STATISTICS 200 EXAM 1

PRINT NAME_____ _____ _____
(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 6 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 _____ 18

Page 2 _____ 28

Page 3_____ 18

Page 4 _____ 10

Page 5 _____26

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 Wednesday night and exams will be returned in class on Thursday.

Stat 200 Exam 1 September 26, 2017

Question 1 pertains to a pre-natal screening test for detecting Down Syndrome.

(10 points total) Screening tests aren't perfect. If a fetus has Down Syndrome there is a 90% chance the test will detect it, but when the fetus does not have Down Syndrome, there's a 7% chance that the screening test will incorrectly return 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 fetus …
- **i)** does not have Down Syndrome
- **ii)** has Down Syndrome
- **iii)** has a 50% chance of having Down Syndrome
- **iv)** has a 90% chance of not having Down Syndrome

b) *(2 points)* A Type I error can only occur ….
- **i)** when the H_0 is false.
- **ii)** when the H_0 is true.
- **iii)** when the significance level is set above 5%.
- **iv)** when the significance level is set below 5%.

c) *(2 points)* What's the probability of a fetus with Down Syndrome testing negative? _____%
What type of error would this be? ***Circle one:*** **i)** Type I error **ii)** Type II

d) *(2 points)* What's the probability of a fetus without Down Syndrome testing positive? _____%
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 if a new drug works either better or worse than the old one. Subjects are randomly assigned to treatment (the new drug) and control (the old drug).

a) Fill in the blanks to complete the null and alternative hypotheses below:
H_0: The difference in mean responses between the treatment and control groups in population ___ 0
Choose one: i) > ii) < iii) = iv) ≠
H_A: The difference in mean responses between the treatment and control groups in population ___ 0
Choose one: i) > ii) < iii) = iv) ≠

b) When we set the significance level $\alpha = 0.05$ we are setting the probability of making a Type I error= _____
Circle one: **i)** 0% **ii)** 1% **iii)** 2.5% **iv)** 5% **v)** 95% **vi)** 99% **vii)** not enough info

and the probability of making a Type II error = _____.
Circle one: **i)** 0% **ii)** 1% **iii)** 2.5% **iv)** 5% **v)** 95% **vi)** 99% **vii)** not enough info

c) If we set $\alpha = 0.10$ (null cut-off at 10%) 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

d) Repeat (c) above with a 1-sided $\mathbf{H_A}$ 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 *(28 points total)*

Part I (16 pts.) Do men lie about their height on dating sites? On a popular dating site the average height of the thousands of male members is **71" with a SD=3"** and the heights roughly followed the normal curve. I believe the men are inflating their heights and their real average is only **69.5" or lower with the same SD= 3"**. I decide to do a significance test by randomly sampling **16** men from the dating site.

a) (3 pts.)Assuming the null to be true, I'd expect the sample average to be _____ inches with a $\mathbf{SE_{avg}}$ = _____inches. (Show work for SE.)

b) (3 pts.) Assuming the alternative to be true, I'd expect the sample average to be ≤ _____ inches with $\mathbf{SE_{avg}}$ = _____ inches.

c) (2 pts) What is the effect size in inches? D = _____inches

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

e) (6 pts.) If I set the significance level $\alpha = 5\%$, what is the Power of the test?

i) First find $|Z_\alpha|$. $|Z_\alpha|$ = _____.

ii) So $|Z_\beta|$ =_____ which means β= _____%, so Power =_____%.

Question 3 Part II (12 pts.)
Now suppose we keep the same $\alpha = 5\%$, same SD = 3" and the same effect size (D) as before, but we want power=90%.

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 = _____%, and $|Z_\beta|$ = ____

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

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

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

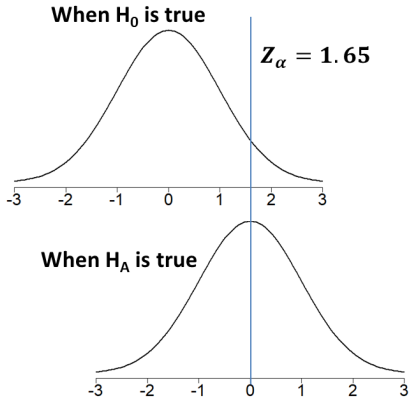
Round n to 2 decimal places, not to nearest whole number as usual.

