STAT 200 Exam 1
STATISTICS 200 EXAM 1

"STAT" Key

V-day, 2017

PRINT NAME____________________ ________________ Key

(Last name) (First name) (net ID, not UIN)

Circle Section: L1 11:00-12:20 pm or L2 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 (6 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 ______ 20

Page 2 ______ 25

Page 3______ 28

Page 4 & 5 ______ 27

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

STAT 200 Exam 1 V-day, 2017

Question 1 pertains to airport security screening tests designed to prevent passengers from bringing weapons on planes. *(10 points total)* Screening tests aren't perfect. When the passenger has a weapon there's 99% chance the alarm will correctly go off, but when the passenger does not have a weapon, there's a 12% chance that the alarm will incorrectly go off.

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

a) *(2 points)* The conventional null hypothesis (H_0) is that the passenger ...
- **i)** has a 12% probability of having a weapon.
- **ii)** has a 1% probability of having a weapon.
- **iii)** does have a weapon.
- **iv)** Does not have a weapon.

b) *(2 points)* A Type I error can occur ....
- **i)** Only when the H_0 is false.
- **ii)** Only when the H_0 is true.
- **iii)** When either H_0 is false or H_A is false, since errors can only occur under false hypotheses.

if switched

-2

c.e.

c) *(2 points)* A passenger comes in without a weapon, what is the probability of the screening test making a
i) Type I error 12 %? ii) Type II error 0 %?

d) *(2 points)* A passenger comes in with a weapon, what is the probability of the screening test making a ....
i) Type I error 0 %? ii) Type II error 1 %?

e) *(2 points)* If you change the cut-off of the screening test to decrease the probability of a Type I error, what happens to the probability of a Type II error?
i) Increases **ii)** Decreases **iii)** Stays the Same

Question 2 *(10 points)* A significance test is performed to analyze the results of a randomized experiment to see if some drug worked. Subjects are randomly assigned to treatment and control. The null and alternative hypotheses are the usual:

H_0: The difference in cure rates between the drug and the placebo = 0
H_A: The difference in cure rates between the drug and the placebo > 0

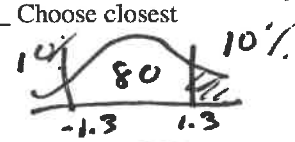
a) Suppose the null is true, what's the chance the researchers are going to make the wrong decision if they set the significance level $\alpha = 0.05$?
i) 0% **ii)** 1% **iii)** 2.5% **iv)** 5% **v)** 95% **vi)** 99% **vii)** not enough info

b) Suppose the null is false, what's the chance the researchers will make the wrong decision if they set the significance level α=0.05?
i) 0% **ii)** 1% **iii)** 2.5% **iv)** 5% **v)** 95% **vi)** 99% **vii)** not enough info (need specific alternative)

set $\alpha = 10\%$

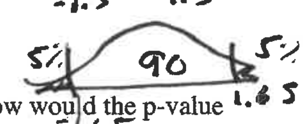
10% in 1 tail

c) If we set α=0.10 (null cut-off at 10%) 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

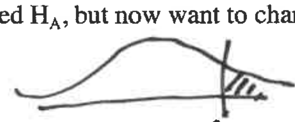


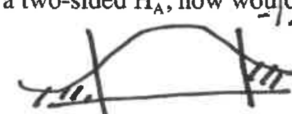
10% distributed over 2 tails

d) Repeat (c) above with a 2-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



e) If we computed a p-value using a one-sided H_A, but now want to change to a two-sided H_A, how would the p-value change?
- i) p-value would stay the same.
- ii) p-value would be divided by 2
- iii) p-value would be multiplied by 2
- iv) We'd have to re-do the experiment since p-values are no longer valid if you change the alternative hypothesis.


1-sided


2-sided

~~sample test stat~~

sample → test-stat → p-value
1-sided says you're assuming sample will land far from hypothesized null value in 1 direction so p = area in 1 tail.
2-sided Ha says sample will land far from hyp. null value in either direction, so ~~[illegible]~~ p-value = area in both tails.

