AP STATS
Midterm Review Chapter 1 and 2

Name Key
Date ____ Hour ____

Chapter 1 Review

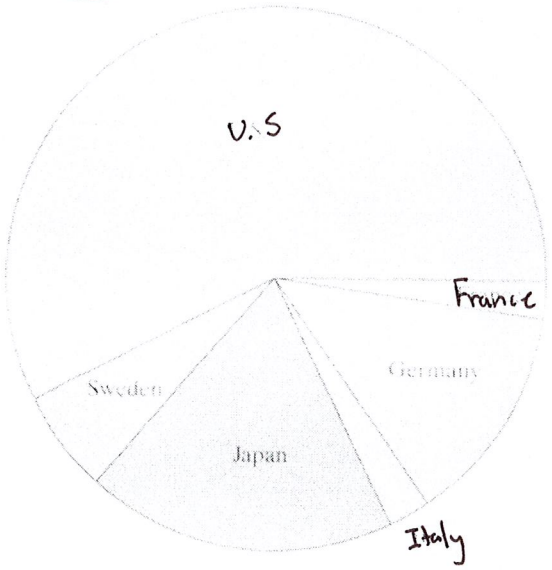
T1.1. You record the age, marital status, and earned income of a sample of 1463 women. The number and type of variables you have recorded is

- (a) 3 quantitative, 0 categorical
- (b) 4 quantitative, 0 categorical
- (c) 3 quantitative, 1 categorical
- (d) 2 quantitative, 1 categorical
- (e) 2 quantitative, 2 categorical

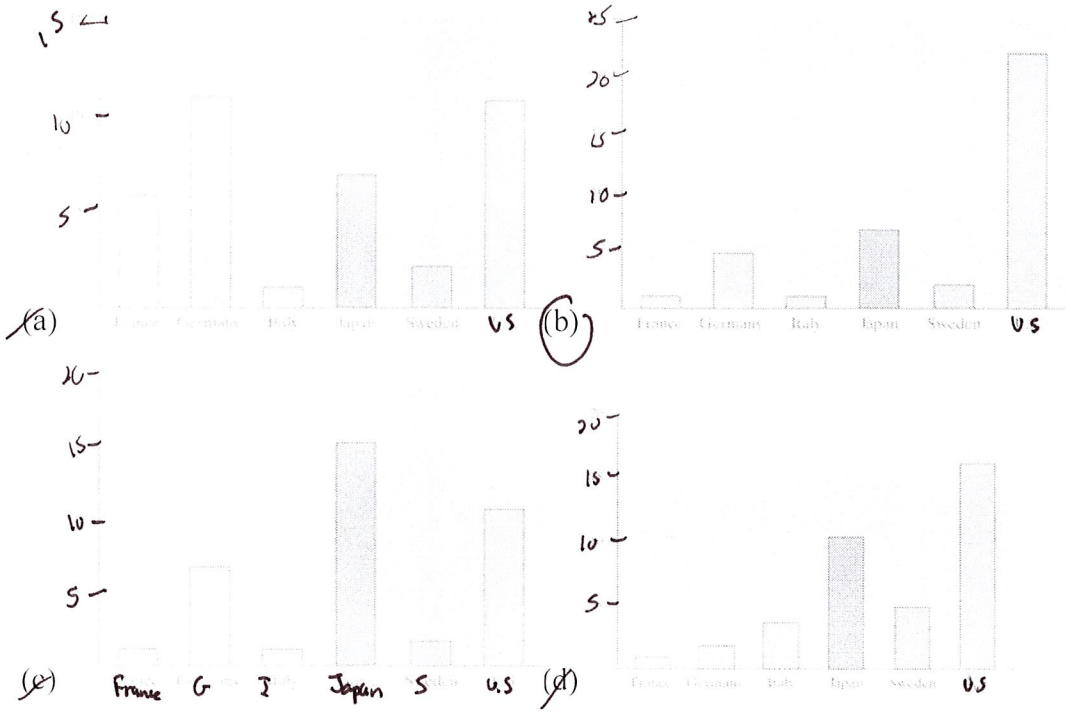
T1.2. Consumers Union measured the gas mileage in miles per gallon of 38 vehicles from the same model year on a special test track. The pie chart provides information about the country of manufacture of the model cars tested by Consumers Union. Based on the pie chart, we conclude

Pie Chart Does not tell us anything about gas mileage

- (a) Japanese cars get significantly lower gas mileage than cars from other countries.
- (b) U.S. cars get significantly higher gas mileage than cars from other countries.
- (c) Swedish cars get gas mileages that are between those of Japanese and U.S. cars.
- (d) Mercedes, Audi, Porsche, and BMW represent approximately a quarter of the cars tested.
- (e) More than half of the cars in the study were from the United States.



