

**Midterm 3 - STAT 301 Section 501/502 version 1
Spring 2012**

Name:

UIN:

Signature:

1. Do not open this test until told to do so.
2. Turn in your exam with your answers circled when you are done with the exam. You should not take the exam with you.
3. This is a closed book examination. You may use three cheat sheets, that you have brought with you, and the tables. You should have no other printed or written material with you on the exam.
4. You have 50 minutes to work on this exam. There are 6 questions.
5. You may use a calculator but not a phone during the exam.
6. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or teaching assistant for clarification.
7. Good Luck!!!

(1) Your aim is to design an experiment and collect data and construct a 95% confidence interval for the mean. You know that the standard deviation of a population is $\sigma = 3$.

(i) Your colleagues ask you to briefly explain what the confidence interval tells you about the mean. Explain this in one sentence. [1]

(ii) What sample size should you use such that the margin of error is at most 0.5? [1]

(iii) Suppose you were incorrectly told that the standard deviation was $\sigma = 3$, and in fact the standard deviation of the population is $\sigma = 5$. What effect would this have on the margin of error of the CI for the sample size calculated in (ii). [1]

(2) Fred wants to use statistics to ‘prove’ his research. As he learnt in his statistics class, he stated his research hypothesis as the alternative and the opposite of that as his null hypothesis. After collecting the data, the test at the 5% level and is unable to reject the null. Disappointed with this result, he goes and collects another sample does the test at the 5% level and again is unable to reject the null. He does this 5 times and on the fifth attempt rejects the null. He then writes a paper saying that there is evidence to reject the null hypothesis, because his p-value was less than 5%.

Are you convinced by his results? Comment on the validity of the result? [2]

(3) During your statistics class you heard a lot about t-distributions with this and that degrees of freedom. The main reason we use a t-distribution instead of a normal distribution is to account for the extra variability in estimating the standard deviation rather than using the true standard deviation. Suppose the sample size is very large (say over 100), in general

does it matter whether you use a normal distribution or t-distribution, give a reason for your answer? [2]

- (4) A tomato packing company advertises that the mean weight of their tomato boxes is 300g (it is also known that standard deviation of the weights is 10g). It is okay if the mean weight is greater than what they claim but it is an offense if the mean is less than what is claimed (and risks closing the company down). Therefore, every year an inspector takes a sample of 30 boxes of tomatoes to check their claims.

(Assume in this question that the standard deviation is known, therefore there is no need to use anything but the normal distribution).

(i) State the null and the alternative for this scenario. [2]

(ii) The inspector does the test at the 5% level. Based on this information, give the cut off point (critical region) such that if the average weight is less than this value we reject the null. [2]

Hint: remember to calculate the z-value and use the correct mean and standard deviation to transform back.

(iii) If the sample average is less than this threshold, then we make the decision to accept the alternative. Will this decision be right or wrong? Explain your answer. [1]

(iv) What is the probability that the sample mean will be greater than the threshold given in part (ii) if the true mean packing weight of tomato boxes is 300g? If you do not have a number for part (ii), then use the number 290 instead. [1]

- (v) Suppose that an employee has a grudge against the company. He recalibrates the machine such that its mean weight of tomatoes is now $\mu = 295g$. Using your answer in (ii), do you think it is likely that the inspector will detect the change in mean and close down the company (explain your answer)? [1]

- (5) A new feed has been developed that is believed to cause a greater weight gain in calves weight over conventional feeds. To statistically test these claims a farmer random sampled 20 sets of twins (calves). To one twin he gave conventional feed and the other calf he gave the new feed. The farmer did not know what test to do so he did both a paired t-test and an independent sample t-test. The statcrunch outputs of both tests are summarised below (X corresponds to the weight on conventional feed and Y to the weight on the new feed).

- (i) By selecting the correct output (see Figures 1 and 2), do we reject the test at the 5% level? [1]

- (ii) Using the output construct a 95% CI for the difference in the population means. [1]

- (iii) The manufacturer of the new feed claims that the mean weight increase over the conventional feed is at least 1.5 pounds, is there any evidence to substantiate these claims (evidence here being at the 5% level)? [1]

Hint: Part (ii) can help you answer this question.

- (iv) Using the plot in Figure 3 explain why there is a difference in the results of the tests for the paired t-test and the independent sample t-test. [1]

Hypothesis test results:

μ_1 : mean of X

μ_2 : mean of Y

$\mu_1 - \mu_2$: mean difference

$H_0 : \mu_1 - \mu_2 = 0$

$H_A : \mu_1 - \mu_2 < 0$

(without pooled variances)

Difference	Sample Mean	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	-0.74636257	0.54786366	28.843489	-1.3623145	0.0918

Figure 1: Independent sample t-test

Hypothesis test results:

$\mu_1 - \mu_2$: mean of the paired difference between X and Y

$H_0 : \mu_1 - \mu_2 = 0$

$H_A : \mu_1 - \mu_2 < 0$

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
X - Y	-0.74636257	0.2778224	19	-2.6864736	0.0073

Figure 2: Paired t-test

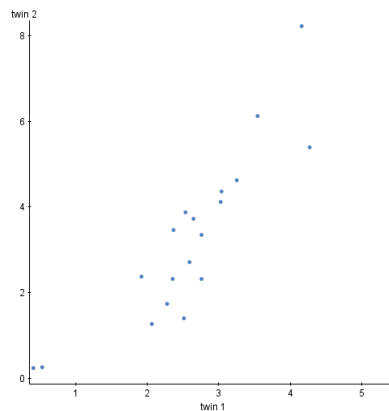


Figure 3:

- (6) A survey was conducted to see whether there is a difference in the preference of fruit juices of those living on the east and west coast. 1000 people were surveyed and the 2 by 2 table summarising the data is given below.

	East	West	Subtotal
Orange		300	400
Apple			600
Subtotal	600	400	1000

- (i) Fill in the missing cells in the table above. [1]

- (ii) If there is no association between the location of a person and their fruit juice preference, on average how many of the West coast people in the survey would you expect to like orange juice and how many would you expect to like Apple juice? [1]

- (iii) Without doing a formal test, by using the table and your answer to (ii), do you believe there is an association, give a reason for your answer? [1]