

(13) 1. A study was conducted to see whether supplementation with creatine improved soccer skills in young soccer players. Twenty male soccer players (15 – 19 years old) participated in the study. They were randomly assigned to 2 groups of 10 each. Group 1 took a creatine-monohydrate supplement, and Group 2 took a placebo, each for 7 days. Before and after the supplementation protocol, each subject underwent a series of soccer skill tests: dribble test, sprint-power test, endurance test, and vertical jump test. The researchers found that the creatine group improved more in all soccer skills, significantly more than the placebo group (in which there was not a significant improvement).

2 (a) What is the research question? Does supplementation with creatine improve athletic performance (specifically soccer skills).

2 (b) Describe the sample: 20 male soccer players, ages 15-19

2 (c) What is the (implied) Population? Young male soccer players

2 (d) What is the Treatment variable? Creatine supplementation
 The treatment variable of interest is (circle one) CATEGORICAL NUMERICAL

2 (e) What is the Response variable? Soccer skills
 The treatment variable of interest is (circle one) CATEGORICAL NUMERICAL *ok*

Poorly worded

This is ambiguous - depends on what was measured in "skills"

3 (f) Can we say that taking creatine supplements CAUSED the soccer players to improve their skills? Why or why not?
Yes, since this was a controlled experiment, we can conclude cause-and-effect.

(8) 2. A study was done examining the use of vitamin C for breast-feeding mothers. The children whose mothers had chosen to take high doses of vitamin C had a 30% lower risk of developing allergies.

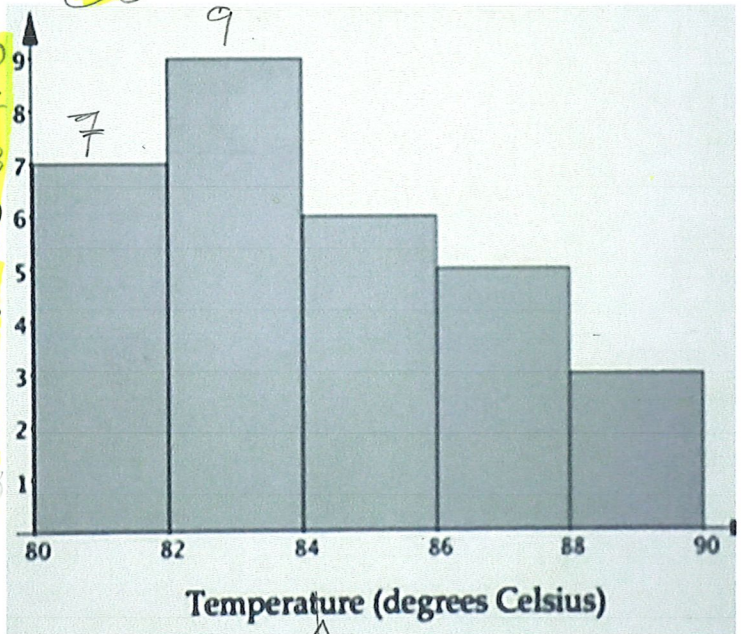
(4) Does this study tell us that taking vitamin C (for breast-feeding mothers) causes a reduction in allergies in their children? Why or why not?
No, we can't conclude there is a cause-and-effect relationship because this is an observational study.

(4) What is one possible confounder that links mothers who choose to take high doses of vitamin C to children with fewer allergies? Answers will vary!
Mothers who choose to supplement with Vitamin C may be more health-conscious in general. For instance, smoking could be a confounder in that fewer health-conscious mothers would smoke, thus showing lower allergy rates in their children.

3. (12 pts) The daily high temperatures (in degrees Fahrenheit) were recorded in SLO over 30 days in the summer. The histogram shows the distribution of temperatures over those days.

Use the histogram to answer the following questions.

relative frequency



(2) (a) What is the shape of the distribution?

skewed right

(2) (b) Which would be greater for this distribution, the mean or the median (you do NOT have to find either the mean or median!)

The mean is greater

(2) (c) Which would it be more appropriate to describe the center and variation of this data set: (circle one)

the mean and standard deviation

OR

the median and IQR

(3) (d) How many days had a high that was less than 84 degrees?

