STAT 3011
Fall 2019
Exam 2 (B)
Time Limit: 120 Minutes

Name (Print): SOLUTION

Student ID: ______________________

Instructions:

- Do *not* begin or turn this page until you are instructed.
- Enter all requested information on the top and bottom of this page, and put your initials on the top of every page, in case the pages become separated.
- This exam contains 16 pages (including this cover page and the multiple choice answer sheet). Check to see if any pages are missing. There are 1 multiple choice problem (with 14 questions) and 3 short answer problems.
- The exam is closed book. You may *not* use your books, or any wireless device on this exam.
- You may use a calculator and one sheet of paper (size A4 or 8.5" by 11") with formulas or other notes on both sides. You may *not* share calculators or notes!
- Show all your work on each problem for full credit except multiple choice problems. The following rules apply:
  - *Organize your work*, in a reasonably neat and coherent way, in the space provided. Work scattered all over the page without a clear ordering will receive very little credit.
  - *Mysterious or unsupported answers will not receive full credit* for short answer problems. A correct answer, unsupported by calculations, explanation, or algebraic work will not receive full credit; an incorrect answer supported by substantially correct calculations and explanation may still receive partial credit.
  - If you need more space, use the back of the pages; clearly indicate when you have done this.

Honesty Statement and Pledge:

I have not given or received any aid or assistance to or from any other student in this course during the exam period. Everything I have written on this exam represents my own work and knowledge. I sign this knowing that infringements on the University's Academic Honest policy may result in failure or expulsion.

Signed By: ______________________ **Date:** ______________________

Problem I. (40 points) **Multiple Choice**

Choose the ONLY ONE correct answer for each question. Circle your answers to all questions in the answer sheet provided. (NO explanation is needed).

1. (3 points) Let p_I and p_M be the graduation rates from public high school in Iowa and Minnesota, respectively. Suppose a 95% Confidence Interval of $p_I - p_M$ is (0.003,0.089), then:

   (A) It is likely p_I is less than p_M.

   *** (B) **It is likely p_I is greater than p_M.**

   (C) It is likely p_I is equal to p_M.

   (D) It is likely p_M is greater than p_I.

2. (3 points) Suppose we want to test the claim that organic produce is more expensive than conventional produce. To test this, we sampled 25 different fruits and vegetables. For each fruit or vegetable, we took the price of its organic and conventional version. What type of test should we use to verify this claim?

   (A) Independent two sample z-test for difference of proportions.

   *** (B) **Matched pairs t-test for mean of difference.**

   (C) One sample z-test for proportion.

   (D) Independent two sample t-test for difference of means.

3. (3 points) Consider testing $H_0 : p_1 = p_2$ vs. $H_a : p_1 \neq p_2$, which of the following R commands produce the p-value if the test statistic is 1.85?

   (A) `pnorm(1-1.85)`

   (B) `pnorm(0.025)`

   (C) `pnorm(1.85)`

   *** (D) `2*(1-pnorm(1.85))`

4. (3 points) Which of the following scenario satisfies assumptions for an independent two sample t-test?

   *** (A) **A random sample of home runs hit by 60 players in the American League and 60 players in the National League.**

   (B) A sample of 15 volunteers' systolic blood pressures before and after taking medication.

   (C) A random sample of heads from coin flips of a quarter and a penny with 15 successes and 15 failures in each group.

   (D) A random sample of family sizes of families from Minnesota and Wisconsin, the sample size from each group was 4, and the data do not appear normal.

5. (3 points) Which of the following results in a small standard error of the mean?

*** (A) **A larger sample size**
(B) A larger population mean
(C) A larger population standard deviation
(D) None of the above.

6. (3 points) In a two-sided hypothesis with $H_a : \mu \neq \mu_0$, the p-value is found to be 0.04. Which of the following is true?

(A) The 90% confidence interval would contain μ_0
(B) The 95% confidence interval would contain μ_0
(C) The 99% confidence interval would NOT contain μ_0
*** (D) **None of the above**

7. (3 points) In a two-sided hypothesis with $H_a : \mu > \mu_0$ with 30 observations, p-value is found to be 0.02. Which of the following R commands produce the value(s) of test statistic.