Question 4 (6 pts)
The two histograms below represent the distribution of the test-statistic when H_0 is true and when H_A is true.
Fill in the 3 blanks below with the correct numbers.

a) What is the probability of mistakenly rejecting the null when H_0 is true? _____%

b) Judging from the 2 histograms, what is the distance between the hypothesized values of H_0 and H_A in Standard Units? D_Z = _____

c) What is the Power of the test? _____%



Question 5 *(12 pts.)* **pertains to the following study:**
In a study on multitasking, researchers divided 262 Stanford undergraduates into high multi-taskers and low multi-taskers based on their answers to questionnaires. Both groups of students were then given a series of cognitive tests designed to measure the essential abilities needed to successfully multitask (filter information, store and retrieve relevant information, switch tasks efficiently, etc.). In every cognitive category those who multitasked the most scored the lowest!

a) (2 pts.) This study is an example of a … ***Choose the best answer:***
- **i)** Randomized controlled double-blind experiment
- **ii)** Non-Randomized Controlled Study
- **iii)** Randomized Controlled Experiment without a placebo

b) (2 pts.) Does the study show that multitasking ***causes*** cognitive deficits? ***Choose one:***
- **i)** Yes, since there is clear evidence that those who multitasked the most scored the lowest in all cognitive categories.
- **ii)** No, the study only shows that multitasking is ***associated*** with cognitive deficits. The study does not prove whether multitasking does or does not cause cognitive deficits.
- **iii)** No, the study shows that multitasking is ***associated*** with, but ***definitely does not*** **cause** cognitive deficits.

c) (6 pts.) Below are either **confounders** that ***mix up*** the study making it look like multitasking impairs cognitive functions when it really doesn't, **causal links** that explain **how** multitasking **causes** cognitive impairment, or neither.

i) Information Overload--People who multitask a lot get bombarded with so much information that they lose the ability to separate the important information from the irrelevant.

Choose one: Confounder Causal Link Neither

ii) Age- Stanford students are young and more likely to successfully handle multitasking.

Choose one: Confounder Causal Link Neither

iii) Low attention spans—People who have difficulty focusing may be more likely to become multi-taskers to begin with, and also more likely to perform poorly on cognitive tests.

Choose one: Confounder Causal Link Neither

d) *(2 pts.)* Suppose we think gender may be a confounder. How can we eliminate its possible confounding effects? ***Choose one****:*
i) Split the data by gender and compare the test results of high multi-taskers to low multi-taskers within each gender.
ii) Split the data into high and low multi-taskers and compare the test results of females to males within each group.
iii) Split the data by gender and compare the % of multi-taskers in each group.

Question 6 (10 pts.)

A study published in the March 4, 2015 issue of the Journal of the American Medical Association evaluated whether peanut consumption might be more effective than peanut avoidance in preventing the development of peanut allergies in infants who are at high risk for the allergy. 640 infants aged 4 to 11 months with severe eczema and egg allergies (high risk indicators for peanut allergy) were ***randomly assigned*** to either consume (treatment) or avoid peanuts (control) until 5 years of age. The results were striking—17.2% of the children in the peanut-avoidance group tested positive for peanut allergy while only 3.2% of the group in the peanut-consumption group tested positive.

a) (2 pts.) Which of the following best describes this study:

- **i)** A randomized controlled experiment
- **ii)** An observational study with controls
- **iii)** A non-randomized controlled experiment

b) (2 pts.) Does the study show that eating peanuts helped prevent the children in the study from developing a peanut allergy?

- **i)** No, it only shows that there is an association between peanut consumption and reduced rate of peanut allergy since many environmental, cultural, social and biological factors contribute to both diet and allergic responses.
- **ii)** No, simply assigning children to 2 groups without considering the consequences of how peanut consumption or peanut avoidance may confer nutritional advantages limits any causal conclusions.
- **iii)** Yes, the study is strong evidence that peanut consumption helped prevent peanut allergy in these children although the causal mechanism can only be inferred.

c) (4 pts.) Which of the following could confound the results? Circle Yes or No for each.

