Today:

- Finish material on standard deviations from last time
- Section 8.4: Binomial random variables

Homework:

Chapter 8: #33, 39 (use computer)

Standard Deviation for a Discrete Random Variable

The **standard deviation** of a random variable is essentially the average distance the random variable falls from its mean over the long run.

If X is a random variable with possible values $x_1, x_2, x_3, \ldots,$ occurring with probabilities $p_1, p_2, p_3, \ldots,$ and **expected value** $E(X) = \mu$, then

$$\text{Variance of } X = V(X) = \sigma^2 = \sum (x_i - \mu)^2 p_i$$

$$\text{Standard Deviation of } X = \sigma = \sqrt{\sum (x_i - \mu)^2 p_i}$$

Example 8.13 *Stability or Excitement*

Two plans for investing \$100 – which would you choose? (Pictures on board)

Plan 1		Plan 2	
X = Net Gain	*Probability*	*Y = Net Gain*	*Probability*
\$5,000	.001	\$20	.3
\$1,000	.005	\$10	.2
\$0	.994	\$4	.5

Expected Value for each plan:

Plan 1:

$E(X) = \$5{,}000 \times (.001) + \$1{,}000 \times (.005) + \$0 \times (.994) = \$10.00$

Plan 2:

$E(Y) = \$20 \times (.3) + \$10 \times (.2) + \$4 \times (.5) = \$10.00$

Example 8.13 *Stability or Excitement (cont)*

Variability for each plan:

Plan 1			Plan 2		
$(X-\mu)^2$	p	$(X-\mu)^2p$	$(Y-\mu)^2$	p	$(Y-\mu)^2p$
$(\$5{,}000 - \$10)^2 = \$24{,}900{,}100$	.001	\$24,900.1	$(\$20 - \$10)^2 = \$100$	.3	\$30
$(\$1{,}000 - \$10)^2 = \$980{,}100$	.005	\$4,900.5	$(\$10 - \$10)^2 = 0$	.2	0
$(\$0 - \$10)^2 = 100$	.994	\$99.4	$(\$4 - \$10)^2 = \$36$	.5	\$18

Plan 1: $V(X) = \$29{,}900.00$ and $\sigma = \$172.92$

Plan 2: $V(X) = \$48.00$ and $\sigma = \$6.93$

The possible outcomes for Plan 1 are much more variable. If you wanted to ***invest cautiously***, you would choose **Plan 2**, but if you wanted to have the ***chance to gain a large amount of money***, you would choose **Plan 1**.

Notes about standard deviation

- Similar to when we used standard deviation for data in Chapter 2, it is most useful for *normal* random variables, which we will cover on Friday.
- In general, useful for comparing two random variables to see which is more spread out. *Examples*:
  - Two cities both have average yearly temperature of 65 degrees, but one has s.d of 5 degrees and the other has s.d. of 20 degrees. Which would you prefer?
  - One investment fund has average rate of return over many years of 8%, and s.d. of 2%. The other has average of 10%, but s.d. of 20%. The second one is higher on average, but is much more volatile.

Section 8.4: Binomial Random Variables

What do the following random variables have in common?

Ex 1: A fair coin is flipped 10 times,
X = number of heads.

Ex 2: Ten births are observed at a hospital,
X = number of boys, assume P(B)=.5

Ex 3: A student takes a true/false test with 10 questions,
just guessing, X = number correct

Ex 4: Suppose half of all adults think genetically modified food is unsafe. Take a random sample of 10 adults, X = number (out of the 10 polled) who think this.

What do these have in common?

Each of these random variables has the exact same probability distribution function!

For instance, in each case, $P(X = 0) = (½)^{10}$
$P(X = 1)$ is same for all of them, and so on.

In each case, X is called a *binomial random variable with n=10 and p=½.*

It is the outcome of a *binomial experiment*.

Properties of a Binomial Experiment

1. There are *n* "trials" where *n* is determined in advance. (10 Coin flips, births, T/F questions, adults polled)

2. There are *two possible outcomes* on each trial, called "success" and "failure" and denoted S and F. (Heads/tails; Boy/girl; Right/Wrong, Unsafe/not unsafe)

3. The *outcomes are independent* from one trial to the next. Knowledge of one does not help predict the next one. (True for all 4 examples.)

