Stat 2470, 2/24 Discussion Questions Name _____

Instructions: Attempt to answer these questions by reading the textbook or with online resources before coming to class on the date above.

1. What is a statistic? How does it differ from a parameter?

a parameter defines the distribution of the population. a statistic is an estimate of the parameter based on a sample.

2. Define sampling distribution.

a sampling distribution is the distribution of values for an estimate of a parameter taken from repeated sampling.

3. Describe when simulation experiments can be used and what do they tell us?

when we know which statistic we wish to measure; how the population is distributed, the size of the sample being taken, and the number of replications (how many samples of size n) are to be taken

4. What happens to the shape of a sampling distribution as the sampling size increases?

it becomes narrower (less variability) and more normal.

5. What is the expected value and variance of the sampling distribution for the mean of a sample?

the expected value of the mean of the sampling distribution is the mean itself, μ. & the variance is σ^2/n.

6. What does the Central Limit Theorem say and why is it so important in statistics?

if n is sufficiently large, then $\bar{X}$ (sampling distribution of the mean of random samples X_i) has approximately normal distribution w/ $\mu_{\bar{x}} = \mu$, and $\sigma_{\bar{x}}^2 = \sigma^2/n$

7. What is the rule of thumb for applying the central limit theorem? What are some examples of situations when larger or smaller sample sizes might be needed (or allowed) for the theorem to apply?

if $n > 30$, then CLT can be used
however, if normal already, smaller sizes may work,
if very non-normal, larger sizes may be needed.

8. What is a point estimate? What notation do we use to indicate a value is a point estimate of a parameter θ?

is a single number regarded as a reasonable estimate for a parameter.

if parameter is Θ, the point estimate designated as $\hat{\Theta}$

9. What are some common point estimates? For p, μ, σ?

$\hat{p} = \frac{x}{n}$, $\hat{\mu} = \bar{x} = \sum \frac{x_i}{n}$, $\hat{\sigma} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

10. What is the primary downside of using a point estimate?

no information on accuracy is available

11. What does it mean for an estimator to be unbiased?

$\hat{\Theta}$ is unbiased if $E(\hat{\Theta}) = \Theta$ for every possible value of Θ.

12. What is the formula for the bias of $\hat{\theta}$?

$\text{bias} = E(\hat{\Theta}) - \Theta$

13. Give at least two examples of unbiased estimators. Give at least two examples of biased estimators.

$\hat{p} = \frac{x}{n}$, $\hat{\mu} = \frac{\sum x_i}{n} = \bar{x}$

14. What is the principle of unbiased estimation?

When choosing among several estimators, select one that is unbiased.

15. What is the minimum variance unbiased estimator (MVUE) of θ? When does this principle apply?

if there is more than one unbiased estimator for a parameter θ, choose the one w/ the least variance.

16. What does the MVUE principle mean for estimating μ?

it means use $\bar{x}$ for $\hat{\mu}$, rather than $\tilde{x}$, or $\bar{x}_{tr,\%}$

17. We can improve the value of a point estimate by providing information about its variance, or the size of its possible error. What is the standard error for $\hat{p}$ and $\hat{\mu}$?

the standard error for $\hat{p}$ $(\sigma_{\hat{p}}) = \sqrt{\frac{pq}{n}}$

the standard error for $\hat{\mu}$ $(\sigma_{\hat{\mu}}) = \frac{\sigma}{\sqrt{n}}$

18. If normal methods for estimating the standard error fail, then a bootstrap technique can be employed. Briefly describe this method.

when the distribution is known, collect data to get $\hat{\theta}$. Then use the computer to simulate samples and $\hat{\theta}^{*(B)}$ estimates. Then find the mean of the bootstrap samples $= \bar{\theta}^*$

then the standard error is $s_{\hat{\theta}} = \sqrt{\frac{1}{B-1}\sum(\hat{\theta}_i^* - \bar{\theta}^*)^2}$

(for large B, use B instead of B-1)

typically used when the standard error of the sampling distribution is difficult to calculate.