Stat Key

EXAM 1 Stat 100 **Fall 2018**

Cover Sheet Questions ***(1 pt.)***

1) What's your name? _____ _____

(Last name) (First name)

2) What's your net ID (email)? _____

3) Which Section are you in?

Circle one: **i)** L1 (TR at 12:30 pm) **ii)** L2 (MWF at noon) **iii)** ONLINE

Write answers in appropriate blanks. All multiple-choice questions have exactly one answer. If you circle more than one answer you will automatically be marked wrong.

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).

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 _____ 14

Page 2 _____ 8

Page 3 _____ 15

Page 4 _____ 16

Page 5 _____ 25

Page 6 _____ 22

WARNING- The exams look alike but you are sitting next to people who actually have different version. Copying from anyone is equivalent to giving a signed confession.

All Cheating including being caught with a non-permissible calculator or formula sheet will result in a 0 and an academic integrity violation on your University record.

Cover Page _____ 1 *for answering cover page questions 1-3 correctly!*

Total Score _____

There is NO CLASS on Thursday or Friday!

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.

Page 1 - Stat

Exam 1 Stat 100 Fall 2018

Question 1 (6 pts.) We compared Exam 1 scores of two groups of Stat 100 students: Those who followed the directions and filled out their cover sheet correctly and those who didn't. The average exam score for the 1350 students who filled out their cover sheets correctly was significantly higher than the average exam score of the 350 students who did not (not even counting the 1 point they lost for not filling it out). observational study

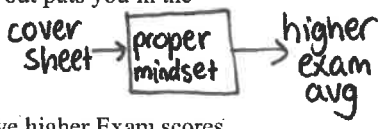
a) (2 pts.) Based only on the given information, can you conclude that correct cover sheet completion caused the higher exam scores? **Choose one:**

i. No, we can only conclude that it is *associated* with higher exam scores. It definitely couldn't cause them because association rules out causation.
(ii.) No, we can only conclude that it is *associated* with higher exam scores. It may or may not be the cause.
iii. Yes, we can conclude causation because the data arose naturally. We didn't set out to prove or disprove that following directions mattered so the data is unbiased and can be trusted.
iv. Yes, there's strong evidence that it was the cover sheet completion that caused the higher exam scores since everything else was the same.

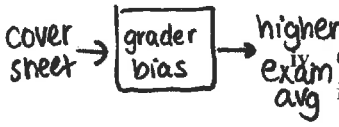
b) (4 pts.) Below are either possible causal links, confounders or neither. (Choose answer based only on given info.)

i. Serious Students who follow directions: Serious Students who follow directions are both more likely to do well on exams and more likely to fill out their cover sheets correctly.
Choose one: a) Causal Link (b) Confounder c) Neither

ii. Getting in the proper mind set: Reading the cover sheet carefully and properly filling it out puts you in the proper frame of mind to perform better on the actual exam.
Choose one: (a) Causal Link b) Confounder c) Neither



iii. Higher Homework Average: Students with a higher homework average also tend to have higher Exam scores.
Choose one: a) Causal Link b) Confounder (c) Neither no mention of cover sheet



iv. Grader Bias: Cover sheet mistakes are annoying and waste graders' time so graders may grade exams with incorrect cover sheets more harshly. **Choose one:** (a) Causal Link b) Confounder c) Neither

Question 2 (8 pts.) A group of 500 men on Medicare (age 65+) participated in a study for a new pill to treat Erectile Dysfunction (inability to perform in the bedroom). Half of the men were randomly assigned to take the new drug and half were given a plain sugar pill. Neither the patients nor the doctors who evaluated them knew who was in which group. Both groups (and their partners) were delighted with their improvement, but there was no difference between the 2 groups.

a) (2 pts.) How would you best describe this study? ***Choose one:***

i) Randomized Single-Blind Experiment (ii) Randomized Double-Blind Experiment iii) Observational Study

b) (2 pts.) What kind of bias exists in these results? ***Choose one:***

i. Subject Bias- the patients could tell which group they were.
ii. Evaluator Bias- the evaluators could tell which group the subjects were in
iii. Both Subject and Evaluator Bias
(iv.) iv. No Bias

c) (2 pts.) Choose the best conclusion … ***Choose one:***

i) Only the drug works. ii) Only the sugar tablet works. (iii) They both work equally well.

d) (2 pts.) Suppose 50 of the men were very old and the researchers want to make sure that these very old men were 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 men into 2 groups (the 50 men who are very old and the 450 men who aren't). Then randomly assign half of each group to treatment and half to control.
ii) They should randomly assign half of the 500 men to treatment and half to control. This will ensure that the men will be evenly divided on all characteristics relevant to the response including age.
iii) Randomly assign half of the 500 men to treatment and half to control. Check to see if the very old men are evenly divided. If not, you can rearrange them without introducing bias as long as you do it before treatment starts

Page 2- Stat

Question 3 (6 pts.)
Would students in Stat 100 learn better if they were allowed Formula Sheets during exams?

