Estefanía Bonilla Rodríguez

**Taller 4- REGRESIÓN LINEAL**

Ho = el coeficiente es 0. No hay relación entre las 2 variables.

1. *Realice una gráfica de cajas y bigotes de la variable salary\_06 (variable de resultado) para las dos categorías de la variable select (variable explicativa). ¿Qué indica el diagrama sobre las diferencias entre los participantes y no participantes?*

El diagrama indica que en el caso de los no (0) participantes la mediana del salario se encuentra aproximadamente entre 230.000 y 240.000 pesos. Y el salario máximo de los no participantes llegó hasta 1’000.000 de pesos.

En cuanto a los participantes (1) el valor de la mediana del salario es más alto y se encuentra aproximadamente entre 290.000 y 300.000 pesos, hay una mayor dispersión en el primer cuartil. Se puede evidenciar que el 50% de los participantes logro un mejor salario que los no participantes. De igual forma el salario máximo (en casos particulares) fue de 2’000.000 de pesos.

1. ***“Regresión 1”***
2. ¿Cuál es el coeficiente de select? Describa qué quiere decir ese coeficiente.

El coeficiente de select es 39.701, quiere decir que los participantes ganan 39.701 más que los no participantes.

1. ¿El coeficiente es estadísticamente significativo? ¿Por qué? Al responder esta pregunta indique cuál es la hipótesis nula, cuál es el valor p, responda si puede rechazar esta hipótesis nula a los niveles de confianza tradicionales (90%, 95%y 99%) y por qué, y en caso que pueda rechazarla en favor de cuál hipótesis se rechaza.

Si es estadísticamente significativo ya que se rechaza la Ho de que no hay una relación entre las dos variables. Esto se explica ya que teniendo en cuenta los niveles de confianza tradicionales el valor de p= 0. 000000127 es menor a los respectivos niveles de significancia, por lo cual se puede decir que la variable “select” y salary\_06 si se relacionan.

1. Calcule un intervalo de confianza para este coeficiente a un nivel de confianza del 99 %. ¿Qué quiere decir este intervalo?:

Este intervalo hace referencia a que con un nivel de confianza del 99% el coeficiente en la población se encuentra entre 20378,21 y 59024,65.

1. ¿La respuesta en b. es consistente con la respuesta en c.? ¿Por qué?

Si es consistente debido a que se rechaza la Ho en la cual el coeficiente es igual a 0, y en este intervalo no se encuentra el 0.

1. e. De acuerdo con el R2 ¿Qué porcentaje de la variación en salary\_06 es explicado por select?

Según R2, la variable select solo explica el 0,8285% en la variación de la variable de resultado salary\_06.

1. ***“Regresión 2”. Estimen una nueva regresión en la que, a las variables incluidas en la regresión estimada en el punto 2, agreguen las variables dwomen y age\_s, y adicionalmente incluya una interacción entre las variables select y dwomen. Llame a esta regresión, “regresion2”.***
2. Usando la siguiente función: stargazer (regresion1, regresion2, type ="text", out = "regresiones.txt"), cree una tabla que le permita compararlas dos regresiones. Copie y pegue esa tabla en el archivo de Word.



1. Cuál es el nuevo coeficiente de la variable select ¿Describa qué quiere decir este coeficiente? ¿Es estadísticamente significativo? ¿Por qué? Al responder esta pregunta indique cuál es la hipótesis nula, cuál es el valor p, responda si puede rechazar esta hipótesis nula a los niveles de confianza tradicionales (90%, 95%y 99%) y por qué, y en caso que pueda rechazarla en favor de cuál hipótesis se rechaza.

Al controlar con las variables dwomen y age\_s el coeficiente es 37.742,5. Quiere decir que los participantes ganan 37.742,5 más que los no participantes. El coeficiente si es estadísticamente significativo debido a que el valor p=0.000554 menor a todos los niveles de significancia en los niveles de confianza (90%, 95%, 99%); por lo que se rechaza la Ho de que no hay una relación entre las variables select y salary\_06.

1. Describa los coeficientes de las dos variables que agregó e indique si esos coeficientes son estadísticamente significativos y por qué. Siga las orientaciones de la pregunta anterior.
* Controlando la edad y la participación en el programa, el coeficiente de dwomen hace referencia a que las mujeres ganan 89.357,2 menos que los hombres. Es estadísticamente significativo porque se rechaza la Ho debido a que el valor de p= 0.0000000000000002 es menor a cualquiera de los niveles de significancia.

