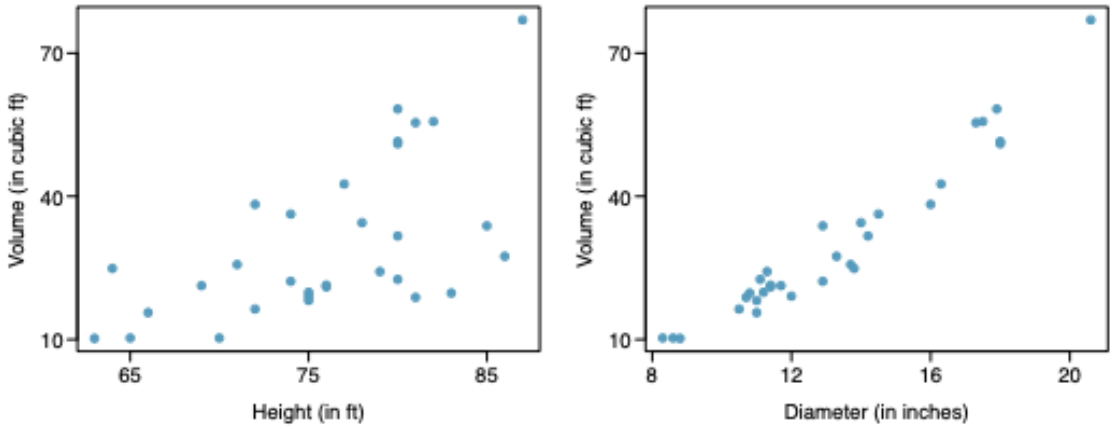
Estadística

U5 5A

David Aarón Ramírez Olmeda

Abril 2023

7.12 Trees. The scatter plots below show the relationship between height, diameter, and volume of timber in 31 felled black cherry trees. The diameter of the tree is measured 4.5 feet above the ground.



(a) **Describe the relationship between volume and height of these trees.**

The scatterplot between volume and height shows a positive relationship, which means that as the height of the tree increases, so does the volume of timber. However, the relationship is not very strong, as there is quite a bit of variability in the data.

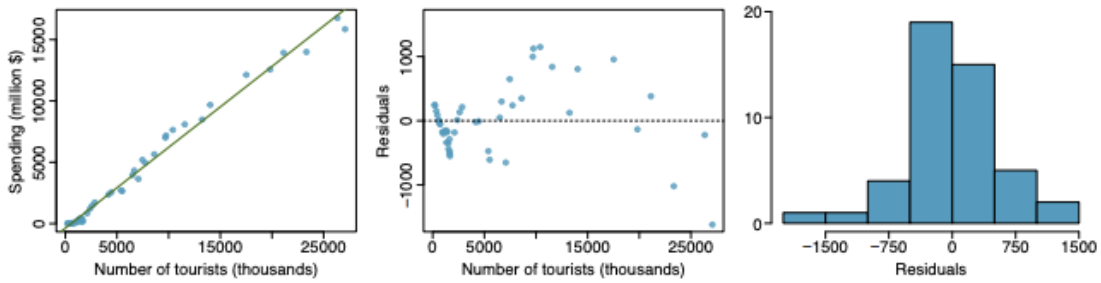
(b) **Describe the relationship between volume and diameter of these trees.**

The scatterplot between volume and diameter shows a stronger positive relationship than the one between volume and height. The relationship appears to be quite strong, as the data points are more tightly clustered around a straight line.

(c) **Suppose you have height and diameter measurements for another black cherry tree. Which of these variables would be preferable to use to predict the volume of timber in this tree using a simple linear regression model? Explain your reasoning.**

Based on the scatterplots, it would be preferable to use diameter to predict the volume of timber in a black cherry tree using a simple linear regression model. The relationship between diameter and volume is stronger than the relationship between height and volume.

7.23 Tourism spending. The Association of Turkish Travel Agencies reports the number of foreign tourists visiting Turkey and tourist spending by year. Three plots are provided: scatterplot showing the relationship between these two variables along with the least squares fit, residuals plot, and histogram of residuals.