To answer that question we did 2 studies.

trust results

Study A—We **randomly** assigned half the Stat 100 students to the Formula Sheet Group and half to the No Formula Sheet group.
Study B—We acted like a doctor and "prescribed" formula sheets to those students I thought really needed them and didn't prescribe them to those I thought would do well without them. obs. study
All students took the same exams and here are the results:

	Study A--- Randomized		Study B---Non-Randomized	
	# of Students	Average Exam Score	# of Students	Average Exam Score
Formula Sheet Group	500	80%	400	74%
No Formula Sheet Group	500	92%	600	94%

a) (2 pts.) Both studies found that the No Formula group did much better than the Formula group, but the randomized design saw only a 12% difference whereas the Non-Randomized Design showed a 20% difference. What possible reason could account for that?
Choose one:

(i) In the non-randomized study, we chose the stronger students to be in the No Formula group and the weaker students to be in the Formula group, so the No Formula group did better both because they were stronger students to begin with and because not having a Formula sheet made them learn better.

ii) In the non-random study, we showed how tailoring the study method to fit the student works better and therefore causes a more dramatic improvement.

iii) In the non-randomized studies, more students were assigned to the No-Formula group than to the Formula group so that could account for the increased difference.

b) (2 pts.) Which study is more likely to have confounders? ***Choose one:*** **i)** Study A **(ii)** Study B **iii)** They're equally likely

c) (2 pts.) Judging from both studies, would you conclude that there is good evidence for the following statements?

i) Students learn better when they are allowed formula sheets based on their needs.
Circle yes or no: Yes **(No)**

ii) Students seem to learn better when they are not allowed to rely on formula sheets.
Circle yes or no: **(Yes)** No

Question 4 (2 pts.) A study was done to compare the effectiveness of high dose vs low dose chemotherapy to treat esophageal cancer. The subjects were 2000 adults. Half were randomly assigned to take a high dose pill daily and half assigned to take the low dose pill daily. In every other way the 2 groups received the same medical care.

The table below gives the 5-year survival rate for "adherers" and "non-adherers" in the high and low dose groups. Adherers regularly took the drug at least 2/3 of the time while non-adherers took the drug less than 2/3 of the time.

	High Dose		**Low Dose**	
	Number	**5-year survival rate**	**Number**	**5-year survival rate**
Adherers	500	84%	900	78%
Non-Adherers	500	60%	100	68%
Total	1000	72%	1000	77%

To assess which dosage is more effective, which two percents in the table above should you compare?

(2 pts.) **Fill in both blanks with the correct percents.** 72 % vs. 77 %

*always compare overall results when you have a randomized controlled experiment

Page 3 - Stat

Question 5 (7 pts.) **pertains to the following study:**

A Swedish study found that heart attack patients have lower death rates when they eat chocolate. The study compared the records of 1,169 patients recovering from a heart attack and tracked them for eight years. Those who reported eating chocolate regularly were less likely to die after 8 years than those who ate no chocolate. And the more chocolate they ate the higher the benefit.

chocolate → lower death rates

a) (2 pt.) 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.) The study reported that they controlled for gender. This means they thought gender might be a confounder so they eliminated its confounding effect. How did they do that? ***Choose one:*** Stratification

i) At the beginning of the study, they divided the patients into males and females and then randomly divided the males and females equally between the chocolate and no chocolate groups.