4. The probability of a "success" *remains the same* from one trial to the next, and this probability is denoted by p. The probability of a "failure" is $1-p$ for every trial.

Note that $n = 10$ and $p = ½$ for each example given.

NOTE: $p = ½$ is not always the case!
For example, multiple choice test with 5 choices, student is just guessing, p = 1/5.

A **binomial random variable** is defined as X = number of successes in the n trials of a binomial experiment.

Examples that are *not* binomial experiments:

1. A team plays 12 games in the season, X = number won.
   p = Probability of win does *not* stay the same
   Condition #4 does not hold.

2. Woman decides to have children until she has one girl or 4 children, whichever comes first.
Number of "trials" is not fixed in advance (Condition #1).

3. Deal a poker hand of 5 cards, X = number of aces.
Cards are drawn *without replacement* so outcomes are NOT independent (also, p changes). (Conditions #3, #4)

Once you recognize a binomial random variable, the pdf is always given by this formula (so you don't have to rely on Chapter 7 rules each time!):

Probability of exactly *k* successes:

$$\Pr(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, 2, \ldots, n.$$

Factorial notation: $n! = 1 \times 2 \times 3 \times \ldots \times (n-1) \times (n)$
$0! = 1$, by convention.

EX: n = 3, p = ½

$$\Pr(X = 2) = \frac{3!}{2!(3-2)!}(.5)^2(1-.5)^{3-2} = 3(\frac{1}{8}) = \frac{3}{8}$$

$$\Pr(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

How this formula is found (using Chapter 7 rules):

- Individual string of k successes and $(n – k)$ failures has probability $p^k(1\text{-}p)^{n\text{-}k}$

- There are $\frac{n!}{k!(n-k)!}$ possible ways to get k successes

Example: $n = 3$, $k = 1$, $\frac{n!}{k!(n-k)!} = 3$; {SFF, FSF, FFS}

Finding binomial probabilities using a computer (need to know how to find pdf and cdf):

Excel – See page 297
BINOMDIST(k,n,p,false) for the pdf
BINOMDIST(k,n,p,true) for the cdf

EX: BINOMDIST(2,3,.5,false) would give 3/8 = .375.

