STAT303 Secs 506–508
Fall 1999
Exam #3

Instructor: Julie Hagen Carroll

1. **Don't EVEN open this until you are told to do so.**
2. There are 20 multiple-choice questions on this exam, each worth 5 points. There is partial credit. Please mark your answers **clearly** on the scantron. Multiple marks will be counted wrong.
3. You will have 60 minutes to finish this exam.
4. If you are caught cheating or helping someone to cheat on this exam, you both will receive a grade of **zero** on the exam. You must work alone.
5. This exam is worth 100 points, and will constitute 20% of your final grade.
6. Good luck!
7. Be sure to mark your section number, test form and PRINT the section and computer number for Thursday's class in the space beside where you sign your name. Also, PRINT your name at the top of this exam and include your Thursday section and computer number. You will get your scantron and exam back.

1. Which of the following is true?

   *A. The t curve is always centered at zero.

   The t curve, like the Z curve has a mean = 0.

   B. The z curve is used instead of the t curve to be more conservative.

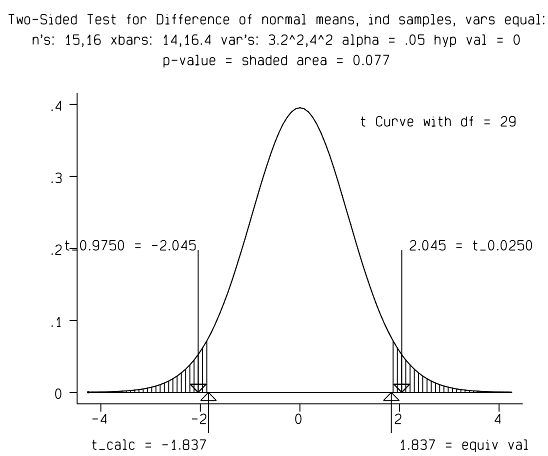
   Conservative means harder to reject or wider intervals, so the t is more conservative than the z.

   C. The t curve is used when we must estimate the mean.

   We always have to estimate the mean, μ. If we knew μ, we wouldn't need to be running a hypothesis test are making a confidence interval. We use a t when we must also estimate the standard deviation, σ, with s.

   D. All of the above are true.

   E. Exactly two of the above are true.



2. What hypotheses are being tested in the graph above?

   The null and alternative hypotheses ALWAYS have the same parameter(s) and value. The null ALWAYS has an equal sign, and the alternative sign must be $\neq$, $>$, or $<$, depending on what you are trying to prove! The title tells you the hypotheses. You should ALWAYS figure them out FIRST. The *name* of the parameter(s) being tested in the 1st line of the title, the *value* being tested is at the end of the 2nd, and the 1st word gives the *sign* in the alternative.

   A. $H_0 : \mu = 0$ vs. $H_A : \mu \neq 0$

   We are testing the difference of TWO means, not just one.

   B. $H_0 : \mu_1 = 14$ vs. $H_A : \mu_2 = 16.4$

   These are the values of the $\overline{x}$'s which are NEVER in the hypotheses.

   C. $H_0 : \mu_1 = 0$ vs. $H_A : \mu_2 = 0$

   The hypotheses must have the SAME parameter.

   *D. $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$

   We are testing whether the DIFFERENCE of TWO means is 0 are not. If $\mu_1 - \mu_2 = 0$, then $\mu_1 = \mu_2$.

   E. $H_0 : \mu_1 - \mu_2 = 0.077$ vs. $H_A : \mu_1 - \mu_2 \neq 0.077$

   The value here is the p-value, not the hypothesized value.

```
90% |Lower Limit = 10.177573
    |Upper Limit = 11.822427

95% |Lower Limit = 10.020018
    |Upper Limit = 11.979982

99% |Lower Limit = 9.7120853
    |Upper Limit = 12.287915
```

3. Given the confidence intervals above, what is the correct range of the p-value for testing $H_0 : \mu = 10.2$ vs. $H_A : \mu \neq 10.2$?

   Use the bottom of the Review Sheet to determine the range. Since the lower endpoints is BELOW 10.2, the value falls within the 90% interval and so the p-value for testing 10.2 is > 0.10.

