STAT303: 509–511
Spring 1999
Exam #2
Form A

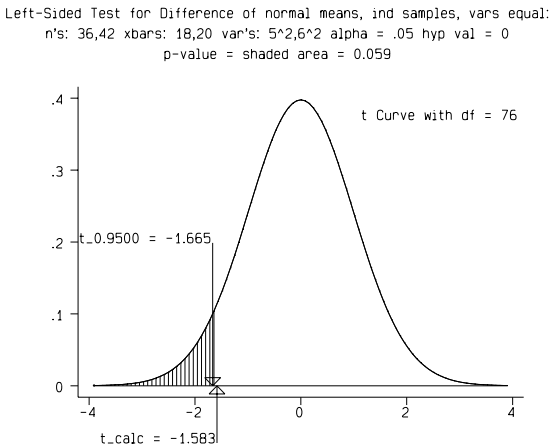
Instructor: Julie Hagen Carroll

Name:_____

1. **Don't even open this until you are told to do so.**
2. Be sure to write your instructor's name in the space provided and your name beneath.
3. There are 20 multiple-choice questions on this exam, each worth 5 points. There is partial credit. Please mark your answers **clearly** on the exam. Multiple marks will be counted wrong.
4. You will have 60 minutes to finish this exam.
5. 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.
6. Good luck!

1. Which of the following is FALSE?

   A. For a symmetric unimodal population, both the sample mean, $\bar{X}$, and the sample median, $\tilde{X}$, are unbiased statistics for estimating the population mean, μ.
   B. Unbiasedness guarantees that any value of the statistic will be close to the true value of the population even if the variance of the sampling distribution of the statistic is large.
   C. When the population distribution is symmetric with heavy tails compared to the normal curve (like the t curve with 3 degrees of freedom), the sample median, $\tilde{X}$ is better than the sample mean, $\bar{X}$, for estimating the population mean, μ.
   D. If the population distribution is normal, then $\bar{X}$ has a smaller variance than any other unbiased statistic for estimating μ.
   E. One might choose a biased statistic over an unbiased one if the bias of the first is small and its sampling distribution has a small variance.

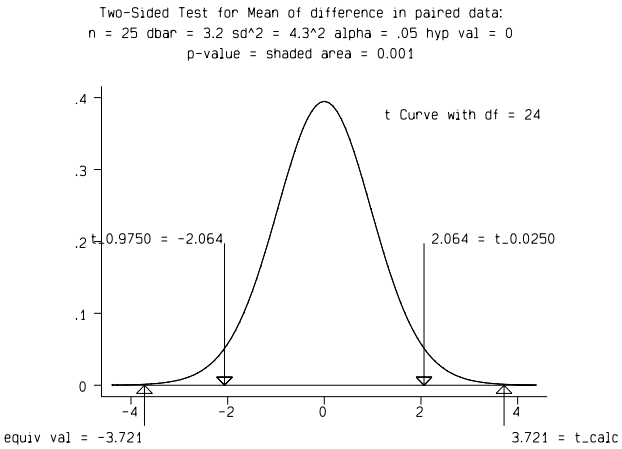


2. What is the correct conclusion for the output above?

   A. Reject H_0 at the 10% level, and conclude that μ_1 is less than μ_2.
   B. Reject H_0 at the 5% level, and conclude that μ_1 is less than μ_2.
   C. Fail to reject H_0 at the 1% level, and conclude that μ_1 is less than μ_2.
   D. Exactly two of the above are correct.
   E. All of the above (excluding D.) are correct.

3. In order to estimate the proportion of students at a particular university who favor the sale of beer on campus, a random sample of 100 students is selected. Of the selected students, 43 support the sale of beer. Let π denote the true proportion of the university's students who favor the sale of beer on campus. A 90% confidence interval for π based on this sample is 0.35, 0.51). Which of the following is the best interpretation of this interval?

