PRICING SOFT DRINK WITH A WHOLE-SALE (USING TEMPERATURE VS SALES DATASET)

Mayuri.A
2348133

Pricing soft drinks with a whole-sale (using temperature vs sales dataset)

Mayuri.A

10-11-2023

Problem statement:

We have to analyze if there is any relationship between temperature in Celsius and sales of soft drinks in a particular city in India. This data is recorded at random throughout the year from different shops across the city. A particular wholesaler wants to make deals with different shops for its pricing in different months. Let us use the given dataset of 30 days and do a complete simple linear regression analysis.

Objective:

a. Comment about the different steps involved in building a simple linear regression model
b. Plot the scatter diagram for the data and find the coefficient of correlation. What do you infer from the scatter plot?
c. Estimate the parameters of a simple linear regression model and fit a regression line. Interpret the results.
d. Test the significance of the regression coefficient and interpret the results.
e. What are the different ways in which u can assess the quality of the fit.

Introduction:

Given that there exists a pricing concern with the soft drinks that are to be supplied to different shops in a city, we are going to build a simple linear regression model. and understand the relation between sales of soft drinks and the temperature.

Variables of interest:

Dependent Variable: sales of the soft drink in lakhs

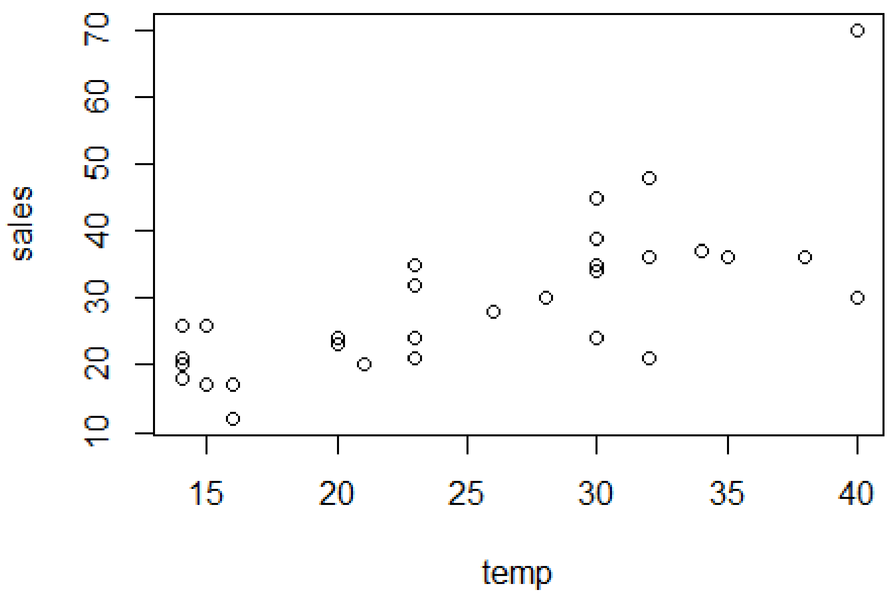
Independent Variable: temperature in Celsius

Analysis:

Step1: Visualization and Correlation

Let us plot the graph between the dependent and independent variables to visualize their relation and their correlation to understand the strength of the relation.

```
temp=c(23,30,16,30,32,14,40,14,20,15,30,32,38,34,35,21,23,26,28,23,23,30,16,30,32,14,40,14,20,15)
sales=c(35,39,12,34,36,20,70,18,23,17,45,48,36,37,36,20,21,28,30,32,24,35,17,24,21,26,30,21,24,26)
plot(temp,sales)
```




```
cor(temp,sales)

## [1] 0.7147753
```

Interpretation: we get a linear relationship between the independent and dependent variables. and a strong positive linear relationship is observed from the karl pearson correlation coefficient with 0.7147.

