

**Question 1 (10 pts)** A study published in the August 15, 2017 issue of *Mayo Clinic Proceedings* tracked 44,000 people aged 20 to 87 for an average of about 16 years and found that those who drank 4 or more cups of coffee a day were 21% more likely to die than those who drank less than 4 cups a day. The risk was 50% higher for heavy coffee drinkers under 55 years of age.

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

- ☒ i) An observational study with controls
- ☐ ii) A randomized controlled experiment
- ☐ iii) A non-randomized experiment with historical controls

b) (2 pts.) Does the study show that drinking 4 or more cups of coffee a day caused the higher death rate?

- ☐ i) No, the study was conducted over such a long time period that it's difficult to determine whether it was the original coffee drinking itself or something *else* about the coffee (for example, the way it was brewed) that caused the higher death rate.
- ☐ ii) Yes, particularly for young people, the study clearly shows that excessive coffee drinking caused an increased risk of death.
- ☒ iii) No, it's possible that coffee drinkers share other traits (besides the coffee) that could put them at a higher risk of dying.
- ☐ iv) No, you cannot conclude causation without a proven causal mechanism. The study does provide strong evidence that it's the coffee that's raising the death rate and not something else, but it fails to explain how or why.

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

- ☐ i) At the beginning of the study, they divided the patients into smokers and non-smokers and then randomly divided the smokers and non-smokers equally between the coffee and no coffee groups.
- ☐ ii) Throughout the study they eliminated anyone who smoked from the study.
- ☒ iii) At the end of the study, they stratified on smoking, and compared the death rate of coffee drinkers to non-coffee drinkers within each smoking level (non-smokers, light smokers, heavy smokers).

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

- i) Increased popularity of coffee- The study was conducted over a 16-year time period that coincided with an enormous increase in coffee consumption. a) confounder b) causal link ☒ c) neither
- ii) Caffeine—Excessive caffeine intake from 4 cups of coffee per day raises health risks because it increases a person's heart rate and blood pressure, which increase one's risk of death. a) confounder ☒ b) causal link c) neither
- iii) Unhealthy Diet – The study stated that people who drank 4 or more cups of coffee were also more likely to have an unhealthy diet that could increase one's risk of death. ☒ a) confounder b) causal link c) neither
- iv) Pre-existing-conditions- Some members of the study may have had pre-existing conditions or illness that would cause them to die sooner. a) confounder b) causal link ☒ c) neither.

**Question 2 (4 pts.)** A country club gives a pass-fail golf test every year to professional and amateur golfers. Professionals have a much higher % passing than amateurs. The club members were happy that the overall % passing went up from 68% in 2007 to 70% in 2017 and wanted to know which group contributed to the improved rate.

	2007				2017			
	Number	# Passes	# Failures	% Passing	Number	# Passes	# Failures	% Passing
Professionals	100	92	8	92%	100	90	10	90%
Amateurs	300	180	120	60%	100	50	50	50%
Overall Total	400	272	128	68%	200	140	60	70%

a) (2 pts.) Which group's % passing went up from 2007 to 2017? **Choose one:** a) Prof. b) Amat. ☒ c) Neither d) Both

b) (2 pts.) Is it possible for each group's % passing to go down if their overall % passing goes up?

☒ i) Yes, it's possible because the overall makeup of the club has changed from 25% to 50% professionals which raises the overall % passing even though both groups % passing declined.

ii) No, it's not possible. If the overall passing rate goes up, then at least one group's passing rates must go up.

**Question 3** pertains to the following study: (6 pts.)

A study was done to test whether Ginkgo biloba (GB) could alleviate symptoms of Alzheimer's and dementia. The 52-week study randomly assigned half of the patients take GB daily and half to take a placebo. Neither the subjects nor evaluators knew who was in each group. At the end of the study, there was significant evidence that GB improved the cognitive performance and the social functioning of the patients for 6 months to 1 year.

