SC212 HW-2

Matt Bass

Section 2.1:

2.8 (554,665 / 2,220,087) = 0.24983930809

2.12:

a. (70/100) = 0.7
b. (50/100) = 0.5
c. (10/50) = 0.2
d. (30/70) = 0.42857142857

2.20:

a. sample
b. (20,293/100,952) = 0.20101632459
c. (23,405/100,952) = 0.23184285601
d. I would say, at least in terms of concussions it is not conclusive that riding a bicycle is more dangerous to children in the US than playing football for multiple compounding reasons. To start out with even though I am not sure the exact year this data was taken but from looking at child-stats.gov/americaschildren/tables/pop1.asp The US population of children 5-18 averages around 49 million children for the past 40 years. So I can say that over the year there might be a difference is the amount of children getting concussions from riding a bike as compared to playing football. This is especially true since the difference in number between the two groups is (23405 −20293) = 3112 or (3112 /100952) = 0.0.03082653142 So I would say there is not a significant difference between the groups, this is also because this is an observational study and we can not control who does what activities and there might be some external factors such as underlying characteristics of kids who bike vs play football

2.34:

a. Both male and female would most likely say 0 drinks for personal use
b. Both male and female most likely response for typical use is 5-6
c. No their perceptions of typical use don't match reality

2.50: certificate

Section 2.2:

2.54 A and H

2.58: For A the mean is larger than the median,
For B the mean is approximately equal to the mean
For C the mean is approximately equal to the mean
For D the mean is approximately equal to the mean

2.68: The mean is 124.8
The median is 123.5
There are no outliers

2.80:

a. Skewed Right

b. The mean is 1 and the median is 5.3 because I said that this data would be skewed to the right meaning there are some outliers with much higher values than the rest of the data.

2.90: Relatively Symmetric

2.93:

a. Women who have been married appear to have the greatest mean which makes sense because a married women has a husband who can help support more children, and unmarried women seem to have the mean most different from the median because the data is heavily skewed

b. m = 2 for married women and m = 0 for never married women

Section 2.3:

2.94:

a. Mean is 17.36364 , STD is 5.731888

b.

Min.	1st Qu.	Median	Mean.	3rd Qu.	Max.
10.00	13.50	17.00	17.36	20.50	28.00

2.101:

a. V

b. III

c. IV

d. I

e. VI

f. II

2.104: Mean = 68. STD = 3

2.116: Z score = 1.72

2.120: 980 to 1020

2.140:

a. 1,1,1,9,9,9

b. 5,5,5,5,5,5

2.150:

a. Mean = 10.5238 STD = 10.32501 **(**tg**)**

b. zscore = 6.728924

Section 2.4:

2.154:

a. W
b. X
c. Y
d. Z

2.156:

a. Skewed Right
b. None
c. 60

2.162a: Outlier range 10 – 1.5(6) =7 and 16+3=19 and all the numbers in both tails are outliers

2.176:

a. Yes there appears to be an association because the high-quality hives get many more circuits on general than the lower quality sites

b. High Quality outliers range: lower = 7.5 – (1.5)(122.5-7.5) = -165
upper = 122.5 + (1.5)(122.5-7.5) = 295

Low Quality outliers range: lower = 0 – (1.5)(42.5) = -165
upper = 42.5 + (1.5)(42.5) = 295

From the IQR I can see that both high and low quality groups have no lower outliers (which makes sense because it is impossible for a bee to do a negative amount of circuits) However both groups both have a few higher valued outliers.

c. 60.5

d. Highest qual z score = (440 – 90.5)/94.6 = 3.69450317125
Lowest qual z score = (185 – 30) / 49.5 = 3.13131313131

The highest quality group has the largest relative outlier

e. No it is not

2.184: To start out with the categorical variable I chose was sex and the quantitative variable I chose was weight. To start out with there were some NA entries for weight (which makes sense as some people might not feel comfortable responding) so the first thing I did was drop all rows in the data set with NA in the height column. Just from looking at the boxplots of heights between males and females it was as I expected that males on average have a greater height than females however it was a bit surprising to see so many high outliers in the female group that were greater in height than any sample in the male group. The female group overall had much more deviation / variance in its data which we will also see form the 6 number summaries below:

