The next question refers to the list {0, –2, 10, –5 , 5, 16, –3, –13}.

1. What are the mean, median, standard deviation, and inter-quartile range of the list?

mean = 8

median ⇒ $-13, -5, -3, -2 \mid 0, 5, 10, 16$

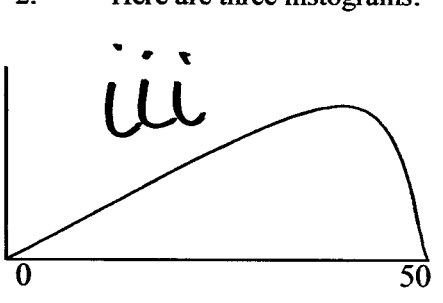
$= -1$ $\quad \frac{-2+0}{2} = -1$

SD ⇒ $\sqrt{\frac{588 - \left(\frac{64}{8}\right)}{8}} = 8.5147$

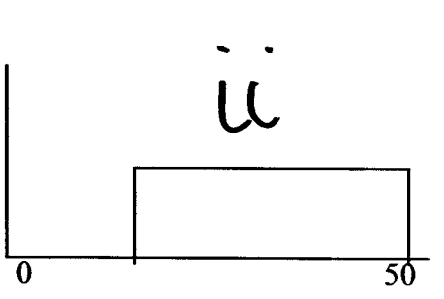
(where $0^2 + -2^2 + 10^2 + -5^2 + 5^2 + 16^2 + -3^2 + -13^2 = 588$)

Interquartile Range = $\left(\frac{10+5}{2}\right) - \left(\frac{-5+-3}{2}\right) = 7.5 - (-4) = 11.5$

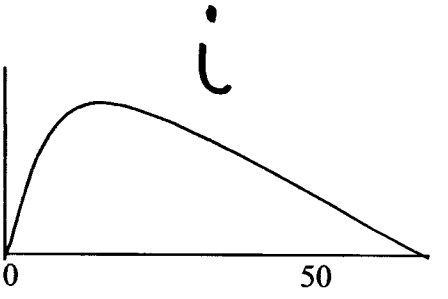
2. Here are three histograms:



(A)



(B)



(C)

Match **each** histogram above to the best choice listed below:

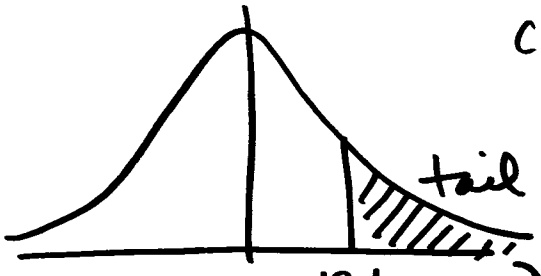
i. The average is bigger than the median.
ii. The average is approximately the same as the median
iii. The average is smaller than the median
iv. Cannot determine the median for this graphic
v. Cannot determine the average for this graphic

3. Indicate whether the following statements are true or false

	T	F	Statement
A.	X		Randomization is necessary to prevent selection bias in experiments
B.	X		Control Groups are necessary in experiments so we can compare the results from a treatment group properly
C.		X	Control Groups in experiments are always given a placebo
D.	X		Observational studies can establish association
E.	X		Observational studies are misleading due to confounding
F.	X		Confounding in experiments is a result of selection bias

The next four questions use information from this statement, but each question is separate (i.e. you can get the first one wrong and it won't affect the others): The Medical College Admissions Test (MCAT) is constructed to be normally distributed with a mean of 9 and a standard deviation of 2. Approximately 20,000 people take the test every year. SHOW YOUR WORK FOR FULL CREDIT.

4. Harvard Medical School only considers applicants with a mean of 12.1 or greater. How many of the test takers qualify for Harvard?



convert 12.1 to z

$z = \frac{12.1-9}{2} = 1.55$

means 87.89% in the middle

so

$\frac{100-87.89}{2} =$ 6.055% qualify

5. The lowest 1% of test takers enroll at Dr. Nick Riviera's School of Medicine. At and below what score is the lowest 1%?

↑ lowest 1% implies 98% in the "middle"

for table A 105

for 98% in the "middle"

$z \approx 2.35$

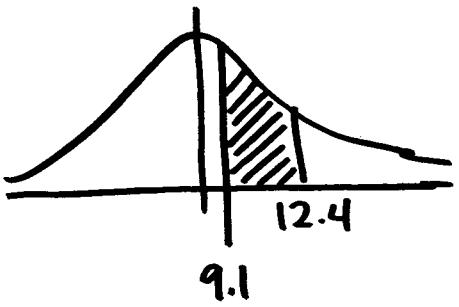
$-2.35 = \frac{X-9}{2}$

solve for X

gives ≈ 4.3

so the lowest 1% below 4.3

6. You decided to take the MCAT and got an 12.4. Your cousin, who went to USC, also took the MCAT got a 9.1. What percentage of test takers have scores between yours and your cousin's?



