February 12, 2018

STATISTICS 200 EXAM 1

PRINT NAME________________________ __________________ KEY
(Last name) (First name) (net ID, not UIN)

Circle Section: ONL or L1 2:00-3:20 pm

Write answers in appropriate blanks. When no blanks are provided **CIRCLE** your answers.
SHOW WORK when requested.

No notes or books are allowed. Calculators (including graphing ones) are allowed.
Do not use your own scrap paper. If you need some, ask me.

For ALL Questions using the normal table: You may "round" z scores and percents to fit the closest line on the normal table and you may round percents on the table to the nearest whole number.

Make sure you have all 5 pages (7 problems) including the Normal table.

DO NOT WRITE BELOW THIS LINE

The numbers written in each blank below indicate how many points you missed on each page. The numbers printed to the right of each blank indicate how many points each page is worth.

Page 1 ______ 25

Page 2 ______ 27

Page 3 ______ 28

Page 4 ______ 20

Formulas (you might not need them both.)

$$SD_{errors} = \sqrt{1-r^2} * SD_y$$

$$SE_{slope} = \frac{SD_{errors}}{\sqrt{n} * SD_x} = \frac{\sqrt{1-r^2}}{\sqrt{n}} * \frac{SD_y}{SD_x}$$

Score ________

Scores will be posted on Compass Tuesday night.

There's lecture tomorrow but not on Friday.

Stat 200 Exam 1 February 12, 2018

Question 1 pertains to a screening test for prostate cancer *(10 points total)*

If a patient has prostate cancer there's a 90% chance that the screening test will correctly give a positive result. But if the patient does not have prostate cancer, there's a 20% chance that the test will *incorrectly* give a positive result.

→ Let's put this situation into the context of a Hypothesis test.

a) *(2 points)* The conventional null hypothesis (H_0) is that the patient …
  [✓ **i)**] does not have prostate cancer
  ii) has prostate cancer
  iii) may have prostate cancer at significance level 5%
  iv) has a 10% chance of having prostate cancer

b) *(2 points)* A Type II error can only occur ….
  [✓ **i)**] when the patient has prostate cancer.
  ii) when the patient does not have prostate cancer.
  iii) Either when the patient has prostate cancer and the significance level is set too high or when the patient does not have cancer and the significance level is set too low.

c) *(2 points)* If a person has prostate cancer, what's the probability he'll test negative? 10 %
What type of error would this be? ***Circle one:*** **i)** Type I error [✓ **ii)** Type II]

d) *(2 points)* If a person does not have prostate cancer, what's the probability he'll test positive? 20 %
What type of error would this be? ***Circle one:*** [✓ **i)** Type I error] **ii)** Type II

e) *(2 points)* If you change the cut-off of the screening test to decrease the probability of false negatives what would happen to the probability of false positives? [✓ **i)** Increases] **ii)** Decreases **iii)** Stays the Same

Question 2 *(8 points)* A significance test is performed to analyze the results of a randomized experiment to determine whether students learn more or less from watching a lecture online compared to attending the same lecture in person. Subjects are randomly assigned to treatment (online lecture) and control (in person lecture) and then given the same exam afterwards.

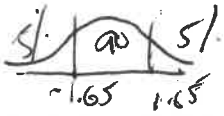
1 **a)** Fill in the blanks to complete the null and alternative hypotheses below:
H_0: The difference in mean exam scores between the treatment and control groups in the population = 0
  Choose one: i) > ii) < [✓ iii) =] iv) ≠
H_A: The difference in mean exam scores between the treatment and control groups in the population ≠ 0
  Choose one: i) > ii) < iii) = [✓ iv) ≠]

1 **b)** A significance level of $\alpha = 0.02$ means when the null is true the probability of making a Type I error= __________.
Circle one: **i)** 0% **ii)** 1% [✓ **iii)** 2%] **iv)** 4% **v)** 96% **vi)** 98% **vii)** not enough info

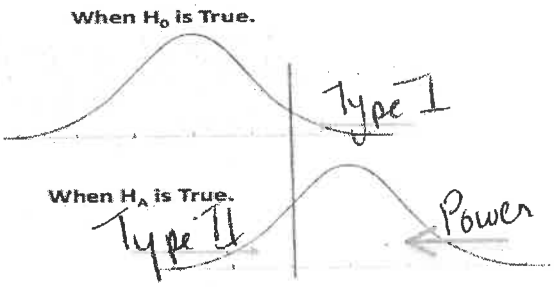
1 and when the null is false the probability of making a Type II error = _________.
Circle one: **i)** 0% **ii)** 1% **iii)** 2% **iv)** 4% **v)** 96% **vi)** 98% [✓ **vii)** not enough info]

