MIDTERM: MATH 209 - STATISTICS

Problem 1. Define the notion of a random sample S of a population X and give at least one example.

Problem 2. Below is a list vertical jumps (in cm) for 24 athletes who compete in various different sports.

28	45	55	46	70	72	62	32	75	68	74	54
50	62	63	31	55	68	73	55	60	69	68	92

Part A. Using # of classes = 4, create a frequency table for the given data set, **and** use your frequency table to construct a histogram.

Part B. Using the data set above, calculate the quartiles Q_i where $i = 0, 1, 2, 3, 4$, and construct a box-and-whiskers plot for this data set.

Part C. Calculate the 15th percentile of this data set.

Problem 3. Consider the sample $S = \{101.1, 102.0, 103.2, 99.7\}$ which represent mean temperatures in Fahrenheit of four patients over a particular period of time. Calculate the sample mean $\bar{x}$ and the sample standard deviation σ_x of the data set S.

Problem 4. Consider a data set S with 51 signals of value 1 and 49 signals of value -1. Calculate the mean $\bar{x}$ of S and the standard deviation σ_x. What would happen to the mean and standard deviation if we added into our data set a bunch of signals of value 0?

Problem 5. Define $\binom{n}{2}$ to be the value $n(n-1)/2$. Calculate the following:
Part A. Calculate $\binom{2}{2}$, $\binom{3}{2}$, $\binom{4}{2}$, $\binom{5}{2}$, $\binom{6}{2}$

Part B. Calculate $\sum_{n=2}^{6} \binom{n}{2}$

Part C. Let $p = 0.5$. Calculate $\sum_{i=2}^{5} p^i$ and $\sum_{i=2}^{5} \binom{i}{2} p^i$.

Part D. Does the function $f(x) = \binom{x}{2} p^x$ for $x = 1, 2$ describe a finite probability distribution on the set $\{1, 2\}$? Why or why not?

Part E. Out of a population of 52 people, how many possible samples of size 2 are there?

Problem 6. Let A and B be two events and assume $P(A) = 0.35$ and $P(B) = 0.45$. Assume further that $P(A|B) = 0.80$.
Part A. Find $P(A \cap B)$ using the general multiplication rule.

Part B. Find $P(A \cup B)$ using the general addition rule.

Part C. Find $P(A^c)$ using the subtraction rule.

Problem 7. Consider the following table formed by a population of 70.

	A	B
X	22	17
Y	18	13

Part A. Find $P(A),\ P(B),\ P(X),\ P(Y)$.

Part B. Find the probability of $X \cap B$ and $X \cup B$.

Part C. Find the conditional probability of $P(X|A)$ and $P(B|Y)$.

Problem 8. Consider the following game: by betting 1 dollar, you have a 35% chance of winning the game which gives you a reward of $2.50. Also, assume you have a 15% chance to neither win nor lose any money (i.e., your $1 is returned) and assume that you have a 50% of losing the game (i.e., you lose your original dollar). Draw a probability tree and calculate the expectation of this game. Would you play it? What would happen to your stash if you played the game a whole bunch of times (i.e., if you had enough money to play the game 1 million times)?

Problem 9. An unbalanced coin is tossed 10 times with probability of heads $p = 0.73$. Calculate $P(x > 8)$ using the binomial distribution.

Problem 10. Consider the binomial distribution with $n = 15$ and $p = 0.80$. Calculate the mean μ and standard deviation σ of this distribution. Use this to find the probability $P(\mu - 2\sigma < x < \mu + 2\sigma)$.

Problem 11. Calculate the *standard deviation* s_x of the data set $S = \{5, 6, 10, 16, 21, 30\}$ using the formulas below:

(a) Calculate the *sum of squares*:

$$SS_x = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

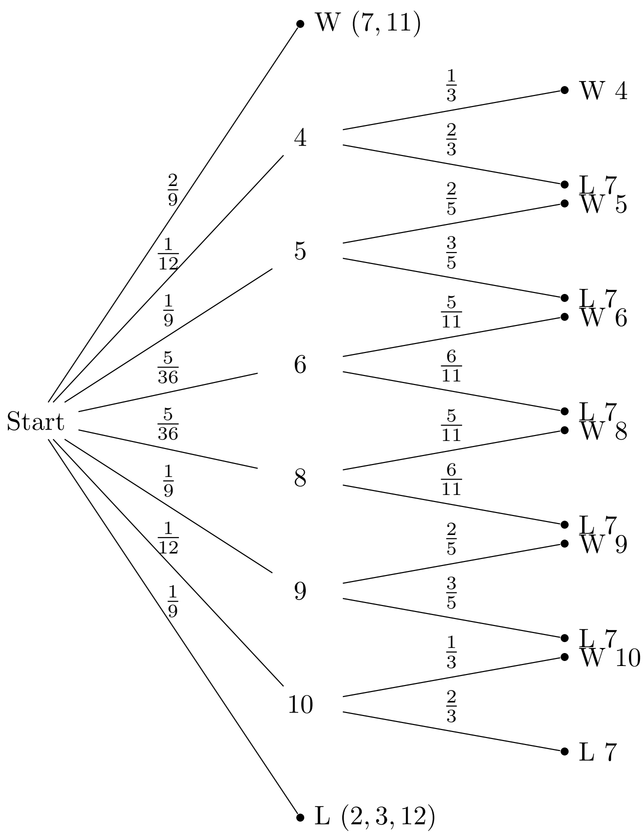
(b) Calculate the *variance*:

$$s_x^2 = \frac{SS_x}{n-1}$$

(c) Calculate the *standard deviation*:

$$s_x = \sqrt{s_x^2}$$

Problem 12. Below is a probability tree, which describes the probabilities of each possible outcome of the game of Craps which we denote as f. Calculate the expectation $\mathbb{E}(f)$ of this game f.



REMARK 0.1. (A description of the above probability tree.) The gambling game Craps is played by rolling two fair six-sided dice and summing the sides up to get a number between 2 and 12. A player automatically wins (W) if they roll a 7 or 11. The player automatically loses (L) if they get a 2, 3, or 12. Alternatively, if they roll another number x (–i.e., x could be any number in $\{4, 5, 6, 8, 9, 10\}$) they must roll and re-roll again and again until they either roll the **same** number x a second time (this counts as a win (W)) or they roll a 7 (which counts as a loss (L)). If the player bets \$1 then on a win (W), the player doubles their money to \$2 making a profit of \$1. If the player loses (L), the player forfeits their original bet of \$1.

Problem 13. Explain the main differences between a discrete probability distribution and a continuous probability distribution. Give at least one example of each.

Problem 14. An important continuous distribution for describing rare events is the so-called *Poisson Distribution* given by the formula $f(k;t) = \frac{t^k e^{-t}}{k!}$ where e is Euler's constant. In a binomial distribution problem if n is large and p is close to 0, then the probability of k successes $Pr(x = k)$ is approximately $f(k;t)$ with $t = n \cdot p$. This is known as *law of small numbers or rare events*. In practical applications, this is a useful fact. For example, if in Problem 3 our drug has **non-**efficacy p equal 0.1% and $n = 1,000,000$, we may use this approximation to make quick calculations concerning the number of patients who will get better on the drug. In this vein, apply this statement to quickly give approximate answers to the following statements:

A) The probability that the drug effectively treats all 1 million patients?

B) The probability that the drug effectively treats at least three patients?

C) The probability that at least 1,000 patients don't get better through the use of the drug (they may get better for other reasons outside the scope of the experiment).

Problem 15. State Chebyshev's Inequality and use it to explain why we need to know both the mean μ and the standard deviation σ of a data set.