)/(1.257)
cat(t_calculado)
```

```
## 5.414479
```

Remplazando:

$t = \frac{18.806-12}{1.257}$

$t = \frac{6.806}{1.257}$

$t \approx 5.41$

Luego, se compara con el valor crítico de la distribución t con $n-k-1$ grados de libertad para determinar si rechazamos o no la hipótesis nula, donde n es el número de observaciones, k es el número de variables explicativas y 1 es el intercepto.

Dado que se está utilizando un nivel de significancia de 0.05 y el test es de dos colas, necesitaremos encontrar los valores críticos $t_{\frac{\alpha}{2}}$ y $-t_{\frac{\alpha}{2}}$ para un nivel de significancia de 0.025 en cada cola de la distribución t, con 675 grados de libertad.

```
t_critico <- qt(1 - 0.025, 675)
cat(t_critico)
```

```
## 1.963485
```

```
t_critico_neg <- -qt(1 - 0.025, 675)
cat(t_critico_neg)
```

```
## -1.963485
```

Si el t calculado es mayor que 1.9634 o menor que -1.9634, se rechaza la hipotesis nula. De lo contrario, no tendríamos suficiente evidencia para rechazar la hipótesis nula.

Dado que el valor calculado de $t \approx 5.58$ es mayor en valor absoluto que 1.96349, podemos concluir que cae fuera del intervalo crítico. Por lo tanto, tenemos suficiente evidencia estadistica para rechazar la hipótesis nula. En otras palabras, es plausible afirmar que el valor de β_1 no es igual a 12.

Pregunta 5

Realice un test estadístico para probar la hipótesis nula conjunta de que $\beta_2 = 0$ y $\beta_3 = 0$. ¿Cuál es la interpretación económica de esta hipótesis? ¿Es plausible esta hipótesis en la población? Explique cómo realizó el test.

Para probar la hipótesis nula conjunta de que $\beta_2 = 0$ y $\beta_3 = 0$, se usará una *prueba F.* Estas pruebas nos permiten evaluar si todas las variables independientes incluidas en el modelo tienen un efecto conjunto significativo en la variable dependiente.

- Hipotesis nula conjunta: $H_0 : \beta_2 = \beta_3 = 0$
- Hipotesis alterna: $H_1 : \textit{Al menos uno de los } \beta_2 \textit{ o } \beta_3 \neq 0$

Para realizar esta prueba, calculamos un estadístico de prueba basado en la distribución F. Si el valor del estadístico F calculado es lo suficientemente grande, rechazamos la hipótesis nula y concluimos que al menos una de las variables incluidas en el modelo tiene un efecto significativo en la variable dependiente.

Primero ajustamos el modelo de regresión lineal con todas las variables independientes:

Modelo no restringido:$\widehat{\text{atndrte}} = \hat{\beta}_0 + \hat{\beta}_1 \text{priGPA} + \hat{\beta}_2 \text{ACT} + \hat{\beta}_3 \text{fresh} + \hat{\beta}_4 \text{priGPA} \times \text{frosh}$

Luego comparamos este modelo con el modelo restringido que excluye las variables ACT y frosh:

Modelo restringido: $\hat{\text{atndrte}} = \hat{\beta}_0 + \hat{\beta}_1 \text{priGPA} + +\hat{\beta}_2 \text{priGPA} \times \text{frosh}$

La comparación entre estos dos modelos nos dará el estadístico F para realizar la prueba de hipótesis.

$F_c = \frac{(SRC_r - SRC_{nr})/SRC_{nr}}{(n-k-1)/q}$

Alternativamente:

$F_c = \frac{(R^2_{nr} - R^2_r)}{(1-R^2_{nr})} \times \frac{(n-k-1)}{q}$

```
modelo_no_restringido <- rg2
modelo_restringido <- lm(atndrte~priGPA+priGPA:frosh,attend)
stargazer(modelo_no_restringido,modelo_restringido,type = "text")
```

