Chapter7HW: OpenIntro Statistics

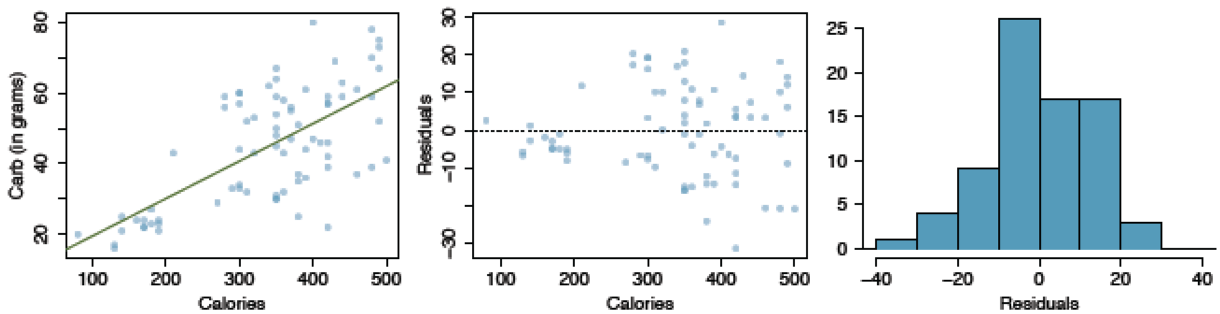
Walt Wells, Fall 2016

Practice: 7.23, 7.25, 7.29, 7.39

Graded

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) that Starbucks food menu items contain.

There is seems to be a positive correlation with a mild upward trend between calories and carbohydrates.

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

Explanatory = Calories; Response = Carbs (in grams)

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

We might want to predict how many grams of carbs some food contains, given a certain calorie count.

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

While the data fit a linear trend, and the residuals appear nearly normal, we may not meet the criteria for constant variablity. We can see that as calories increase so to does the variability of carbs.

7.26 Body measurements, Part III.

Exercise 7.15 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.

```
b1 <- round(0.67 * (9.41/10.35), 4)
b0 <- round(b1 * -107.20 + 171.14, 4)
```

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

$$\hat{y} = 105.8445 + 0.6091 * shouldergirth$$

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

For each additional cm of shoulder girth, we would expect an additional 0.6091 of height. If the individual had 0 shoulder girth, we would expect a height of 0.6091.

(c) Calculate R2 of the regression line for predicting height from shoulder girth, and interpret it in the context of the application.

The R2 is the square of our given correlation, or 0.4489. In this context, a reduction of about 65% of the data's variability can be expected using a linear model with shoulder girth.

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

```
b0 + b1 * 100
```

```
## [1] 166.7545
```

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

```
160 - 166.7545
```

```
## [1] -6.7545
```

The residual implies that our model overestimated the prediction by the amount above.

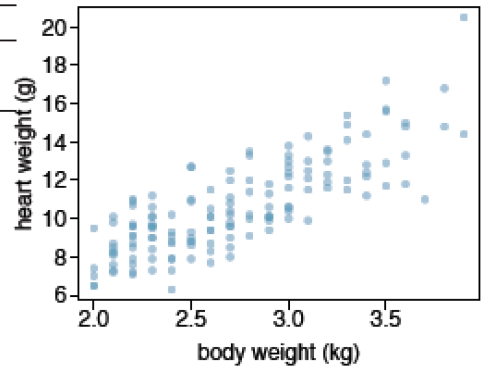
(f) A one year old has a shoulder girth of 56 cm. Would it be appropriate to use this linear model to predict the height of this child?

Because this value is an outlier and isn't considered in the model creation ($>$ 4 SD away from mean), it would be inappropriate to predict the height using our model.

7.30 Cats, Part I.

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.

$$\hat{y} = -0.357 + 4.034 * bodyweight$$

(b) Interpret the intercept.

If a cat's body weight were 0 kg, we would expect their heart to weight -0.357 grams.

(c) Interpret the slope.

For each additional kg of body weight, we expect a cat's heart to weigh and additional 4.034 grams.

(d) Interpret R2.

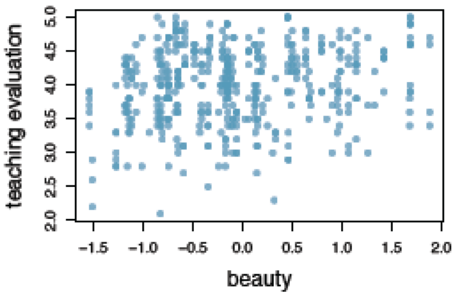
A reduction of about 64% of the data's variability can be expected using a linear model prediciting heart weight using body weight.

