LESSON 6: THE BINOMIAL DISTRIBUTION

TABLE OF CONTENTS

THE FOUR CONDITIONS OF THE BINOMIAL SETTING

The binomial distribution describes the behavior of a count variable *X* if the following four conditions apply:

(1) The number of trials (or observations) *n* is fixed.
(2) Each trial (or observation) is independent.
(3) Each trial (or observation) can have one of only two possible outcomes ("success" or "failure").
(4) The probability of "success" *p* is the same for each outcome.

If these conditions are met, then *X* has a binomial distribution with parameters *n* and *p*, abbreviated *B(n,p)*.

How do we recognize a Binomial Distribution?

As the name implies, in a <u>bi</u>nomial setting, each time the researcher asks a question (or conducts a trial) there are only <u>two</u> possible outcomes: "Success" or "Failure" ("Yes" or "No;" "Red" or "not Red;" etc.).

Essentially, if the researcher's question can be rephrased as a "Yes or No" question, you can safely assume, in this course, that you are dealing with a binomial distribution. The clincher is that the probability of a "yes" must be constant. Put another way, *each* time the researcher asks the question, the probability the answer is "yes" will be the same. **We will let *p* equal the probability of a "yes".** The reason *p* will be constant is, in a binomial distribution, **each trial is *independent***. Consequently, the probability of a "yes" will not change.

If you are given a percentage (or proportion) in a problem, you have been given a value of *p*. That almost certainly indicates you have a binomial distribution. All you need now is a value for *n*.

Once you have identified a problem is binomial, IMMEDIATELY list three things:

Let n = the sample size, or number of trials (this will be given)

Let p = the probability of "success" (or "yes") on any given trial (usually given)

Let X = the *number* of "successes" (or "yeses") (this you will simply define yourself)

The sample space is always $X = 0, 1, 2, 3, \ldots n$.

If you can't list a value for n and p, then you do not have a binomial distribution!

A binomial setting is a *counting* problem. All we are doing is counting successes, X. Consequently, every time, be sure to write out the sample space for X to help you visualize the problem. $X = 0, 1, 2, 3, \ldots n$. (The values of X *always* start at 0 and count up to n the sample size.) This is because X is simply counting the number of yeses. If you ask a "yes or no" question 10 times ($n = 10$), it is possible *nobody* says yes ($X = 0$), exactly *one* person says yes ($X = 1$), exactly *two* people says yes ($X = 2$), and so on, all the way to all *ten* people saying yes ($X = 10$). But, there is no way 2.7 people would say yes, or 6.8 people. X could never equal 2.7 or 6.8.

The random variable X for a binomial distribution is a *discrete* quantitative variable. X is always counting the number of "successes." You can have exactly 0 successess, exactly 1 success, exactly 2 successes, etc., all the way to as much as exactly n successes. Always remind yourself about the sample space for X in a binomial problem:

$$X = 0, 1, 2, \ldots n$$

THREE SITUATIONS WE IMMEDIATELY KNOW ARE BINOMIAL

(1) You are **rolling a fair die** a fixed number of times and are counting how many times you get a specific number (perhaps you want to count how many times you get a "2"). This is immediately a binomial problem (since it is either "YES, you do get the number" or "NO, you don't," and the probability of rolling the number is the same each time you roll (since each roll is independent). Let's assume you were trying to roll "2s":

n = the number of rolls (this will be given)
$p = 1/6$ (since there is always a one-in-six chance of rolling a "2" every time)
X = the *number* of "2s" you roll = 0, 1, 2, … n

(2) You are **tossing a fair coin** a specific number of times and counting how many times you get "heads" (or "tails", whichever). This is immediately a binomial problem (since it is either "YES, you do get heads" or "NO, you don't," and the probability of getting "heads" is the same each time you toss (since each toss is independent).

n = the number of tosses (this will be given)
$p = 1/2 = .5$ (since there is always a 50/50 chance of "heads" on a fair coin)
X = the *number* of "heads" you toss = 0, 1, 2, … n

(3) You are **guessing on a test**. This is immediately a binomial problem (since it is either "YES, you do guess right" or "NO, you don't"), and the probability of guessing right is the same for each question (since each guess is independent). If it is a True/False test, then you have a 50/50 chance of guessing a question right ($p = 0.5$). If it is a multiple-choice test with *four* choices in each question, then you have a one-in-four chance of guessing a question right ($p = 1/4 = 0.25$). **In general, if each question has k choices, then $p = 1/k$.**

n = the number of questions on the test (this will be given)
$p = 1/k$ (since there are k choices and only one is correct)
X = the *number* of correct guesses= 0, 1, 2, … n

Make sure you have memorized these three situations and can immediately list the values of n, p and X. That is what is so crucial about these three: they will not tell you the question is binomial, and they will expect you to be able to establish the value of p yourself. For all other binomial questions, you can bet you will be given a percentage, p, tipping you off it is binomial.

BINOMIAL NOTATION

The parameters of the binomial distribution are n and p. (Memorize that fact.) If X has a binomial distribution, a prof will often summarize that fact as $\boldsymbol{X \sim B(n, p)}$. For example, if I know X has a binomial distribution with parameters $n = 10$ and $p = .6$, I could write: $\boldsymbol{X \sim B(10, .6)}$.

Computing a Binomial Probability

In a binomial probability question make sure that you list the sample space ($X = 0, 1, 2, \dots n$), then box in the specific value (or values) of X the question is interested in. Let's call that value k ($X = 0, 1, 2, \dots \boxed{k} \dots n$). Then:

$$\boldsymbol{P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}}$$

$\binom{n}{k}$ is pronounced "n choose k" and computes the number of **combinations** of k successes you can get from n trials. For example, $\binom{5}{2}$ computes the number of ways you can have 2 successes from 5 trials. To compute this on your calculator you will use the "$\boxed{\text{nCr}}$" button. Most calculators have "$\boxed{\text{nCr}}$" (often above a button where you will need to use "2nd F" or "Shift" to access it). $\binom{5}{2}$ tells you to type 5, then press the "$\boxed{\text{nCr}}$" button, then type 2 ("5 $\boxed{\text{nCr}}$ 2"). If you have done this correctly on your calculator, you will get an answer of 10: $\binom{5}{2} = \mathbf{10}$.

If you can't find "$\boxed{\text{nCr}}$" on your calculator, or can find it but can't seem to get it to work properly, I bet you have a Texas Instruments calculator. TI calculators seem to have the motto: "Why make things straightforward?" If you can't find the "$\boxed{\text{nCr}}$" button at all, do you perhaps have a "$\boxed{\text{PRB}}$" button? $\boxed{\text{PRB}}$ stands for "Probability" and, if you press that button, you will be given a menu that shows "nCr", among other things. To get the "nCr" use your arrow buttons to underline it, then press "equals". So, to compute "5 $\boxed{\text{nCr}}$ 2", you would type 5, then go to the "$\boxed{\text{PRB}}$" menu to get "nCr", and then type in 2. Again, you should get the answer of 10, if things were computed properly.

Or, if you do have an "nCr" button, but simply pressing "5 $\boxed{\text{nCr}}$ 2" does not seem to work (you don't get 10), then you will have to use a button like this: "$\boxed{x \rightleftarrows y}$" (it shows an *x* and *y* pointing at each other). This button is used to have the calculator hold one number temporarily while you feed in another one. The TI-36X is a model of calculator that computes "nCr" this way. To compute $\binom{5}{2}$ or 5 $\boxed{\text{nCr}}$ 2 using this method, you would type in 5, then press the "$\boxed{x \rightleftarrows y}$" button; now type in 2, then press the "$\boxed{\text{nCr}}$" button. Again, if done correctly, the answer should be 10.

For those of you who do not have "nCr" at all, you can work it out from scratch. (But really, why don't you just go and buy yourself a better calculator? If it doesn't have "$\boxed{\text{nCr}}$" then it probably doesn't have Stat modes either, and is making your life in this course far too complicated.) This requires the use of the *factorial* function which is denoted by an exclamation mark "!". For example, "5!" reads "5 factorial". A factorial tells you to count down from the given number all the way to 1 and multiply all the numbers together:

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = \mathbf{120}$$

$$7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = \mathbf{5040}$$

(Note: $\mathbf{0! = 1}$, by definition, and $\mathbf{1! = 1}$ as well)

You do not have to compute a factorial by hand. *Every* calculator has a factorial button (not counting those free ones you get from the bank). (It will be denoted "$\boxed{n!}$" or "$\boxed{x!}$" and is usually accessed by the "2nd F" or "Shift" button, or it will be in the "$\boxed{\text{PRB}}$" menu for you Texas Instruments people.) For example, to compute 5!, you will simply type 5, then press the factorial button "$\boxed{n!}$"on your calculator and, after you press equals, the answer should be 120.

You can compute $\binom{n}{k}$ using factorials: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

For example: $\binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5!}{2!3!}$. To compute this on your calculator, divide 5! by <u>both</u> 2! and 3!. You will type in 5, then press your "factorial" button, then press your "divide by" button, then type 2 and press "factorial", then press "divide by" again, and type in 3 and press "factorial", and finally press "equals".

$$\binom{5}{2} = \frac{5!}{2!3!} = 5! \div 2! \div 3! = \mathbf{10}$$

$$\binom{49}{6} = \frac{49!}{6!43!} = 49! \div 6! \div 43! = \mathbf{13{,}983{,}816}$$

(By the way, $\binom{49}{6}$ is computing all the possible ways you could get exactly 6 "yeses" from 49 trials, which is also counting all the different combinations of 6 numbers you could choose from 49 numbers. Which is to say, it is counting all the possible combinations there are for the Lotto 6/49. Naturally, it would be quicker to compute this using the "nCr" button:

$$\binom{49}{6} = 49 \ \boxed{\text{nCr}} \ 6 = \mathbf{13{,}983{,}816}$$

For example, let's say we have a binomial problem where $\boldsymbol{n = 15}$ and $\boldsymbol{p = .4}$. So, we can list our sample space as $X = 0, 1, 2, 3, \ldots 15$. (In prof-talk, we could say $X \sim B(15, .4)$). Finally, let's say we have been asked to find $\boldsymbol{P(X = 6)}$. Which is to say, we would box in the number 6 in our sample space, like so: $X = 0, 1, 2, 3, \ldots \boxed{6} \ldots 15$. $\boldsymbol{k = 6}$.

