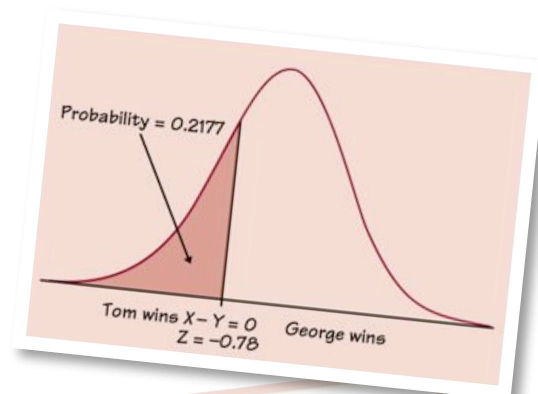
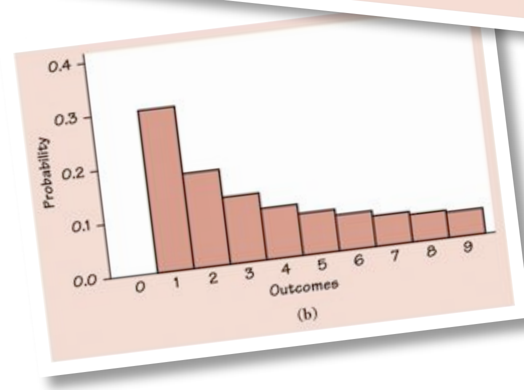
Chapter 7

Random Variables





In Chapter 6, we learned that a "random phenomenon" was one that was unpredictable in the short term, but displayed a predictable pattern in the long run. In Statistics, we are often interested in numerical outcomes of random phenomena. In this chapter, we will learn to define random variables to describe numeric outcomes of random phenomena as well as how to calculate the means and variances of such random variables.

Random Variables:

- 7.1: Discrete and Continuous Random Variables
- 7.2: Means and Variances of Random Variables

AP Statistics Chapter 7: Random Variables

"It seems that to make a correct conjecture about any event whatever, it is necessary to calculate exactly the number of possible cases and then to determine how much more likely it is that one case will occur than another." ~Jakob Bernoulli

Tentative Lesson Guide					
Date		**Stats**	**Lesson**	**Assignment**	**Done**
Mon	12/11	7.1	Discrete RandomVariables	Rd 391-395 Do **1-5**	
Tues	12/12	7.1	Continuous RandomVariables	Rd 397-403 Do **6-12, 14-16**	
Wed	12/13	7.2	Mean/Variance of R.V.	Rd 407-411 Do **22-26, 29**	
Thu	12/14	7.2	Rules for Mean/Variance	Rd 418-423 Do **34-39, 41**	
Fri	12/15	7.2	Combining Random Variables	Rd 424-427 Do **42-46, 49-50**	
Mon	12/18	Rev	Review	Practice Problems	
Tues	12/19	Rev	Review	Practice Problems	
Wed	12/20	**Exam**	**Exam Chapter 7**	**Online Quiz Due**	
Thurs	12/21		***Casino Lab***		
Fri	12/22		***Casino Lab***		
		Have a Great Happy Festive Winter Season Break!			

Note:
The purpose of this guide is to help you organize your studies for this chapter.
The schedule and assignments may change slightly.

Keep your homework organized and refer to this when you turn in your assignments at the end of the chapter.

Class Website:
Be sure to log on to the class website for notes, worksheets, links to our text companion site, etc.

http://web.mac.com/statsmonkey

Don't forget to take your online quiz!. Be sure to enter my email address correctly!
http://bcs.whfreeman.com/yates2e

My email address is:

jmmolesky@isd194.k12.mn.us

Chapter 7 Objectives and Skills:

These are the expectations for this chapter. You should be able to answer these questions and perform these tasks accurately and thoroughly. Although this is not an exhaustive review sheet, it gives a good idea of the "big picture" skills that you should have after completing this chapter. The more thoroughly and accurately you can complete these tasks, the better your preparation.

Random Variables

☐ A Random Variable is a variable whose outcome is a numerical value of a random phenomenon.

☐ When describing a random variable "X", be sure to note the probability distribution, showing the values X takes on and their respective probabilities.

Discrete and Continuous Random Variables

☐ A "discrete" random variable has a countable number of possible values.

☐ We can display the probability distribution of a discrete random variable using a probability histogram. The height of each bar represents the probability of the outcome.

☐ A "continuous" random variable takes on all possible values in an interval of numbers.

