

Lab 10a: Sampling Distributions

Stat 131A, Fall 2018

Learning Objectives:

- Distribution of sample proportions
- Distribution of sample means

Problem 1

In the 2000 presidential election, newspapers reported that Obama received 40% of the white male vote. If this is true, would you be surprised to see less than 35% support for Obama in a local exit poll of 100 white male voters?

- a. Yes, this is surprising. Random samples this large will not usually have this much error.
- b. Yes, this is surprising. We expect 40% of those polled to support Obama.
- c. No, this is not surprising. This sample result is slightly more than one standard error below 40%.
- d. It is impossible to tell because the sample is not randomly selected.

Problem 2

The popular candy Skittles comes in five colors. According to the Skittles website, the five colors are evenly distributed in the population of Skittle candies, so each color makes up 20% of the population.

Suppose that we purchase a small bag of Skittles. Assume that this size bag always has 20 candies. In this particular bag, 6 are green: 6 out of 20 is 30%. Is this a surprising result?

- a. Yes, this result is surprising. Random samples with this much error are unusual.
- b. Yes, this result is surprising. We expect 20% of candies to be green.
- c. No, this result is not surprising. It is a little bit more than one standard error above 20%.
- d. No, this result is not surprising. We expect random samples to vary.

Problem 3

One year, there were 252 trading days on the Nasdaq stock market, and the Amazon stock went up 131 of them: $131/252 \approx 0.52$ or 52%. A statistician attaches a standard error to this proportion as follows:

$$\bullet \text{ SE}(\text{proportion}) = \sqrt{0.52 \times 0.48} / \sqrt{252} \approx 3\%$$

Is this the right SE? Answer yes or no, and explain.

Problem 4

Imagine that you have a very large barrel that contains tens of thousands of M&M's. According to the official M&M website, 20% of the M&M's produced by the Mars Corporation are orange. 5 students each take a random sample of 50 M&M's and record the percentage of orange in each sample.

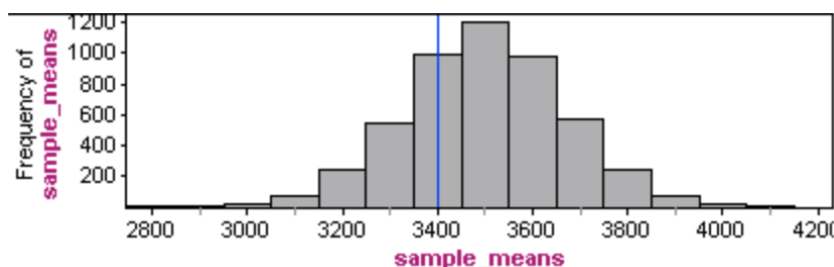
Which sequence is the most plausible for the percent of orange candies obtained in these samples?

- a. 20%, 20%, 20%, 20%, 20%
- b. 15%, 25%, 22%, 20%, 28%
- c. 5%, 80%, 8%, 65%, 70%
- d. any of the above

Problem 5

Suppose that babies in a town had a mean birth weight of 3,500 grams in 2005. This year, a random sample of 9 babies has a mean weight of 3,400 grams.

The following graph shows the sampling distribution for samples of size 9 from a simulation where $\mu = 3,500$. It looks different because we zoomed in. This changed the scale for the sample means. We also marked the actual sample mean of 3,400 grams.

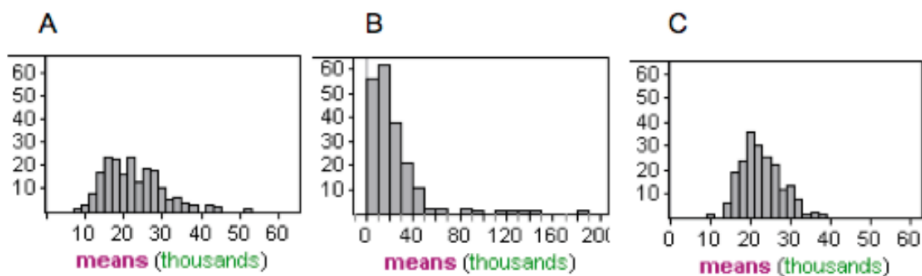


- a. Are these statements valid or invalid?

