5.1 Review and Preview

Theoretical Probability: using the probability of an event occurring based on what would happen in theory

Empirical Probability: using real sample data to find the probability of an event occuring

5.2 Probability Distributions

Random variable:

Probability distribution:

→ often expressed as a graph, table, or formula

→ Ex: 12 jurors are to be randomly selected (without bias) from a population in which 80% of the jurors are Mexican-American.

If we let x = number of Mexican-American jurors among 12 jurors

Then x is a random variable because its value depends on chance. The possible values of x are 1, 2, 3, …, 12. The following table lists the values of x along with the corresponding probabilities. Probability values that are very small (ex: 0.000000123) are represented by 0+.

x (Mexican Americans)	P(x)
0	0+
1	0+
2	0+
3	0+
4	0.001
5	0.003
6	0.016
7	0.053
8	0.133
9	0.236
10	0.283
11	0.206
12	0.069

Discrete random variable:

→ Ex #1: Let x = the # of eggs that a hen lays in a day

→ Ex #2: Let x = the # of students in this class

Continuous Random Variable:

→ Ex #1: Let x = the amount of milk a cow produces in a day

→ Ex #2: Let x = the measure of voltage for a particular smoke detector battery

Examples of Determining if each of the following is a discrete random variable, continuous random variable, or neither:

1. The speed of an airplane __________________________

2. The number of ships in Pearl Harbor on any given day _____________________

3. The distance a golf ball travels after being hit with a driver ________________

4. The eye colors of the players on the NY Giants __________________________

5. The number of lightning strikes in Rocky Mountain National Park on a given day __________________________

Requirements for a Probability Distribution:

1. There is a numerical random variable x and its values are associated with corresponding probabilities.
2. $\sum P(x) = 1$
3. $0 \leq P(x) \leq 1$

→ Because the probability distribution is based on a discrete random variable, we will not use class boundaries for the horizontal scale but rather the bars will represent each discrete value.

→ The vertical scale shows probabilities instead of relative frequencies

→ Ex #1: Does the table on the right describe a probability distribution?

x	P(x)
0	0.2
1	0.5
2	0.4
3	0.6

→ Ex #2: Suppose we have a function for a probability distribution where P(x) = x/3 where x can be 0, 1, or 2. Does this function follow the requirements for a probability distribution?

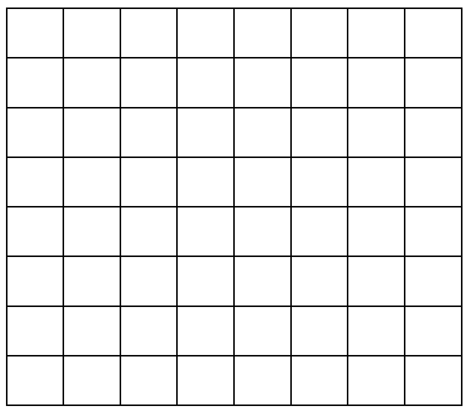
Probability Histogram:

➔ Ex #1: Use the probability distribution about the 12 jurors to create the following probability histogram:

x (Mexican Americans)	P(x)
0	0+
1	0+
2	0+
3	0+
4	0.001
5	0.003
6	0.016
7	0.053
8	0.133
9	0.236
10	0.283
11	0.206
12	0.069

→ Ex #2: The following table describes the probability distribution for the number of girls in two births. Create a probability histogram.

x	P(x)
0	0.25
1	0.5
2	0.25



Parameters of a Probability Distribution

- ➔ The probability histogram can give us insight into the nature or shape of the distribution

- ➔ Use the following formulas to find the mean, variance, & standard deviation of data from a probability distribution

Mean: $\mu = \sum[x \bullet P(x)]$

Variance: $\sigma^2 = \sum[(x-\mu)^2 \bullet P(x)]$

Variance: $\sigma^2 = \sum[x^2 \bullet P(x)] - \mu^2$

Standard deviation: $\sigma = \sqrt{\sum[x^2 \bullet P(x)] - \mu^2}$

-***Note:*** Evaluate $[x^2 \bullet P(x)]$ by first squaring each value of x, then multiplying each square by the corresponding probability P(x), then adding those products together.