2 **c)** If we set α=0.05 (null cut-off at 5%) for a 2-sided H_A then the critical value of our test-statistic, Z*= _______
Choose closest answer. i) 0.85 ii) 1.3 iii) 1.65 [✓ iv) 2] v) 2.35 vi) 2.6

2 **d)** Repeat (c) above with a 1-sided H_A keeping all else the same. Choose closest answer.
i) 0.85 ii) 1.3 [✓ iii) 1.65] iv) 2 v) 2.35 vi) 2.6



Question 3 *(7 pts.)* Look at the histograms below. *(3 points)* Label the 3 areas (indicated by arrows) that represent Type I and Type II errors and Power by writing "Type I", "Type II", or "Power" above each arrow.



(4 points) Which of the following will increase the Power of the test? Circle either "yes" or "no".

a) Increasing the probability of a Type I error [✓ i) Yes] ii) No
b) Increasing D, the effect size (H_A-H_0) [✓ i) Yes] ii) No
c) Increasing the sample size [✓ i) Yes] ii) No
d) Increasing the SD i) Yes [✓ ii) No]

Stat 200 Exam 1 February 12, 2018

Question 4 *(27 points total)*

Part I (15 pts.) Suppose the average number of times students reported texting during lecture is 5 with a SD = 3. I think the true average is 6.6 texts or higher, so I decide to choose 6.6 texts for my alternative hypothesis and assume the SD is still 3 texts. I decide to do a significance test by randomly choosing **25** students to be carefully observed in lecture and record how many times they text.

a) (3 pts.)Assuming the H_0 to be true, I'd expect the sample average to be 5 texts with a SE_{avg} = 0.6 texts. (Show work for SE.)

$3/\sqrt{25} = 0.6$

b) (3 pts.) Assuming H_A to be true, I'd expect the sample average to be = 6.6 texts with a SE_{avg} = 0.6 texts

C.e

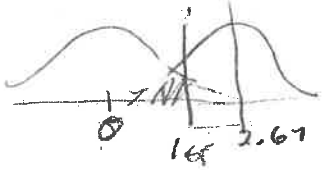
c) (2 pts) What is the effect size ? D = 1.6 texts C.e a+b

66.-5

d) (2 pts.) What is the effect size in Standard Units? D_z = 2.67. *Show work.*

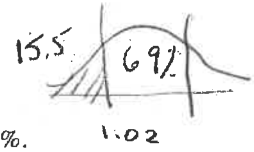
$D = D_z \cdot SE$

$D_z = 1.6/0.6$ C.e. a+c



e) (5 pts.) If I set the significance level α = 5%, what is the Power of the test?

2pt. i) First find $|Z_\alpha|$. $|Z_\alpha|$ = 1.65.



3pts ii) So $|Z_\beta|$ = 1.02 which means β= 15.5 %, so Power = 84.5 %.

c.c. 2.67 − 1.65 = 1.02

accept 15-16 accept 84-85

Question 4 Part II (12 pts.)

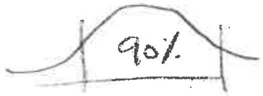
Now suppose we keep the same α = 5%, same SD = 3 texts and the same effect size (D) as before, but we want power=95%.

a) (2 pts.) Will we need a larger sample size than in Part I? [✓ i) Yes] ii) No

Calculate what size sample you'll need.

b) (4 pts.) First, compute β and $|Z_\beta|$. β = 5 %, and $|Z_\beta|$ = 1.65

100-95 = 5



c) (2 pts.) What is the effect size in Standard Units? D_z = 3.3. *Show work.*

C.e from b

1.65 + 1.65

C.e from Part I+I

d) (2pts.) SE_{avg} = ______ (Round to 3 decimal places). *Show work.*

more space

$1.6 = 3.3 \cdot SE \rightarrow SE = \frac{1.6}{3.3} = 0.485$

e) (2 pts.) How large an n will give us that small a SE_{avg}? *Show work.*

n = 38.26

Round n to 2 decimal places, not up to ~~nearest~~ whole number as usual.

$0.485 = \frac{3}{\sqrt{n}}$

C.e from d

$\sqrt{n} = \frac{3}{0.485}$

Stat 200 Exam 1 February 12, 2018

Question 5 *(28 points total)*

	Avg	*SD*	
Husband	180 lbs.	30 lbs.	*r = 0.2*
Wife	150 lbs.	20 lbs.	