   A. The probability that π is between 0.35 and 0.51 is 0.90.
   B. If we were to repeatedly sample 100 students from the university and calculate a new CI for each sample, then roughly 90% of these will contain 0.43.
   C. If we were to repeatedly sample 100 students from the university and calculate a new CI for each sample, then roughly 90% of these will contain π.
   D. If we were to repeatedly sample 100 students from the university and calculate a new CI for each sample, then roughly 90% of these will between 0.35 and 0.51.
   E. If we were to repeatedly sample 100 students from the university and calculate a new CI for each sample, then roughly 90% of these will contain p, the sample proportion.

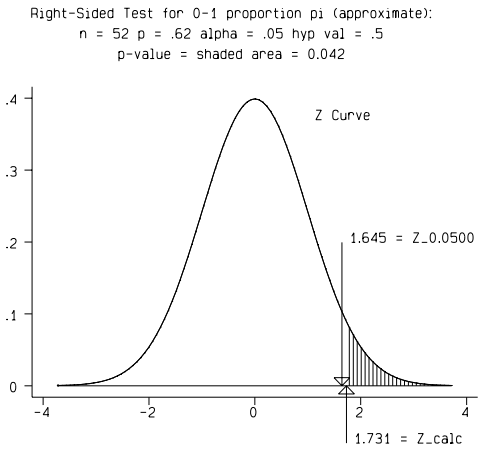


4. What are the correct hypotheses for the output above?

   A. $H_0: \ \mu_1 - \mu_2 = 0$ vs. $H_A: \ \mu_1 - \mu_2 \neq 0$
   B. $H_0: \ \mu_1 - \mu_2 \neq 0$ vs. $H_A: \ \mu_1 - \mu_2 = 0$
   C. $H_0: \ \mu_{diff} = 0$ vs. $H_A: \ \mu_{diff} \neq 0$
   D. $H_0: \ \pi_1 - \pi_2 = 0$ vs. $H_A: \ \pi_1 - \pi_2 \neq 0$
   E. Both A. and C. are correct.

5. What is the approximate distribution of the sample proportion of successes, $p = X/n$, if $X \sim Bin(n = 40, \pi = 0.7)$, *i.e.*, the sample size, $n = 40$, and the true proportion of successes, $\pi = 0.7$?

   A. $p \sim N(0.7, 0.072^2 = 0.00525)$
   B. $p \sim N(40, 0.7)$
   C. $p \sim N(0.7, 0.7^2)$
   D. You can't say p is normal since the original data is discrete.
   E. You can't say p is normal since $n(1-\pi) < 5$.



6. What is the consequence of making a Type II error in the output above?

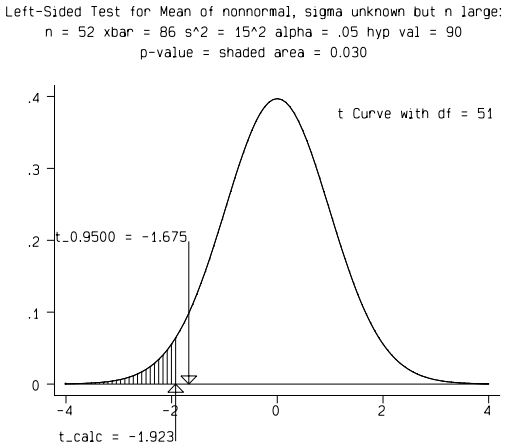
   A. We would conclude that the true mean is greater than 0.5 when it really is not.
   B. We would conclude that there is not enough evidence to conclude that the true mean is greater than 0.5 even though it really is greater than 0.5.
   C. We would conclude that the true proportion is greater than 0.5 when it really is not.
   D. We would conclude that the true mean is not 0.5 when it really is.
   E. We would conclude that there is not enough evidence to conclude that the true proportion is greater than 0.5 even though it really is greater than 0.5.

7. What is $P(\overline{X}_5 < 15)$ if $\overline{X}_5 \sim N(25, 10^2)$?

   A. -2.24
   B. 0.0125
   C. 0.1587
   D. 0.9875
   E. It's not normal since n is only 5.

8. What is the purpose of *random sampling*?

   A. It keeps anyone from knowing the winning lottery numbers beforehand.
   B. It reduces the variability between samples.
   C. It keeps the sample from being biased.
   D. All of the above.
   E. Exactly two of the above are true.



