STATISTICS 100 EXAM 1 Stat Key **Fall 2016**

PRINT NAME_____ _____ _____
(Last name) (First name) net ID

CIRCLE SECTION please! **L1 (TR 12:30pm) L2 (TR 3:30pm) S1 (MWF) ONLINE**

Write answers in appropriate blanks. When no blanks are provided **CIRCLE** your answers.
SHOW WORK when requested, otherwise no credit.
Do NOT use scrap paper.

Make sure you have all 7 pages including the normal table (11 problems).

For Questions 9 -11 use 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.

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 _____ 12

Page 2 _____ 16

Page 3 _____ 20

Page 4 _____ 25

Page 5 _____ 14

Page 6 _____ 13

Total Score _____

Online Section
Last names: A-D
150 Animal Science
(50 exams)

There IS No Class tomorrow!

Scores will be posted on Compass by Friday morning and exams returned in class next week. Online students may pick up their exam in 23 Illini Hall during office hours next week.

Question 1 (12 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)** An observational study with controls
- **ii)** A non-randomized controlled experiment
- **iii)** A randomized controlled experiment

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

- **i)** Yes, the study is strong evidence that peanut consumption helped prevent a peanut allergy in these children although the causal mechanism can only be inferred.
- **ii)** 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.
- **iii)** 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.

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

- **i)** 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
- **ii)** 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 peanut allergy. **a.** Yes **b.** No
- **iii)** 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 peanut allergies. **a.** Yes **b.** No
- **iv)** 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

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)** 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.
- **ii)** 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.
- **iii)** 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.

e) (2 pts.) Over the course of the 5 year study some children assigned to eat peanuts quit eating them and some children assigned to not eat peanuts started eating them. Which represents the best comparison?

- **i)** Compare those who ate peanuts to those who didn't since that's the original plan.
- **ii)** Compare those assigned to eat peanuts to those assigned to not eat peanuts to maintain the balance of the 2 groups.
- **iii)** Compare those assigned to eat peanuts who consistently ate them, to those assigned to not eat peanuts who consistently didn't eat them.

Question 2 pertains to the following study: (8 pts.)

One day, while scrolling through Facebook this summer, I came across an article entitled: "The Secret to a Long Marriage Is Drinking Together." As a newlywed, I quickly became interested and clicked on the article. It stated that couples who drink together, stay together. Over 2,000 couples were involved and they found that couples who reported drinking alcohol together even just a few times each year were less likely to get on each others' nerves and more likely to have a positive outlook on their marriage. The study highlighted that what's most important isn't how much the couples drink, but whether they BOTH drink. If both partners drank, they were more likely to have a happier marriage than if just one of them drank.

a) (2 pts.) Based only on the information above, this study is an example of .... ***Choose one:***

- **i)** A randomized controlled double blind experiment.
- **ii)** A non-randomized controlled experiment with historical controls.
- **iii)** An observational study
- **iv)** A randomized controlled experiment that was not double blind and did not have a placebo

b) (2 pts.) What can we conclude from this study? ***Choose one:***

- **i)** We see that there is an association between drinking alcohol together and happy marriages, however we aren't sure if there is a causal relationship- there could be other variables confounding the data.
- **ii)** Drinking alcohol together causes couples to stay together! Steve and I should drink together as much as possible if we want to remain happy in our marriage.
- **iii)** Drinking alcohol and happy marriages are not related to each other at all.
- **iv)** When one partner drinks and the other doesn't, it causes marital problems.

c) (4 pts.) State whether the following are confounders, causal links, or neither:

- **i)** Shared Interests- Couples with shared interests are both more likely to drink together and to engage in other activities together that lead to happier marriages.
  a) confounder **b)** causal link **c)** neither.
- **ii)** Endorphins- Alcohol releases endorphins making you feel happier which leads to happier marriages.
  a) confounder **b)** causal link **c)** neither.
- **iii)** Commitment - Marital happiness depends on a strong commitment from both partners to continually work on resolving their problems rather than running away from them.
  a) confounder **b)** causal link **c)** neither
- **iv)** Communication—Clear communication about expectations going into the marriage and the long term goals contribute to longer lasting and happier marriages.
  a) confounder **b)** causal link **c)** neither

Question 3 (8 pts.)

A study compared the success rate of two treatments designed to help smokers quit smoking. Subjects were classified as either heavy smokers or moderate smokers before treatment began. Heavy smokers have a harder time quitting. The table below gives the results for the two treatments.

	Treatment A				Treatment B			
	Number	**# Successes**	**# Failures**	**% Success**	**Number**	**#Successes**	**# Failures**	**% Success**
Heavy Smokers	400	200	200	50%	100	30	70	30%
Moderate Smokers	600	600	0	100%	900	870	30	97%
Overall Total	1000	800	200	80%	1000	900	100	90%

a) (2 pts.) Based only on the information above Do you think this was a randomized experiment?