We can use our formula to compute this probability:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$P(X=6)=\binom{15}{6}(.4)^6(1-.4)^{\overbrace{\cancel{15-6}}^{9}}=\mathbf{.2066}$$

DON'T BE A WIMP! Be efficient and compute this probability all in one step in your calculator. Then there is nothing to fear when you do a binomial problem. If you don't know that 1 – .4 is .6 in your head, don't even do that. Leave it for the calculator. The only thing you must be sure to compute is the $n - k$ power. In this case, be sure to compute 15 – 6 and write the power as 9.

Recall, as I showed you back in Lesson 5, use the "$\boxed{y^x}$" button (or "$\boxed{x^y}$" button or "$\boxed{\wedge}$" button) to raise .4 to the power of 6 and to raise (1 – .4) to the power of 9. Also, be sure to use your open and close bracket buttons on your calculator where I show the brackets in the formula. Thus, if I were computing the above probability in my calculator, I would do this (the words I enclose in square brackets "[ ]" and boldface are telling you to press the appropriate button or buttons on your calculator to perform that function):

[open bracket] 15 **[nCr]** 6 **[close bracket] [times] [open bracket]** .4 **[close bracket]**
[raised to the power of] 6 **[times] [open bracket]** 1 **[subtract]** .4 **[close bracket]**
[raised to the power of] 9 **[equals]**

And your calculator should give you the answer .2065976… = **.2066**

This may look a little scary, but all I am doing is typing into the calculator exactly what I have written on paper, except I am being careful to include the "times" signs that are understood. Feel free to write $P(X=6)=\binom{15}{6}\times(.4)^6\times(1-.4)^9$ if you want to make sure you don't forget those "times" signs.

1. Thirty-five percent of the voters in the last election voted Liberal. If you randomly selected ten voters from the last election, what is the probability exactly four of them voted Liberal?

We have been given a *percentage* which means we have been given a *p*! We suspect this is a **binomial distribution**. (Either it is "*yes*, the person voted Liberal" or "*no*, they did not", and each person's probability they did vote Liberal is p **= 35%**.) Then they give us an "*n*" (n **= 10**). List n, p and X for this problem:

n **= 10** voters selected

p **= 35% = .35** voted Liberal

$X = \text{the } \textit{number} \text{ who voted Liberal} = 0, 1, 2, 3, \underset{k}{\boxed{4}}, \ldots 10$

In prof-talk: $X \sim B(10, .35)$. I boxed in the number 4 in my X sample space because we are asked to find the probability *exactly* 4 people voted Liberal. Find $P(X = 4)$.

Use the binomial probability formula:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$P(X=4)=\binom{10}{4}(.35)^4(1-.35)^{10-4}=\binom{10}{4}\times(.35)^4\times(1-.35)^6=\mathbf{.2377}$$

Solution to Question 1

There is a .2377 probability exactly four of them voted Liberal.

Here is how you would type all that into your calculator:

[open bracket] 10 **[nCr]** 4 **[close bracket] [times] [open bracket]** .35 **[close bracket] [raised to the power of]** 4 **[times] [open bracket]** 1 **[subtract]** .35 **[close bracket] [raised to the power of]** 6 **[equals]**

And your calculator should give you the answer .237668… = **.2377**

2. A die is rolled seven times.

(a) What is the probability we roll a three four times?

(A) 0.0156 (B) 0.2857 (C) 0.4286 (D) 0.5714 (E) 0.8988

We are **rolling a die** a specific number of times (seven times in this case) and counting how many times we get a "three". This is a **binomial distribution**. (Either it is "*yes*, we get a three" or "*no*, we do not", and, each roll, the probability of getting a three is the same: 1/6.) List n, p and X for this problem:

$n = 7$ rolls of the die

$p = 1/6$ chance of a three

X = the *number* of threes $= 0, 1, 2, 3, \underset{k}{\boxed{4}}, 5, 6, 7$

In prof-talk: $X \sim B(7, 1/6)$. I boxed in the number 4 in my X sample space because we are asked to find the probability of rolling a three *four* times. Find $P(X = 4)$.

Use the binomial probability formula:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$P(X=4)=\binom{7}{4}\left(\frac{1}{6}\right)^4\left(1-\frac{1}{6}\right)^{7-4}=\binom{7}{4}\times\left(\frac{1}{6}\right)^4\times\left(1-\frac{1}{6}\right)^3=\mathbf{.0156}$$

Solution to Question 2(a)

There is a .0156 probability we roll a three four times.
The correct answer is (A).

Here is how you would type all that into your calculator:

[open bracket] 7 **[nCr]** 4 **[close bracket] [times] [open bracket]** 1 **[divide by]** 6 **[close bracket] [raised to the power of]** 4 **[times] [open bracket]** 1 **[subtract]** 1 **[divide by]** 6 **[close bracket] [raised to the power of]** 3 **[equals]**

And your calculator should give you the answer .0156285… = **.0156**

2. A die is rolled seven times.

(b) What is the probability you get at least one 5?

(A) 0.2791 (B) 0.6093 (C) 0.7209 (D) 0.3907 (E) 1

Obviously, this is still a **binomial distribution**. (Either it is "*yes*, we get a five" or "*no*, we do not", and, each roll, the probability of getting a five is the same: 1/6.) List n, p and X for this problem:

$\boldsymbol{n} = \mathbf{7}$ rolls of the die

$\boldsymbol{p} = \mathbf{1/6}$ chance of a five

$\boldsymbol{X}$ = the *number* of fives= 0, $\boxed{\mathbf{1, 2, 3, 4 , 5, 6, 7}}$

I have boxed in from 1 to 7 in my X sample space because we are asked to find the probability of *at least one* five, meaning we want one *or more* fives. We could get 1 five, 2 fives, 3 fives, … all the way to 7 fives (we can't get more than 7 fives because we are only rolling the die 7 times). Find $\boldsymbol{P(X \geq 1)}$.

The binomial probability formula can only compute the probability of one x-value at a time. That means we would have to compute $P(X = 1)$, then $P(X = 2)$, then $P(X = 3)$, all the way to $P(X = 7)$ separately, and *add* all the results together. This would obviously be a bit tedious.

However, we know **the probability of the entire sample space must be exactly 1** (as is the case for all probability models, as we discussed in Lesson 5). **Specifically, in a binomial distribution, we know:**

$$\boxed{P(X = 0) + P(X = 1) + P(X = 2) + \ldots + P(X = n) = 1.}$$

Consequently, if the probability we desire has caused us to box in a large part of the sample space, rather than directly compute the probability, we can compute the probability of the *complement* instead. Then we can merely subtract from 1 to get the answer. Put another way, if there are a lot of x-values inside the box we have drawn, we can compute the probability of the x-values *outside* the box and subtract from 1.

P(*x*-values *inside* the box) = 1 – *P*(*x*-values *outside* the box)

In this problem, we have boxed in from 1 to 7: $X = 0, \boxed{1, 2, 3, 4, 5, 6, 7}$. That means the only number *outside* the box is 0. (Only rolling no fives at all would not satisfy the desire for *at least one* five; $X = 0$ is the complement of the question). We can work out $P(X = 0)$ instead, then subtract that result from 1 to get $P(X \geq 1)$.

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$P(X=0)=\binom{7}{0}\left(\frac{1}{6}\right)^0\left(1-\frac{1}{6}\right)^{7-0}=\binom{7}{0}\times\left(\frac{1}{6}\right)^0\times\left(1-\frac{1}{6}\right)^7=\mathbf{.2791}$$

Make sure, when you are typing this into your calculator, that you type in the "0" values when called for. Students often make the mistake of saying, "0 is nothing anyway, so I won't bother to type that in." WRONG! If you don't type in the power of 0, for example, you will not get the correct answer. (Because any number raised to the power of 0 is 1; if you type $\frac{1}{6}$ into your calculator and don't raise it to the power of 0, you are multiplying by one-sixth when you should really be multiplying by 1 since one-sixth to the power of 0 is 1.)

Be careful! We aren't finished yet. We have found the probability of 0 fives is .2791, but we were asked to find the probability of *at least one* five.

P(at least one five) = 1 – .2791 = **.7209**.

Put another way, $\mathbf{P(X \geq 1) = 1 - P(X = 0) = 1 - .2791 = .7209}$.

Solution to Question 2(b)

There is a .7209 probability we roll at least one five.
The correct answer is (C).

3. A student is writing a multiple-choice Statistics exam. Each question has 5 choices and only one choice is correct. There are a total of 20 questions on the exam. If the student is simply guessing on every single question:

(a) What is the probability he just barely passes the exam (i.e. gets exactly 50%)?

(A) 0.5000 (B) 0.1762 (C) 0.0026 (D) 0.0020 (E) 0.0002

(b) What is the probability he passes the exam?

(A) 0.5000 (B) 0.1762 (C) 0.0026 (D) 0.0020 (E) 0.0002

We are **guessing on a test**. This is a **binomial distribution**. (Either it is "*yes*, we are correct" or "*no*, we are not", and each question the probability of guessing correctly is the same: one-in-five, since there are five choices.) List n, p and X for this problem:

$\boldsymbol{n} = \mathbf{20}$ questions

$\boldsymbol{p} = \mathbf{1/5 = .2}$ chance of guessing correctly

$\boldsymbol{X}$ = the *number* of correct guesses= 0, 1, 2, 3, … $\boxed{10}$, 11, … 20

We are asked to find the probability of getting exactly 50%. In a 20 question exam, 10 correct is 50%; that's why I boxed in $\boldsymbol{k} = \mathbf{10}$ above. Find $\boldsymbol{P(X = 10)}$.