* Controlando la participación y el sexo el coeficiente de age\_s significa que cuando la edad de los participantes aumenta en un año el salario aumenta en 4.277,7. No es estadísticamente significativo en el nivel de confianza del 99% ya que p=0.015 y no es menor al nivel de significancia respectivo. No obstante en los niveles de confianza del 95% y 90% si es estadísticamente significativo, rechazando la Ho.
1. Describa el coeficiente de la interacción entre select y dwomen e indique si ese coeficiente es estadísticamente significativo y por qué. Siga las orientaciones dela pregunta b.

Dejando constante la participación (variable select) el número de mujeres que participan en el programa ganan 203,6 menos que los hombres participantes. El coeficiente no es estadísticamente significativo en ninguno de los niveles de confianza debido a que el valor p= 0.9 siendo mayor a los tres niveles de significancia respectivos. Esto quiere decir que no hay un efecto particular para las mujeres que participan respecto de los hombres participantes, no hay una diferencia relevante por lo que se acepta la Ho.

1. De acuerdo con el R2 ¿Qué porcentaje de la variación ensalary\_06 es explicado por la segunda regresión? Comparada con la regresión 1, está regresión es mejor o peor explicando la variación de salary\_06.

R2 hace referencia a que la variación de salary\_06 puede ser explicada por la regresión 2 en un 5,1%. En comparación con la regresión1 (0.8285%.),la segunda explicaría mejor la variación de salary\_06 al tener un mayor porcentaje.

1. ***Repita todo lo realizado en los puntos 2 y 3 usando como variable dependiente o de resultado profit\_06.***

***Punto 2* “Regresión 3”**

* 1. ¿Cuál es el coeficiente de select? Describa qué quiere decir ese coeficiente.

El coeficiente de select es -3.850. Quiere decir que los participantes tienen 3.850 menos en ganancias que los no participantes.

* 1. ¿El coeficiente es estadísticamente significativo? ¿Por qué? Al responder esta pregunta indique cuál es la hipótesis nula, cuál es el valor p, responda si puede rechazar esta hipótesis nula a los niveles de confianza tradicionales (90%, 95%y 99%) y por qué, y en caso que pueda rechazarla en favor de cuál hipótesis se rechaza.

El coeficiente no es estadísticamente significativo en todos los niveles de confianza (90%, 95%y 99%), debido a que el valor p=0.187 siendo igual o mayor a los distintos niveles de significancia respectivos. Por lo tanto no se rechaza la Ho de que no hay una relación entre las 2 variables.

* 1. Calcule un intervalo de confianza para este coeficiente a un nivel de confianza del 99 %. ¿Qué quiere decir este intervalo?:

A un nivel de confianza del 99% el coeficiente en la población puede encontrarse entre -11.370 y 3.670.

* 1. ¿La respuesta en b. es consistente con la respuesta en c.? ¿Por qué?

Si debido a que la Ho es que el coeficiente es igual a 0 y este valor se encuentra entre el rango calculado el en punto c de -11.370 y 3.670.

* 1. De acuerdo con el R2 ¿Qué porcentaje de la variación en profit\_06 es explicado por select?

Según R2, la variable select solo explica el 0,02287% en la variación de la variable profit\_06.

**Punto 3 *“Regresión 4”. Estimen una nueva regresión en la que, a las variables incluidas en la regresión estimada en el punto 2, agreguen las variables dwomen y age\_s, y adicionalmente incluya una interacción entre las variables select y dwomen. Llame a esta regresión, “regresion4”.***

1. Usando la siguiente función: stargazer (regresion1, regresion2, type ="text", out = "regresiones.txt"), cree una tabla que le permita compararlas dos regresiones. Copie y pegue esa tabla en el archivo de Word.



1. Cuál es el nuevo coeficiente de la variable select ¿Describa qué quiere decir este coeficiente? ¿Es estadísticamente significativo? ¿Por qué? Al responder esta pregunta indique cuál es la hipótesis nula, cuál es el valor p, responda si puede rechazar esta hipótesis nula a los niveles de confianza tradicionales (90%, 95%y 99%) y por qué, y en caso que pueda rechazarla en favor de cuál hipótesis se rechaza.

