

HOMEWORK 13

Due: March 17

1. A fundraising organization typically gets a return from about 5% of the people on their mailing list. To see what the response rate might be for future appeals, they hired a statistician to do a simulation using sample sizes of 20, 50, 100, and 200. For each sample size the statistician simulated 1000 mailings with from a population with a response rate of 5% and constructed a histogram of the 1000 sample proportions, shown below.

- a. Of the 1000 simulated samples of size 100 that were taken, about how many produced a response rate that was between 7% and 9%?

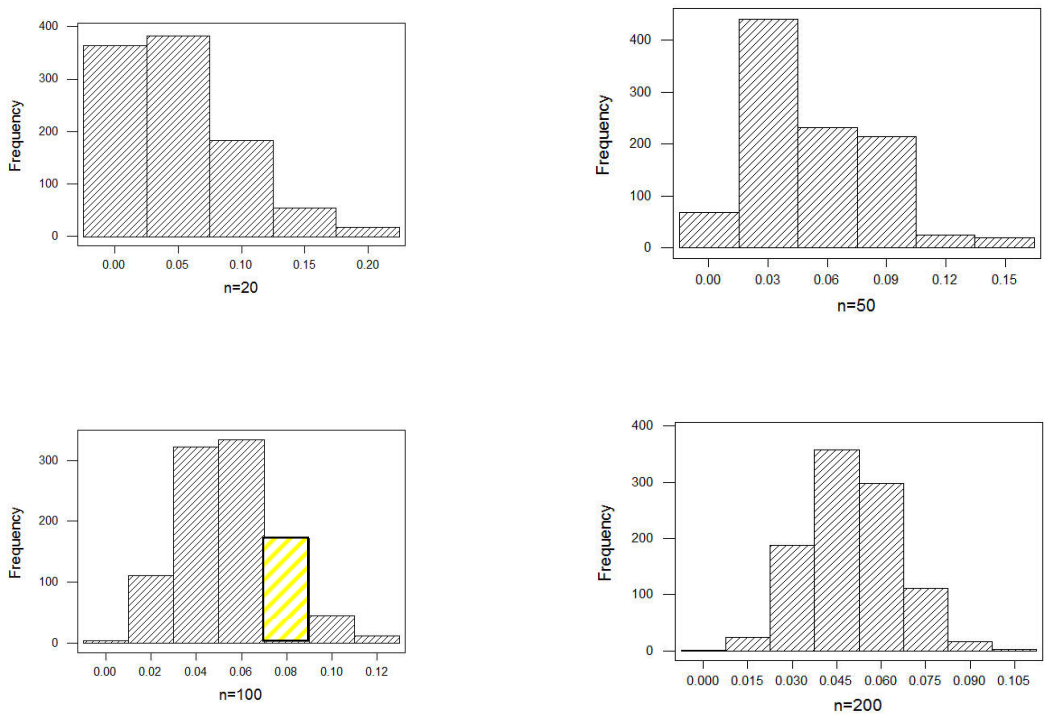
Looking at the histogram of $n = 100$, we see that about 190 of the simulated samples produced a response rate that was between 7% and 9%. (See highlighted rectangle.)

- b. Explain how these histograms demonstrate what the Central Limit Theorem says about the sampling distribution for a sample proportion. Be sure to talk about shape, center, and spread.

Shape: as the sample size increases, the shape of the sampling distribution gets closer and closer to a bell-shaped curve.

Center: the center is about the same for all four distributions. The center of the sampling distribution doesn't depend on the sample size.

Spread: as the sample size increases, the spread of the distribution decreases.



According to the Central Limit Theorem for Sample Proportions, what should the theoretical means and standard deviations be for each of these sample sizes? Fill in their values in the table below, using three decimal places for each:

Mean : $\mu_{\hat{p}} = p = 5\% = 0.05$

S.d. = $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

n	20	50	100	200
Mean:	0.050	0.050	0.050	0.050
SD:	0.049	0.031	0.022	0.015

- b. The actual means and standard deviations of the sampling distributions shown in the histograms above are as follows:

n	20	50	100	200
Mean:	.049	.052	.050	.050
SD:	.048	.031	.021	.015

How close are those theoretical values to what was observed in these simulations? Would you say that using simulations is a reasonable way to obtain information about sampling distributions?

The actual values from the simulation are very close to the theoretical values. Based on this, I would say using simulations is a reasonable way to obtain information about sampling distributions.