- **i)** Cultural/Ethnic differences- Peanuts and peanut oil are popular in West African and Southeast Asian cuisines, groups that have a relatively low incidence of peanut allergies. **a.** Yes **b.** No
- **ii)** Health Benefits – Peanuts are a relatively healthy snack food. Children who eat peanuts may be healthier in general and less likely to develop allergies. **a.** Yes **b.** No
- **iii)** Pre-existing Health Problems- The children all had severe health problems to begin with making it difficult to discern whether or not it was the peanuts or pre-existing conditions that led to the development of a peanut allergy. **a.** Yes **b.** No
- **iv)** Overactive Immune System- Children with overactive immune systems are both more likely to have egg allergies (like the children in the study) and to develop a peanut allergy. **a.** Yes **b.** No

d) (2 pts.) 40 of the 640 infants showed evidence (by a skin-prick test) of already having a peanut allergy before they were even assigned to treatment or control. The researchers want to make sure that the 40 children are exactly evenly divided between the treatment and control groups but they don't want to introduce bias. What should they do?

- **i)** They should divide the infants into 2 groups (40 with pre-existing peanut allergy, and 600 without). Then randomly assign half of each group to treatment and half to control.
- **ii)** Randomly assign half of the 640 infants to treatment and half to control. This will ensure the infants will be evenly divided on all characteristics relevant to the response including pre-existing peanut allergy.
- **iii)** Randomly assign half of the 640 infants to treatment and half to control. In the unlikely event that the 2 groups are not balanced then, the researchers should balance the groups taking into account all variables to be as objective as possible.

.

Question 7 (26 points total)
Is there a correlation between how much you study and your GPA at UI? To find out we took a random sample of 36 UI undergrads and got the following results. (Assume the population scatter plot follows a liner trend.)

	Avg	***SD***
GPA	3	0.5
# Study Hours	4	2

r = 0.28

a) (2 pts.) Our best estimate of the **slope** of the regression equation predicting GPA from study hours for all UI undergrads is β = _____pts/study hr *Show work*.

b) (2 pts.) SE_{slope} = _____ pts/study hr. *Show work. Round to 2 decimal places.*

c) (2 pts.) Which one of the following assumption(s) about the errors are needed for the above SE to be valid?
Circle all that are needed **i)** Independence **ii)** Equal Variability **iii)** Normality

d) (4 pts.) Find a **95% CI** for β using the normal curve. (Assume the sample size is large enough to use normal curve.)
(Fill in 2 blanks below)

95% Confidence Interval = _____ ± _____ * SE_{slope}

e) (2 pts.) Can you be 95% confident that there is a non-zero correlation between study hours and GPA in the population?
i) Yes, because the above interval includes 0.
ii) No, because the above interval includes 0.
iii) Yes because the upper limit of the confidence interval is positive.
iv) No, because the above interval does not include 0.

f) (6 pts.) Now let's compute the Z test-statistic for testing H_0: $\beta = 0$ against H_A: $\beta \neq 0$

i) Z= _____. ii) p-value= _____% iii) Conclusion: Reject Null at $\alpha = 0.05$?
Show work. Circle one: Yes No

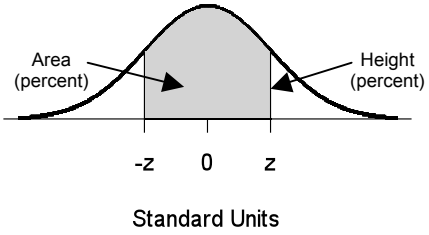
g) (2 pts.) Are the conclusions drawn in e and f, logically equivalent? In other words, will the conclusion from e and f always be the same? i) Yes ii) No

h) (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) $\sqrt{\frac{35}{36}}$

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

j) (2 pts.) The p-value would be _____ than the one computed in (f) .
i) smaller ii) larger iii) exactly the same

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