☐ We can display the probability distribution of a continuous random variable with a density curve.

☐ All continuous probability distributions assign a probability of zero to each individual outcome. Probabilities are defined over ranges of values.

Means and Variances

☐ The mean, or "expected value" of a random variable is the average value we would expect to see in a long run of repeated observations.

☐ For a discrete RV, $E(X) = \mu_X = \sum x_i p_i$

☐ The variance of a random variable is a measure of the amount of variability we would expect to see in a long run of repeated observations.

☐ For a discrete RV,

$$Var(X) = \sigma_X^2 = \sum (x_i - \mu_X)^2 p_i$$

☐

Rules for Means and Variances

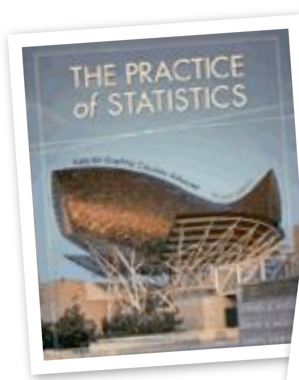
☐ We often work with combinations or linear transformations of Random Variables. We can calculate the means and variances of these new random variables using the following formulas:

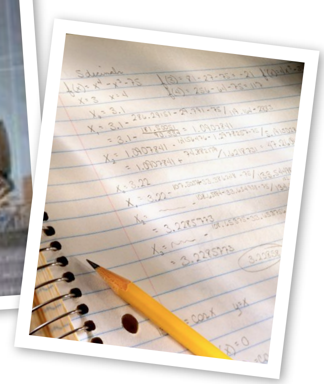
$$\mu_{a+bX} = a + b\mu_X$$

$$\sigma_{a+bX}^2 = b^2\sigma_X^2$$

$$\mu_{X \pm Y} = \mu_X \pm \mu_Y$$

$$\sigma_{X \pm Y}^2 = \sigma_X^2 + \sigma_Y^2$$





7.1: Introduction to Random Variables - "Greed"

The point of this game is to gain as many points as possible. Points are earned by rolling two dice–the sum of the dice equals the points earned for the roll. To earn points, a player must be actively "in the game". Once a player opts "out of the game," they maintain their current point total, but can not earn any more points until the next round. Players who are "in the game" continue earning points until they opt out or until the "greed point" is rolled, whichever occurs first. The "greed point" determines the end of a round. Any player "in the game" when the greed point is rolled lose all points earned for that round.

The Game of Greed

- ☑ The teacher should establish a "greed point" that will end the round. For example, if a "2" or "9" is rolled, the round ends and all players in the game lose their points for that round.
- ☑ Everybody stands up. The teacher rolls two dice. The result is the starting score for all students.
- ☑ If a student is satisfied with the score, they can opt out and keep those points for their total.
- ☑ If a student wishes to continue earning points, they can stay in the game for another roll.
- ☑ Continue rolling. After each roll, students may opt out and keep their points or stay in the game to earn more points for the round.
- ☑ The round ends when all students opt out or when the greed point is rolled.

Use the following table to track your scores. Add additional lines if necessary. Add all 4 round totals to determine your game score.

Roll	Score	Roll	Score	Roll	Score	Roll	Score
1		1		1		1	
2		2		2		2	
3		3		3		3	
4		4		4		4	
5		5		5		5	
6		6		6		6	
7		7		7		7	
8		8		8		8	
9		9		9		9	
10		10		10		10	
11		11		11		11	
12		12		12		12	
13		13		13		13	
14		14		14		14	
15		15		15		15	
Total		**Total**		**Total**		**Total**	
						Game Total	

We learned in Chapter 6 that rolling dice is a random phenomenon. While we can't predict *exactly* what will come up on a particular roll, we can be reasonably sure what the distribution of sums will look like for a long series of rolls.

Your task is to use what you know about rolling dice to create a strategy for the game of greed.

7.1: Discrete Random Variables

When we roll two dice, we can define the sum to be a random variable, *X*. *X* can take on any value from 2 through 12. Since we don't know exactly what sum will appear on a given roll, we call *X* a *random variable*.

Random Variable:

Discrete Random Variable:

Probability Distribution of a Discrete Random Variable:

Consider rolling two dice. Define *X* to be the sum of the two dice. Construct the probability distribution of *X* and display it with a **probability histogram**.

Consider flipping a coin 4 times and recording H or T. Define *X* to be the number of Heads flipped. Construct the probability distribution of *X* and use it to answer the following questions:

P(X=0)=

P(X=0 or 1)=

P(X>2)=

P(X≤3)=

P(X≥1)=

7.1: Continuous Random Variables

Rolling dice and flipping coins result in random variables whose outcomes are *countable*. Some situations result in outcomes that can take on any value over a given interval.