9. What does the $p - value$ tell us in the output above?

   A. There is a 3% chance of making a Type I error.
   B. There is a 3% chance of making a Type II error.
   C. 3% of the time we will make a Type II error.
   D. 3% of the time we will get a sample mean of 86 or less even though the true population mean is 90.
   E. 3% of the time we will get a sample mean of 90 or less even though the true population mean is 86.

10. Why is a *paired* t-test (Case 10) preferred over either 2-sample t-tests (Cases 8 and 9)?

    A. A paired t-test uses half as much data.
    B. A paired t-test has more power since it uses a smaller standard deviation.
    C. A paired t-test is more conservative.
    D. A paired t-test is not preferred since it doesn't use independent samples.
    E. Exactly two of the above are correct reasons.

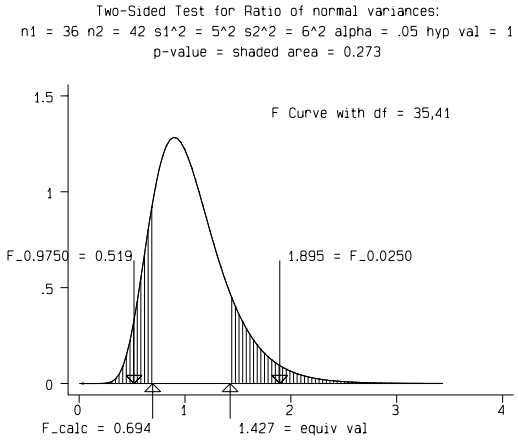
11. What do we mean when we say we reject H_0 at the 5% level?

    A. We are admitting that we will reject H_0 5% of the time.

B. We know that H_0 will be true 5% of the time.
C. The probability of getting our data (or worse) will happen less than 5% of the time even though H_0 is true.
D. All of the above are true.
E. Exactly two of the above are true (excluding D.).

12. According to the Central Limit Theorem, if you take many samples of size n from a given population and compute the sample mean, $\bar{X}$, for each sample, then the resulting distribution of $\bar{X}$ will

A. look like the distribution of the parent (original) population, provided n is large enough.
B. look like a normal distribution, provided n is large enough.
C. have the same variance as the parent population no matter what the value of n is.
D. have the same mean as the parent population no matter what the value of n is.
E. have the same variance as the parent population, provided n is large enough.



13. Case 8 in *StataQuest* requires the assumption that the variances are equal. If the output above was used to determine whether this assumption was valid, which of the following is true?

A. This output can't be used to validate the assumption since Case 8 is a test for means.
B. The assumption is valid since we reject H_0.
C. The assumption is invalid since we reject H_0.
D. The assumption is valid since we fail to reject H_0.
E. The assumption is invalid since we fail to reject H_0.

14. Many consumers pay careful attention to stated nutritional contents on packaged foods when making purchases. It is therefore important that the information on the packages be accurate. A random sample of $n = 12$ frozen dinners of a certain type was selected from production during a particular period, and the average caloric content and standard deviation was determined. What is the appropriate test to use to determine whether the true mean caloric content is equivalent to the advertised amount?

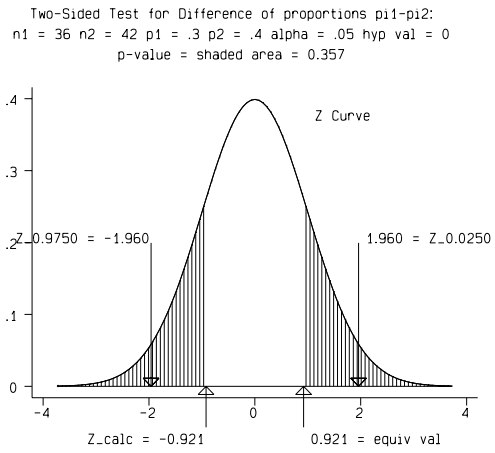
A. Case 1: a small sample z test
B. Case 2: a small sample t test
C. Case 3: a large sample z test (*StataQuest* actually does a t test)
D. Case 6: a one-sample test for proportion
E. A non-parametric test since the sample size is small.