T1.3. Which of the following bar graphs is equivalent to the pie chart in Question T1.2?



(e) None of these.

T1.4. Earthquake intensities are measured using a device called a seismograph, which is designed to be most sensitive to earthquakes with intensities between 4.0 and 9.0 on the Richter scale. Measurements of nine earthquakes gave the following readings:

4.5 L 5.5 H 8.7 8.9 6.0 H 5.2

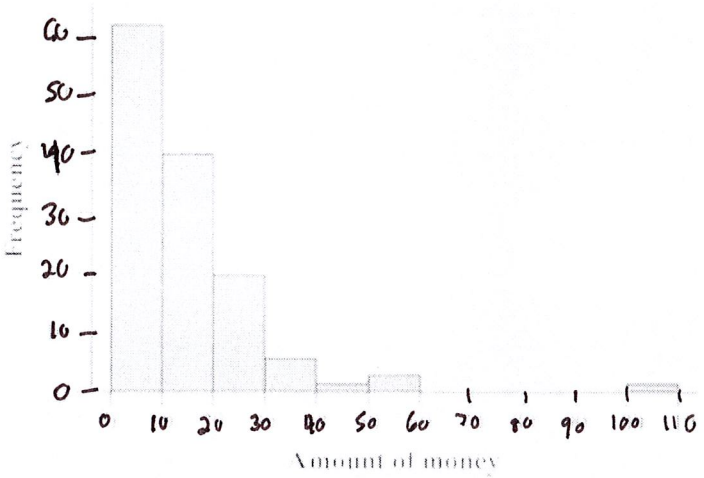
where L indicates that the earthquake had an intensity below 4.0 and an H indicates that the earthquake had an intensity above 9.0. The median earthquake intensity of the sample is

L 4.5 5.2 5.5 6.0 8.7 8.9 H H

n

- (a) 5.75.
- (b) 6.00.
- (c) 6.47.
- (d) 8.70.
- (e) Cannot be determined.

Questions T1.5 and T1.6 refer to the following setting. In a statistics class with 136 students, the professor records how much money (in dollars) each student has in his or her possession during the first class of the semester. The histogram shows the data that were collected.



T1.5. The percentage of students with less than \$10 in their possession is closest to

- (a) 30%.
- (b) 35%.
- (c) 50%.
- (d) 60%.
- (e) 70%.

about 62 out of 136

$62/136 \approx .456 \approx 46\%$

just under 1/2

T1.6. Which of the following statements about this distribution is *not* correct?

- (a) The histogram is right-skewed. correct
- (b) The median is less than \$20. yes
- (c) The *IQR* is \$35. incorrect
- (d) The mean is greater than the median. correct skewed right pulls mean to the right so it will be greater than the median
- (e) The histogram is unimodal. yes correct

Q_1 is between 0 and 10

Q_3 is between 20 and 30

So Largest IQR in is 30-0=30

T1.7. Forty students took a statistics examination having a maximum of 50 points. The score distribution is given in the following stem-and-leaf plot:

```
0 | 28
1 | 2245
2 | 01333358889
3 | 001356679
4 | 2244|4466788
5 | 000
```

40(.75) = 30
since even # of students
Q_3 falls between 30th + 31st observation

$Q_3 = 44$

The third quartile of the score distribution is equal to

(a) 45 (b) 44 (c) 43 (d) 32 (e) 23

T1.8. The mean salary of all female workers is \$35,000. The mean salary of all male workers is \$41,000. What must be true about the mean salary of all workers?

- (a) It must be \$38,000.
- (b) It must be larger than the median salary.
- (c) It could be any number between \$35,000 and \$41,000.
- (d) It must be larger than \$38,000.
- (e) It cannot be larger than \$40,000.