Step2: Fit the model

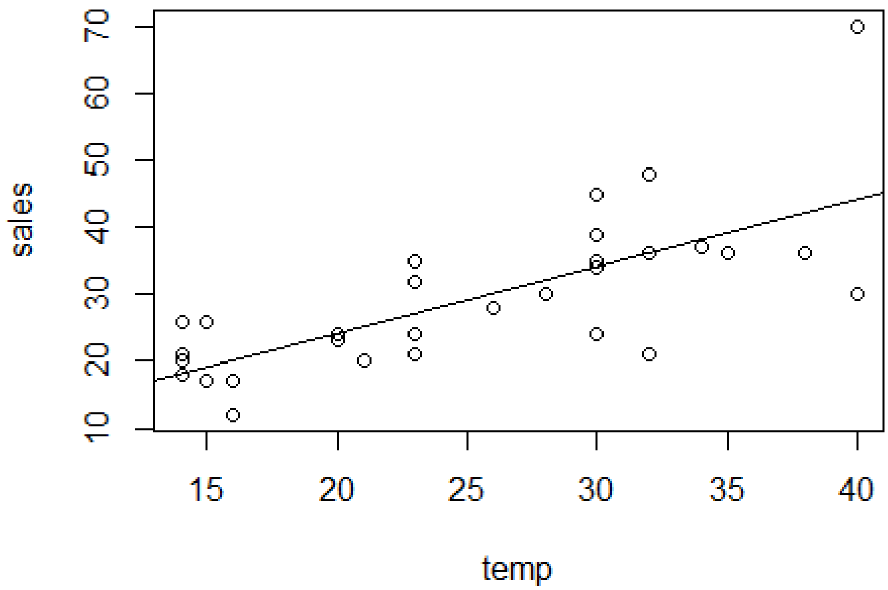
```
reg1=lm(sales~temp)
reg1 #fitting a simple regression model

##
## Call:
## lm(formula = sales ~ temp)
##
## Coefficients:
## (Intercept)        temp
##      4.3575      0.9951
```

Interpretation: we got the value of beta(0)=4.3575 and the value of beta(1)=0.9951 based on sample. since the sign of beta 1 is positive, there is a positive relationship between the variables. for each additional 1-degree increase in temperature, there is a 0.9951 increase in sales of ice cream ~1 ice cream.

Step 3: Fitting the regression line

```
plot(temp,sales)
abline(reg1) #fitting the regression line Y=4.3575+0.9951*x into the existing scatter plot.
```



points-original values of Y vs line-fitted values of Y. The deviation from each point and line will give the errors.

Now that we have understood the variables, visualized the relationship and measured using Karl-person. We have also estimated beta(0) and beta(1) using a sample. however, let us see if it is a significant value and check for its goodness of fit.

Step4: Summary

the overall fit of the regression model can be obtained with a summary command. which is required to devise the hypothesis and goodness of fit

```
summary(reg1)

##
## Call:
## lm(formula = sales ~ temp)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -15.2002  -3.2703  -0.7396   4.2445  25.8391
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.3575     4.8866   0.892     0.38
## temp          0.9951     0.1840   5.408  9.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.245 on 28 degrees of freedom
## Multiple R-squared:  0.5109, Adjusted R-squared:  0.4934
## F-statistic: 29.25 on 1 and 28 DF,  p-value: 9.097e-06
```

Hypothesis Testing:

H0= Beta(1)=0

H1= Beta(1)=! 0

From the above table, the p-value in the line of units is 9.1e-06, which is less than 0.05 level of significance. hence we reject the null hypothesis, and hence beta (1)=! 0 . thus, there exists a significant linear relationship between the variables., ie there is a significant linear relationship between the temperature of the day and sales of ice cream.

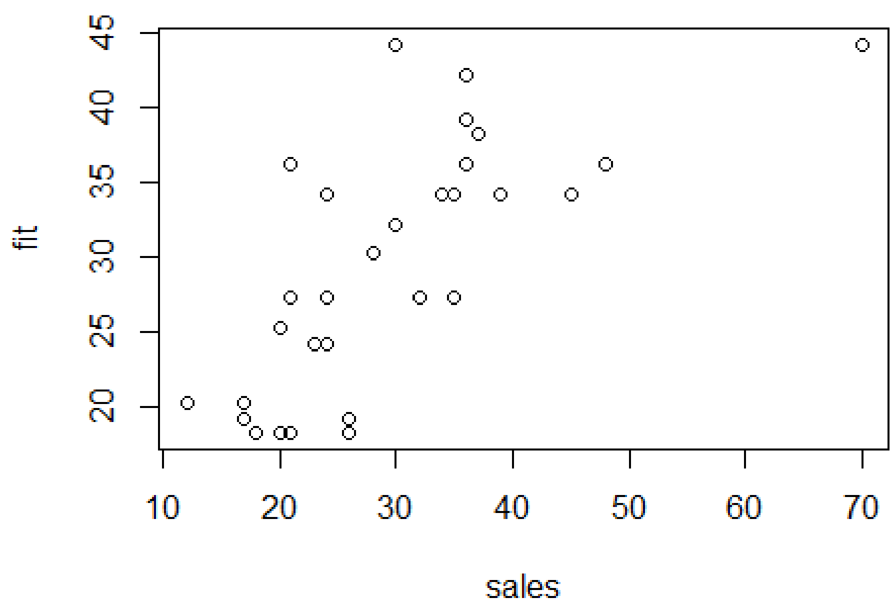
The goodness of fit: if t value 5.408>> is significantly larger than t(calc)=2.048, there exists a very strong linear relation. alternatively, since the p-value is very small compared to 0.05, there exists a very strong linear relation. This is also observed in the graph where the error is very small (Y and Y hat are very close to each other.); hence the model is a good fit.