El coeficiente es -15.101, lo cual hace referencia a que los participantes tienen 15.101 menos en utilidades que los no participantes. Si es estadísticamente significativo ya que el valor de p=0,000475 siendo menor a todos los niveles de significancia en sus respectivos niveles de confianza.

Al ser estadísticamente significativo se rechaza la Ho de que no hay una relación entre las variables select y profit\_06; en favor de la hipótesis alternativa.

1. Describa los coeficientes de las dos variables que agregó e indique si esos coeficientes son estadísticamente significativos y por qué. Siga las orientaciones de la pregunta anterior.
* Dejando constantes la edad y la participación en el programa, el coeficiente de dwomen hace referencia a que las mujeres tienen 24.192 menos utilidades que los hombres. Es estadísticamente significativo ya que el valor de p= 0.0000000134 es menor a los niveles de significancia en los niveles de confianza 90%, 95%, 99%; por ello se rechaza la Ho de que no hay relación entre el sexo y profit\_06

* Controlando la participación y sexo el coeficiente significa que cuando la edad de los participantes aumenta en un año la utilidad aumenta en 2.531. Es estadísticamente significativo ya que p= 0.000276 es menor a los niveles de significancia de los niveles de confianza tradicionales, por lo que se rechaza la Ho en favor de la alternativa si hay una relación entre la edad y profit\_06
1. Describa el coeficiente de la interacción entre select y dwomen e indique si ese coeficiente es estadísticamente significativo y por qué. Siga las orientaciones dela pregunta b.

Dejando constante la participación (variable select) el número de mujeres que participa tiene 20.351 más de utilidad que los hombres participantes.

El coeficiente es estadísticamente significativo ya que en los niveles de confianza del 90%, 95%, 99% porque el valor de p= 0.000487 siendo menor a los respectivos niveles de significancia. Por ello se rechaza la Ho de que el efecto del programa es el mismo para hombres y mujeres en favor de la hipótesis alternativa de que el programa si tiene un efecto especifico en las mujeres.

1. De acuerdo con el R2 ¿Qué porcentaje de la variación profit\_06 es explicado por la segunda regresión? Comparada con la regresión 1, está regresión es mejor o peor explicando la variación de profit\_06.

R2 explica que la variación en profit\_06 puede ser explicada en un 1,28% por la segunda regresión. Por eso, esta regresión explica mejor la variación de profit\_06 al tener un mayor porcentaje que la “regresión 3” (0,02287%).

1. ***[Esta pregunta vale 30% del taller] De acuerdo con los resultados obtenidos en los resultados anteriores, considera que el programa jóvenes en acción tuvo impacto en el salario y las ganancias de las personas. Justifique su respuesta y asegúrese de mencionar los efectos y diferencias estimadas en las preguntas anteriores. La respuesta debe contener por lo menos dos párrafos.***

El programa tuvo impacto positivo en el salario, inicialmente la gráfica de cajas y bigotes muestra que la mediana del salario es más alta para los participantes. De igual forma, los resultados de la regresión 1 evidencian que los participantes lograron un salario de 39.701 más alto que los no participantes. Esto también ocurre al controlar variables como la edad y el sexo, los participantes continúan teniendo un mayor salario ganando 37.743 más que los no participantes. En conclusión se acepta la hipótesis alternativa de que el hecho de participar o no, si influye en el nivel salarial de los jóvenes; ya que hacer parte del programa significó tener un salario más alto. Lo anterior puede ser sustentado con el valor de R2 (controlando las variables) según el cual, select puede explicar la variación del salario en un 5,1%, apoyando que si existe una relación entre el nivel salarial y el haber participado o no en el programa.

Sin embargo, cuando se analiza la variable profit\_06 el impacto no fue el esperado. El programa tuvo un efecto negativo en las ganancias de los participantes debido a que sus ganancias fueron menores en 3.850 que los no participantes, como lo muestra el coeficiente de select en la regresión 3. Cuando se dejan constantes las variables age\_s y dwomen la conclusión es la misma, el coeficiente de la variable select evidencia que los participantes tuvieron 15.101 menos ganancias que los no participantes. Por otra parte, se debe tener en cuenta que al aplicar la interacción las mujeres participantes ganan 20.351 más que los hombres participantes, entonces el programa tuvo un efecto diferente en las mujeres. Finalmente, es preciso aclarar que aunque el valor R2 (manteniendo constantes las variables) es un porcentaje bajo, el coeficiente de select demuestra que si hay una relación entre las ganancias y el hecho de ser participante o no.