(e) Calculate the correlation coefficient.

The correlation coefficient is calculated as 0.8041144.

7.40 Rate my professor.

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 effetiveness 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.

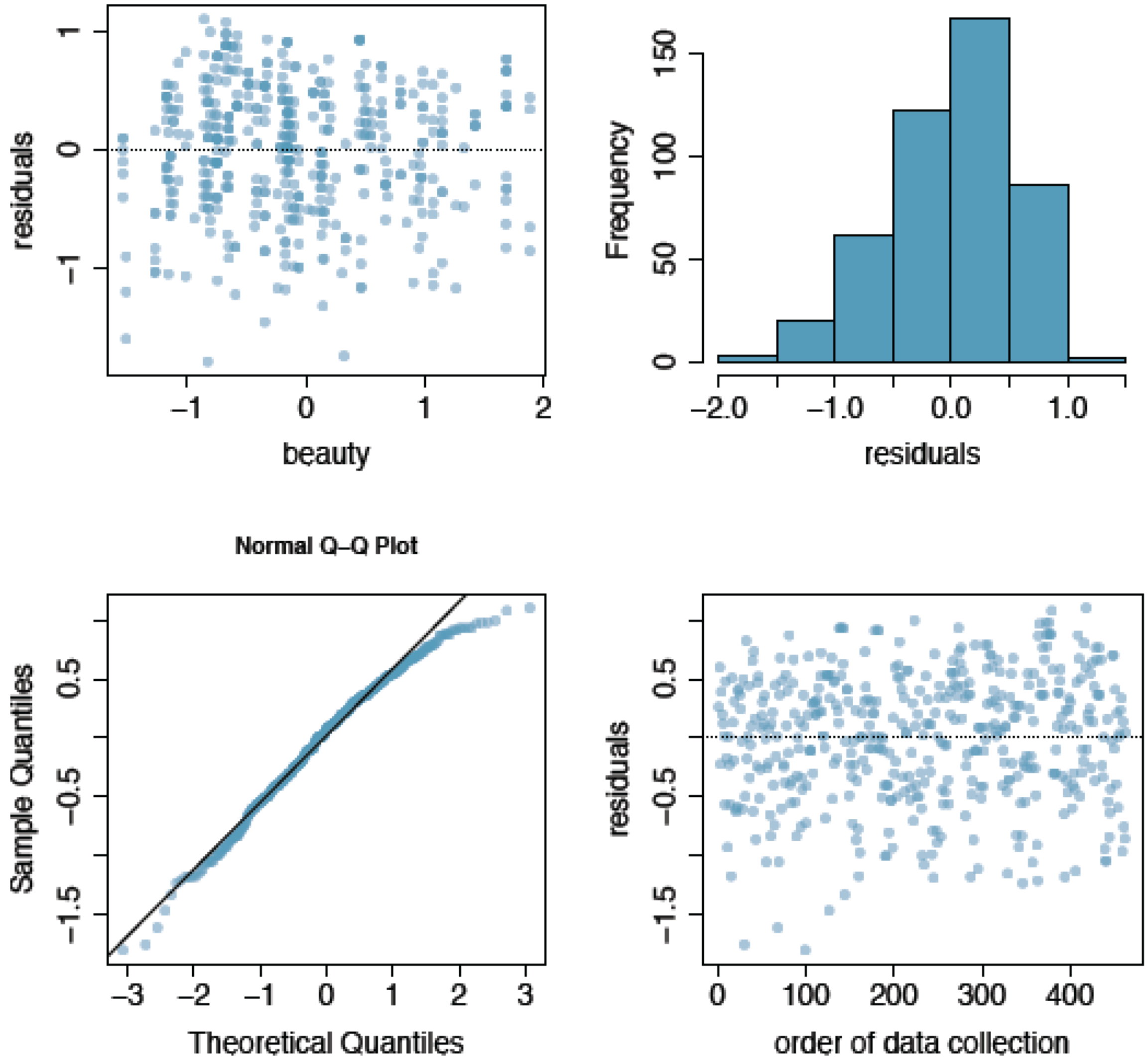
```
b0 <- 4.010
b1 <- 4.13 * .0322
```

The slope is 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.

There is convincing evidence that the slope is positive. The Pvalue of the output for a onesided test is small, suggesting we should reject a H_O and conclude that there is convincing evidence that as a teacher's beauty score increases so to does the likelihood of a better teaching evaluation score. We assume the regression output supports this interpretation of H_O and H_A.

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



It seems that all the conditions are met. The residuals are nearly normal and show relatively constant variability. We believe these are independant observation. And the data shows a minor linear trend.