   *A. p-value > 0.10

   B. $0.10 > p$-value > 0.05

   C. $0.05 > p$-value > 0.01

   D. p-value< 0.01

   E. You need a test statistic value to determine the p-value

4. What would happen to the confidence intervals in the previous problem if we had use a larger sample size?

   Increasing n *decreases* the standard deviation of the statistic and therefore reduces the width of the interval. This also would make the test statistic larger in absolute value (further from 0) since the sd is always in the denominator.

A. Each of the confidence intervals would be narrower.

B. Each of the confidence intervals would be closer to the width of the z interval for the same level, asssuming these are t intervals.

`Since the df=`$n-1$` for the 1-sample case (it's always dependent on the sample size), increasing `n` increases the df and we move down the `t`-table closer to the `z` values.`

C. Each of the confidence intervals would be more likely to include the hypothesized value, 10.2.

`There are two ways you could think of this: 1. Since the intervals are narrower, they would contain `*less*` values and thus less likely to cover 10.2 or 2. The confidence level, how often in repeated sampling you get intervals that actually cover the true parameter, is not affected by `n`, so the likelihood wouldn't change. Either way this statement is not true.`

D. All of the above are true.

*E. Exactly two of the above are true.

`Correct since A. and B. are right.`

5. What would be the consequence of a Type II error for the test, $H_0 : \mu = 10.2$ vs. $H_A : \mu \neq 10.2$?

`Type II error: failing to reject a false null.`

A. We would fail to conclude that the true mean was 10.2 when is actually was 10.2.

`This is `*almost*` a 'fail to reject', but the `*not*` is missing AND it says the null is true.`

*B. We would fail to conclude that the true mean was not 10.2 even though is actually was not 10.2.

`This says we failed to reject a FALSE null.`

C. We would conclude that the true mean was 10.2 when is actually was 10.2.

`Actually this is NEVER what happens since it says we concluded the null was true and it really was true. We NEVER conclude anything about the NULL; we only reject or fail to reject it. We make conclusions about the alternative.`

D. We would conclude that the true mean was not 10.2 when is actually was 10.2.

`This says we rejected a TRUE null---a Type I error.`

E. We could fail to conclude that the true mean was 10.2 even though is actually was not 10.2.

`Again, the `*not*` is missing.`

6. What is the advantage of using a paired t-test (Case 10) over either 2 sample t-tests (Cases 8 or 9)?

`Refer to the '`t`-tests handout'.`

A. You only need half as many observations (smaller sample size).

`You need just as many observations it's just that you have 2 from each pair and we make a sample of `*differences*` which has half as many in it.`

*B. You have more power (easier to detect a difference).

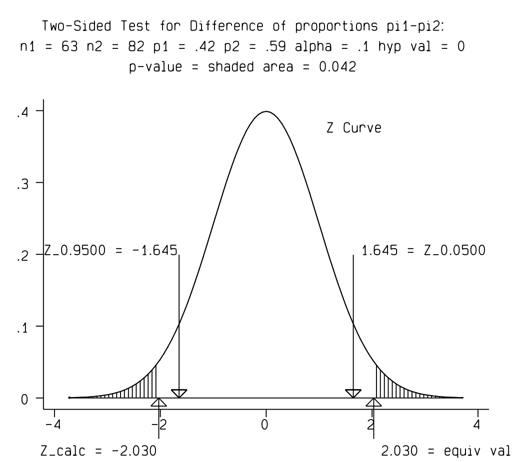
`Yes, the variance is reduced, so smaller differences are now 'detected'.`

C. You have more degrees of freedom (less conservative test).

`Actually, the df for the paired `t`-test is smaller since it uses the number of differences rather than the total number of observations.`

D. All of the above are advantages to the paired t-test.

E. Exactly two of the above are advantages to the paired t-test.



7. What is the best explanation of the p-value for the test above?

   First, you need to know the hypotheses: $H_0 : \pi_1 = \pi_2$ vs. $H_A : \pi_1 \neq \pi_2$. The p-value is how often you will get what you did or worse (in this case less alike since it's a 2-sided test) even though the null is true (in this case the π's are the same).