Use the binomial probability formula:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$P(X=10)=\binom{20}{10}(.2)^{10}(1-.2)^{20-10}=\binom{20}{10}\times(.2)^{10}\times(1-.2)^{10}=\mathbf{.0020}$$

Solution to Question 3(a)

There is a .0020 probability he just barely passes the exam.
The correct answer is (D).

3. (b) What is the probability he passes the exam?

(A) 0.5000 (B) 0.1762 (C) 0.0026 (D) 0.0020 (E) 0.0002

Clearly, this is still a binomial question, but note the difference between (a) and (b). In (a), we were asked to simply find the probability of *barely* passing the exam (getting *exactly* 50%). In (b), we want the probability he passes the exam. That means he could get 10 correct, or 11, or 12, all the way to all 20 questions correct (if he got all 20 correct, he would certainly pass the exam with a mark of 100%). List n, p and X for this problem:

$\boldsymbol{n = 20}$ questions

$\boldsymbol{p = 1/5 = .2}$ chance of guessing correctly

$\boldsymbol{X}$ = the *number* of correct guesses= 0, 1, 2, 3, … 9, $\boxed{\textbf{10 , 11, … 20}}$

I have boxed in from 10 to 20 in my X sample space because we are asked to find the probability of passing the exam or getting *at least* 10 correct. Find $\boldsymbol{P(X \geq 10)}$.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$P(X = 10) = \mathbf{.0020} \text{ (found in part (a))}$$

$$P(X = 11) = \binom{20}{11} \times (.2)^{11} \times (1-.2)^9 = \mathbf{.0005}$$

$$P(X = 12) = \binom{20}{12} \times (.2)^{12} \times (1-.2)^8 = \mathbf{.0001}$$

$$P(X = 13) = \binom{20}{13} \times (.2)^{13} \times (1-.2)^7 = \mathbf{.0000}$$

Since $P(X = 13)$ rounds off to 0.0000, we can bet $P(X = 14)$, $P(X = 15)$ … all the way up to $P(X = 20)$ also round off to 0.0000 since we can see the probabilities are getting smaller each time.

$P(X \geq 10) = .0020 + .0005 + .0001 + .0000 = \mathbf{.0026}$.

Solution to Question 3(b)

There is a .0026 probability he passes the exam.
The correct answer is (C).

4. A seed company has determined its seeds have a 90% chance of germinating. If 20 seeds are planted what is the probability more than 18 will germinate?

(A) 0.270 (B) 0.285 (C) 0.392 (D) 0.608 (E) 0.715

We have been given a *percentage* which means we have been given a *p*. This is a **binomial distribution**. (Either it is "*yes*, the seed will germinate" or "*no*, it will not", and each seed's probability that it germinates is 90%.) List *n*, *p* and *X* for this problem:

n = **20** seeds planted

p = **90% = .9** chance of germinating

X = the *number* of seeds that germinate= 0, 1, 2, 3, … 18, $\boxed{\mathbf{19, 20}}$

I have boxed in 19 and 20 in my *X* sample space because we are asked to find the probability more than 18 will germinate, so that does not include 18 itself. Binomial distributions are discrete, so $\boldsymbol{P(X > 18) = P(X \geq 19)}$.

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$\left.\begin{aligned} P(X=19)&=\binom{20}{19}\times(.9)^{19}\times(1-.9)^{1}=\mathbf{.2702}\\ P(X=20)&=\binom{20}{20}\times(.9)^{20}\times(1-.9)^{0}=\mathbf{.1216}\end{aligned}\right\}\text{ add these together}$$

$P(X > 18) = P(X \geq 19) = .2702 + .1216 = .3918 = \mathbf{.392}$.

Solution to Question 4

There is a .392 probability more than 18 seeds will germinate.
The correct answer is (C).

5. It is known 75% of the executives at a major multinational corporation are male. In a random sample of 8 executives from this corporation, what is the probability 3 or 4 of them are female?

(A) 0.2942 (B) 0.0865 (C) 0.2076 (D) 0.0231 (E) 0.1096

We have been given a *percentage* which means we have been given a *p*. This is a **binomial distribution**. READ CAREFULLY! We were told 75% of the executives are male, but the question wants to know the probability of 3 or 4 females. Since the question is asking for females, **let *X* = the number of females**. *X* and *p* should always be talking about the same thing, so ***p* should be the percentage of females**. This is no problem. If 75% are male, then we know 25% are female (the complementary percentage).

Always make sure *p* and *X* are counting the same thing. Let the question guide you as to what *X* is counting. If *X* is counting females, then *p* must be the percentage of females. Change *p* to fit *X*; do not change *X* to fit a given *p*.

$n = 8$ executives selected

$p = 25\% = .25$ are female

X = the *number* of females= 0, 1, 2, $\boxed{3, 4}$, 5, 6, 7, 8

Find **$P(X = 3 \text{ or } 4)$**.

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$\left.\begin{aligned} P(X=3)&=\binom{8}{3}\times(.25)^3\times(1-.25)^5=\mathbf{.2076}\\ P(X=4)&=\binom{8}{4}\times(.25)^4\times(1-.25)^4=\mathbf{.0865}\end{aligned}\right\}\text{ add these together}$$

$P(X = 3 \text{ or } 4) = .2076 + .0865 =$ **.2941** which is closest to .2942 in the choices given.[*]

Solution to Question 5

There is a .2942 probability 3 or 4 of the executives are female.
The correct answer is (A).

---

[*] The only reason we did not get exactly 0.2942 is because we were rounding off to four decimal places. If we had kept five or more decimal places during our calculations, our answer would have rounded off to exactly 0.2942.

6. An airline determines 97% of the people who booked a flight actually show up in time to take their seat. Assuming this is true, what is the probability, in a randomly selected sample of 12 people who independently booked various flights, no more than 10 of them showed up?

We have been given a *percentage* which means we have been given a p. That means this is a **binomial distribution**. (Either it is "*yes*, the people show up" or "*no*, they don't", and each person's probability of showing up is 97%.) List n, p and X for this problem:

n **= 12** people who booked flights

p **= 97% = .97** chance they will show up

X = the *number* who show up = $\boxed{\mathbf{0, 1, 2, \ldots 10,}}$ 11, 12

<u>No more than</u> 10 will show up means 10 or less. Find $P(X \le 10)$.

There is no way we are going to compute $P(X = 0)$ all the way up to $P(X = 10)$ and add them all up. Instead, we will compute the complement instead. Recall:

***P*(*x*-values *inside* the box) = 1 – *P*(*x*-values *outside* the box)**

We will compute $P(X = 11)$ and $P(X = 12)$, then subtract that answer from 1.

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$\left.\begin{aligned} P(X=11)&=\binom{12}{11}\times(.97)^{11}\times(1-.97)^{1}=\mathbf{.2575}\\ P(X=12)&=\binom{12}{12}\times(.97)^{12}\times(1-.97)^{0}=\mathbf{.6938}\end{aligned}\right\}\text{ add these together}$$

$P(X = 11 \text{ or } 12) = .2575 + .6938 = \mathbf{.9513}$.

Therefore, $\mathbf{P(X \le 10) = 1 - .9513 = .0487}$.

Solution to Question 6

There is a .0487 probability no more than 10 will show up.

7. Determine whether the variable X has a binomial distribution in each of the following cases. If it does, explain why and determine the values of the parameters n and p (or define the parameters n and p if the actual value is unknown). If X does not have a binomial distribution, explain why not.
   - (a) You randomly select ten cards from a deck of playing cards, with replacement.
     X = number of "Diamonds" selected.
   - (b) You randomly select ten cards from a deck of playing cards, with replacement.
     X = number of "Queens" selected.
   - (c) You randomly select ten cards from a deck of playing cards, without replacement.
     X = number of "Clubs" selected.
   - (d) You roll a fair die repeatedly.
     X = number of rolls until you get a "6".
   - (e) You randomly select one resident from each of the 30 apartments in a specific building.
     X = number of females selected.
   - (f) A boy asks the same girl to go on a date every day for a week.
     X = number of times the girl agrees to go on a date with the boy.
   - (g) An airplane carrying 200 passengers has two emergency exits, one at the front of the plane and one over the wings. The plane makes an emergency landing.
     X = number of passengers that exit the plane through the front emergency exit.
   - (h) I am looking at a specific row in the random digits table. There are 100 rows on the random digit table we are given, and each row has 40 digits.
     X = number of times there is a "0" in the row I am looking at.
   - (i) Weights of free-range roasting chickens follow a normal distribution with mean 1.955 kg and standard deviation 0.325 kg. You select a random sample of 12 chickens.
     X = number of chickens that weigh less than 1.5 kilograms.
   - (j) Every morning (Monday to Friday), I buy one cup of coffee at *Tim Horton's* during their *Roll Up The Rim to Win* contest.
     X = number of times in one week I win a prize.

Remember the four conditions for a Binomial Setting as you answer each of these questions.

7. (a) You randomly select ten cards from a deck of playing cards, with replacement.

X = number of "Diamonds" selected.

This is binomial. You are selecting $n = 10$ cards. Each trial has only two outcomes: each card either is a "Diamond" or it is not. And, since we are sampling with replacement, the chance of selecting a diamond is the same each time. There are four suits in the deck, each with an equal number of cards (Hearts, Diamonds, Clubs, Spades; 13 cards in each suit: *A*, 2-10, *J*, *Q*, *K*) and one of them is Diamonds, so there is $p = 13/52$ **or** $1/4$ **or** 0.25[*] chance of selecting a diamond. $X \sim B(10, 0.25)$.

Solution to Question 7(a)

This distribution is binomial with $n = 10$ and $p = 0.25$.

7. (b) You randomly select ten cards from a deck of playing cards, with replacement.

X = number of "Queens" selected.

This is binomial. You are selecting $n = 10$ cards. Each trial has only two outcomes: each card either is a "Queen" or it is not. And, since we are sampling with replacement, the chance of selecting a queen is the same each time. There are four queens in the deck, so there is $p = 4/52$ **or** $1/13$ **or** 0.0769 chance of selecting a queen. $X \sim B(10, 1/13)$.

