Math 247: Test 3
(Wright, Fall 2018)

Name: KEY

Class Time: ______________________

Part I: (50 points) Do this part individually. You will turn in your own copy of this work.

- *This exam is due at the beginning of class on Tuesday, November 4, 2018. Be sure that all answers are written in your own words; i.e., do not write verbatim the same answer as another student.*
- *You are welcome to work with other students in the class (and I encourage you to do so!) but please do not ask tutors or other instructors to answer the questions on the exam for you.*

Scoring will be based on neatness and organization of your work, accuracy, and thoughtful, well-written answers using complete sentences!

1. (3 pts) If you wanted to obtain a Simple Random Sample (SRS) of 30 Business majors at Cuesta College, what would you have to do? Assume there are about 400 students who are Business majors at Cuesta and use that number to describe the process you would have to go through.

   To get an SRS of 30 business majors, you would have to number the 400 business majors from 1 to 400, then use a Random Number Generator to select 30 of these numbers. Your sample would consist of these 30 students.

   If instead you just used 30 students in your Accounting class as the sample, what would this sample be called?

   Sample of Convenience

   If you wandered around campus asking people whether they were Business majors, then using those people as your sample, would this be a raondom sample? Why or why not?

   No. Not every business major would be equally likely to be chosen. You would exclude distance students from your sample and also be biased in who you chose to talk to.

2. (3 pts) Deepak hosts a podcast and he is curious how much his listeners like his show. He decides to conduct an online poll. He asks his listeners to visit his website and participate in the poll. He finds 89% of 200 people who responded say they love his show.

   OK voluntary response bias

   What type of bias does this survey have? (Specific name of type of bias)

   Self-selection bias — the researcher (Deepak) didn't choose the subjects; they chose themselves.

   Is the bias positive (overestimate) or negative (underestimate)? Explain your answer.

   The bias will most likely be positive, since the people who take the time to respond will likely have strong feelings in order to bother to go to the website and take the survey.

3. (6 pts) A company with a fleet of 150 cars found that the emissions systems of 7 out of the 22 they randomly chose and tested failed to meet pollution control guidelines. They want to use this sample to test whether more than 20% of the fleet is out of compliance.

2 (a) What would the hypotheses be? (Write them in words and using appropriate symbols).

$H_0: p = .20$ The proportion of cars that failed to meet guidelines (=success) is .20.

$H_a: p > .20$ The proportion is more than .20.

3 (b) If you wanted to use the z-Test for One Proportion, would this scenario satisfy the conditions for the test? Write down each of the conditions that must be checked and explain why or why not it is satisfied.

$\boxed{n=22}$ $\boxed{p=.20}$ $\boxed{1-p=.80}$

1. Random Sample? Yes, stated in problem
Independent observations? yes, one car failing the test won't affect another car's result.

2. Large Sample? Expected successes ≥ 10?
$E = np = 22(.20) = 4.4 \geq 10$ NO
use $\hat{p}$ $(-\frac{1}{2})$ NO!
Expected failures ≥ 10?
$E = n(1-p) = 22(.8) = 17.6 \geq 10$ yes

3. Large population?
Pop of cars $\geq 10n = 10(22) = 220$ cars NO! The population is too small!

1 (c) Is it appropriate to use the z-Test for 1 Proportion in this study? Why or why not?

No! The condition of Large Sample is not met AND the population is too small (see work in (b)) If the conditions aren't met, the results of a Z-Test aren't valid.

4. (3 pts) Donated blood is tested for infectious diseases and other contaminants. Since most donated blood is safe, it saves time and money to test batches of donated blood rather than test individual samples. A certain test is performed to see if a certain toxin is present, and the entire batch is discarded if the toxin is detected. This is like using a null and an alternative hypothesis to determine whether to discard or keep the batch. The hypotheses being tested could be stated as:

Type I: Reject H_0 when it's true
$H_a \Rightarrow$ batch has toxin, throw it out!
$H_0 \Rightarrow$ no toxin in reality!

*H*0: The batch does not contain the toxin.
*H*a: The batch contains the toxin.

What would be the consequence of a Type I error in this context?Choose 1 answer:

(a) The batch is discarded when it actually contains the toxin.
(b) The batch is discarded when it actually doesn't contain the toxin.
(c) The batch is kept when it actually contains the toxin.
(d) The batch is kept when it actually doesn't contain the toxin.

