Math 10 - Exam 1 Topics

- **Types and Levels of data**
  - Categorical, Discrete or Continuous
  - Nominal, Ordinal, Interval or Ratio

- **Descriptive Statistics**
  - Stem and Leaf Graph
  - Dot Plot (Interpret)
  - Grouped Data
    - Relative and Cumulative Relative Frequencies
    - Histogram
    - Ogive
  - Mean, Median, Mode
    - Skewness
  - Range, Variance, Standard Deviation
    - Empirical Rule
    - Z-Scores
  - Percentiles, Quartiles
    - Interquartile Range
    - Box Plot
  - Correlation
    - Bivariate Data
    - Scatterplot
    - Correlation Coefficient
  - Outliers
    - Identifying
    - Effect of outliers on Descriptive Statistics

- **Experimental Design**
  - Steps of a Statistical Process
  - Observational Study
    - Representative Sample
    - Sampling Methods
  - Experiment
    - Explanatory Variable
    - Response Variable
    - Blinding
    - Placebos

- **Probability**
  - Empirical, Classical or Subjective
  - Terms and Laws of Probability
    - Events and Outcomes
    - Sample Space
    - Complement
    - Unions and Intersections
    - Additive Rule
    - Conditional Probability
    - Tree Diagram
    - Multiplicative Rule
    - Independence
  - Changing the conditionality
  - Contingency (Two way) Tables
    - Marginal Probabilities
    - Joint Probabilities
    - Conditional Probabilities
    - Constructing table

- **Discrete Random Variables**
  - Mean and Standard Deviation
  - Probability distribution function (pdf)
  - Probability problems
  - Binomial Distribution

- **Continuous Random Variables**
  - Mean and Standard Deviation
  - Probability density function (pdf)
  - Probability and Percentile problems for Normal Distribution

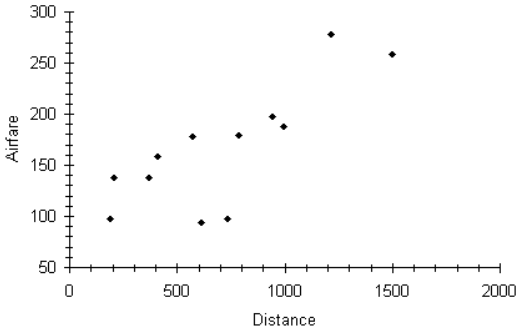
- **Central Limit Theorem**
  - pdf of the random variable $\overline{X}$ (3 important parts)
  - pdf of sample proportion $\hat{p}$
  - Probability Questions

- You must bring a picture ID to the exam.
- You may bring 4 pages of **HANDWRITTEN** notes to the exam.
- Bring your probability tables (Binomial Normal, etc)
- Bring Pencil, Calculator and your notes to the exam–**no sharing** is allowed during the exam. No cell phone calculators.
- Cell Phones, tablets, smart watches, PDAs, and other electronic devices must be **turned off** and **put away**.
- Manage your time so you can **attempt every question**.

Practice Questions for Exam 1

1. 10% of American adults have Type II diabetes. A test has been developed that has a 80% chance of correctly detecting this disease, but has a false positive rate of 15%.
    a. Draw a tree diagram of all possibilities, where the first branch represents person having Type II diabetes (positive D+ or negative D-) and the second branch represents the test (positive T+ or negative T-).
    b. What percentage of American adults will TEST positive for the Type II diabetes?
    c. Given an adult tests positive for the disease, what is the probability the adult actually has Type II diabetes?

2. The data shown in the scatter plot is the distance traveled and the airfare for 12 flights on Delta Airlines:
    a. Which of the following is a reasonable estimate of the correlation coefficient? (Circle one answer)
        1.0 0.8 0.0 -0.5
    b. What does this graph tell us about distance and airfare
    c. What is the type and level of distance traveled?
        Type (Circle One) Categorical Discrete Continuous
        Level (Circle One) Nominal Ordinal Interval Ratio



3. You have a 70% chance of being on time to class today and a 80% chance of being on time to class tomorrow. Assume these two days are independent events.
    a. Find the probability of being on time to class both today **and** tomorrow.
    b. Find the probability of being on time to class at least once today **or** tomorrow.