(A) `qt(0.01, df=29), qt(0.99, df=29)`
(B) `1-pt(0.02, df=29)`
*** (C) `qt(.98, df=29)`
(D) `qt(0.02,df=29)`

8. (3 points) A diet pill company advertises that 75% of its customers lose 10 pounds or more within 2 weeks. You suspect the company of falsely advertising the benefits of taking their pills. Based on a random sample of 100 customers, 65 lost 10 pounds or more within 2 weeks. The value of test statistic is 2.5. Which of the following is the correct interpretation of the test statistic?

(A) If the true proportion of customers lose weight is 75%, the probability of making type I error is 2.5%.
(B) If the true proportion of customers lose weight is 75%, then the probability of obtaining as or more extreme test statistic that we observed is 2.5%.
*** (C) **If the true proportion of customers lose weight is 75%, the observed sample proportion 65% is 2.5 standard errors above the hypothesized proportion.**
(D) None of the above.

9. (3 points) A study considered a drug (a pill called Sumatriptan) for treating migraine headaches in a random sample of 60 subjects. The study observed each subject two times when he or she had a migraine headache. The subject received the drug at one time and a placebo at the other time. The order of the treatment was randomized. For each subject, the response was whether the drug or the placebo provided better pain relief. Which of the following is the correct set of hypotheses in testing whether the pill has any effect?

*** (A) $H_0 : p = 0.5$ **vs** $H_a : p \neq 0.5$

(B) $H_0 : \hat{p} = 0.5$ vs $H_a : \hat{p} \neq 0.5$

(C) $H_0 : \hat{p}_1 = \hat{p}_2$ vs $H_a : \hat{p}_1 \neq \hat{p}_2$

(D) $H_0 : p_1 = p_2$ vs $H_a : p_1 \neq p_2$

10. (3 points) Which error, Type I or Type II, would be considered more serious for decisions in the following tests?

i A medical diagnostic procedure, such as mammogram to detect breast cancer.

ii A trial to test a murder defendant's claimed innocence, when conviction results in the death penalty. In the American criminal justice system, a defendant is innocent until proven guilty.

(A) Type I error is worse in both Case i and ii.

(B) Type II error is worse in both Case i and ii.

(C) Type I error is worse in Case i and Type II error is worse in Case ii.

*** (D) **Type II error is worse in Case i and Type I error is worse in Case ii.**

11. (3 points) A $t-$distribution is NOT:

(A) more spread out than the standard normal distribution.

(B) having 0 as its mean.

(C) characterized by degrees of freedom.

*** (D) **used for constructing a CI for population proportion.**

12. (3 points) We are trying to estimate the mean of a population, denoted by μ. Suppose we draw a random sample from the population with size n, denoted by $\{X_1, \cdots, X_n\}$, and $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$ is the sample mean. We use $\bar{X}$ as an point estimator of μ. Which one of the following statements is true?

(A) Bias of $\bar{X}$ is random.

(B) $\bar{X}$ is a biased estimator of μ.

*** (C) $\bar{X}$ **is an unbiased estimator when we set** $n = 1$

(D) As we increase n, then bias of $\bar{X}$ will decrease.

13. (3 points) You are given R commands and outputs as following:

```
> val<-c(1,2,3,4,5,6,7,8,9)
> sd(val)
[1] 2.7386
> t.test(x=val, conf.level=0.97, alternative="two.sided")

One Sample t-test

data:  val
t = 5.4772, df = 8, p-value = 0.0005894
alternative hypothesis: true mean is not equal to 0
97 percent confidence interval:
 2.595667 7.404333
sample estimates:
mean of x
        5
```

Then, for random variable T the follows $t-$distribution with degrees of freedom 8, we know the probability $P(T_8 < t) = 0.985$, t is:

(A) 2.738

(B) 2.500

*** (C) **2.633**

(D) 2.404

14. (1 point) Did you remember to circle your answers on the last page (Multiple Choice Answer sheet)?

