Test Exercise 1

Abolfazl Babanazari

4/2/2022

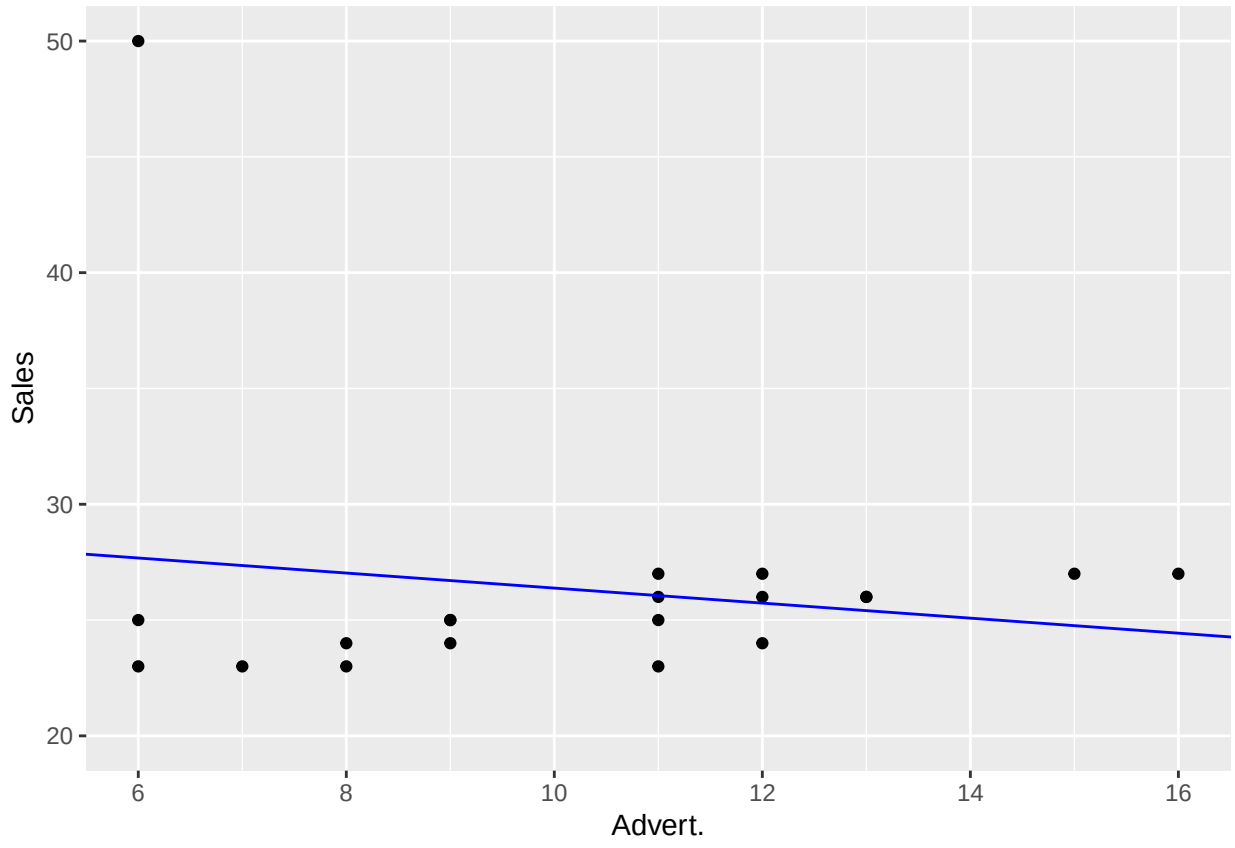
Question 1

Make the scatter diagram with sales on the vertical axis and advertising on the horizontal axis. What do you expect to find if you would fit a regression line to these data?

Answer

There is an anomaly in the data ((6, 50)) which seems to effect the linear relation between other points. In the below plot, blue line shows fitted regression line which is quite strange.

I think we should eliminate that point from the dataset, and then regression model demonstrates a meaningful progressive correlation between advertising and sales amount.



Question 2

Estimate the coefficients a and b in the simple regression model with sales as dependent variable and advertising as explanatory factor. Also compute the standard error and t-value of b. Is b significantly different from 0?

Answer

I have used R-functions to calculate the parameters.

```
model <- lm(Sales ~ Advert., data = df)
```

The results are as follows

- The intercept of regression (α) is equal to 29.63 and the slope (β) is equal to -0.32
- The t-value and standard error of β are respectively -0.71 and 0.46.