The mean salary of all workers will be somewhere between the mean salaries of the 2 groups seperately; where it will be between these 2 groups depends on how many workers are in each individual group

Questions T1.9 and T1.10 refer to the following setting. A survey was designed to study how business operations vary according to their size. Companies were classified as small, medium, or large. Questionnaires were sent to 200 randomly selected businesses of each size. Since not all questionnaires in a survey of this type are returned, researchers decided to investigate the relationship between the response rate and the size of the business. The data are given in the following two-way table:

Size	Response	No Response	Total
Small	125	75	200
Medium	81	119	200
Large	40	160	200

125/200 = .625

T1.9. What percent of all small companies receiving questionnaires responded?

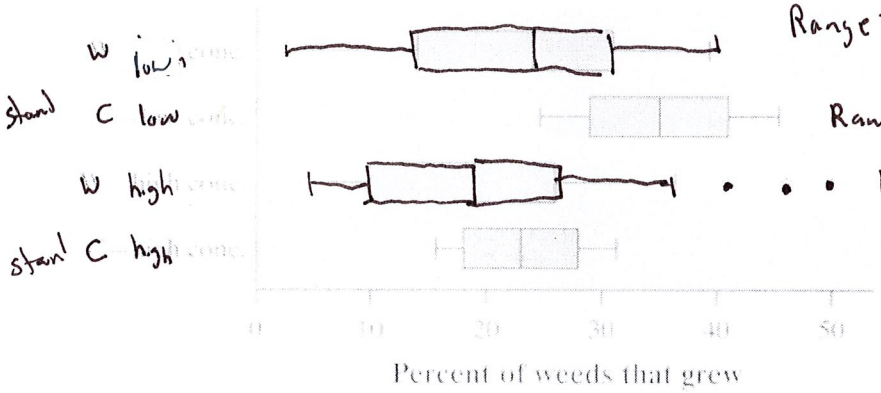
(a) 12.5% (b) 20.8% (c) 33.3% (d) 50.8% (e) 62.5%

T1.10. Which of the following conclusions seems to be supported by the data?

- (a) There are more small companies than large companies in the survey.
- (b) Small companies appear to have a higher response rate than medium or big companies.
- (c) Exactly the same number of companies responded as didn't respond.
- (d) Small companies dislike larger companies.
- (e) If we combined the medium and large companies, then their response rate would be equal to that of the small companies. 81+40=122

$$\frac{122}{200} \neq \frac{125}{200}$$

T1.11. An experiment was conducted to investigate the effect of a new weed killer to prevent weed growth in onion crops. Two chemicals were used: the standard weed killer (C) and the new chemical (W). Both chemicals were tested at high and low concentrations on a total of 50 test plots. The percent of weeds that grew in each plot was recorded. Here are some boxplots of the results. Which of the following is *not* a correct statement about the results of this experiment?



Range = 40-0 = 40 IQR = 31-12 = 19

Range = 45-25 = 20 IQR = 48-30 = 18

Range = 37-2 = 35 IQR = 25-10 = 15

Range = 31-15 = 16 IQR = 28-18 = 10

- (a) At both high and low concentrations, the new chemical (W) gives better weed control than the standard weed killer (C). true
- (b) Fewer weeds grew at higher concentrations of both chemicals. true
- (c) The results for the standard weed killer are less variable than those for the new chemical. true
- (d) High and low concentrations of either chemical have approximately the same effects on weed growth. High concentrations appear to have better weed control (fewer weeds growing) than lower concentrations
- (e) Some of the results for the low concentration of weed killer W show fewer weeds growing than some of the results for the high concentration of W. true

Section II: Free Response *Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.*

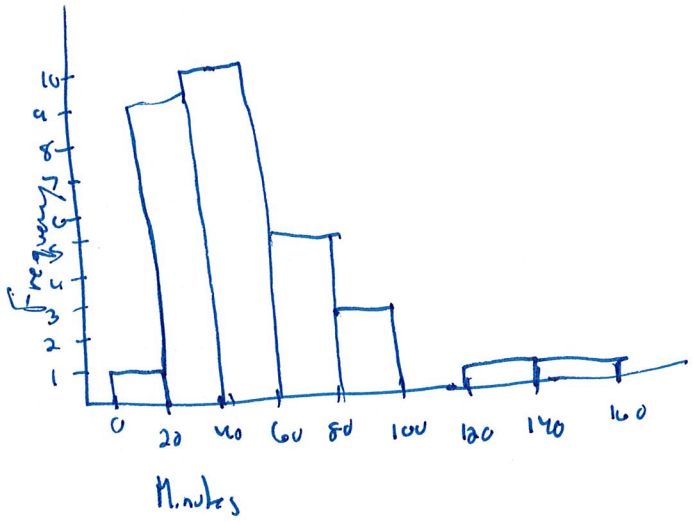
T1.12. You are interested in how much time students spend on the Internet each day. Here are data on the time spent on the Internet (in minutes) for a particular day reported by a random sample of 30 students at a large high school:

7 20 21 25 25 26 28 30 32 35
0 0 0 0 0 0 0 0 0 0
72 75 77 78 79 83 87 88 155 151

Q1

- (a) Construct a histogram of these data.