   A. 4.2% of the time we will get a true H_0.

      Not even a possibility. The null is either true or not.

   B. 4.2% of the time we will get 0.42 and 0.59 for the sample proportions.

      This is this particular sample. 4.2% of time we will get p's further apart even when the π's are equal.

   C. 4.2% of the time we will get differences of 0 when the true proportions are 0.42 and 0.59.

      Close, but backwards. We will get a difference like we did $(0.42 - 0.59)$ or worse when the true difference is 0.

   D. 4.2% of the time we will get differences of 0.42 and 0.59 when the true proportons are 0.

      Again close. The proportions are NOT ZERO if H_0 is true but equal.

   *E. 4.2% of the time we will get differences at least as large as $0.42 - 0.59$ when the true difference is 0.

8. Why would you prefer to run a one-sided tests of hypotheses instead of a two-sided?

   *A. If you only cared about whether you were better (bigger or smaller, whichever would be appropriate), then the one-sided test would give you more power.

      A 1-sided test puts all α on the side you care about where the 2-sided test divides it in half (half on each side). This means you don't have to be as far away to be in the part that is rejected.

   B. If you had a sample mean that was less than what you were testing against, you should run a left-sided test.

      The hypotheses and α are decided BEFORE you take a sample, so you could not know this.

   C. If you knew that the null hypothesis was true, you could still get a rejection with a one-sided test.

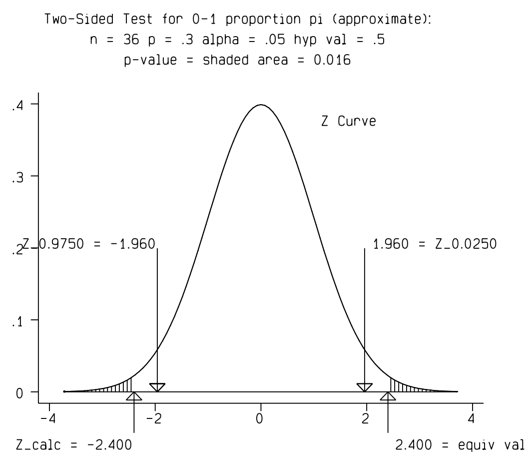
      If you knew the null was true you wouldn't be bothering with a test of hypotheses!

   D. You will always get a rejection with a one-sided test if you rejected with a two-sided test.

      This is true AS LONG AS you picked the correct side! The p-value is the area in the direction of the alternative's sign. If you have a sample mean that is less than the hypothesized value but you're doing a right-sided test, you would have a large p-value ($pv > 0.5$).

   E. The one-sided test requires less data to get a rejection.

      Possibly you could find an example where this was true, but you must decide the hypostheses and α BEFOREHAND, so technically you would never find it.



9. If we have created 90, 95 and 99% confidence intervals using the data above, which intervals would have contained the hypothesized proportion, $\pi = 0.5$?

   Refer to the bottom of the Review Sheet. Since the p-value for testing 0.5 is 0.016 which falls in C: $0.05 > pv > 0.01$, 0.5 would fall within the 99% confidence interval only.

   A. all three
   B. only the 90%
   *C. only the 99%

D. both the 90 and 95%
E. both the 95 and 99%

10. A 95% confidence interval for the true mean test score on this exam is $(68.95, 81.05)$. Which of the following is true?

`Definition: in repeated sampling, approximately 95% of the 95% CI's created will actually contain the true mean test score on this exam.`

*A. The sample mean test score is 75.

`The sample mean is the center of the interval which is 75 for this interval.`

B. You have a 95% chance of making between a 69 and 81.

`CI's are only for the means, not individuals. Plus, for any particular interval, there is no longer any probability; it either contains the true or not.`

C. 95% of the class will make between a 69 and an 81.

`Again, wrong, not the definition of a CI.`

D. 95% of all the different classes will have an average between 68.95 and 81.05.

`Although this at least is about averages, it is still the wrong definition.`

E. Exactly two of the above are true.

11. Suppose we run a hypothesis test at the 5% significance level. Which of the following is true?

A. If we repeatedly sampled the data and ran the same test, we will reject about 5% of the time.

`If the null was FALSE, we would hopefully reject much more than just 5% of the time!`