How do the weights of husbands and wives correlate? Suppose a random sample of 36 married couples from Illinois yielded results given in the table. (Assume the population scatter plot follows a linear trend.)

a) (2 pts.) Our best estimate of the **slope** of the regression equation for predicting husbands' weights from their wives' weights for all married couples in Illinois is β = 0.3 *Show work.* $0.2\left(\frac{30}{20}\right) = 0.3$

b) (2 pts.) SE_{slope} = 0.245 *Show work. Round to 3 decimal places.* $SE_{slope} = \frac{\sqrt{1-0.04}}{6} \cdot \left(\frac{30}{20}\right) = 0.245$

c) (2 pts.) Find a **92% CI** for the population slope using the normal curve. (Fill in 2 blanks below)

92% Confidence Interval = 0.3 ± 1.75 * SE_{slope}

Ce from a

d) (6 pts.) Suppose you wanted to use SE^+ and the t-curves instead of the normal curve to compute confidence intervals. How would you adjust your answer to part c above?

i. (2pts.) To find SE^+_{slope}, you'd multiply SE_{slope} by 1.03. $\sqrt{\frac{36}{34}}$ = Ok if give $\sqrt{\frac{36}{34}}$ instead of decimal
Fill in blank with a number, round to 2 decimal places.

ii. (2 pts.) To find the critical value of t (called t*) corresponding to a 92% CI, you'd look at the t curve with how many degrees of freedom? 34

iii. (2pts.) How would t* compare to z*, the critical value of z, used in part c above? **Choose one.**
[✓ a) t* > z*] b) t* < z* c) t* = z*

e) (2 pts.) Can you be 92% confident that there is a non-zero correlation between husbands' and wives' weights in the population?
i) Yes, because the above interval includes 0.
[✓ ii) No, because the above interval includes 0.]
iii) No, because the above interval does not include 0.

f) (6 pts.) Now let's compute the Z test-statistic for testing H_0: **β = 0** against H_A: **β ≠ 0**

Round to 2 decl

i) Z= 1.22 ii) p-value= 23 % iii) Conclusion: Reject Null at α = 0.05?
Show work. Circle one: Yes [✓ No]

$$\frac{r}{\sqrt{1-r^2}}\sqrt{n} = \frac{0.2}{\sqrt{1-0.04}} \times 6 = 1.22$$

g) (2 pts) Now, let's do a t- test. To compute the t-statistic you'd multiply the Z stat by....
i) $\sqrt{\frac{36}{34}}$ [✓ ii) $\sqrt{\frac{34}{36}}$] iii) $\sqrt{\frac{36}{35}}$ iv) $\sqrt{\frac{35}{36}}$

h) (2 pts.) To compute the p-value you'd have to look at the t curve with degrees freedom = 34.

i) (2 pts.) The p-value would be ________ than the one computed in (fii) .
i) smaller [✓ ii) larger] iii) exactly the same

j) (2 pts.) If the alternative was H_A: **β > 0** instead of **β ≠ 0** would you get the same p-values?
i) Yes ii) No, they'd be twice the size. [✓ iii) No, they'd be half the size.]

Question 6 *(14 pts.)*
A study published in the Feb 18, 2004 issue of the Journal of the American Medical Association compared pharmacy and medical records of 10,219 women and found that women who filled 25 or more prescriptions for antibiotics over a 17 year period received breast cancer diagnoses at twice the rate as those who took no antibiotics. The study concluded that high antibiotic usage increases one's risk of breast cancer.

a) *(2 pts.)* Which of the following statements best describes this study? ***Circle one:***
- i) This was a randomized controlled experiment without a placebo.
- [✓ ii)] This was an observational study with controls.
- iii) This was a randomized controlled double-blind experiment.
- iv) This was a non-randomized controlled experiment with a placebo.

b) *(2 pts.)* Based on the results of this study alone, which of the following statements is best? ***Circle one.***
- i) High antibiotic use causes an increased risk of breast cancer.
- [✓ ii)] High antibiotic use is *associated* with and *may* cause increased breast cancer risk.
- iii) High antibiotic use is *associated* with but does *not* cause increased breast cancer risk.
- iv) Having cancer is likely to cause increased use of antibiotics.

