

Math 58 / 58B - Introduction to (Bio)Statistics

your name here

due Friday, April 10, 2020

Homework 8

How to include a screen shot (as .png) into a markdown file:

```
<p>
![caption] (MyScreenShotThatLivesInTheSameDirectoryAsThisFile.png)
</p>
```

Note: look at the pdf to see all the images. Then remove those lines of code from your own assignment, because you don't have the images on your computer, so it won't compile for you unless you remove those lines.

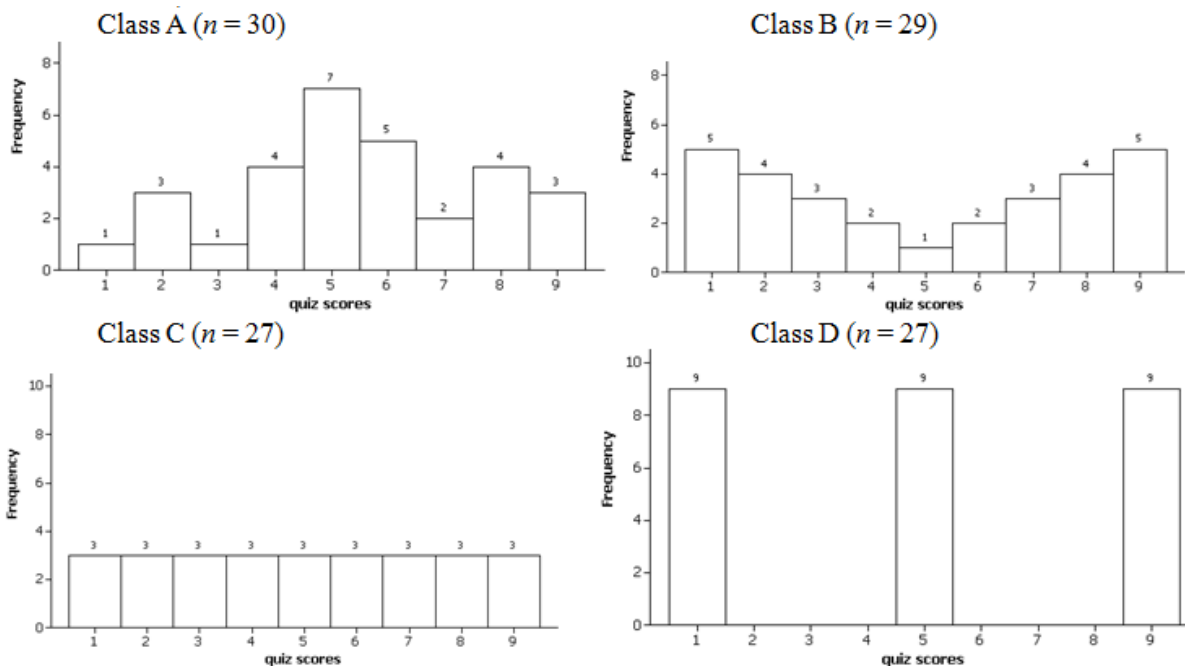
Assignment Summary (Goals)

understanding of and working with:

- quantitative data
- standard deviation
- sampling distribution of the mean
- t-scores

1. Hypothetical Quiz Scores¹

Consider the four histograms of (hypothetical) quiz scores in four classes (scores are integers ranging from 1 to 9):

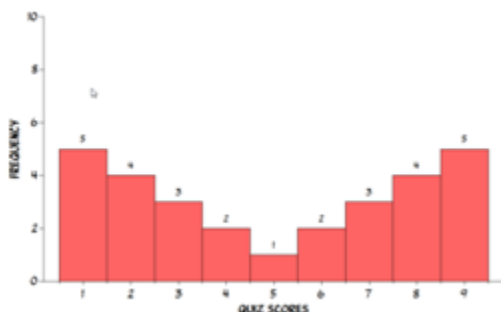


¹From ISCAM, HW 0.2

- (a) Between class A and class B, which has more variability in quiz scores? [Hint: Think which will have the larger standard deviation. Explain your reasoning.]
- (b) Between classes C and D, which has more variability in quiz scores? Explain your reasoning.

2. Properties of Center and Spread²

The histogram displays the (hypothetical) quiz scores for a class of $n = 29$ students.



Suppose we were to give every student 5 bonus points.

- (a) How would the mean change? The median?
- (b) How would the standard deviation change?

Note: You should explain your answers to (a) and (b) *without* carrying out the calculations to find these new values.

3. Practice Problem³

- (a) Use the Central Limit Theorem to estimate the probability that the sample mean of **20 randomly selected** passengers exceeds 159.57lbs, assuming a normal population with mean 167lbs (μ) and standard deviation 35lbs (σ).
- (b) Is the probability you found in (a) larger or smaller than the probability you found (that is, that the investigation found) for 47 passengers? Explain why your answer makes intuitive sense.
- (c) Repeat (a) assuming a uniformly distributed population of weights. (Do not use normal probabilities, instead, use population 3 and the applet. You can include a screen shot or just describe in words.) How do the two probabilities from (a) and (c) compare? [Hint: Would it have been appropriate to use the CLT to answer this question?]

copy and paste pop 3 into the applet:

<http://www.rossmanchance.com/iscaam2/data/WeightPop3.txt>

- (d) Explain why the calculation in (a) does not estimate the probability of the Ethan Allen sinking with 20 passengers (that is, does not estimate the probability of the Ethan Allen going over its weight limit). [Hint: this problem is *not* about the shape of the population distribution... think about the number of passengers.]

