EXAM 1

Practice A

- There are 110 possible points on this exam. The test is out of 100.

- You have one class period to complete this exam, but you should be able to complete it in less than that

- Please turn off all cell phones and other electronic equipment.

- Be sure to read all instructions and questions carefully.

- Remember to show all your work. You may print your formulas in Excel using the Show Formulas option in the Formulas tab. Printed versions of your work showing formulas **and** showing the results counts as showing your work. But you must include both with your test for "showing your work" to count this way. Write your name on both print outs.

- Try all questions! You get zero points for questions that are not attempted.

- Note the last sheet lists all the equations you will need for this exam.

- *Please print clearly and neatly.*

Part I: Matching. *Write the letter from the column on the right which best matches each word or phrase in the column on the left. You will not use all the options on the right and you cannot use the same option more than once.*
    2 points each.

1. ___ Coefficient of variation

2. ___ Mean

3. ___ Median

4. ___ Mode

5. ___ Sample

6. ___ Standard deviation

7. ___ Population

A. Problem: will have error
B. Problem: when samples with unusual observations change
C. Problem: additional calculation needed
D. Problem: impractical to get data for
E. Problem: when used to compare sample with very different averages
F. Problem: samples with large outliers
G. Problem: will have accuracy issues
H. Problem: samples of continuous data
I. Problem: samples with a low standard deviation

Part II: Multiple Choice. *Choose the best answer to the following.*
    4 points each.

8. Consider this hypothetical sample data on how many exterior doors (in other words, all doors that lead outside) each house in the United States has. Based on this information available in the sample, what is the mean number of doors? (Data are also available in Practice Exam 1 Data Set 1).

| Number of Exterior Doors	Percent of Doors in the U.S.
1	10%
2	40%
3	25%
4	11%
5	3%
6	1%
Unknown	10%

   a. 2.00
   b. 2.30
   c. 2.56
   d. 3.50
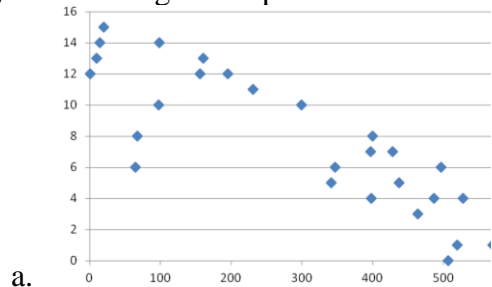   e. None of the above

9. Imagine you're tallying the results of customer surveys. After averaging three surveys, the average customer rating is 7.8 out of ten. After including a fourth survey, the average increases to 8.1. Based on the law of large numbers, what would expect the average of all customer surveys to be?
   a. Above 8.1
   b. 8.1
   c. Below 8.1 but above 7.8
   d. 7.8
   e. Below 7.8

10. Which of the following is an example of categorical data?
    a. A person's income at the age of 25.
    b. Which state a person was born in.
    c. If a person owns a car or not.
    d. B & C
    e. None of the above

11. When making a histogram in Excel, it might be tempting to use the histogram button which makes a histogram with one click. This can be a helpful tool if you're trying to get an intuitive idea of what the data look like but it shouldn't be used for something others would see. Why?
    a. Because the bins would not be easy to read.
    b. Because the bins would be too large.
    c. Because the bins would be awkwardly defined.
    d. A & C
    e. None of the above

12. Suppose you had some data concerning daily oil production (in gallons) for 500 different wells in the United States. If you wanted to get an idea for the distribution of production to see how common really good and really bad wells are, which data display would be most appropriate?
    a. Scatterplot
    b. Histogram
    c. Pie chart
    d. Box plot
    e. None of the above

13. It's not unusual for lottery winners to win the lottery again. Lottery winners, unsurprisingly, tend to like playing the lottery and will often spending their winnings on even more tickets. Which of the following is this an example of?
    a. Gambler's fallacy
    b. Hot hand fallacy
    c. Central tendency
    d. Standard deviation
    e. None of the above

14. Use the Practice Exam 1 Data Set for this question. It includes hypothetical data on a hypothetical grocery store chain called The Happy Spud. What is the mean for amount (in thousands) that Happy Spud spent on advertising in the East region?
    a. 7.10
    b. 8.20
    c. 8.36
    d. 9.36
    e. None of the above

15. Which of the following is a weakness of presenting data in a pie chart?
    a. It is intuitively difficult to tell what the chart represents
    b. People have trouble interpreting round objects
    c. It's hard to tell which section is largest
    d. A & C
    e. None of the above

16. Use the Practice Exam 1 Data Set 1 for this question with Happy Spud franchises. Create a scatterplot with Annual profit, in thousands and Number of competing stores in district with Number of competing stores in district on the horizontal axis. Which of the following is your resulting scatterplot?

a.

c.

b.

d.
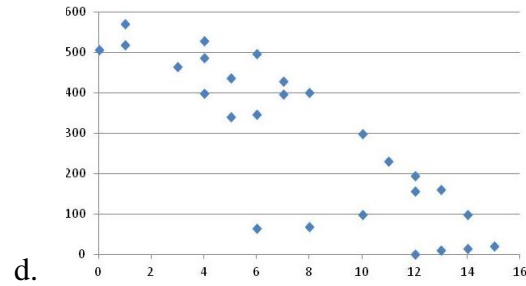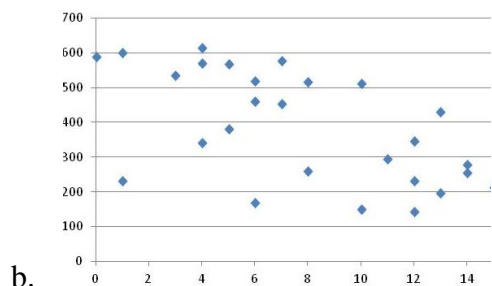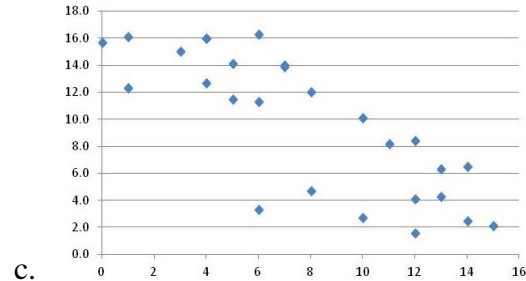
    e. None of the above

