Midterm 3 - STAT 301
Fall 2019

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 **16** 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) Botonists are trying to understand how African Bush lillies (a type of plant) propogate their seeds. They believe it is done by samango monkeys, who first sit by the plant, then pick the fruit and eat it (while sitting by the plant) and finally spit the seeds out.

It is known that if the monkeys spit the seeds too close to the "parent" plant, then there will be competition for resources (which is bad for both the parent and child plants). In order for the parent and child plants not to compete for resources, the seed must be *more* than 63cm from the plant (source: Biotropica, Volume 51, 2019).

Botonists hypothesis that the monkeys spit the seeds **more** than 63 cm. To test this, they measured the distance of 20 seeds spat out by monkeys. The data is summarized below.

One sample T hypothesis test:
μ : Mean of variable
$H_0 : \mu = 63$
$H_A : \mu > 63$

Hypothesis test results:

Variable	Sample Mean	Std. Err.	DF	T-Stat	P-value
SeedDistance	66.325225	0.49932466	19	6.6594456	<0.0001

(1) Is there any evidence to suggest the mean spit distance is greater than 63cm. State the hypothesis of interest and the conclusion of the test?

(A) $H_0 : \mu \leq 63$ vs $H_A : \mu > 63$, since the p-value is less than 0.01% there is not enough evidence to suggest that mean is greater than 63cm.

(B) $H_0 : \mu \leq 63$ vs $H_A : \mu > 63$, since the p-value is less than 0.01% there is strong evidence to suggest that mean is greater than 63cm.

(C) $H_0 : \mu \leq 66.3$ vs $H_A : \mu > 66.3$, since the p-value is less than 0.01% there is strong evidence to suggest that mean is greater than 66.3cm.

(D) $H_0 : \mu \geq 63$ vs $H_A : \mu < 63$, since the p-value is less than 0.01% there is not enough evidence to suggest that mean is less than 63cm.

(E) $H_0 : \mu \leq 63$ vs $H_A : \mu < 63$, since the p-value is less than 0.01% there is strong evidence to suggest that mean is less than 63cm.

(2) However, botonists also believe that in order that the "child" plants have the same microclimate as the parent plant, the seeds must be spat **less** than 68cm from the parent plant. Using the data above is there any evidence that mean spitting distance of the monkeys is **less** than **68**cm (test $H_0 : \mu \geq 68cm$ vs $H_A : \mu < 68cm$)?

(A) The p-value is 3.36%, there is **not** enough evidence to suggest the mean is **less** than 68cm.

(B) The p-value is about 3.36%, there **is** sufficient evidence to suggest the mean is **greater** than 68cm.

(C) The p-value is 6.7%, there is **not** enough evidence to suggest the mean is **less** than 68cm.

(D) The p-value is between 0.1-0.25%, there **is** evidence to suggest the mean is **less** than 68cm.

(E) The p-value is between 99.75-99.9%, there **is not** evidence to suggest the mean is less than 68cm.

(3) Construct a 95% confidence interval for the mean distance of the seeds spat out by the monkeys.

(A) [61.95,64.05] (B) [66.08,66.55] (C) [59.72,72.93] (D) [65.44,67.16] (E) [65.27,67.37]

(4) Recently, scientists investigated whether 12 weeks of blueberry concentrate supplementation **improved** cognitive function in healthy older adults. Participants were randomised to consume either 30 mL blueberry concentrate providing 387 mg antioxidants (22 participants) or an isoenergetic placebo (21 participants). To check for baseline differences. At the start of the study (before either the blueberry concentrate or placebo was taken) the participants were given various tasks. For the congnitive test, at the **start** of the study 9 out of 22 in the blueberry group passed the test whereas 9 out of 21 in the placebo group passed the test (the corresponding **p-value for the difference is 55%**).

After 12 weeks the participants were given another cognitive test. At the **end** of the study 21 out of 22 in the blueberry group passed the test, whereas 10 out of 21 in the placebo group passed the test (the corresponding **p-value for the difference is 0.107%**). Using these results what can one conclude about the results in the study?

(A) The sample sizes are too small to draw any meaningful conclusion.

(B) There is evidence to suggest that Blueberry juice **increased** cognitive function.