5. (3 pts) Regulations from the Environmental Protection Agency say that soil used in play areas should not have lead levels that exceed 400parts per million (ppm). Before beginning construction at a new site, an agent will take a sample of soil and run a significance test on the mean lead level in the soil. If the mean lead level in the sample is significantly higher than 400 ppm then the soil is deemed unsafe and construction cannot continue. Here are the hypotheses for this test:

Type II: Fail to reject H_0 when H_0 is false
Conclude soil is safe when it's not

$H0$: $\mu = 400$ ppm (soil is safe):
Ha: : $\mu > 400$ ppm (soil is unsafe)
(where μ is the mean lead level in the soil at the new site).

construction continues — soil is not safe

What would be the consequence of a Type II error in this setting? Choose 1 answer.

(a) Construction continues when the soil is actually safe.
(b) Construction stops when the soil is actually safe.
(c) Construction continues when the soil is unsafe.
(d) Construction stops when the soil is actually unsafe.

6. (10 pts) Nanda saw a report that claimed 57% of US adults primarily get their news from television. She was curious how faculty at Cal Poly get the news, so she surveyed a random sample of 100 Cal Poly professors and made a 95% confidence interval to estimate the proportion of professors who get their news from TV.
Her resulting interval was (0.44, 0.60).

True, all, pop $1\frac{1}{2}$ pts.

2 (a) Interpret the confidence interval in the context of the problem.
We are 95% confident that the true (population) proportion of all Cal Poly profs who get their news from television is between 44% and 60%.

1 (b) Graph the confidence interval on a number line.

p = proportion who get news from T.V.

2 (c) How many people in Nanda's sample get their news from television? (Hint: what is the center of the confidence interval?)
$\hat{p}$ is in the middle of the confidence interval
$\hat{p} = \frac{.44 + .60}{2} = \frac{1.04}{2} = .52$ | 52% of 100 = $\boxed{52 \text{ people}}$ in the Sample get news from T.V.

2 (d) What is the margin of error? (Again, reason this out from the confidence interval)
Margin of Error = $M = .08$ (Find using either $M = .60 - .52 = .08$
or $M = .52 - .44 = .08$

2 (e) What conclusion can Nanda make based on this confidence interval? Choose 1 answer.
a. The proportion of Cal Poly faculty who get their news from television is significantly higher than the proportion of overall US adults.
b. The proportion of Cal Poly faculty who get their news from television is significantly lower than the proportion of overall US adults.
c. The proportion of Cal Poly faculty who get their news from television is not significantly different from the proportion of overall US adults.
(Because .57 is captured by the confidence interval

7. (14 pts) A survey conducted five years ago by the health center at a college showed that 15% of the students smoked at the time. After implementing a smoking ban, a new survey was conducted to determine whether the percentage of smokers percentage has changed. A random sample of 200 students from the college was taken, and it was found that 21 of them smoke. Do these data provide evidence to suggest that the percentage of students who smoke now has changed after the implementation of the smoking ban? Use a significance level of .10.

2 Hypotheses (in words and using correct symbols):

$H_0: p = .15$ The proportion of smokers now is the same as 5 years ago (15%)

$H_a: p \neq .15$ The proportion of smokers has changed.

Parking Lot

$P_0 = .15$
$n = 200$ students
$X = 21$

1 Sample proportion (use correct notation):

$\hat{p} = \frac{21}{200} = .105 = 10.5\%$

Assume the conditions are met for performing a z-Test for 1 Proportion. Perform the hypothesis test using StatCrunch. Write the results below:

Hypothesis test results: (round to 3 decimal places)

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	21	200	0.105	.0252	-1.782	.0747

1

1 Find the Standard Error by hand (show work!) and confirm that it matches the results from StatCrunch.

$SE = \sqrt{\frac{P_0(1-P_0)}{n}} = \sqrt{\frac{.15(.85)}{200}} = .0252$ matches!