① $\frac{12.4-9}{2} = 1.70 = z$ area is 91.09

you

divide by 2 = 45.55%

② $\frac{9.1-9}{2} = .05 = z$ area is 3.99

COUSIN

divide by 2 ≅ 2.0

area between ⇒ 45.55 − 2.0 = 43.55 ≈ 43.6%

subtract areas.

7. After thinking it over, you decide not to apply to Medical School, but apply to Law School instead. And to your surprise, the UCLA Law School is willing to consider applicants with a valid MCAT score – with the following condition: All applicants must add 23 to their MCAT Score first and then multiply that score by 5. So for example, you got a 12.4, your new score is 177.

If you apply UCLA rules to ALL the MCAT scores, what are the new mean, median and standard deviation?

OLD MEAN = 9

OLD MEDIAN = 9 (because normal mean = median)

OLD SD = 2

New mean = (23+9) * 5 = 160

New median = (23+9) * 5 = 160

New SD = 2 x 5 = 10

standard deviation is not affected by addition

8. A study was conducted on the sleep patterns of infants in the United States. A sample of 25 infants was drawn at random with an average hours slept of 13 hours in a 24-hour period. Suppose the standard deviation is 3.5 hours.

a. The population for this study is (verbal answer)

ALL US INFANTS

b. The sample for this study is size (numeric answer)

25

c. The desired parameter being estimated by this study is (verbal answer)

Average hours slept in 24 hours

d. The statistic calculated from the study is (numeric answer)

$\bar{X} = 13$ hours

OR

$S = 3.5$ hours

9. In a hypothetical experiment, a new drug was compared with "standard therapy" treatment for patients suffering from inoperable cancer. The result measured was difference in survival time (in months). Which of the following best describes the primary reason to randomize patients into treatment or control groups? (Choose one best answer)

(a) to create two groups that are similar at baseline on both known and unknown factors associated with survival time.

(b) prevent bias introduced when the patients know what type of treatment they are receiving

(c.) prevent bias introduced when the investigators know what type of treatment the patients are receiving

(d.) Both b and c

To determine the prevalence of sexually transmitted diseases (STD) and high risk sexual behavior for STD among adolescent males admitted to a juvenile detention facility, a survey was obtained from interviews. The results are tabled below.

Table 1. Behavioral variables in 966 subjects

Variable	Mean (SD)	Range	Median
Age at first coitus	12.3 (2.0)	6-18	13
No. lifetime partners	13.7 (16.8)	1-100	8
No. partners past 4 months	2.9 (3.4)	0-30	2
No. weeks since last sex	5.8 (15.1)	1-260	2

Note: SD is the abbreviation for standard deviation

10. Of the four variables in the table, which has the most symmetrical (normal-like) distribution based on the statistics presented? (Choose **one** best answer.)

a. age at first coitus
b. number of lifetime partners
c. number of partners in the past 4 months
d. number of weeks since last sex
e. none of the above, they are all non-normal based on the statistics given

11. All of the variables in Table 1 are quantitative and discrete (circle one)

TRUE FALSE Bad question

The next three questions refer to the list {-8, -5, -3, 0,1, 3, 0, 4}.

12. The mean is:
    (a) -1.0
    (b) 0
    (c) 0.5
    (d) 1.0
    (e) none of the above.

-8, -5, -3, 0, 0, 1, 3, 4

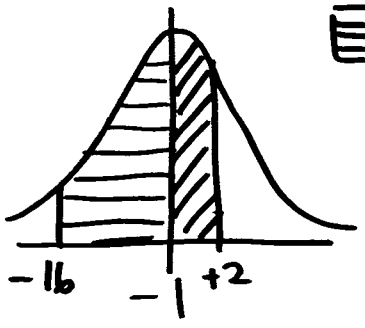
13. The median is:
    (a) –1.0
    (b) 0
    (c) 0.5
    (d) 1.0
    (e) none of the above.

14 The inter-quartile range is:
    (a) -4
    (b) 4
    (c) -6
    (d) 6
    (e) none of the above.

15. The SD can never be larger than the mean:
    (a) True
    (b) False

The next three questions refer to this statement, but each question is separate (i.e. you can get the first one wrong and that won't affect the others): Corporate securities (or publicly traded stocks) are an investment opportunity for individuals as well as institutions. The 10,000 stocks available for investment to U.S. residents are normally distributed with a mean one-year return of -1% and a standard deviation of 12%. SHOW YOUR WORK FOR FULL CREDIT.

