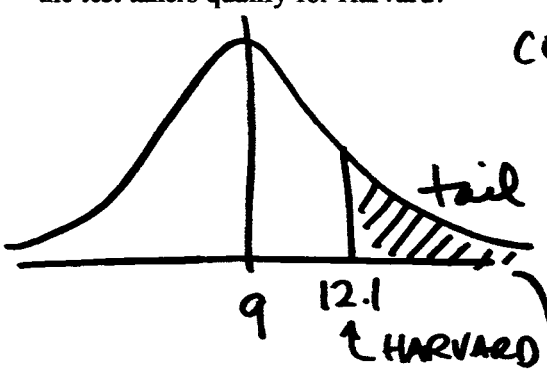
1. Indicate whether the following statements are true or false

	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"

$z \approx 2.35$

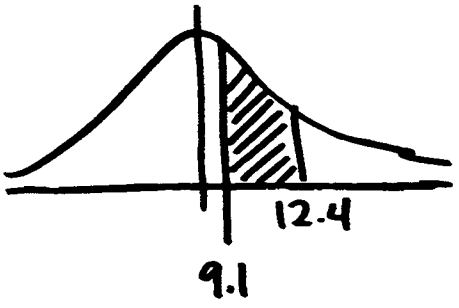
$-2.35 = \frac{X-9}{2}$

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.

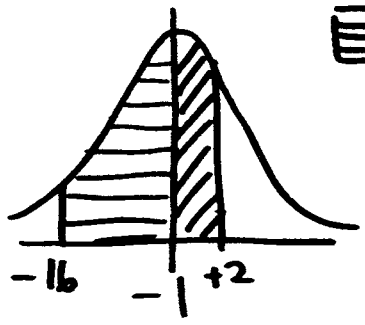
11 [redacted]. The SD can never be larger than the mean:

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

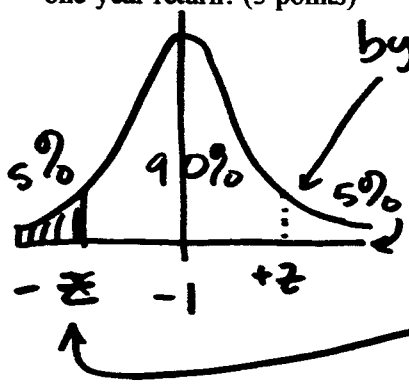
area is $\frac{78.87}{2} = 39.435$

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)

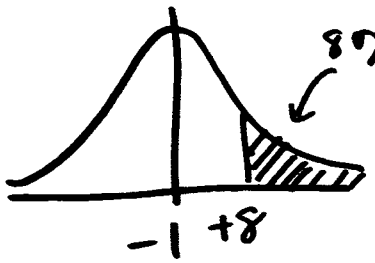


by symmetry, 5% in the OTHER tail gives 90% in the middle. A z w/ 90% in the middle is ≈ 1.65 and -1.65

$-1.65 = \frac{x-(-1)}{12}$ 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. 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 b/c not everyone has the same opportunity to get into the sample

2) Non-Response Bias is evident, only 369/1500 responded, much less than 50%

16. 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		✓		✓

17. 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

b. What is the response or outcome variable?

GPA

c. Is this an observational study or an experiment?

OBS. STUDY, EXERCISE WASN'T "ASSIGNED"

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 between exercisers + non-exercisers

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

a. What is the mean of this list?

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

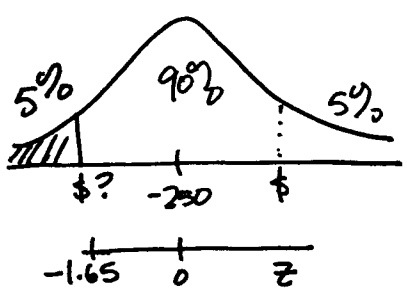
b. What is the standard deviation of this list?

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

$$\approx 3.24$$

The next 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.

19. 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"?