$D = D_z \cdot SE_{ave}$

STAT 200 Exam 1 V-day, 2017

Question 3 *(25 points total)*

Part I (17 pts.) A University claims the ACT scores of its freshman class of 7,000 students roughly follows the normal curve with an average = 29 and SD = 3. I believe University is inflating the ACT average and it's really only 27.5 or lower with the same SD =3. I decide to do a significance test by randomly sampling 25 students from the class.

a) (3 pts.) Fill **in all 6** blanks below with the **correct numbers** for H_0 and H_A.

H_0: The average of the 7,000 students = 29 with a SD= 3.

H_A: The average of the 7,000 students ≤ 27.5 with a SD= 3.

b) (2 pts.)Assuming the null to be true, I'd expect the sample average to be 29 with a SE_{avg} = 0.6
(Show work for SE.)

$\frac{SD}{\sqrt{n}} \quad \frac{3}{\sqrt{25}} = 0.6$

c) (2 pts.) Assuming the maximum alternative to be true, I'd expect the sample average to be 27.5 with SE_{avg} = 0.6.

d) (2 pts) The effect size in ACT points is D = 1.5 points.

e) (2 pts.) The effect size in Standard Units is D_z = 2.5. *Show work.*

$\frac{1.5}{0.6} = 2.5$

f) (4 pts.) If I set the significance level α = 5%, what is the Power of the test?

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

ii) So $|Z_\beta|$ = 0.85 which means β= 20 %, so Power = 80 %.
(19.765) (80.235)

$|Z_\beta| = 2.5 - 1.65$ $\quad Z_\alpha + Z_\beta = D_z$

$D_z - |Z_\alpha|$

g) (2 pts.) Would we have improved the accuracy of our test if we had computed SE^+ instead of SE and used a t-test instead of z-test?

i)Yes, because n < 30 ii) No because population SD is known iii) Not enough information

Question 3 Part II (8 pts.)
Now suppose we keep the same α = 5%, same SD = 3 pts and the same effect size (D) as before. What sample size do we need to get power = 95%?

a) (2pts.) First, compute β and $|Z_\beta|$. β = 5 %, and $|Z_\beta|$ = 1.65

b) (2 pts.) The effect size in Standard Units is D_z = 3.3

1.65 + 1.65

c) (2pts.) SE_{avg} = 0.45 (Round to 2 decimal places). *Show work.*

$D = D_z \cdot SE_{ave} \quad SE_{ave} = \frac{1.5}{3.3} = 0.4545$

d) (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.

$SE_{avg} = \frac{SD}{\sqrt{n}} \Rightarrow n = \left(\frac{SD}{SE_{avg}}\right)^2 = \left(\frac{3}{0.4545}\right)^2 = 43.57$

$\text{accu}\left(\frac{3}{0.45}\right)^2 = 44.44$

Question 4 (14 pts.) A study examined the medical records of 4 million patients admitted to hospital emergency rooms over a 10-year period and found that those admitted during the weekend had a much higher death rate than those admitted on weekdays.

a) (2pts.)This study is an example of …. *Choose one:*
i) Randomized Controlled Experiment ii) Non-Randomized Controlled Experiment iii) Observational Study

b) (2pts.) Does the study show the quality of hospital emergency room care for the 4 million patients in the study was worse during the weekends than weekdays? ***Choose one:***
i) Yes, the study shows that the quality of care was worse over the weekends but does not indicate exactly how it was worse.
ii) No, it only shows that there is an association between weekend emergency room admissions and more deaths. It doesn't show that the type of care given over the weekend is responsible for more deaths. W H A

c) (10 pts.) Circle whether each of the following is a possible **confounder, causal link or neither**.

Weekend Hospital Adm.

i) Mortality rates vary by hospital. **a.** Confounder **b.** Causal Link **c.** Neither

$\boxed{\text{Serious Accidents}} \rightarrow \boxed{\text{Death}}$

ii) People may have more serious accidents over the weekend than during the week. **a)** Confounder **b.** Causal Link **c.** Neither

iii) Emergency rooms may be understaffed over the weekend. **a.** Confounder **b)** Causal Link **c.** Neither

W. H. A → less doctors → more death

iv) Maybe people who aren't that sick are more likely to wait until Monday to go to the emergency room because they'd rather not waste their weekends in the hospital while those who are so sick they have no choice are more likely to go on weekends.
a) Confounder **b.** Causal Link **c.** Neither

