STAT303
Fall 2013
Exam #3
Form A

Instructor: Julie Hagen Carroll

November 19, 2013

1. **Don't even open this until you are told to do so.**
2. Remember to turn your phone off now.
3. Please turn your hats around backwards or take them off.
4. Please put your backpack and other things along the walls or at the front of the room.
5. You need a gray, 81/2 $\times$ 11" scantron, pencil, calculator and you may have 5 sheets of notes.
6. There are 20 multiple-choice questions on this exam, each worth 5 points. There is partial credit. Please mark your answers **clearly**. Multiple marks will be counted wrong.
7. You will have 60 minutes to finish this exam.
8. If you have questions, please write out what you are thinking on the back of the page so that we can discuss it after I return it to you.
9. 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.
10. When you are finished please make sure you have filled in your name and marked your FORM (A, B, C or D) and 20 answers, then turn in JUST your scantron.
11. Good luck!

1. If H_0 is false, some possible reasons for failing to reject are

   A. the difference between the hypothesized mean and the sample mean is not big enough to detect. The two means are too close to be able to tell a difference.
   B. the sample size is too small.
   C. the standard deviation is too large. This makes the curves too wide, thus overlap.
   D. All of the above could cause a Type II error.
   E. If H_0 is false, we have to reject.

2. If $X \sim N(15, 8^2)$, how likely are you to get a *sample mean* less than 12 if the sample size is 25?

   Xbar ~ N(15, (8/sqrt(25))^2), sigma = 1.6

   P(Xbar < 12) = P(Z < (12-15)/1.6) = P(Z < -1.88)
   Look it up on the negative page of the Z table.
   The main point is that you need the sigma of xbar.

   A. 0.3520
   B. 0.0301
   C. 0.32
   D. 0.6480
   E. 0

3. Which of the following is true?

   A. If $\alpha = 0.05$, we will reject only 5% of the time. ONLY if Ho is true. We hope to reject all the time. It's what we're trying to do with a hypo test.
   B. One-sided tests have more power than 2-sided tests because there doesn't need to be as big of a difference in means to be rejected. The pv for a 2-sided test is twice that of a 1-sided which makes it harder to reject.
   C. If the hypothesized value falls within a 95% confidence interval for the true mean, then it must fall within the 90% (using the same data). 95% CI's are wider, so they contain more values.
   D. All of the above are true.
   E. Only two of the above are true.

```
90% CI:  (0.34964, 2.95036)    out
95% CI:  (0.08136, 3.21864)    out
99% CI:  (-0.47572, 3.77572)   in
```

2 sample CI's are for the difference in means. If mu1 = mu2, then mu1 -mu2 = 0, so 0 is the number to look for in the intervals.

4. What is the correct range of the p-value for testing $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$ given the three confidence intervals for the difference of the true means above?

   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. We can't use confidence intervals here because we don't have a hypothesized value to compare.

5. Using the same scenario, which of the following is the *best* interpretation of the power of the test above?

   Power is how often we correctly reject (a false) Ho and conclude the alternative is true.

   A. Power is the probability that we correctly conclude $\mu_1 \neq \mu_2$.
   B. Power is the probability that we ~~in~~correctly conclude $\mu_1 \neq \mu_2$.
   C. Power is the probability that we correctly conclude ~~$\mu_1 = \mu_2$~~.
   D. Power is the probability that we ~~in~~correctly conclude $\mu_1 = \mu_2$.
   E. Confidence intervals don't have power.

6. Which of the following is true about the sampling distribution of the sample mean, $\overline{X}$?

   A. The shape is always dependent on the sample size. If the parent population is normal, the sample size doesn't matter.
   B. The center is always dependent on the sample size. Never!
   C. The spread is always dependent on the sample size. sigma(xbar) = sigma(x)/sqrt(n)
   D. All of the above are true.
   E. Two of the above are true.

7. The manager of an automobile dealership is considering a new bonus plan to increase sales. Currently, the mean sales rate per salesperson is five automobiles per month. Let μ_{old} be the true mean before the bonus plan, μ_{bonus} the true mean with the bonus plan, and μ_d the true mean difference paired by salesman. The correct set of hypotheses to test the effect of the bonus plan is

   Because we have data 'by salesman', we should run a paired test.
   If the true paired difference is greater than 0, then the plan worked.
   Pairing allows for the best and worse salesmen to show improvement even though their number of sales is quite different.

