STAT 350 MidTerm 2 (11/5/2013) Fall 2013

Section (circle): 9:30 am 11:30 am 3:30 pm

Name: __Key______________________________________

INSTRUCTIONS

1. Calculators (using the programmable features will not help you since ALL work including integration steps are required to receive full credit) and your cheat sheet are allowed. Exam is closed book/closed notes.
2. There are 5 pages front and back for the exam in addition to the cover sheet. You may tear off the last page (tables) to make your work easier. However, this sheet must be returned.
3. All numeric answers must have at least 3 decimal places.
4. Work is required to receive credit. Partial credit will be given for work that is partially correct. Points will be deducted for incorrect work even if the final answer is correct.
5. If I cannot read your work, it will be marked wrong.
6. Please attach your cheat sheet, table and any scrap paper used to the exam.
7. Good Luck!

Question	Possible	Score
1	18	
2	6	
3	22	
4	32	
5	10	
6	14	
7	7	
Total	109	

(18 pts.) 1. Multiple choice questions. Please circle your answer. There is only one correct answer for each choice. (2 pts. each)

a) Suppose we are testing the hypothesis H_0: $\mu = 0.5$ vs H_a: $\mu \neq 0.5$ and the only information we are given is a sample size of n = 60. What would be the type I error when testing the above hypothesis?

A) You claim that the true mean is 0.5 when it really is 0.5.
B) You claim that the true mean is 0.5 when it is really not 0.5.
C) You claim that the true mean is not 0.5 when it really is 0.5.
D) You claim that the true mean is not 0.5 when it is really not 0.5.

b) Suppose that we do not have a random sample, the results from our confidence interval and hypotheses tests are still good as long as we have a large n so that the Central Limit Theorem applies.

A) True   B) False

c) If two events are mutually exclusive they can also be independent.

A) True   B) False

d) Let two confidence intervals (15, 18) and (14, 19) be calculated from the same data. The interval (15,18) has the lower confidence level because it is narrower.

A) True   B) False

e) If our decision is that we fail to reject the null hypothesis, it is because the null hypothesis is true.

A) True   B) False

f) In a test of statistical hypotheses, the P-value tells us:
A) if the null hypothesis is true.   B) if the alternative hypothesis is true.
C) the largest level of significance at which the null hypothesis can be rejected.
D) the smallest level of significance at which the null hypothesis can be rejected.

g) A 95% confidence interval for the mean μ of a population is computed from a random sample and found to be 9 ± 3. We may conclude:

A) there is a 95% probability that μ is between 6 and 12.
B) there is a 95% probability that the true mean is 9 and there is a 95% chance that the true margin of error is 3.
C) if we took many, many additional random samples and from each computed a 95% confidence interval for μ, approximately 95% of these intervals would contain μ.
D) all of the above.

h) An engineer knows that 12% of the power supplies in his lab are defective. Let X = the number of failed power supplies that she has to try to find the second good power supply. X follows a ________ distribution.

A) binomial   B) Poisson   C) neither binomial nor Poisson

i) Which of the following codes could be used to calculate a lower confidence bound with a confidence level of 90%?

SAS

A)
```
proc ttest data=a SIDES=1;
  var Mean;
  run;
```

B)
```
proc ttest data=a SIDES=U alpha=0.1;
  var Mean;
  run
```

C)
```
proc ttest data=a SIDES=U;
  var Mean;
  run;
```

D)
```
proc ttest data=a SIDES=1 level=90;
  var Mean;
  run;
```

MiniTab

A) Stat → Basic Statistics → 1-Sample t → (Samples in columns: Mean → Options (Confidence level: 0.1, Alternative: greater than)

B) Stat → Basic Statistics → 1-Sample t → (Samples in columns: Mean → Options (Confidence level: 90, Alternative: greater than)

C) Stat → Basic Statistics → 1-Sample t → (Samples in columns: Mean)

D) Stat → Basic Statistics → 1-Sample t → (Samples in columns: Mean → Options (Significance level: 0.1, Alternative: greater than)

(6 pts.) 2. Please fill in the following table. Each correct answer is worth 1 pt.
Joey owns a race track for radio controlled cars. He is interested in certain aspects of his business and proceeds to collect data to answer the questions he has. For each part below identify whether it involves 1 sample or 2 sample independent (2I) or 2 sample paired (2P) and whether the distribution is a z or t.