- i) A sample with a mean of 3,400 is unlikely because it is 100 grams below the population mean of 3,500.
 - ii) A sample with a mean of 3,400 is likely because samples with smaller means occurred fairly often.
- b. Which conclusion is the most reasonable?
- i) The actual sample is 100 grams less than the town's mean birth weight last year. So this sample gives strong evidence that the town's mean birth weight is less than 3,500 grams this year.
 - ii) The sample gives strong evidence that the mean birth weight for the population of all babies is 3,400 grams. There has been a decrease in birth weight from the 2005.
 - iii) A sample with a mean birth weight of 3,400 grams would not be surprising if the mean birth weight in the population is 3,500 grams. So the sample does not provide convincing evidence that the town's mean birth weight is less than 3,500 grams this year.

Problem 6

The distribution of incomes is strongly skewed to the right for individuals in the U.S. The following histograms represent mean income from 200 samples randomly selected from the U.S. population. One histogram is based on samples of size of $n = 4$, one on samples of size of $n = 40$, and one on samples of size of $n = 100$.



- a. Which histogram represents mean incomes from samples of size 4?
- i) A
 - ii) B
 - iii) C
- b. Which histogram represents mean incomes from samples of size 40?
- i) A
 - ii) B
 - iii) C

Problem 7

The Federal Pell Grant Program gives grants to low-income undergraduate students. According to the National Postsecondary Student Aid Study conducted by the U.S. Department of Education in 2008, the average Pell grant award for 2007-2008 was \$2,600. Assume that the standard deviation in Pell grant awards was \$400.

- a. If we randomly sample 20 Pell grant recipients and record the mean Pell grant award for the sample, then repeat the sampling process many, many times, what is the mean and standard error of the sample means?
- b. Which of the following random samples is more surprising? Why?
 - i) Sample A: a random sample of 9 Pell grant recipients with a mean award amount of \$2750.
 - ii) Sample B: a random sample of 36 Pell grant recipients with a mean award amount of \$2750.

Problem 8

We randomly select 100 Pell grant recipients from two states. State A is a relatively small state with approximately 4,000 Pell grant recipients. State B is a large state with approximately 200,000 Pell grant recipients.

Suppose that the mean and standard deviation in individual Pell grants is approximately the same for both states: $\mu = \$2,600$ and $\sigma = \$800$.

For which state is the sample mean most likely to be within \$80 of \$2,600?

- a. Equally likely because $\sigma = \$800$ for both states.
- b. State A because the sample represents a larger segment of this small population.
- c. State B because there is less variability in larger populations so estimates from samples are more accurate.

Problem 9

A box contains 10,000 tickets. The numbers on these tickets average out to 50, and the SD is 20.

- a. One hundred tickets are drawn at random with replacement. The average of these draws will be around _____, give or take _____ or so.
- b. What are if 100 draws are made without replacement?

- c. What if 100 draws are made without replacement, and there are only 100 tickets in the box?

Problem 10

A box of tickets averages out to 75, and the SD is 10. One hundred draws are made at random with replacement from this box. Show your work.

- a. Find the chance (approximately) that the sample mean will be in the range 65 to 85.
- b. Repeat, for the range 74 to 76.

Problem 11

One hundred draws are made at random with replacement from a box. The average of the box is 3.1.

- a. True or False: the expected value for the average of the draws is exactly equal to 3.1. If this cannot be determined from the information given, what else do you need to know, and why?
- b. What is the SE for the average of the draws? If this cannot be determined from the information given, what else do you need to know, and why?