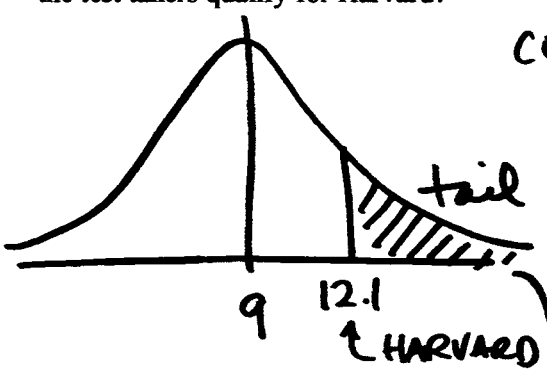
1. Indicate whether the following statements are true or false

	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.

2. 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

3. 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"

$z \approx 2.35$

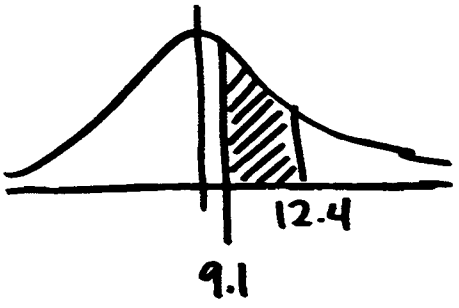
$-2.35 = \frac{X-9}{2}$

solve for X

gives ≈ 4.3

so the lowest 1% below 4.3

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

divide by 2 = 45.55%

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

divide by 2 ≅ 2.0

area between ⇒ 45.55 − 2.0 = 43.55 ≈ **43.6%**

subtract areas.

5 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

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

6. 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

7. All of the variables in Table 1 are quantitative and continuous (circle one)

TRUE FALSE

10/21/02

Statistics 10 Exam 1 Review Exam is [redacted]

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

8 [redacted]. 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

9 [redacted]. The median is:

(a) –1.0
(b) 0
(c) 0.5
(d) 1.0
(e) none of the above.

10 [redacted] The inter-quartile range is:

(a) -4
(b) 4
(c) -6
(d) 6
(e) none of the above.

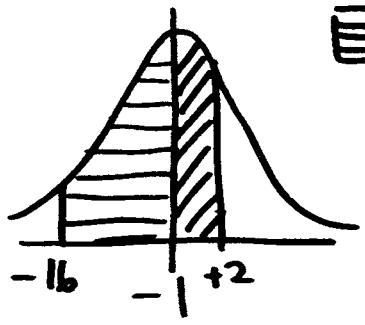
11 [redacted]. The SD can never be larger than the mean:

(a) True
(b) False

10/21/02

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.

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



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

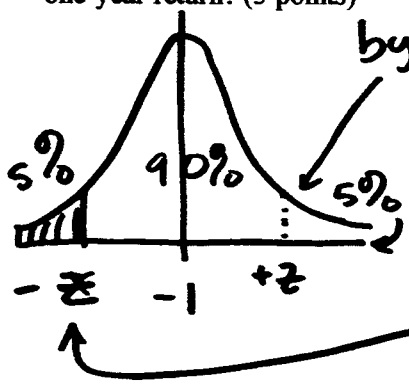
area is $\frac{78.87}{2} = 39.435$

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

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

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

13. 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)

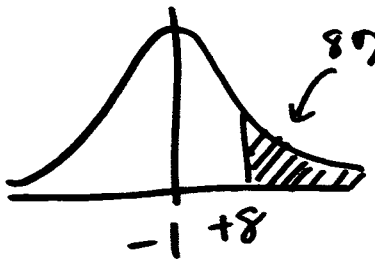


by symmetry, 5% in the OTHER tail gives 90% in the middle. A z w/ 90% in the middle is ≈ 1.65 and -1.65