B. If we repeatedly sampled the data and ran the same test, we will fail to reject about 95% of the time.

`This is just the complement of A., so it's also not true.`

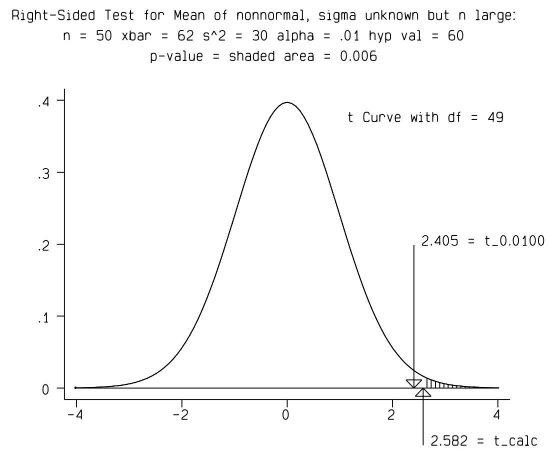
C. If we repeatedly sampled the data and ran the same test, we will make a Type II error about 95% of the time.

α and β are NOT complements. They are areas of different curves: α is a proportion of the curve using the hypothesized value; β is a proportion of a curve using a different value (H_0 is false!)

*D. If we repeatedly sampled the data and ran the same test, we will make a Type I error about 5% of the time.

α is how often we (or how willing we are to) make a Type I error.

E. Exactly two of the above are true.



12. In the test above, what would happened if we had gotten a sample mean, $\overline{x} = 60.5$, instead?

First of all, $H_0 : \mu = 60$ vs. $H_A : \mu > 60$, the p-value = 0.006, and the sample mean, $\overline{x} = 62$. If we got a sample mean *closer* to the hypothesized value, we'd be *more* likely to believe the null and less likely to reject it. This means the p-value would be larger!

A. We would have tested different hypotheses since we had different data.

`The hypotheses are determined BEFORE you gather any data!`

B. We would have used a different α-level since our data was closer to the hypothesized value, 60.

α is also determined BEFOREHAND.

*C. We would have gotten a different p-value since we had different data.
D. All of the above are true.
E. Exactly two of the above are true.

13. Suppose you are interested in the caloric content of hamburgers. You took a random sample from 3 different chains and calculated the following confidence intervals: Chain 1: (248.9,307.8); Chain 2: (263.5,309.3); Chain 3: (307.1,349.7).

Which of the following would be the most appropriate conclusion?

`Since there is` *some* `overlap, we have to allow that there is some possibility that the 3 chains have a common mean caloric content value.`

A. The true mean caloric content for Chain 3 is more than that of Chain 1 or 2.

`Although it is true that the sample mean is higher for Chain 3, the variability of the data allows how possible common values.`

B. The true mean caloric content for Chain 3 is more than that of Chain 1 but not Chain 2.

`Same reasoning as A.`

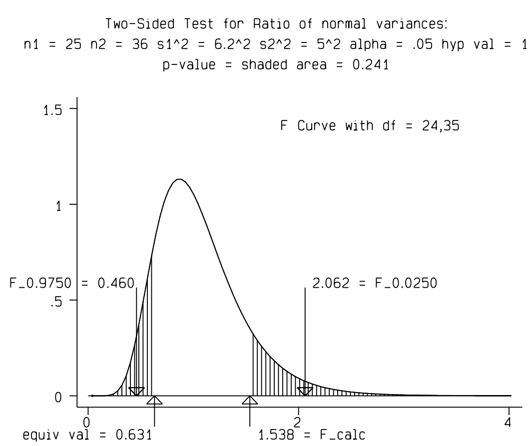
C. The true mean caloric content is different for all three chains.

`The overlap negates this possibility, even though the overlap is slight.`

*D. The true mean caloric content is possibly the same for all three chains.