- c. Looking at the histograms, at which of the four sample sizes used would you be comfortable using the Normal model as an approximation for the sampling distribution?

n = 100

That histogram looks roughly symmetric

2. Remember that the simulations in #1 assumed that the population response rate (the parameter) was 5%. But in reality, each time that the fundraising organization does a mass mailing, this response rate may be different according to what the mailing is asking for, as well as seasonal and economic factors. Suppose now that the organization does a test mailing for a new appeal they plan to send out soon. The test mailing involves sending the appeal to 100 names that are randomly selected from its mailing list. They receive a reply from 8 of them, for a response rate of 8%.

- a. Based on the relevant histogram shown in #1, if the response rate for the population will really be 5%, would it be unusual to obtain a sample response rate of 8% in 100 appeals?

No, it 8% response rate doesn't seem to be unusual. The response rate was between 7% and 9% about 19% of all the mailings (about 190/1000).

b. Do the test mailing results give convincing evidence that the response rate for this appeal will be higher than 5% when it is sent out to everyone on the mailing list (the population)? Explain.

No, from the histogram for $n = 100$, we can see that the 8% response rate is not unusual. The response rate of 8% in the 100 appeals can happen just by chance quite often. Since it's not an unusually high response rate, we don't have convincing evidence that the response rate will be higher than 5% when it is sent out to the population.

c. Let p be the population response rate for the new appeal. State the null and alternative hypotheses that correspond to the question described in part a, both in words, and in terms of p .

H_0 : the population response rate for the new appeal will be 5%. $p = 0.05$

H_a : the population response rate for the new appeal will be higher than 5%. $p > 0.05$

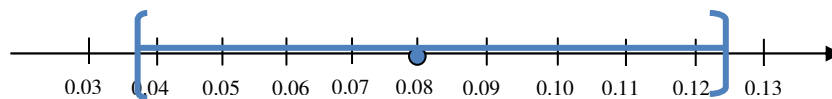
d. Inference (estimation): Based on the test mailings, the point estimate of population response rate that the organization will achieve for this mailing is obviously 8%. Find an *interval estimate* for the population response rate. Use the formula *point estimate \pm margin of error*. You will soon learn a precise formula for calculating the margin of error; for now use two s.d.'s, using the standard deviation of the sampling distribution for $n = 100$ given in #1. Work in decimal but then convert your final answer back to percentage form.

Point estimate = 0.08

The s.d. for $n = 100$ is 0.022.

Thus, the interval is $0.08 \pm 2(0.022) = (0.036, 0.124)$

Write out the interval estimate in interval form: (,) and sketch it on a number line.



e. Does the interval estimate include the value 5%? How does this relate to your answer to part b?

Yes, 5% is in the interval, so it's a plausible value. Same conclusion we had in part b. Even though the sample proportion was 8%, our null hypothesis is still believable; we cannot reject it.

3. Redo Problem #2 assuming the test mailing had involved a sample size of 200 rather than 100.

a. Based on the relevant histogram shown in #1, if the response rate for the population will really be 5%, would it be unusual to obtain a sample response rate of 8% in 100 appeals?

Now looking at the histogram for $n = 200$, the 8% response rate does seem to be a little be unusual. It's farther away from the mean.

b. Do the test mailing results give convincing evidence that the response rate for this appeal will be higher than 5% when it is sent out to everyone on the mailing list (the population)? Explain.

From the histogram for $n = 200$, we can see that the 8% response rate is more unusual than it was with $n = 100$. It could happen just by chance that the response rate is 8% with a sample size of 200, but the chance for this to happen is less than it was with $n = 100$. So here I am more convinced that the response rate will be higher than 5% when it is sent out to the population.

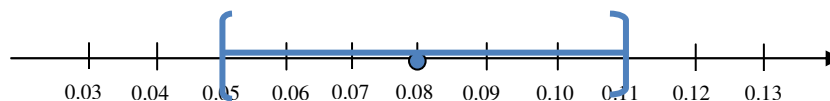
c. Let p be the population response rate for the new appeal. State the null and alternative hypotheses that correspond to the question described in part a, both in words, and in terms of p .