*** (A) **No, but I will now.**

*** (B) **Yes.**

*** (C) **Yes.**

*** (D) **Yes.**

Problem II. (20 points) Be sure to show all work for full credit.

A group of students at the University of Minnesota wants to study how students at the university commute. They randomly select 50 students on campus, and ask each of them two questions:
(1) Do you drive to school or not?
(2) How long do you spend on commuting everyday?

The following R commands and outputs may or may not be helpful.

```
qnorm(0.99) = 2.33
qnorm(0.98) = 2.05
qt(0.99,df=49) = 2.40
qt(0.98,df=49) = 2.11
```

1. (7 points) Suppose that 20 of those 50 students answered they drive to school. Construct a 98% confidence interval for p, which is the proportion of students who drive to school. Interpret the interval in context of the problem.(Note: show the formula, your calculation and result, and your interpretation.)

$\hat{p} = \frac{20}{50} = 0.4$

(1-α) CI has form: $\hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = 0.4 \pm 2.33\sqrt{\dfrac{0.4(1-0.4)}{50}}$
$= 0.4 \pm 2.33 \times 0.06928$
$= 0.4 \pm 0.1614$
$= (0.23857, 0.561427)$

Interpretation: We are 98% confident that the true proportion p is between (0.239,0.561).

2. (7 points) Suppose the sample mean commute time of those 50 students is 50 minutes, and the sample standard deviation s is 5 minutes. Construct a 98% confidence interval for μ, the mean commute time for all students in the university. Interpret the result in context of the problem. (Note: show the formula, your calculation and result, and your interpretation.)

(1-α) CI has form: $\bar{x} \pm t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}$
Plug in values, we get $50 \pm 2.40 \cdot \frac{5}{\sqrt{50}} = (48.303, 51.697)$
We are 98% confident that the mean commute time is between 48.303 and 51.697 minutes.
In the long run, we repeatedly collect a random sample of size 50 and construct 98% confidence intervals, 98% of those intervals will contain the true mean commute time.

The following R commands and outputs are copied from the previous page.

```
qnorm(0.99) = 2.33
qnorm(0.98) = 2.05
qt(0.99,df=49) = 2.40
qt(0.98,df=49) = 2.11
```

3. (6 points) Now suppose you want to conduct a new survey. How many students should you survey to get :
a. a 98% confidence interval for p with margin of error no more than 0.1. Use the sample proportion from Question 1 as an educated guess for the true population proportion.

b. a 98% confidence interval for μ with margin of error no more than 1 minute. Use the sample standard deviation from Question 2 as an educated guess for the true population standard deviation.

For p: $n = p^* \cdot (1 - p^*) \cdot (\frac{z_{\alpha/2}}{m})^2$
Plug in values, get $n = 0.4 \cdot 0.6 \cdot (2.33/0.1)^2 = 130.3$**, round up,** n **is at least 131.**

For μ: $n = (\frac{z_{\alpha/2} \cdot s^*}{m})^2$
Plug in values, get $n = (2.33 \cdot 5/1)^2 = 135.7$**, round up,** n **is at least 136.**

Problem III. (20 points) Be sure to show all work for full credit.

A random sample of 50 male students in current freshman class has an average height of 175.2 centimeters and standard deviation of 7.9 centimeters (cm). Let μ be the population mean height of male students. It is known that male height is normally distributed.

The following R commands and outputs may or may not be helpful.

```
qt(0.95,df=49) = 1.675
qt(0.90,df=49) = 1.300

pt(2.767, df=49,lower.tail=FALSE) = 0.004
pt(2.417, df=49, lower.tail = FALSE) = 0.0097

pnorm(2.767, lower.tail=FALSE) = 0.0028
pnorm(2.417, lower.tail = FALSE) = 0.0078
```

1. (5 points) Calculate a 90% confidence interval for the population mean height of male students. Use the sample information provided above. Remember to state assumptions. Interpret the result in context of problem.