4. The following data represent the daily births at a hospital for 20 days

14	15	15	17	17	19	19	21	23	25
26	27	31	35	36	47	48	59	70	99

a) Construct a stem and leaf diagram of the data
b) Calculate the interquartile range for this data set.
c) Calculate the median for this data set.
d) Make a box plot for the data
e) Using the rule that potential outliers are more than 1.5 IQRs from the box, determine if 99 is an outlier.
f) **Without calculating**, what can you say about the mean births for this Hospital.**(check one answer below)**?
    - ❑ The mean is greater than the median.
    - ❑ The mean is less than the median.
    - ❑ The mean is about the same as the median.
    - ❑ None of the above – no way to know without calculating.

5. The following data represents the hours per week worked outside of school by 200 randomly selected night students at a community college:

Hours	Frequency	Relative Freq	C.R.Freq
1-8	20		
9-16	30		
17-24	50		
25-32	60		
33-40	30		
41-48	10		

a) In the space above, determine the relative frequencies and cumulative relative frequencies.
b) Sketch a relative frequency **histogram**, showing **all horizontal and vertical** labels.
c) Sketch a cumulative relative frequency **ogive**, showing **all horizontal and vertical** labels.
d) Estimate the median from the graph.
e) What percentage of the night students work 32 hours per week or less?
f) **Without calculating** but **explaining your reasoning**, which of the following is a reasonable estimate for the standard deviation? a) 0.5 b) 1 c) 10 d) 50

6. Determine if each of the following data are categorical, continuous or discrete (circle one for each)

a.	Number of fatalities from a tsunami:	categorical	continuous	discrete
b.	Time spent in traffic:	categorical	continuous	discrete
c.	Number of Songs on your I-pod:	categorical	continuous	discrete
d.	Your student number	categorical	continuous	discrete
e.	Names of cities in California with a Walmart:	categorical	continuous	discrete
f.	Price per gallon of gasoline:	categorical	continuous	discrete
g.	Number of Courses taken in a year.	categorical	continuous	discrete
h.	Tons of steel used by a manufacturer:	categorical	continuous	discrete

7. 1000 students (500 morning, 300 afternoon, 200 night) were asked how often they use the campus library. The results are summarized in the table below:

	Never uses library	*Sometimes uses library*	*Frequently uses libray*	***Total***
Morning	200	250	150	**600**
Afternoon	80	145	75	**300**
Night	80	10	10	**100**
Total	**360**	**405**	**235**	**1000**

a. Find the following probabilities:
   i) A randomly selected student never uses the library.

   ii) A randomly selected student is a night student and frequently uses the library.

   iii) **Given** the student is an afternoon student, the student never uses the library.

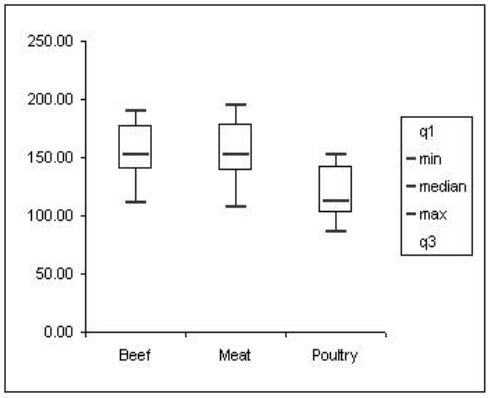
b. Are "Afternoon Student" and "Never uses library" Independent Events? **Justify and explain your answer**.

c. Would the probabilities generated from this data be classical, empirical or subjective probability?

8. These descriptive statistics and boxplots were generated from data representing **calories per serving** for three types of hotdogs: All **Beef**, Mixed **Meat** and **Poultry.**

Descriptive statistics

	Beef	*Meat*	*Poultry*
count	20	17	17
mean	156.85	158.71	118.76
sample variance	512.66	636.85	508.57
sample standard deviation	22.64	25.24	22.55
minimum	111	107	86
maximum	190	195	152
range	79	88	66
median	152.5	153	113
mode	149	#N/A	102



a. Compare the mean to the median calories for the **Meat** group. Is the result consistent with the shape of the box plot? Explain your answer.

b. If the data is approximately bell shaped, between what two values of calories would you expect to find about 95% of the **Beef** data?

c. Which of the three groups has the most variability in calories per serving? Explain your answer.

d. Hebrew National All Beef Hotdogs had 190 calories per serving. Calculate and interpret the z-score for Hebrew National Hotdogs using the **Beef** Category data.

e. Determine the probability a randomly selected **Poultry** Hot Dog exceeds 113 calories.

f. Compare the three groups and draw at least two conclusions from the results.