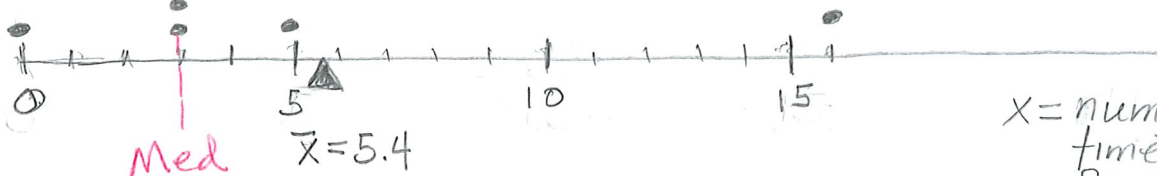
16 days

(3) (e) Relabel the y-axis with relative frequencies.

(out of 30) $1 \rightarrow \frac{1}{30} = .033$
 $2 \rightarrow \frac{2}{30} = .067$ etc.

(14) 4. Suppose you select a sample of five students and ask them how many times they check social media in a typical day. Their answers are 0, 3, 3, 5, 16

(1) a) Construct a dotplot for this data



(2) b) Find the sample mean for the data. Use the proper notation for the mean! Plot it on the dotplot.

$$\bar{x} = \frac{\sum x}{n} = \frac{0+3+3+5+16}{5} = \frac{27}{5} = 5.4$$

(2) c) Find the sample median for the data. Plot it on the dotplot.

Median = 3

4. Continued...

(1) d) Which statistic is a more "typical" value for this data: the mean the median

(1) e) Which value appears to be an outlier?

16 checks

(2) f) What effect did the outlier have on the mean (in terms of the mean being in the "center" of the data)?

The outlier pulled the mean up to the right and away from the more typical value of 3 checks.

(1) g) Because of this effect, we say that the mean is not "resistant"

(4) h) Find the sample standard deviation by hand. Show work! $\bar{x} = 5.4, n = 5$

x	$x - \bar{x}$	$(x - \bar{x})^2$
0	-5.4	29.16
3	-2.4	5.76
3	-2.4	5.76
5	-.4	.16
16	10.6	112.36
$\sum x - \bar{x} = 0$ check!		$\sum (x - \bar{x})^2 = \text{153.2} = 153.2$

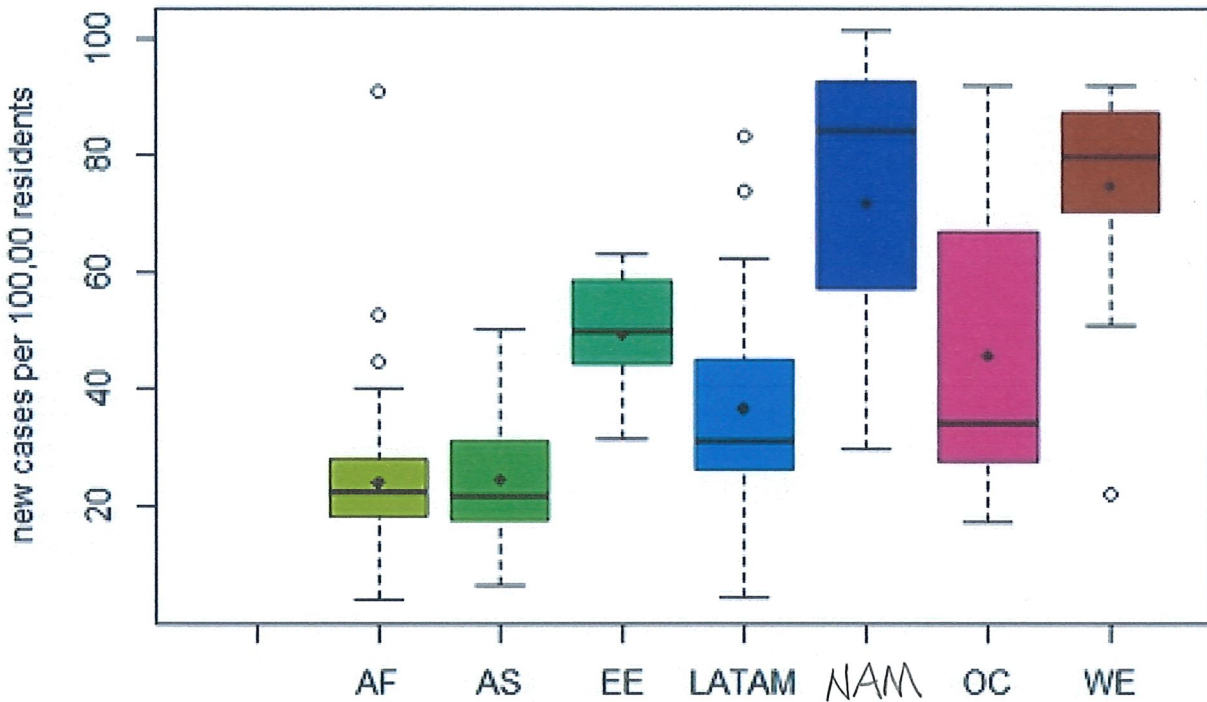
$$\text{Sample S.D.} = S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{\frac{153.2}{4}} = \sqrt{38.3} = 6.189$$