	1 or 2I or 2P	z or t
a) Eight customers race both a truck and a buggy. For each he measures lap times for both their truck and buggy. He wants to know if the buggies are faster than the trucks if the same person is driving them. It is determined that s is 2.4	2P	t
b) He measures the lap time for all 22 four-wheel drive cars. He would like to know if their average lap time is less than 19 seconds. It is known that the population standard deviation is 2.1 seconds.	1	z
c) He has a total of 28 racers; 14 buggy racers and 14 truck racers. For each racer he records how far they traveled to visit his track. He wants to know if buggy racers are willing to travel farther than truck racers. From the data, it is determined that s is 12.2 miles.	2U	t

(22 pts.) 3. The following data was obtained by determining the cylinder strength for various types of columns. Each type of column was cured under moist conditions (M) and laboratory drying conditions (LD). The strengths of the columns depend on the structural material that they are composed of in addition to the curing time.

Type	1	2	3	4	5	6	7	8	9	10	$\bar{x}$	s
M	82.6	87.1	89.5	88.8	94.3	80.0	86.7	92.5	97.8	90.4	88.97	5.28
LD	86.9	87.3	92.0	89.3	91.4	85.9	89.4	97.8	94.3	92.0	90.63	3.65
M - LD	-4.3	-0.2	-2.5	-0.5	2.9	-5.9	-2.7	-5.3	3.5	-1.6	-1.66	3.18

(5 pts.) a) Which method should be used in this problem, paired or independent? Please explain your answer.

Paired should be used because I stated in the description that there is an extraneous variable, the structure. This situation has two-samples, therefore, every column needs to be measured via the two curing times. The two samples cannot be the same cylinder because each cylinder can only be cured once. Stating that the data is organized as paired is not a valid reason because you need to decide BEFORE you perform the experiment whether paired or independent is appropriate.

(7 pts.) b) Calculate the 99% lower bound for this data. The Satterwaith df = 16.002 . No interpretation is needed.

$$\mu > \bar{x} - t^*_{column}(df)\frac{s_D}{\sqrt{n}} = -1.66 - (2.821)\frac{3.18}{\sqrt{10}} = -4.497$$

$t_{column}*(df) = t*_{0.01}(10-1) = 2.824$
this is a bound therefore the column = α

Note: I gave you the df for independent, that did not mean that you needed to use it.

(5 pts.) c) What assumptions did you make to perform you analysis in b)? How would you test these assumptions?

1) SRS
2) the differences are normal. To test this, generate a normal quantile plot and histogram on the differences.

I did not take off if you did not include the word ‘differences’.

(5 pts.) d) Without performing the hypothesis test, what would the decision be (Step 6)? Please explain your answer.

If you performed the hypothesis test, you did not get credit for this part. If you generated the confidence interval, you lost points because the hypothesis needs to be compared to what you have already generated.

The answer is fail to reject because 0 is greater than -4.497.

You had to mention that you are comparing to $\Delta = 0$. Note: $\bar{x}$ has to be greater than the lower bound because of the equation. However, we know what the sample is stating, all inference is concerned with the populations.

(32 pts.) 4. A manufacturer of a certain type of hydrogenated vegetable oil claims that the melting point is 95 °F. To check this assumption, a customer took 50 samples of this oil and determined that the average melting point was $\bar{x}$ = 94.51 and the population standard deviation was σ = 1.20.

(7 pts.) a) Construct a 99% CI for this data.