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

$x \approx -20.8\%$

14. 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

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

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

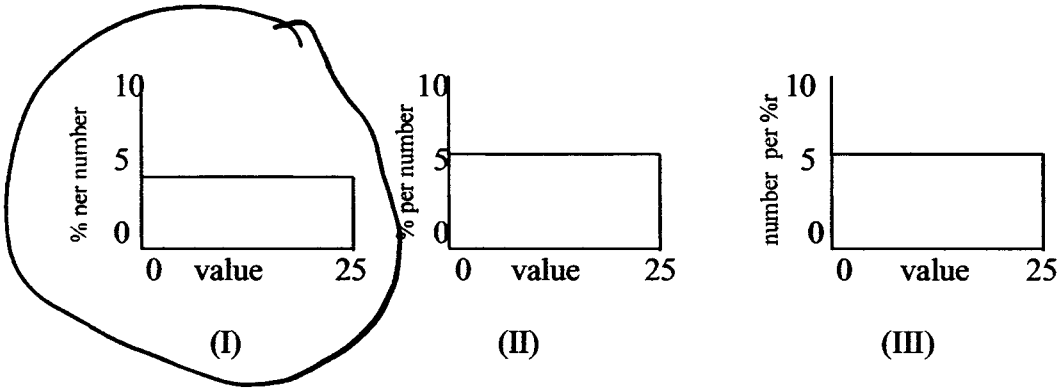
15. The formula for converting temperature readings into degrees Celsius (a temperature reading on the metric scale) is:

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.

16. 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)



17.Explain the choice you made in problem #16. Why is it right? (be brief your answer should fit in the space below)

I. $25 \times 4 = 100\%$ total area Y axis is properly labeled

II. is wrong because $5 \times 25 = 125\%$

III. is both incorrectly labeled AND $5 \times 25 = 125\%$

18. 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. 1) Selection Bias here b/c not all students have the same opportunity to be chosen

2) Non-response Bias is present, only 369/1500 responded (less than 50%)

19. Classify the following variables as either categorical or numerical by checking the correct box, if it is a numerical variable, further classify the variable as either discrete or continuous:

	Variable	Categorical	Numerical	Discrete	Continuous
A	Hair Color	✓			
B	Frozen Food Brand	✓			
C	Number of students in a classroom		✓	✓	
D	Your age		✓		✓

20. To study the effects of exercise on the grades of college students, a researcher wishes to compare the grade point averages of students at randomly selected colleges across the United States. The researcher selects students at random and after interviewing them to find out who exercises and who does not, chose 644 students of each (exercisers and non-exercisers). The researcher made sure the two groups of 644 were similar in racial composition, gender, major, and every subject had accumulated at least 120 units towards graduation. There were a total of 1,288 students in the study from approximately 40 colleges, their overall GPA was 3.22. The average GPA for the students who exercised was 3.34 and the standard deviation was .36.

a. What is the "treatment"?

Exercise (not assigned by researcher)

b. What is the response or outcome variable?

GPA

c. Is this an observational study or an experiment?

observational study

d. From this study, an example of a sample statistic is:

3.34 (sample mean) or .36 (sample SD)

e. What is the population of interest?

all college students

f. What is the parameter of interest in this study?

difference in GPA among exercisers + non exercisers

21. The next questions refer to the list {-4, -9, 0, -3}.

a. What is the mean of this list?