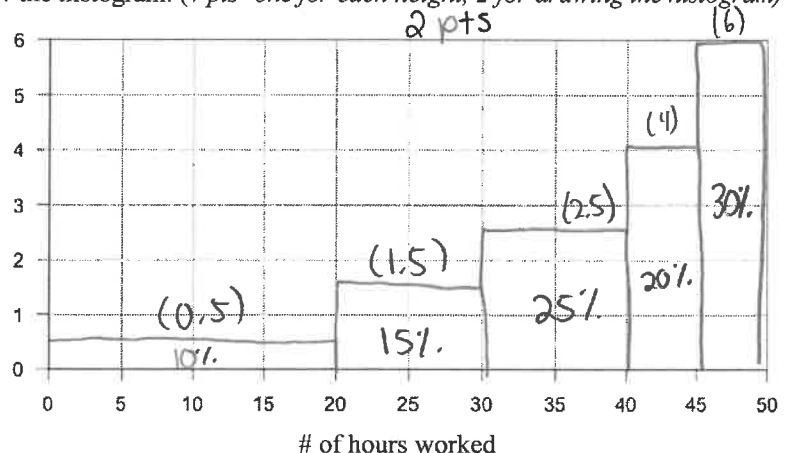
- a) (2 pts.) What type of bias could be present in this study **Choose one:**  
☒ i) No systematic bias   ii) Subject Bias   iii) Evaluator Bias   iv) Selection Bias   v) ii, iii, and iv
- b) (2 pts.) Which of the following could confound the results? **Choose one:**  
 i) Forgetfulness-- Patients with dementia may forget to take the GB on a regular basis.  
 ii) Increased Attention-- Participation in the study increased the attention these patients received. They felt less neglected and therefore more cognitively active.  
 iii) More motivated-- Those who volunteered to be in the GB group were probably more conscientious and motivated to begin with since they actively sought a remedy for their condition.  
☒ iv) All of the above  
☐ v) None of the above
- c) (2 pts.) Not everyone in the treatment and control group adhered to the program and took their medicine/placebo. Which comparison is best when analyzing the final data?  
☒ i) Compare everyone assigned to take the GB to everyone assigned to take the placebo.  
 ii) Compare everyone who actually took the GB to the placebo group.  
 iii) Compare only those who took the GB regularly to only those who took the placebo regularly.

**Question 4** (14 pts)

Below is a distribution table for the number of hours per week that employees at a company work. The first row says that 10% of the employees work 0 to 20 hours per week.

- a) Fill in the column for the height of the blocks then draw the histogram. (7 pts- one for each height, 2 for drawing the histogram)

Hours	%	Height of Block (% per work hour)
0-20	10	0.5 1pt
20-30	15	1.5 1pt
30-40	25	2.5 1pt
40-45	20	4 1pt
45-50	30	6 1pt



- b) (2 pts) What is the median number of hours worked? 40 hours
- c) (2 pts) Is the average less than, greater than, or equal to the median?  
**Circle one:** i) greater than   ☒ ii) less than   iii) equal to
- d) (3 pts) Some of the employees actually worked over 50 hours per week. If we changed those who worked 50 hours per week to numbers greater than 50, how would the median, average and SD change?

The median would: a) increase   b) decrease   ☒ c) stay the same

The SD would: ☒ a) increase   b) decrease   c) stay the same

The average would: ☒ a) increase   b) decrease   c) stay the same

For the following data sets below check whether you think the histogram would have a long left-hand tail, a long right-hand tail, or be fairly symmetrical.

*Next to each survey question check the box that best describes its histogram.*

Survey Question	Long Left-Hand Tail	Long Right-Hand Tail	Fairly Symmetrical
a) Family income in the U.S. today		✓	
b) Exam scores in a class where the average is ten points lower than the median.	✓		
c) IQ scores where the average and median are about 100.			✓

**Question 6 (4 pts)**

Suppose a set of SAT scores have an average of 540 and a SD of 60. Fill in all 4 blanks with numbers (NOT words!).