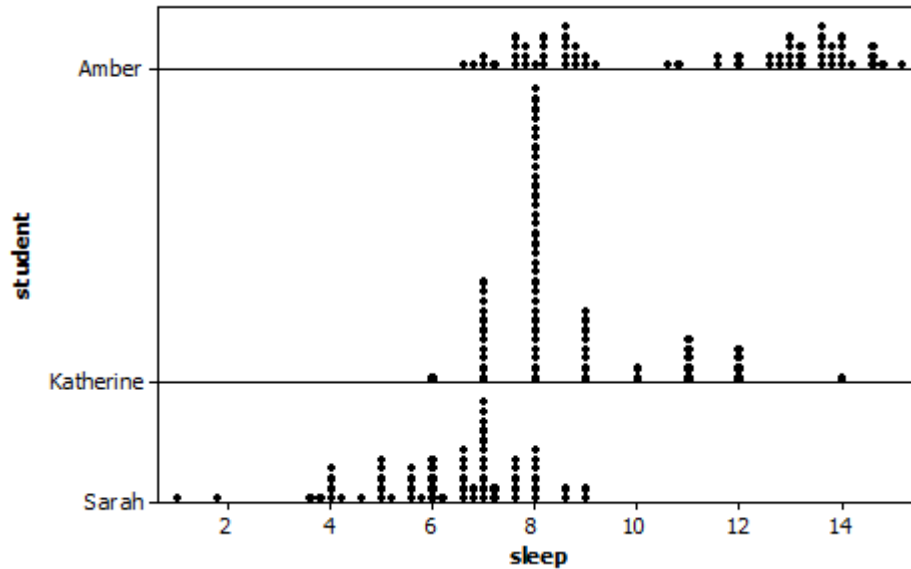
4. Student Sleep Times⁴

The dotplots display the distribution of sleeping times (per day, in hours) of three college students (Amber, Katherine, Sarah) for a nine-week period in the fall of 2004.

²From ISCAM, HW 2.17

³From ISCAM, PP 2.4A

⁴From ISCAM, HW 0.3



- One of these students developed mononucleosis during the term and so was told to get as much rest as possible for several weeks. Which student do you think this is? Explain your reasoning.
- One of these students is the mother of two small children. Which student do you think this is? Explain your reasoning.
- Which student recorded her sleeping times only to the nearest hour? Explain.
- Which student generally got the most sleep? Which generally got the least?

5. Sleeping Students (cont)⁵

Reconsider the students' sleeping times from the Chapter 0 Exercises (SleepStudents.txt).

- Determine the five-number summary of sleeping times for each student (min, 25%, median, 75%, max).
- Create boxplots of these students' sleeping times on the same scale. Comment on what these boxplots reveal.
- What does the dotplot reveal about Amber's sleeping times that the boxplot does not?

Some code you might find useful, to run, make sure to change eval to TRUE:

```
studentSleep = read.table("http://www.rossmanchance.com/iscam2/data/SleepStudents.txt",
                           sep="\t", header=TRUE, na.strings="*")
```

```
studentSleep %>%
  select(sleep.stacked., student) %>%
  group_by(student) %>%
  summarize(meanslp = mean(sleep.stacked.), medslp = median(sleep.stacked.),
            sdslep = sd(sleep.stacked.), iqrslep = IQR(sleep.stacked.),
            slp25 = quantile(sleep.stacked., 0.25))
```

```
## # A tibble: 3 x 6
##   student  meanslp medslp sdslep iqrslep slp25
##   <fct>    <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Amber     10.9   11.5  2.77  5.25  8.25
```

⁵From ISCAM, HW 2.9

```
## 2 Katherine      8.57      8      1.58      1      8
## 3 Sarah          6.25      6.5     1.61      1.88    5.38
```

```
studentSleep %>%
  filter(student == "Amber") %>%
  filter(sleep.stacked. < 10) %>%
  summarize(badsleep = n())
```

```
##   badsleep
## 1         28
```

5. Sleeping Students (cont)⁶

Reconsider the students' sleeping times from Exercise 9 (SleepStudents.txt).

- Calculate the mean and standard deviation of sleeping times for each student.
- For each student, determine the proportion of the 63 sleeping times that fall within one standard deviation of the mean.

see the examples just below and also the lines of code above for question 4 (c).

```
group_by the student
then summarize to calculate the means and standard deviations

then filter for each student one at a time.
then filter by nights of sleep that are inside upper or below the lower
filter(value > low# & value < up#) %>% ...
```

- For which student does the empirical rule (see part(b)) appear to hold most closely? For that student, determine the proportion of sleeping times that fall within two standard deviations of the mean.
- Suppose that Katherine gets 10 hours of sleep in a particular night. How many hours more than her mean is this? Also calculate the t-score for this value. [NOTE: the smiley face in this problem is a single night of sleep, i.e., the number 10. That is, *one* night of sleep above some predetermined value (her average). As always, the denominator should measure the variability associated with the smiley face ... so here, the denominator is measuring the variability of a single observation.]
- Suppose that Amber gets 13 hours of sleep in a particular night. How many hours more than her mean is this? Also calculate the t-score for this value.
- Which of these (10 hours for Katherine or 13 for Amber) is higher above that student's mean? Which has the higher t-score? Explain why your answers are not the same.

⁶From ISCAM, HW 2.10