Midterm 3 - STAT 301
Spring 2018

Name:

UIN:

Signature:

Version A:

1. Do not open this test until told to do so.
2. Please **read** each question carefully.
3. This is a closed book examination, However you may use the cheat sheet provided and the tables you have brought with you. You should have no other printed or written material with you on the exam. But scrap paper is allowed.
4. You have 60 minutes to work on this exam. There are **15** multiple choice questions.
5. On the scantron please state the version of exam that you have.
6. You may use a calculator in the exam.
7. If there is no correct answer or if multiple answers are correct, select the **best** answer.
8. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or course assistant for clarification.
9. Do all tests at the **5%** level unless specified otherwise.
10. Some questions are very easy (don't make them more complicated than they are), make sure you get these correct.
11. Please only give one answer per question (the one that is closest to the solution).
12. Good Luck!!!

(1-5) Nutty Narrow Bridge (in Washington State) is the world's narrowest bridge and is designed specifically for squirrels to cross a road safely. A counter is placed on the bridge to monitor the number of squirrels crossing the bridge each day.

Prior to 2014, the mean number of squirrels that crossed the bridge each day was 90.5. However, during 2014 some neighbouring forests were cut down. Over a 100 day period in 2015 the number of squirrels crossing the bridge (each day) was recorded. The data is summarized below.

One sample T hypothesis test:
μ : Mean of variable
H_0 : μ = 90.5
H_A : μ ≠ 90.5

Hypothesis test results:

Variable	Sample Mean	Std. Err.	DF	T-Stat	P-value
squirrels	85.22	2.1036563	99	-2.5099157	0.0137

probability	0.3	0.15	0.10	0.05	0.025	0.01	0.005
t^*	0.53	1.042	1.29	1.66	1.98	2.36	2.63

Table 1: Critical values for t-distribution with 99df

(1) A nature group is concerned that after 2014 there has been a **drop** in the number of squirrels crossing the bridge. What is the hypothesis of interest and the result of the test?

(A) $H_0 : \mu \leq 90.5$ vs $H_A : \mu > 90.5$, the p-value is 0.685%. Thus there **is** evidence that the number of squirrels using the crossing has **increased**.

(B) $H_0 : \mu \leq 90.5$ vs $H_A : \mu > 90.5$, the p-value is 99.315%. Thus there **is** evidence that the number of squirrels using the crossing has **increased**.

(C) $H_0 : \mu \geq 90.5$ vs $H_A : \mu < 90.5$, the p-value is 0.685%. Thus there **is** evidence that the number of squirrels using the crossing has **dropped**.

(D) $H_0 : \mu \geq 90.5$ vs $H_A : \mu < 90.5$, the p-value is 99.315%. Thus there **no** is evidence that the number of squirrels using the crossing has **dropped**.

(E) $H_0 : \mu = 90.5$ vs $H_A : \mu \neq 90.5$, the p-value is 1.37%. Thus there **is** evidence that the number of squirrels using the crossing has **increased**.

(2) The city council pest control department suspects the number of squirrels crossing the bridge has **changed** after 2014. What is the hypothesis of interest and the result of the test?

(A) $H_0 : \mu = 90.5$ vs $H_A : \mu \neq 90.5$, the p-value is 0.685%. Thus there is **no** evidence that the number of squirrels using the crossing has **changed**.

(B) $H_0 : \mu = 90.5$ vs $H_A : \mu \neq 90.5$, the p-value is 99.315%. Thus there **is** evidence that the number of squirrels using the crossing has **increased**.

(C) $H_0 : \mu \neq 90.5$ vs $H_A : \mu = 90.5$, the p-value is 1.37%. Thus there **is** evidence that the number of squirrels using the crossing has **remained the same**.

(D) $H_0 : \mu = 90.5$ vs $H_A : \mu \neq 90.5$, the p-value is 1.37%. Thus there **is** evidence that the number of squirrels using the crossing has **changed**.

(3) A squirrel advocacy group suspects that mean number of squirrels crossing the bridge has dropped below $\mu = 88$. What is the hypothesis of interest and the results of the test (at the 5% level)?

(A) $H_0 : \mu \geq 88$ vs $H_A : \mu < 88$. The p-value is between 5-10%, there is **no** evidence of a drop below 88.

(B) $H_0 : \mu \leq 88$ vs $H_A : \mu > 88$. The p-value is between 90-95%, there **is** evidence of a drop below 88.

(C) $H_0 : \mu \geq 85.22$ vs $H_A : \mu < 85.22$. The p-value is between 90-95%, there is **no** evidence of a drop below 85.22.

