

Homework 4: Sections 3.1 - 3.4

KEY

STA209-04: Applied Statistics

February 15, 2019

From the Book:

Due Date: February 22, 2019

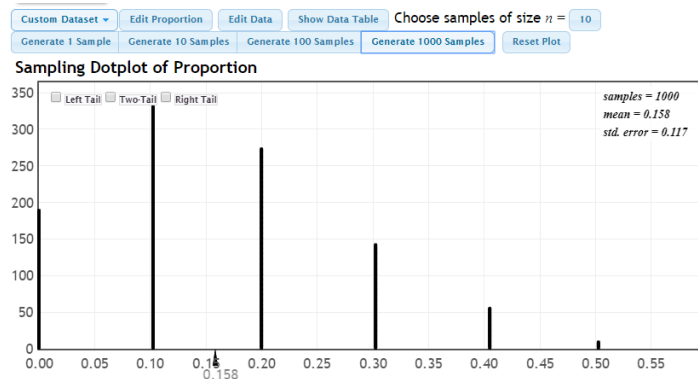
Questions: 3.31, 3.39, 3.63, 3.67, 3.74, 3.91, 3.133

3.31) [4 pts]

- (a) [1 pts] Various answers possible here. You should see a list of 10 datapoints (which were sampled randomly) and a sample mean computed from these data.
- (b) [1 pts] Same as above, but the random sample should contain 20 observations.
- (c) [1 pts] The population mean enrollment is 53.54. Students should comment on how close their sample means from (a) and (b) are to this population mean.
- (d) [1 pts] Sketches of the sampling distribution will vary, but each should be roughly symmetric and bell-curved in shape. Each sketch should also be centered at the population mean of 53.54.

3.39) [4 pts]

- (a) [1 pts] Student should use StatKey to take random samples of size 10 to form a sampling distribution for a proportion. Below is what I obtained:



From this sampling distribution, we see that the standard error for the sample proportions is 0.117, the sample proportion furthest from 0.155 is 0.50 and it is $0.5 - 0.155 = 0.345$ units away.

- (b) [1 pts] A plot similar to above should be generated with $n = 20$. The standard error should be provided (and should be smaller), the datapoint furthest from 0.155 should be mentioned (should be smaller than 0.5), and its distance from 0.155 should be mentioned (should be smaller than before).
- (c) [1 pts] A plot similar to above should be generated with $n = 50$. The standard error should be provided (and should be even smaller), the datapoint furthest from 0.155 should be mentioned (should be smaller than previous), and its distance from 0.155 should be mentioned (should be smaller than before).

- (d) [1 pts] As the sample size increases, the accuracy of our sample proportion in estimating the population proportion also increases.

3.63) [2 pts] The correct interpretation of this interval is that we are 95% confident that at least 83% of US adults and at most 89% of US adults consider a car a necessity. In other words, we are 95% confident that the interval contains the true proportion of US adults who believe a car is a necessity.

3.67) [4 pts]

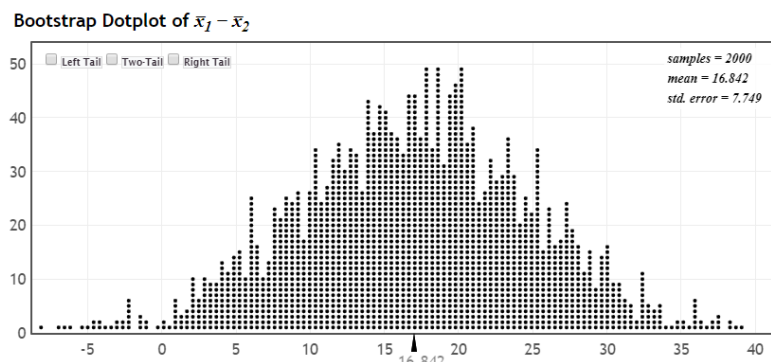
- (a) [1 pts] We are 95% confident that the mean difference in response time between gamers and non-gamers is at least -1.8 seconds and at most -1.2 seconds.
- (b) [1 pts] Given this confidence interval, it is not plausible that gamers and non-gamers are the same in response time. Since the interval is entirely negative and does not contain the value of 0 (which would imply no difference), it is more plausible that gamers have faster response times.
- (c) [1 pts] We are 95% confident that the mean difference in accuracy score between gamers and non-gamers is at least -4.2 and at most 5.8.
- (d) [1 pts] Since this interval covers the value of 0 (which implies no difference) and is roughly symmetric about this value, it is plausible that gamers and non-gamers are basically the same in accuracy.

3.74) [7 pts]

- (a) [1 pts] This statement fails to address that the confidence interval provides a plausible range for the population average and not each individual student. Furthermore, inferences made from the confidence interval are only generalizable to students within her college.
- (b) [1 pts] She should be 100% sure that the mean pulse rate for the sample is contained in her interval. The interval was constructed using the sample mean as its center, so it is guaranteed to contain it.
- (c) [1 pts] The confidence level does not apply to the range of the confidence interval. She should be 100% confident about the range of her interval given that that is what she computed.
- (d) [1 pts] The confidence level of a confidence interval does not indicate the percentage of your population datapoints contained in your interval.
- (e) [1 pts] This confidence interval was formed by taking a random sample of students within her college. It is not generalizable to all US college students.
- (f) [1 pts] The interpretation here is incorrect for reasons similar to (d). Here, it is a bit worse given that it does not make sense to describe a mean pulse rate for an individual student. The random sample contained single measurements of pulse rates for each student - not a series of measurements that could be used to compute an average for every student.
- (g) [1 pts] The confidence level refers to our confidence that the interval will contain the true population mean. It does not refer to the percentage of sample estimates that it will contain.

3.91) [2 pts]

We estimate the mean difference using our sample and obtain a value of 17.12. We use StatKey to generate the bootstrap distribution below:



Using the standard error (7.749), we compute a 95% bootstrap confidence interval: $17.12 \pm 2 * 7.749 = (1.62, 32.62)$. This interval suggests that the interferon gamma production is increased among tea drinkers. We are 95% confident that the mean difference is at least 1.62 and at most 32.62.

3.133) [7 pts]

- (a) [1 pts] The population parameter of interest is ρ , the correlation between weight gain over one month of bingeing and a person's weight 2.5 years later. The best estimate for this parameter is the sample correlation, $r = 0.21$.
- (b) [1 pts] To obtain a single bootstrap sample, we would sample with replacement from this original sample until 18 cases were obtained. We would make sure to maintain the ordered pairs in our data for this resampling.
- (c) [1 pts] Using this bootstrap sample, we would compute the sample correlation.
- (d) [1 pts] The standard error would be found by computing the standard deviation using all of the bootstrap statistics.
- (e) [1 pts] Given our original sample correlation ($r = 0.21$), the 95% confidence interval would be: $0.21 \pm 2 * 0.14 = (-0.07, 0.49)$
- (f) [1 pts] Given the interval in (e), there is some evidence for a positive correlation. The majority of our interval is to the right of 0. However, since our interval does contain 0, it is also (slightly) plausible that there is no correlation at all.
- (g) [1 pts] A 90% confidence interval would be narrower.