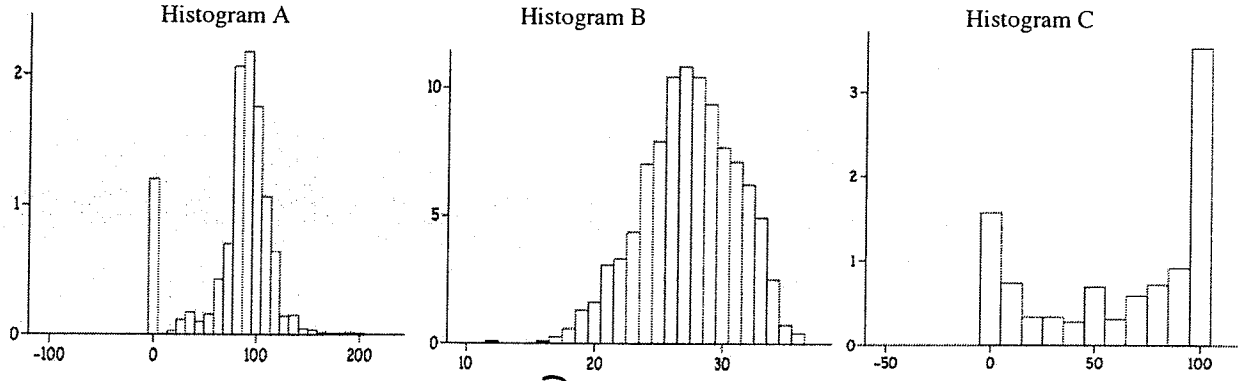
- **i.** Yes, since it's highly unlikely both groups would end up with exactly1000 subjects otherwise.
- **ii.** No, since it's highly unlikely that randomization would result in 40% heavy smokers in A and only 10% in B.
- **iii.** There are arguments for both sides, it's hard to tell without more information..

b) (2 pts.) Which treatment had a higher success rate for heavy smokers?
Choose one: i) A ii) B iii) Not enough info

c) (2 pts.) Which treatment had a higher success rate for a moderate smokers?
Choose one: i) A ii) B iii) Not enough info.

d) (2 pts.) Based only on the information given, which treatment has a better chance of success for someone who smokes?
Choose one: i) A ii) B iii) It depends on whether they're heavy or moderate smokers.

Question 4 pertains to the 3 histograms below representing your survey responses to 3 questions: (6 pts.)
What is your ACT score?, What's the fastest speed you've ever driven (in mph)? and What percent of your college costs are your parents paying for?



a) (4 pts) Which graph represents ACT scores? B Which graph represents % of college costs? C

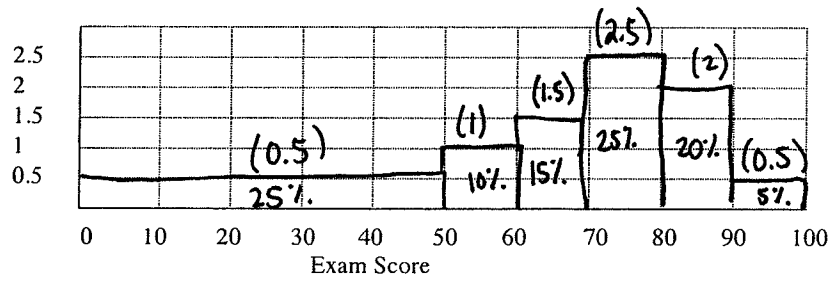
b) (2 pt.) I wrote the average and median of Histogram A down, but I forgot to label them. Here are the 2 numbers: 81.23 and 90. Which is which?
i) 90 is the average (ii) 90 is the median iii) Cannot be determined

Question 5 (14 points)
A distribution table for the scores on a Physics exam is shown below on the left. The second row says that 10% of the students scored between 50 and 60. Fill in the blanks in the % and height columns then draw the histogram below.

a) (3 pts.) **Fill in table** ½ point for each blank

Points	%	Height (% per point)
0-50	25	0.5
50-60	10	1
60-70	15	1.5
70-80	25	2.5
80-90	20	2
90-100	5	0.5

b) (3 pts.)**Draw Histogram** ½ point for each block



c) (2 pts) What is the median score? 70

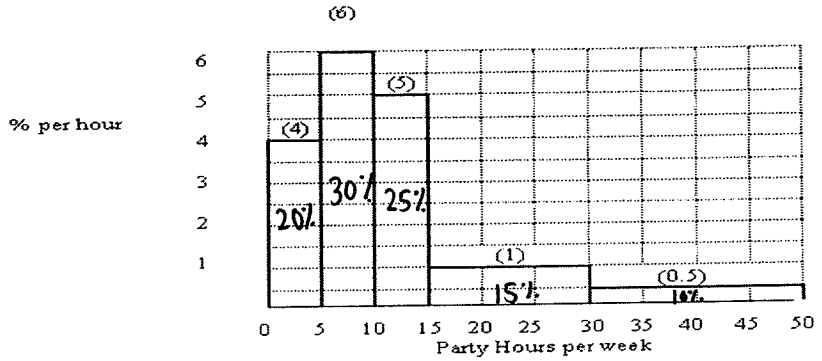
d) (2 pts.) Is the average >, <, or = the median? <

e) (4 pts) If 10 pts were added to everyone's score (so the highest possible score was 110) how would it affect the average, median and SD? ***Check the appropriate boxes below.***