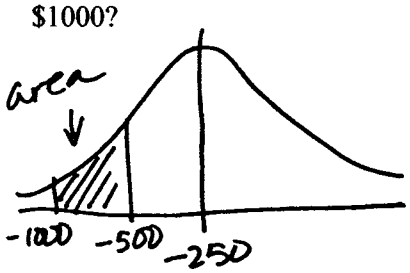


when "middle" = 90% $z = \pm 1.65$

$$-1.65 = \frac{X-(-250)}{700}$$

$$X = -1405$$

20. What percentage of adults age 21 and over had gambling losses of at least $500 but not more than $1000?



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

about 27.37%

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

about 70.63%

area is $\frac{70.63-27.37}{2}$

21. What is the median gambling income?

If mean = -250
Then median = -250
for normally dist variable

The Dull Computer Company manufacturers its own computers and delivers them directly to customers who order them via the Internet. Dull's market dominance has arisen from its quick delivery and competitive pricing. The CEO (Chief Executive Officer) of Dull has stated publicly that if customers make unassisted online purchases of their computers, the computers will have a mean delivery time of 45 hours from the time of purchase (with a standard deviation of 11 hours) and have a mean cost of $1,588 with a standard deviation of $401. Please assume that the cost of all the computers is normally distributed

A consumer research organization decided to test the CEO's mean delivery time claim by purchasing 100 computers from Dull at randomly selected times and days. The 100 purchases were randomly divided into two groups: 49 were purchased by telephone and involved talking to a live salesperson, the remaining 51 were unassisted online purchases. The delivery time of the 51 had a mean of 52 hours with a standard deviation of 16 hours and they also had a mean cost of $1,503 with a standard deviation of $678. 11 of the 51 computers were delivered in less than 45 hours.

22A. The population of interest to the consumer research organization is

(a) all Dull Computers
(b) all Dull Computers purchased online and unassisted
(c) 100 computers purchased from Dull by the consumer research organization
(d) 49 computers purchased by telephone
(e) 51 computers purchased online and unassisted

22B. The parameter of greatest interest to the consumer research organization is

(a) 52 hours
(b) 16 hours
(c) 45 hours
(d) 11 hours
(e) $1,503
(f) $1,588
(g) $678
(h) $401

22C. The sample of interest to the consumer research organization is

(a) all Dull Computers
(b) all Dull Computers purchased online and unassisted
(c) 100 computers purchased from Dull by the consumer research organization
(d) 49 computers purchased by telephone
(e) 51 computers purchased online and unassisted

22D. The statistic of greatest interest to the consumer research organization is

(a) 52 hours
(b) 16 hours
(c) 45 hours
(d) 11 hours
(e) $1,503
(f) $1,588
(g) $678
(h) $401

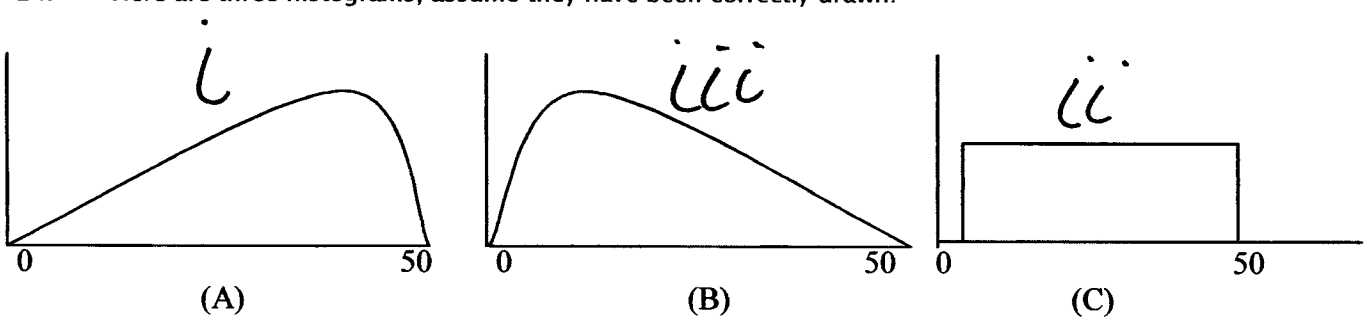
22E. The median cost of all Dull computers purchased online and unassisted is

i. $1, 503
ii. $1, 588
iii. Greater than $1,503 and also greater than $1,588
iv. Less than $1,503 and also less than $1,588
v. Greater than $1,503 but less than $1,588
vi. Not possible to calculate the median from the information provided

