*In-class test _____/80 points*          *Take home test _____/20 points*

1. A study was conducted to see whether <u>supplementation with creatine</u> improved <u>soccer skills</u> 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 in all soccer skills, significantly more than the placebo group (in which there was not a significant improvement).

    (a)  What is the research question?

    (b) Describe the sample:

    (c) What is the (implied) Population?

    (d) What is the Treatment variable?

           The treatment variable of interest is (circle one)     CATEGORICAL        NUMERICAL

    (e)  What is the Response variable?

           The treatment variable of interest is (circle one)     CATEGORICAL        NUMERICAL

    (f) Can we say that taking creatine supplements CAUSED the soccer players to improve their skills? Why or why not?

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.

    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?

    What is one possible confounder that links mothers who choose to take high doses of vitamin C to children with fewer allergies?

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.

(a) What is the shape of the distribution?

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

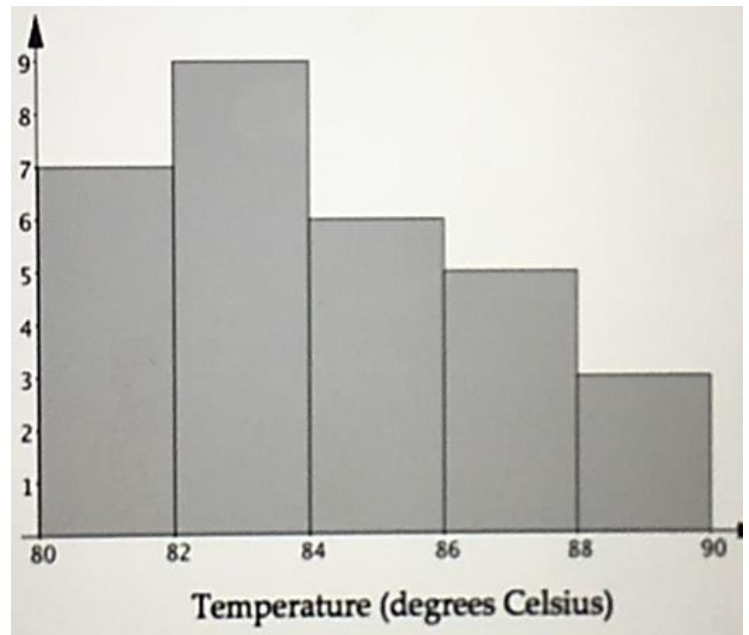(c) Which would it be more appropriate to describe the center and variation of this data set: (circle one)

    the mean and standard deviation

        OR

    the median and IQR

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

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



Temperature (degrees Celsius)

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
   a) Construct a dotplot for this data

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

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

4. Continued…

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

    e) Which value appears to be an outlier?

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

    g) Because of this effect, we say that the mean is not _____

    h) Find the sample standard deviation by hand.  Show work!

5. (3 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)
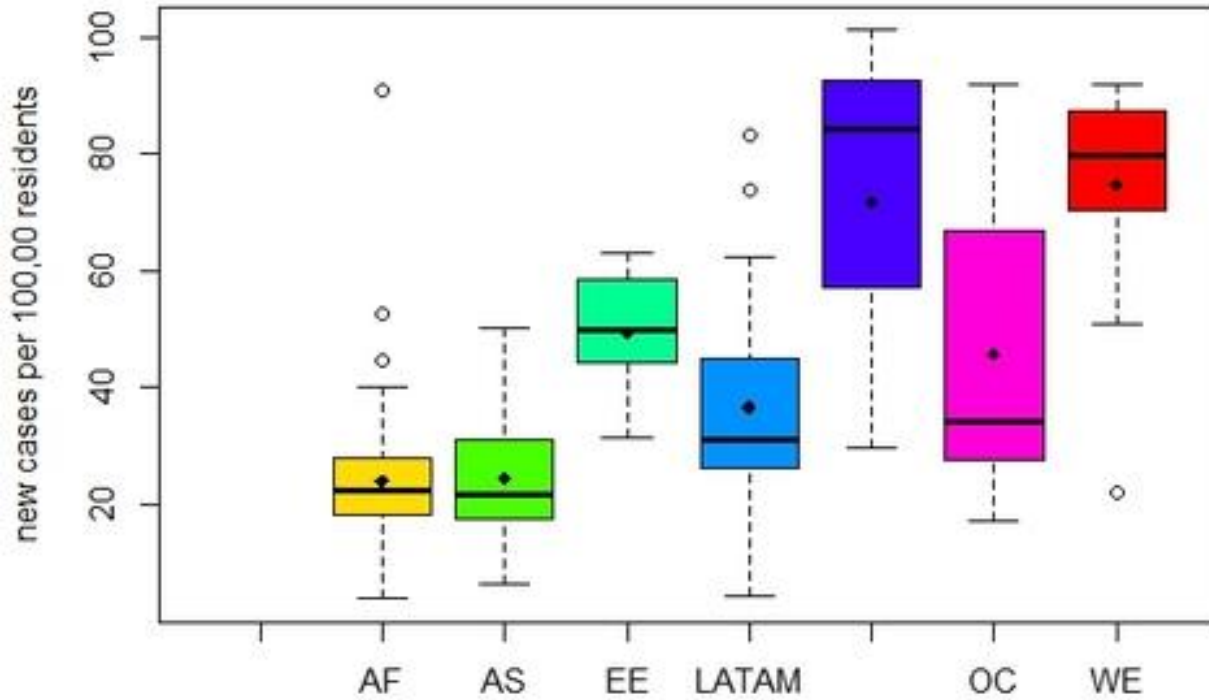
Which regions show the most skewing?