	Increase	Decrease	Stay the Same
i) Average would …	X		
ii) Median would …	X		
iii) SD would …			X

Question 6 pertains to the histogram below. (15 pts.)

The histogram below depicts the number of hours per week students reported partying last semester. Class intervals include the left-endpoint but not the right. (For example someone who partied 5 hours is included in the 5-10 block, not the 0-5 block.) The height of each block is given in parentheses.



a) (4 pts.). What percentage fell in the following intervals?

i) 5-10 hours 30 % ii) 10-15 hours 25 % iii) 15-30 hours 15 % iv) 30-50 hours 10 %

b) (2 pts.) What percent of the students partied 3 hours per week? 4 %
(Assume an even distribution throughout each interval.)

c) (2 pts.) What is the median number of party hours per week? 10 hours

d) (2 pts.) The median is _____ the average *Choose one:*
i) less than **ii)** greater than **iii)** equal to **iv)** cannot be determined

e) (2 pts.) The 90th percentile is....(In other words, 90% of the students party less than how many hours per week?)
Choose one:
i) 5 hours **ii)** 15 hours **iii)** 30 hours **iv)** 40 hours **v)** 45 hours

f) (3 pts.) If everyone who partied 30 hours or more cut their weekly party hours by half (so 30 would be reduced to 15) would the average, median and SD change?
Check the appropriate boxes below.

	Increase	Decrease	Stay the Same
i) Average would		X	
ii) Median would			X
iii) SD would ...		X	

Question 7 pertains to the following list of 5 numbers: 3, 7, 11, 0, 9 (8 pts.) 0, 3, 7, 9, 11

a) (1 pts.) The median is 7

b) (1 pts.) The average is 6

c) (3 pts.) The deviations from the average are -6 -3 1 3 5

d) (3 pts.) Compute the SD. *(Show work. Circle answer.)*

$$\frac{36+9+1+9+25}{5} = 16 \qquad \sqrt{16} = (4)$$

Question 8 (2 pts.)
A list of 4 numbers has the following deviations from the average: -5, 8, 7, -10 (2 pts.) Fill in the blank with the missing deviation.

Rounding Instructions for Questions 9-11. 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.

Question 9 (14 pts.) pertains to Math SAT scores that are normally distributed with an **average=500 and SD=100**

a) (2 pts.)Approximately **95%** of math SAT scores are between 300 and 700 also accept 305 - 695
(Fill in the blanks with SAT scores, NOT z scores)

For parts b-d, mark the correct z scores on the curves below and shade the appropriate region.
(Pay attention to the markings given on the horizontal axis, and place your z score so that it fits the scale given. The markings range from -3 to 3)

b) (3 pts.) If a student scored **310** on the math SAT, what is his ***Z score*** and what % of the people scored ***below*** him?
Show work and write your answers in the blanks provided.

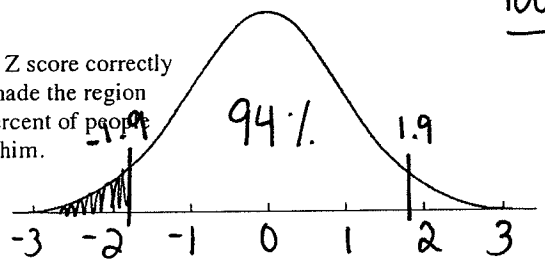
i) (1 pt) **Z score =** -1.9 **ii)** (1 pt) 3 **% below 310**

Show work:

$Z = \frac{310-500}{100} = -1.9$

$\frac{100-94}{2} = 3$

iii) (1 pt) Mark the Z score correctly on the curve and shade the region representing the percent of people who scored below him.

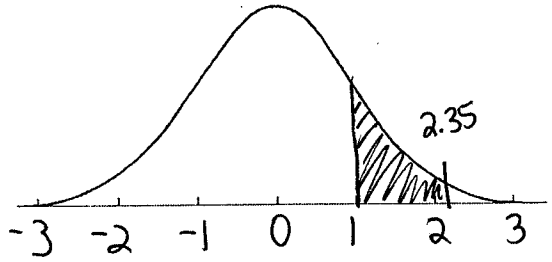


c) (4 pts.)What **percent** of the SAT scores fall in the interval **600 to 735**?

