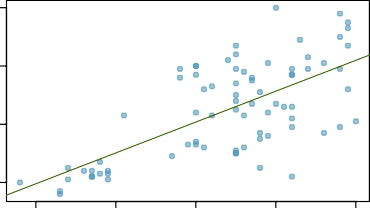
Chapter 8 - Introduction to Linear Regression

**Nutrition at Starbucks, Part I.** (8.22, p. 326) 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.

80

60

Carbs (grams)

40

20

100 200 300 400 500

Calories

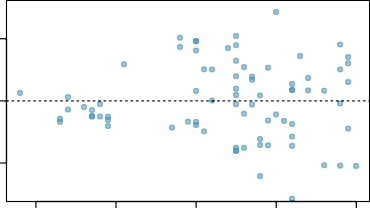
20

0

Residuals

−20

25

20

15

10

5

0

100 200 300 400 500

Calories



−40 −20 0 20 40

Residuals

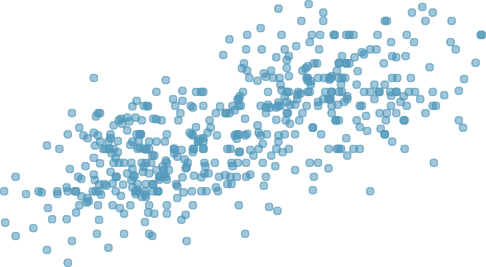
1. Describe the relationship between number of calories and amount of carbohydrates (in grams) that Starbucks food menu items contain.
2. In this scenario, what are the explanatory and response variables?
3. Why might we want to fit a regression line to these data?
4. Do these data meet the conditions required for fitting a least squares line?

**Answer**

1. There is a direct correlation between calorie content and carbohydrate levels, meaning as carbohydrates increase, so does the calories.
2. The x-axis represents calories, which is the explanatory variable, while the y-axis represents carbohydrates, the response variable.
3. Here, our goal is to estimate the carbohydrate amount using the calorie count as a basis for prediction.
4. The data follows a linear trend, and the residuals seem to be approximately normal, but we are unable to maintain constant variance.

**Body measurements, Part I.** (8.13, p. 316) Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender for 507 physically active individuals. The scatterplot below shows the relationship between height and shoulder girth (over deltoid muscles), both measured in centimeters.

200



190

180

Height (cm)

170

160

150

90 100 110 120 130

Shoulder girth (cm)

1. Describe the relationship between shoulder girth and height.
2. How would the relationship change if shoulder girth was measured in inches while the units of height remained in centimeters?

**Answer**

1. The connection between shoulder girth and height is generally such that as shoulder girth increases, height tends to increase as well. In most cases, this indicates that individuals with larger shoulder girths are usually taller.
2. Even if shoulder girth were measured in inches while height stayed in centimeters, the relationship between the two would remain unchanged.

**Body measurements, Part III.** (8.24, p. 326) Exercise above introduces data on shoulder girth and height of a group of individuals. The mean shoulder girth is 107.20 cm with a standard deviation of 10.37 cm. The mean height is 171.14 cm with a standard deviation of 9.41 cm. The correlation between height and shoulder girth is 0.67.

1. Write the equation of the regression line for predicting height.
2. Interpret the slope and the intercept in this context.
3. Calculate *R*2 of the regression line for predicting height from shoulder girth, and interpret it in the context of the application.
4. A randomly selected student from your class has a shoulder girth of 100 cm. Predict the height of this student using the model.
5. The student from part (d) is 160 cm tall. Calculate the residual, and explain what this residual means.
6. A one year old has a shoulder girth of 56 cm. Would it be appropriate to use this linear model to predict the height of this child?

**Answer**

1. The predicted height (ŷ) can be calculated using the formula: 105.8445 plus 0.6091 times the shoulder girth. ŷ =105.8445 + 0.6091 ∗ ‘shouldergirth‘
2. For each additional centimeter of shoulder girth, the height would increase by 0.6091 cm. If a person's shoulder girth were zero, the expected height would be 0.6091 cm.
3. R = 0.67  
    R² = R × R  
   Therefore, R² equals 0.4489.
4. R(100) = B₀ + B₁ × 100  
    R(100)= 166.7545
5. Residual = Actual - Predicted

Res = 160 - 166.7545  
Res = -6.7545

The model predicted the individual's height to be higher than it actually is.

1. This can only be accomplished through extrapolation. We are assuming that a linear relationship can be established with data that hasn't been previously analyzed.

**Cats, Part I.** (8.26, p. 327) The following regression output is for predicting the heart weight (in g) of cats from their body weight (in kg). The coefficients are estimated using a dataset of 144 domestic cats.