Assumptions : Random sample, normal population / large sample size
t**-multiplier** $= t_{\alpha/2,n-1} = t_{0.05,49} \approx 1.675$

90% CI for μ is: $175.2 \pm 1.67 * 7.9/\sqrt{50} = (173.3342, 177.0658)$

We are 90% confident that the average height of female students is between 173.3342 cm and 177.0658 cm.
OR
We are 90% confident that the interval $(173.3342, 177.0658)$ **will contain the average height of female students.**
OR
In the long run, 90% of the 90% CI's will contain the average height of female student.

For Question 2 through 6: It is known that the average height of male students in the freshman class has historically been 172.5 cm. You want to test whether there has been a change in the average height of male students in freshman class. Use the sample information of the present freshman class provided in the Problem III description.

2. (2 points) State the assumptions.

Random sample, normal population or large sample size.

3. (3 points) State the null and alternative hypotheses. Is this test one-sided or two-sided?

H_0: $\mu = \mathbf{172.5cm}$
H_a: $\mu \neq \mathbf{172.5cm}$
Two sided

4. (2 points) Calculate the test statistic. Does the test-statistic follow the z-distribution or a t-distribution?

$t = \frac{175.2-172.5}{7.9/\sqrt{50}} = 2.417$
t-distribution.

The following R commands and output are copied from Problem III description.

```
qt(0.95,df=49) = 1.675
qt(0.90,df=49) = 1.300

pt(2.767, df=49, lower.tail=FALSE) = 0.004
pt(2.417, df=49, lower.tail = FALSE) = 0.0097

pnorm(2.767, lower.tail=FALSE) = 0.0028
pnorm(2.417, lower.tail = FALSE) = 0.0078
```

5. (3 points) Calculate the p-value. Draw the conclusion and interpret in context of problem. Use the significance level is 0.05.

p**-value**$= 2 \cdot P(T_{49} > 2.417) = 2 \cdot 0.0097 = 0.0194$

The p-value is less than 0.05, so we reject null hypothesis and conclude that $\mu \neq 172.5cm$.
There is enough evidence to conclude that there has been a change in the average height of male students.

6. (2 points) **(Multiple choice)** How will the probability of type II error change if the significance level is changed from 0.05 to 0.1?

(A) Increase
(B) Decrease
(C) Not change
(D) Unknown

(B) Decrease

The following R commands and outputs are copied from the Problem III description.

```
qt(0.95,df=49) = 1.675
qt(0.90,df=49) = 1.300

pt(2.767, df=49,lower.tail=FALSE) = 0.004
pt(2.417, df=49, lower.tail = FALSE) = 0.0097

pnorm(2.767, lower.tail=FALSE) = 0.0028
pnorm(2.417, lower.tail = FALSE) = 0.0078
```

7. (3 points) Suppose it is known that the true population standard deviation of current male height is 6.9 cm. This time, a random sample of 20 male students in freshman class has an average height of 175.2 centimeters (cm).

   Based on this information, you want to test the same hypothesis (whether there has been a change in the average height of male students) again.

   a. Calculate the test statistic.
   b. Is this test statistic z-distribution or t-distribution?

$z = \frac{175.2-172.5}{6.9/\sqrt{20}} = 1.75$

z-distribution.

Problem IV. (20 points) Be sure to show all work for full credit.

Sociologists at the University of Minnesota want to know if opinion of the importance of religion differs between students at private colleges and universities and students at public colleges and universities.
They took a random sample of 60 students from private colleges and universities and a random sample of 65 students from public colleges and universities. From the sample, 45 out of the 60 students from private schools felt religion was important and 38 out of the 65 students from public schools felt religion was important.

Let p_{pr} and p_{pu} denote the proportion of private school, public school students who say that religion is important to them.

```
qnorm(0.975) = 1.96, qnorm(0.95) = 1.65, qnorm(0.9) = 1.28
prorm(0.95) = 0.83, pnorm(0.6633) = 0.75, pnorm(-1.955) = 0.0253
```

1. (2 points) What are the standard errors of $\hat{p}_{pr}$ and $\hat{p}_{pu}$?

$$\mathbf{se}(\hat{p}_{pr}) = \sqrt{\frac{\hat{p}_{pr}(1-\hat{p}_{pr})}{n_{pr}}} = \sqrt{\frac{(45/60)(1-45/60)}{60}} = 0.0559$$
$$\mathbf{se}(\hat{p}_{pu}) = \sqrt{\frac{\hat{p}_{pu}(1-\hat{p}_{pu})}{n_{pu}}} = \sqrt{\frac{(38/65)(1-38/65)}{65}} = 0.0611$$

2. (5 points) Construct and interpret a 95% confidence interval of $p_{pr} - p_{pu}$. In your work, include the point estimate of the difference, the z-multiplier, and the standard error.