(ii) At the end of the study, they stratified on gender, and compared the death rate of chocolate eaters to non-chocolate eaters within each gender.

iii) Throughout the study they kept track of those who failed to adhere in both groups whether they were male or female and made sure to compare the death rate of everyone in the original chocolate group to everyone in the original non-chocolate group.

c) (3 pts.) Circle whether the following are confounders, causal links, or neither:

chocolate → [antioxidants] → lower death rate

i) (1 pt.) Flavonoid Antioxidants: Chocolate contains flavonoid antioxidants that are widely believed to have beneficial cardiovascular effects. ***Choose one:*** a) confounder (b) causal link c) neither

ii) (1 pt.) Eating Disorders – People with eating disorders are less likely to eat chocolate and eating disorders take a toll on your cardiovascular health ***Choose one:*** (a) confounder b) causal link c) neither

iii) (1 pt.) Chocolate Type- Dark chocolate is said to have more health benefits than white chocolate ***Choose one:*** a) confounder b) causal link (c) neither.

chocolate ← [eating disorders] → lower death rate

Question 6 (8 pts.) **pertains to the following list of 6 numbers:** 2, -3, -1, 5, 3, 6

a) (2 pts.) The average is 2, and the median is 2.5 list #'s in order first: -3, -1, 2, 3, 5, 6

Step 1

b) (2 pts.) The **deviations** from the **average** are 0, -5, -3, 3, 1, 4 Step 2 Fill in the 4 missing deviations.

c) (2 pts.) The sum of the deviations from the average should = 0. *Fill in the blank with a number.*

d) (2 pts.) Compute the Standard Deviation. Round your answer to 2 decimal places. ***No work, no credit. You may start with the deviations you found in part (b).***

Step 3: square the deviations

0, 25, 9, 9, 1, 16

cont error from b

Step 4: take the avg of the squared deviations

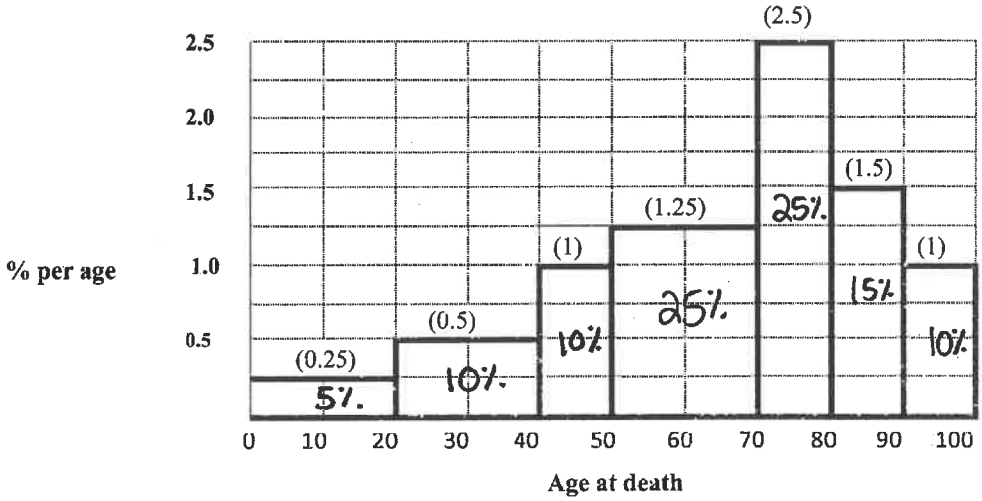
$$\text{avg} = \frac{0+25+9+9+1+16}{6} = 10$$

Step 5: take the square root

$\sqrt{10} = 3.16$ accept

Page 4- Stat

Questions 7 (16 pts.) pertain to the histogram below showing the age at death of a large population. The height of each block is given in parentheses. (Assume an even distribution throughout each interval.)



Area = Width × Heig

a) (2 pts.) What percent of the population died in the 70-80 interval?
Choose one: **i)** 10% **ii)** 15% **iii)** 20% **iv)** 25% **v)** 30%

b) (2 pts.) The median is closest to …
Choose one: **i)** 40 **ii)** 50 **iii)** 60 **iv)** 70 **v)** 80

c) (2 pts.) The median is _____ the average. long left hand tail ⇒ avg < med so med > avg
Choose one: **i)** less than **ii)** greater than **iii)** equal to **iv)** cannot be determined

d) (2 pts.) The 25th percentile is ***Choose one:*** **i)** 20 **ii)** 25 **iii)** 30 **iv)** 40 **v)** 50

e) (2 pts.) The percent of the population who died at 75 years is closest to…..
Choose one: **i)** 1% **ii)** 1.25% **iii)** 1.5% **iv)** 2% **v)** 2.5%