0 to <20 1
20 to <40 9
40 to <60 10
60 to <80 5
80 to <100 3
100 to <120 0
120 to <140 1
140 to <160 1



- (b) Are there any outliers? Justify your answer.

Formula for outlier
Lower cutoff $Q_1 - 1.5 \times IQR$
Upper cutoff $Q_3 + 1.5 \times IQR$

Q_1 = median of lower 15 values so 8th data value

Q_3 = median of upper 15 values so 23rd data value

$Q_1 = 30$
$Q_3 = 77$
$IQR = 77 - 30 = 47$

$30 - 1.5(47) = -40.5$ no values below -40.5
$77 + 1.5(47) = 147.5$ 1 value above

so 151 is an outlier

- (c) Would it be better to use the mean and standard deviation or the median and *IQR* to describe the center and spread of this distribution? Why?

Since there is an outlier skewing the distribution to the right it would be better to use the median and IQR to describe center + spread.

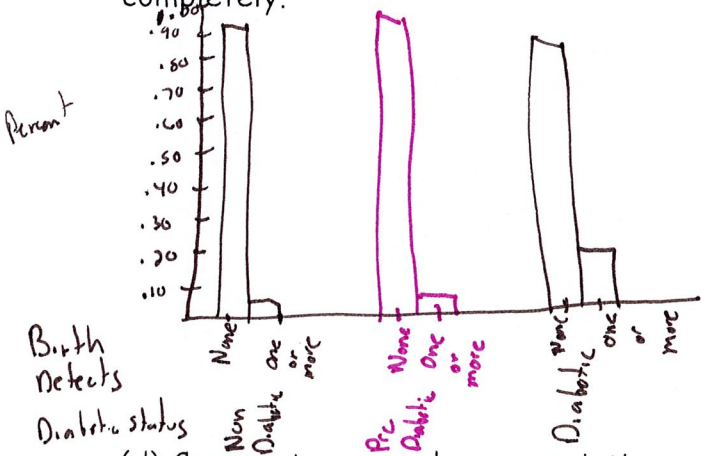
T1.13. A study among the Pima Indians of Arizona investigated the relationship between a mother's diabetic status and the appearance of birth defects in her children. The results appear in the two-way table below.

	Diabetic Status			
Birth Defects	Nondiabetic	Prediabetic	Diabetic	Total
None	754	362	38	1154
One or more	31	13	9	53
Total	785	375	47	1207

- (a) Fill in the row and column totals in the margins of the table.
- (b) Compute (in percents) the conditional distributions of birth defects for each diabetic status. (given)

	Diabetic Status		
Birth Defects	Non Diabetic	Pre Diabetic	Diabetic
None	$\frac{754}{785} = .961$	$\frac{362}{375} = .965$	$\frac{38}{47} = .809$
One or More	$\frac{31}{785} \approx .039$	$\frac{13}{375} = .035$	$\frac{9}{47} \approx .191$

- (c) Display the conditional distributions in a graph. Don't forget to label your graph completely.

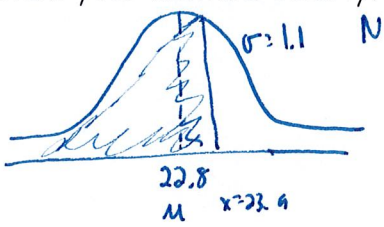


- (d) Comment on any clear associations you see.

Non Diabetics + Pre diabetics appear to have babies with birth defects at about the same rate. Those with diabetes have babies with birth defects at a higher rate (almost 16% higher)

12. The army reports that the distribution of head circumference among male soldiers is approximately Normal with mean 22.8 inches and standard deviation 1.1 inches.