-Round your answers by carrying one more decimal place than the number of decimals used for the random variable (x) unless more precision is necessary

- ➔ Ex: Use the previous example of choosing 12 jurors & the following probability distribution to find the mean, variance, & standard deviation.

x (Mexican Americans)	P(x)	$x \cdot P(x)$	$x^2 \cdot P(x)$
0	0+		
1	0+		
2	0+		
3	0+		
4	0.001		
5	0.003		
6	0.016		
7	0.053		
8	0.133		
9	0.236		
10	0.283		
11	0.206		
12	0.069		

-Mean: $\mu = \sum[x \bullet P(x)]$

-Variance: $\sigma^2 = \sum[x^2 \bullet P(x)] - \mu^2$

-Standard deviation: σ

Expected Value of a Discrete Random Variable:

$$E = \sum [x \bullet P(x)]$$

→ theoretical mean outcome for infinitely many trials
- ie: Expected value is the average value that we would expect to get if the trials would continue indefinitely

→ plays an important role in decision theory

→ the mean of a discrete random variable is the same as its expected value

→ Ex: When selecting 12 jurors from the Hidalgo County Population, the mean number of Mexican-Americans is 9.6 so the expected value of the number of Mexican-Americans is also 9.6

→ Ex: The following table describes the probability distribution for the number of girls in two births. Find the mean, variance, standard deviation, and expected value for the probability distribution.

x	P(x)	$x \cdot P(x)$	$x^2 \cdot P(x)$
0	0.25		
1	0.5		
2	0.25		

-Mean: $\mu = \sum [x \bullet P(x)]$

-Variance: $\sigma^2 = \sum [x^2 \bullet P(x)] - \mu^2$

-Standard deviation: σ

-Expected value = μ

→ You can use expected value to compare the average gain and loss of two different bets:

Ex #1: If you bet $1 in Kentucky's Pick 4 lottery game, you either lose $1 or gain $4,999. (The winning prize is $5,000 but your $1 bet is not returned so the net gain is $4,999). The game is played by selecting a 4-digit number between 0000 and 9999. If you bet $1 on 1-2-3-4, what is your expected value of gain or loss?

Event	***x***	***P(x)***	***x•P(x)***
Lose			
Gain (net)			
Total			

Ex #2: The probabilities and payoffs for betting $1 on the number 7 in roulette are summarized in the following table. (Remember, there are 38 possible outcomes in roulette). Find the expected value of gain or loss.

Event	x	P(x)	x · P(x)
Lose	-$1	37/38	
Gain (net)	$35	1/38	

Making Sense of Results: There are 2 approaches for determining whether a value of a random variable x is unusually low or unusually high.

1. Range Rule of Thumb:

Ex: Use the range rule of thumb to determine whether a jury consisting of 7 Mexican-Americans among 12 jurors is usual or unusual?
(We found the mean to be 9.6 and the standard deviation to be 1.4)

2. Rare Event Rule of Inferential Statistics: If, under a given assumption (such as that a coin is fair), the probability of a particular observed event (such as 992 out of 1000 coin tosses) is extremely small, we conclude that the assumption is probably not correct.

→ Unusually high # of successes:

→ Unusually low # of successes:

x (Mexican Americans)	P(x)
0	0+
1	0+
2	0+
3	0+
4	0.001
5	0.003
6	0.016
7	0.053
8	0.133
9	0.236
10	0.283
11	0.206
12	0.069

Ex: If 80% of those eligible for jury duty in Hidalgo County are Mexican-American, then a jury of randomly selected people should have around 9 or 10 who are Mexican-American. Is 7 Mexican-American jurors among 12 an unusually low number? Could this suggest discrimination in the selection process?

Ex: The following table describes the probability distribution for the number of girls in two births.

x	P(x)
0	0.25
1	0.5
2	0.25

Determine if exactly 2 girls would be considered usual or unusual using the range rule of thumb. (We found the mean to be 1.0 and the standard deviation to be 0.7)

Determine if exactly 2 girls would be considered unusually high among 2 children using the rare event rule.

5.3 Binomial Probability Distributions

Binomial Probability Distribution: results from a procedure that meets the following requirements…

1. The procedure has a fixed # of trials

2. The trials must be independent (the outcome of any trial doesn't affect the probabilities in other trials)

3. Each trial must have all outcomes classified into 2 categories—success & failure

4. The probability of success remains the same in all trials

→ Notation for Binomial Probability Distributions:

S	
F	
$P(S) = p$	
$P(F) = q = 1 - p$	
n	
x	
p	
q	
P(x)	

→ Be sure that x and p both refer to the same category being called a success (where a success doesn't necessarily represent that something is good but rather that a certain outcome occurs).

→ Use the 5% Guideline for Cumbersome Calculations: When sampling without replacement, consider events to be independent if the sample size n is less than 5% of the population size N: $n \le 0.05N$

→ Ex #1: If we need to select 12 jurors from a population that is 80% Mexican-American, & we want to find the probability that among 12 randomly selected jurors, exactly 7 are Mexican-Americans.

a. Does this procedure result in a binomial distribution?

