

Práctica Dirigida N°2 de Análisis de Regresión

Asencios Menacho Luz

2024-09-18

Caso: Ventas de ropa de la tienda "TrendWear"

El gerente general de una cadena de tiendas de ropa "**TrendWear**" desea predecir el ingreso mensual por la venta de ropa (miles de dólares), considerando ciertas variables que pueden afectarlas. Las variables del conjunto de datos incluyen:

Ventas: Monto de venta mensual (en miles de dólares) de la tienda.

Publicidad: Monto gastado en campañas publicitarias por la tienda, (en miles de dólares).

Empleados: Número de empleados en la tienda.

Promociones: Número promedio de cupones de promociones utilizadas por los clientes al mes.

Los datos de 100 tiendas elegidas al azar a nivel nacional se presentan en el archivo PD2.xlsx.

```
library(readxl)
```

a) Realice el análisis descriptivo (medidas de resumen y gráficos)

```
# Cargar Los datos
```

```
PD2 <- read_excel("PD2.xlsx")
```

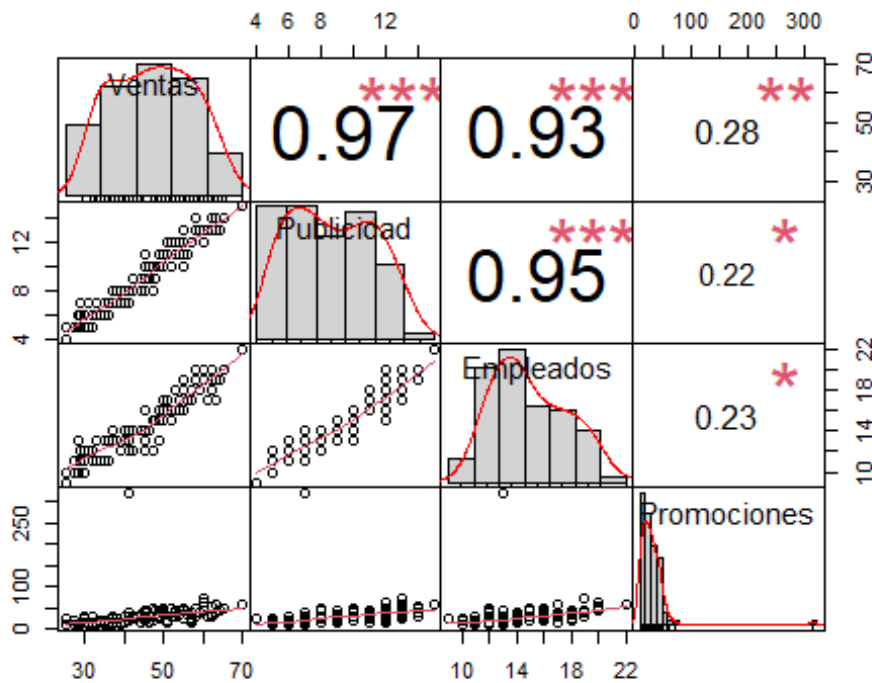
```
cor(PD2)
```

```
##           Ventas Publicidad Empleados Promociones
## Ventas      1.0000000  0.9662531 0.9306445  0.2814501
## Publicidad  0.9662531  1.0000000 0.9451205  0.2167756
## Empleados   0.9306445  0.9451205 1.0000000  0.2302557
## Promociones 0.2814501  0.2167756 0.2302557  1.0000000
```

Puede existir multicolinealidad entre Ventas y Empleados debido a su alta correlación.

```
#-- Gráficos (Matriz de dispersión)
```

```
pairs(PD2)
```

- b) Estime el modelo de regresión lineal múltiple para predecir la variable Ventas en función de todas las variables predictoras. Interprete el coeficiente correspondiente a la variable “Gasto de publicidad”

```
# Ajustar el modelo
modelo <- lm(Ventas ~ Publicidad + Empleados + Promociones, data = PD2)

modelo

##
## Call:
## lm(formula = Ventas ~ Publicidad + Empleados + Promociones, data =
## PD2)
##
## Coefficients:
## (Intercept)  Publicidad  Empleados  Promociones
##      6.22785      3.24410      0.56251      0.02498

# Resumen del modelo
summary(modelo)

##
## Call:
## lm(formula = Ventas ~ Publicidad + Empleados + Promociones, data =
## PD2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -6.513 -1.944 0.108 1.833 6.821
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.227854  1.876136   3.320 0.00128 **
## Publicidad  3.244097  0.302027  10.741 < 2e-16 ***
## Empleados   0.562507  0.292109   1.926 0.05710 .
## Promociones 0.024977  0.008854   2.821 0.00582 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.785 on 96 degrees of freedom
## Multiple R-squared:  0.9413, Adjusted R-squared:  0.9395
## F-statistic: 513.6 on 3 and 96 DF, p-value: < 2.2e-16
```

El coeficiente de la variable “Publicidad” indica el aumento promedio en las ventas por cada unidad adicional gastada en publicidad, manteniendo constantes las otras variables. Por ejemplo, si el coeficiente es 2, significa que por cada mil dólares adicionales en publicidad, las ventas aumentan en promedio 2 mil dólares.