f) (2 pts.) If everyone lived 1 more year, the average would …. ***Choose one:***
i) Increase by 1 year. **ii)** Increase by 0.01 years **iii)** Increase 1 5% **iv)** Stay the Same **v)** Decrease

g) (2 pts.) and the SD would …. ***Choose one:***
i) Increase by 1 year. **ii)** Increase by 0.01 years **iii)** Increase 1 5% **iv)** Stay the Same **v)** Decrease

h) (2 pts.) If you knew the average and SD of the ages displayed in the histogram above, would it be appropriate to use the normal approximation to figure what percentage of the ages fell within various intervals?
Choose one:
i) Yes, because we know that the histogram represents the age at death of a large population.
ii) Yes, because the ages at death range from 0 to 100.
iii) No, because the histogram of the ages is not close enough to following the normal curve; it has a long left-hand tail
iv) Maybe, depending on whether the ages were randomly drawn from a larger population.

*only use the normal curve if the histogram looks normal

Page 5 - Stat

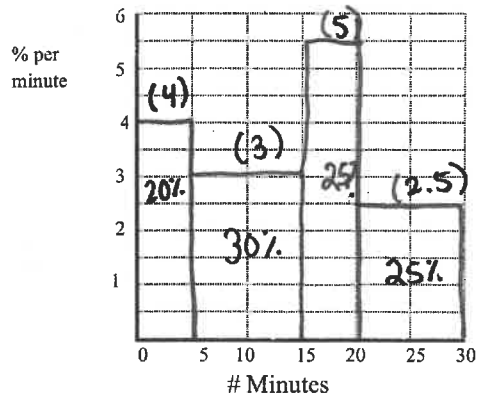
Question 8 (10 pts.) 5,000 teenage Americans were asked the question: "How much time, in minutes, passes between when your alarm first goes off and you get out of bed?" The results are summarized in the table below.

Area = Width × Height

a) (4 pts.) **Fill in the 4 blanks in the table below:**

Minutes	Area %	Height of Block (% per minute)
0-5	20	4
5-15	30	3 (1 pt.)
15-20	25	5 (1 pt.)
20-30	25 (1 pt.)	2.5 (1 pt.)

b) (2 pts.) **Draw the histogram on the graph below.**



c) (2 pts.) What is the median number of minutes? 15

d) (2 pts.) The number of people who answered 0-5 minutes is 20% the number of people who answered 5-15 minutes. 30%

Choose one: **i)** less than **ii)** more than **iii)** the same as

Question 9 (12 pts.)

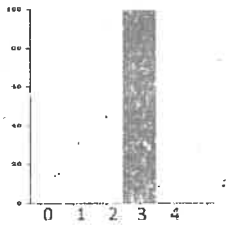
A list of 10 numbers has an average = 6, median= 4, and SD = 2. For (a-d) below, calculate the new average, median, and SD after the original list has been changed. **For a-c, give your answer as a number (not a word like "increases".**

	New Average	New Median	New SD
a) 4 is added to every number on the original list.	Write a **number.** 10	Write a **number.** 8	Write a **number.** 2
b) Every number on the original list is multiplied by **negative** 3.	Write a **number.** -18	Write a **number.** -12	Write a **number.** 6
c) Change every number to a Z score by subtracting 6 and dividing by 2.	Write a **number.** 0	Write a **number.** -1	Write a **number.** 1
d) Every number on the original list remains the same, EXCEPT that 20 is added to the largest number.	**Choose one:** (i) Increases ii) Decreases iii) Stays the same	**Choose one:** i) Increases ii) Decreases (iii) Stays the same	**Choose one:** (i) Increases ii) Decreases iii) Stays the same

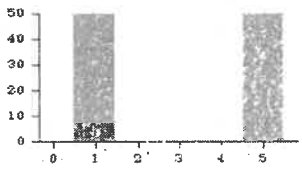
Ch. 4 Rules

Question 10 (3 pts.) **pertains to the 3 histograms below. Each histogram represents a data set of 10 whole numbers.**

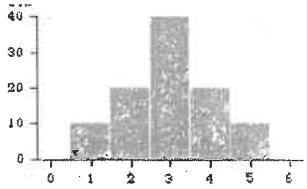
Histogram A



Histogram B



Histogram C



a) (1 pt.) Which histograms have an average of 3? ***Choose one:*** i) Only A ii) Only B (iii) All iv) None

b) (1 pt.) Which histogram has a SD = 0? ***Choose one:*** (i) Only A ii) Only B iii) All iv) None

c) (1 pt.) Which histogram has the largest SD? ***Choose one:*** i) A (ii) B iii) C