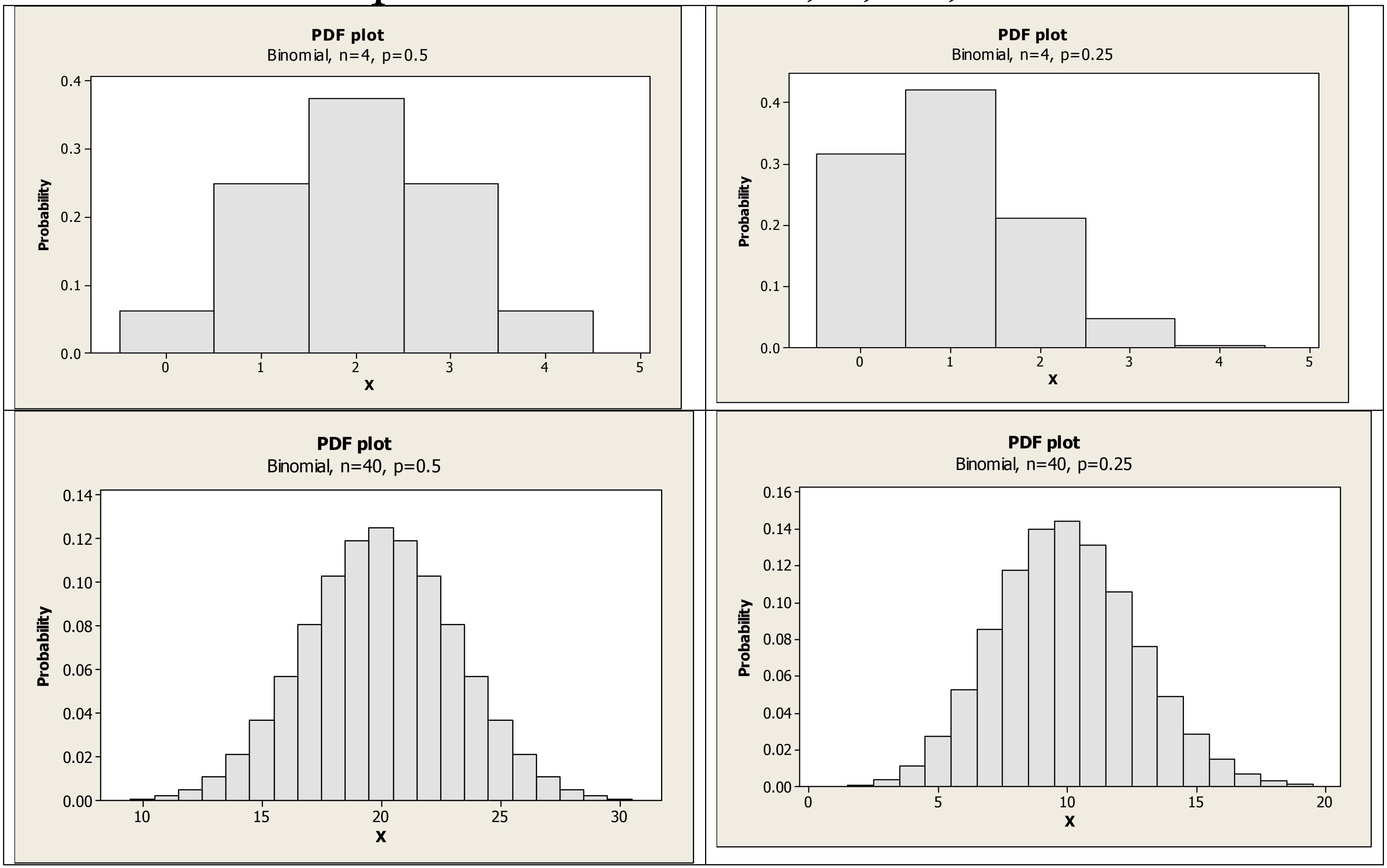
R Commander: See instructions linked to website.
Distributions →Discrete distributions → Binomial distribution → Binomial probabilities
(then fill in n and p in the popup box)

Mean and standard deviation for binomial random variables:

Mean = expected value of X = E(X) = μ = np
Variance = $\sigma^2 = np(1-p)$, so standard deviation = $\sqrt{np(1-p)}$

Example:
n = 10, p = 0.2
mean = (10)(0.2) = 2
standard deviation = $\sqrt{10(.2)(.8)} = \sqrt{1.6}$ = 1.265
(not much use for now, but will be very useful soon)

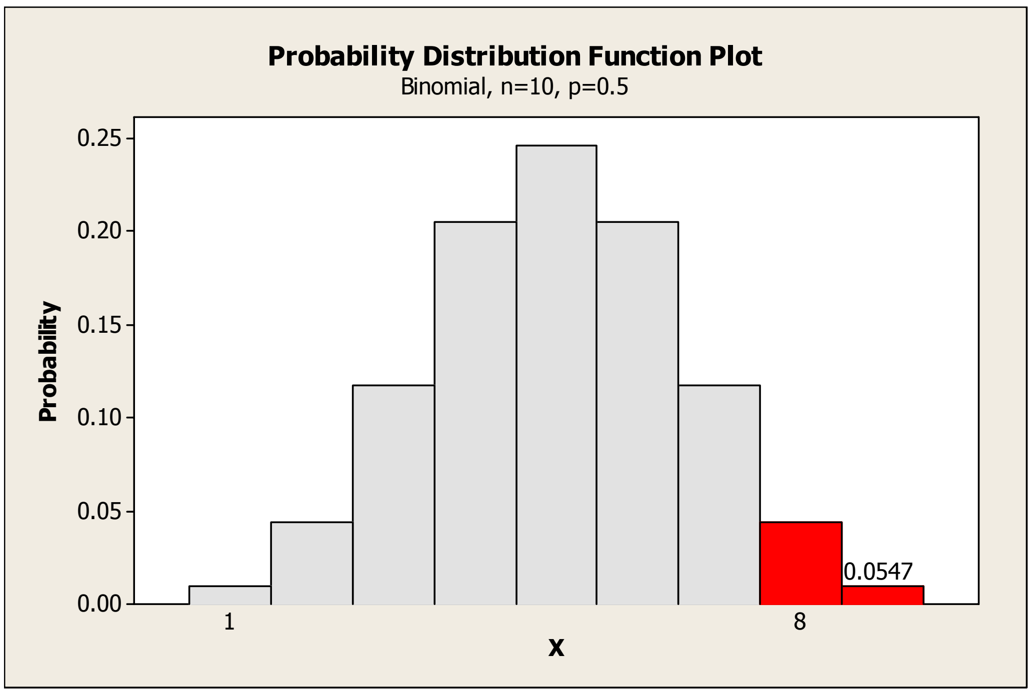
Pictures of binomial pdfs for n = 4, 40 and p = .5, .25
Expected values are 2, 1, 20, 10



