STAT 401 Cassandra Davies Oct. 2023

1.(a) The single predictor that most accurately predicts sales is Youtube. I came to this conclusion by analyzing multiple models using each predictor (Youtube, Facebook, Newspaper) individually, in a separate simple linear model for each predictor, versus Sales, which is the response value on the y-axis. The R² values for each predictor are: Youtube R² = 0.6119, Facebook R² = 0.332, Newspaper R² = 0.05212, which indicates that Youtube is the most accurate predictor with the highest R² value. I used the following code in R Studio to find the R² values for each predictor:

Imagen de la pantalla de un celular de un mensaje en letras blancas

Descripción generada automáticamente con confianza baja

The linear function for the expected value of the response variable is:

Yhat = b0 + b1\*xp ; where b0 is the population intercept and b1 is the slope estimator.

(b) Here is a scatterplot of Sales versus the selected predictor Youtube (below left):

Gráfico, Gráfico de dispersión

Descripción generada automáticamente Texto

Descripción generada automáticamente con confianza media

(c) The regresssion standard error is 3.91 on 198 degrees of freedom; the coefficient of determination is 0.6119; the population slope is 0.047537. (See figure above right)

(d) Here is a plot that shows the fitted values versus the residuals (below left) and the predictor variable Youtube versus the residuals (below right):

Gráfico, Gráfico de dispersión

Descripción generada automáticamente Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Here are the histogram and the QQ plot for the same simple linear model for Youtube:

Gráfico, Histograma

Descripción generada automáticamente Gráfico, Gráfico de líneas

Descripción generada automáticamente

(e) The intercept estimate is 8.439112, the slope estimate is 0.047537, and both are shown in this code summary from R Studio (below left):

Texto

Descripción generada automáticamente con confianza media Texto

Descripción generada automáticamente con confianza baja

2.(a) Here is the code I used in R Studio to summarize the full multiple linear model data, which includes all three predictor variables: Youtube, Facebook, Newspaper (see above right).

(b) The single predictor that I decided to remove from the model is Newspaper. As noted in 1.(a), when tested in a simple linear model, Newspaper’s R² value was the lowest of the three predictors. In addition to the multiple linear model summarized in 2.(a), I tested three multiple linear models, each with a different combination of two of the three predictors.

Here are the model summaries for Youtube+Facebook (below left),

Facebook+Newspaper (below center), Newspaper+Youtube (below right):

Tabla

Descripción generada automáticamente Texto

Descripción generada automáticamente Texto

Descripción generada automáticamente

As you can see when comparing the three summary figures from R Studio (directly above), the Adjusted R² values further validate the findings of the R² values from the simple linear models conducted for each predictor variable in 1.(a). For example, the Adjusted R² value for Youtube+Facebook is 0.8962, the Adjusted R² value for Facebook+Newspaper is 0.3259, and the Adjusted R² value for Newspaper+Youtube is 0.6422. The model without Newspaper has the highest Adjusted R² value, and so we can confidently remove Newspaper from the full model in our future analysis.

(c) After dropping Newspaper as as unnecessary predictor variable, we are ready for the final evaluation and interpretation of the multiple linear model of Sales response to two predictor variables Youtube+Facebook. The regresssion standard error is 2.018 on 197 degrees of freedom; the coefficient of determination is 0.8962; the population slope is 0.11687, averaged from 0.04575 and 0.18799; the intercept estimate is 3.50532; the slope estimate is 0.11687.

(See figure at top of next page.)

Tabla

Descripción generada automáticamente

3.The simple linear model of Sales versus Youtube has an R² value of 0.6119. The multiple linear model of Sales versus Youtube+Facebook has an Adjusted R² value of 0.8962. This indicates that the multiple linear model of Sales versus Youtube+Facebook is the model that fits the data best.

The 95% confidence interval for the expected sales is $16.55 to $17.11. I found this value by taking the mean, or average value of each predictor and using that as a data frame input for the predict function with confidence interval in R Studio. Here’s the code in R Studio:

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

The 95% prediction interval for the expected sales is $12.84 to $20.82. I found this value by taking the mean, or average value of each predictor and using that as a data frame input for the predict function with prediction interval in R Studio. Here’s the code in R Studio:

Texto

Descripción generada automáticamente

The 95% prediction interval for an individual Y-value with the observation (Youtube, Facebook, Newspaper) = (50.9, 42.1, 87.6) is $9.65 to $17.76. Here is the code in R Studio I used to find this:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

This is my own unaided work.

* Cassandra Davies