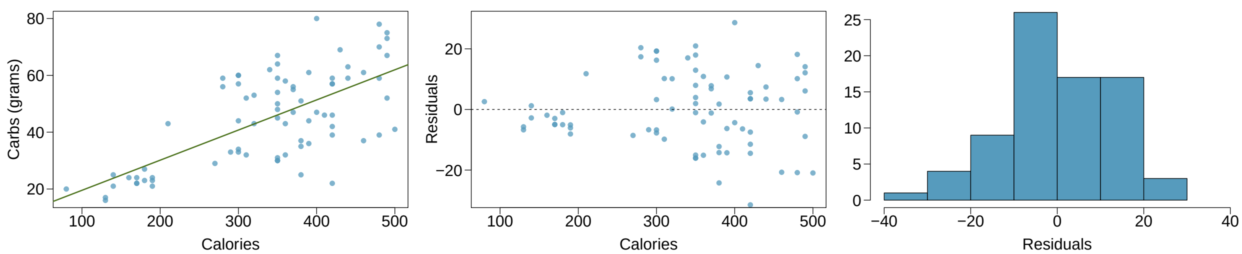
Homework 8

Irene Jacob

2020-10-31

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.

The relationship between number of calories and amount of carbohydrates (in grams) that Starbucks food menu items contain is linear.

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

Explanatory variable: Calories Response variable: carbohydrates.

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

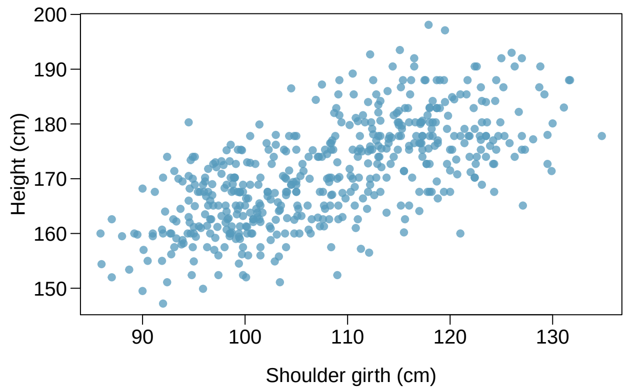
Regression line would be the best method to understand the carbohydrate consumption based on the calories.

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

These data do not meet the conditions required for fitting a least squares line.

Body measurements, Part I.

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.

Height increases as soulder girth increases.

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

The relationship will remain the same as the unit change does not matter.

Body measurements, Part III.

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.

```r
x_m = 107.2
x_sd = 10.37
y_m = 171.14
y_sd = 9.41
c = 0.67

s = c*(y_sd/x_sd)
s
```

```
## [1] 0.6079749
```

```r
i = y_m - s*x_m
i
```

```
## [1] 105.9651
```

Height = 105.9651+(0.6079749*Shoulder Girth)

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

slope: 0.6079749 intercept: 105.9651

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

```r
q = lm(hgt ~ bdims$sho_gi, data = bdims)
summary(q)$r.squared
```

```
## [1] 0.4432035
```

Regression line explains 44.3% of the observed variation.

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

```r
s_g_100 <- i + s * 100
s_g_100
```

```
## [1] 166.7626
```

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

```r
160 - s_g_100
```

```
## [1] -6.762581
```

This means that the model overestimated the height of the student.

(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?

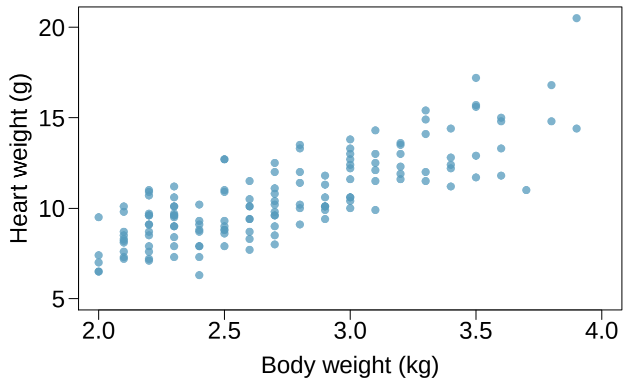
No.

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

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



(a) Write out the linear model.

heart weight= -0.357 + 4.034 * body weight

(b) Interpret the intercept.

A cat with body weight = 0kg has an expected Heart Weight of -0.357g. A cat with 0kg is not possible.

(c) Interpret the slope.

For each 1kg increase in body weight there is a 4.034 increase in heart weight.