$\boxed{\text{Sicker People}} \rightarrow \boxed{\text{WHA}}$, $\boxed{\text{Sicker People}} \rightarrow \boxed{\text{Death}}$

v) Less experienced doctors may be more likely to be assigned to the weekend shifts. **a.** Confounder **b.** Causal Link **c.** Neither

$\boxed{\text{W.H.A}} \rightarrow \boxed{\text{worse Doctors}} \rightarrow \boxed{\text{Death}}$

Question 5 *(14 pts.)* **pertains to the following study:**
(Hypothetical) A nation-wide study examining cholesterol levels in women found that women with 4 or more children had significantl higher levels of cholesterol than women with only 2 or fewer children.

a) *(2 pts.)*Is this study an observational study or a designed experiment? **i)** Observational Study **ii)** Designed Experiment

b) *(2 pts.)* Does the study show that having more children raises cholesterol levels? Circle answer.
i) Yes, it shows definite causation although the exact causal mechanism isn't explained.
ii) No, it only shows that there is an association between more children and higher cholesterol. It does not show that one causes the oth

c) *(8 pts.).* Which of the following describe possible confounders? *Circle "Yes" if possible confounder given the description, "No" if not possible confounder given the description.*

i. Income- Women with more children tend to be poorer (poorer to begin with, not as a result of more children), and cholesterol levels tend to be higher among the poor. **a)** Yes **b)** No

$\boxed{\text{Poor}} \rightarrow$ More Children, $\boxed{\text{Poor}} \rightarrow$ High Cholesterol

ii. Stress- More children cause extra stress that could lead to general poor health including raised cholesterol levels. (For example, women with 4 children may have less time and energy to prepare healthy food and to exercise) **a)** Yes **b)** No

More kids → Stress → Higher Cholesterol

iii. Genetic Predisposition- Cholesterol levels are strongly affected by heredity, so the more children you have the more likely they will be to exhibit a propensity towards high cholesterol. **a)** Yes **b)** No

iv. Age- Women with more children tend to be older (on the average), and cholesterol rises with age. **a)** Yes **b)** No

d) *(2 pts.)* Suppose we think that geographical region is a confounder since both family size and cholesterol levels are strongly influenced by region (for example, South Carolina has both high levels of cholesterol and high birth rates while Colorado has low levels of both.) How can we minimize the possible confounding effects of geographical region? ***Choose one:***
i) Split the data by geographical region and compare the % of women with 4 or more children to the % with 2 or less children within each region.
ii) Split the data by geographical region and compare the % of women with high cholesterol within each region.
iii) Split the data by geographical region and compare the cholesterol levels of women who have 4 children in one region (say South Carolina) to the cholesterol levels of women who have 2 or less children in another region (say Colorado).
iv) Split the data by geographical region and compare the cholesterol levels of women who have 4 or more children to the cholesterol levels of women with 2 or less children within each region. (to eliminate influence of region)

STAT 200 Exam 1 V-day, 2017

Question 6 (27 points total)

How do the number of hours students spend partying per week correlate with their GPA? Suppose a random sample of 49 UI undergrads yielded the following results. (Assume the population scatter plot follows a liner trend.)

		Avg	*SD*	
y	***GPA***	3.1	0.4	***r = - 0.3***
x	***# Party Hours***	9.3	4	

a) (2 pts.) Our best estimate of the **slope** of the regression equation for all UI undergrads is β = ______pts/party hr
Show work.

$r\ SD_y / SD_x = -0.3 \left(\frac{0.4\ pts}{4\ hr}\right) = -0.03$

b) (2 pts.) SE_{slope} = 0.014 pts/party hr. *Show work. Round to 3 decimal places.*

$\frac{\sqrt{1-(-0.3)^2}}{\sqrt{49}} \times \left(\frac{0.4\ pts}{4\ hr}\right) = 0.0136 \frac{pts}{party\ hr}$

c) (2 pts.) Which one of the following assumptions is NOT needed for the above SE to be valid?
Choose one:

i) Independence of the Errors **ii)** Equal Variability of the Errors **iii)** Normality of the Errors

d) (3 pts.) Find a **96% CI** for the population slope using the normal curve. (Fill in 2 blanks below)