15. Previous to last week's election, Student Gov't wanted to estimate what percent of the student body was going to vote. This percentage is fairly consistent from year to year, around 80%, but they wanted an accurate estimate for this year. They decided they wanted a 99% confidence interval with a margin of error of only 1%, *i.e.*, they wanted it to within 1%. What size sample should they use?

A. 258
B. 10651
C. 16641
D. 10650
E. 16640

16. Suppose we perform a hypothesis test at the 5% level. Which of the following statements is/are correct?

A. There is a 5% chance that our decision is correct.
B. There is a 5% chance that we made a Type I error.
C. Under repeated sampling and performing the same test on the same population, we would reject H_0 approximately 5% of the time.
D. Under repeated sampling and performing the same test on the same population, we would make a Type I error approximately 5% of the time.
E. Exactly two of the above are correct.



17. Using the data represented above, which of the following is/are correct?

    A. 0 would be in a 90% confidence interval for the difference of two population proportions.
    B. 0.358 would be in a 90% confidence interval for the difference of two population proportions.
    C. π_1 and π_2 would both be in a 90% confidence interval for the difference of two population proportions.
    D. p_1 and p_2 would both be in a 90% confidence interval for the difference of two population proportions.
    E. Both C. and D. are correct statements.

```
Var|Obs Mean  Std. Err.  [Conf.Interval]
---+-------------------------------------
90%| 33   58  1.392621  55.64105 60.35895
95%| 33   58  1.392621  55.16332 60.83668
99%| 33   58  1.392621  54.18633 61.81367
```

18. What is the best statement for the range of the p-value for testing $H_0 : \mu = 55$ vs. $H_A : \mu \neq 55$?

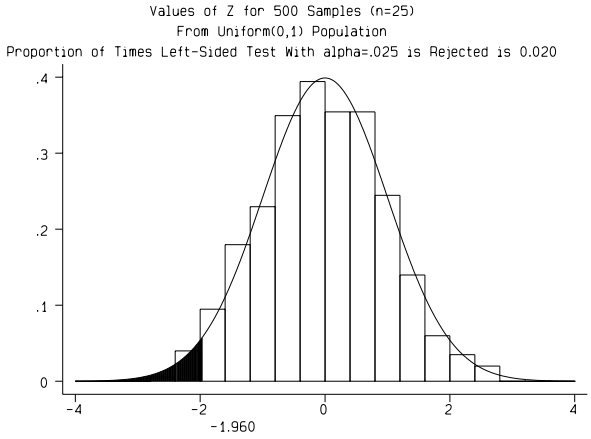
    A. p-value > 0.10
    B. $0.10 >$ p-value > 0.05
    C. $0.05 >$ p-value > 0.01
    D. $0.01 >$ p-value
    E. You can't determine the p-value without doing a hypothesis test.

19. Suppose a wholesale electronics company conducted a test for the true proportion of defective walkmans. At a significance level, $\alpha = 0.05$, they rejected the null hypothesis. If they had conducted a 10% significance level test instead, what would have been their conclusion? Select the best answer.

    A. Since the data wouldn't change, the conclusion wouldn't change.
    B. There is no way of knowing without rerunning the test.
    C. We would fail to reject the null since the p-value is less than 0.05.
    D. We would reject the null since the p-value is less than 0.05.
    E. They would have made a Type I error.

20. What is $P(p_{30} < 0.15)$ if $p_{30} \sim N(0.25, 0.079^2)$?

    A. We can't be sure p_{30} is normally distributed since $\pi \neq 0.5$.
    B. -1.26
    C. 0.1038
    D. 0.8962
    E. 0 since the z-score $= -6.94$



21. The 'Proportion of Times Left-Sided Test with alpha =.025 is Rejected' above is an estimate of

    A. the power of this test.
    B. the α-level for this test.
    C. the β for this test.
    D. All of the above are true.
    E. None of the above are true.

1B,2A,3C,4E,5A,6E,7C,8E,9D,10B,11C,
12B,13D,14E,15B,16D,17A,18C,19D,20C,21B