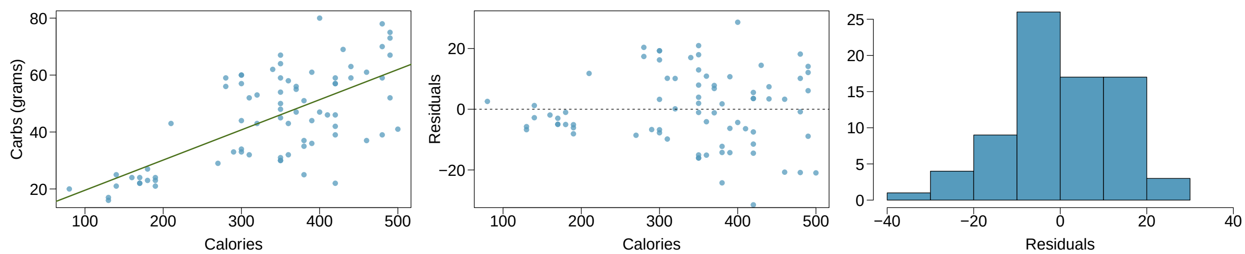
Chapter 8 - Introduction to Linear Regression

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.



(a) Describe the relationship between number of calories and amount of carbohydrates (in grams) that Starbucks food menu items contain.

Answer: There is a positive linear relationship between number calories and carbs.

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

Answer: The explanatory variable is Calories and the response variable is Carbs

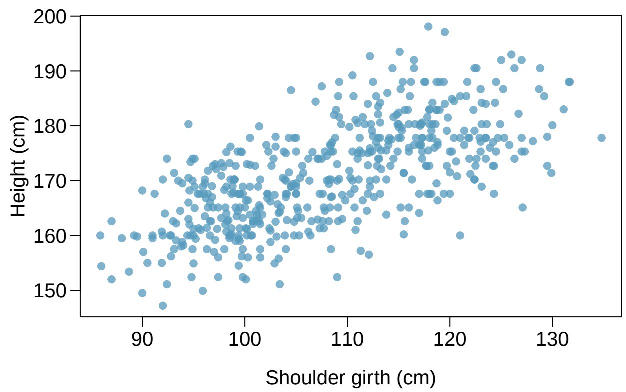
(c) Why might we want to fit a regression line to these data?

Answer: Predict the carbs in the menu based on the calories.

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

Not completely. There is linearity as the relationship between calories and carbs. The residuals appear to follow a normal distribution. However, the variability is not constant.

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.



(a) Describe the relationship between shoulder girth and height.

Answer: There is a positive, linear association between the two variables.

(b) How would the relationship change if shoulder girth was measured in inches while the units of height remained in centimeters?

Answer: The relationships would become more linear.

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.

(a) Write the equation of the regression line for predicting height.

The equation is y = 105.8391 + 0.6091498 * x

```
var1 <- 0.67 * (9.41/10.35)
var2 <- (var1 * -107.20) + 171.14

var1
```

```
## [1] 0.6091498
```

```
var2
```

```
## [1] 105.8391
```

(b) Interpret the slope and the intercept in this context.

Answer: For every centimeter in shoulder girth there is a predicted additional .6091498 cm to height.

(c) Calculate R^2 of the regression line for predicting height from shoulder girth, and interpret it in the context of the application.

The R^2 is 0.4489. The value says that 44.89% of the variability in height is explained by shoulder girth.

```
0.67^2
```

```
## [1] 0.4489
```

(d) A randomly selected student from your class has a shoulder girth of 100 cm. Predict the height of this student using the model.

```
guess_ht <- 105.9651 + 0.6091498*100
guess_ht
```

```
## [1] 166.8801
```

Answer: 166.8801

(e) The student from part (d) is 160 cm tall. Calculate the residual, and explain what this residual means.

Residual = -6.8801. The model overestimated the height.

```r
residual <- 160 - 166.8801
residual
```

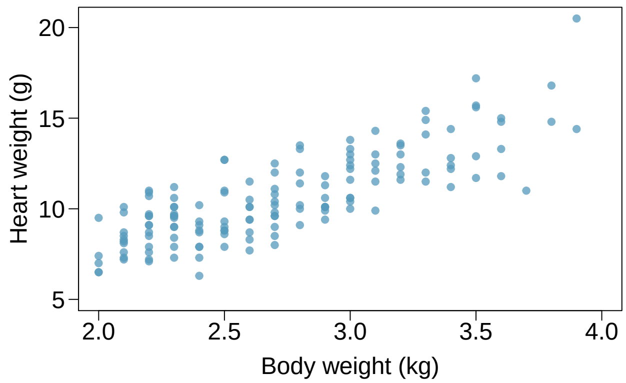
```
## [1] -6.8801
```

(f) 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?

The girth of the child is outside of the scope of the linear model, so it would not be appropriate.

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
	$s = 1.452$	$R^2 = 64.66\%$	$R^2_{adj} = 64.41\%$	



(a) Write out the linear model.

The linear model: y = -0.357 + 4.034 * x

(b) Interpret the intercept.