```
##
## ======================================================================
##                                      Dependent variable:
##                           --------------------------------------------
##                                            atndrte
##                                     (1)                     (2)
## ----------------------------------------------------------------------
## priGPA                           18.806***               14.130***
##                                   (1.257)                 (1.101)
##
## ACT                              -1.713***
##                                   (0.169)
##
## frosh                             10.542
##                                   (6.493)
##
## priGPA:frosh                      -2.622                  2.057***
##                                   (2.694)                 (0.605)
##
## Constant                         70.572***               44.067***
##                                   (4.198)                 (2.989)
##
## ----------------------------------------------------------------------
## Observations                        680                     680
## R2                                 0.302                   0.196
## Adjusted R2                        0.298                   0.194
## Residual Std. Error          14.283 (df = 675)       15.306 (df = 677)
## F Statistic               73.056*** (df = 4; 675) 82.627*** (df = 2; 677)
## ======================================================================
## Note:                                      *p<0.1; **p<0.05; ***p<0.01
```

Una vez que tengamos ambos modelos ajustados, comparamos sus bondades de ajuste utilizando la prueba F para determinar si el modelo con todas las variables independientes es significativamente mejor que el modelo restringido.

```
f_calculado <- (0.302-0.196)/(1-0.302) * (680-4-1)/2
cat(f_calculado)
```

```
## 51.25358
```

Remplazando:

$F_c = \frac{(0.302-0.196)}{(1-0.302)} \times \frac{(680-4-1)}{2}$

$F_c = \frac{0.106}{0.698} \times \frac{675}{2} \approx 51.25$

Después de calcular el estadístico F, utilizando los modelos ajustados, obtenemos un valor para el estadístico F. Luego, comparamos este valor con el valor crítico de la distribución F para determinar si es significativo.

Si el valor calculado de F es mayor que el valor crítico de la distribución F, rechazamos la hipótesis nula conjunta de que $H_0 : \beta_2 = \beta_3 = 0$, lo que indica que al menos una de las variables "ACT" o "frosh" es significativa en la explicación de las variaciones en el porcentaje de clases asistidas.

```r
f_critico <-  qf(1-0.05,2,675)
cat(f_critico)
```

```
## 3.009067
```

Al comparar el valor calculado del estadístico F con el valor crítico, utilizando un nivel de significancia de 0.05, se eobserva que el valor calculado del estadístico F es mayor al valor crítico. Por lo tanto, se rechaza la hipótesis nula conjunta.

Lo anterior nos indica que al menos una de las variables "ACT" o "frosh" es significativa en la explicación de las variaciones en el porcentaje de clases asistidas. Por lo tanto, al menos una de estas variables tiene un efecto significativo en el porcentaje de clases asistidas cuando se considera junto con otras variables en el modelo.

Finalmente, podemos concluir que tanto la puntuación ACT como el indicador de si el estudiante es de primer año son relevantes para explicar las variaciones en la asistencia a clases.

Pregunta 6 Si el o la estudiante A tiene priGPA = 3.1, ACT = 21 y no es de primer año (fresh = 0) y el o la estudiante B tiene priGPA = 2.1, ACT = 26 y no es de primer año (fresh = 0), ¿cuál es la diferencia predicha en sus tasas de asistencia?

Modelo usado:

$\widehat{\text{atndrte}} = \hat{\beta}_0 + \hat{\beta}_1 \text{priGPA} + \hat{\beta}_2 \text{ACT} + \hat{\beta}_3 \text{fresh} + \hat{\beta}_4 \text{priGPA} \times \text{frosh}$

```r
# Parametros del modelo
b0 <- 70.572
b1 <- 18.806
b2 <- -1.713
b3 <- 10.542
b4 <- -2.622

# Estudiante A
priGPA_A <- 3.1
ACT_A <- 21
frosh_A <- 0

# Estudiante B
priGPA_B <- 2.1
ACT_B <- 26
frosh_B <- 0
```

```r
# % de asistencia predicha para el estudiante A
atndrte_A <- b0 + b1 * priGPA_A + b2 * ACT_A + b3 * frosh_A + b4 * priGPA_A * frosh_A

# % de asistencia predicha para el estudiante B
atndrte_B <- b0 + b1 * priGPA_B + b2 * ACT_B + b3 * frosh_B + b4 * priGPA_B * frosh_B

# Diferencia de asistencia predichas
diferencia <- atndrte_A - atndrte_B

print(diferencia)
```

```
## [1] 27.371
```

Según las caracteristicas predichas, el estudiante A, en promedio, asiste a clases un $26,71\%$ más que el estudiante B.