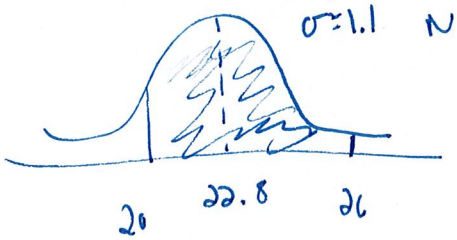
(a) A male soldier whose head circumference is 23.9 inches would be at what percentile? Show your method clearly.



normcdf(−10,000, 23.9, 22.8, 1.1) ≈ .84

≈ 84th percentile

(b) The army's helmet supplier regularly stocks helmets that fit male soldiers with head circumferences between 20 and 26 inches. Anyone with a head circumference outside that interval requires a customized helmet order. What percent of male soldiers require custom helmets? Show your work, including a well-labeled sketch of a Normal curve.



normcdf(20, 26, 22.8, 1.1) = .993

So 99% are within 20-26, which are regularly stocked

so 1 − .99 = .007

.7% of soldiers require custom helmets

(c) Find the interquartile range for the distribution of head circumference among male soldiers. Show your method clearly.

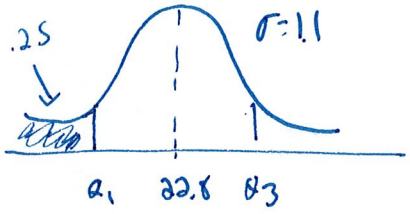
IQR = Q3 − Q1 N(22.8, 1.1)

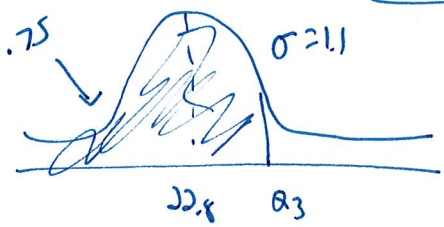
Q3 = 75th % ≈ 23.54

Q1 = 25th % = 22.06

IQR = 23.54 − 22.06

IQR ≈ 1.48 in





InvNorm(.25, 22.8, 1.1) ≈ 22.06

InvNorm(.75, 22.8, 1.1) ≈ 23.54

13. A study recorded the amount of oil recovered from the 64 wells in an oil field. Here are descriptive statistics for that set of data from Minitab.

Descriptive Statistics: Oilprod

Variable	N	Mean	Median	StDev	Min	Max	Q1	Q3
Oilprod	64	48.25	37.80	40.24	2.00	204.90	21.40	60.75

Does the amount of oil recovered from all wells in this field seem to follow a Normal distribution? Give appropriate statistical evidence to support your answer.

If data is normally distributed mean should be "pretty close to the median
In our case, the mean is larger than median. (mean is 48.25, and median 37.80)
This indicates our distribution is skewed right.
We also can see the skewed rightness when comparing distances between min and median and max and median. From min to median the distance is 37.80 - 2 = 35.80 From median to max the distance is 204.90 - 37.80 = 167.10
There is a greater spread in the upper half of distribution, indicating distribution is skewed right.

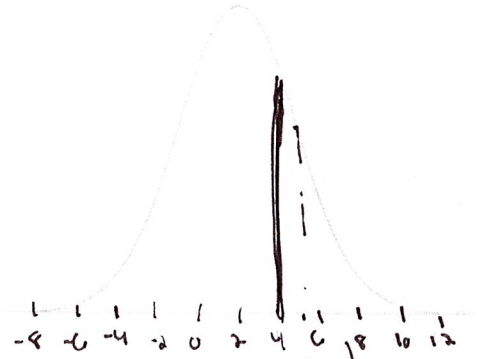
∴ not normally distributed

Chapter 2

1 Many professional schools require applicants to take a standardized test. Suppose that 1000 students take such a test. Several weeks after the test, Pete receives his score report: he got a 63, which placed him at the 73rd percentile. This means that

(a) Pete's score was below the median.
(b) Pete did worse than about 63% of the test takers.
(c) Pete did worse than about 73% of the test takers.
(d) Pete did better than about 63% of the test takers.
(e) Pete did better than about 73% of the test takers.

