Midterm 2 - STAT 301
Spring 2013

Name:

UIN:

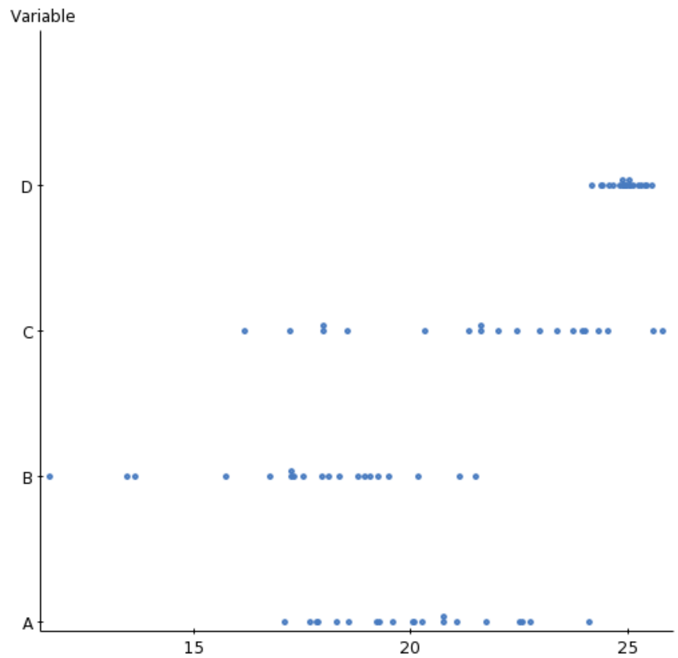
Signature:

Version A:

1. Do not open this test until told to do so.
2. This is a closed book examination, However you may use one single-sided sheet of formulas that you have brought with you and the tables. You should have no other printed or written material with you on the exam. But scrap paper is allowed.
3. You have 60 minutes to work on this exam. There are 15 multiple choice questions.
4. On the scantron please state the version of exam that you have.
5. You may use a calculator in the exam.
6. If there is no correct answer or if multiple answers are correct, select the **best** answer.
7. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or course assistant for clarification.
8. Please only give one answer per question (the one that is closest to the solution).
9. Good Luck!!!

blank

(1-2) [FRUIT PACKING] Four fruit packing machines (machine A, B, C and D) pack limes in bags which are suppose have a mean weight of 20 ounces. The fruit packing company is worried that the machines are not functioning well and the mean weight is not 20 ounces. To investigate this claim the company takes 20 bag samples from each of the four machines. A dot plot of the data and summary statistics is given in Figure 1.



Summary statistics:

Column	n	Mean	Range	Min	Max
A	20	20.10465	7.015831	17.09554	24.11137
B	20	17.66841	9.825138	11.660146	21.485283
C	20	21.783342	9.606618	16.18483	25.791447
D	20	24.9408	1.410344	24.161917	25.57226

Figure 1:

Use Figure 1 to answer next two questions.

(1) [FRUIT PACKING] We want to the the hypothesis $H_0 : \mu = 20$ against $H_A : \mu \neq 20$. Which sample (machine) will give the smallest p-value?

(A) Machine A.

(B) Machine B.

(C) Machine C.

(D) Machine D.

(E) It's unclear without the standard errors.

(2) [FRUIT PACKING] We want to test the hypothesis $H_0 : \mu = 20$ against $H_A : \mu > 20$. Which machine will give the largest p-value?

(A) Machine A.

(B) Machine B.

(C) Machine C.

(D) Machine D.

(E) It's unclear without the standard errors.

(3) The summary statistics of one of the lime packing machines is given in figure 2. Based

Summary statistics:

Column	n	Mean	Std. Dev.	Std. Err.
B	20	17.66841	2.505967	0.56035125

Figure 2:

on the summary statistic (using a t-distribution with 19 degrees of freedom). You want to test the hypothesis $H_0 : \mu = 20$ against $H_A : \mu < 20$ - give the closest t-value and p-value.

(A) The t-transform is -0.93 and the p-value is between 15% to 20%.