9. From samples of a total of 2100 young (18-24 year old) White, Black and Latino men taken in January 2010 in the U.S., the unemployment rate of each sample was determined as given in the following table. (*2013, Urban Institute, The Labor Market Performance of Young Black Men in the Great Recession*). The study used stratified sampling. The Urban Institute concluded that young black men have a higher unemployment during the recession than their white and Latino peers.

Race/Ethnicity	Unemployment Rate
White	15.6%
Black	30.0%
Hispanic	26.9%

   a. What is the population and what is the sample?

   b. Identify the steps of the statistical process:

      - Ask a question that can be answered with sample data.
      - Determine the information needed.
      - Collect sample data that is representative of the population.
      - Summarize, interpret and analyze the sample data.
      - State the results and conclusion of the study.

10. A study was conducted to examine the effects of active recovery (AR), massage (MR), and cold water immersion (CR) on performance of repeated bouts of high-intensity cycling separated by 24 hours. A sample of physically active men aged 18–30 were randomly assigned to one of four groups. Each group performed an intense 18-minute cycling workout after which each underwent a 15-minute recovery period. In the 15 minutes, the first group (AR) continued to cycle at a low level, the second group (MR) received leg massage, the third group (CR) immersed their legs in a bath of cold water. The last group simply sat and rested. The next day the subjects did the same intense 18-minute cycling workout. Each exercise was done on a cycle ergometer so that the work level (measure in kilojoules) was calculated for each. The researchers found that on the second day, that there was no difference in the performance level of the subjects in the AR, MR and CR, but that the subjects who just sat in a chair to rest did less work than the other groups. (*Journal of Strength and Conditioning Research (2004; 18 [4], 855-60*).

    a. What is the explanatory variable?

    b. What is the response variable?

    c. Which groups are the treatment groups?

    d. Is there a control group? If so, which one?

    e. Is there blinding in this experiment? Explain your answer.

11. 70% of students at a large New York University receive some financial aid.

    a. If 4 students are randomly selected, determine the probability that **exactly 2** students in the sample receive some financial aid.
    b. If 4 students are randomly selected, determine the probability that **less than 2** students in the sample receive some financial aid.

    .

12. The random variable X follows the probability distribution function as shown to the right:
    a. Determine P(X=3)
    b. Determine the population mean.
    c. Determine the population variance

x	P(x)
0	0.1
1	0.3
2	0.2
3	
4	0.1

13. 40% of students at a college use the cafeteria.
    a. If 9 students are randomly sampled, determine the probability that less than 3 use the cafeteria. If 9 students are randomly sampled and X represents the number of students in the sample who use the cafeteria, find the **mean** and **standard deviation** of X.

14. Find the 30^{th} percentile for the cooking time for oatmeal which follows a Normal Distribution with a mean of **4** and a standard deviation of **3**

15. Students' exam scores for a course follow a Normal Distribution with **μ=70** and **σ=10.**

    a. Find the probability a randomly selected student scores a **75 or more**.

    b. Find the exam score which is the **25^{th} percentile** of this distribution.

    c. You take a random sample of 40 students. Find the probability the **sample mean** is between 68 and 72.

    d. Would your answer for part c be different if the probability distribution of "exam scores" did not follow a Normal distribution? **Explain your answer**.

16. The age of a grove of walnut trees follow a Normal Distribution with **μ=50 years** and **σ=15 years.**

    a. Find the probability that the age of a randomly selected tree is between 40 and 70 years.

    b. Find the probability of a randomly selected tree has lived exactly 45.231789 years.

    c. Find the 30^{th} percentile of this distribution.

17. 35% of students at De Anza College plan to transfer to San Jose State. 200 students are randomly selected and the sample proportion $\hat{p}$ will be calculated.
    a. Determine the expected value and standard deviation of the sample proportion.
    b. Determine that the condition for normality is satisfied.
    c. Determine the probability the sample proportion exceeds 0.40.