

LAB10 : PRUEBA GLOBAL DEL MODELO MULTILINEAL. PRUEBA INDIVIDUAL PARA LOS COEFICIENTES. INTERVALO DE CONFIANZA PARA LOS PARÁMETROS

Aranda Huerta, Milene & Escriba Flores Daniel

2025-05-21

Contents

| | | |
|----------|--|----------|
| 1 | Caso Practico | 1 |
| 2 | Pregunta 1: Lectura - Regresion | 2 |
| 2.1 | Lectura de Datos | 2 |
| 2.2 | Análisis de regresión múltiple | 3 |
| 2.3 | Ecuación de la regresión | 3 |
| 3 | Pregunta 2: Prueba Global e individual - Intervalo de Confianza | 3 |
| 3.1 | Prueba Global e individual | 4 |
| 3.1.1 | Interpretación de la prueba global: | 4 |
| 3.1.2 | Pruebas Individuales (t) | 5 |
| 3.2 | Intervalo de Confianza | 5 |
| 3.2.1 | Interpretación de los intervalos de confianza | 5 |
| 3.3 | Nuevo modelo de regresión múltiple | 6 |
| 3.4 | Verificar que todos los coeficientes sean significativos | 6 |
| 3.4.1 | Interpretación | 7 |
| 3.5 | Pronósticos con las nuevas variables | 7 |
| 4 | Conclusiones Finales | 8 |

1 Caso Practico

Frank Castle desea ampliar sus negocios a otras ciudades como Arequipa, Cusco, Trujillo y Piura. Para preparar su presentación al banco local, le gustaría comprender mejor los factores que hacen que una tienda

de descuento en particular sea productiva. Este debe hacer todo el trabajo por cuenta propia, por lo que no será capaz de estudiar todas las tiendas de descuento. Por tanto, selecciona una muestra aleatoria de 15 tiendas y registra las ventas diarias promedio, el espacio de piso (área), el número de cajones de estacionamiento y el ingreso medio de las familias en la región por cada una de las tiendas.

Deseando predecir las ventas diarias promedio con base a las otras variables, haga lo siguiente:

2 Pregunta 1: Lectura - Regresion

a. Lea la base de datos y haga un análisis de regresión múltiple. Además, escriba la ecuación de regresión.

2.1 Lectura de Datos

```
library(readxl)
library(pander)
library(equatiomatic)

##
## Attaching package: 'equatiomatic'

## The following object is masked from 'package:datasets':
##
##      penguins

data = read_excel("tiendas.xlsx", sheet="ventas")
data = data[-1]
data = as.data.frame(data)

# Usar pander para formatear la tabla
pander(data, caption = "Tabla de ventas diarias promedio, area de
      tienda, cajones de estacionamiento e ingreso de familias")
```

Table 1: Tabla de ventas diarias promedio, area de tienda, cajones de estacionamiento e ingreso de familias

| ventas | area | cajones | ingreso |
|--------|------|---------|---------|
| 1840 | 532 | 6 | 44 |
| 1746 | 478 | 4 | 51 |
| 1812 | 530 | 7 | 45 |
| 1806 | 508 | 7 | 46 |
| 1792 | 514 | 5 | 44 |
| 1825 | 556 | 6 | 46 |
| 1811 | 541 | 4 | 49 |
| 1803 | 513 | 6 | 52 |
| 1830 | 532 | 5 | 46 |
| 1827 | 537 | 5 | 46 |
| 1764 | 499 | 3 | 48 |

| ventas | area | cajones | ingreso |
|--------|------|---------|---------|
| 1825 | 510 | 8 | 47 |
| 1763 | 490 | 4 | 48 |
| 1846 | 516 | 8 | 45 |
| 1815 | 482 | 7 | 43 |

2.2 Analisis de regresión multiple

Para llevar a cabo la regresión múltiple utilizamos la función `lm()` escribiendo primero la variable dependiente y luego las independientes

```
modelo = lm(formula= ventas ~ area +
             cajones +
             ingreso,
             data=data)

modelo$coefficients
```

```
## (Intercept)      area      cajones      ingreso
## 1480.7446116    0.7314989    9.9914874   -2.3082627
```

2.3 Ecuacion de la regresión

Ecuación de regresión:

Ventas = 1480.745 + 0.7314989 * Población + 9.991487 * Presupuesto + -2.308263 * Ingreso

```
# Versión básica
extract_eq(modelo)
```

$$\text{ventas} = \alpha + \beta_1(\text{area}) + \beta_2(\text{cajones}) + \beta_3(\text{ingreso}) + \epsilon \quad (1)$$

```
# Versión con valores de coeficientes
extract_eq(modelo, use_coefs = TRUE, coef_digits = 5)
```

$$\hat{\text{ventas}} = 1480.74461 + 0.7315(\text{area}) + 9.99149(\text{cajones}) - 2.30826(\text{ingreso}) \quad (2)$$

3 Pregunta 2: Prueba Global e individual - Intervalo de Confianza

b. Aplique la prueba global e individual para los coeficientes y construya intervalos de confianza para los mismos con un 96% de confiabilidad. Interprete los resultados obtenidos.

