*YOUR TA’S NAME*:

*Lecture Worksheet*

*Tuesday 10/6/2020*

**MAIN POINTS OF LECTURE**

1. A sampling distribution is a theoretical distribution of many means or proportions, taken from many independent random samples of size n
2. Sampling distributions of means and proportions…
	1. …are normal (regardless of the shape of the distribution of the variable in question)
	2. …have expected values equal to the population mean () or proportion (p)
	3. …have variance 
3. Central Limit Theorem: If n is sufficiently large, then the sample means from many random samples from a population with mean **Y and variance **2Y are approximately normally distributed with mean **Y and variance${σ\_{Y}^{2}}/{\sqrt{n}}$
4. Proportions: If we generate many random sample of the same size, then the sampling distribution of the several sample proportions (“*p*-hats”) will be centered over the population proportion (p) and will have a standard deviation of $ \sqrt{\left(p(1-p)/n\right)}$ (but only if there is some fixed *p*, if the samples are random, and if both *np* and *n*(*p*-1) are 5 or more).
	1. Because the standard deviation includes the unknown quantity p, we estimate the standard *deviation* using the standard *error* $ \sqrt{\left(\hat{p}(1-\hat{p})/n\right)}$.
	2. As n increases, the variability of the sampling distribution will decline
5. Means: If we generate many random sample of the same size, then the sampling distribution of the several sample means (“*x*-bars”) will be centered over the population mean (Y) and will have a standard deviation of $σ\_{Y}/\sqrt{n}$ .
	1. Because the standard deviation includes the unknown quantity **Y, we estimate the standard *deviation* using the standard *error* $s\_{Y}/\sqrt{n}$.
	2. *As n increases*, the variability of the sampling distribution will decline
6. **Main Insight of Entire Course**: Because sampling distributions of means and proportions are normal, we can calculate the probability that one random sample falls within plus or minus Z standard deviations from the true population mean or proportion. By extension, **we can calculate the probability that the true population mean or proportion falls within Z standard deviations of one sample’s observed proportion or mean** … and one sample is all we ever really have
7. Computing Z scores for sample means requires knowledge of the standard deviation of the sampling distribution of means ($σ\_{Y}/\sqrt{n}$). When we use standard errors of the sampling distribution of means to approximate the standard deviation of this sampling distribution, we use t scores instead of Z scores … more on this later

**QUESTIONS**

1. [From the recorded lecture]: I selected one random sample of 200 people and got a sample proportion (p-hat) of 0.4. What is the probability that my p-hat differs from p by more than 0.035?

With that sample proportion and sample size, the standard error of the sampling distribution =$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}=\sqrt{\frac{0.4(1-0.4)}{200}}=0.035$. The probability of getting a sample proportion that differs from the population proportion by more than 0.035 is the same as the probability of getting a Z score that differs from zero by 0.035/0.035=1.

P(Z<-1)=0.1587 and P(Z>1)=0.1587, so P(Z<-1 or Z>1)=0.1587+0.1587=0.3174. There is a 31.74% chance of getting a sample proportion that differs from the population proportion by that much.

1. [From the recorded lecture]: I selected one random sample of 1,000 people and got a sample mean (x-bar) of 200 with a standard deviation, s, of 50. What is the probability that my x-bar differs from m by more than 3?

With that sample mean and sample size, the standard error of the sampling distribution =$\sqrt{{s^{2}}/{n}}=\sqrt{{50^{2}}/{1000}}=1.581$. The probability of getting a sample mean that differs from the population mean by more than 3 is the same as the probability of getting a Z score that differs from zero by 3/1.581=1.90.

P(Z<-1.90) = 0.0287 and P(Z>1.90)=0.0287, so P(Z<-1.90 or Z>1.90) =0.0287+0.0287 = 0.0574. There is a 5.74% chance of getting a sample mean that differs from the population mean by that much.

You might have elected to use a t distribution with n-1=999 degrees of freedom (denoted as t999) to solve this problem instead of a Z distribution. If you did that, the t score would still be 1.90. P(t999<-1.90) is between 0.025 and 0.05 and P(t999>1.90) is also between 0.025 and 0.05, so P(t999<-1.90 or t999>1.90) is somewhere between 0.05 and 0.10. Thus, there is between a 5% and 10% chance of getting a sample mean that differs from the population mean by that much. (Same answer as above, if less precise!)

1. [From the recorded lecture]: I selected one random sample of 20 people and got a sample mean (x-bar) of 200 with a standard deviation, s, of 50. What is the probability that my x-bar differs from  by more than 18?

With that sample mean and sample size, the standard error of the sampling distribution =$\sqrt{{s^{2}}/{n}}=\sqrt{{50^{2}}/{20}}=11.18$. The probability of getting a sample mean that differs from the population mean by more than 18 is the same as the probability of getting a t score that differs from zero by 18/11.18=1.61.

Note that because the sample size is smaller than 50 or so and because this is a question about means, we must use a t distributions with n-1 degrees of freedom (denoted as t19) instead of the Z distribution.

P(t19<-1.61) = between 0.05 and 0.1 and P(t19>1.61) also = between 0.05 and 0.1, so P(t19<-1.61 or t19>1.61) = somewhere between 0.1 and 0.2. There is between a 10% and 20% chance of getting a sample mean that differs from the population mean by that much.

1. [From the synchronous class session]: You drew a random sample of 100 cats (because you wanted to have 100 cats on your farm) and observed that 15% of them are very mean cats. What is the probability that in the population of cats, between 10% and 20% are very mean?

See the recording of the synchronous class session; I lay out the answer in detail there.

1. [From the synchronous class session]: You randomly selected 35 people and observed that they go to the bathroom an average of 4.5 times per day with a standard deviation of 1.0 times. What is the probability that the population mean number of times that people go to the bathroom per day is between 3.9 and 5.1?

See the recording of the synchronous class session; I lay out the answer in detail there.