Female Heights:	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Sd
	59.00	64.00	65.00	65.69	67.00	83.00	3.272213

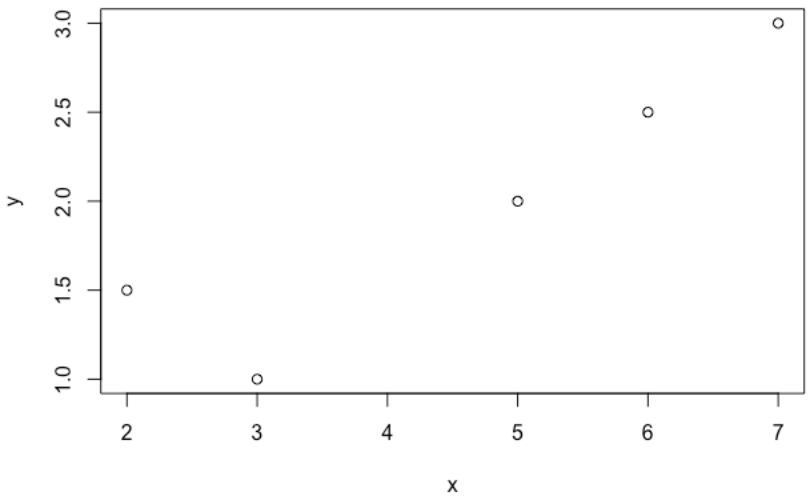
male Heights:	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Sd
	61.00	69.00	71.00	70.85	73.00	79.00	3.071803

Section 2.5:

2.188: d

2.190: c

2.200:



2.202: 0.9149914

2.220:

a. Positive and means they are associative**,** it appears that when the ecological footprint of a country goes up so does the countries happiness (but im guessing this is due to underlying confounding variables in particularly that wealthy countries tend to have a greater ecological impact and also are “happier” which can also be drawn from looking at the country region groups and their locations

since western countries are in the top right of the graph. There is not a strong positive correlation thought I would say it is weakly positive correlated

b. The happiness and ecological footprint of a country in the bottom left corner of the graph are both low. However we would considered a low ecological footprint to be a good while low happiness is considered bad

c. 2.5

d. For ecological footprints 0-6 there appears to be a strong associate with ecological footprint and happiness however for footprints 6-10 there does not appear to be much of an association so that shows us that an increasing ecological footprint only has a positive effect in increasing happiness up to the point of around an ecological footprint of 6. This means that there are diminishing returns to increasing an ecological footprint to a certain extent so many of the countries with a higher ecological footprint than that should try to reduce their footprint to that point of 6 without having to worry about a reduction in happiness

e. It processes a low ecological footprint and high happiness

f. In region 4 (Sub-Saharan Africa countries we can see that all of them fall within the range of a 0-4 ecological footprint as well as a 2-6 range of happiness. In this region we see that an increasing ecological footprint has a large increase on a country in that groups happiness which makes sense because many of these countries are some of the least developed in the world so and infrastructure and other pollution related activities tend to have a much has a much substantial effect on all the variables such as wealth and health that contribute to a countries happiness (especially when they don't have much as stated)

g. For countries in the bottom left efforts should be focused on increasing their happiness (which will probably take a little hit in the ecological footprint) and for countries in the top right their efforts should be focused towards reducing their ecological footprint.

Unit A Synthesis:

A.18:

a. The cases are the tagged penguins. . The variables are the type of tag (categorical), number of chicks (quantitative), survival or not (categorical), and length of time on foraging trips (quantitative).

b. We can strongly infer that it is

c.

1. histogram and calculate an average number of chicks for both groups (metal or electronic) this can also info can be gained from distribution of the histogram also 2 box plots could be uses

2. Bar Graph and survival rate of the 2 groups

3. histogram and calculate an average foraging for both groups of types of band this can also info can be gained from distribution of the histogram also 2 box plots could be uses

4. Scatter plot and I would calculate the correlation of the 2 vars

5. histogram and calculate an average foraging for both groups (survived and not survived) this can also info can be gained from distribution of the histogram also 2 box plots could be uses