Solution to Question 7(b)

This distribution is binomial with $n = 10$ and $p = 4/52 = 1/13$.[†]

---

[*] You could write whichever you prefer. You could use 13/52, 1/4 or 0.25. They wouldn't care. That's true for all of the parts to this question. You can use a fraction or decimal for p and you can reduce the fraction or leave it unreduced. Don't think that matters.

[†] I chose to use a fraction for p instead of a decimal because 1/13 is perfect; whereas 0.0769 is a rounded off answer. That's just me, though. They would accept either. If you are using a decimal though, as always, never round to less than four decimal places unless specifically instructed.

7. (c) You randomly select ten cards from a deck of playing cards, without replacement.

X = number of "Clubs" selected.

This is NOT binomial. You are selecting $\boldsymbol{n = 10}$ cards. Each trial has only two outcomes: each card either is a "Club" or it is not. BUT, since we are sampling without replacement, the chance of selecting a club will change each time. The first time, we have 13 clubs to choose from in the full 52-card deck (chance of club is 13/52). But, the second selection leaves us with only 51 cards to choose from. AND, we don't even know how many clubs remain. It DEPENDS on whether we selected a club the first time or not. There is either a 12/51 or 13/51 chance of a club. Each selection will be from a dwindling supply of cards. The value of p is not constant.

Solution to Question 7(c)

This distribution is NOT binomial. The trials are not independent, and the value of p (the chance of selecting a club) changes each trial.

7. (d) You roll a fair die repeatedly.

X = number of rolls until you get a "6".

This is NOT binomial. There are no fixed number of trials, n. We have no idea how many times we will roll the die.

Solution to Question 7(d)

This distribution is NOT binomial. There are no fixed number of rolls n.

7. (e) You randomly select one resident from each of the 30 apartments in a specific building.

X = number of females selected.

This is NOT binomial. You are selecting $n = 30$ residents. Each trial has only two outcomes: each resident either is a female, or they are not. You could even argue that the trials are independent since there is no reason that the gender of the residents in one apartment would have any relationship to the gender in another apartment. BUT, the probability of selecting a female p is NOT the same for each apartment. It DEPENDS on how many residents are in a given apartment and how many of them are female.

DON'T YOU DARE SAY THAT $p = 50\%$!!

Just because a resident is either male or female does not make it 50/50. Maybe everyone in one apartment is female (100% chance); maybe the next apartment has all males (0% chance of a female); maybe the next apartment has 5 residents and one of them is female (1/5 or 20% chance of a female).

Solution to Question 7(e)

This distribution is NOT binomial. The chance of selecting a female is not the same for each apartment. It depends on the proportion of females in the apartment, and that will vary in each apartment. The value of p is not constant for each trial.

7. (f) A boy asks the same girl to go on a date every day for a week.

X = number of times the girl agrees to go on a date with the boy.

This is NOT binomial. He is asking for a date $n = 7$ times. She will say yes, or no. BUT the trials are NOT independent. The chance of success depends on many changing factors. Does she like him? Does she have another commitment on one of those days? Etc.

Solution to Question 7(f)

This distribution is NOT binomial. Each trial is DEPENDENT.

7. (g) An airplane carrying 200 passengers has two emergency exits, one at the front of the plane and one over the wings. The plane makes an emergency landing.

X = number of passengers that exit the plane through the front emergency exit.

This is NOT binomial. There are $n = 200$ passengers. Each trial has only two outcomes: each passenger will either exit through the front door, or they will not. BUT, each trial is NOT independent. It DEPENDS on how close the passenger is to the front door. Obviously, someone near the front door is almost 100% likely to exit out the front door. But, someone near the back of the plane is much more likely to exit out the wing door, with close to a 0% chance of exiting out the front door.

Solution to Question 7(g)

This distribution is NOT binomial. The chance of exiting out the front emergency exit DEPENDS on how close the passenger is to that exit.

7. (h) I am looking at a specific row in the random digits table. There are 100 rows on the random digit table we are given, and each row has 40 digits.

X = number of times there is a "0" in the row I am looking at.

This is binomial. Since there are 40 digits in the row I am looking at, that means $n = 40$ trials (it is irrelevant that there are 100 rows since I am looking at one specific row). Each trial has only two outcomes: as I look at each digit in the row, it either is a "0" or it is not. There are ten different digits any particular digit could be (0, 1, 2, 3… 9) and only one of those digits is "0": the first digit could be 0, or it might be any of 1 through 9; the second digit could be 0, or 1 through 9; etc. up to the fortieth digit I look at. So there is $p = 1/10$ **or** 0.1 chance of any specific digit in the row being a 0. $X \sim B(40, 0.1)$.

Solution to Question 7(h)

This distribution is binomial with $n = 40$ and $p = 0.1$.

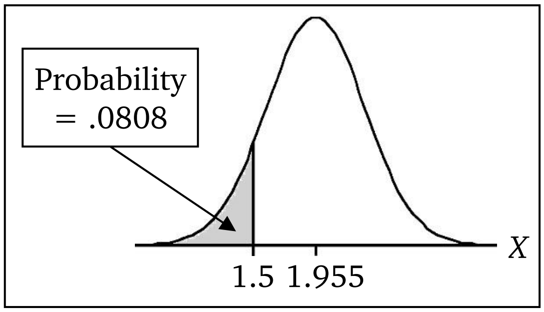
7. (i) Weights of free-range roasting chickens follow a normal distribution with mean 1.955 kg and standard deviation 0.325 kg. You select a random sample of 12 chickens.

X = number of chickens that weigh less than 1.5 kilograms.

This is binomial. You are selecting $n = 12$ chickens. Each trial has only two outcomes: each chicken either weighs less than 1.5 kg, or it does not. We can safely assume each chicken is independent of the others since we are selecting randomly. The hard part is establishing the value of p. That's where the normal distribution comes into play.

We know the weights are **normal** with $\mu = 1.955$ **kg** and $\sigma = 0.325$ **kg**. We can compute the probability that a randomly selected chicken weighs less than 1.5 kg. Use an X-Bell Curve and find $P(X < 1.5)$.

Draw an X-Bell curve centred at $\mu = 1.955$ with $\sigma = 0.325$ and shade in the region to the left of 1.5.



When $x = 1.5$:

$$z = \frac{x-\mu}{\sigma} = \frac{1.5-1.955}{0.325} = \mathbf{-1.4}$$

Consulting Table A, the area to the left of $z = -1.4$ is **.0808**.

So, we now know that, for any randomly selected chicken, there is a 0.0808 chance (or 8.08%) that the chicken ways less than 1.5 kg. That is p! There is $p = 0.0808$ chance of selecting a chicken that ways less than 1.5 kg. And that would be the chance for each of the 12 chickens we select. $X \sim B(12, 0.0808)$.

Solution to Question 7(i)

This distribution is binomial with $n = 12$ and $p = 0.0808$.

7. (j) Every morning (Monday to Friday), I buy one cup of coffee at *Tim Horton's* during their *Roll Up The Rim to Win* contest.

X = number of times in one week I win a prize.

This is binomial. You are buying $\boldsymbol{n = 5}$ cups of coffee. Each trial has only two possible outcomes: for each cup you either win a prize, or you do not. We can assume the 5 cups we get are random, so whether each cup is a winner or not is independent. And there is a fixed percentage of cups that are winners. We would have to be shown the fine print of the contest rules to know exactly what percentage that is, but it certainly exists. For example, if 10% of the cups that are produced for the contest are winners, then there is a 10% chance any randomly selected cup is a winner. We just don't have enough information to know the exact value of p since we have not been told what percentage of cups are winners. $\boldsymbol{X \sim B(5, p)}$.

Solution to Question 7(j)

This distribution is binomial with $n = 5$ and
p = whatever proportion of cups produced for the contest are winners.

8. A newspaper reports that one in three drivers routinely exceed the speed limit. Assuming this is true, we select a random sample of 30 drivers.

(a) What is the probability exactly half of them exceed the speed limit?

(b) What is the mean number of drivers in a sample of this size who routinely exceed the speed limit, and what is the standard deviation?

This question is sneaky. We have to realize the phrase "one in three" is giving us the fraction 1/3, and **a fraction is just another way of giving a percentage, and a percentage is a *p*!** Thus, we have been given **$p = 1/3$** of drivers routinely exceed the speed limit. This is a **binomial distribution**. (Either it is "*yes*, they routinely speed" or "*no*, they do not", and each driver's probability of speeding is 1/3.)

Another twist is being asked to find the probability exactly half speed. Since we have a random of sample of **$n = 30$** drivers, half is 15, so we want **$P(X = 15)$**. List *n*, *p* and *X* for this problem:

$n = 30$ drivers

$p = 1/3$ chance of being a speeder

X = the *number* who speed = 0, 1, 2, $\boxed{15}$ … 30

Find **$P(X = 15)$**.

$$P\left(X=k\right)=\binom{n}{k}p^{k}\left(1-p\right)^{n-k}$$

$$P\left(X=15\right)=\binom{30}{15}\times\left(\frac{1}{3}\right)^{15}\times\left(1-\frac{1}{3}\right)^{15}=\mathbf{.0247}$$

Solution to Question 8(a)

There is a .0247 probability exactly half
the sample of drivers routinely exceed the speed limit.

THE MEAN AND STANDARD DEVIATION OF A BINOMIAL DISTRIBUTION

Let X be a random variable from a binomial distribution with parameters n and p. Which is to say $X \sim B(n, p)$:

$$\text{The mean of } X = \mu_X = np$$

$$\text{The standard deviation of } X = \sigma_X = \sqrt{np\,(1-p)}$$

8. (b) What is the mean number of drivers in a sample of this size who routinely exceed the speed limit, and what is the standard deviation?