1. Is the # of trials fixed?

2. Are the trials independent?

3. Does each trial have 2 categories of outcomes?

4. Does the probability of success remain the same in all trials?

Binomial distribution?

b. Identify the values of n, x, p, q

n =

x =

p =

q =

→ Ex #2: If we want to select 2 girls from a family with 5 children?

a. Does this procedure result in a binomial distribution?

1. Is the number of trials fixed?

2. Are the trials independent?

3. Does each trial have 2 categories of outcomes?

4. Does the probability of success remain the same in all trials?

Binomial distribution?

b. Identify the values of n, x, p, q

$n =$

$x =$

$p =$

$q =$

3 Methods for Finding the Probabilities corresponding to the Random Variable x in a Binomial Distribution

*Remember all probabilities should be rounded to ______________________________.

1. *Using the Binomial Probability Formula*

$$P(x) = \frac{n!}{(n-x)!x!} \bullet p^x \bullet q^{n-x}$$ for x = 0, 1, 2, …, n

where n = # of trials
x = # of successes among n trials
p = probability of success in any one trial
q = probability of failure in any one trial (q = 1 – p)

Ex #1: Use the binomial probability formula to find the probability of getting exactly 7 Mexican-Americans when 12 jurors are randomly selected from a population that is 80% Mexican-American.

Find P(7) given that n =______ x = _____ p = _______ q = _____

Ex #2: If we want to select 2 girls from a family with 5 children?

Find P(2) given that n =______ x = _____ p = _______ q = _____

2. *Using Table A-1 in Appendix A*

1. Locate n and the corresponding value of x that is desired
2. Align that row with the proper probability of p by using the column across the top.
3. A very small probability is indicated by 0+ (such as 0.000064)

*Using the table is often more efficient; however, it only contains a limited number of values of n and p so it won't always work.

Ex #1: Use Table A-1 to find the following binomial probabilities:

a. The probability of exactly 2 girls out of a family of 5 children

b. The probability of at least 2 girls out of a family of 5 children

Ex #2: Use Table A-1 to find the following probabilities:

a. The probability of rolling exactly one five out of 8 rolls of a die

b. The probability of no more than two fives out of 8 rolls of a die

3. *Using the TI-83/84 Plus Graphing Calculator*

1. Press STAT Edit 1: Edit
2. Enter the possible values of x into L1
3. With the cursor on L2, press DISTR A: binompdf(
4. Enter the value of n, comma, then the value of p and close the parenthesis
5. When you press ENTER, a list of probabilities should appear in L2

* Using the calculator is the easiest way to find binomial probabilities but you still must add the individual probabilities for each x if necessary.

Ex #1: Try it with the previous example about the number of Mexican-American jurors chosen on a jury of 12

x (Mexican Americans)	**P(x)**
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

What is the probability of getting exactly 7 Mexican-American jurors?

What is the probability of 7 or fewer Mexican-American jurors?

Ex #2: Use the data about the number of girls in a family of 5 children. Create a probability distribution. Then find the binomial probabilities that follow.

<u>x</u>	**<u>P(x)</u>**

What is the probability of selecting exactly 2 girls out of 5 children?

What is the probability of selecting at least 2 girls out of 5 children?

<u>To find the sum of binomial probabilities with the TI-84 graphing calculator:</u>

On the main calculation screen:

--Press 2nd Stat to get the LIST menu
--Arrow over to MATH
--Choose 5:Sum(
--Press 2nd Vars to get the DISTR menu
--Choose A:binompdf(
--Enter the values (n, p, x) and press ENTER
--if there is more than one x value, they must be typed in { } brackets separated by commas

Ex #1: Try it for the probability of selecting at most 7 Mexican-American jurors out of a jury of 12

Ex #2: Try it for the probability of selecting at least 2 girls out of 5 children

5.4 Parameters for Binomial Distributions

The formulas given previously for discrete probability distributions can be greatly simplified for binomial distributions:

	Discrete probability distributions	Binomial distributions
Mean:	$\mu = \sum [x \bullet P(x)]$	$\mu = np$
Variance:	$\sigma^2 = \sum [x^2 \bullet P(x)] - \mu^2$	$\sigma^2 = npq$
Standard deviation:	$\sigma = \sqrt{\sum [x^2 \bullet P(x)] - \mu^2}$	$\sigma = \sqrt{npq}$

You can still use the same range rule of thumb to determine if values are unusual:

Minimum usual value: _______________

Maximum usual value: _______________

*Note: Because a binomial distribution is a particular type of discrete probability distribution, we could use those formulas, but if we know the values of n and p, it is much easier to use the formulas for the binomial distributions.

