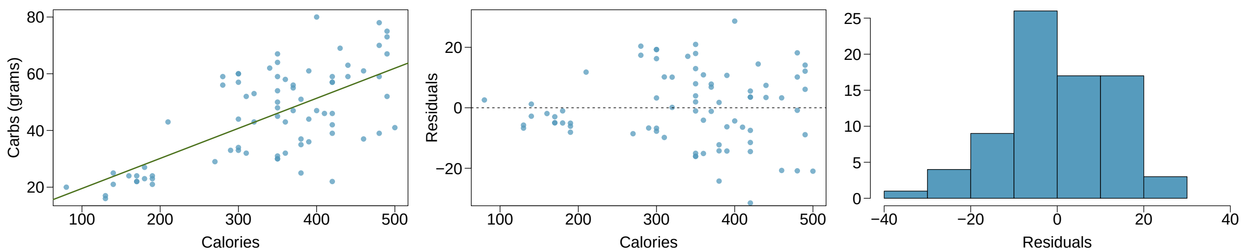
Chapter 8 - Introduction to Linear Regression

Joshua Registe

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.

There is a positive relationship between carbs and calaries in the starbucks food menu items. As calories increase, it seems carbs also increase.

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

The explanatory variable in this scenario are the calories, and the response variable are the carbohydrates.

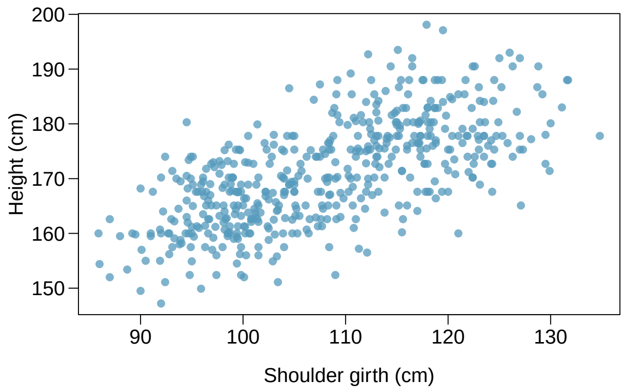
(c) Why might we want to fit a regression line to these data?

We might want to fit a regression line to these data in order to creade a model that helps us predict the carb content of a menu item based on the calories

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

Yes these meet the requirements for least square line, points are spread around residuals with mild biases and the distribution is sufficiently normal

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.19 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.

the relationship between girth and height is positive. as shoulder girth increases, there is an increase in height according to the scatter. the summary in question 3a below shows the distribution ranges and our estimated regression coefficients for this relationship. the positive slope confirms direction of relationship.

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

If girth was measured in inches, this should not change the relationship between girth and height.

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.

```r
M1<-lm(hgt~sho.gi, data = bdims)
summary(M1)
```

```
##
## Call:
## lm(formula = hgt ~ sho.gi, data = bdims)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.2297  -4.7976  -0.1142   4.7885  21.0979
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 105.83246    3.27245   32.34   <2e-16 ***
## sho.gi        0.60364    0.03011   20.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.026 on 505 degrees of freedom
## Multiple R-squared:  0.4432, Adjusted R-squared:  0.4421
## F-statistic:   402 on 1 and 505 DF,  p-value: < 2.2e-16
```

based on above, the equation can be written as Height_pred = 105.83 +0.603*Girth

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

The intercept implies that at a shoulder girth near 0, we can expect the height of the person to be around 105.83 cm. the slope of 0.6 shows the direction and strenth of direction implying that this for every 1 unit increase in girth, there is a 0.6 increase in height.

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

```r
(rsquared<- summary(M1)$r.squared)
```

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

(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
predict(M1, newdata = data.frame(sho.gi = c(100)))
```

```
##        1
## 166.1969
```

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

```
(res = abs(160-166))
```

```
## [1] 6
```

this residual represents how far our model prediction is from the actual observation in cm

(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?

```
predict(M1, newdata = data.frame(sho.gi = c(56)))
```

```
##        1
## 139.6365
```

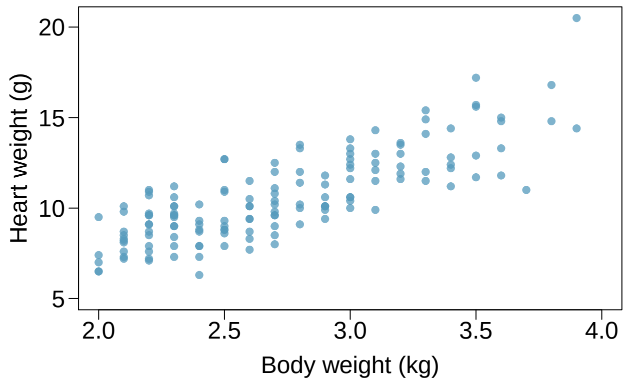
it would not be appropriate because 1 year olds are outliers in predicting height and would not measure 139 cm

---

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

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



(a) Write out the linear model.

the linear model based on the model table can be written as:

heartweight_pred = 4.034*bodywt - 0.357

(b) Interpret the intercept.

the intercept at -0.357 implies that at a body weight of 0, the heart weight is 0, theres some uncertainty around the incercept with a PR>t of 0.6

(c) Interpret the slope.

The slope implies that for every unit increase in body weight (kg) there is a 1/4th increase in heart weight (g)

(d) Interpret R^2.

the R2 shows a 64.66% linear correlation coefficient between the body weight and heart weight and this is going in a positive direction

(e) Calculate the correlation coefficient.

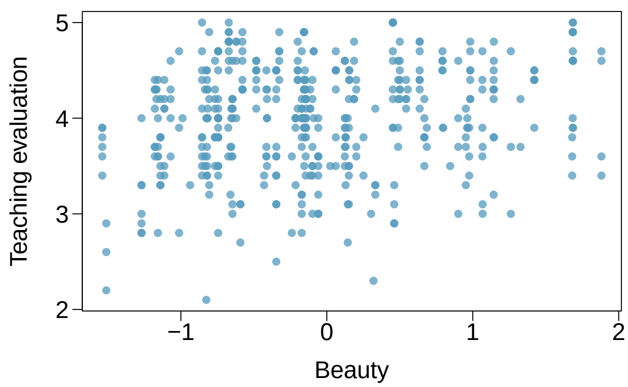
the correlation coefficient can be measured by the square root of our R2 as