17. Ira wants to learn how people view video game violence and decides to collect a sample. Which of the following methods would result in an accurate sample?
    a. While standing outside a major metro station on a weekend, offer people a chance to win a prize if they complete a short survey.
    b. While standing outside a video game store on a weekend, ask people to take a detailed survey.
    c. While standing outside a busy bus stop Monday morning and evening, request people take a short survey.
    d. All of these will result in a very inaccurate sample.
    e. It is impossible to get an accurate sample; there is always inaccuracy.

18. Which of the following is an example of cross-sectional data?
    a. Unemployment rates in the United States, from 1800 to 1850.
    b. Profitability of various German bakeries founded between 1976 and 2016.
    c. A company's monthly revenues for the past five years.
    d. A & C
    e. None of the above

19. How does breast milk compare with feeding infants formula? Due to ethical constraints, studies addressing this question cannot randomly assign families one source of food or another. Instead, they compare outcomes (e.g. IQ) of children whose parents breast fed with those of children whose parents bottle fed. But parents who breast feed are very different from parents who bottle feed. What kind of sample bias is this?
    a. Self-selection
    b. Survivorship
    c. Undercoverage
    d. B *or* C; it is impossible to tell with the information provided
    e. It could be any; it is impossible to tell with the information provided

Part III: Short Answer. *Answer the following.*
16 points each.

20. You've been put in charge of promotion and advertising for a new line of energy drinks, XTREME CAFFINE!, at the beverage company you work for. One of your first tasks is to create a website for XTREME CAFFINE!. The data below indicate the growth rate of unique visits each month after launch (the data are also available in the Practice Exam 1 Data Set 1):

Month	Growth of Visits	Month	Growth of Visits
May	48%	September	11%
June	30%	October	6%
July	25%	November	-2%
August	12%	December	4%

The website was launched in April, with 15,000 unique visitors. By the end of the year, how many unique visits are there? What is the average growth rate over this eight-month time span? (For the second question, round to the nearest two decimal places.)

Show your work; if you used Excel to answer this question, write what you put into Excel so I know how you got the answer you did.

21. Using the information below (the data are also available in the Practice Exam 1 Data Set 1), determine which stock has a more volatile price.[1] (When doing your calculations, round to the nearest cent.) Show your work.

Month (2014)	Microsoft (Start of the Month)	Apple (Start of the Month)
January	$35.99	$555.68
February	$37.74	$502.61
March	$37.92	$523.42
April	$41.15	$537.76

Show your work; if you used Excel to answer this question, write what you put into Excel so I know how you got the answer you did.

22. As China grows wealthier, alcohol consumption is on the rise. (*Economist* "The Spirit Level")[2] China's alcohol consumption per person rose from 2.5 liters in 1978 to 6.7 liters in 2010. For the *Economist* article:

> But the countrywide statistics hide a grimmer picture. More than half the Chinese population [does not drink alcohol at all]. Those who do drink often do so to great excess. Male Chinese drinkers down far more than Japanese ones, and almost as much as notoriously sozzled British, Australian or Irish boozers.

Based on the information provided, what is the median alcohol consumption per person in China? What's advantageous about using the median here? What's disadvantageous about using it?

_____
_____
_____
_____
_____
_____
_____
_____
_____

[1] Data from Yahoo! Finance; Apple issued a 7:1 stock split in June 2014 which is why its stock price is much lower now.

[2] http://www.economist.com/news/china/21611118-chinese-are-drinking-more-spirit-level

Equation and Information Sheet

Function or Command	Result
ABS	Returns the absolute value of an input
AVERAGE	Arithmetic mean of an array
CTRL + `	Show formulas
CTRL + F	Find
CTRL + P	Print
CTRL + X	Cut highlighted area
CTRL + C	Copy highlighted area
CTRL + V	Past highlighted area
CTRL + Z	Undo
F4	Makes cell reference absolute
GEOMEAN	Geometric mean of an array (adjustments must be added manually)
LARGE	Larger values of an array (k=1 is largest, k=2 is second largest, k=3 is third largest…)
MAX	Maximum value of an array
MEDIAN	Median of an array
MIN	Minimum value of an array
MODE	Mode of an array
QUARTILE	Returns the 0th to 4th quartile of an array
SQRT	Finds the square root of the value in question.
SMALL	Smaller values of an array (k=1 is smallest, k=2 is second smallest, k=3 is third smallest…)
STDEV.S	Standard deviation of a sample
SUMPRODUCT	The summed product of two or more arrays.

Geometric Mean

$$Geometric\ Mean = \sqrt[n]{\prod_{i=1}^{n}(1 + x_i)} - 1$$

Standard deviation of a sample

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Weighted Average

$$Weighted\ Average = \frac{\sum_i^n (w_i x_i)}{\sum_i^n w_i}$$

Coefficient of Variation

$$CV_{sample} = \frac{s}{\bar{x}}(100)$$