c) *(8 pts.)* Below are either confounders that mix up the study, causal links that explain the conclusion, or neither.
Circle which is which based only on the given information.
- i) Age of first pregnancy- women who have their first child after the age of 35 are more likely to get breast cancer.
  a) Confounder b) Causal Link [✓ c) Neither]
- ii) Destruction of Protective Bacteria- antibiotics kill healthy bacteria that may help prevent breast cancer.
  a) Confounder [✓ b) Causal Link] c) Neither
- iii) Underlying Immune Problem- a weak immune system leads both to frequent infections necessitating antibiotics *and* also to a higher cancer risk.
  [✓ a) Confounder] b) Causal Link c) Neither
- iv) Regular Check-ups- Women who regularly go to the doctor are both more likely to be prescribed antibiotics and more likely to receive a breast cancer diagnosis (especially for slow growing cancers that are unlikely to lead to serious health problems.)
  [✓ a) Confounder] b) Causal Link c) Neither

d) *(2 pts.)* Suppose the researchers thought that income was a possible confounder since high income women tend to take more antibiotics and tend to get more breast cancer. To separate out the effects of income from the effects of antibiotics researchers should ... **Circle one:**
- i) split the data into high, middle and low income groups and compare the antibiotic usage between the 3 groups.
- [✓ ii)] split the data into high, middle and low income groups and compare the cancer rate of those who took a lot of antibiotics to those who took no antibiotics within each group.
- iii) split the data into high and low antibiotic users and compare the cancer rates between the groups.
- iv) split the data into 2 groups—breast cancer and no breast cancer and compare antibiotic usage between the 2 groups.

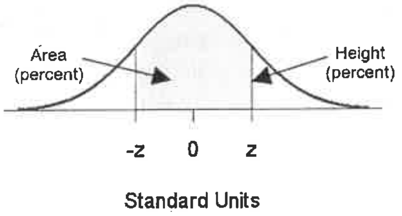
Question 7 *(6 pts,)*
Two experiments were done comparing the effects of listening to classical music versus rap music while studying. All the students in both experimental designs were given an identical 2-hour lesson and then allowed time to study for a short exam.
In Design A students themselves chose to study either listening to classical or rap.
In **Design B the students were randomly assigned** to study either listening to classical or rap.
Design A found that the classical study group scored significantly higher on the exam than the rap group did. Design B found no significant difference in exam scores between the 2 groups. **The overall exam average in both designs was the same.**

a) Which design had randomized controls? A only [✓ B only] Both Neither

b) Which design is more likely to have confounders? [✓ A] B Both are equally likely

c) Which conclusion is best supported by the evidence? ***Circle one***
- i) Students learn better when they are able to choose their own music while studying.
- [✓ ii)] Students who choose classical are different in more ways than just their musical tastes than students who choose rap.
- iii) Classical music seems to enhance learning better than rap music.

STANDARD NORMAL TABLE



z	***Area***		*z*	***Area***		*z*	***Area***
0.00	0.00		1.50	86.64		3.00	99.730
0.05	3.99		1.55	87.89		3.05	99.771
0.10	7.97		1.60	89.04		3.10	99.806
0.15	11.92		1.65	90.11		3.15	99.837
0.20	15.85		1.70	91.09		3.20	99.863
0.25	19.74		1.75	91.99		3.25	99.885
0.30	23.58		1.80	92.81		3.30	99.903
0.35	27.37		1.85	93.57		3.35	99.919
0.40	31.08		1.90	94.26		3.40	99.933
0.45	34.73		1.95	94.88		3.45	99.944
0.50	38.29		2.00	95.45		3.50	99.953
0.55	41.77		2.05	95.96		3.55	99.961
0.60	45.15		2.10	96.43		3.60	99.968
0.65	48.43		2.15	96.84		3.65	99.974
0.70	51.61		2.20	97.22		3.70	99.978
0.75	54.67		2.25	97.56		3.75	99.982
0.80	57.63		2.30	97.86		3.80	99.986
0.85	60.47		2.35	98.12		3.85	99.988
0.90	63.19		2.40	98.36		3.90	99.990
0.95	65.79		2.45	98.57		3.95	99.992
1.00	68.27		2.50	98.76		4.00	99.9937
1.05	70.63		2.55	98.92		4.05	99.9949
1.10	72.87		2.60	99.07		4.10	99.9959
1.15	74.99		2.65	99.20		4.15	99.9967
1.20	76.99		2.70	99.31		4.20	99.9973
1.25	78.87		2.75	99.40		4.25	99.9979
1.30	80.64		2.80	99.49		4.30	99.9983
1.35	82.30		2.85	99.56		4.35	99.9986
1.40	83.85		2.90	99.63		4.40	99.9989
1.45	85.29		2.95	99.68		4.45	99.9991