E. We cannot determine any conclusion since we don't have a method for testing three means.

`We must use ANOVA to determine whether multiple means are the same.`



14. What is being tested in the graph above?

`Notice the title say 'Ratio of Two Variances' being equal to 1.`

A. $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$

`Not variances.`

*B. $H_0 : \sigma_1^2 = \sigma_2^2$ vs. $H_A : \sigma_1^2 \neq \sigma_2^2$

`If the ratio =` σ_1^2/σ_2^2 `= 1, then` $\sigma_1^2 = \sigma_2^2$.

C. $H_0 : \mu_1/\mu_2 = 1$ vs. $H_A : \mu_1/\mu_2 \neq 1$

`Not variances.`

D. $H_0 : \mu_1 - \mu_2 = 1$ vs. $H_A : \mu_1 - \mu_2 = 1$

`Not variances.`

E. $H_0 : \sigma_1^2 - \sigma_2^2 = 1$ vs. $H_A : \sigma_1^2 - \sigma_2^2 \neq 1$

`What would a difference of 1 tell us? We use ratios for variances. Remember variances are` *scale* `changes, so it makes sense to multiply or divide them.`

15. The paper "Undergraduate Marijuana Use and Anger" (*J. of Psychol.*(1988)) reported that a sample of 47 frequent marijuana users had a mean anger expression score of 42.72 with a standard deviation of 6.05. If we wanted to test whether frequent marijuana users have a higher mean anger expression score than the general public, which type of test should we do?

A. Case 1 using the general public's mean anger expression score and standard deviation as the hypothesized value and σ, respectively.

`This is plausible IF we know the general public's sd and believe it is the same for marijuana users.`

B. Case 2 using the general public's mean anger expression score as the hypothesized value and the standard deviation from our sample.

`This would be more likely than A since it is doubtful that the sd's are the same value if the means are not, BUT we have a large sample.`

C. Case 3 using the general public's mean anger expression score as the hypothesized value and our sample standard deviation as σ since our sample size is more than 30.

`This would be better than B since we have a larger sample. There really isn't any difference in` *StataQuest*, `only in the blue textbook.`

D. Case 6 using the proportion of the general public with anger expression scores more than 42.72 as the hypothesized value.

`What would we use as our sample proporiton? We are talking about averages,` μ`'s, not` π`'s.`

*E. Case 9 and take a large sample from the general public and calculate a mean and standard deviation.

`This is the best choice because we may find the general public's score is not as reported, and therefore we have a 'truer' test.`

16. Which of the following is the best interpretation of the power of a test of hypotheses?

Power is defined as $1 - \beta$, so it is how often a false null is detected(rejected).

A. The power of the test is the proportion of times you reject.

Since power is a probability, we need to have repeated samples PLUS we need to know that the null is false.

B. Under repeated sampling, the power of the test is the proportion of times you reject.

If the null were true, we would expect to reject about α% of the time.

*C. Under repeated sampling, the power of the test is the proportion of times you reject a false H_0.

D. Under repeated sampling, the power of the test is the proportion of times you reject a true H_0.

Since is the definition of α.

E. The power of the test measures how false H_0 is.

Power is, in a sense, a measure of 'falseness', but this is not the definition. 'Falsenes' would be determined by how different the null and true values are.

17. When should you use a significance level of 1% instead of 5%?

If a Type I error is more critical, use a small $\alpha = 0.01$. If a Type II error is more critical, use a larger $\alpha = 0.10$. Otherwise, use $\alpha = 0.05$.

*A. when you want to keep the chance of making a Type I error low.

B. when you want to keep the chance of making a Type II error low.

You would use a larger α here.

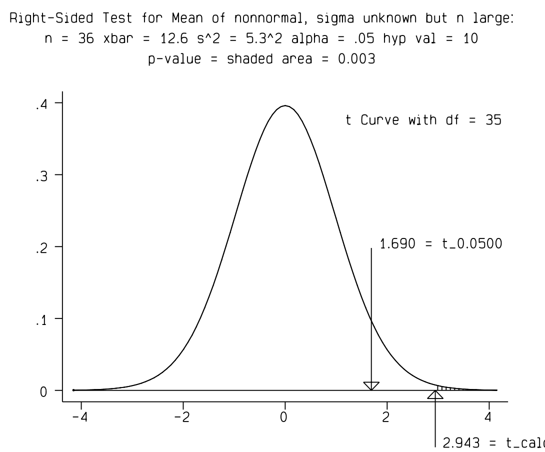
C. when you want to keep the chance of making a Type II error high.

