# STA130 W7(Review)

Monday, October 28, 2019 2:10 PM

**Goal:** Come up with a range of plausible values for what the true value of the parameter could be.
What do we need to achieve this?: What we really want is the **sampling distribution** of the statistic...
Why ?

- Tells us about the distribution of other values of the statistic that we could have obtained from diff. samples(we don't know this, we need to estimate it)

How can we estimate the sampling distribution of the statistic?

1. Get many samples from the population, calculate the statistic for each one, look at the distribution of these statistics(we need to have all data from the population)
2. If don't have all data from the population, get many **bootstrap samples** from the original sample, calculate statistic each time, and look at distribution of values

Environmental scientists are interested in estimating the mean mercury content of fish in a lake. They collect a random sample of 50 fish from the lake, measure the mercury content for each, and calculate the average mercury content for these 50 fish. The 99% bootstrap confidence interval is (0.82, 1.13). How many of the following are valid interpretations?
(i) We are 99% certain that each fish has between 0.82 and 1.13 ppm of mercury.
'each fish' - CI does not tell us about individual fish, but rather about the population mean.
(ii) We expect 99% of the fish to have between 0.82 and 1.13 ppm of mercury.
Not a statement about population mean, but values for fishes
(iii) We are 99% confident that the confidence interval of (0.82, 1.13) includes the true mean of the mercury content of fish in the lake.
(iv) There is a 99% chance that the true mean mercury content of fish in the lake is between 0.82 and 1.13 ppm.
'99% chance' - CI does not tell us the probability that the true value lies between 10.82 and 1.1.3
(v) If we got another sample of 50 fish from the lake, we are 99% confident that the sample mean would be between 0.82 and 1.13.
The samples does not have relationships - 'sample mean' - there is not reason to think that the sample mean from sample 2 would fall from confidence interval for sample one

A medical doctor is investigating the efficacy of two treatments on blood pressure. She randomly assigns patients to treatment 1 and treatment 2 and compares the median change in blood pressure between the two groups.
Which of the following ggplot2 geometries would be most appropriate to explore the distribution of changes in blood pressure for individuals in the two treatment groups?
(A) geom_bar()
(B) geom_histogram()
(C) geom_boxplot()
(D) geom_point()
(E) facet_wrap()
(F) More than one of the above
B + E create 2 graphs for comparison
Boxplots show the median

Which of the following statistical methods would be most appropriate to help the doctor compare the median change in blood pressure for treatment 1 and treatment 2?
(A) Statistical test for one proportion
(B) Randomization test
(C) Bootstrap confidence interval
(D) More than one of the above
(E) None of the above

A medical doctor is investigating the efficacy of two treatments on blood pressure. She randomly assigns patients to treatment 1 and treatment 2 and compares the median change in blood pressure between the two groups.
The doctor carries out a statistical test and the p-value is 0.001. How many of the following are valid interpretations of the p-value?
(A) The p-value is the probability of observing a difference in median changes in blood pressure as large or larger than she observed.
(B) The p-value is the probability of observing a difference in median change of BP between the two treatment groups at least as large has she observed, if both treatments are equally effective.
(C) The p-value is the probability that both treatments work equally well
(D) The p-value is the probability that the median change in blood pressure is different in the two treatment groups.