La función `summary()` nos permite obtener los resultados del estadístico F para la prueba global y los estadístico t para las pruebas individuales.

Para la prueba global, si el p-valor es menor a 0.05, se debe rechazar la hipótesis nula H_0 de que todos los coeficientes de la regresión son ceros y en caso contrario aceptar H_0 .

Asimismo, para la prueba individual, si el p-valor asociado a un coeficiente es menor a 0.05, se debe rechazar la hipótesis nula H_a de que el coeficiente de la regresión es cero y en caso contrario aceptar H_0 .

3.1 Prueba Global e individual

```
resumen = summary(modelo)
resumen

##
## Call:
## lm(formula = ventas ~ area + cajones + ingreso, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.5076  -8.3760  -0.2526   10.3255   16.3206
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1480.7446    126.3042   11.724 1.48e-07 ***
## area          0.7315     0.1633    4.479 0.000933 ***
## cajones       9.9915     2.6000    3.843 0.002733 **
## ingreso      -2.3083     1.5947   -1.447 0.175655
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.42 on 11 degrees of freedom
## Multiple R-squared:  0.8354, Adjusted R-squared:  0.7905
## F-statistic: 18.6 on 3 and 11 DF,  p-value: 0.0001285

# Estadístico F calculado
F_calculado <- resumen$fstatistic[1]
F_calculado

##      value
## 18.60356

# Nivel de confianza: 96% (alfa = 0.04)
F_critico <- qf(0.96, 3, 11)
print(F_critico)

## [1] 3.909611
```

3.1.1 Interpretación de la prueba global:

F calculado: 18.6 F crítico (96%): 3.91 p-valor: 0.0001285

Como el F calculado (18.6) es mayor que el F crítico (3.91), rechazamos la hipótesis nula con un 96% de confianza de que todos los Betas son iguales a 0. Esto significa que el modelo en su conjunto es estadísticamente significativo y al menos uno de los predictores (área, cajones o ingreso) tiene un efecto significativo sobre las ventas. Asimismo, el p-valor muy pequeño (0.0001285) confirma esta conclusión.

3.1.2 Pruebas Individuales (t)

- Intercepto:
p-valor: 1.48e-07 (0.000000148) Como p-valor ($1.48e-07$) < 0.04 , rechazamos la hipótesis nula. El intercepto es estadísticamente significativo con un 96% de confianza.
- Área (0.7315):
p-valor: 0.000933 Como p-valor (0.000933) < 0.04 , rechazamos la hipótesis nula. El coeficiente de área es estadísticamente significativo con un 96% de confianza.
- Cajones (9.9915):
p-valor: 0.002733 Como p-valor (0.002733) < 0.04 , rechazamos la hipótesis nula. El coeficiente de cajones es estadísticamente significativo con un 96% de confianza.
- Ingreso (-2.3083):
p-valor: 0.175655 Como p-valor (0.175655) > 0.04 , no rechazamos la hipótesis nula. El coeficiente de ingreso no es estadísticamente significativo con un 96% de confianza.

3.2 Intervalo de Confianza

Para encontrar los intervalos de confianza para el intercepto y los parámetros del modelo de regresión múltiple poblacional, usamos la función `confint()`, ingresando como parámetros el modelo y el nivel de confianza. Si alguno de los intervalos contiene al cero (0), este no es significativo para el modelo y por tanto se puede excluir del mismo.

```
intervalos <- confint(modelo, level = 0.96)
print(intervalos)
```

```
##              2 %          98 %
## (Intercept) 1186.6908675 1774.798356
## area        0.3512426    1.111755
## cajones     3.9384156    16.044559
## ingreso     -6.0208893    1.404364
```

3.2.1 Interpretación de los intervalos de confianza

Al examinar estos intervalos de confianza del 96%:

- Intercepto: [1186.59, 1774.90]
No contiene el cero El intercepto es estadísticamente significativo con un 96% de confianza
- Área: [0.35, 1.11]
No contiene el cero El coeficiente de área es estadísticamente significativo Con 96% de confianza, por cada unidad de aumento en el área, las ventas aumentan entre 0.35 y 1.11 unidades
- Cajones: [3.93, 16.05]
No contiene el cero El coeficiente de cajones es estadísticamente significativo Con 96% de confianza, por cada cajón adicional, las ventas aumentan entre 3.93 y 16.05 unidades

- Ingreso: [-6.02, 1.40]

Contiene el cero El coeficiente de ingreso no es estadísticamente significativo Como el intervalo incluye el cero, no podemos descartar la posibilidad de que el verdadero efecto del ingreso sea nulo, lo que sugiere que su influencia en las ventas no es relevante en este modelo y *podría ser excluida*.