(B) The t-transform is -0.93 and the p-value is between 30% and 40%.

(C) The t-transform is 4.17 and the p-value is less than 0.05%.

(D) The t-transform is -4.17 and the p-value is less than 0.1%.

(E) The t-transform is -4.17 and the p-value is less than 0.05%.

(4) A professor is investigating whether students who attend class do better in exams than those who do not attend class. He wants to test the hypothesis $H_0 : \mu_C - \mu_{NC} = 0$ against $H_A : \mu_C - \mu_{NC} > 0$ (where μ_C is the mean score of students who attend class and μ_{NC} is the mean score of students who do not attend class). He collects the data from his own class, by taking attendance of students and comparing with their corresponding grade. He compares the average grade of students who attend class with those who don't, for this sample on average students who attended class got 8% more than students who did not attend class. Using this information and the corresponding standard error, the professor calculated the p-value to be 2%. Which statement(s) are correct.

(A) He can reject the null hypothesis at the 5% level, this suggests that that it is unlikely we can explain the 8% difference by random chance, and attending class improves the exam performance of students.

(B) Since the students who attended his class got on average 8% higher it is highly likely that this professor is a good teacher.

(C) He can reject the null hypothesis at the 5% level, this suggests that it is unlikely we can explain the 8% difference by random chance, and there is evidence to suggests that the students who choose to attend class get higher grades. However, we can not say that attending class improves the grade, because there could be other factors involved.

(D) He cannot reject the null hypothesis at the 5% level. Therefore, the difference in grades can easily be explained by random chance (though we do know that technically we are not allowed to accept the null).

(E) None of the above.

(5) I would like to construct a confidence interval to locate the mean height for a population of mammals. I draw a random sample, and the 99% confidence interval for the mean height is $[20, 52]$ (for this interval the margin or error = 16). I would like to construct an interval which has a margin of error = 8. What can I plausibly do to obtain a margin of error = 8?

(A) Increase the sample size by factor eight.

(B) Increase the sample size by factor four.

(C) Since the critical values for the 99% and 80% confidence intervals are 2.56 and 1.28 respectively, we can decrease the level of confidence from a 99% level to 80% level.

(D) Either do (B) or (C).

(E) Either do (A) or, alternatively, reduce the standard deviation of the population.

(6) A dieting club is concerned that the mean number of calories in Big Fat Chocolate diet bar is more than the 250 calories that is advertised. Let μ_B denote the mean number of calories in a Big Fat Chocolate diet bar. They collect a sample of 40 Big Fat Chocolate diet bars and measure the number of calories in each bar (by burning them in controlled conditions) and find that the sample mean $\bar{X} = 260$. Identify the correct hypothesis:

(A) $H_0 : \mu_B = 250$ against $H_A : \mu_B > 250$.

(B) $H_0 : \mu_B > 250$ against $H_A : \mu_B = 250$.

(C) $H_0 : \mu_B = 250$ against $H_A : \mu_B \neq 250$.

(D) $H_0 : \bar{X} = 260$ against $H_A : \bar{X} > 260$.

(E) $H_0 : \bar{X} = 250$ against $H_A : \bar{X} > 250$.

(7) Antecdotal evidence suggests that Texans are less likely to be vegetarian than Californians. I draw a random sample to see whether there is evidence of this. What hypothesis should I use?

(A) H_0 : The number of Texans who are vegetarian is equal to the number of Californians who are vegetarian against H_A : The number of Texans who are vegetarian is less than the number of Californians who are vegetarian.

(B) H_0 : The percentage of Texans who are vegetarian is equal to the percentage of Californians who are vegetarian against H_A : The percentage of Texans who are vegetarian is less than the percentage of Californians who are vegetarian.

(C) H_0 : The proportion of Texans who are vegetarian is less than the proportion of Californians who are vegetarian against H_A : The proportion of Texans who are vegetarian is equal to the proportion of Californians who are vegetarian.

(D) Two of the above.

(E) None of the above.