Continuous Random Variable:

Probability Distribution of a Continuous Random Variable:

According to a recent AP poll, approximately 40% of American adults indicated they used the internet to get news and information about political candidates. Suppose 40% of all American adults use this method to get their political information. What would happen if you randomly sampled a group of 1500 American adults and asked them if they used the internet to get this information. Define *X* to be the % of your sample that would respond that the internet was their primary source.
We will learn in chapter 9 that the distribution of *X* is approximately N(0.4, 0.01265). Use this information to sketch the probability distribution of *X* and answer the following questions:

If you conducted a survey of 1500 American Adults, what % would you expect use the internet as their primary source?

What is $P(X \geq 0.42)$?

What is $P(X \leq 0.35)$?

What is P(your result is within 5% of the *actual* % who use the internet as a primary source)?

7.1: Random Variable Practice

1. Suppose each of four randomly selected customers purchasing a monkey shaving kit at a certain store chooses either an electric (E) or a gas (G) model. Assume that these customers make their choices independently of one another and that, according to Consumer Reports, 40% of all customers select an electric model while 60% tend to choose the gas model. One possible purchase outcome for the four is EGGE. Mmmmmmm…."egge". Find the probability that at least two of the four customers will choose the electric model. Be sure to define your random variable, provide a probability distribution, and sketch a histogram of the probabilities when answering the question.

2. Accurate labeling of packaged meat is difficult because of weight decrease due to moisture loss (defined as a percentage of the package's original net weight). Suppose that moisture loss for a package of chicken breasts is N(4%, 1.1%) as suggested in the paper "Drained Weight Labeling for Meat and Poultry: An Economic Analysis of a Regulatory Proposal" (J. of Consumer Affairs (1980): 307-325). Let X denote the moisture loss for a randomly selected package.
   a) What is the probability X is at most 4.2%?
   b) What is the probability X is at least 6.9%?
   c) What is the probability moisture loss differs from the mean value by at least 1.3%
   d) Find a moisture loss value such that 90% of packages have a loss below that value.

3. Suppose a random variable is uniformly distributed between the values $3<X<7$. Sketch the density curve describing this variable's probability distribution.
   a) Find the height of the uniform distribution
   b) Find $P(X>4.5)$
   c) Find $P(3.2<X<6.4)$
   d) Find $P(X\leq 5.2)$

7.2: Mean and Variance of a Discrete Random Variable

Adapted from "Introduction to Statistics and Data Analysis," Peck, Olson, Devore

We have seen how to picture the probability distribution of random variables through the use of histograms (discrete) and density curves (continuous). However, it is also helpful to have numerical descriptions of these variables. The mean of a random variable describes where it is centered, while the variance and standard deviation describe the extent to which it spreads out about the center. Calculating these measures for discrete random variables requires the following formulas:

Mean of a Discrete Random Variable X:

The mean of a discrete random variable is also called the ______________ ______________. Why?

To calculate the mean, we must consider the probability that each outcome can occur. Since outcomes are not always equally likely, their probabilities need to be factored in when calculating the mean.

Mean of a Discrete Random Variable: $\mu_X =$

$\mu_X =$

At 1 min after birth and again at 5 min, each newborn child is given a numerical rating called an Apgar score. Possible values of this score are 0, 1, 2, ... , 9, 10. The score is determined by five factors: muscle tone, skin color, respiratory effort, strength of heartbeat, and reflex, with a high score indicating a healthy infant. Let the random variable X denote the Apgar score (at 1 min) of a randomly selected newborn infant, and suppose X has the following probability distribution:

x	0	1	2	3	4	5	6	7	8	9	10
P(x)	.002	.001	.002	.005	.02	.04	.17	.38	.25	.12	.01

Use the formula for the mean value of X. That is, find the average Apgar score that is approached as child after child is rated. Show all work:

The mean value provides only a partial summary of a probability distribution. For a more complete understanding of the distribution, we must consider the variability. This is described numerically by the variance and standard deviation.

Variance and Standard Deviation of a Discrete Random Variable X:

Recall that the variance and standard deviation involve calculations based on squared deviations from the mean. That is, each observation is a certain distance from the mean. That distance is called a deviation. To calculate the variance, the deviations are squared and averaged. However, in a random variable, we must take into account the likelihood of each outcome occurring when calculating this measure. Therefore, we multiply each squared deviation by its probability to factor in its weight.