Yes, this is a consequence, but you should NEVER *want* to make an error.

D. when you want to keep the power of the test high.

You would use a larger α here.

E. Exactly two of the above are good reasons.



18. What would be the result of a Type I error in the test represented above?

$H_0 : \mu = 10$ vs. $H_A : \mu > 10$ Type I error: reject a true null.

A. We would claim that the true mean was 12.6 when it really was only 10.

We cannot claim anything about the *sample* mean. This is sort of the right idea, though, since the null is true ($\mu = 10$), but we claimed something else (reject).

B. We would claim that the true mean was 10 when it really was 12.6.

This says we 'failed to reject', but the null is false.

*C. We would claim that the true mean was more than 10 when it really was only 10.

This says 'reject the null' even though it is true.

D. We would claim that the true mean was more than 10 when it really was 12.6.

This says 'reject the null', but it is false. This would have been a correct decision(although, we don't say anything about the sample mean so this isn't exactly correct).

E. We would claim that the true mean was not 10 when it really was 10.

`This would be a Type I error IF the alternative had been` $\neq$`!`

19. The lifetime of 60W GE light bulbs are known to follow an exponential distribution. GE wants to test its own claim that the average lifetime of its 60W light bulbs is at least 1000 hours, so they test 10 bulbs at random (they don't want to destroy too many) and calculate the mean and standard deviation of their sample. Which type of test should they perform?

`It says that the data is NOT normal! and they only took a small sample.`

A. Case 1 since they know the true standard deviation of the light bulb lifetimes.

`The data is not normal, plus it doesn't say they know` σ.

B. Case 2 since they have a small sample.

`Again, the data is not normal.`

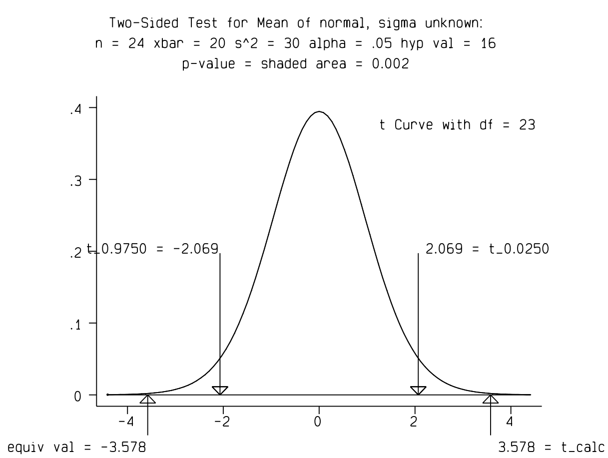
*C. A nonparametric test since they have a small sample.

D. Case 5 since the data is not normal.

`This is a test for a proportion, not a mean.`

E. Case 9 using the known mean and standard deviation as the second sample statistics.

`IF they had taken a large enough sample, they could have done this test.`



$H_0 : \mu = 16$ vs. $H_A : \mu \neq 16$. `Since the p-value is quite small, 0.002, we would reject` H_0 `and conclude the alternative is true.`

20. What conclusion can be made from the test above?

*A. The true mean is not 16.

B. The true mean is not 20.

`We cannot say anything about the sample mean. If we had used 20 as the hypothesized value, we would have had a p-value = 1 and so wouldn't have rejected.`

C. The true mean is 20.

`Again, we don't claim anything about the sample mean. Just because we reject 16 as a plausible value, we cannot say` $\mu = 20$. `It just happened to be one particular sample value.`

D. We have insufficient evidence to conclude that the true mean is 16.

`'Insufficient evidence' means we 'failed to reject', but we don't claim the null is true.`

E. We have insufficient evidence to conclude that the true mean is not 16.

`This would have been the correct conclusion had we gotten a large` $(> \alpha)$ `p-value.`

1A,2D,3A,4E,5B,6B,7E,8A,9C,10A,11D
12C,13D,14B,15E,16C,17A,18C,19C,20A