(D) $H_0 : \mu \geq 88$ vs $H_A : \mu < 88$. The p-value is 1.32%, there is **no** evidence of a drop below 88.

(E) $H_0 : \mu \geq 88$ vs $H_A : \mu < 88$. The p-value is 1.32%, there **is** evidence of a drop below 88.

(4) Construct a **95%** confidence interval for the mean number of squirrels crossing the bridge after 2014.

(A) [84.8,85.6] [B] [81.7,88.7] [C] [86.3,94.7] [D] [87.0,94.0] [E] [81.1,89.4]

(5) The QQplot of the data and the sample mean (based on sample size $n = 100$) is given in Figure 1 (see next page). Which statement(s) is correct?

(A) The distribution of the data is heavy tailed.

(B) The sample mean is close to normal.

(C) The QQplots show that the p-values derived in Questions 1-3 are correct.

(D) The QQplots show that we **do not** have the stated 95% confidence in the interval constructed in Question 4.

(E) [A], [B] and [C].

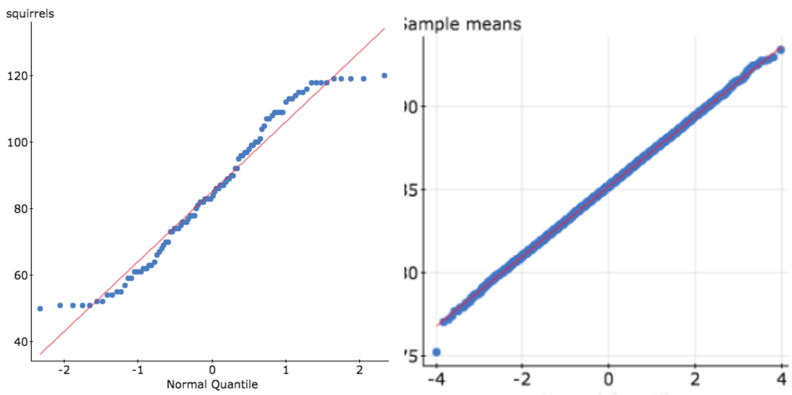


Figure 1: Left: QQplot of Data. Right: QQplot of sample mean.

(6) Gestational diabetes is diagnosed if the mean level of sugar in the blood is **greater than 140**. Every expectant women is tested for gestational diabetes. Each women gives four blood samples and using the blood samples ones tests $H_0 : \mu \leq 140$ against $H_A : \mu > 140$. The test is done at the 5% significance level.

Which statement is correct?

(A) If **1000** healthy women **without** gestational diabetes took the test, on average **950** would be **falsely** diagnosed with gestational diabetes.

(B) If **1000** healthy women **without** gestational diabetes took the test, on average **50** would be **falsely** diagnosed with gestational diabetes.

(C) If **1000** women **with** gestational diabetes took the test, on average **950** would be **falsely** diagnosed with gestational diabetes.

(D) If **1000** women **with** gestational diabetes took the test, on average **50** would be **falsely** diagnosed with gestational diabetes.

(7) Recently there has been research on links between Anxiety disorders (in particular Obsessive Compulsive Disorder; OCD for short) and the diversity of gut flora (diversity of bacteria in the gut).

In a recent study conducted by PhD student Jasmine Turna, the gut flora of 11 OCD patients and 12 healthy controls underwent gut microbiome analysis. The following results were presented at the Anxiety and Depression Conference, 2017.

"The average diversity of gut flora in the 11 OCD patients was **lower** than the average diversity of gut flora in the 12 healthy patients (1.2 in the OCD patients and 1.9 in the healthy control group). The result of an independent two-sample t-test gave a

p-value <0.1%."

Let μ_{ocd} denote the mean level of gut flora in OCD patients and $\mu_{control}$ denote the mean level of gut flora in the healthy controls. Which statement(s) is/are correct?

(A) In the test $H_0 : \mu_{control} - \mu_{ocd} \leq 0$ vs $H_A : \mu_{control} - \mu_{ocd} > 0$ the p-value is less than 0.1%. Thus there is **clear evidence** to suggest that patients with OCD tend to have **less** diverse gut flora than healthy people.

(B) In the test $H_0 : \mu_{control} - \mu_{ocd} \leq 0$ vs $H_A : \mu_{control} - \mu_{ocd} > 0$ the p-value is less than 0.1%. Thus there is **no** evidence to suggest that patients with OCD tend to have **less** diverse gut flora than healthy people.