16. What percentage of stocks had one-year returns between -16% and +2%? (5 points)



between -16 and -1 $\Rightarrow$ $z = \frac{-16-(-1)}{12} = -1.25$

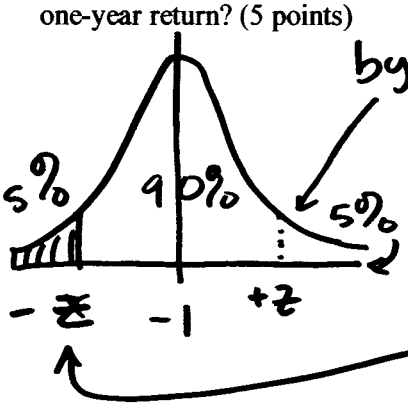
area is $\frac{78.87}{2} = 39.435$

between -1 and $+2$ $\Rightarrow$ $z = \frac{+2-(-1)}{12} = 0.25$

area is $\frac{9.74}{2} = 9.87$

combined area is $39.435 + 9.87 = 49.31\%$

17. A stock is at 5th percentile (i.e. 5% of the stocks have returns lower than this stock), what is its one-year return? (5 points)

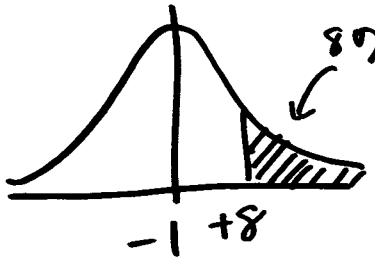


by symmetry, 5% in the OTHER tail gives 90% in the middle. A z w/ 90% in the middle is ± 1.65 and -1.65

$-1.65 = \frac{x-(-1)}{12}$ solve for x

$x \approx -20.8\%$

18. In order to meet your retirement goals, you need to buy stocks that have a return of 8% or more. Approximately how many stocks out of the 10,000 qualify? (5 points)



$z = \frac{+8-(-1)}{12} = +.75$

table gives 54.67 in the middle

so upper tail is $\leftarrow$

$\frac{100-54.67}{2} = 22.665\%$ and

22.665% of $10{,}000 = 2{,}266.5 \approx$ 2267 stocks

Ten temperature readings from a freezer unit (in degrees Fahrenheit):

-9 -10 +1 +2 -5 0 0 +3 -11 +1

19. What are the average and median of this list? (5 points)

A. average = -2.8 and median = 0
B. average = -2.8 and median = -2.5
C. average = -3.5 and median = 0
D. average = -3.5 and median = -2.5
E. average = 0 and median = -3.5

20. The formula for converting temperature readings into degrees Celsius (a temperature reading on the metric scale) is:

Celsius = 5/9 (Fahrenheit - 32)

Which statement below is correct about a dataset of temperature readings after the change of scale from Fahrenheit to Celsius? (5 points)

A. The mean will change, but the median will remain the same
B. The mean and median will change, but the standard deviation will remain the same.
C. The range will remain the same but the standard deviation will change.
D. The mean and standard deviation will change, but the median will remain the same.
E. All of the above are false.

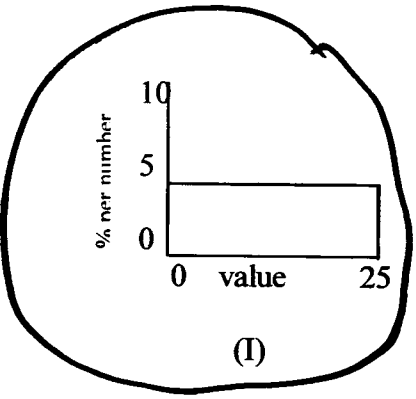
21. A medical observational study investigated the safety of anesthetics used in surgery. Records of over 1,000,000 operations performed in major hospitals showed the following death rates:

Anesthetic	1	2	3
Death Rate	1.7%	1.8%	5.6%

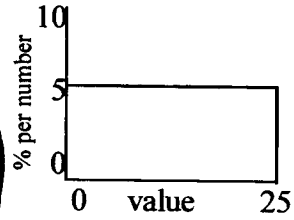
Clearly, anesthetic 3 appears dangerous, but follow-up studies showed that we could not conclusively determine this. What kinds of problems might be present in the original study?

A. Confounding Factors
B. Bias
C. Simpson's Paradox
D. A, B, and C are present
E. Only 2 of the 3 (A, B, C) are present

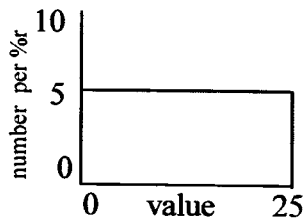
22. Three students sketched the histogram for lottery numbers which are uniformly (or evenly distributed) ranging from 0 to 25, only one histogram is correct, which one? (circle only one of them)



(I)



(II)



(III)

23. Explain the choice you made in problem #22. Why is it right? (be brief your answer should fit in the space below)

I. → 25 × 4 = 100% total area and Y axis is properly labeled

II is wrong because 25 × 5 = 125%

III is incorrectly labeled and 25 × 5 = 125%

24. A professor made a careful sample survey to estimate the percentage of USC undergraduates living at home. Two assistants were stationed at the "Tommy" Trojan statue (it's on the main plaza) and instructed to interview all students who passed by at specified times. Many students would not speak with the assistants, in fact, only 369 out of 1500 approached, did. As it turned out, 39% of 369 students interviewed said they live at home. Does the investigator's procedure give a probability sample of USC students?

First, answer yes or no and then explain your reasons for your choice. This does not need to be a long answer.

No. Knowing nothing else we can say Selection Bias is present here (not all students will pass through the main plaza)

Non Response Bias is present, only $\frac{369}{1500}$ for less than 50% responded