Stat 2470, 3/10 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. How are intervals based on T different than intervals based on Z?

they are basically the same, but a bit wider due to the additional uncertainty of having S instead of σ

2. Why do we care about the degrees of freedom of the student t distribution?

Sample size minus one

$n-1$

3. Why is this a more conservative estimate (and more accurate) for small sample sizes when σ is not known?

the t-distribution has more probability in the tails and so allows better for estimating from estimates

4. What happens to the shape of the student t distribution as the degrees of freedom increases?

converges to the normal distribution

5. How do we find t critical values in the calculator?

for TI-84's w/ it $invT(\alpha/2, df) = t_{\alpha/2,df}$ $df = n-1$ in this chapter under DISTR

if there is no invT, then you'll need to use $tcdf(-E99, guess, df)$ until you get $\alpha/2$, or $1-\alpha/2$. Then the "guess" is the value for $t_{\alpha/2,df}$.

6. What is the formula for the confidence interval based on t?

$$\left(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right)$$

7. How do confidence intervals based on t vs. those based on z differ in terms of their width at the same level of confidence?

t intervals are a bit wider for small sample sizes.

8. What is the prediction interval formula for a single observation?

$$\bar{X} \pm t_{\alpha/2,n-1}\, s\sqrt{1+\frac{1}{n}}$$

9. What kind of assumptions are in the background when we use t confidence intervals?

using s instead of σ
distribution approximately normal

10. What is a tolerance interval?

Captures a % of the population (individual values) and level of confidence

$\bar{X} \pm$ (tolerance critical value) $\cdot s$

11. Compare the tolerance critical values to the z and t values for the 95% case.

ex. $n = 5$ 95% confidence, captures 95% of population 5.079
$n = 10$ 95% confidence, 95% of population 3.379
$n = 40$ " " " 2.445

converges to 1.96 as sample size increases.

12. What are some strategies for dealing with very non-normal distributions?

bootstrapping is an option
see Statistical Intervals by Hahn, Gerald & William Meeker

13. What is a statistical hypothesis?

a claim or assertion either about the value of a parameter, several parameters or the form of a population distribution.

14. What is the null hypothesis in a hypothesis test and how is it notated?

the claim that is initially assumed to be true, H_0

i.e. $H_0: \mu = 175$

15. What is the alternative hypothesis and how is it notated?

the claim we are testing (contradicts the assumption) H_a

i.e. $H_a: \mu > 175$

16. Why do we say "reject" and "fail to reject" H_0 rather than "reject" and "accept" H_0?

we said "fail to reject" since H_0 is an assumption, but one which we are not gathering evidence to support, much as in court we do not assert that a failure to convict "proves" innocence.

17. What assumption(s) do we make when testing a hypothesis?

we assume H_0 is true and reject if H_0 is "implausible" according to some preset standard.

18. What is the general rule for setting up the null hypothesis in terms of claiming equalities, inequalities, etc.

H_0 should include equality $=, \leq$ or $\geq$

H_a should not include equality $\neq, >$, or $<$

19. What are the general test procedures in terms of the test statistic, critical values and rejection regions?

① calculate the test statistic by comparing the assumption to the data collected.

② based on what we consider to be the threshold for "implausibility" of H_0, we calculate the critical region boundary to compare to test statistic

③ if statistic is beyond the critical value reject H_0, if not, fail to reject H_0

20. How do these procedures relate to P-values and levels of significance? (You may need to jump ahead in the chapter to answer this.)

the level of significance is used to calculate the critical value. However, we can convert the test-statistic to a p-value. The chance we could get the value (or a more extreme one) given the assumptions of H_0 and compare that directly to the significance level. If less plausible reject; if more plausible, fail to reject.

21. What are the two types of errors in a hypothesis test?

Type I error rejects H_0 when it is true

Type II error fails to reject H_0 when it is fails

22. In legal proceedings, we have the following hypotheses: H_0: the suspect is innocent and H_a: the suspect is guilty. Explain a Type I and Type II error in this context.

Type I error sends an innocent man to prison

Type II error lets a guilty man go free.

23. What does α mean in terms of errors in a hypothesis test? By what term is α usually referred to?

α is the significance level; the threshhold level of implausibility we use to reject H_0.

this is related to the chance of a Type I error.

24. What does β mean in terms of errors in a hypothesis test?

β is the chance of a Type II error

25. $1 - \alpha$ is confidence. What is the term for $1 - \beta$?

power

26. How are α and β related for a fixed sample size and test statistic?

$$n = \left[\frac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'}\right]^2$$ for a one-tailed test

$$n = \left[\frac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'}\right]^2$$ for a two-tailed test