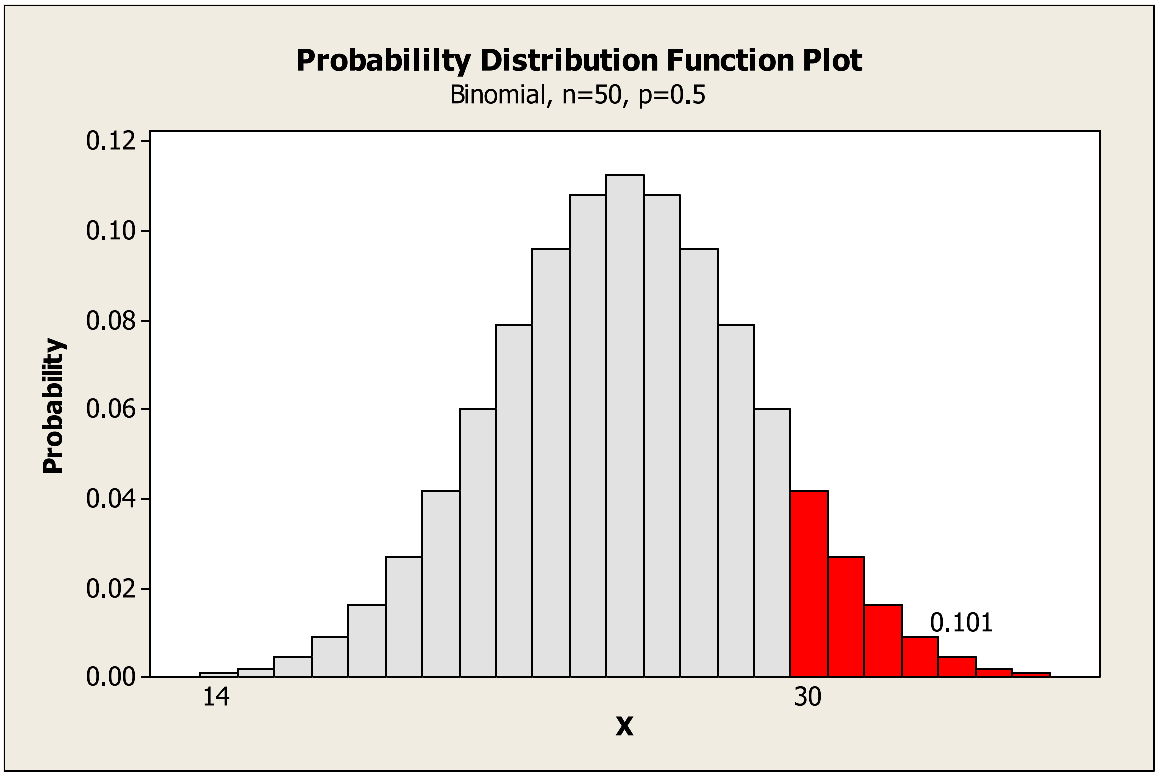
For binomial, the CDF often more interesting than PDF.

Ex: Test has 10 questions, pass if 80%, 8 or more, correct.

Find $P(X = 8, 9, 10) = P(X \geq 8) = 1 - P(X \leq 7) = 1 - \text{cdf}$ for $X = 7$, which is **1 −.94531 = .0547** (if just guessing)



Now suppose test has 50 questions, you need 60% correct to pass, so need 30 questions correct. If just guessing, $P(X \geq 30) = 1 - P(X \leq 29) = 1 - .899 = .101$ $= P(30)+P(31)+P(32)+\ldots.+P(50)$



Ex: Political poll with n = 1000. Suppose *true p = .48* in favor of a candidate.
X = number in poll who say they support the candidate.
X is a binomial random variable, n = 1000 and p = .48.