We already established we were given $n = 30$ randomly selected drivers, $p = 1/3$ routinely exceed the speed limit, and $X =$ **the number who routinely exceed the speed limit**; which is to say $X \sim B(30, 1/3)$.

$$\text{The mean of } X = \mu_X = np = 30 \times 1/3 = 10$$

$$\text{The standard deviation of } X = \sigma_X = \sqrt{np\,(1-p)} = \sqrt{10 \times \left(1 - \frac{1}{3}\right)}$$

We already know $np = 10$ from computing the mean.

$$\sigma_X = \sqrt{6.6666\ldots} = 2.5820$$

Solution to Question 8(b)

Thus, in a random sample of 30, the mean number of drivers who routinely exceed the speed limit is 10 with a standard deviation of 2.5820.

Which is to say, if we picked 30 drivers at random, we would expect 10 of them to be speeders, give or take 2.6 (since the standard deviation is 2.5820 ≈ 2.6).

THE DISTRIBUTION OF *X* IN A BINOMIAL DISTRIBUTION

Consider a binomial question where you were asked to find the probability X is between 50 and 90 inclusive; which is to say $P(50 \le X \le 90)$. You were given:

$n = 200$

$p = .3$

$X = 0, 1, 2, 3, \dots 49, \boxed{50, 51, \dots 89, 90}, 91, \dots 199, 200$

If we were to use the binomial probability formula to compute this probability, we would be there all day. We would have to compute $P(X = 50) + P(X = 51) + \dots + P(X = 90)$. Surely there is an easier way!

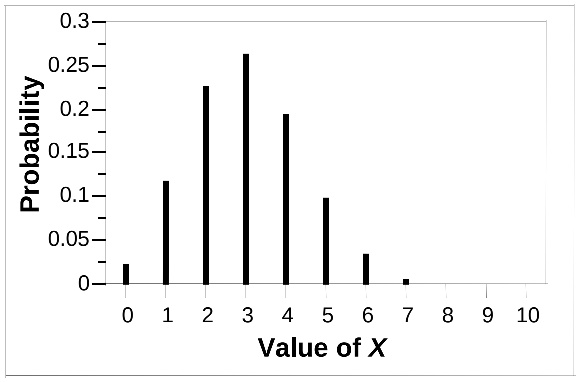
Let's take a closer look at the binomial distribution. **The binomial distribution has two *parameters*: *n* and *p*.** Changing the value of n and/or p affects the shape, centre and spread of the distribution of X. On the next couple of pages, I show you histograms of various binomial probability distributions. You will note that, instead of the usual rectangles on all the histograms, there are instead very slender lines with large gaps between them. This is because the **random variable *X* in a binomial distribution is discrete**. Consequently, we know $X = 0$, or 1, or 2, etc., but it could never take on "in-between" values like 1.7 or 2.3. Those gaps in the histograms emphasize X is *never* those kinds of values.

Essentially, each histogram shows the various values X could have together with the probability of each value. First, I show you what happens if we change p while leaving n constant. Then I show you what happens if we change n but leave p constant. Observe how the shape of the binomial distribution changes.

The binomial distribution does not have one consistent shape. The shape varies depending on the value of *n* and/or *p*. The only definite thing we can say is, if *p* = .5, the binomial distribution is symmetric.

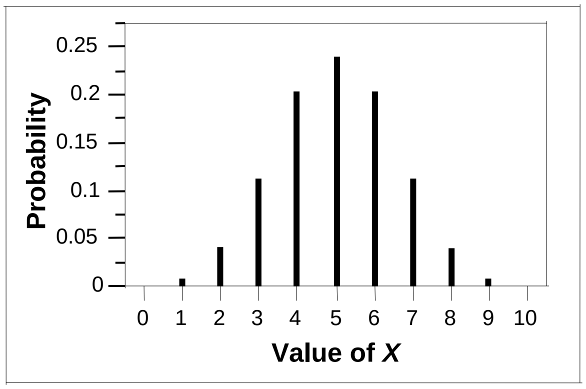
Figure 1: Binomial Distributions where $n = 10$, but p is changing.

$n = 10, p = \underline{.3}$



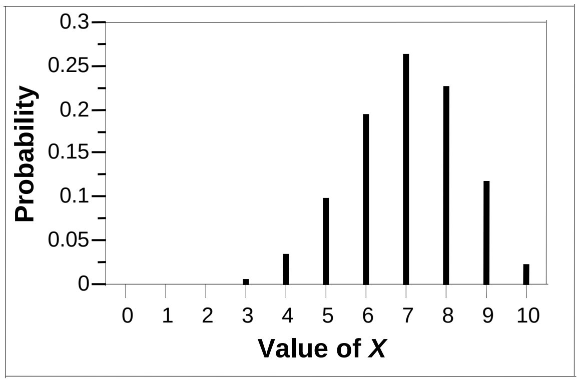
When $p < .5$, the distribution of X peaks quite early (since low values of X are more likely), making the binomial distribution right-skewed (for small values of n at least, but see the next page).

$n = 10, p = \underline{.5}$



When $p = .5$, the distribution of X peaks in the middle (since middle values are most likely).
If $p = .5$ the binomial distribution is symmetric.

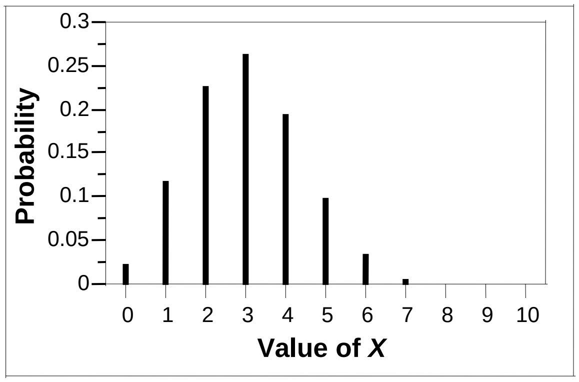
$n = 10, p = \underline{.7}$



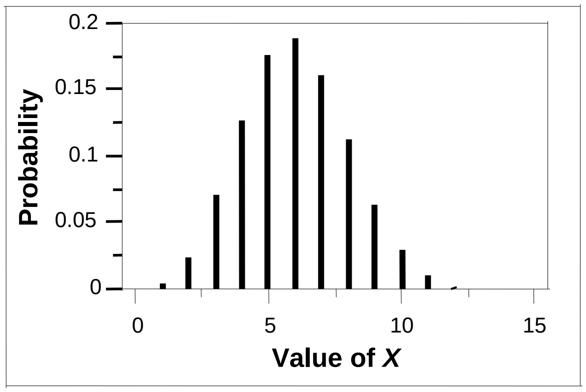
When $p > .5$, the distribution of X peaks quite late (since high values of X are more likely), making the binomial distribution left-skewed (for small values of n at least, but see the next page).

Figure 2: Binomial Distributions where $p = .3$ but n is changing.

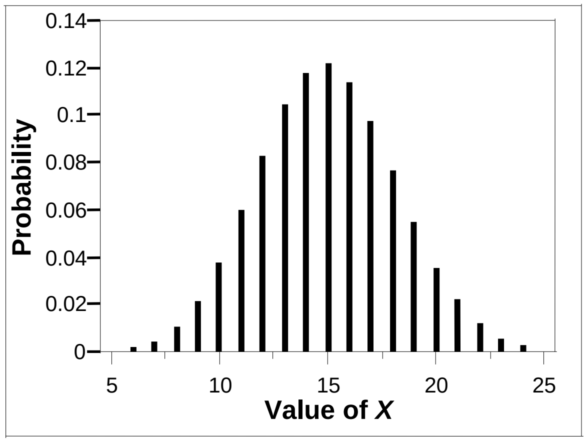
$n = \underline{10}, p = .3$



$n = \underline{20}, p = .3$



$n = \underline{50}, p = .3$



Observe how, as n changes from 10 to 20 to 50, the distribution of X with $p = .3$ starts to look more and more symmetrical and, indeed, starts to look almost bell-shaped. **As n gets large, all binomial distributions become approximately normal, regardless of the value of p. If p is near .5 (50%), n does not have to be all that large for the binomial distribution to become close to normal. However, the further away from 50% p is, the larger n has to be before the distribution starts to become approximately normal.**

Rule of Thumb:

If $np \geq 10$, and if $n(1 - p) \geq 10$, the distribution of X in a binomial distribution will be approximately normal. This means we could draw an X-bell curve to approximate binomial probabilities.

What we learn from these histograms is a binomial distribution starts to look more and more bell-shaped as n gets larger. How quickly it starts to look bell-shaped depends on the value of p. (We saw, when $p = .5$, the binomial distribution is symmetric and looks almost bell-shaped for even small values of n. When p is not .5, the binomial distribution is skewed (the further p is from .5, the stronger the skewness); then we would need a fairly large sample size before the distribution starts to look bell-shaped.

Our Rule of Thumb is:

If the population size N is at least 10 times the sample size n ($N \geq 10n$)[*] and if $np \geq 10$ and $n(1-p) \geq 10$, then the binomial random variable X has an approximately normal distribution. This means we can use an X-bell curve centred at μ_X (recall $\mu_X = np$) and can approximate binomial probabilities by changing X-scores into z-scores and using Table A.

(Note, some professors say use of z is valid if $np \geq 5$ and $n(1-p) \geq 5$.)

Recall my example gave us $n = 200$ and $p = 0.3$. Then $\mathbf{np = 200 \times .3 = 60} \geq 10$ and $\mathbf{n(1-p) = 200 \times .7 = 140} \geq 10$. Thus, by our rule of thumb, we can assume X has an approximately normal distribution. This means we can draw an X-Bell Curve and use it to solve any probabilities. However, to draw a bell curve, we also need to know μ and σ for the distribution. **We know, for the binomial distribution, the mean of $X = \mu_X = np$ and the standard deviation of $X = \sigma_X = \sqrt{np(1-p)}$.**

---

[*] Students often wonder why the population has to be at least ten times bigger than the sample. This seems counterintuitive. After all, wouldn't it be great if we had a huge sample, say a sample that was half the size of the population, instead of no larger than one-tenth the size? Well, of course, that would be great. The larger the sample the better. However, if the sample were that large, the formula we use to compute the standard deviation of X would be very inaccurate. It would actually overestimate the size of the spread. Consequently, we would actually have a much better margin of error than the formulas suggest. In these cases, statisticians introduce what is called the "finite population correction factor". You need to know nothing about this formula. Suffice to say you will never need it. As if you would ever have a sample size that is nearing one-tenth the size of the whole population anyway. You can safely assume any sample is a spit in the ocean in comparison to the population you are examining.