Which regions show the least variation?

Which regions have outliers?

What would these outliers represent (describe in words)?

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?

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

   10.35    15.52    19.73    27.36    54.53

(a) Which number is the third quartile?

(b) Find the IQR.

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

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

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

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.
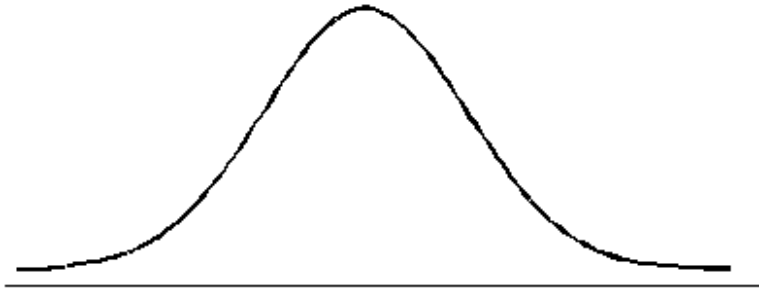


(a) Fill in the curve with the mean and show how the standard deviation fits into the picture.

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

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

(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!)

**Math 247: Test 1  (Take Home)**                     Name:_____
(20 points)

In-Class Test_____/80                 Take-Home Test_____/20

- *This exam is due <u>at the beginning of class</u> on Thursday, 9/6/18. You may work with other people in the class but not with tutors, other instructors, etc.  <u>Be sure that all answers are written in your own words;</u> i.e., do not write verbatim the same answer as another student.*
- *Your work can be typed or NEATLY handwritten.*
- *Your work should incorporate all of the technology work mentioned below; i.e., **<u>copy and paste the StatCrunch work into your write-up.</u>***

*Scoring will be based on organization of your work, accuracy, and thoughtful, well-written answers.*
**Mention how to put the data into one column ONLY!!!**
*Answer all of the following questions <u>on another piece of paper</u> and <u>attach to this cover sheet</u>.  Use complete sentences in your answers!*

Twenty-five Cuesta students were asked how many hours they usually sleep on a weeknight.  The table shows their responses.

| Sleep (hours) | | | | | |
|---|---|---|---|---|---|
| 6 | 8 | 8 | 8 | 6 | 8 |
| 9 | 8 | 6 | 6 | 9 | 9 |
| 8 | 9 | 5 | 8 | 11 | 14 |
| 8 | 7 | 7 | 8 | 10 | 8 |
| 8 | | | | | |

| Sleep (hours) |
|---|
| 6 |
| 9 |
| 8 |
| 8 |
| 9 |
| 11 |
| 8 |
| 8 |
| 9 |
| 7 |
| 10 |
| 8 |
| 8 |
| 6 |
| 5 |
| 7 |
| 8 |
| 9 |
| 8 |
| 6 |
| 8 |
| 8 |
| 12 |
| 8 |
| 6 |

1.) Use StatCrunch to graph a dotplot of the data.

   Describe the distribution, including shape, spread, potential outliers, and "typical value".

2.) Use StatCrunch to find the Summary Statistics" for the data.

- Explain whether or not it be appropriate to apply the Empirical Rule to this data set. Write a sentence or two (complete sentences, please!) to explain and reference the dotplot in your explanation.

- By hand, showing work, find the range of sleep times that are 1 standard deviation from the mean, 2 standard deviations from the mean, 3 standard deviations from the mean:

  Fill these values by hand on your dotplot.

- Determine how many actual data values (using the given sleep data above) fall in each of the ranges above, then compute what percent of the total (total of 25 data values) is in each range:  Report these numbers and percentages as shown below:

  Actual data within 1 standard deviation of the mean: _____/25 = _____ %
  Actual data within 2 standard deviations of the mean: _____/25 = _____ %
  Actual data within 3 standard deviations of the mean: _____/25 = _____ %

- Relate the actual data percentages to those predicted by the Empirical Rule?  Include the predicted values from the Empirical Rule as well as the percentages you found in part (e).

3.) By hand, find the outlier limits (show work!) and explain what this tells you about any potential outliers.

- Use StatCrunch to graph a boxplot and explain whether or not it confirms your outlier analysis.  The "explanation" can be written directly on the boxplot graph!

- By hand, Find the z-score for the outlier and interpret it in the context of the problem.  In your explanation, connect the z-score to what it means to be an outlier.