Step 5: Fitting the observed and predicted values

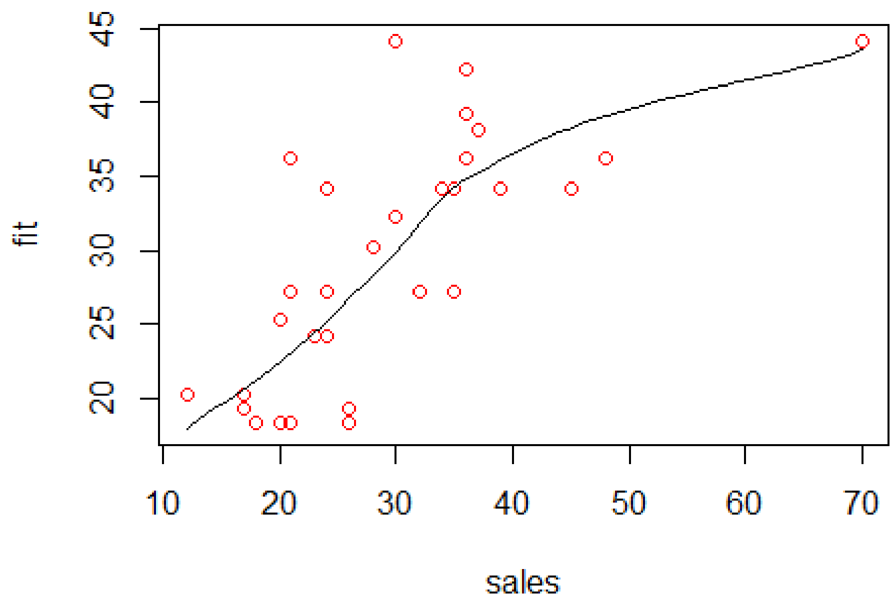
```
fit=fitted.values(reg1)
fit
```

```
##        1        2        3        4        5        6        7        8
## 27.24447 34.21007 20.27887 34.21007 36.20025 18.28870 44.16093 18.28870
##        9       10       11       12       13       14       15       16
## 24.25921 19.28378 34.21007 36.20025 42.17076 38.19042 39.18550 25.25430
##       17       18       19       20       21       22       23       24
## 27.24447 30.22973 32.21990 27.24447 27.24447 34.21007 20.27887 34.21007
##       25       26       27       28       29       30
## 36.20025 18.28870 44.16093 18.28870 24.25921 19.28378
```

```
plot(sales,fit)
```




```
scatter.smooth(sales,fit,col='red')
```



##

Step6: Coefficient of determination

```
r=cor(sales,fit)
r

## [1] 0.7147753

r^2

## [1] 0.5109037
```

r^2=0.51 ie,- 51% of the data's total variability is explained by the independent variable of the dependent variable. (note: at least 50% of data should be explained for consideration of prediction.) for it to be a model that can be considered as a good fit. according to the r^2, the model is a good fit for the data.4

Conclusion:

We have developed a linear regression model of Y=4.3575+0.9951*X +E. This model has been tested for the hypothesis that H0= Beta (1)=0 H1= Beta(1)=! 0, and we have rejected the null hypothesis and concluded that beta 1 is not zero; thus, there exists a significant relationship between the variables. also, by calculating the correlation of determination r^2=0,51, 51% of the data's total variability is explained by the independent variable of the dependent variable.

Also, we can say that, indeed, the shopkeeper can model his pricing based on these two variables and can make a significant change in the price during months that are hotter.