(8-9) [CALVES] Our aim is to investigate how calf weights change from 1/2 a week after birth to one week after birth. The results of a matched paired t-test are given in Figure 3, you will need this information to answer the following two questions. Let μ_1 be the mean weight at week 0.5 and μ_2 be the mean weight at week 1.

Hypothesis test results:

$\mu_1 - \mu_2$: mean of the paired difference between wt 0.5 and Wt 1

$H_0 : \mu_1 - \mu_2 = 0$

$H_A : \mu_1 - \mu_2 \neq 0$

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
wt 0.5 – Wt 1	1.0909091	0.53232527	43	2.049328	0.0466

Figure 3:

(8) Assume all the tests are done at the 5% level, which statement(s) are correct?

(A) If we test the hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against $H_A : \mu_1 - \mu_2 > 0$, the p-value is about 2.3% and there is evidence to suggest that the mean weight has dropped from week 0.5 to week 1.

(B) If we test the hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against $H_A : \mu_1 - \mu_2 < 0$, the p-value is about 2.3% and there is evidence to suggest that the mean weight has increased from week 0.5 to week 1.

(C) If we test the hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against $H_A : \mu_1 - \mu_2 < 0$, the p-value is about 2.3% and there is evidence to suggest that the mean weight has decreased from week 0.5 to week 1.

(D) If we test the hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against $H_A : \mu_1 - \mu_2 \neq 0$, we are unable to reject the null at the 5% level.

(E) The Statcrunch output is for two-sided tests, therefore none of the above statements can be determined from the output.

(9) To answer the following question you will need the critical t-values corresponding to the upper tail probability for the t-distribution with 43 degrees of freedom. They are given in the table below.

probability	0.15	0.10	0.05	0.025	0.01
t^*	1.049	1.30	1.68	2.01	2.41

Based on this information, what is a 95% confidence interval for the mean difference in weight from week 0.5 to week 1.

(A) The 95% confidence interval for the mean difference in weight is $[1.09 - 2.01 \times \frac{0.53}{\sqrt{44}}, 1.09 + 2.01 \times \frac{0.53}{\sqrt{44}}]$.

(B) The 95% confidence interval for the mean difference in weight is $[1.09 - 1.68 \times 0.53, 1.09 + 1.68 \times 0.53] = [0.1996, 1.9804]$.

(C) The 95% confidence interval for the mean difference in weight is $[0 - 2.01 \times 0.53, 0 + 2.01 \times 0.53] = [-1.0653, 1.6053]$.

(D) The 95% confidence interval for the mean difference in weight is $[1.09 - 2.01 \times 0.53, 1.09 + 2.01 \times 0.53] = [0.0247, 2.1553]$.

(E) The 95% confidence interval for the mean difference in weight is $[1.09 - 2.01 \times 0.53 \times \sqrt{44}, 1.09 + 2.01 \times 0.53 \times \sqrt{44}]$.

(10) Pretend that the 95% confidence interval for the difference in the mean weights of calves from week 0.5 to week 1 is $[0.1, 1]$. What is the correct interpretation of this confidence interval?

(A) Suppose we took 200 different simple random samples (SRS) of calves and for each SRS we constructed a confidence interval. About 190 of these confidence intervals would contain the mean difference in weights.

(B) The calves come from a simple random sample (SRS), therefore there is a 95% chance the difference in means lies in the interval $[0.1, 1]$.

(C) Because the calves come from a SRS, about 95% of the weight gain will lie in the interval $[0.1, 1]$. This interval will help us to diagnose calves that are unhealthy.

(F) Two of the above above statements.

(E) None of the above.

(11) Pretend that the 95% CI for the mean difference in weight is $[0.1, 1]$. We test the two hypothesis $H_0 : \mu = 0.9$ against $H_A : \mu \neq 0.9$ and $H_0 : \mu = 0.5$ against $H_A : \mu \neq 0.5$. Which statement is correct?

(A) Both 0.9 and 0.5 lie in the 95% confidence interval $[0.1, 1]$, therefore there is no evidence to reject either of the nulls at the 5% level.

(B) Both 0.9 and 0.5 lie in the 95% confidence interval $[0.1, 1]$, therefore there is no evidence to reject either of the nulls at the 1% level.