Ex #1: Find the mean, variance, & standard deviation for the numbers of Mexican-Americans on a jury of 12 people selected from the population that is 80% Mexican-American.

n =

p =

q =

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

Standard deviation: $\sigma = \sqrt{npq}$

Ex #2: Find the mean, variance, & standard deviation if during a period of 11 years, 870 people were selected for duty on a grand jury in Hidalgo County, TX.

n = _______ p = _____ q = _______

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

Standard deviation: $\sigma = \sqrt{npq}$

Based on those numbers & the range rule of thumb, determine if the actual result of 339 Mexican-American jurors chosen is unusual. Does this suggest that the selection process discriminated against Mexican-Americans?

Minimum usual value: $\mu - 2\sigma =$

Maximum usual value: $\mu + 2\sigma =$

Ex #3: The brand name of McDonald's has a 95% recognition rate. A special focus group consists of 12 randomly selected adults to be used for extensive market testing. For such random groups of 12 people, find the mean, variance, and standard deviation for the number of people who recognize the brand name of McDonalds.

Based on the values you just found and the range rule of thumb, determine if it would be unusual to have only 6 people in the focus group of 12 people that recognized the brand name McDonalds.

5.5 Poisson Probability Distributions

Poisson Distribution:

→ The interval can be time, distance, area, volume, or other similar unit
→ The probability of the event occurring over an interval is given by the following formula:

$$P(x) = \frac{\mu^x \bullet e^{-\mu}}{x!}$$

where $e \approx 2.71828$

μ = the mean number of occurrences of the event over the intervals

→ Often used for describing the behavior of rare events (with small probabilities)

Ex: radioactive decay, arrivals of people in a line, eagles nesting in a region, patients arriving at an emergency room, internet users logging onto a website

→ *Requirements for the Poisson Distribution:*

1. The random variable x is the number of occurrences of an event over some interval
2. The occurrences must be random
3. The occurrences must be independent of each other
4. The occurrences must be uniformly distributed over the interval being used

→ *Parameters of the Poisson Distribution:*

--the mean is μ

--the standard deviation is $\sigma = \sqrt{\mu}$

→ *A Poisson distribution differs from a binomial distribution in the following 2 ways:*

1. The binomial distribution is affected by the sample size n and the probability p, whereas the Poisson distribution is affected ____________
____.

2. In a binomial distribution, the possible values of the random variable x are 0, 1, …, n, but a Poisson distribution ____________________________
____.

→ Ex: In analyzing hits by V-1 buzz bombs in World War II, South London was subdivided into 576 regions, each with an area of 0.25 km^2. A total of 535 bombs hit the combined area of 576 regions.

a. If a region is randomly selected, find the probability that it was hit exactly twice.

b. Based on the probability found in part a, how many of the 576 regions are expected to be hit exactly twice?

→ Ex: For a recent period of 100 years, there were 530 Atlantic hurricanes. Assume that the Poisson distribution is a suitable model.

a. Find the mean number of hurricanes per year:

b. If P(x) is the probability of x Atlantic hurricanes in a randomly selected year, find P(9):

<u>Using Poisson Distribution as an Approximation to the Binomial Distribution:</u>

The Poisson Distribution may be used to approximate the binomial distribution when n is large and p is small.

- Requirements for Using the Poisson Distribution as an Approximation to the Binomial Distribution:
  - $n \geq 100$
  - $np \leq 10$

- mean is found using $\mu = np$

- Ex #1: In Kentucky's Pick 4 lottery game, you pay \$1 to select a sequence of 4 digits (such as 2283). If you play this game once every day, find the probability of winning exactly once in 365 days.

- Ex #2: In the Maine Pick 4 game, you pay \$0.50 to select a sequence of four digits, such as 2-4-4-9. If you play this game once every day, find the probability of winning at least once in a year with 365 days.

■ Using the Graphing Calculator to Find the Poisson Distribution:

1. Press 2nd VARS to get to DISTR menu
2. Scroll down and select C: poissonpdf(
3. Press ENTER
4. Enter the value of μ , comma, the value of x, then close the parenthesis
5. Press ENTER

Ex #1: Try using the graphing calculator to find the Poisson distribution for the previous example about winning Kentucky's Pick 4 lottery game once out of 365 days.

Ex #2: Use the graphing calculator to find the Poisson distribution for the previous example about winning the Maine Pick 4 game at least once out of 365 days.