**Actividad: Análisis de regresión: Salarios en una empresa**

Se cuenta con una base de datos de los empleados de una empresa. Hay indicios de que la empresa tiene problemas de discriminación salarial por género y se desea soportar esta idea con análisis de regresión.

Hacer un análisis de regresión de la variable tomando a Y=salario en función de otras variables independientes como

X1= género. (X1=1 mujer, X1=0 hombre)

X2= años de experiencia previos a la entrada a la empresa

X3= años de experiencia en la empresa

X4= nivel educativo. Esta es una variable categórica. Hay cinco niveles: 1, 2, 3, 4 y 5. Se usarán cuatro variables binarias: X42 (nivel 2), X43 (nivel 3), X44 (nivel 4) y X45 (nivel 5). El nivel 1 es la referencia, es decir, cuando una persona tiene nivel educativo 1, X42=X43=X44=X45=0.

**Datos**

Discriminacionsalarial.csv

**Regresión 1: Salario vs género**

La ecuación de regresión es:

*Salario = 78784 12874X1 (R2= 0.04)*

Es decir, que si es hombre (variable X1=0) gana en promedio $78784, si es mujer (X1=1) $65910. Hay una diferencia de $12874 en promedio.

**Regresión 2: Salario vs género, experiencia previa y experiencia actual**

Ahora añadamos 2 variables: X2: experiencia previa y X3: experiencia en la empresa.

La ecuación de regresión correspondiente es:

Salario = \_31937 - 8975X1 + 3308X2 + 2708X3 *(R2= 0.77 )*

Es decir, separando entre hombres y mujeres (reemplace el valor 1 para mujeres y 0 para los hombres en la ecuación) las ecuaciones son:

Salario mujer = 22962 + 3308X2 + 2708X3

Salario hombre = 31937 - 3308X2 + 2708X3

Se concluye que si hombres y mujeres tienen la misma experiencia previa y en la empresa hay una diferencia de 8975 en el salario entre hombres y mujeres.

**Regresión 3: Salario vs género, experiencia previa, experiencia actual y nivel educativo**

Como R2 aún es bajo, se ingresará otra variable X4=nivel educativo. Para ello, convertir los datos de la variable X4 en variables binarias X42, X43, X44 y X45.

La ecuación de regresión correspondiente es:

Salario = 17924 -8202X1 + 2926X2 + 2593X3 - 5978X42 + 14511X43 + 25984X44 + 64490X45 *(R2= 0.97)*

Es decir, separando entre hombres y mujeres (reemplace el valor 1 para mujeres y 0 para los hombres en la ecuación) las ecuaciones son:

Salario mujer = 9722 + 2926X2 + 2593X3 - 5978X42 + 14511X43 + 25984X44 + 64490X45

Salario hombre = 17924 + 2926X2 + 2593X3 - 5978X42 + 14511X43 + 25984X44 + 64490X45

Se concluye que, si hombres y mujeres tienen la misma experiencia previa y en la empresa y el mismo nivel educativo, hay una diferencia de \_\_\_\_\_\_\_\_\_\_\_\_en el salario entre hombres y mujeres.

¿Qué concluye? Existe discriminación por género?

Si hay discriminación por género.

**ANEXO**

**CODIGO EN R**

#Decisiones bajo incertidumbre en las organizaciones

#Actividad: Análisis de regresión para verificar si hay discriminación por género

#Lectura de datos

datos=read.csv(file.choose(),sep=",",header=T)

#Nombramiento de las varibles

y=datos$Salario.anual

x1=datos$genero

x2=datos$Experiencia.previa

x3=datos$experiencia.en.la.empresa

x4=datos$Nivel.educativo

#Regresion de salario vs género

r1=lm(y~x1)

summary(r1)

#Regresión de salario vs genero y experiencia previa y actual

r2=lm(y~x1+x2+x3)

#Expresamos x4 como una variable categorica

x4=as.factor(x4)

#Regresion de salario vs genero, experiencia previa, experiencia actual y nivel educativo

r3=lm(y~x1+x2+x3+x4)

summary(r3)

**Resultados**

> datos=read.csv(file.choose(),sep=",",header=T)

> y=datos$Salario.anual

> x1=datos$genero

> x2=datos$Experiencia.previa

> x3=datos$experiencia.en.la.empresa

> x4=datos$Nivel.educativo

> r1=lm(y~x1)

> summary(r1)

Call:

lm(formula = y ~ x1)

Residuals:

Min 1Q Median 3Q Max

-65385 -20435 -3697 16846 85115

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 78785 3216 24.499 < 2e-16 \*\*\*

x1 -12875 4211 -3.058 0.00253 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 29650 on 202 degrees of freedom

Multiple R-squared: 0.04424

> r2=lm(y~x1+x2+x3)

> x4=as.factor(x4)

> r3=lm(y~x1+x2+x3+x4)

> summary(r3)

Call:

lm(formula = y ~ x1 + x2 + x3 + x4)

Residuals:

Min 1Q Median 3Q Max

-23376.5 -878.3 -248.8 595.2 20628.7

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 17924.34 1998.06 8.971 2.38e-16 \*\*\*

x1 -8202.56 824.26 -9.951 < 2e-16 \*\*\*

x2 2926.54 84.81 34.506 < 2e-16 \*\*\*

x3 2593.66 66.96 38.733 < 2e-16 \*\*\*

x42 -5978.84 2278.77 -2.624 0.00938 \*\*

x43 14511.87 1924.88 7.539 1.72e-12 \*\*\*

x44 25984.41 2015.28 12.894 < 2e-16 \*\*\*

x45 64490.08 2681.09 24.054 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5747 on 196 degrees of freedom

Multiple R-squared: 0.9652