Prueba de Regresión

Estrella Guerra, Danilo David

2024-09-11

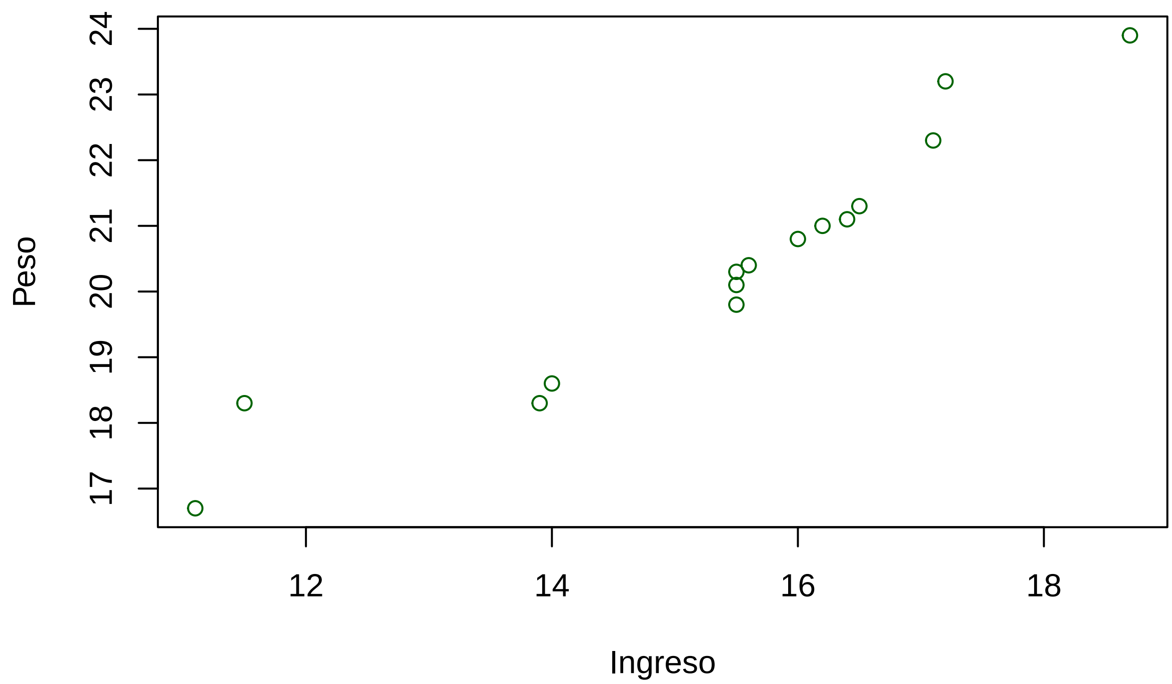
El alcalde de una municipalidad solicitó que se utilicen los datos recolectados de los hogares de esa municipalidad y encargó al personal de R&Dgo Analytics que estime el modelo de Regresión Lineal Simple para explicar el peso (en kg) de papel o cartón desechado al mes/familia en función del ingreso familiar mensual (en miles de soles). Se obtuvieron los siguientes datos:

```r
# Datos proporcionados
Y <- c(18.3, 16.7, 21, 21.1, 23.9,
          18.6, 22.3, 20.8, 23.2, 21.3, 20.4, 20.3, 18.3, 20.1, 19.8)
X <- c(11.5, 11.1, 16.2, 16.4,
             18.7, 14, 17.1, 16, 17.2, 16.5, 15.6, 15.5, 13.9, 15.5, 15.5)

# Creación del data frame
datos <- data.frame(X, Y)

# Visualización del data frame
print(datos)
```

```
##       X    Y
## 1  11.5 18.3
## 2  11.1 16.7
## 3  16.2 21.0
## 4  16.4 21.1
## 5  18.7 23.9
## 6  14.0 18.6
## 7  17.1 22.3
## 8  16.0 20.8
## 9  17.2 23.2
## 10 16.5 21.3
## 11 15.6 20.4
## 12 15.5 20.3
## 13 13.9 18.3
## 14 15.5 20.1
## 15 15.5 19.8
```

Diagrama de Dispersión

```r
plot(datos$X,datos$Y,
     col="darkgreen", xlab = "Ingreso", ylab = "Peso")
```



Pregunta 1

A un nivel de significación de 0.05, responda las siguientes preguntas:

a) La pendiente estimada del modelo de regresion lineal simple es:

```r
lm(Y~X, data = datos)
```

```
##
## Call:
## lm(formula = Y ~ X, data = datos)
##
## Coefficients:
## (Intercept)            X
##      6.6919       0.8917
```

b) La suma de cuadrados del error en el cuadro de Análisis de Varianza es:

```r
# Ajustar el modelo de regresión lineal simple
modelo <- lm(Y~X, data = datos)

# Realizar el análisis de varianza
anova_modelo <- anova(modelo)

# Mostrar el cuadro de ANOVA
print(anova_modelo)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value       Pr(>F)
## X          1 46.282  46.282  110.46 0.0000001007 ***
## Residuals 13  5.447   0.419
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Extraer la Suma de cuadrados del Error (SCE)
SCE <- anova_modelo["Residuals", "Sum Sq"]

round(SCE, 3)
```

```
## [1] 5.447
```

c) El valor del coeficiente de determinación (en términos porcentuales) es igual a:

```r
summary(modelo)
```

```
##
## Call:
## lm(formula = Y ~ X, data = datos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7869 -0.3150 -0.1595  0.2347  1.3532
##
## Coefficients:
##             Estimate Std. Error t value    Pr(>|t|)
## (Intercept)  6.69193    1.31559   5.087    0.000209 ***
## X            0.89173    0.08485  10.510 0.000000101 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6473 on 13 degrees of freedom
## Multiple R-squared:  0.8947, Adjusted R-squared:  0.8866
## F-statistic: 110.5 on 1 and 13 DF,  p-value: 0.0000001007
```

El valor del coeficiente de determinación (R^2) Multiple R-squared: 0.8947 y en porcentaje es 89.47%

Pregunta 2

Si se estima un intervalo de confianza para el coeficiente de correlación para dos variables X e Y y se encuentra que su límite inferior es positivo y su límite superior es negativo.

Solo existen 3 posibles casos:

$$IC(\beta_1) = [-; +]$$

$$IC(\beta_1) = [-; -]$$

$$IC(\beta_1) = [+; +]$$

Porlo tanto podemos deducir que el intervalo ha sido mal estimado.

Pregunta 3

Se quiere evaluar la significancia de la pendiente del un modelo de regresión lineal simple en base a una muestra de tamaño 22. Por lo tanto el valor crítico (con aproximación a 3 decimales) a un nivel de significación de 0.03 para contrastar el estadístico F es:

```
round(qf(1-0.03, 1, 20), 3)
```

```
## [1] 5.458
```

Dado que el F-statistic (110.5) es mayor que tu valor crítico (5.458), y el p-value es extremadamente pequeño (0.0000001007), puedes concluir que la relación entre el ingreso mensual de las familias y el peso del cartón desechado es estadísticamente significativa.