- a) If 10 points were added to each score, the new average would be 550 and the new SD would be 60
- b) If 10% were added to each of the original scores, the new average be 594 and the new SD would be 66  
(\*adding 10% is the same as multiplying by 1.1)

**Question 7 pertains to the following 2 lists of numbers (13 pts)**

List A : 0, 6

- a) The average of List A = 3
- b) The median of List A = 3
- c) The deviations from the average of List A are: -3, 3
- d) (2 pts) Compute the SD of List A. (No work, no credit.) Circle your answer. 1 pt work, 1 pt answer

$$\sqrt{\frac{9+9}{2}} = \textcircled{3}$$

List B: 0, 3, 6

- a) The average of List B = 3
- b) The median of List B = 3
- c) The deviations from the average of List B are: -3, 0, 3
- d) (2 pts) Compute the SD of List B. (No work, no credit.) Circle your answer and round to 2 decimal places.

$$\sqrt{\frac{9+9}{3}} = \textcircled{2.45}$$

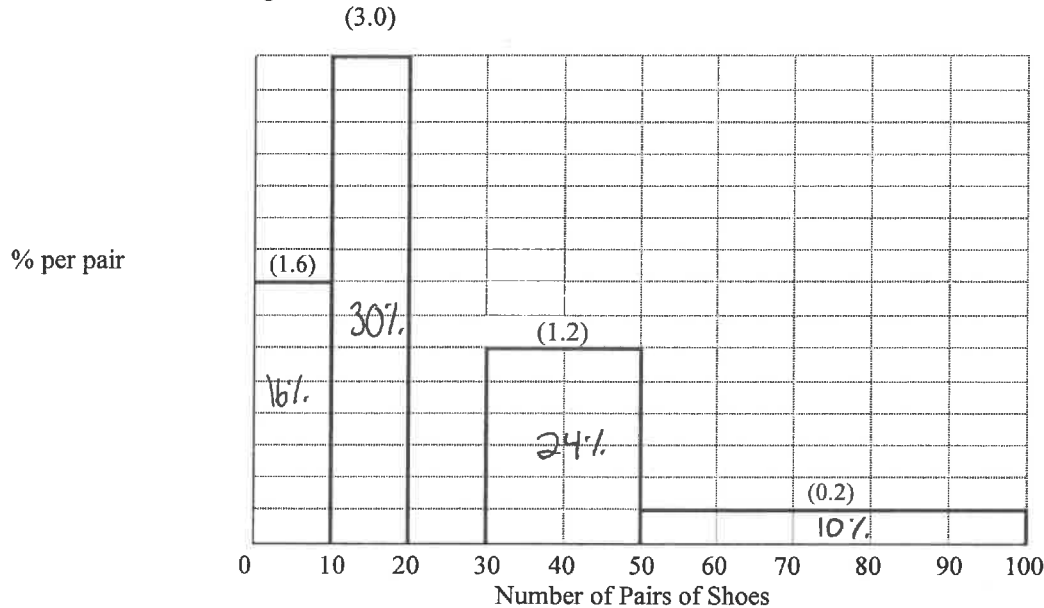
or 2.44

1 pt work, 1 pt answer

**Question 8 pertains to the histogram below. (18 pts)**

The figure below is a histogram for the number of pairs of shoes owned by 400 college women (roughly based on past survey data). Class intervals include the right-endpoint but not the left. (For example, someone who owns 10 pairs of shoes would fall in the 0-10 block not the 10-20 block.) The height of each block is given above the block in parentheses.

Assume even distributions throughout each interval.



- a) (4 pts) What percent of the women owned shoes in the intervals below?

0-10 pairs 6 %      10-20 pairs 30 %      30-50 pairs 24 %      50-100 pairs 10 %

- b) (2 pts) The block over the 20-30 interval is missing. How tall must it be? 2

- c) (2 pts) What **percentile** is the median? 50<sup>th</sup> percentile. Fill in the blank with a number.

- d) (1 pt) Is the average greater than, less than or equal to the median? avg > med  
**Circle one:** ☒ i) greater than      ii) less than      iii) equal to

- e) (2 pts) Which interval has **less** people, 30-50 or 50-100, or are they the same?

**Circle one:** i) 30-50      ☒ ii) 50-100      iii) same

- f) (2 pts) The percent of subjects in the study who own 5 pairs of shoes is closest to.....