2. For the Normal distribution shown, the standard deviation is closest to



(a) 0 (b) 1 (c) 2 (d) 4 (e) 5

3. Rainwater was collected in water collectors at 30 different sites near an industrial complex, and the amount of acidity (pH level) was measured. The mean and standard deviation of the values are 4.60 and 1.10, respectively. When the pH meter was recalibrated back at the laboratory, it was found to be in error. The error can be corrected by adding 0.1 pH units to all of the values and then multiplying the result by 1.2. The mean and standard deviation of the corrected pH measurements are

$\mu = 4.60 \quad \sigma = 1.10$

(a) 5.64, 1.44 (b) 5.64, 1.32 (c) 5.40, 1.44 (d) 5.40, 1.32 (e) 5.64, 1.20

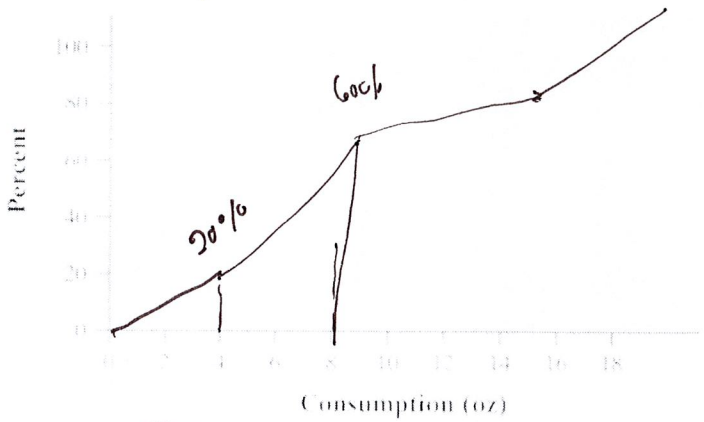
$\mu_{corrected} = (4.60 + 0.1) \cdot 1.2 = 5.64$

$\sigma_{corrected} = 1.10 \cdot 1.2 = 1.32$

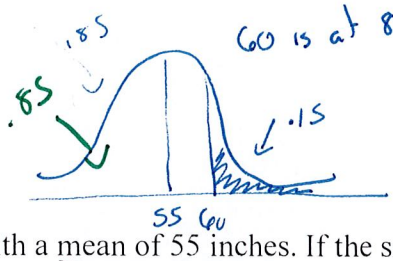
addition + multiplication affect mean

multiplication only affects standard deviation

4. The figure shows a cumulative relative frequency graph of the number of ounces of alcohol consumed per week in a sample of 150 adults. About what percent of these adults consume between 4 and 8 ounces per week?



(a) 20% (b) 40% (c) 50% (d) 60% (e) 80%



60 is at 85th o/o file

5. The average yearly snowfall in Chillyville is Normally distributed with a mean of 55 inches. If the snowfall in Chillyville exceeds 60 inches in 15% of the years, what is the standard deviation?

so 60 has a z-score of 1.04

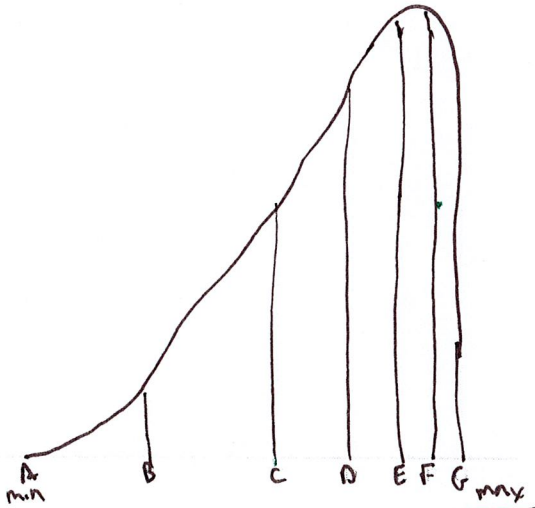
(a) 4.83 inches (b) 5.18 inches (c) 6.04 inches cannot be computed from the given information. (d) 8.93 inches (e) The standard deviation

z score = $\frac{\text{value} - \text{mean}}{\sigma}$: $1.04 = \frac{60-55}{\sigma}$: $1.04 = \frac{5}{\sigma}$ $\sigma = \frac{5}{1.04}$

Inv norm(.85) = 1.04

6. The figure shown is the density curve of a distribution. Five of the seven points marked on the density curve make up the five-number summary for this distribution. Which two points are *not* part of the five-number summary?

