

Math 160, REVIEW SHEET, Exam 1
Chapter 1, 2, 3
Vanden Eynden

1. Determine whether the given value is a statistic or a parameter.
 - a) After inspecting all 13,000 kg of meat stored at the Wurst Sausage Co, it was found that 63% of the meat was spoiled.
 - b) A sample of 120 employees is selected, and the average age is found to be 37 years.
2. Determine whether the value is from a discrete or continuous data set
 - a) The median men's shoe size is 9.5.
 - b) The average weight of an apple is 5.1 oz.

3. Determine which of the four levels of measurement is most appropriate.

1. Nominal 2. Ordinal 3. Interval 4. Ratio

- a) Tips earned by cocktail waitresses
- b) Survey response of "good", "better", "best"
- c) Nationalities of survey respondents
- d) Temperatures of the ocean at various depths

4. Suppose someone needs to randomly select a sample of boxes of Red Vines licorice. Each day, thousands of boxes are made at a processing plant. After they come off the production line, they are placed in large shipping boxes that fit 100 boxes of licorice. They are then stored in the warehouse for several days before being shipped out. For each of the following sampling techniques, identify the type of sampling that was used. Just put the appropriate number of the sampling technique in the blank.

1. Convenience Sampling 2. Stratified Random Sampling
3. Cluster Sampling 4. Systematic Random Sampling

_____ a) The sampler takes the top 3 shipping boxes off the shelf in the warehouse and uses all of the licorice inside them as his sample.

_____ b) The large shipping boxes are numbered 01 – 44. Boxes 07, 22, and 39 are selected. Every box of Red Vines in each of the selected shipping boxes is used as part of the sample.

_____ c) The boxes produced on machine 1 are thought to be different from the other machines. A portion of sample is randomly selected from the 400 boxes produced on machine 1. The rest of the sample is chosen randomly from the other machines.

_____ d) The sampler takes every 25th box of Red Vines as he goes through the large shipping boxes.

5. 88 smokers participate in an experiment in which 2 treatments are used: a new antidepressant drug and a placebo. Because it's thought that the success in quitting may depend on how long someone has been a smoker, the 88 participants are first divided into "less than 10 years smoking" and "more than 10 years smoking". Then each smoker is randomly assigned to either the drug or the placebo. What type of experiment is this?

6. A study conducted in Fall 2005 compared colleges that tested for drugs with colleges that did not, finding higher levels of students using drugs in those that were testing. Is this an observational study or an experiment? What type of observational study or experiment is it?

7. Consider the given frequency distribution below showing home sale prices in the city of Summerhill for the month of June.

Sale Price (in thousands of dollars)	Frequency (No. of homes sold)
80 – 110.9	2
111 – 141.9	5
142 – 172.9	7
173 – 203.9	10
204 – 234.9	3
235 – 265.9	1

- A. What is the class width?
- B. What are the class midpoints? List in increasing order.
- C. What are the class boundaries? List in increasing order.
- D. Find the mean sale price. Is this number the actual mean or an approximation?

8. A. A sample of 14 middle school girls was taken, and each was asked to estimate the length of the longest phone call (in minutes) they had made in the last week. Here are the results of the sample:
7 11 13 15 18 23 42 46 48 48 51 59 67 73

Calculate the mean and median of the sample of call lengths. MEAN _____

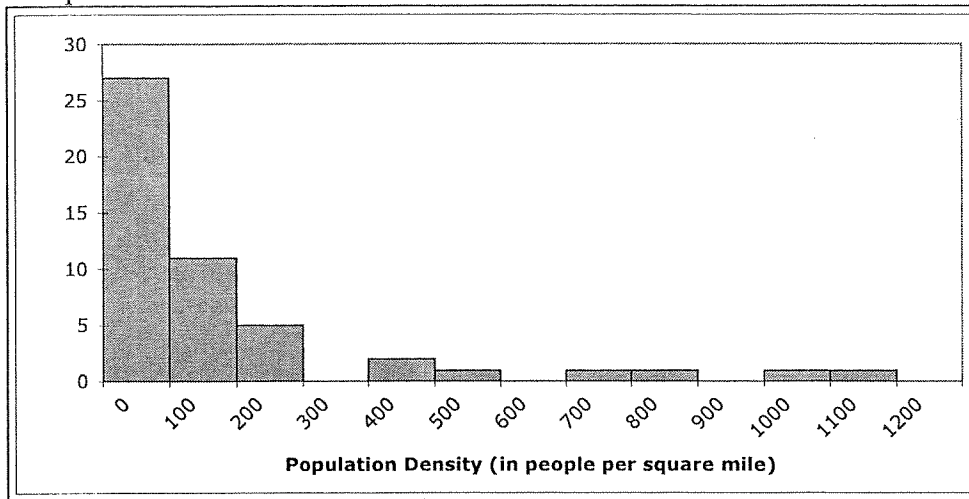
MEDIAN _____

B. What type of plot would we use to *compare* these 14 girls' responses to a sample of boys' responses?

C. Suppose a mistake was made in recording the two lowest values. It is unclear as to which one was recorded incorrectly. What we do know is that the incorrect one should have been 10 minutes higher. If the mean and median were re-calculated after the error was corrected, what could you expect of the two numbers?

- a) The new mean would be unaffected because it is a measure of exact center
- b) The new median would be unaffected because it is a measure of exact center

9. Pictured below is the distribution of the population densities by state from the 2000 U.S. census. They looked at the number of people per square mile for each of the 50 states. The lowest happened to belong to Alaska (1 person per square mile) and the highest was New Jersey (1134 people per square mile). Use the graph to answer the questions below.