95% CI:

$$
\begin{aligned}
&(\hat{p}_{pr} - \hat{p}_{pu}) \pm z_{0.05/2}\sqrt{\frac{\hat{p}_{pr}(1-\hat{p}_{pr})}{n_{pr}} + \frac{\hat{p}_{pu}(1-\hat{p}_{pu})}{n_{pu}}} \\
=&(\frac{45}{60} - \frac{38}{65}) \pm 1.96\sqrt{\frac{45/60(1-45/60)}{60} + \frac{38/65(1-38/65)}{65}} \\
=&(0.750 - 0.585) \pm 1.96\sqrt{\frac{0.75(1-0.75)}{60} + \frac{0.585(1-0.585)}{65}} \\
=&(0.0026, 0.3273)
\end{aligned}
$$

We are 95% confidence the proportion of private school students that feel religion is important is between 0.0026 and 0.3273 higher than students at public schools.

3. (2 points) **(Multiple Choice)** What is the center of this confidence interval?

   (A) $p_{pr} - p_{pu}$
   (B) $n_{pr} - n_{pu}$
   (C) 0
   (D) $\hat{p}_{pr} - \hat{p}_{pu}$

(D) $\hat{p}_{pr} - \hat{p}_{pu}$ **is at the center.**

4. (7 points) Conduct the five steps of a level 0.05 test that the proportion of students at private schools believe religion is important is greater than the proportion of students at public schools. That is test, $p_{pr} > p_{pu}$ at level 0.05.

1. **Assumptions: These were random samples and there are at least 5 successes and 5 failures in each group.**
2. **Hypotheses:** $H_0 : p_{pr} = p_{pu}$ **vs.** $H_0 : p_{pr} > p_{pu}$ **or** $H_0 : p_D = 0$ **vs.** $H_0 : p_D > 0$**, where** $p_D = p_{pr} - p_{pu}$**.**
3. **Test Statistic:**

$$\begin{aligned}\hat{p} &= \frac{45+38}{60+65} = 0.664 \\ z &= \frac{0.75 - 0.585 - 0}{\sqrt{0.664(1-0.664)(1/60+1/65)}} \\ &= 1.95\end{aligned}$$

4. **p-value:** $P(Z > z) = P(Z > 1.95) = 1 - P(Z < 1.95) = 1 - \texttt{pnorm}(1.95) \approx$ `pnorm(-1.955)` $=$ **0.0253.**
5. **Conclusion: The** p**-value is less than 0.05, we therefore have evidence to conclude the proportion of students at private schools believe religion is important is greater than proportion of students at public schools who believe religion is important.**

5. (2 points) Based on your conclusion from the previous question, what type of error could you be making? Why?

Type I error because we rejected H_0**.**

6. (2 points) Would we reject the null hypothesis at level 0.1? Why?

Yes, the p-value is less than 0.1.

Name: ______________________

Lecture Section: 001 (9:05 am) Yang | 005 (8:00 am) Burrell | 009 (12:20 pm) Park | 013 (10:10 pm) Park | 017(5:10 pm) Xue
(**Circle One**)

Question	Answer			
1	A	B	C	D
2	A	B	C	D
3	A	B	C	D
4	A	B	C	D
5	A	B	C	D
6	A	B	C	D
7	A	B	C	D
8	A	B	C	D
9	A	B	C	D
10	A	B	C	D
11	A	B	C	D
12	A	B	C	D
13	A	B	C	D
14	A	B	C	D

Please do NOT write in the following table. This is for grading purpose only!

Question	I (40)	II (20)	III (20)	IV (20)	100
Score					