(C) There is strong evidence to suggest **differences** between the two groups at the **start** of the study.

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

(5-6) [DUCKS] The weights of three types of geese/ducks are compared. Namely, canadian geese, muscovy ducks and snow geese, 10 of each animal is sampled. The dot plot of the weights in each group is given below (the x-axis corresponds to the weight if the duck).

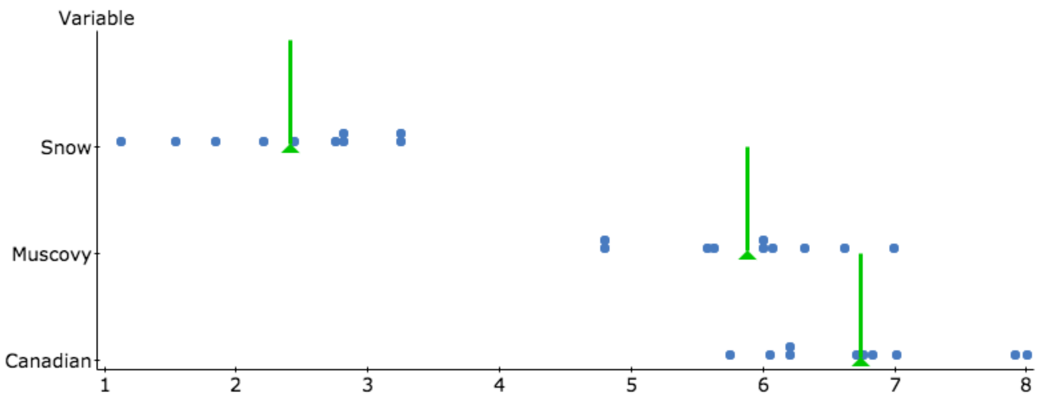


Figure 1: Average Snow = 2.4, Average Muscovy = 5.8, Average Canadian = 6.7.

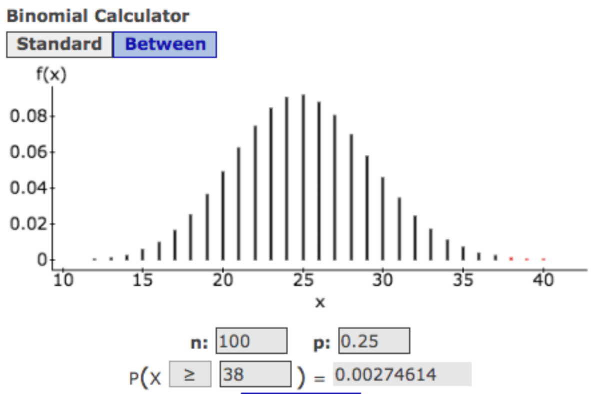
(5) Use the plot in Figure 1 to identify the correct p-value for each hypothesis. Pay attention to the alternative hypothesis.

	$H_0 : \mu_{\text{Canadian}} - \mu_{\text{Muscovy}} \leq 0$ $H_A : \mu_{\text{Canadian}} - \mu_{\text{Muscovy}} > \mathbf{0}$	$H_0 : \mu_{\text{Canadian}} - \mu_{\text{Snow}} \leq 0$ $H_A : \mu_{\text{Canadian}} - \mu_{\text{Snow}} > \mathbf{0}$	$H_0 : \mu_{\text{Muscovy}} - \mu_{\text{Snow}} \geq 0$ $H_A : \mu_{\text{Muscovy}} - \mu_{\text{Snow}} < \mathbf{0}$
A	less than 0.1%	10-20%	10-20%
B	1-10%	less than 0.1%	less than 0.1%
C	1-10%	less than 0.1%	more than 99.9%
D	90-99%	more than 99.9%	less than 0.01%
E	greater than 50%	10-20%	less than 5%

(6) Use the plot in Figure 1 to identify the correct p-value for each hypothesis. Read each hypothesis very **carefully**.