Example 1

Given X is a binomial random variable with $n = 200$ and $p = .3$, find the probability X is between 50 and 90, inclusive; which is to say, find $P(50 \le X \le 90)$.

Solution to Example 1

First, like all binomial problems, I list my givens and my X sample space, boxing in the relevant X-values:

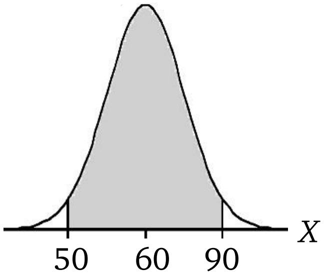
$\boldsymbol{n = 200}$

$\boldsymbol{p = .3}$

$\boldsymbol{X} = 0, 1, 2, 3, \ldots 49, \boxed{\mathbf{50, 51, \ldots 89, 90}}, 91, \ldots 199, 200$

Since it would be far too tedious to compute the sum of all these probabilities using the binomial formula, we consider using the normal approximation to the binomial distribution instead. We compute $np = 200 \times .3 = 60$ and $n(1 - p) = 200 \times (1 - .3) = 140$. They are both greater than or equal to 10 (considerably!). This confirms we are allowed to use the normal approximation.

This binomial distribution has a mean of $\mu_X = np = 200 \times .3 = 60$ ($\boldsymbol{\mu_X = 60}$) and a standard deviation of $\sigma_X = \sqrt{np(1-p)} = \sqrt{60 \times (1-.3)} = \sqrt{42} = 6.4807$ $\left(\boldsymbol{\sigma_X = \sqrt{42}}\right)$.[*] We are asked to find $P(50 \le X \le 90)$, so we can draw an X-Bell curve centred at $\mu = 60$ and shade the area between 50 and 90. That area is the *approximate* probability we are after.



Recall, as we learned back in Lesson 4, to standardize an *X*-Bell Curve, we use the formula $\boldsymbol{z = \dfrac{x - \mu}{\sigma}}$. First, we must change these 2 x-scores into z-scores:

---

[*] I prefer to use $\sqrt{42}$ for the standard deviation to maintain perfect accuracy in my later calculations. It is just as easy to type that into the calculator during calculations as 6.8407 and then there is no rounding off.

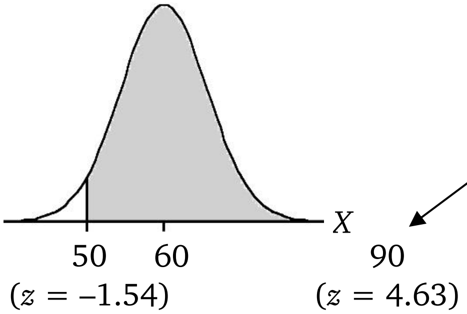
When **$x = 90$**,

$$z = \frac{x-\mu}{\sigma} = \frac{90-60}{\sqrt{42}} = 4.63$$

z = 4.63 is off the end of Table A. This tells us 90 can be considered to be way off the end of the bell curve. Redraw the diagram.

When **$x = 50$**,

$$z = \frac{x-\mu}{\sigma} = \frac{50-60}{\sqrt{42}} = -1.54$$



90 is way over here, off the end of the bell curve, so the area between 50 and 90 amounts to simply the area to the right of 50.

Consulting Table A, we see , when z = –1.54, the Left Area = .0618, so the Right Area = 1 – .0618 = **.9382**.

The approximate probability X is between 50 and 90, inclusive, is .9382. Which is to say, $P(50 \le X \le 90) = 0.9382$.

If you have boxed in a lot of X-values in a binomial probability problem, you can bet you will be allowed to use the *normal approximation*. If you are not certain, use your Rule of Thumb to determine if X has an approximately normal distribution.

However, if the random variable X in a binomial distributionis approximately normal, then it will also follow that $\hat{p}$, the sample proportion is also approximately normal. And it is more straightforward for us to use $\hat{p}$ in a normal approximation. Let's take a look at the distribution of the sample proportion, $\hat{p}$.

THE DISTRIBUTION OF THE SAMPLE PROPORTION, $\hat{p}$

The two parameters of a binomial distribution are n and p. Recall, **a percentage is a proportion is a p**. Frequently, a statistician will have to estimate the value of p. In these cases, we take a sample and compute the **sample proportion, $\hat{p}$**. The classic way you will be given $\hat{p}$ is in the form: "We took a sample of size ***n*** and counted ***x*** said yes to our question."

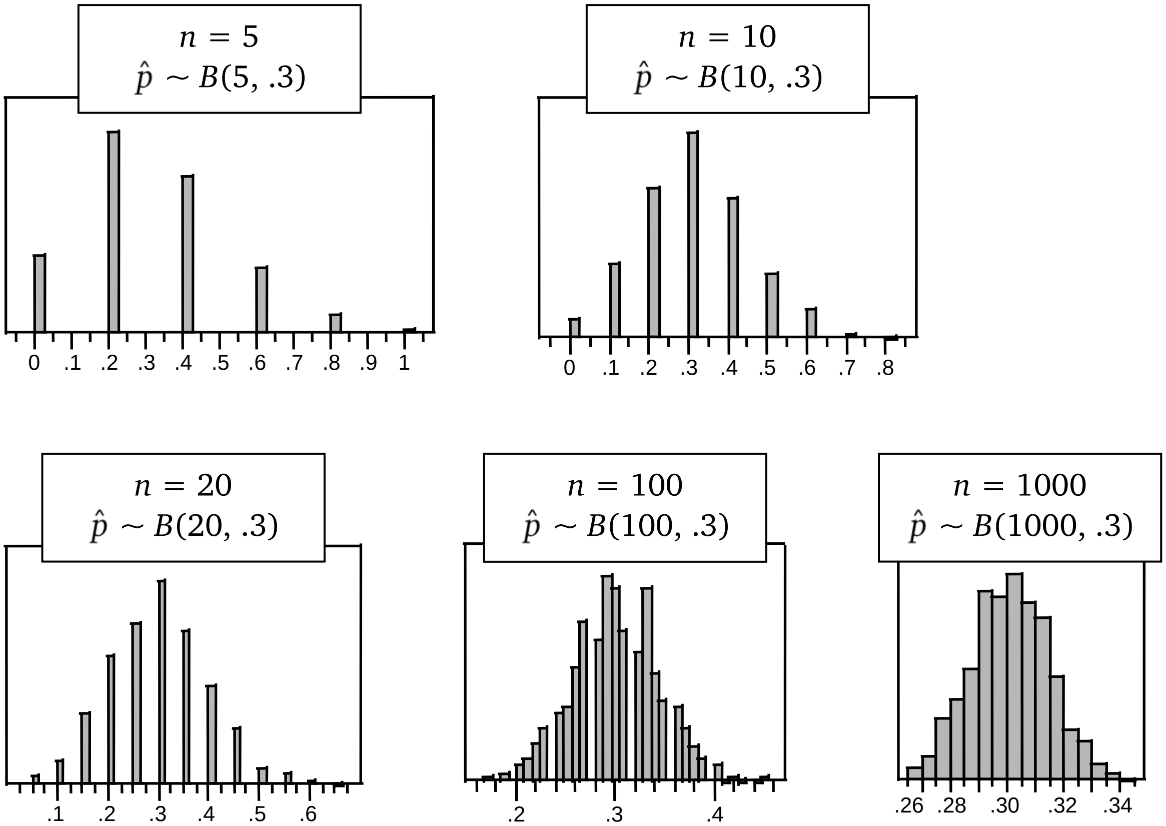
$$\hat{p} = \frac{x}{n}$$

For example: A random sample of **1000** dentists found **330** use Crest toothpaste. We have been given $n = 1000$ and $x = 330$. Therefore, $\hat{p} = \frac{x}{n} = \frac{330}{1000} = .33$. We can say 33% of our *sample* use Crest.

When we select a random sample of size n and compute $\hat{p}$, the sample proportion, that is *not* the one and only possible answer for $\hat{p}$. In the example above, my random sample of n = 1000 dentists revealed $\hat{p}$ = .33 use Crest, but, if you selected a *different* random sample of the same size, you could quite possibly get a different answer for $\hat{p}$ since your sample is not likely to contain the exact same members of the population. Your sample of 1000 dentists might have 29% = .29 who use Crest; someone else's sample might have 31% = .31 who use Crest. We could repeatedly take samples, and each time compute the sample proportion, $\hat{p}$, until we have thousands and thousands of answers for $\hat{p}$. We could then visualize the distribution of $\hat{p}$, by making a histogram showing all of these answers, and how frequent the various values are.

In Figure 3 below, I show the distribution of $\hat{p}$ drawn from a population where the true percentage was 30% (**p = 30% = 0.3**). (Let's pretend the true percentage of dentists who use Crest is 30%.) I used *JMP*™ to simulate selecting a sample and computing $\hat{p}$ one thousand separate times. First, *JMP*™ did this using a sample size of n = 5, then I did it again for n = 10, n = 20, n = 100, and, finally, n = 1000.

Figure 3: Distribution of one thousand $\hat{p}$ values from a **binomial** population where $\boldsymbol{p = .3}$ but $\boldsymbol{n}$ varies; i.e., $\hat{p} \sim B(n, .3)$.