$Z = \frac{600-500}{100} = 1$ $Z = \frac{735-500}{100} = 2.35$

i) (2pts.)Translate interval into Z scores 1 **to** 2.35

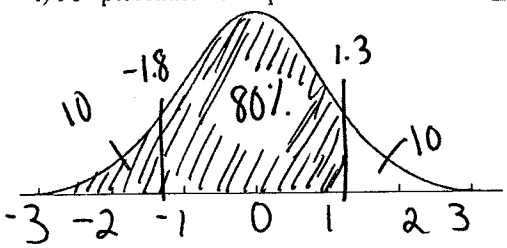
ii) (2pts.) Mark the z scores on the curve below, shade the interval and calculate the percent. 15 %



$\frac{98-68}{2} = 15\%$

d) (3 pts.) What score corresponds to the **90th percentile**? Mark the 90th percentile on the curve below, find the corresponding z-score, and the corresponding SAT score. ***Show work.***

i) 90th percentile corresponds to middle area= 80 %, → **Z score=** 1.3 and **SAT score =** 630



value = avg + (z)(SD)
= 500 + (1.3)(100)

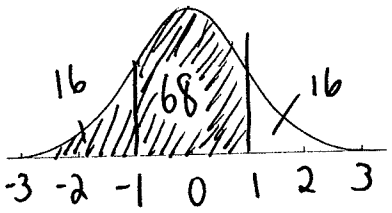
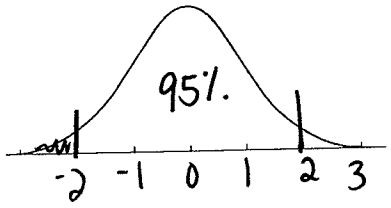
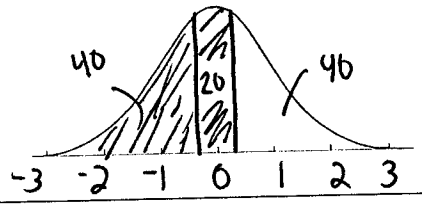
e) (2 pts.) What SAT score corresponds to the **10th percentile**? (Hint: You don't need to re-draw the curve from part (d), no need to show work.)

Z score= -1.3 **SAT score =** 370

value = 500 + (-1.3)(100)

Question 10 (7 pts.)

According to survey data, **male** Stat 100 students have an ***average height of 70" and a SD of 3"*** and the histogram of their heights is close to the normal curve. Consider male Students A, B and C below. For each student you're given either his height or his z score or his percentile. For Student A you're given height and you need to find his Z score and his percentile, for Student B you're give his Z score and you need to find his height and percentile, and for Student C you're given his percentile and you need to find his Z score and his height. ***Fill in the 7 missing blanks below. Show work where indicated.***

Height in inches	**Z score**	**Percentile** **Use the normal table to find the middle area.** **Mark the z score accurately on the horizontal axis**
a) Student A is 73" $Z = \frac{73-70}{3} = 1$	(1 pt) **Z=** 1 **Show work:**	(1 pt) 84 **th percentile** 
b) (1 pt) 64 **"** **Show work:** val = 70 + (-2)(3)	**Student B has a Z score = -2**	(1 pt) 2.5 **th percentile** 
c) (1 pt) 70.75 **"** (Don't round answer) **Show work:** val = 70 + (0.25)(3)	(1 pt) **Z=** 0.25	**Student C is in the 60th percentile which** (1pt) **Corresponds to a middle area=** 20 **%** 

Question 11 (6 pts)

The scores on a Chemistry exam follow the normal curve with an average of 55 and a SD of 20. The professor wants to give the top 15% of the class A's, the next 35% of the class B's, and the rest C's.

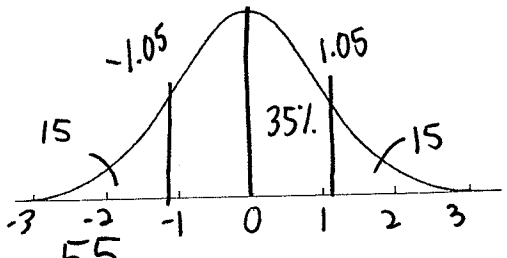
a) (4 pts.) What should be the cut-off between an A and B? Mark the top 15% on the curve and label the Z score.

i) Z score cut-off is Z= 1.05

Show work.

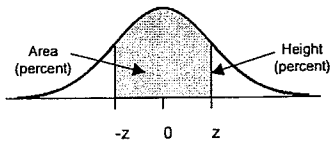
val = 55 + (1.05)(20)

ii) Exam Score cut-off = 76



b) (2 pts.) What exam score should be the cut-off between a B and a C? 55

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