$\sigma = 4.8$ in



A = min

G = max

min
Q_1
med
Q_3
max

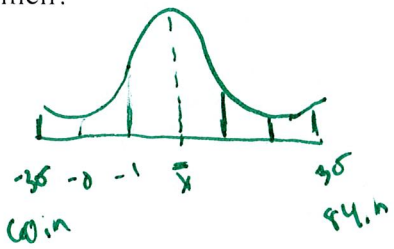
- not enough area between A and B to constitute 25 o/o
- not enough area between F and G to constitute 25 o/o
B and F are too close to the ends of distribution

(a) B and E (b) C and F (c) C and E (d) B and F ~~(e) A and G~~

60 in 84 in

7. If the heights of American men follow a Normal distribution, and 99.7% have heights between 5′0″ and 7′0″, what is your estimate of the standard deviation of the height of American men?

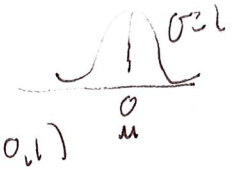
(a) 1″ (b) 3″ (c) 4″ (d) 6″ (e) 12″



8. Which of the following is *not* correct about a standard Normal distribution?

(a) The proportion of scores that satisfy $0 < z < 1.5$ is 0.4332. ✓ true normcdf(0, 1.5, 0, 1)
(b) The proportion of scores that satisfy $z < -1.0$ is 0.1587. ✓ true normcdf(-10,000, -1)
(c) The proportion of scores that satisfy $z > 2.0$ is 0.0228. ✓ true normcdf(2, 10000)
(d) The proportion of scores that satisfy $z < 1.5$ is 0.9332. ✓ true normcdf(-10,000, 1.5)
(e) The proportion of scores that satisfy $z > -3.0$ is 0.9938. ✗ false normcdf(-3, 10,000) = .9987

$\mu = 0$, $\sigma = 1$



Questions 9 and 10 refer to the following setting. Until the scale was changed in 1995, SAT scores were based on a scale set many years ago. For Math scores, the mean under the old scale in the 1990s was 470 and the standard deviation was 110. In 2009, the mean was 515 and the standard deviation was 116.

9. What is the standardized score (z-score) for a student who scored 500 on the old SAT scale?

(a) −30 (b) −0.27 (c) −0.13 (d) 0.13 (e) 0.27

$z = \frac{\text{score} - \text{mean}}{\sigma} = \frac{500 - 470}{110} = \frac{30}{110} = .27$

10. Jane took the SAT in 1994 and scored 500. Her sister Colleen took the SAT in 2009 and scored 530. Who did better on the exam, and how can you tell?

(a) Colleen—she scored 30 points higher than Jane.
(b) Colleen—her standardized score is higher than Jane's.
(c) Jane—her standardized score is higher than Colleen's.
(d) Jane—the standard deviation was bigger in 2009.
(e) The two sisters did equally well—their z-scores are the same.

Jane's z-score $= \frac{500 - 470}{110} = .27$

Colleen $= \frac{530 - 515}{116} = \frac{15}{116} = .13$

Section II: Free Response

11. As part of the President's Challenge, students can attempt to earn the Presidential Physical Fitness Award or the National Physical Fitness Award by meeting qualifying standards in five events: curl-ups, shuttle run, sit and reach, one-mile run, and pull-ups. The qualifying standards are based on the 1985 School Population Fitness Survey. For the Presidential award, the standard for each event is the 85th percentile of the results for a specific age group and gender among students who participated in the 1985 survey. For the National award, the standard is the 50th percentile. To win either award, a student must meet the qualifying standard for all five events.

Jane, who is 9 years old, did 40 curl-ups in one minute. Matt, who is 12 years old, also did 40 curl-ups in one minute. The qualifying standard for the Presidential award is 39 curl-ups for Jane and 50 curl-ups for Matt. For the National award, the standards are 30 and 40, respectively.

Jane Matt
Girls = 39 curlups is 85th %tile Boys = 50 curlups is 85th %tile
Girls = 30 curlups is 50th %tile Boys 40 curlups is 50th %tile

(a) Compare Jane's and Matt's performances using percentiles. Explain in language simple enough for someone who knows little statistics to understand.

Jane did 40 curlups. Matt did 40 curl ups.

Jane performed better. She did more curlups than 85% of the girls her age. This means she qualified for both the Presidents award and the National award.

Matt did more curlups than 50% of the boys his age but less than 85% of the boys his age. He only qualified for the national award.

So Jane only had less than 15% do better than her and Matt had less than 50% do better than him.

(b) Who has the higher standardized value (*z*-score), Jane or Matt? Justify your answer.

Since Jane's score is at a higher %tile than Matt's, her standardized score will also be higher than Matt's.