Stat 2470, 3/24 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. When comparing two samples, what is the natural estimator for $\mu_1 - \mu_2$? (the difference between the two population means?)

$\bar{X}_1 - \bar{X}_2$ or $\bar{x} - \bar{y}$

depending on notation used

2. What is the formula for the standard deviation for the difference between the two means?

$$\sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

3. What is the formula for the test statistic for the hypothesis test?

$$Z = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \quad \text{or} = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

4. In our testing procedures, what does Δ_0 represent?

the value of the difference between the two means

5. Our calculator is set up to easily test two samples when $\Delta_0 = 0$. How can we use the calculator to do the test when $\Delta_0 \neq 0$?

rearrange the statement $\mu_1 - \mu_2 = \Delta_0$ to be $\mu_1 = \mu_2 + \Delta_0$

and adjust the value of $\bar{x}_2$ (or $\bar{y}$) accordingly

6. When is the drawback of using observational studies to try to establish causation?

we can only see relationships not causation

Causation requires experiments so that lurking variables can be controlled for.

7. How are randomized controlled experiments better than retrospective studies?

accounts better for lurking variables and the bias of our own memories

8. How do we calculate β for the two-sample Z-test? In what ways does it differ from the one-sample test?

one-tailed: $\Phi\left(z_\alpha - \frac{\Delta' - \Delta_0}{\sigma}\right)$ or $1 - \Phi\left(-z_\alpha - \frac{\Delta' - \Delta_0}{\sigma}\right)$

two-tailed: $\Phi\left(z_{\alpha/2} - \frac{\Delta' - \Delta_0}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\Delta' - \Delta_0}{\sigma}\right)$

9. What are the conditions on using a Z-test with s_1 and s_2 (instead of σ_1 and σ_2) in terms of the required sample sizes?

both m & $n > 40$
and normality

10. What is the formula for confidence intervals for the difference between two means?

$$\bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

11. What is the point estimate at the center of the interval?

$\bar{X} - \bar{Y}$

12. How can we use confidence intervals to conduct hypothesis tests?

if assumption for H_0 (or for sample) is inside the confidence interval, fail to reject; if outside, then reject H_0 ← depending on center of interval based

13. How can we do a two-sample confidence interval (using z) in our calculator?

2-SampZInt under Stat → Tests (TI-84)

14. What is the formula for the required sample sizes for some width of the confidence interval, assuming the two sample sizes are the same?

$$n = \frac{4z_{\alpha/2}^2(\sigma_1^2 + \sigma_2^2)}{w^2}$$

15. How is the formula for the T-test statistic for the two sample means different than the two-sample Z-test?

it's not, except that it uses s_1 and s_2 instead of σ_1 and σ_2

16. Calculating the degrees of freedom for the two-sample T-test is quite complicated. What is the formula? How is the value of n rounded?

$$df = \nu = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}}$$

(a more conservative calculation is to use the smaller sample size minus 1)

17. An even more conservative way to calculate the degrees of freedom (but one that is considerably less complicated) is to use the smaller of the two sample sizes (minus one). What are the downsides of this approach?

it will make it harder to reject H_0 since the t critical value will be larger for the smaller sample size

18. How do we conduct this test in our calculator when $\Delta_0 \neq 0$? What are the steps?

for H_0: $\mu_1 - \mu_2 = \Delta_0 \Rightarrow \mu_1 = \mu_2 + \Delta_0$ adjust the value of $\bar{X}_2$ by adding the value of Δ_0 to it.

keep all other values the same.

19. Why is the order of subtraction when setting up our procedures so important?

because it affects the direction of the inequality in one-tailed tests.

20. What is the formula for the pooled T-test?

$$s_p^2 = \frac{m-1}{m+n-2} s_1^2 + \frac{n-1}{m+n-2} s_2^2 \quad \text{then } T = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{s_p^2\left(\frac{1}{m} + \frac{1}{n}\right)}}$$

21. What assumption does a pooled T-test make? Is this assumption reasonable in most cases?

That the standard deviations are the same.

Not, not really reasonable most of the time one should have good reason to make that assumption