96% Confidence Interval = -0.03 ± 2.05 * SE

96

-2.05 2.05

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

i. (2pts.) To find SE^+, you'd multiple SE by $\sqrt{\frac{49}{47}} \approx 1.02$
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? 47

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*

Question 6f is on the next page.

bc t curve is fatter in tails and lower in middle

Don't confuse t-stat with t*

t-stat < z-stat by factor of $\sqrt{\frac{n-2}{n}}$

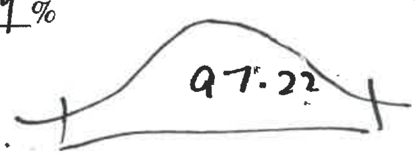
but t* > z* (for same CI)

Question 6 cont.

f) (12 pts.) Compute the z and t test-statistics for testing H_0: **β = 0.** (*Round your final answers to 2 decimal places, but don't round during intermediate steps.*) $r = -0.3$ $n = 49$

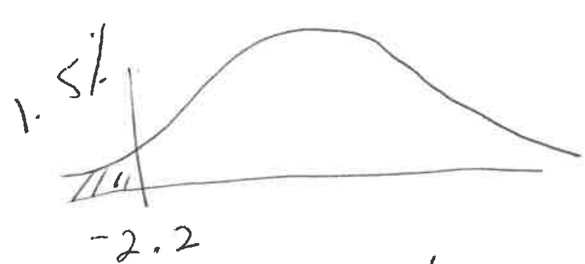
Fill in the table below

	Z $= \frac{r\sqrt{n}}{\sqrt{1-r^2}}$	**t** $= \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$
Compute the values of the 2 test statistics. *Show work below your answers.*	(*2 pts.*) **Z** = -2.2 $\frac{-.3\cdot\sqrt{49}}{\sqrt{1-(-0.3)^2}} = -2.204 \simeq -2.2$	(*2 pts.*) **t** = -2.16 $\frac{-.3\cdot\sqrt{47}}{\sqrt{1-(-0.3)^2}} = -2.156$ $t = Z\sqrt{\frac{47}{49}}$

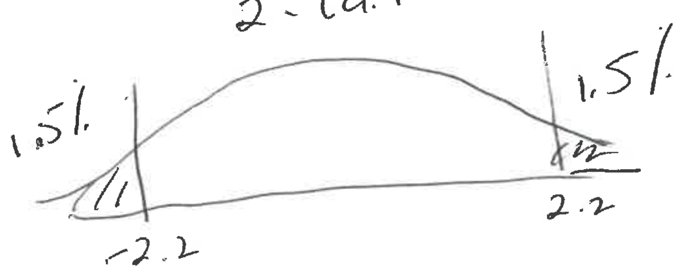
Use the normal table to calculate the p-value for the Z test. Assume $\mathbf{H_A}$: $\boldsymbol{\beta < 0}$ for both the Z and t tests.	
Z (*2 pt*) p-value= 1.39 % accept 1.5% ce 	**t** (*2 pts*) How many degrees of freedom? 47 (*2 pts*) Will the p-value be >, < or equal to the p-value for Z? i) > ii) < iii) =

Same as above, but now assume the H_A for both the Z and t test is 2-sided: $\mathbf{H_A}$: $\boldsymbol{\beta \neq 0}$	
Z (*1 pt*) p-value is … ***Choose one*** **i)** double the % for the 1-sided test **ii)** half the % for the 1-sided test **iii)** it depends on the value and sign of the Z-stat.	**t** (*1 pt*) p-value is … ***Choose one*** **i)** double the % for the 1-sided test **ii)** half the % for the 1-sided test **iii)** it depends on the degrees of freedom

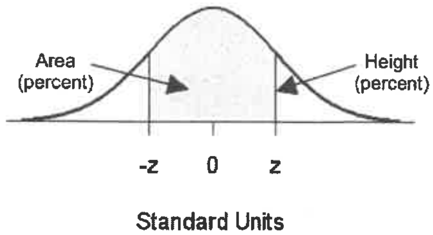
1-tail



2-tail



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