(a) **Describe the relationship between number of tourists and spending.**

Based on the scatterplot, there appears to be a positive relationship between the number of tourists visiting Turkey and tourist spending. As the number of tourists increases, so does the amount of money spent.

(b) **What are the explanatory and response variables?**

The number of tourists is the explanatory variable and tourist spending is the response variable.

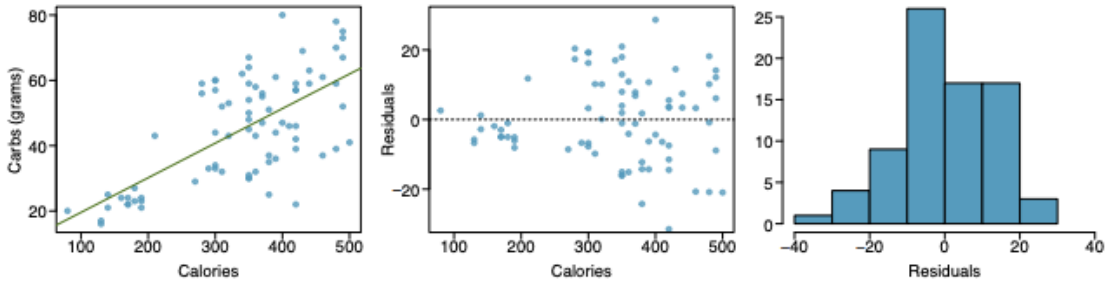
(c) **Why might we want to fit a regression line to these data?**

We might want to fit a regression line to these data to model the relationship between the number of tourists and tourist spending. This could help us make predictions about future tourist spending based on the number of tourists visiting Turkey.

(d) **Do the data meet the conditions required for fitting a least squares line? In addition to the scatterplot, use the residual plot and histogram to answer this question.**

Based on the residual plot and histogram provided, the data seem to meet the conditions required for fitting a least squares line. However, there does seem to be some variability in the spread of the residuals as the number of tourists increases, which may indicate some heteroscedasticity. Since the variability does not appear to be too extreme, a least squares line may still be appropriate for these data.

7.24 Nutrition at Starbucks, Part I. The scatterplot below shows the relationship between the number of calories and amount of carbohydrates (in grams) Starbucks food menu items contain. Since Starbucks only lists the number of calories on the display items, we are interested in predicting the amount of carbs a menu item has based on its calorie content.



(a) Describe the relationship between number of calories and amount of carbohydrates (in grams)

There does not seem to be a strong relationship between the number of calories and amount of carbohydrates, however, there is a general trend where as the number of calories increases, the amount of carbohydrates also tends to increase.

(b) In this scenario, what are the explanatory and response variables?

The number of calories is the explanatory variable and the amount of carbohydrates is the response variable.

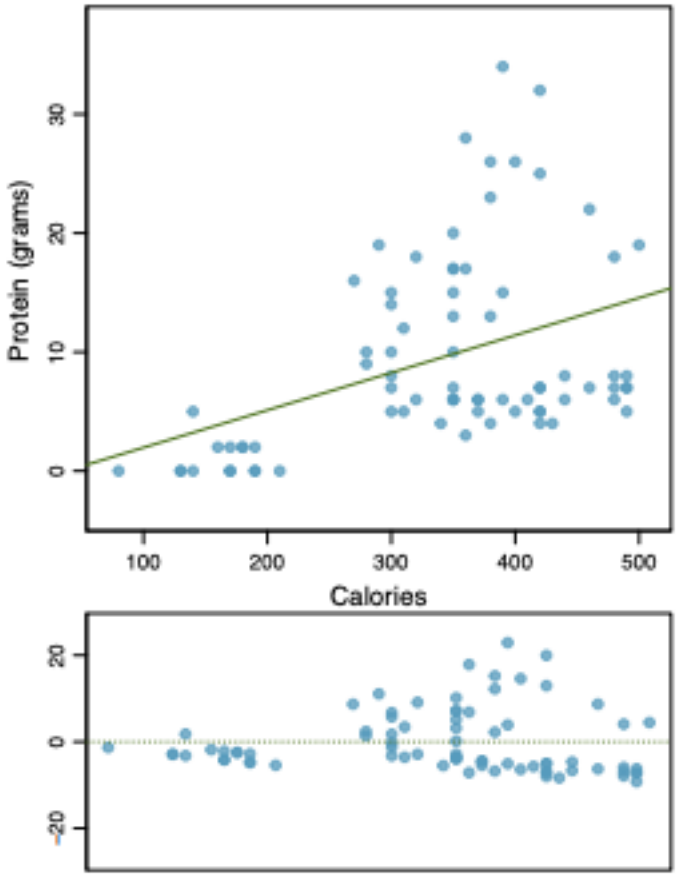
(c) Why might we want to fit a regression line to these data?