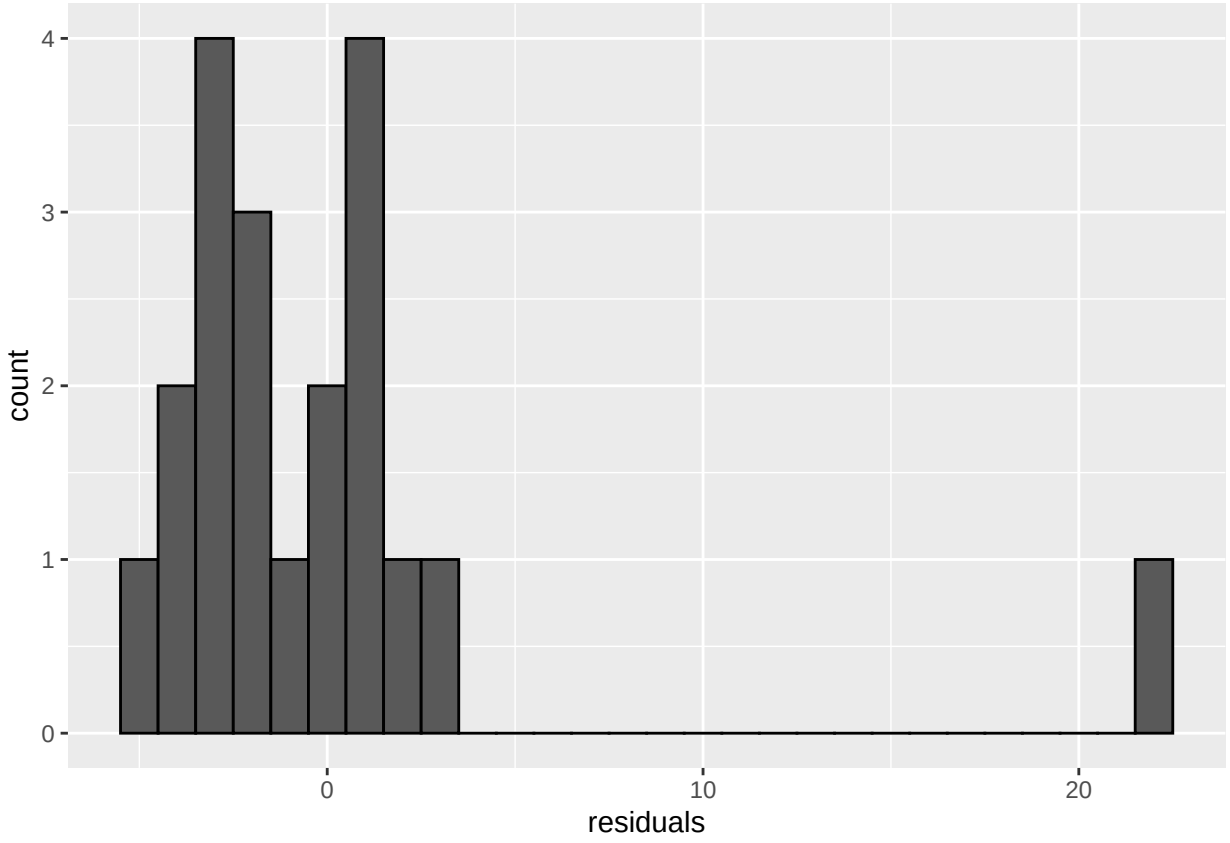
By knowing above results, we can calculate 95% confident interval, which is (-1.24, 0.59). The interval contains zero, therefore `Advertise.` is not significant than zero.

Question 3

Compute the residuals and draw a histogram of these residuals. What conclusion do you draw from this histogram?

Answer

In the first glimpse the plot doesn't look like if it's normal distributed, However without considering that one out bar, the overall plot is roughtly bell shaped.



Question 4

Apparently, the regression result of part (b) is not satisfactory. Once you realize that the large residual corresponds to the week with opening hours during the evening, how would you proceed to get a more satisfactory regression model?

Answer

By removing this point and applying another linear regression to the data, hopefully results would be more explanatory.