(C) In the test $H_0 : \mu_{control} - \mu_{ocd} \geq 0$ vs $H_A : \mu_{control} - \mu_{ocd} < 0$ the p-value is less than 0.1%. Thus there is **clear** evidence to suggest that patients with OCD tend to have **more** diverse gut flora than healthy people.

(D) The research suggests that people with less diverse gut flora will develop an anxiety disorder, such as OCD.

(E) [C] and [D].

(8-11) Recently there has been research in the role that probiotics may have on depression (probiotics are live "good" bacteria). A conjecture by a group of psychiatrists is that probiotics may **reduce the length** of a depressive episode in clinically depressed patients.

To investigate this, a random sample of 20 hospitalized patients with clinical depression participated in a study. In order to ensure that any effect observed was not due to a placebo, neither the participants nor their doctors were told about the aims of the study. At the start of the study (before they were given probiotics) the length of a depressive epsiode for each participant was recorded (in days). Then for six weeks high levels of probiotics were given to the patients. At the end of the study (after they were given probiotics for 6 weeks) the length of each of the participants depressive epsiodes (in days) were again recordered. The data is summarized below.

Paired T confidence interval:
$\mu_D = \mu_1 - \mu_2$: Mean of the difference between Before and After
95% confidence interval results:

Difference	Mean	Std. Err.	DF	L. Limit	U. Limit
Before - After	2.2216426	0.23129088	19	1.7375452	2.70574

(8) Why must a matched/paired t-test (method) be used?

(A) For each individual, the difference between the length of depressive episodes before and after treatment gives information on the effectiveness of the treatment.

(B) We can only use a matched paired method when the sample size is small.

(C) The matched paired method corrects for the lack of normality of the data

(D) [B] and [C]

(9) Construct a 95% confidence interval for the mean difference in the length of depressive episodes before and after treatment (with probiotics).

(A) 2.22 [B] [2.11,2.33] [C] [1.82,2.61] [D] [1.74,2.71] [E] [0.015,0.978]

(10) Is there any evidence in the data to suggest that the use of probiotics may reduce the length of a depressive episode? State the hypothesis of interest and the results of the test.

(A) $H_0 : \mu_B - \mu_A \leq 0$ vs $H_0 : \mu_B - \mu_A > 0$. The p-value is 9.65%. There is **no** evidence at the 5% level to suggest that probiotics **reduce** the length of depressive episodes.

(B) $H_0 : \mu_B - \mu_A \leq 0$ vs $H_0 : \mu_B - \mu_A > 0$. The p-value is less than 0.05%. There is strong evidence to suggest that probiotics **reduce** the length of depressive episodes.

(C) $H_0 : \mu_B - \mu_A \geq 0$ vs $H_0 : \mu_B - \mu_A < 0$. The p-value is less than 0.05%. There is strong evidence to suggest that probiotics **increase** the length of depressive episodes.

(D) Using the data, we reject the null $H_0 : \mu_B - \mu_A \leq 0$ and accept the alternative, at **both** the 5% and the 0.1% significance levels.

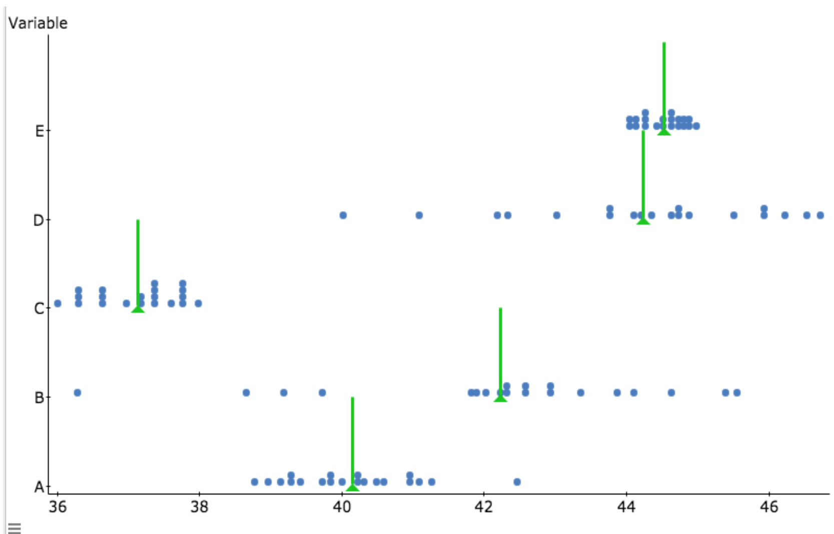
(E) (B) and (D).