As the sample size n gets larger, the distribution of $\hat{p}$ becomes more and more bell-shaped. It also starts to look more and more continuous (as opposed to the very obviously discrete distribution it has for small values of n). Also, as the Law of Large Numbers predicts, the spread of the distribution gets narrower and narrower as n gets larger (note how the $n = 5$ histogram spreads from 0 to 1, while the $n = 20$ histogram spreads from about .05 to .65, and the $n = 1000$ histogram spreads from about .26 to .34). It is also fair to say, as n gets large, $\hat{p}$ has an approximately normal distribution.

We see $\hat{p}$ frequently has values near .3, but the shape of the distribution and how spread out it is depends on the sample size n. The first histogram, shows the distribution of $\hat{p}$ when $n = 5$. (We could say I randomly selected 5 dentists from a binomial population where $p = .3$ and counted the number, X, who said, "Yes, I do use Crest.") X could be 0, 1, 2, 3, 4, or 5. Then, I computed $\hat{p}$. If you think about it, even though the true percentage is $p = .3$, it is impossible for $\hat{p}$ to ever be .3 exactly in a sample this small. There are only six possible answers we could get for $\hat{p}$ in a sample of size $n = 5$: $\hat{p} = \frac{x}{n} = \frac{0}{5} = 0$; $\hat{p} = \frac{x}{n} = \frac{1}{5} = .2$; $\hat{p} = \frac{2}{5} = .4$; $\hat{p} = \frac{3}{5} = .6$; $\hat{p} = \frac{4}{5} = .8$; $\hat{p} = \frac{5}{5} = 1$. But not all of these values are equally likely. Since $p = .3$, we would expect our sample to give us results near .3. Which is to say, $\hat{p} = .2$ or $\hat{p} = .4$ should occur more often than $\hat{p} = .8$ or $\hat{p} = 1$.

For $n = 5$, we see the distribution of $\hat{p}$ is right skewed and centred around .3. It is also quite clearly a discrete distribution, since $\hat{p}$ can only ever take on the six values 0, .2, .4, .6, .8, and 1, as I demonstrated above.

However, as I increase the sample size to $n = 10$, 20, 100, and 1000, each time computing one thousand random $\hat{p}$ values, we see the distribution starts to look more symmetric, and it also starts to look more and more continuous (because there are so many values $\hat{p}$ could have in a larger sample: if $n = 1000$, $\hat{p} = \frac{x}{n} = \frac{0}{1000} = 0$; $\hat{p} = \frac{x}{n} = \frac{1}{1000} = .001$; $\hat{p} = \frac{x}{n} = \frac{2}{1000} = .002$; … $\hat{p} = \frac{x}{n} = \frac{999}{1000} = .999$; $\hat{p} = \frac{x}{n} = \frac{1000}{1000} = 1$).

All the distributions stay centred at the true percentage $p = .3$, but they start to have a much narrower spread as n gets larger. This is the **Law of Large Numbers** in action.[*] As our sample size, n, gets larger, our statistic comes closer and closer to the parameter it is estimating. **Here, as n gets larger, $\hat{p}$ comes closer and closer to p.**

We can proceed to compute the mean of $\hat{p}$. Remember, when we are talking about the <u>mean</u> of $\hat{p}$, we are talking about computing $\hat{p}$ thousands and thousands of times, then averaging

---

[*] I will elaborate on the Law of Large Numbers in Lesson 7.

all of those results. It is mathematically certain, if we were able to repeat the sampling procedure endless amounts of times, and compute endless amounts of sample proportions, $\hat{p}$, **the mean of $\hat{p} = \mu_{\hat{p}} = p$.** In my dentists example above, I said $p = 30\%$, so it is irrelevant that my one sample of 1000 dentists, produced $\hat{p} = 33\%$. I know if I repeatedly took samples of 1000 dentists, and computed $\hat{p}$ each time, then the average of all those sample proportions would be 30%, the true percentage. Additionally, like all statistics $\hat{p}$ has a variability (a spread). **The standard deviation of $\hat{p} = \sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$.**

We also see, as n gets larger, $\hat{p}$ has an approximately normal distribution. How large n has to be depends on the true value of p. When $p = .5$ or something near it, the distribution of $\hat{p}$ will already be symmetric (but not continuous). Here, even if n is not very large at all, $\hat{p}$ tends to have an approximately normal distribution. If p has a value quite far away from .5, the distribution tends to be strongly skewed at first, and only when n is quite large does $\hat{p}$ start to look approximately normal. The Rule of Thumb for the approximately normal distribution of $\hat{p}$ is the same Rule of Thumb we used to assume the binomial random variable X has an approximately normal distribution.

Our Rule of Thumb is:

If the population size N is at least 10 times the sample size n ($N \geq 10n$) and if $np \geq 10$ and $n(1-p) \geq 10$, then $\hat{p}$ has an approximately normal distribution, meaning we have a $\hat{p}$-bell curve centred at p (since $\mu_{\hat{p}} = p$) and can change $\hat{p}$-scores into z scores.

(Again, some professors say use of z is valid if $np \geq 5$ and $n(1-p) \geq 5$.)

If $\hat{p}$ is normally distributed, that means we can draw a $\hat{p}$-bell curve and standardize $\hat{p}$ scores into z scores. The $\hat{p}$ standardizing formula is: $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$.

THE NORMAL APPROXIMATION TO THE BINOMIAL DISTRIBUTION

Let's summarize the key formulas:

Let X be a random variable from a binomial distribution with parameters n and p. Which is to say $X \sim B(n, p)$:

$$\text{The mean of } X = \mu_X = np$$

$$\text{The standard deviation of } X = \sigma_X = \sqrt{np\,(1-p)}$$

$$\text{The mean of } \hat{p} = \mu_{\hat{p}} = p.$$

$$\text{The standard deviation of } \hat{p} = \sigma_{\hat{p}} = \sqrt{\frac{p\,(1-p)}{n}}\,.$$

If our rule of thumb suggests $\hat{p}$ has an approximately normal distribution, we can draw a $\hat{p}$-bell curve and standardize our $\hat{p}$ score into a z-score:

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} \quad \text{which is to say} \quad z = \frac{\hat{p} - p}{\sqrt{\dfrac{p\,(1-p)}{n}}}$$

9. In Big City only 35% of the voters in the last election were in favour of a one-time levy to cover the cost of sewer upgrades. During the current campaign a random sample of 575 voters will be selected. Assume the opinion has not changed since the last election.
   (a) What is the mean and standard deviation of the number of voters sampled in favour of the levy?
   (A) 0.35; 0.020 (B) 0.65; 0.020 (C) 201.25; 11.44
   (D) 373.75; 11.44 (E) 350; 13.46
   (b) What is the mean and standard deviation of the proportion of voters sampled in favour of the levy?
   (A) 0.35; 0.020 (B) 0.65; 0.020 (C) 201.25; 11.44
   (D) 373.75; 11.44 (E) 350; 13.46
   (c) What is the probability the sample proportion will be between 30% and 40% in favour of the levy?
   (d) What is the probability a random sample of 575 voters finds at least 225 in favour of the levy?

We are given a percentage, so we are given a p. **p = 35% = .35**. This is a **binomial** distribution. List your n, p and X. We are told:

n = 575 voters

p = 35% = .35 in favour of the sewer levy

X = the number in favour of the levy = 0, 1, 2, ... 575.

At first glance, questions (a) and (b) look identical. What is the difference? The first asks for the mean and standard deviation of the number of voters sampled, while the second asks for the mean and standard deviation of the proportion of voters sampled. In binomial distributions, **we let X = the number who said "yes",** but we can also compute **$\hat{p}$ = the proportion of our sample who said "yes".**

Which mean do they want in a Binomial setting? μ_X or $\mu_{\hat{p}}$?

In binomial distribution problems, make sure you ask yourself, "Are we interested in X, the number of yeses, or $\hat{p}$, the proportion of yeses, in our sample?" Do we want to know the mean of X, μ_X? or the mean of $\hat{p}$, $\mu_{\hat{p}}$?

In that respect, NEVER MISS THE WORD "PROPORTION" IN A BINOMIAL QUESTION! That almost certainly suggests the question is talking about $\hat{p}$ stuff. If there is no mention of the word "proportion" or the symbol $\hat{p}$, you can safely assume they are talking about X by default.

There are lots of words they can use when talking about X (the "number," the "amount," the "count," the "frequency," and many more. It is much easier to watch for them asking for the mean or standard deviation of a sample proportion, $\hat{p}$. If they are not clearly talking about $\hat{p}$, then it must be X.

9. (a) What is the mean and standard deviation of the number of voters sampled in favour of the levy?

(A) 0.35; 0.020 (B) 0.65; 0.020 (C) 201.25; 11.44

(D) 373.75; 11.44 (E) 350; 13.46

The number of voters in favour of the levy is X. **The mean of X** is $\mu_X = np = 575 \times .35 =$ **201.25**. The correct answer must be (C). **The standard deviation of X** is $\sigma_X = \sqrt{np(1-p)} = \sqrt{201.25 \times (1-.35)} = 11.4373$ **= 11.44**, as expected.

Solution to Question 9(a)

The correct answer is (C).

9. (b) What is the mean and standard deviation of the proportion of voters sampled in favour of the levy?

(A) 0.35; 0.020 (B) 0.65; 0.020 (C) 201.25; 11.44

(D) 373.75; 11.44 (E) 350; 13.46

Part (b) is talking about the proportion of voters sampled, $\hat{p}$. **Don't miss the word "proportion"! Whenever we see the word "proportion" in a binomial problem, we can bet we will be dealing with $\hat{p}$.**

The mean of $\hat{p}$ is $\mu_{\hat{p}} = p = .35$[*]. The correct answer for (b) must be (A). **The standard deviation of $\hat{p}$** is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.35\times(1-.35)}{575}} = .0199 = \mathbf{.020}$.

Solution to Question 9(b)

The correct answer is (A).

9. (c) What is the probability the sample proportion will be between 30% and 40% in favour of the levy?