(d) Interpret R^2.

$R^2 = 64.66\%$.

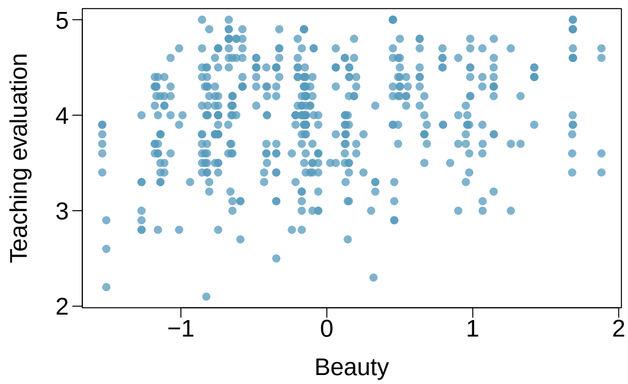
(e) Calculate the correlation coefficient.

sqrt(0.6466)= 0.8041144

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 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



(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.

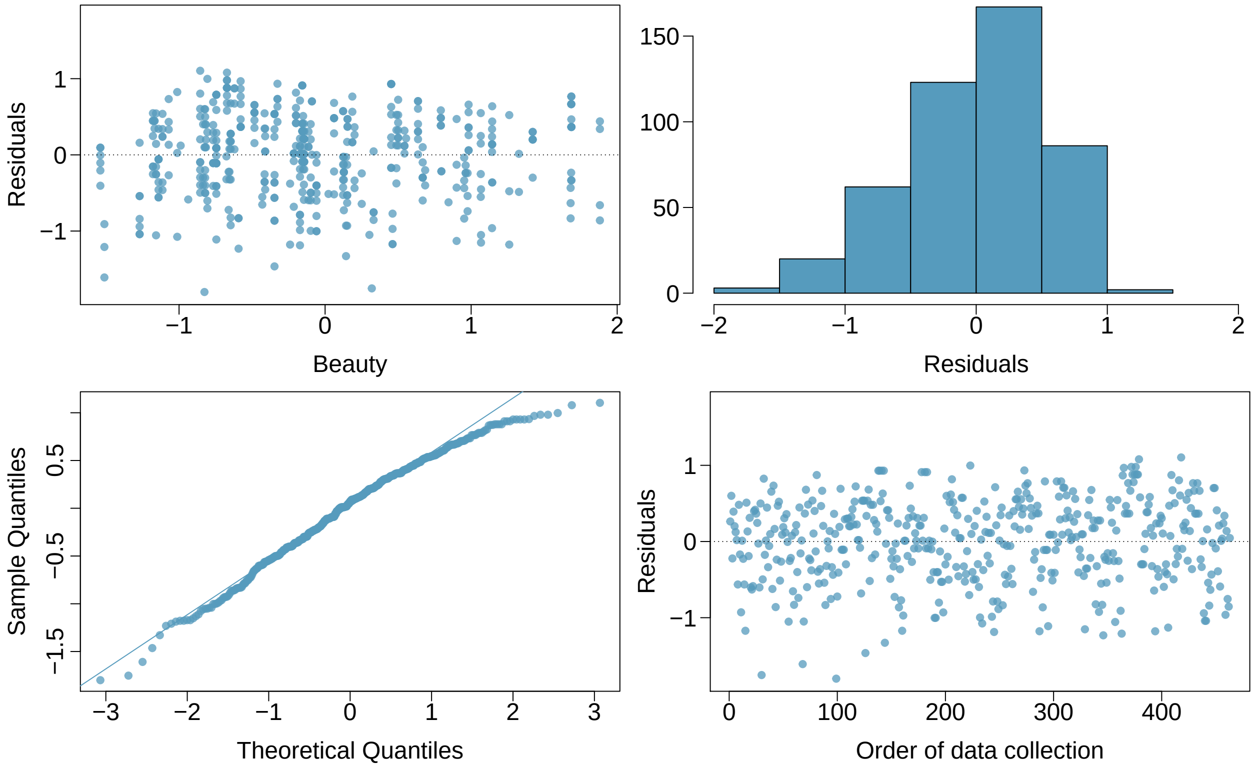
```
x_m = -0.0883
y_m = 3.9983
i = 4.010

s = (y_m - i) / x_m
s
```

```
## [1] 0.1325028
```

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

Yes because the slope is positive.

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



The conditions of linear regression are met. The observation is independent of each other and random. Histogram is fairly normal. The data is linear.