(11) Match the following three experiment with the correct test.

1. To test the efficacy of a memory enhancing drug, 20 randomly sampled twins were analysed. One twin was placed in the Control group while the other was placed in the drug group (altogether there are 10 in the Control group and 10 in the drug group).
2. To see whether the protein content in bread decreases with age, 30 loaves of breaded were baked. Immediately after baking the amount of protein was measure in each bread and then three days later the amount of protein was measured.
3. To see whether the protein content in bread decreases with age, 60 loaves of breaded were baked. They split the loaves into two groups, each of size 30. In the first group the amount of protein was measured immediately after baking. In the second group the amount of protein was measured 3 days after baking.

	[1]	[2]	[3]
(A)	Matched paired t-test	Independent two Sample t-test	Independent two Sample t-test
(B)	Matched paired t-test	Independent two Sample t-test	Matched paired t-test
(C)	Independent two Sample t-test	Matched paired t-test	Independent two Sample t-test
(D)	Matched paired t-test	Independent two sample t-test	One sample t-test
(E)	Matched paired t-test	Matched paired t-test	Independent two Sample t-test

(12-13) Below a dotplot of 5 data sets (labelled A-E) is given (20 observations in each data set). The vertical line is the sample mean.

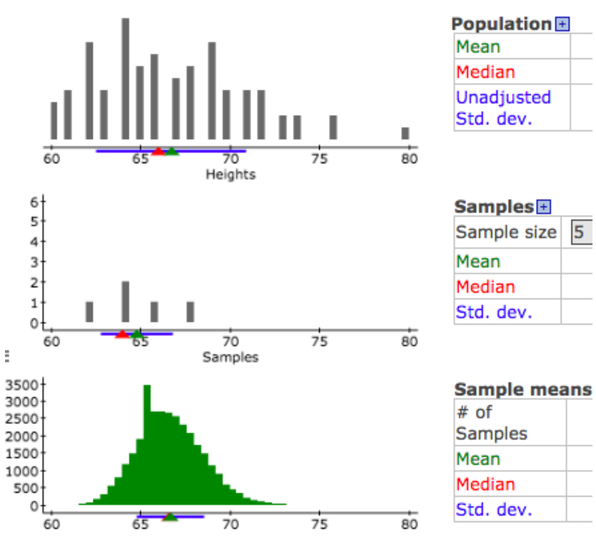


(12) We test the hyothesis $H_0 : \mu \leq 40$ vs $H_A : \mu > 40$. Which data set has the smallest p-value.

(13) We test the hyothesis $H_0 : \mu \leq 40$ vs $H_A : \mu > 40$. Which data set has the largest p-value.

(14) 5 students are randomly sampled, their sample mean is **64.8** inches.

To understand the distribution of the sample means we use the sampling distribution applet in Statcrunch. Which statement(s) are correct?



(A) **64.8** is a number drawn from the histogram in the top plot.

(B) **64.8** is a number drawn from the histogram in the middle plot.

(C) **64.8** is a number drawn from the histogram in the bottom plot.

(D) The distribution of the sample mean based on 5 is highly left skewed.

(E) [A] and [D].

(15) Suppose the weight of **healthy** hens are normally distributed with mean **6** pounds and standard deviation **two** pounds (population standard deviation is $\sigma = 2$).

A farmer has a coop and is monitoring her hens. The average weight of the **10** hens in her coop is $\bar{x} = 7$ pounds. Is there any evidence to suggest that the mean weight of the farmer's hens is **less** than 6 pounds? State the hypothesis of interest and the corresponding p-value.

Remember, do not use the t-distribution use the normal distribution.

(A) $H_0 : \mu \geq 6$ vs $H_A : \mu < 6$, the p-value is **98.4%**, there is evidence the mean is **less** than 6 pounds.

(B) $H_0 : \mu \geq 6$ vs $H_A : \mu < 6$, the p-value is **5.7%**, there is **no** evidence the mean is **less** than 6 pounds.

(C) $H_0 : \mu \geq 6$ vs $H_A : \mu < 6$, the p-value is **94.3%**, there is **no** evidence the mean is **less** than 6 pounds.

(D) $H_0 : \mu \leq 6$ vs $H_A : \mu > 6$, the p-value is **5.7%**, there is evidence the mean is **less** than 6 pounds.

(E) $H_0 : \mu \geq 7$ vs $H_A : \mu < 7$, the p-value is **5.7%**, there is **no** evidence the mean is **less** than 6 pounds.