   A. $H_0 : \mu_{bonus} = 5$ vs. $H_A : \mu_{bonus} > 5$
   B. $H_0 : \mu_d = 0$ vs. $H_A : \mu_d > 0$
   C. $H_0 : \mu_d = 5$ vs. $H_A : \mu_d > 5$
   D. $H_0 : \mu_{bonus} = \mu_{old}$ vs. $H_A : \mu_{bonus} > \mu_{old}$
   E. $H_0 : \mu_{bonus} = \mu_{old}$ vs. $H_A : \mu_{bonus} \neq \mu_{old}$

8. A researcher wants to know if tougher sentencing laws have had a positive effect in terms of deterring crime. He plans to select a sample of states which have enacted a "3 strikes" law and compare violent crime rates before the law was enacted and two years later. The 1-sample t statistic from a sample of $n = 19$ observations for the 1-sided test has the value $t = {}^{-}1.93$. Based on this information

   A. We would reject at the 5 and 10% levels and conclude that tougher sentencing laws have had a positive effect in terms of deterring crime.
   B. We would ~~fail to reject~~ at the 1% level and conclude that tougher sentencing laws have had a negative effect in terms of deterring crime.
   C. We would fail to reject at the 1% level but could only say that tougher sentencing laws have not had a positive effect in terms of deterring crime.
   D. A and B are correct.
   E. A and C are correct.

H0: no effect on crime Ha: laws have a positive effect
crime rate >= crime rate <

df = 18	0.688	0.862	1.067	1.330	1.734	2.101

-1.93
0.05 > pv > 0.025

reject at 5 and 10%
FTR at 1%

C is almost correct, but the correct statement is
we cannot prove that the laws have a positive effect.
It could be that we just didn't have conclusive evidence even though
the laws did make a difference.

9. I have a 95% confidence interval for μ the true mean of (2.786, 9.352). Which of the following are plausible statements?

We can only hypothesize what the mean is not (or more or less).

A. At the 5% level of significance, I can conclude that the true mean is between ~~2.786 and 9.352.~~

Since 10 is not in a 95% CI, it is not in the narrower, 90% CI, so we would reject.

(B.) At the 10% level of significance, I can conclude that the true mean is not 10.

We don't know if 2.7 would be in a 99% CI.

C. A 99% confidence interval would not include 2.7 because it's not in the 95% interval.

6.069 is the sample mean because it is the center of the interval and would be in any interval.

D. No conclusion can be made about 6.069 at the 10% level because I don't know if 6.069 would be in the 90% confidence interval or not.

E. More than one of the above are true.

10. What would be a Type I error in the previous example using (2.786, 9.352) as a 95% confidence interval for the true mean? reject a true Ho

Ho true = mean is 10, Ho false = mean is NOT 10
reject = claim it's not 10, FTR = can't prove it's not 10.
We can never say IS because that would be saying H0 is true.

A. claiming that 10 ~~is~~ in the interval when it's not

B. claiming that the true mean ~~is~~ 10 when it's not

(C.) claiming that the true mean is not 10 when it is 10 (reject / Ho false)

D. claiming that 6.069 is the true mean when it's not

(E.) failing to capture 10 in the interval when it should be in the interval

This actually works too. 'failing to capture' means rejecting with this data and if it should be in the interval, Ho is true.

11. A researcher wants to know if calcium is an effective treatment for lowering blood pressure. He assigns one randomly chosen group of subjects to take calcium supplements; the other group will get placebo. At the end of the treatment period, he measures the difference in blood pressure. The 50 members of the calcium group had blood pressure lowered by an average of 25 points with standard deviation 10 points. The 50 members of the placebo group had blood pressure lowered by an average of 15 points with standard deviation 8 points. Which of the following is true?

A. The true difference in means is 10 points.

B. They should have run a matched pairs test with this data instead of a 2-sample t-test.

C. If we can assume the standard deviations are close enough to be pooled, the correct degrees of freedom are 99.

D. We need to know that neither sample has outliers.

(E.) None of the above are true.

n1 = 50, xbar1 = 25, s1 = 10, n2 = 50, xbar2 = 15, s2 = 8
The SAMPLE mean difference is 25 - 15 = 10. We don't know what the true mean difference is; we can only hypothesize.
This can't be a matched pairs test because the data is independent. There is no link between members of the 2 groups.
The degrees of freedom for a 2-sample t-test is the minimum of n1 - 1 and n2 - 1, so 49. The df for a pooled t test is n1 + n2 - 2 = 98.
Since both samples are well over 30, outliers shouldn't be a problem.

12. The manager at a movie theater would like to estimate the true mean amount of money spent by customers on popcorn only. He selects a simple random sample of 36 receipts and calculates a 92% confidence interval for true mean to be ($12.45, $23.32). The confidence interval can be interpreted to mean that, in the long run:

(A.) 92% of similarly constructed intervals would contain the population mean.

B. 92% of similarly constructed intervals would contain the ~~sample~~ mean. true mean

C. 92% of all customers who buy popcorn spend between ~~$12.45 and $23.22~~.

D. 92% of the time the true mean will be between ~~$12.45 and $23.22.~~

Never these numbers unless we use the word 'confident'.

E. Popcorn is way too expensive at this theater.

13. The water diet requires you to drink two cups of water every half hour from when you get up until you go to bed, but eat anything you want. Four adult volunteers agreed to test this diet. They are weighed prior to beginning the diet and six weeks after. Their weights in pounds are:

Person	1	2	3	4	mean	s.d.
Before	180	125	240	150	173.75	49.56
After	170	130	215	152	166.75	36.09
Diff	10	-5	25	-2	7	13.64

The last row gives the mean and standard deviation of the four differences, (not the differences of the two means and standard deviation) so you don't need to calculate them, just the test statistic. What are the correct degrees of freedom and p-value for testing whether the diet worked or not?