**Circle one:** ☒ i) 1.6%      ii) 3%      iii) 5%      iv) 8%      v) 15%

- g) (2 pts) Would it be appropriate to use the normal approximation with this data to calculate percentages within different intervals? **Circle one:**

- i) Yes, if we knew the average and the SD we would lose little accuracy in using the normal curve.  
 ii) Yes, normal approximations are always accurate in approximating percentages.  
 iii) Yes, because even though the histogram is far from normal, it will become normal after the data is converted to z-scores.  
☒ iv) No, this histogram is not close to normal, so the normal approximations would not be close to the real percentages.

- h) (3 pts) Suppose the Stat 100 team gives every single woman 2 pairs of shoes. How would that affect the SD, median, and average in the histogram above?

- The SD would: i) increase      ii) decrease      ☒ iii) stay the same
- The median would: ☒ i) increase      ii) decrease      iii) stay the same
- The average would: ☒ i) increase      ii) decrease      iii) stay the same

**Rounding instructions for Question 9:** 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.

**Question 9**

The histogram of the heights of the 668 women who filled out last semester's survey roughly follows the normal curve with an average of 64 inches and a SD of 4 inches.

- a) (2 pts.) Approximately 68% of the heights are between 60 inches and 68 inches  
(Put the smaller number first) 1 pt 1 pt

- b) (2 pts.) How tall is someone who is 2 SD's above average? 72 inches  
 $z = 2$

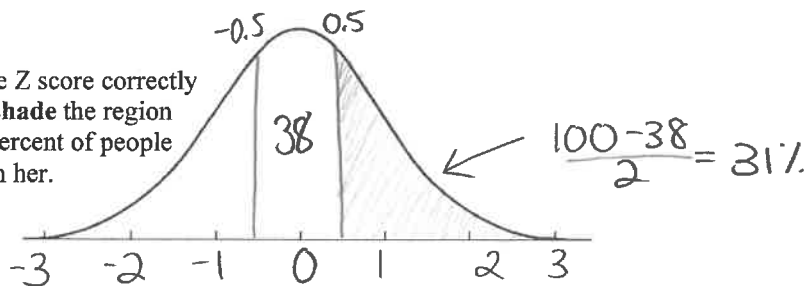
- c) (3 pts.) If a student is 66", what is her Z score and what percent of the people are taller than her?  
Show work and write your answers in the blanks provided.

- i) (1 pt) Z score = 0.5 ii) (1 pt) 31 % taller  
cont error

Show work:

$$z = \frac{66 - 64}{4} = 0.5$$

- iii) (1 pt) Mark the Z score correctly on the curve and shade the region representing the percent of people who are taller than her.

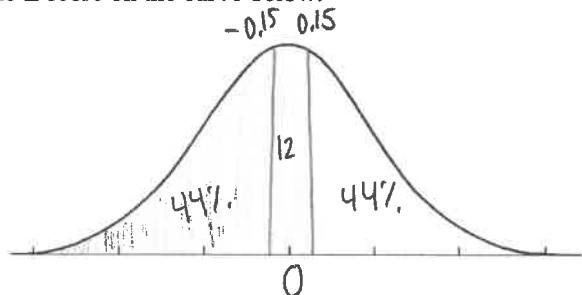


- d) (2 pts) What percentile is 66 inches? (no work necessary- use your work from part c)

$$38 + 31 = \textcircled{69^{\text{th}}}$$

- e) (3 pts) If a student is in the 56<sup>th</sup> percentile how tall is she in inches? Fill in the 3 blanks below.

56<sup>th</sup> percentile corresponds to middle area = 12 %,  $\rightarrow$  Z score = 0.15 and her height = 64.6 inches  
Mark the Z score on the curve below. (don't round your answer)



cont error

$$\begin{aligned} \text{Value} &= \text{avg} + (z)(\text{SD}) \\ &= 64 + (0.15)(4) \\ &= 64.6 \end{aligned}$$

- f) (2 pts) If a student is in the 44<sup>th</sup> percentile, how tall is she in inches? 63.4 inches. (Hint: See part e)