H_0 : the population response rate for the new appeal will be 5%. $p = 0.05$
 H_a : the population response rate for the new appeal will be higher than 5%. $p > 0.05$

d. Inference (estimation): Based on the test mailings, the point estimate of population response rate that the organization will achieve for this mailing is obviously 8%. Find an *interval estimate* for the population response rate. Use the formula *point estimate \pm margin of error*. You will soon learn a precise formula for calculating the margin of error; for now use two s.d.'s, using the standard deviation of the sampling distribution for $n = 100$ given in #1. Work in decimal but then convert your final answer back to percentage form.

Point estimate = 0.08
The s.d. for $n = 200$ is 0.015.
Thus, the interval is $0.08 \pm 2(0.015) = (0.05, 0.11)$

Write out the interval estimate in interval form: (,) and sketch it on a number line.



e. Does the interval estimate include the value 5%? How does this relate to your answer to part b?

5% is right on the edge of the interval. This gives the same conclusion we had in part b. If the response rate is REALLY 5%, then it should be in the confidence interval. Since it's right on the edge, it is harder to believe that the response rate will be 5%.

4. Assume that 30% of students at a university wear contact lenses. We randomly pick 100 students. Let \hat{p} represent the proportion of students in this sample who wear contacts.

- a. Describe and draw the sampling distribution model for the proportion of students in such a random sample who wear contacts. Make sure to mention shape, center, and standard deviation in your description. Also, be sure to verify that the conditions are met before using the Normal model to describe the sampling distribution of the sample proportion.

$$n = 100, p = 30\% = 0.30$$

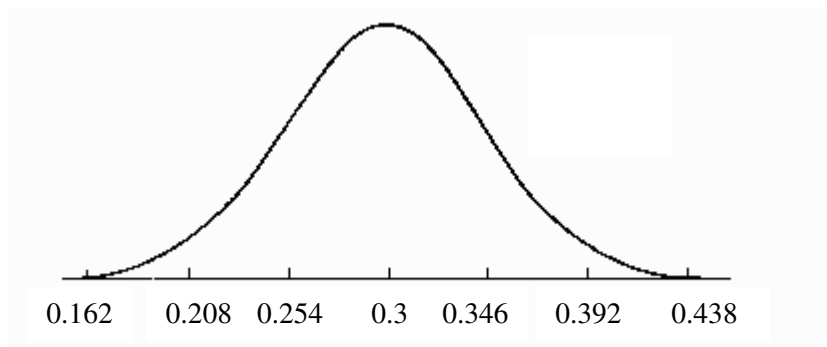
$$\text{Conditions: } np = 100(0.3) = 30 > 10$$

$$n(1-p) = 100(1-0.3) = 70 > 10$$

The conditions are satisfied.

Since the conditions are satisfied, we can assume that the shape of the sampling distribution of the sample proportions will be approximately normal with mean $p = 0.3$, and standard

deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3(1-0.3)}{100}} \approx 0.046$



- b. What's the approximate probability that more than one third of this sample wear contacts?

$1/3$ is about 0.33. That is less than 1 standard deviation above the mean. More precisely,

$$z = \frac{0.33 - 0.3}{0.046} = 0.652$$

Then using either the z-table, or the calculator (upper tail: $\text{normalcdf}(0.652, 99999) = 0.2572$), we can say that the probability that more than one third of this sample wear contacts is about 26%.

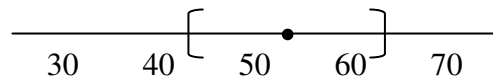
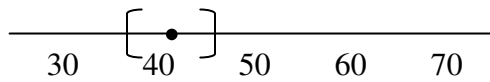
- c. What's the approximate probability that less than 20% of this sample wear contacts?

20% is 0.2. That is about 2 standard deviations below the mean. More precisely,

$$z = \frac{0.2 - 0.3}{0.046} = -2.17$$

Then using either the z-table, or the calculator (lower tail: $\text{normalcdf}(-999999,-2.17)=0.015$), we can say that the probability that less than 20% of this sample wear contacts is about 1.5%.

5. Two interval estimates of a certain parameter are shown below. One is based on a sample size of 350, and the other is based on a sample size of 1200. Both samples were taken randomly. Indicate which interval estimate goes with which sample size, and indicate the locations of the two point estimates:



As we know, larger sample size implies that the standard deviation of the sampling distribution is smaller. Then, since the margin of error is twice the s.d., the margin of error will be smaller, meaning narrower confidence intervals. Thus, the first figure must be the one that used the sample size of 1200, and the second one show the confidence interval with sample size of 350.

The point estimates are always in the middle of the interval, so the points in the middle indicate the point estimate (either the sample mean or the sample proportion).