We might want to fit a regression line to these data to predict the amount of carbohydrates in a Starbucks item based on its calorie content, this could be useful for people who are interested in monitoring their carbohydrate intake.

(d) Do these data meet the conditions required for fitting a least squares line?

The data seem to meet the conditions required for fitting a least squares line. But we cannot confirm whether the residuals are normally distributed or whether the variability of the residuals is constant across all levels of the explanatory variable without additional information.

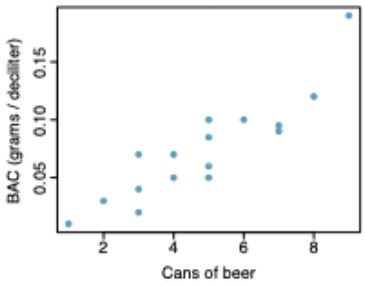
7.27 Nutrition at Starbucks, Part II. Exercise 7.24 introduced a data set on nutrition information on Starbucks food menu items. Based on the scatterplot and the residual plot provided, describe the relationship between the protein content and calories of these menu items, and determine if a simple linear model is appropriate to predict amount of protein from the number of calories.



Based on the scatterplot there seems to be a weak positive linear relationship between the protein content and calories, although there is a lot of variability in the data.

The residual plot shows some curvature, which suggests that a simple linear model may not be appropriate for predictions. There appears to be some heteroscedasticity in the residuals, which supports this conclusion. It may be necessary to consider a more complex model.

7.36 Beer and blood alcohol content. Many people believe that gender, weight, drinking habits, and many other factors are much more important in predicting blood alcohol content (BAC) than simply considering the number of drinks a person consumed. Here we examine data from sixteen student volunteers at Ohio State University who each drank a randomly assigned number of cans of beer. These students were evenly divided between men and women, and they differed in weight and drinking habits. Thirty minutes later, a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood. The scatterplot and regression table summarize the findings.



	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0127	0.0126	-1.00	0.3320
beers	0.0180	0.0024	7.48	0.0000

(a) Describe the relationship between the number of cans of beer and BAC.

The scatterplot shows a strong positive linear relationship between the number of cans of beer and BAC. As the number of cans of beer increases, so does BAC.

(b) Write the equation of the regression line. Interpret the slope and intercept in context.

The equation of the regression line is:

BAC = -0.0127 + 0.018*(x)

For each additional can of beer, BAC is expected to increase by about .018

(c) Do the data provide strong evidence that drinking more cans of beer is associated with anincrease in blood alcohol? State the null and alternative hypotheses, report the p-value, and state your conclusion

The null hypothesis is that there is no association between the number of cans of beer and BAC, while the alternative hypothesis is that there is a positive association between the two variables. The p-value for the slope coefficient is less than 0.05, indicating strong evidence that the number of cans of beer is positively associated with BAC.

(d) The correlation coefficient for number of cans of beer and BAC is 0.89. Calculate R2 and interpret it in context.

The coefficient of determination R2 is 0.792, which means that 79.2% of the variability in BAC can be explained by the number of cans of beer consumed. This indicates a strong relationship between the two variables.

(e) Suppose we visit a bar, ask people how many drinks they have had, and also take their BAC. Do you think the relationship between number of drinks and BAC would be as strong as the relationship found in the Ohio State study?

The relationship between number of drinks and BAC may not be as strong in a bar setting as it was in the Ohio State study. This is because there are many other factors that can influence BAC, such as weight, gender, drinking habits, etc. Also, reported number of drinks may not be accurate, as people may underestimate the number of drinks they have consumed.