(C) Both 0.9 and 0.5 lie in the 95% confidence interval $[0.1, 1]$, therefore there is no evidence to reject either of the nulls at the 10% level.

(D) Since both 0.9 and 0.5 lie in the confidence interval, the results of the test suggest that the mean can only be 0.5 or 0.9.

(E) Two of the above.

(12) Your friend is asking your advice on an article that he is reading. The article said *The test was done at the 5% significance level.* Your friend asks you to explain what is meant by a 5% significance level.

(A) If the null hypothesis is true and the test is done 200 times, we will falsely reject the null about 10 times.

(B) There is a 5% chance the null hypothesis is true.

(C) If the null hypothesis is true and any we repeat experiment and the corresponding test many times, eventually we will falsely reject the null.

(D) Two of the above.

(E) None of the above.

(13) A pregnant women is diagnosed with gestational diabetes if her mean glucose level is over 140mg one hour after drinking a sugary drink. She is diagnosed as critically ill if her mean level is over 155mg.

A medical test does exist for diagnosing gestational diabetes, but this is very costly. Instead doctors want to first identify risky patients. If the patient is determined to be at risk then they will go on to have a costly medical test to see whether they have gestational diabetes.

Risky patients are identified by taking two blood samples and statistically testing the hypothesis $H_0 : \mu = 130$ against $H_A : \mu > 130$ based on the average of the two blood samples. The test is done at the 5% significance level and if the patient's p-value is less than 5% they are given the medical procedure. It is found that the ability of this test to detect critically ill patient (mean of 155 or over) is only 40% (ability of the test to reject the null when the $\mu = 155$ is 40%), which is very low. Our objective is to see what measures can be taken to increase the probability of detection.

(A) We can improve the ability of the test to detect critically ill patients by decreasing the significance level, the insurance companies will like this because less healthy patients will be given the costly medical test.

(B) We can improve the ability of the test to detect critically ill patients by increasing the significance level, however this means that more healthy patients will given the costly medical test.

(C) Increase the number of samples taken. This will mean the standard error will decrease and thus increase the ability of the test to detect critically ill patients.

(D) Use 155 as my point of rejection, so all averages above 155 will be identified as risky.

(E) Two of the above.

(14) In a statistical test about μ the null hypothesis was not rejected. Based on this conclusion, which of the following statements are true?

(A) It is impossible to have committed both a type I and type II error.

(B) Whether an error was committed or not is unknown, but if an error was made, then it was a type I error.

(C) Whether an error was committed or not is unknown, but if an error was made, then it was a type II error.

(D) (A) and (B)

(E) (A) and (C).

(15) I suspect that my pet dog Woof can predict the outcome of international soccer matches. I place two balls infront of Woof, the red ball represents one team and the blue ball represents the opposition. Woof picks up one of the two balls, the ball he picks up is the team he predicts to win the soccer match.

My assertion that he can predict the outcome of matches comes from the the observation that he correctly predicted the outcome of three consecutive soccer matches last weekend. It is known that the probability that a dog can just randomly predict correctly the outcome of three consecutive soccer matches is 12.5% (ie. the probability of correctly predicting the outcome of three consecutive soccer matches by just random chance is 12.5%).

Which statement is/are correct?

(A) If we test the hypothesis that H_0 : Woof has no predictive ability against H_A : Woof has predictive abilities, we see that the p-value is 12.5%, as this is greater than the 5% significance level, his predictions can easily be explained by random chance, though it does not prove that Woof has no predictive abilities.

(B) This is an example of biased reporting. If I wait long enough (and Woof has a very long life), by random chance he will be able to correctly predict the outcome of 3 or more consecutive soccer matches (even 20 consecutive soccer matches!).

(C) If we test the hypothesis that H_0 : The dog has predictive ability against H_A : The dog has no predictive abilities, we see that the p-value is 12.5%, as this is greater than 5% we can reject the null. This proves that my dog has no predictive abilities.

(D) (B) and (C) are correct.

(E) (A) and (B) are correct.