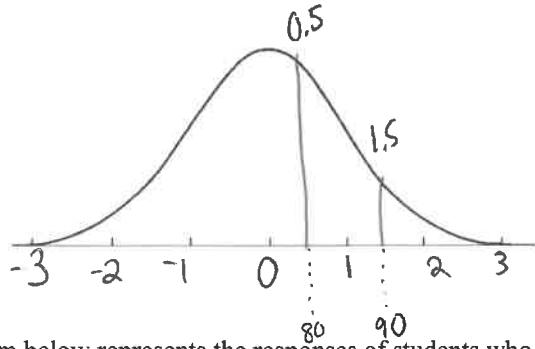
$$z = -0.15$$

(don't round your answer)  
cont error

$$\begin{aligned} \text{val} &= 64 + (-0.15)(4) \\ &= 63.4 \end{aligned}$$

**Question 10** The histogram of the final scores in a large course looked reasonably like a normal curve. The average score was 75 and the SD of the scores was 10. The grades were assigned based on the cutoffs below. Use the normal approximation to answer the following questions. Pick the choice that's closest to your answer. Answers may vary slightly based on rounding.

A	90 or above
B	80-90
C	70-80
D	60-70
F	60 or below



a) (2 pts) The percentage who received a B or higher (80 or above) is closest to:

- i) 40% ii) 38% **iii) 31%** iv) 25% v) 20%

b) (2 pts) The percentage who received a B (80-90) is closest to:

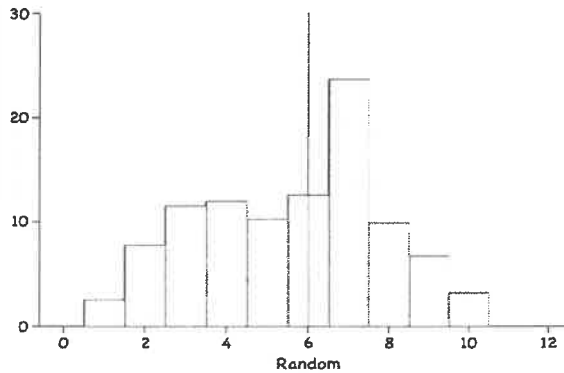
- i) 13% ii) 22% **iii) 24.5%** iv) 38.2% v) 43%

**Question 11** (7 pts)

**Part I--** The histogram below represents the responses of students who answered the following question on Survey 1:  
Pick a random number from this list: 1,2,3,4,5,6,7,8,9,10

\*\*\*\*\*The line shown divides the histogram exactly in half by area.

% per number



a) (1 pt) The percent of the class who chose 7 is closest to ...

- i) 15% ii) 20% **iii) 24%** iv) 28% v) 33%

b) (1 pt) The median = 6 (Give a number.)

**Part II--** Now suppose everyone randomly chose their numbers by a spinner where each number was equally likely.

a) (2 pts.) What percent of the class would you expect to get "7". 10 %.

b) (2 pts.) You would expect the shape of the histogram to be close to ...

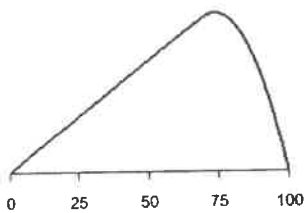
- i) Normal **ii) Uniform with every block the same height** iii) the histogram above

c) (1 pt) How would the SD of the survey and spinner results compare? Which would be larger?

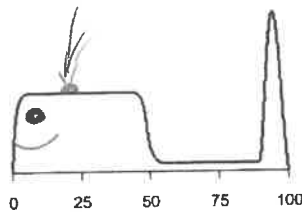
- i) Spinner** ii) Survey iii) The same

### Whale Question

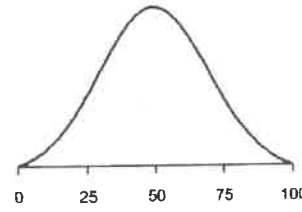
Below are sketches of histograms for test scores in 3 different classes. The scores ranged from 0 to 100.



A



B



C

a) (2 pts) Which histogram most closely resembles a whale?

Choose one: A B C

\*Feel free to draw additional features on your histogram to support your answer. Get creative- any answer could convince us!

accept any answer :)