- A. *Approximately what percent* of the states had population densities below 200 per square mile?
- B. *Approximately how many* states had population densities over 300 people per square mile?
- C. In which interval would the median fall?
 - a) 0-100 b) 100-200 c) 200-300 d) 300-400 e) 25-26 f) 700-1200
- D. Based on the shape of the distribution, which of the following can you conclude?
 - a) The mean population density would probably be *lower* than the median population density.
 - b) The mean population density would probably be *higher* than the median population density.
 - c) The mean and the median population densities should be the approximately the same number.
 - d) The mode would be the most appropriate measure of center whenever you're dealing with numbers.
 - e) The standard deviation would be unaffected by New Jersey's population density.
- F. Is this distribution NORMAL, UNIFORM, SKEWED LEFT, or SKEWED RIGHT?
- G. What is the class width?

10. Consider the following data set giving the ages of 35 members of a track and field team.

15	20	22	24	27	30	34
16	20	22	24	27	31	35
18	20	23	25	28	31	39
18	21	23	25	29	31	42
19	21	24	26	29	33	48

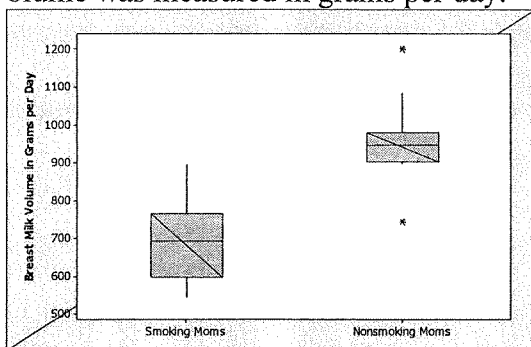
- A. Find the values of the 5-number summary.
- B. Construct a boxplot for the given data.
- C. Based on the 1.5 X IQR criterion, is the 48 year old athlete an outlier?

11. The following stemplots below compares the widths (in millimeters) of male Egyptian skulls unearthed by archaeologists. The skull widths on the left are skulls from approximately 4000 BC, and the skulls on the right were from approximately 150 AD. Change in skull width over time suggests interbreeding. Use the stemplots to answer the questions below.

4000BC		150AD	
11		11	
11	9	11	
12		12	
12	5 6 6 8 8 9	12	6 6 9
13	1 1 1 2	13	0 1 3 4
13	8	13	6 7 8 9
14		14	1
14		14	

- Calculate the means for both sets of skulls.
- Calculate the standard deviations for both sets of skulls.
- For which set of data was the mode 131mm?
- An archaeologist finds another skull that dates around 150 AD with a width of 142 mm. Would the skull's width be considered unusual?

12. Researchers compared the total daily volume of breast milk output by smoking mothers and non-smoking mothers. The breast milk volume was measured in grams per day.

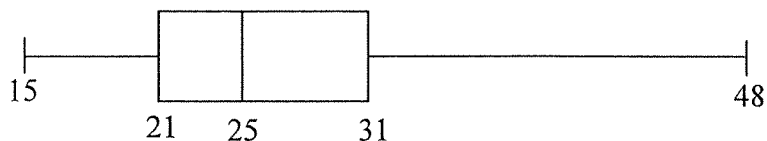


- Which of the following could we conclude about the 2 groups of mothers?
 - The *average* breast milk volume for nonsmokers would be about the same as that of smokers
 - With respect to the middle half of the values, nonsmokers' milk volumes were more variable than smokers.
 - If you ignore the outliers, and focus on the top and bottom 25% of the milk volumes, the non-smokers' volumes appear to be skewed left.
 - The highest milk volume for the smokers was lower than the median volume for nonsmokers
 - The nonsmokers' milk volumes appear to be symmetric
- Which of the following is **NOT** true?
 - Smoking mothers' volumes would have a larger IQR than that of non-smoking mothers
 - For smoking mothers, there is more variability in the top 25% than there is in the bottom 25%
 - The median would be the best measure of center for the non-smoker data
 - The standard deviation would be the best measure of spread since we're dealing with samples

13. The average caloric content of a Starbucks Venti Caramel Macchiato (with regular milk) is 380, and the standard deviation of 12 calories. For the Venti Caffe Mocha, the average is 490 calories, and the standard deviation is 14 calories.
- Calculate the z-scores for a 390 calorie Macchiato and a 500 calorie Caffe Mocha.
 - Which would be more unusual—the 390 calorie Macchiato or the 500 calorie Mocha? Or would they be equally unlikely because both are 10 calories above their respective means?
14. **True or False.** If all the values in a data set are converted to z-scores, the shape of the distribution of the z-scores will be the same as the distribution of the original data.
15. In statistics class, what do these symbols stand for?
- \bar{x}
 - σ
 - s
 - μ

Answer Key:

- parameter, statistic
- discrete, continuous
- 4, 2, 1, 3
- 1, 3, 2, 4
- Randomized Block
- Observational Study, cross-sectional study
- A. 31
B. 95.45, 126.45, 157.45, 188.45, 219.45, 250.45
C. 79.95, 110.95, 141.95, 172.95, 203.95, 234.95, 265.95
D. 168.52 (in thousands) or \$168,521
- 37.2, 44, side-by-side boxplot, b
- 76%, 7 states, a, b, skewed right, 100
- A. Min = 15, Q1 = 21, Q2 = 25, Q3 = 31, Max = 48
B.



- Yes, 48 is an outlier
- 128.7, 133.3, 4.6, 5.0, 4000BC dataset, No, 142 is within 2 standard deviations from the mean, so it is not unusual.
 - d, d
 - $Z = .83$, $Z = .71$, Macchiato
 - True
 - sample mean, population standard deviation, sample standard deviation, population mean