$\hat{p}$ instead of P_0 $-\frac{1}{2}$

1 The SE value is the Standard deviation of the Sampling Distribution of p-hat.

2 Find the Test statistic, z, by hand and confirm that it matches the results from StatCrunch.

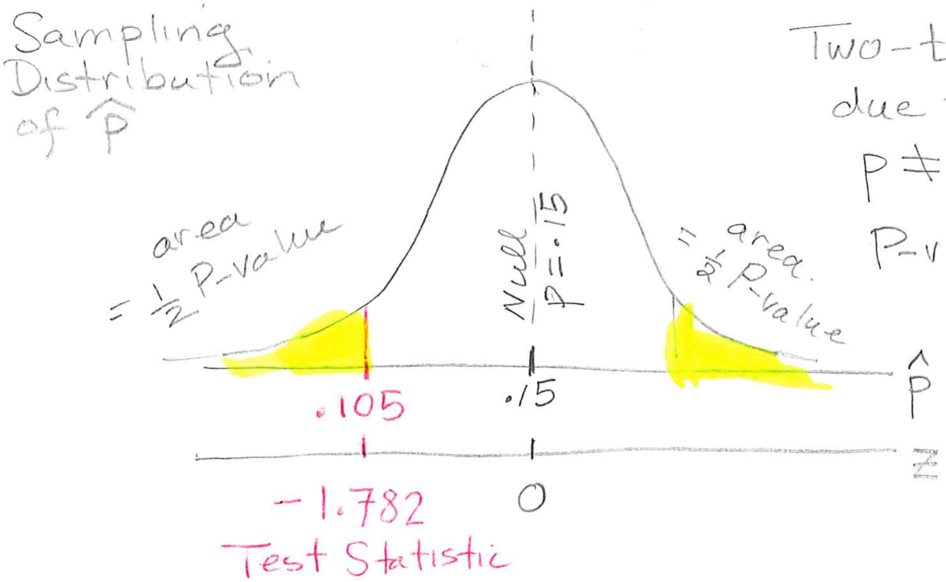
$Z = \frac{\hat{p} - P_0}{SE} = \frac{.105 - .15}{.0252} = -1.786$

note: the difference between this value and StatCrunch's is due to rounding the SE value

#7: continued

Sketch the sampling distribution of $\hat{p}$, with the null hypothesis, sample proportion, and P-value clearly illustrated and labeled. Put the z-axis under the $\hat{p}$ axis, and mark where the Test Statistic is.

3



Two-tailed test due to Ha being $p \neq .15$

P-value = area in BOTH tails = .0747

2 Conclusion of Hypothesis Test (give a thorough answer!):

Reject H_0: P-value = .0747 < .10 = α (level of significance)
Accept Ha There is a 7.47% chance we would have observed 21 smokers in our sample of 200 students IF the proportion of ALL students smoking was still .15 = 15%.
We therefore reject the null and conclude there is sufficient evidence at the .10 level of significance to say there has been a significant change in the proportion of students who smoke.

2 You should have found that the result was statistically significant at the .10 significance level. Can we then say that the smoking ban *caused* a decrease in smoking for the students at this university? Why or why not (I'm looking for a very specific answer here!)

We can't say the ban caused the change in smoking since this is an observational study. The researchers would have to conduct a controlled experiment to determine cause-and-effect.

8. (8 points) (a) Using the same data as #7 find a 90% confidence interval by hand for the proportion of students who smoke. Confirm your results using StatCrunch.

3

"A survey conducted five years ago by the health center at a college showed that 15% of the students smoked at the time. After implementing a smoking ban, a new survey was conducted to determine whether the percentage of smokers percentage has changed. A random sample of 200 students from the college was taken, and it was found that 21 of them smoke."

CI: $\hat{p} \pm z^* SE_{est}$

$SE_{est} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\frac{1}{2}$pt

$= \sqrt{\frac{.105(.895)}{200}}$

$= .0217$

From before

$\hat{p} = .105$

$n = 200$

For 90% confidence level,

$z^* = 1.645$

$= .105 \pm 1.645(.0217)$

$= .105 \pm .036$

$= (.105 - .036, .105 + .036)$

$= (.069, .141)$

	L. Limit	U. Limit
StatCrunch:	.0693	.1407

(matches the "by hand" result.

2 (b) Interpret the confidence interval in words in the context of the problem.

We are 90% confident that the proportion of all students who smoke is between .069 = 6.9% and .141 = 14.1%

3 (c) Explain how the confidence interval supports the conclusion of the hypothesis test, again, in the context of the problem.

Since the confidence interval does not capture .15 = 15% we can, at the 90% confidence level, exclude the possibility that the percentage of students smoking has remained at 15% and conclude the true population proportion has decreased significantly to anywhere between 6.9% and 14.1%