A. df = 7, $0.40 > \text{pv} > 0.30$

B. df = 7, $0.20 > \text{pv} > 0.15$

C. df = 3, $0.40 > \text{pv} > 0.30$

(D.) df = 3, $0.20 > \text{pv} > 0.15$

E. df = 6, $0.40 > \text{pv} > 0.30$

The degrees of freedom for a paired test is the number of differences -1 = 4 - 1 = 3.
'whether the diet worked or not' is NOT a two-sided test. For the diet to work, people need to lose weight, so this is a less than test.
t = (7 - 0)/(13.64/sqrt(4)) = 1.026, df = 3

14. What would be the consequence of a Type II error in the previous test?

failing to reject a false Ho
Ho true = people don't lose weight, Ho false = people do lose
reject = claim diet works, FTR = claim to prove diet works

A. The diet was approved when it really didn't help anyone lose weight. (reject; Ho true --> Type I)

(B.) The diet was deemed a failure (it didn't work) when it actually did help people lose weight. FTR false Ho

C. The diet didn't work for these volunteers, but it should have. (Ho false)

Claims are made to the whole population, not just this sample.

D. The diet actually works, but these volunteers gain 7 pounds on average.

E. Drinking water is good for you, but it doesn't help you lose weight.

15. Which of the following tests would have the most power if the true mean is 15?

To maximize power, use the largest alpha possible, the largest sample size affordable, and choose the hypothesis that's easiest to prove.

If the true mean is 15, it's more likely to get a sample mean greater than 10 (than 12).

A. $H_0 : \mu = 10$ vs. $H_A : \mu \neq 10$, $n = 25$, $\alpha = 0.05$
B. $H_0 : \mu \leq 10$ vs. $H_A : \mu > 10$, $n = 25$, $\alpha = 0.05$
C. $H_0 : \mu \leq 12$ vs. $H_A : \mu > 12$, $n = 25$, $\alpha = 0.05$
D. $H_0 : \mu \geq 12$ vs. $H_A : \mu > 10$, $n = 25$, $\alpha = 0.10$
E. $H_0 : \mu \geq 12$ vs. $H_A : \mu > 12$, $n = 25$, $\alpha = 0.10$

Skip the Ho's. Bad copy job!

16. Which of the following is always true in reference to hypothesis testing using a 1-sample t-test?

A. the sample must be random
B. the data must be normal
C. the purpose is to prove the null false
D. All of the above are true.
E. Only two of the above are true.

If the sample size is large enough the sampling population does not have to be normal.

17. Suppose we sample from a population, $X \sim N(35, 9^2)$, 50 times. We reject 40 of those 50 times. What's the best explanation of what happened?

A. There must have been a sampling error in the 10 fail to rejects.
B. We must have been testing something other than $\mu = 35$.
C. The approximate power of the test is 80%.
D. The α level is too high. It's 20%.
E. There is no explanation for what happened.

There is always the possibility of an error, but it's not due to sampling.

Rejecting 40/50 times implies Ho is false. This means the rejections are the correct decisions; therefore, 40/50 = 80% is the estimated power of the test.

18. Suppose we test $H_0 : \mu = 30$ vs. $H_A : \mu < 30$. Our sample mean is 28 and our resulting p-value is 0.043. Which of the following is true based on the same data?

Reject < 0.05

Since the 0.01 < pv < 0.05 < 0.10, we would reject at the 5 and 10% levels but not at the 1%.

A. 28 would be in a 99% interval but not in the 90 or 95%.
B. 30 would be in a 90% interval but not in the 95 and 99%.
C. At the 1% level, we could say the true mean is not less than 30.
D. At the 1% level, we could say the true mean is not less than 28.
E. None of the above are correct.

28 is the sample mean, so it would be the center of any interval created from this data.

This is backwards. 30 would be in the 95 and 99% CI's, but not in a 90% CI.

The pv for not = would be 0.086, so the hypothesized value = 30 would be rejected only at the 10% level.

19. What does the p-value above tell us?

A. how often we would get a mean of 30 when the true mean is 28
B. how often we would get a mean of 30 or less when the true mean is 28
C. how often we would get a mean of 28 or less when the true mean was at least 30
D. how often we would get a mean of at least 28 when the true mean was 30 or less
E. how often we would get a mean of at least 30 when the true mean was 28 or less

sample data or more like Ha when Ho is true

20. Matched pairs t-test

A. is the same as taking the difference of two 1-sample t-tests.
B. has twice the power of a 1-sample t-test.
C. has more power than a 2-sample t-test because it uses twice as much data.
D. Two of the above are true.
E. None of the above are true.

Matched pairs is the proper test for dependent data. It has more power because one source of variability, the difference between subjects, is eliminated. There is the same amount of data as in a 2-sample test (2 groups), but the groups are related, thus paired.

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