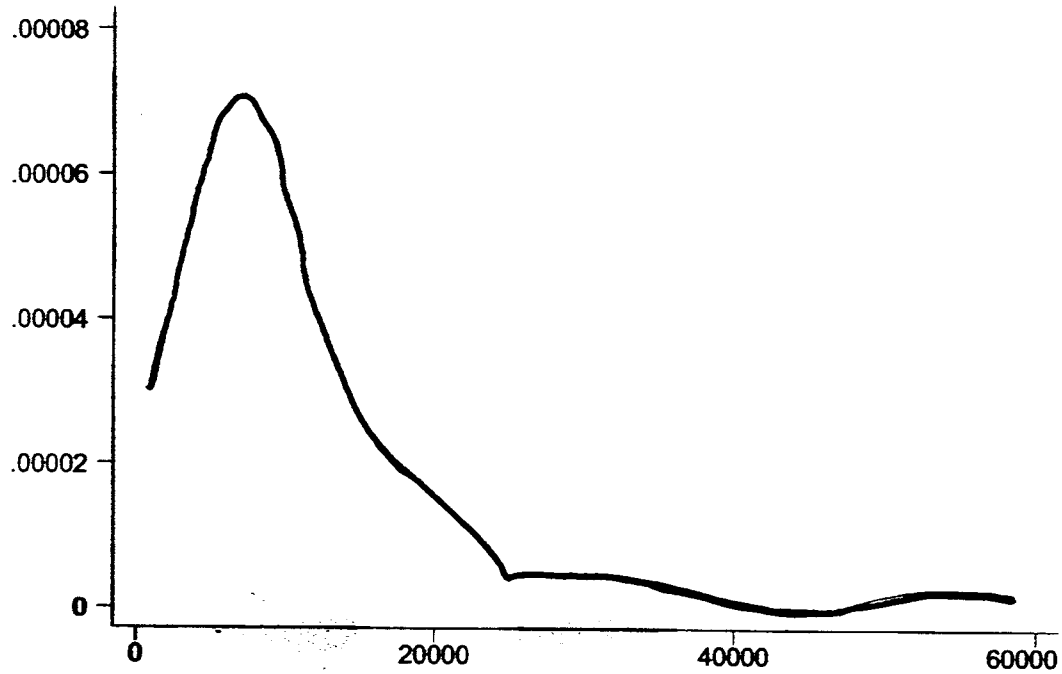
$$\frac{-4 + -9 + 0 + -3}{4} = -4$$

b. What is the standard deviation of this list?

$$SD = \sqrt{\frac{(-4 - -4)^2 + (-9 - -4)^2 + (0 - -4)^2 + (-3 - -4)^2}{4}}$$

$$= \sqrt{\frac{42}{4}} \approx 3.2404$$

22 (5 points) Here is a histogram, please assume it was correctly drawn

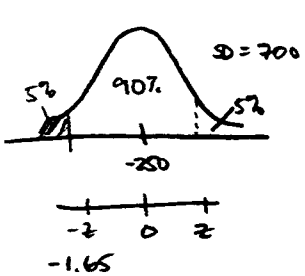


Please answer the following questions about this histogram:

	True	False	Statement
A	X		The mean is greater than the median
B	X		This histogram is right skewed
C		X	This histogram is left skewed
D	X		The total area is 1.0 or 100%
E	X		If the X axis represents dollars, the implied Y axis label is percent per dollar

The last three 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): A recent study showed that the gambling income of adults age 21 and over in the United States from all forms of legalized gambling (e.g. lottery, video poker, horse racing, casinos) is normally distributed with a mean of -250 dollars (a loss) and a standard deviation of $700. SHOW YOUR WORK FOR FULL CREDIT.

23 It is believed that the gamblers with the largest losses, that is those with the lowest 5% of gambling income, should be considered gambling "addicts" and given some kind of treatment. How much money does a gambling adult need to lose to be considered an "addict"? (5 points)



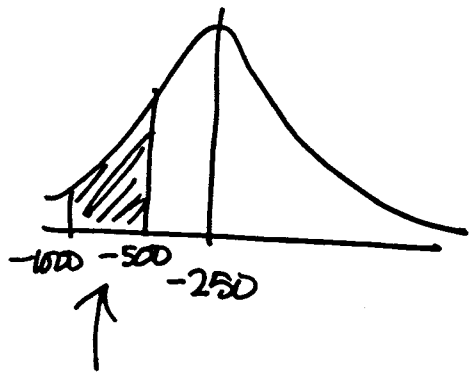
where area = 90%, $z = 1.65$

$$z = \frac{x-\mu}{\sigma} \qquad -1.65 = \frac{x-(-250)}{700}$$

$$-1155 = x + 250$$

$$x = -1405$$

24. What percentage of adults age 21 and over had gambling losses of at least $500 but not more than $1000? (7 points)



the area we want.

mean = -250

SD = 700

$$z_{(-500)} = \frac{-500-(-250)}{700} = -.36$$

about 27.37%

$$z_{(-1000)} = \frac{-1000-(-250)}{700} = -1.07$$

about 70.63%

$$\frac{70.63-27.37}{2} = \boxed{21.63\%}$$

25. What is the median gambling income?

If mean = -250

then median = -250

because mean = median

for a normally distributed variable