$$S = 6.189$$

10
 5. (10 pts) The boxplots below show the distribution of breast cancer rates by region/continent

AF = Africa EE = Eastern Europe AS = Asia
 WE = Western Europe NAM = North America OC = Oceania (Australia, South Pacific, etc.)
 LATAM = Latin America (Mexico, Central and South America)

Breast cancer by continent (mean is black dot)



- (2) Which regions show the most skewing? *NAM (North America) OC (Oceania)* - ambiguous due to outliers - explain about large data sets
- (2) Which regions show the least variation? *AF (Africa has the least variation) (Next are AS (Asia) and EE (Eastern Europe))*
- (2) Which regions have outliers? *Africa, Latin America, Western Europe*
- (2) What would these outliers represent (describe in words)? *Nations in each group that have unusually high (AF, LATAM) or unusually low (WE) breast cancer rates.*
- (2) The chart seems to indicate that less developed parts of the world have healthier women (lower rates of breast cancer). What is a possible confounder for this association?
Lifespan. Less developed countries have shorter lifespans on average, so there are fewer older women hence less breast cancer (lower rates). Discuss "rates"

XC

7. The Five Number Summary for annual rainfall (in inches) for 62 years in San Luis Obispo is as follows:
(14)

10.35 15.52 19.73 27.36 54.53
min Q_1 Med Q_3 max

(2) (a) Which number is the third quartile?

$$Q_3 = 27.36 \text{ inches of rain}$$

(2) (b) Find the IQR.

$$IQR = Q_3 - Q_1 = 27.36 - 15.52 = \boxed{11.84} \text{ inches}$$

(4) (c) Find the Lower Outlier and Upper Outlier Limits. Show work!

$$\begin{aligned} \text{Lower limit} &= Q_1 - 1.5(IQR) = 15.52 - 1.5(11.84) \\ &= -2.24 \text{ inches} \end{aligned}$$

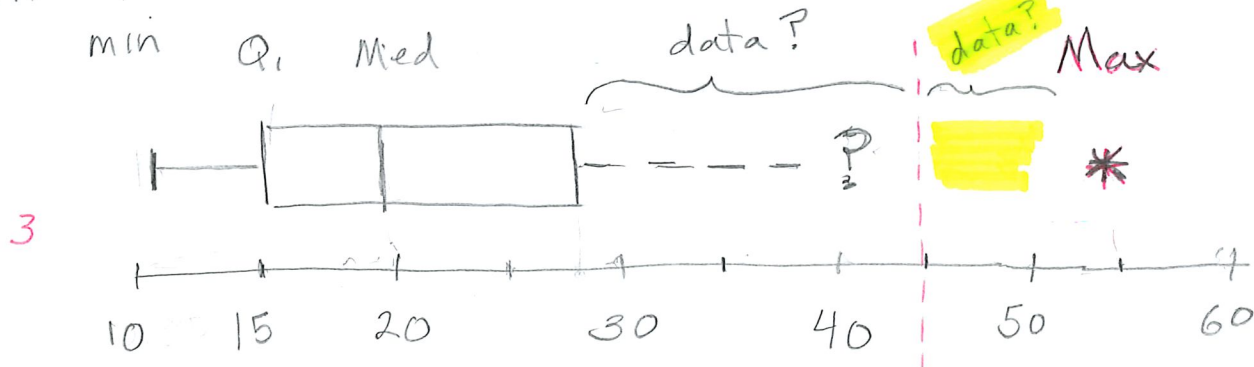
$$\begin{aligned} \text{Upper limit} &= Q_3 + 1.5(IQR) = 27.36 + 1.5(11.84) \\ &= 45.12 \end{aligned}$$

(2) (d) Is either the minimum value or the maximum value from the Five Number Summary an outlier?