With a body weight of 0, there will be a heart weight of -0.357.

(c) Interpret the slope. An increase in body weight, there is a predicted increase of the heart's weight by 4.034g

(d) Interpret R^2.

the Body's weight explains 64.66% of the variability in the heart's weight.

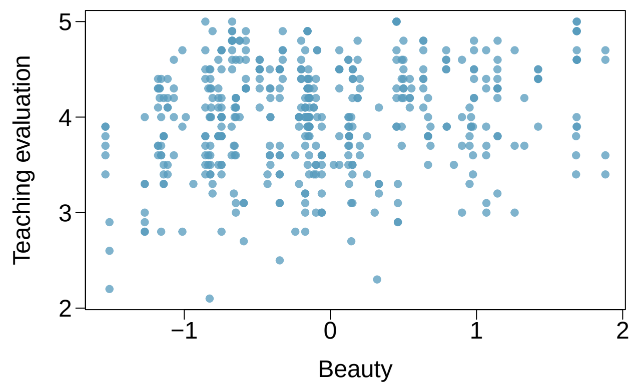
(e) Calculate the correlation coefficient. the correlation coefficient is .8041274

```
cor(cats$Hwt, cats$Bwt)
```

```
## [1] 0.8041274
```

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	$\boxed{}$	0.0322	4.13	0.0000



(a) 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.

Answer: 0.132986

```
slope <- 0.0322 * 4.13
slope
```

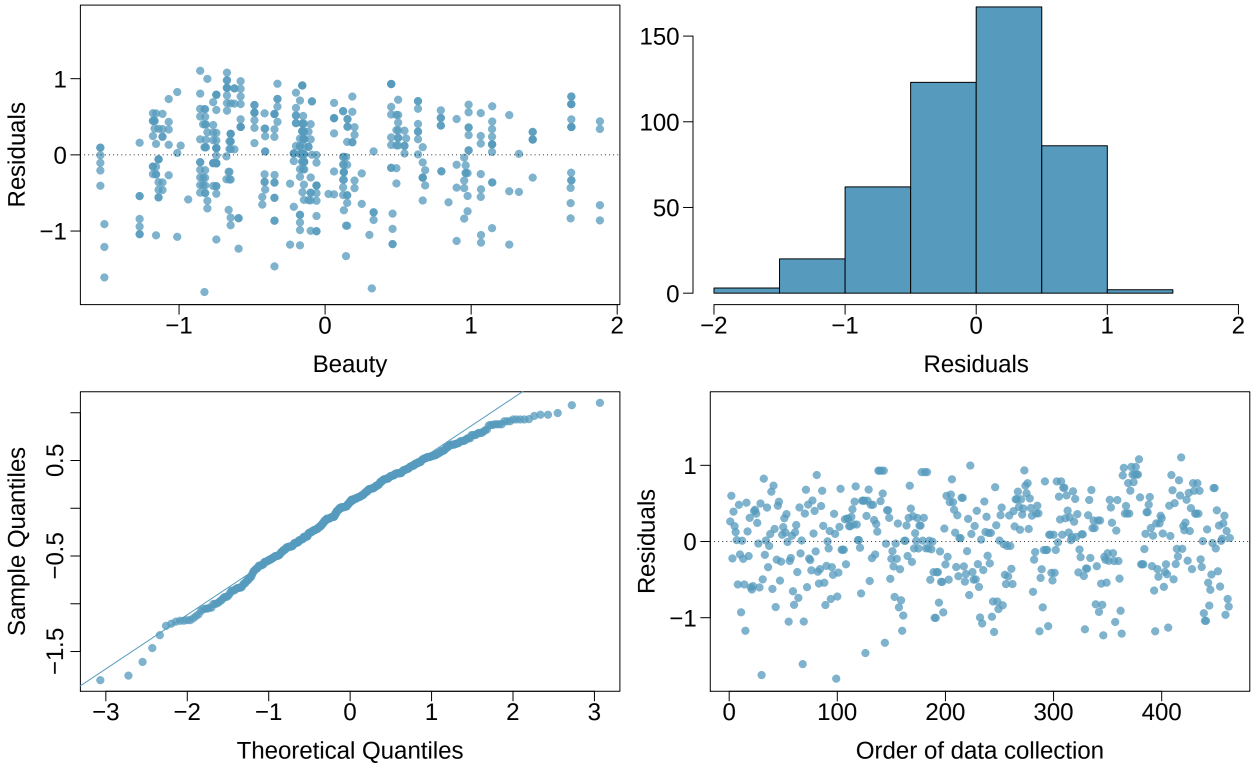
```
## [1] 0.132986
```

(b) Do these data provide convincing evidence that the slope of the relationship between teaching evaluation and beauty is positive? Explain your reasoning.

The slope is positive, but not indicative of a strong relationship.

(c) List the conditions required for linear regression and check if each one is satisfied for this model based on the following diagnostic plots.

Linearity is not clear however, the residuals appear normally distributed and there is variability.


Residuals
Beauty
Residuals
Sample Quantiles
Theoretical Quantiles
Residuals
Order of data collection