- *n* trials = 1000 people
- *"success"* = support, *"failure"* = doesn't support
- Trials are *independent*, knowing how one person answered doesn't change others probabilities
- *p* remains fixed at .48 for each random draw of a person to ask

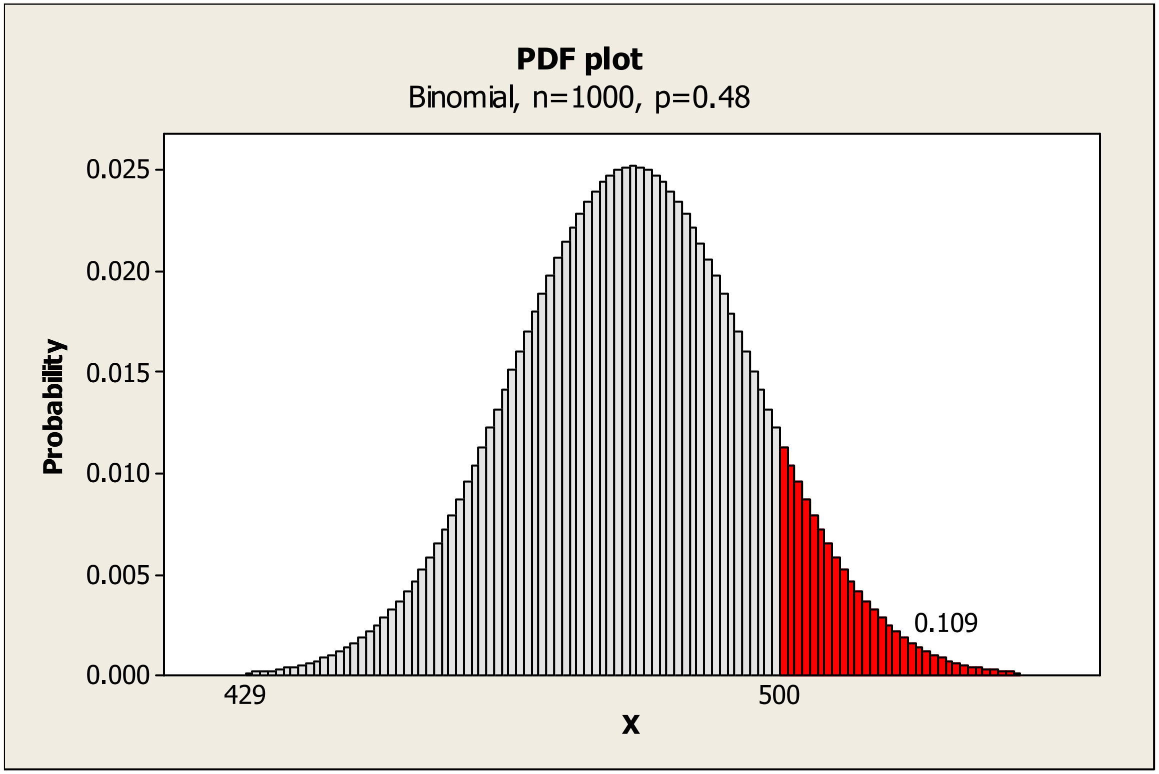
Mean = np = (1000)(.48) = **480**.

Standard dev, = $\sqrt{np(1-p)} = \sqrt{1000(.48)(.52)}$ = **15.8**

What is the probability that *at least half* of the *sample* support the candidate? (Remember only 48% of population supports him or her.)

$P(X \geq 500) = P(X = 500) + P(X = 501) + \ldots + P(X = 1000)$.

Using Excel: $1 - P(X \leq 499) = 1 - .891 = .109$.



In Section 8.7, will learn how to *approximate* this using normal curve.

Note what this says:
In polls of 1000 people in which 48% favor something, the poll will say *at least half favor it* with probability of just over .10 or in just over 10% of polls.