	$H_A : \mu_{\text{Canadian}} - \mu_{\text{Muscovy}} \leq 0$ $H_A : \mu_{\text{Canadian}} - \mu_{\text{Muscovy}} > \mathbf{0}$	$H_A : \mu_{\text{Canadian}} - \mu_{\text{Snow}} = 0$ $H_A : \mu_{\text{Canadian}} - \mu_{\text{Snow}} \neq \mathbf{0}$	$H_A : \mu_{\text{Muscovy}} - \mu_{\text{Snow}} = 0$ $H_A : \mu_{\text{Muscovy}} - \mu_{\text{Snow}} \neq \mathbf{0}$
A	less than 0.1%	10-20%	10-20%
B	1-10%	less than 0.1%	less than 0.1%
C	1-10%	less than 0.1%	more than 99.9%
D	90-99%	more than 99.9%	less than 0.01%
E	greater than 50%	10-20%	less than 5%

(7-8) In a multiple choice exam there were 100 question and each question had **4** choices. Matt claims he guessed the answers to all the questions. But you suspect that he had actually put some effort into answering the questions correctly. Matt scores 38 out of 100 in his exam. The statcrunch output is given below.



One sample proportion summary hypothesis test:
p : Proportion of successes
H_0 : p = 0.25
H_A : p > 0.25

Hypothesis test results:

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	38	100	0.38	0.04330127	3.0022214	0.0013

Figure 2:

(7) Using the output in Figure 2, is there an evidence that he knew some of the material?

(A) $H_0 : p \le 0.25$ vs $H_A : p > 0.25$. The p-value is less than 1% there is **no** evidence he knew some of the material.

(B) $H_0 : p \le 0.25$ vs $H_A : p > 0.25$. The p-value is less than 1% there **is** evidence he knew some of the material.

(C) $H_0 : p \ge 0.25$ vs $H_A : p < 0.25$. The p-value is over than 99.9% there is **no** evidence he knew some of the material.

(D) $H_0 : p \le 0.38$ vs $H_A : p > 0.38$. The p-value is 25% there is **no** evidence he knew some of the material.

(E) $H_0 : p \le 0.38$ vs $H_A : p > 0.38$. The p-value is less than 1% there is **no** evidence he knew some of the material.

(8) There is a small discrepancy between the two p-values in Figure 2. Give a reason(s) for the small difference.

(A) The true p-value is calculated using the binomial distribution, whereas the p-value given in the Statcrunch output is based on the normal distribution, which is an **approximation** of the true p-value.

(B) The binomial distribution for $p = 0.25$ tends to be right skewed. But since the sample size ($n = 100$) is large enough, the resulting binomial is quite close to normal. This is why the *difference* between the two p-values is quite small.

(C) The binomial distribution for $p = 0.25$ is highly left skewed leading to highly unreliable answers. The normal distribution *should* be used instead.

(D) [A] and [B] (E) None of the above.

(9) A patient has gestational diabetes if the mean glucose level, μ, of the patient is over 140. We are looking for evidence of gestational diabetes. The hypothesis is $H_0 : \mu \leq 140$ against $H_A : \mu > 140$.

The glucose level in a blood sample is normally distributed with mean μ and standard deviation $\sigma = 4$. A patient goes to the doctor. **Four** blood samples are taken and the sample mean is evaluated. The sample mean is $\bar{x} = 145$. Is there any evidence at the 5% level that she has gestational diabetes?

(A) The p-value is 0.6%, there is evidence she has gestational diabetes.

(B) The p-value is 10.6%, there is **no** evidence she has gestational diabetes.

(C) The p-value is 0.6%, there is **no** evidence she has gestational diabetes.

(D) The p-value is 2.5%, there is **no** evidence she has gestational diabetes.

(E) The p-value is 99.4%, there is **no** evidence she has gestational diabetes.

(10) Suppose in question (9) the significance level is **increased** from 5 to 10%. Which statement(s) are true?

(A) The proportion of healthy women **falsely** diagnosed with gestational diabetes will **decrease**.

(B) The proportion of healthy women **falsely** diagnosed with gestational diabetes will **increase**.

(C) The proportion of unhealthy women who will be **correctly** diagnosed with gestational diabetes will **increase**.

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

(11) Consider the hypothesis H_0 : innocent vs H_A : guilty, used in a criminal court. Evidence is collected and p-value based on the evidence (under the null) is calculated. We reject the null if the p-value is less than 5%. If we reject the nulll, the person goes to prison. Which statement is correct?