- c) A un nivel de significación de 0.04 ¿Al menos una variable es significativa para el modelo?

Prueba de significancia global del modelo

anova(modelo)

```
## Analysis of Variance Table
##
## Response: Ventas
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## Publicidad  1 11853.3 11853.3 1528.1828 < 2.2e-16 ***
## Empleados   1   36.1   36.1    4.6524 0.033507 *
## Promociones  1   61.7   61.7    7.9569 0.005821 **
## Residuals  96   744.6    7.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Prueba de significancia individual de cada variable

summary(modelo) *# Los valores p en la columna Pr(>|t|) indican la significancia*

```
##
## Call:
## lm(formula = Ventas ~ Publicidad + Empleados + Promociones, data = PD2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.513 -1.944  0.108  1.833  6.821
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.227854   1.876136   3.320  0.00128 **
## Publicidad  3.244097   0.302027  10.741 < 2e-16 ***
## Empleados   0.562507   0.292109   1.926  0.05710 .
## Promociones 0.024977   0.008854   2.821  0.00582 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.785 on 96 degrees of freedom
## Multiple R-squared:  0.9413, Adjusted R-squared:  0.9395
## F-statistic: 513.6 on 3 and 96 DF,  p-value: < 2.2e-16
```

Un valor p menor a 0.04 indica que la variable es significativa al nivel de significancia del 5%.

d) A un nivel de significancia de 0.04, ¿Qué variables predictoras son significativas?

```
modelo_reducido <- lm(Ventas ~ Publicidad + Empleados, data = PD2)
modelo_reducido

##
## Call:
## lm(formula = Ventas ~ Publicidad + Empleados, data = PD2)
##
## Coefficients:
## (Intercept)  Publicidad  Empleados
##          6.0912         3.2418         0.6281
```

e) Estime el modelo que solo incluya a las variables predictoras significativas.

```
# ... (código anterior)

# Predicciones
nuevos_datos <- data.frame(Publicidad = 5, Empleados = 20, Promociones = 10)
prediccion <- predict(modelo_reducido, newdata = nuevos_datos)
print(prediccion)

##          1
## 34.86171
```

Con el modelo obtenido en la pregunta e) f) Estime e interprete un intervalo del 95% de confianza para el coeficiente asociado con la variable “número de empleados”.

```
# Obtener el intervalo de confianza
confint(modelo_reducido, level = 0.95)

##           2.5 %    97.5 %
## (Intercept) 2.23769593 9.944781
## Publicidad  2.62127470 3.862398
## Empleados   0.02978223 1.226347
```

El intervalo de confianza nos indica que con un 95% de confianza, el verdadero efecto del número de empleados sobre las ventas se encuentra dentro de este rango. Si el intervalo no incluye el cero, podemos concluir que el número de empleados tiene un efecto significativo sobre las ventas.

g) Calcule e interprete la medida de bondad de ajuste del modelo.

Un valor de R-cuadrado cercano a 1 indica que el modelo explica una gran proporción de la variabilidad en las ventas.

h) Diagnostique la presencia de multicolinealidad

```
# Matriz de correlación
cor(PD2[, c("Publicidad", "Empleados", "Promociones")])

##               Publicidad Empleados Promociones
## Publicidad      1.0000000  0.9451205    0.2167756
## Empleados       0.9451205  1.0000000    0.2302557
## Promociones     0.2167756  0.2302557    1.0000000

# VIF
library(car)

## Cargando paquete requerido: carData

vif(modelo_reducido)

## Publicidad  Empleados
##    9.367921    9.367921
```

i) ¿Cuánto sería la Venta mensual de una tienda a un 96% de confianza cuando el gasto publicitario de una tienda es de 12 mil USD, el número de empleados es de 18 y los clientes utilizan en promedio 60 cupones al mes?

```
# Nuevos datos
nuevos_datos <- data.frame(Publicidad = 12, Empleados = 18, Promociones = 60)

# Predicción con intervalo de confianza
predict(modelo_reducido, newdata = nuevos_datos, interval = "prediction",
level = 0.96)

##      fit      lwr      upr
## 1 56.29843 50.22104 62.37582
```

El intervalo de predicción nos indica que con un 96% de confianza, la venta mensual de una tienda con esas características estará entre los valores inferior y superior del intervalo.