23. Please indicate whether each statement is true or false **(one point each)**

	True	False	Statement
A	X		The total area of a histogram is always 100% when area is expressed as percentages
B	X		Larger samples are no better than smaller samples at preventing bias
C	X		A histogram is a graphical summary which represents percentages as areas
D		X	Incorporating a density scale in a histogram means that the height of each block divided by the width of the class interval equals the percentage of cases in that class interval
E	X		The area under the histogram between two values is equal to the percentage of cases in a class interval defined by those values
F	X		In a randomized controlled experiment utilizing a placebo, if the control group is comparable to the treatment group, then the difference in the responses of the two groups is likely to be a result of the treatment
G		X	Double blind experiments are better at preventing the placebo effect than blind experiments
H		X	Confounding is not a source of bias
I		X	Probability methods are used in sampling because they are impartial <u>but</u> they cannot minimize bias
J		X	If a large number of persons selected for a sample do not respond, problems of response bias are likely non-response bias

24. Here are three histograms, assume they have been correctly drawn:



Match <u>each</u> histogram above to the best choice listed below: **(2 points each, 6 points total)**

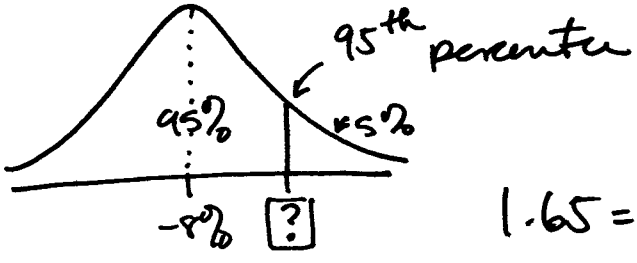
i. The average is smaller than the median.
ii. The average is equal to the median
iii. The average is larger than the median
iv. Cannot determine the average for this graphic
v. Cannot determine the median for this graphic

25. In an observational study (choose one) **(2 points)**

a. Investigators do not assign subjects to treatment or to control groups
b. There isn't a control group
c. Investigators can establish association but not causation
d. Confounding factors cannot be controlled
e. All of the above are true
f. Only A and C are true
g. Only D is false

26. The next questions refer to this statement, but each question is separate (i.e. you can get the first one wrong and it won't affect the others): Los Angeles has been suffering economically as a result of this year's stock market decline and loss of tax revenue. The 20,000 restaurants and eateries in Los Angeles reflect this economic downtown in their percentage change in sales. Suppose the sales for all of these restaurants are normally distributed with an average percentage change in sales of -8% over the last year (they loss 8% of their sales on average) and a standard deviation of 9%. SHOW YOUR WORK FOR FULL CREDIT.

A. The percentage change in sales for your favorite restaurant was at the 95th percentile. What exactly was the value of its percentage change?

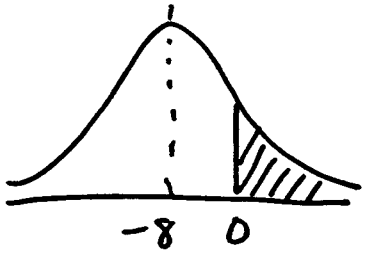


5% in upper "tail" implies a "middle" of 90% which is $z = \pm 1.65$

$$1.65 = \frac{X - (-8)}{9}$$

solve for $X = 6.85\%$

B. What percentage of restaurants experienced no decrease or an increase in sales over the last year?



$$z = \frac{0-(-8)}{9} = \frac{8}{9} = .889$$

~~[illegible]~~ table

A105 lookup 63.19% ("middle")

So $100 - 63.19 = 36.81$ ("tails")

and $\frac{36.81}{2} = \boxed{18.405\%}$

↑ for one "tail"