Variance of a Discrete Random Variable: $\sigma_X^2 =$

Standard Deviation of a Discrete Random Variable: $\sigma_X =$

Consider the distribution of Apgar scores. Calculate the standard deviation for this distribution.

x	P(x)	$(x_i - \mu_X)^2 p_i$
0	.002	
1	.001	
2	.002	
3	.005	
4	.02	
5	.04	
6	.17	
7	.38	
8	.25	
9	.12	
10	.01	
		$\sum (x_i - \mu_X)^2 p_i$

Interpret the mean and standard deviation of X in the context of the problem.

7.2: Rules for Means and Variances

Often in statistics, we need to consider the sum, difference, or linear combination of multiple random variables. We can use several formulas to determine the mean and variance of these situations:

Rules for Means:

$\mu_{a+bX} =$

$\mu_{X \pm Y} =$

Rules for Variances:

$\sigma^2_{X \pm Y} =$

$\sigma^2_{a+bX} =$

Suppose the average SAT-Math score is 625 and the standard deviation of SAT-Math is 90.
The average SAT-Verbal score is 590 and the standard deviation of SAT-Verbal is 100. The composite SAT score is determine by adding the Math and Verbal portions. What is the mean and standard deviation of the composite SAT score?

Mr. Molesky and Mr. Liberty are avid golfers. Suppose Mr. Molesky's scores average 110 with a standard deviation of 10. Mr. Liberty's scores average 100 with a standard deviation of 8. Find the mean and standard deviation of the difference of their scores (Molesky - Liberty). Assuming their scores are normally distributed, find the probability that Mr. M will win on any given day.

7.2: Combining Random Variables Practice

1. Bottle caps are manufactured so that their inside diameters have a distribution that is approximately N(36mm, 1mm). The distribution of the outside diameters of bottles is approximately N(35mm, 1.2mm). If a bottle cap and a bottle are selected at random, what is the probability the cap will fit on the bottle?

2. W.J. Youden (Australian, 1900-1971) weighed many new pennies and found the distribution of weights to be approximately N(3.11g, 0.43g). What are the reasonably likely mean and standard deviation of weights of rolls of 50 pennies?

3. For each million tickets sold, the original New York Lottery awarded one $50,000 prize, nine $5000 prizes, ninety $500 prizes, and nine hundred $50 prizes.
   a) Describe the possible winnings in terms of a random variable and calculate the expected value of a single ticket.

   b) The tickets sold for 50¢ each. How much could the state of New York expect to earn for every million tickets sold?

4. Suppose the amount of propane needed to fill a customer's tank is a random variable with a mean of 318 gallons and a standard deviation of 42 gallons. Hank Hill is considering two pricing plans for propane. Plan A would charge $2 per gallon. Plan B would charge a flat rate of $50 plus $1.80 per gallon. Calculate the mean and standard deviation of the distributions of money earned under each plan. Assuming the distributions are normal, calculate the probability that Plan B would charge more than Plan A.

Random Variables AP Free-Response Problems

1. There are 4 runners on the New High School team. The team is planning to participate in a race in which each runner runs a mile. The team time is the sum of the individual times for the 4 runners. Assume that the individual times of the 4 runners are all independent of each other. The individual times, in minutes, of the runners in similar races are approximately normally distributed with the following means and standard deviations.

	Mean	Standard Deviation
Runner 1	4.9	0.15
Runner 2	4.7	0.16
Runner 3	4.5	0.14
Runner 4	4.8	0.15

a) Runner 3 thinks he can run a mile in less than 4.2 minutes in the next race. Is this likely to happen? Explain.

b) The distribution of team times is approximately normal. What are the mean and standard deviation of this distribution?

c) Suppose the team's best time to date is 18.4 minutes. What is the probability that the team will beat its best time in the next race?

2. A department supervisor is considering purchasing one of two comparable photocopy machines, A or B. Machine A costs $10,000 and Machine B costs $10,500. This department replaces photocopy machines every three years. The repair contract for Machine A costs $50 per month and covers an unlimited number of repairs. The repair contract for Machine B costs $200 per repair. Based on past performance, the distribution of the number of repairs needed over any one-year period for Machine B is shown below.

Number of Repairs	0	1	2	3
Probability	0.50	0.25	0.15	0.10

You are asked to give a recommendation based on overall cost as to which machine, A or B, along with its repair contract, should be purchased. What would your recommendation be? Give statistical justification to support your recommendation.