STAT22000 Summer 2020 Homework 14

All page, section, and exercise numbers below refer to the course text (*OpenIntro Statistics*, 3rd edition, by Diez, Barr, and Cetinkaya-Rundel.).

Reading: Section 7.2-7.4
Problems for Self-Study :

1. Exercise 7.7, 7.19, 7.21, 7.25, 7.27, 7.31, 7.37, 7.41 on p.358-371 where the answers can be found at the end of the book.

2. A number is missing in each of the data sets below. If possible, fill in the blank to make the correlation r equal to 1. If this is not possible, explain why not. *Hint: Make a scatterplot. Under what circumstance will the correlation equal to 1?*

(a)

x	y
1	0
2	2
2	2
4	–

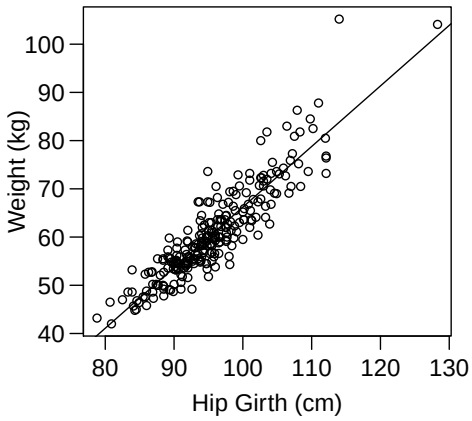
(b)

x	y
1	0
2	2
3	5
4	–

3. The scatterplot on the right shows the weights (in kg) and hip girths (in cm) of 249 physically active women age 18-45. Here is a summary of the data:

	Body weight (kg)	Hip girth (cm)
Mean	60.694	95.603
SD	9.639	6.945

Correlation $r \approx 0.905$



(a) How would the correlation r change if weight was measured in pounds while the units for hip girth remained in centimeters? (1 pound = 0.454 kg).

(b) Write down the equation of the regression line for predicting a woman's weight in kilograms from her hip girth in centimeters.

(c) Interpret the slope and the intercept of the equation in the previous part in this context.

(d) Calculate R^2 of the regression line for predicting weight from hip girth, and interpret it in the context of the application.

(e) A randomly selected female student from your class has a hip girth of 90 cm. Predict the weight of this student using the regression line.

(f) The student in the previous part weighs 55 kg. Calculate the residual, and explain what this residual means.

(g) A one-year-old baby has a hip girth of 52 cm. Would it be appropriate to use the regression line in part (b) to predict the weight of this baby?

(h) Can we use the regression line in part (b) to predict the weight of an adult man with a hip girth of 110 cm? Explain your answer.

(i) Can we use the regression line in part (b) to predict the hip girth of a 35-year old woman who weighs 80 kg? Explain your answer.

(j) Find the equation of the regression line for predicting a woman's hip girth from her weight, and use the equation to predict the hip girth of a 35-year old woman weighs 80 kg.

4. A biologist was interested in the relationship between the velocity at which a beluga whale swims and the tail-beat frequency of the whale. A sample of 19 whales was studied and measurements were made on swimming velocity, measured in units of body lengths of the whale per second and tail-beat frequency, measured in units of hertz (number of beats per second). The data file `BelugaSwim.txt` is posted on Canvas with this exercise.

   (a) Make a scatterplot with tail beat frequency (in Hertz) on x-axis and the swimming speed on y-axis. Label the plot properly. Describe the relationship between the two variables.

   ```
   whale = read.table("BelugaSwim.txt", h=T)
   plot(whale$freq, whale$speed,
        xlab = "Tail Beat Frequency (Hertz)",
        ylab = "Swimming Speed \n(Body Length per Second)")
   ```

   Review Section 3 in Lab #1 `http://www.stat.uchicago.edu/~yibi/s220/labs/lab01.html` *about changing the working directory if you have trouble loading the data file to R.*

   (b) Find the means and the standard deviations of the two variables and their correlation coefficient ($\overline{x}$, $\overline{y}$, s_x, s_y, and r) in R or by a calculator.

   ```
   library(mosaic)
   favstats(~freq, data=whale)
   favstats(~speed, data=whale)
   with(whale, cor(freq, speed))
   ```

   (c) Here we fit a simple linear regression model in R using the `lm()` function, in which `lm` stands for "linear model".

   ```
   lmwhale = lm(speed ~ freq, data=whale)
   ```

   The general syntax to fit a model with the response variable `y` and explanatory variable `x` is `lm(y~x, data=nameofdataset)`. We can save the fitted model by giving it a name. You can name it whatever you like, such as `lmwhale`. We can call a saved model whenever we need it. For example, to get the intercept and slope of the fitted regression line we can type `lmwhale$coef` and then get the following output.

   ```
   > lmwhale$coef
   (Intercept)        freq
   -0.01561813  0.59237262
   ```

   The equation of the regression line is then

   $$\text{predicted speed} = -0.01561813 + 0.59237262 \times (\text{tail beat frequency in hertz})$$

   Verify that the slope and the intercept given by R are $r \cdot s_y/s_x$ and $\overline{y} - (\text{slope}) \cdot \overline{x}$ respectively. Show your computation.

   (d) Add the regression line to the scatter plot using the R command below

   ```
   plot(whale$freq, whale$speed,
        xlab = "Tail Beat Frequency (Hertz)",
        ylab = "Swimming Speed \n(Body Length per Second)")
   abline(lmwhale)
   ```

(e) The R command `summary(lmwhale)` gives a more detailed output for the model.

```
> summary(lmwhale)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.01562    0.07075  -0.221    0.828
freq         0.59237    0.05973   9.917 1.75e-08 ***
```

Test the null hypothesis that the slope of the regression line is 0.8 against a 2-sided alternative. Report the test statistic with degrees of freedom, and the P-value.

(f) Calculate a 95% confidence interval for the slope of the regression line for predicting the swimming speed of a beluga whale (in the number of body lengths of the whale per second) from its tail beat frequency (in hertz), and interpret the interval in context of the data.