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 94.51 \pm (2.576)\frac{1.20}{\sqrt{50}} = 94.51 \pm 0.437 \Longrightarrow (94.073, 94.947)$$

The column to use for z* is 0.005 = 0.001/2

(5 pts.) b) Interpret the interval.

We are 99% confident that the true population mean melting point is between 94.073 °F and 94.947°F.

Note: I gave you this format; if you made a comment about in the long run 99% of the intervals contain μ, you did not get credit.

(15 pts.) c) Perform the appropriate hypothesis test with the appropriate α (be sure to include the 7 steps).

1. μ = population mean melting point.
2. H_o: μ = 95 H_a: μ ≠ 95
3. α = 0.01
4. $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{94.51 - 95}{1.20/\sqrt{50}} = -2.887$
5. P = 2(P(Z > |-2.887|) = 2(1 – P(Z ≤ 2.887)) = 2(1 – 0.9981) = 2(0.0019) = 0.0038
 = 2(P(Z < -2.887) = 2(0.0019) = 0.0038
6. reject H_0 because 0.0038 ≤ 0.01
7. The data does provide strong support (P = 0.0038) to the claim that the population mean melting point is not 95°F.

(5 pts.) d) What sample size would be required for the margin of error of the confidence interval to be 0.1 °F?

$$n = \left(\frac{z^*\sigma}{m}\right)^2 = \left(\frac{(2.576)(1.20)}{0.1}\right)^2 = 30.912^2 = 955.552 \Longrightarrow 956$$

Note I graded on consistency on this question. If you made a mistake in a previous part, I graded you correct if you used the same values.

(10 pts.) 5. Suppose that 17% of students on a campus are mathematics majors and 19% of the students are computer science majors. In addition, 5% of the students have a dual major in both mathematics and computer science.

Abbreviations: M: Math, CS: Computer Science, F: Female

(5 pts.) a) What is the probability that a particular student is a mathematics major or a computer science major.

P(M) = 0.17, P(CS) = 0.19, P(M and CS) = 0.05
P(M or CS) = P(M) + P(CS) – P(M and CS) = 0.17 + 0.19 – 0.05 = 0.31
Remember what 'or' means, it means at least one of A or B.

(5 pts.) b) In addition, 49% of the students are female and 8% of the students are mathematics majors and female. Given that a student is a mathematics major, what is the probability that the student is female?

P(F) = 0.49 (not used), P(M and F) = 0.08

$$P(F|M) = \frac{P(F \text{ and } M)}{P(M)} = \frac{0.08}{0.17} = 0.471$$

(14 pts.) 6. The College Board reports that 2% of the High School students who take the SAT each year receive special accommodations because of documented disabilities. Assume that we have a random sample of 25 students. Note: This is a binomial distribution.

(9 pts.) a) What is the probability that at least two of the students receive special accommodations?

I stated that this was a binomial so points were taken off if you used an approximation.
n = 25, p = 0.02

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1)$$
$$= 1 - \binom{25}{0} 0.02^0 (1 - 0.02)^{25} - \binom{25}{1} 0.02^1 (1 - 0.02)^{24} = 1 - 0.603 - 0.308 = 0.089$$

I did take off points here if you had less than 3 decimal places.

(5 pts.) b) What are the mean and the standard deviation of the number of students that receive special accommodations?

E(X) = np = (25)(0.02) = 0.5

$$\sigma = \sqrt{np(1-p)} = \sqrt{0.5(0.98)} = 0.7$$

(7 pts.) 7. The number of hours that it rains at the Toronto airport has an exponential distribution with a mean of 3. What is the probability that it will rain between 2 and 3 hours on one particular night? Hint: What is the expected value for an exponential distribution?

$$E(X) = \frac{1}{\lambda} \Longrightarrow \lambda = \frac{1}{3}, f(x) = \lambda e^{-\lambda x}$$
$$P(2 < X < 3) = \int_2^2 \frac{1}{3} e^{-x/3} dx = -e^{-x/3}\Big|_2^3 = -e^{-1} + e^{-\frac{2}{3}} = 0.146$$