Estimate Std. Error t value Pr(*>|*t*|*) (Intercept) -0.357 0.692 -0.515 0.607

body wt 4.034 0.250 16.119 0.000

*s* = 1*.*452 *R*2 = 64*.*66% *R*2 = 64*.*41%

*adj*

20

15

Heart weight (g)

10

5

2.0 2.5 3.0 3.5 4.0

Body weight (kg)

1. Write out the linear model.
2. Interpret the intercept.
3. Interpret the slope.
4. Interpret *R*2.
5. Calculate the correlation coefficient.

**Answer**

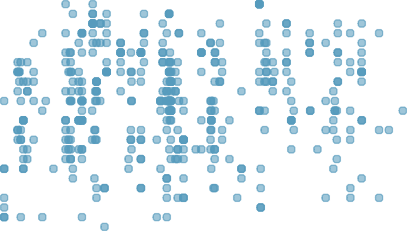
1. B₀ = -0.357  
   B₁ = 4.034  
   The predicted value (ŷ) can be expressed as: ŷ = - 0.357 + 4.034 × ‘body weight’.
2. If a cat's body weight is zero, we would expect its heart to weigh -0.357 grams.
3. For every extra kilogram of body weight, we can anticipate that a cat's heart will weigh an additional 4.034 grams.
4. R² = 64.66%, indicating that 64.66% of the observed data can be accounted for by the linear model presented in (a).
5. R² = 0.6466  
   corcoef = √(R²)  
   corcoef = 0.8041144

**Rate my professor.** (8.44, p. 340) Many college courses conclude by giving students the opportunity to evaluate the course and the instructor anonymously. However, the use of these student evaluations as an indicator of course quality and teaching effectiveness is often criticized because these measures may reflect the influence of non-teaching related characteristics, such as the physical appearance of the instructor. Researchers at University of Texas, Austin collected data on teaching evaluation score (higher score means better) and standardized beauty score (a score of 0 means average, negative score means below average, and a positive score means above average) for a sample of 463 professors. The scatterplot below shows the relationship between these variables, and also provided is a regression output for predicting teaching evaluation score from beauty score.

Estimate Std. Error t value Pr(*>|*t*|*) (Intercept) 4.010 0.0255 157.21 0.0000

beauty 0.0322 4.13 0.0000

5



4

Teaching evaluation

3

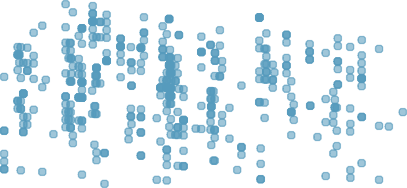
2

−1 0 1 2

Beauty

1. Given that the average standardized beauty score is -0.0883 and average teaching evaluation score is 3.9983, calculate the slope. Alternatively, the slope may be computed using just the information provided in the model summary table.
2. Do these data provide convincing evidence that the slope of the relationship between teaching evaluation and beauty is positive? Explain your reasoning.
3. List the conditions required for linear regression and check if each one is satisfied for this model based on the following diagnostic plots.

150



1

Residuals

100

0

50

−1

−1 0 1 2

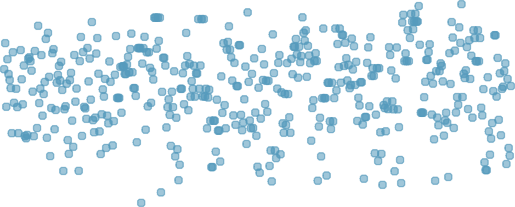
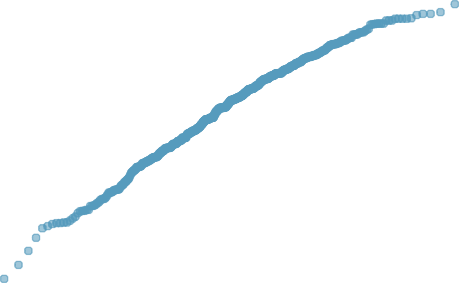
Beauty

0

−2 −1 0 1 2

Residuals

1



Sample Quantiles

0.5

0

−0.5

Residuals

−1

−1.5

−3 −2 −1 0 1 2 3

Theoretical Quantiles

**Answer**

1. B₀ = 4.010  
   B₁ = 4.13 × 0.0322  
     
   The slope value is 0.132986.
2. Upon examining the scatter plot, we observe a random distribution of points with no discernible upward or downward trend. The p-value is reported as zero in the summary table, indicating acceptance of the null hypothesis. So, this suggests there is no relationship between teaching evaluations and physical attractiveness.
3. - A visual assessment of the scatterplot indicates that the residuals are randomly distributed around the horizontal axis, demonstrating that linearity is present.

- The histogram exhibits left skewness, suggesting the potential presence of outliers. However, the residuals appear to follow a nearly normal distribution.

- The scatterplot also shows that the points maintain a constant variance.

- These observations are independent, displaying only a slight linear trend.

1. 0 100 200 300 400

Order of data collection