Page 6 - Stat

Question 11 (22 pts.) According to our survey, the weights of Stat 100 females follow the normal curve with **Average = 135 lbs and SD=24 lbs.** (*You may "round" Z scores and percentages* to fit the closest line on the Normal table and you may round percentages on the table to the nearest whole number.)

a) (7 pts.) What percentage of the female students weigh over 177 lbs?

i) (2 pts.) First, convert **177 lbs** to a Z-score. **Z-score =** 1.75 **(Show work, no work no credit)**

$Z = \frac{val - avg}{SD} = \frac{177-135}{24} = 1.75$

-½ for careless error

ii) (2 pts.) Which histogram's shaded region correctly depicts the percentage of women who weigh over 177 lbs?

Choose one:

a) Histogram A b) Histogram B c) Histogram C d) Histogram D

iii) (2 pts.) What percentage of women weigh over 177 lbs.? 4 % Show calculation.

cont error from i

iv) (1 pt.) Women who weigh 177 lbs are at the 96th percentile of the weight distribution.

cont error from iii

b) (3 pts.) If a student is 0.5 SD's ***below*** average in weight. How much does she weigh and what percentile is she in?

i) (1 pt.) She weighs 123 lbs. **(Show work)**

$Z = -0.5$

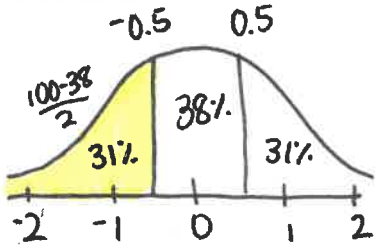
$val = avg + (z)(SD)$

$= 135 + (-0.5)(24)$

$= 123$

ii) (1 pt.) She's in the 31st percentile.

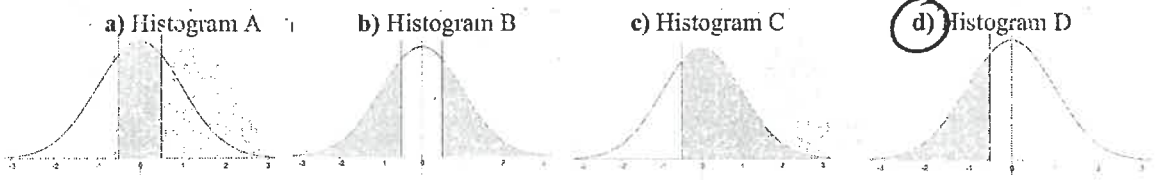
accept not rounded answer



iii) (1 pt.) Which histogram's shaded region correctly depicts the percentile?

Choose one:

a) Histogram A b) Histogram B c) Histogram C d) Histogram D



c) (6 pts.) If a student is in the 60th percentile how much does she weigh?

i) (3 pts.) To find the Z score, you should look at the middle area = 20 % *Fill in the blank with the correct percent.*

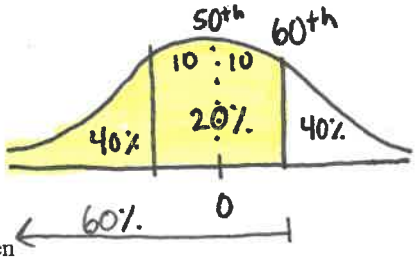
ii) (1 pt.) Z= 0.25

iii) (2 pts.) Weight = 141 lbs. **(Show work.)** continued error from ii

$val = avg + (z)(SD)$

$= 135 + (0.25)(24)$

$= 141$



d) (2 pts.) If you're below average in weight, is your Z score positive or negative?

Choose one: i) positive ii) negative iii) not enough information given

e) (2 pts.) If you're exactly at the 50th percentile in weight then your Z score = 0 and your weight = 135 lbs. (Fill in the two blanks with numbers.)

exactly average

f) (2 pts.) If 2 people have the same Z scores in absolute value but opposite signs then their percentiles must sum to

Choose one: i) 0 ii) 25 iii) 50 iv) 100 v) Not enough info.

* if 2 percentiles sum to 100, your z-scores are opposites