Min value = 10.35 which is within the lower limit, so not an outlier.

Max value = 54.53 which is outside the upper limit, so it is an outlier

(4) (e) Can we tell if there is more than one outlier? Briefly explain and include a boxplot sketch in your answer.

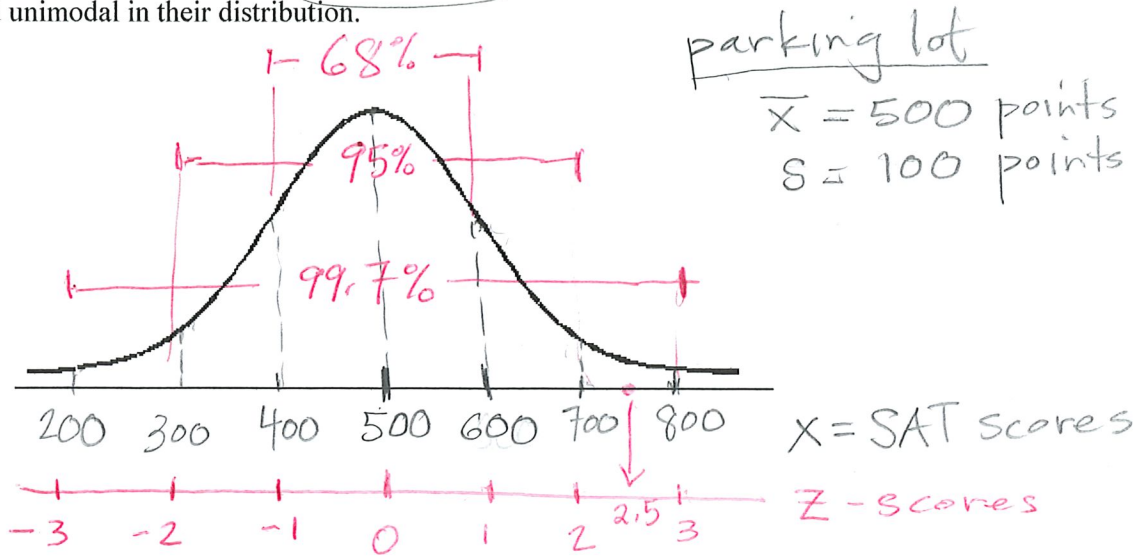


No.

Upper Outlier Limit

1 Since we don't have all the data values, we don't know whether there are values beyond the Upper Limit in addition to the max value.

6. (9 pts) Quantitative SAT scores have a mean of 500 and a standard deviation of 100 and the scores are symmetric and unimodal in their distribution.



(2) (a) Fill in the curve with the mean and show how the standard deviation fits into the picture.

See graph

(2) (b) Between what two values should about 68% of the SAT fall?

About 68% of the scores should fall between 400 and 600 points

(2) (c) Find the z-score for a student who scores a 750 on the SAT. $X = 750$

$$Z = \frac{X - \bar{X}}{S} = \frac{750 - 500}{100} = 2.5$$

$$Z = 2.5$$

You can also see this z-score on the graph above as shown.

(3) (d) ACT scores have a mean of 21 and a standard deviation of 5. Assuming ACT scores have a distribution that is also unimodal and symmetric, which is more unusual: A quantitative SAT score of 750 or an ACT score of 35? (Hint: find the z-score for the ACT score and compare!)

$$\text{SAT} = 750 \text{ points} \Rightarrow Z = 2.5$$

$$\text{ACT} = 35 \text{ points} \Rightarrow Z = 2.8$$

The ACT is more unusual (larger z-score).

ACT Parking Lot

$$\bar{X} = 21 \text{ points}$$

$$S = 5 \text{ points}$$

$$X = 35$$

$$Z = \frac{35 - 21}{5} = 2.8$$