```
sqrt(.6466)
```

```
## [1] 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



(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.

```
summary(m_eval_beauty)
```

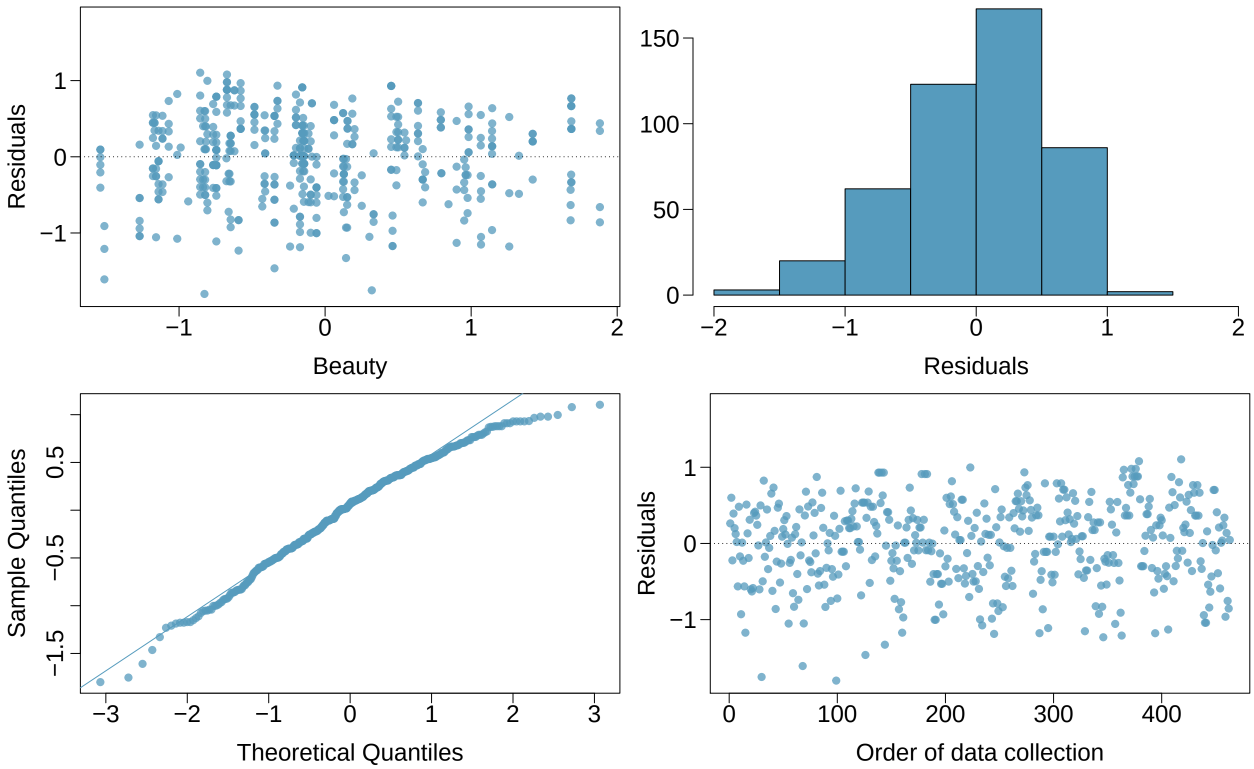
```
##
## Call:
## lm(formula = eval ~ beauty)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.80015 -0.36304  0.07254  0.40207  1.10373
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.01002    0.02551 157.205  < 2e-16 ***
## beauty       0.13300    0.03218   4.133 4.25e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5455 on 461 degrees of freedom
## Multiple R-squared:  0.03574,    Adjusted R-squared:  0.03364
## F-statistic: 17.08 on 1 and 461 DF,  p-value: 4.247e-05
```

slope is 0.133

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

yes with a t of <0.05 there is sufficient evident however this relationship is very small

(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 - observation barely shows a linear trend.

nearly normal - Observation does show normal distributions

constant variability - observation does show constant variability per the residuals and beauty plot

independent observations - The values are independent