Question 5

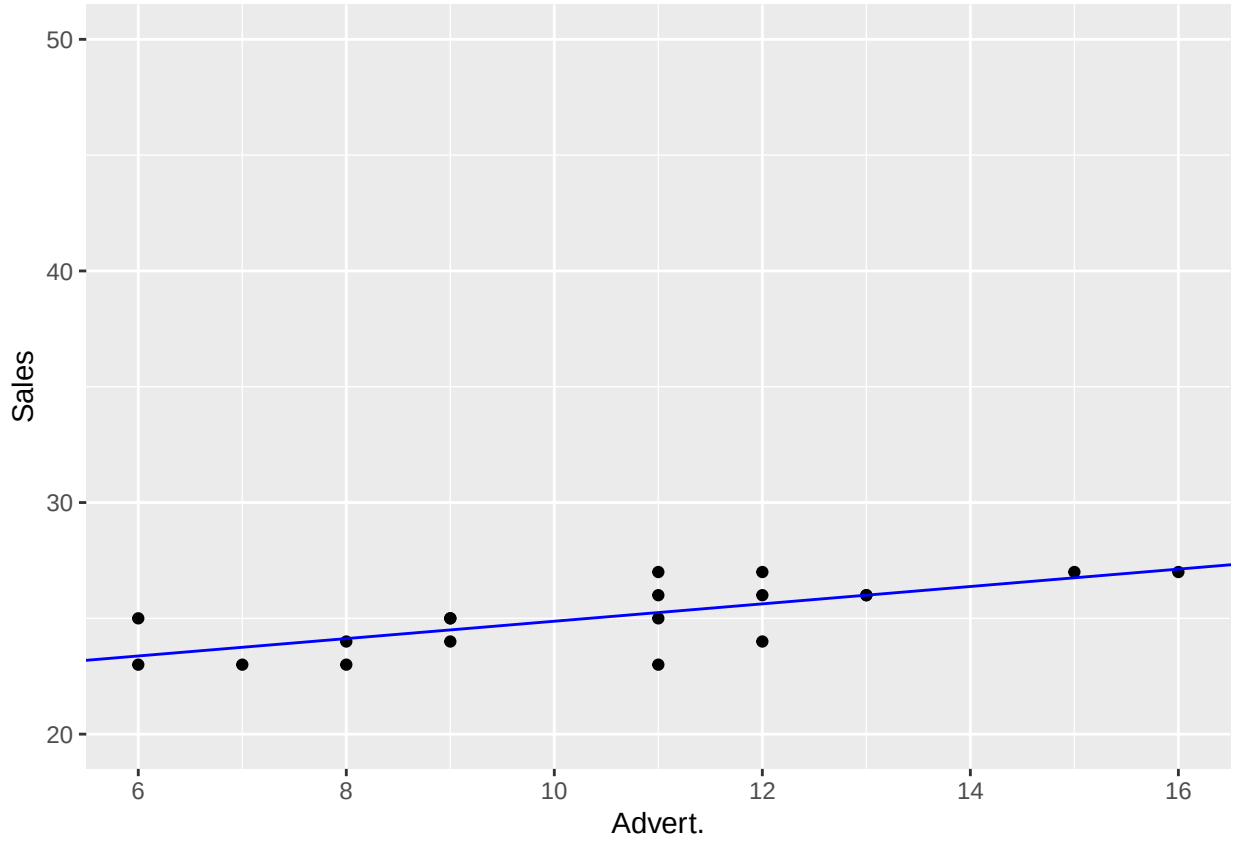
Apparently, the regression result of part (b) is not satisfactory. Once you realize that the large residual corresponds to the week with opening hours during the evening, how would you proceed to get a more satisfactory regression model?

Answer

To do so I have used following commands (note that 12th observation is anomaly)

```
newdf <- df[-c(12), ]

model <- lm(Sales ~ Advert., data = newdf)
```



The results are as follows

- The intercept of regression (α) is equal to 21.12 and the slope (β) is equal to 0.37
- The t-value and standard error of β are respectively 4.25 and 0.09.

- Also R-squared in this cause is 0.52 which is enough to say that `Advertise.` is significant.

By knowing above results, we can calculate 95% confident interval, which is (0.2, 0.55). The interval doesn't contain zero, therefore `Advertise.` is now significant than zero.

Question 6

Discuss the differences between your findings in parts (b) and (e). Describe in words what you have learned from these results.

Answer

In section (b), when regression were trying to fit the points, that anomaly points posed a large residual which disrupted the fitted line. In other words, there were a unseen variable which had a relation with both x and y (violation to A3 and A4).

By removing that point from dataset in section (e), the actual relation between variables reveals.