#Pregunta 3: Nuevo modelo - Pronostico

- De acuerdo a lo anterior, genere un nuevo modelo de regresión para el cual todos los coeficientes sean significativos. Además, realice 10 pronósticos con las nuevas variables donde una tome valores por encima de su máximo y las otras sean constantes.

3.3 Nuevo modelo de regresion múltiple

Excluyemos la variable Ingreso

```
modelo_revisado <- lm(ventas ~ area + cajones, data = data)
summary(modelo_revisado)
```

```
##
## Call:
## lm(formula = ventas ~ area + cajones, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.432  -8.118  -4.612   11.589   18.655
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1342.4902    86.3319   15.550 2.57e-09 ***
## area         0.7727     0.1680    4.599 0.000612 ***
## cajones     11.6338     2.4438    4.761 0.000464 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.02 on 12 degrees of freedom
## Multiple R-squared:  0.804, Adjusted R-squared:  0.7713
## F-statistic: 24.61 on 2 and 12 DF,  p-value: 5.67e-05
```

3.4 Verificar que todos los coeficientes sean significativos

```
# Verificar intervalos de confianza
confint(modelo_revisado, level = 0.96)
```

```
##              2 %          98 %
## (Intercept) 1143.6918280 1541.288526
## area         0.3857567   1.159546
## cajones     6.0064378   17.261077
```

3.4.1 Interpretacion

- Vemos que todos los coeficientes tienen el p-valor menor a 0.04 (para el nivel de confianza del 96%) y que sus intervalos de confianza no contienen el cero.

3.5 Pronósticos con las nuevas variables

Para hacer el pronóstico y cumplir con los criterios, elegiremos que haya 9 cajones de estacionamiento fijos y la area de los pisos sera 10 unidades mas del maximo. Asimismo, los siguiente valores iran en crecimiento de 10 en 10.

```
# Obtener el valor máximo del área
max_area <- max(data$area)

# Crear 10 valores de área empezando desde el máximo y aumentando en incrementos de 10
valores_area <- max_area + seq(10, 100, by = 10)

# Crear conjunto de datos para pronósticos con cajones constante en 9
datos_pronosticos <- data.frame(
  area = valores_area,
  cajones = rep(9, 10) # Valor constante de 9 cajones
)

# Obtener pronósticos
pronosticos <- predict(modelo_revisado, newdata = datos_pronosticos)

# Crear tabla de resultados
tabla_pronosticos <- data.frame(
  Cajones = datos_pronosticos$cajones,
  Area = datos_pronosticos$area,
  Ventas_Pronosticadas = pronosticos
)

# Mostrar la tabla de pronósticos

pander(tabla_pronosticos, caption = "Tabla de Pronosticos con 9 cajones
de estacionamos y Area de piso ascendente")
```

Table 2: Tabla de Pronosticos con 9 cajones de estacionamos y Area de piso ascendente

| Cajones | Area | Ventas_Pronosticadas |
|---------|------|----------------------|
| 9 | 566 | 1885 |
| 9 | 576 | 1892 |
| 9 | 586 | 1900 |
| 9 | 596 | 1908 |
| 9 | 606 | 1915 |
| 9 | 616 | 1923 |
| 9 | 626 | 1931 |
| 9 | 636 | 1939 |
| 9 | 646 | 1946 |
| 9 | 656 | 1954 |

| Cajones | Area | Ventas_Pronosticadas |
|---------|------|----------------------|
|---------|------|----------------------|

4 Conclusiones Finales

El modelo final, basado únicamente en las variables significativas área y cajones, explica satisfactoriamente el comportamiento de las ventas. Al eliminar la variable ingreso (no significativa), se obtuvo un modelo más parsimonioso y robusto. Los pronósticos realizados demuestran que un incremento en el área de venta, manteniendo constante el número de cajones, genera un aumento proporcional en las ventas estimadas, lo que sugiere oportunidades de crecimiento mediante la expansión física del negocio.