(A) 5% of guilty people **go** to prison. (B) 5% of guilty people **do not** go to prison

(C) 5% of innocent people **go** to prison. (D) 5% of innocent people **do not** go to prison.

(E) 95% of guilty people **go** to prison.

(12) We want to construct a confidence interval for the proportion of the population who support vaccinations. It is **known** this proportion must be **greater than 0.8** (but exactly what it is we do not know). What is the minimum sample to ensure that the margin of error of a **90**% CI is at most **2**% (2% = 0.02)?

(A) 1681 (B) 606 (C) 2401 (D) 1076 (E) 96

(13) Psychologists believe that the number of accidents may increase on Friday 13th as compared with other days. To check this they look through the emergency admission records in a hospital. They collect the number of accidents which happen on Friday 13th with the number of number of accidents which happen on the **previous** Friday 6th. They collect data over a two year period (where 6 Friday 13ths occurred). Let μ_{13} and μ_6 denote the mean number of accidents which happen on the 13th and 6th respectively. They test $H_0 : \mu_{13} - \mu_6 \leq 0$ vs $H_A : \mu_{13} - \mu_6 > 0$. By selecting the correct output, what is the conclusion of the test (at the 5% level)?

Two sample T hypothesis test:
μ_1 : Mean of Frid13th
μ_2 : Mean of Frid6th
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 > 0$
(without pooled variances)

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	3.3333333	2.0027759	9.9402006	1.6643567	0.0636

Paired T hypothesis test:
$\mu_D = \mu_1 - \mu_2$: Mean of the difference between Frid13th and Frid6th
$H_0 : \mu_D = 0$
$H_A : \mu_D > 0$

Hypothesis test results:

Difference	Mean	Std. Err.	DF	T-Stat	P-value
Frid13th - Frid6th	3.3333333	1.2292726	5	2.7116307	0.0211

(A) The p-value is 6.36%, there is a **no** evidence that more accidents happen on the 13th.

(B) The p-value is 6.36%, there is **some** evidence that more accidents happen on the 13th.

(C) The p-value is 2.11%, there is a **no** evidence that more accidents happen on the 13th.

(D) The p-value is 97.89%, there is **some** evidence that fewer accidents happen on the 13th.

(E) The p-value is 2.11%, there is **some** evidence that more accidents happen on the 13th.

(14) In a recent survey of 1000 American (conducted by Pew Research) the authors stated that $[66 \pm 3]$% of Americans (with 95% confidence) support legalisations of Marijuana. How do we interprete this poll?

(A) The proportion of the Americans who support legalisation of Marijuana is 66%.

(B) The proportion in this sample who support legalisation of Marijuana is 66%.

(C) The Margin of Error is 3%.

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

(15) 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

(16) Scientists suspect that a **Low** Calcium Diet **increases** iron absorption whereas a **High** Calcium Diet **decreases** iron absorption.

To test their hypothesis they asked for 20 volunteers. They randomly placed them into two groups. Before the start of the trial all of the volunteers had their iron level measured. Then, 10 people were placed on a high calcium diet (for one month) and the other 10 were placed on a low calcium diet (for one month). All volunteers were given iron supplements.

At the end of the trial all the volunteers had their iron level measured again. Let μ_{HC} denote the mean iron level after being on a calcium rich diet and μ_{LC} denote the mean iron level after being on a low calcium diet.

The scientists test the hypothesis $H_0 : \mu_{LC} - \mu_{HC} \leq 0$ against $H_A : \mu_{LC} - \mu_{HC} > 0$. Using the output below what is the result of the test (at the 5% level)?

Two sample T hypothesis test:
μ_1 : Mean of CalciumLow
μ_2 : Mean of CalciumHigh
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 < 0$
(without pooled variances)

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	1.991	0.62343894	17.623898	3.1935766	0.9974

(A) The p-value is 0.26%, there is evidence to suggest that a low calcium diet **increases** iron levels over a high calcium diet.

(B) The p-value is 99.74%, there is evidence to suggest that a low calcium diet **increases** iron levels over a high calcium diet.

(C) The p-value is 0.26%, there is evidence to suggest that a low calcium diet **decreases** iron levels over a high calcium diet.

(D) The p-value is 99.74%, there is evidence to suggest that a low calcium diet **decreases** iron levels over a high calcium diet.