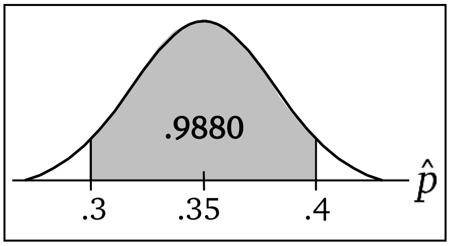
They want the probability the sample proportion is between 30% and 40%. Which is to say, they want the probability $\hat{p}$ is between 30% (.3) and 40% (.4); or, in prof talk: $P(.3 \le \hat{p} \le .4)$. Draw a **$\hat{p}$ -bell curve centred at p = .35** ($\mu_{p} = p = .35$). Mark .3 and .4 on the curve and shade the area between. That area is the probability we are after.

Note: We are certain we have every right to draw a $\hat{p}$-bell curve because $n = 575$ is very large. You can generally take it for granted if n is in the hundreds, the Rule of Thumb will confirm $\hat{p}$ is normally distributed, and not even bother checking. To be safe: we already found

---

[*] You might be saying, "Hey, how can I find the mean of $\hat{p}$, when they never even gave me a value for $\hat{p}$ in the question?" That's irrelevant! We know, before we have taken one single sample, all the sample proportions will average out to the true proportion p. As long as we know p, we know what the mean of $\hat{p}$ will be. Of course, in the real world, we won't know the value of p. But, then we can trust our value of $\hat{p}$, being an unbiased estimate of p, as being a reasonable guess at the true value, since, on average, $\hat{p}$ will be the same as p. Certainly, if we take repeated samples, and compute several $\hat{p}$ values, we could average all of them to come up with a fair guess as to the true p.

$np = 575 \times .35 = 201.25 \geq 10$ back in part (a), and it is obvious $n(1-p) = 575 \times .65$ will be even larger, so $\hat{p}$ is normally distributed.



Use the formula $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$ to standardize the $\hat{p}$ values.

BE CAREFUL! If you round off too much while using this formula, you tend to get inaccurate z scores. Personally, I work out the denominator first for every z score formula, then store the exact answer in a memory in my calculator so that I maintain perfect accuary and don't have to waste time writing anything down. Learn how to use the memory functions on your calculator! If you are unsure or don't feel comfortable doing so, I recommend that you round to no less than SIX DECIMAL PLACES when computing z scores in all formulas. Then you can safely assume your final answer for z rounded off to its usual two decimal places is correct.

When $\hat{p} = .4$: $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}} = \dfrac{.4 - .35}{\sqrt{\dfrac{.35(1-.35)}{575}}} = \dfrac{.05}{.019891} = \mathbf{2.51}$; **Left Area = .9940.**

When $\hat{p} = .3$: $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}} = \dfrac{.3 - .35}{\sqrt{\dfrac{.35(1-.35)}{575}}} = \dfrac{-.05}{.019891} = \mathbf{-2.51}$; **Left Area = .0060.**

The area between .3 and .4 is .9940 – .0060 = **.9880.**

Solution to Question 9(c)

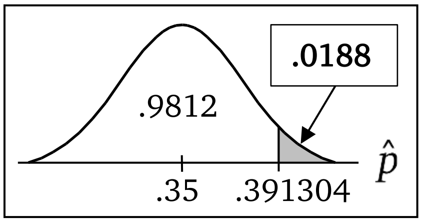
The probability the sample proportion
will be between 30% and 40% is .9880.

9. (d) What is the probability a random sample of 575 voters finds at least 225 in favour of the levy?

Again, we can solve this problem using a $\hat{p}$-bell curve. Remember, in a binomial problem, watch for an n followed by a count x, enabling us to compute the sample proportion, $\hat{p} = \frac{x}{n}$. Here, we have $n = 575$ voters and $x = 225$ are in favour, so we can compute $\hat{p} = \frac{x}{n} = \frac{225}{575} =$ .391304. To find the probability <u>at least</u> 225 are in favour of the levy, is to find the probability $\hat{p}$ is <u>at least</u> .391304, or $P(\hat{p} \geq .391304)$. Draw a $\hat{p}$-bell curve centred at $p = .35$ and shade the area to the right of .391304. Compute the z-score, and use Table A.

When $\hat{p} = .391304$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\frac{225}{575} - .35}{\sqrt{\frac{.35(1-.35)}{575}}} = \frac{.391304 - .35}{.019891} = \mathbf{2.08}$$



The Left Area = .9812. The right area is 1 – .9812 = **.0188.**

Solution to Question 9(d)

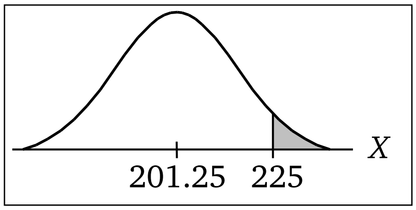
The probability a random sample of 575 will find at least 225 in favour is .0188.

Instead of a $\hat{p}$-bell curve, we could also use an X-bell curve to solve this problem since our Rule of Thumb has assured us the binomial random variable X is approximately normal. Our preference is to us the $\hat{p}$-bell curve for two reasons: (1) that formula will come into play in hypothesis testing for proportions (see Lesson 11), and (2) they give us that formula on the formula sheet.

Nonetheless, we would get the exact same answer using an *X*-bell curve:

We want to find at least 225 are in favour of the levy, or $P(X \geq 225)$.

We learned in part (a) that $\mu_X = 201.25$ and $\sigma_X = 11.4373$. Draw an *X*-Bell curve centred at 201.25 and shade in the region to the right of 225. That area is the approximate probability we are looking for.



When $x = 225$: $z = \dfrac{x-\mu}{\sigma} = \dfrac{225-201.25}{11.4373} = \mathbf{2.08}$

Same z score as we computed above using a $\hat{p}$-bell curve.

You always have a choice when using the normal approximation to the binomial distribution to use either a $\hat{p}$-bell curve or an *X*-bell curve. The answers will be the same either way.

I recommend that you always use the $\hat{p}$-bell curve method to solve any binomial problems that require the normal approximation, as that formula is on your formula sheet given on exams, and it is a formula that shows up in other contexts later in the course. We may as well get used to it now.

WE NOW HAVE TWO KINDS OF BINOMIAL PROBABILITY PROBLEMS

Once you have spotted that a problem is binomial (because you have listed values for n and p), we will use the value of n to focus our approach in solving the problem.

Case 1: When n is small (n ≤ 20, let's say).

That will be our classic binomial problem.

- List what n equals.
- List what p equals.
- List $X = 0, 1, 2, 3, \ldots\ n$ and box in the value(s) you want, k.
- Solve the probability using $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$.

Case 2: When n is huge (n ≥ 100, let's say).[*]

That will certainly be a $\hat{p}$-bell curve problem.

- List what n equals.
- List what p equals.
- List what $\hat{p}$ equals, or compute it knowing that $\hat{p}=\dfrac{x}{n}$
- Draw a $\hat{p}$-bell curve centred at p, mark the value(s) of $\hat{p}$ on the horizontal axis, and shade the relevant area.
- Use $z=\dfrac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}}$ to compute the z score(s) and read Table A to determine the desired probability.

---

[*] What if n is neither small nor huge (between 20 and 80)? It will probably never happen. But, if it does, you will have to use the $np \geq 10$ and $n(1-p) \geq 10$ check to see if it is appropriate to use the $\hat{p}$-bell curve, or not.

10. A mail-order company finds 7% of its orders tend to be damaged in shipment. If 500 orders are shipped:
    - (a) Compute the mean and standard deviation of the number of orders that would be damaged.
    - (b) Find the approximate probability between 30 and 50 orders (inclusive) will be damaged.
    - (c) Find the approximate probability at least 50 orders will be damaged.

We have been given a *percentage* which means we have been given a *p*. This is a **binomial distribution**. We are given p **= 7% = .07** of the orders tend to be damaged and a sample of size n **= 500**. Let **X = the *number* of orders that get damaged**. Which is to say:

n = 500 packages
p = 7% = .07 are damaged
X = the number of damaged packages = 0, 1, 2, 3, … 499, 500

We want to find the mean and standard deviation of *X*, the *number* of orders damaged (not the mean and standard deviation of $\hat{p}$, the *proportion* of orders in the sample damaged.

$$\mu_X = np = 500 \times .07 = \mathbf{35}$$

$$\sigma_X = \sqrt{np(1-p)} = \sqrt{35 \times (1-.07)} = \sqrt{\mathbf{32.55}} = \mathbf{5.7053}$$

Solution to Question 10(a)

The mean and standard deviation of the number of orders damaged are 35 and 5.7053, respectively.

(Basically, we have found that, in a shipment of 500 orders, we would have an average of 35 orders damaged, give or take about 5.7.)

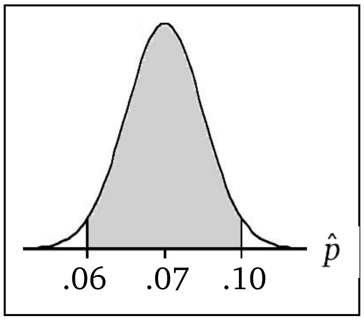
10. (b) Find the approximate probability between 30 and 50 orders (inclusive) will be damaged.

We have a binomial probability problem where **n is huge (n = 500)**. This is a **$\hat{p}$-bell curve** problem. We saw in part (a) that $np = 35 \geq 10$ and it is pretty obvious that $n(1-p)$ would be even higher: $n(1-p) = 500 \times (1 - .07) = 500 \times .93 = 465 \geq 10$. **Use the normal approximation to the binomial distribution.**[*]

So **n = 500** and we want the count **x** to be between **30** and **50**. Compute the two sample proportions, $\hat{p}$.

When **x = 30**, $\hat{p} = \frac{30}{500} = .06$. When **$x$ = 50**, $\hat{p} = \frac{50}{500} = .10$

Draw a $\hat{p}$-bell curve centred at **p = .07**, and shade the area between the two scores of **$\hat{p}$ = .06** and **$\hat{p}$ = .10**.



Now, we must change these 2 $\hat{p}$-scores into z-scores:

When **$\hat{p}$ = .10**, $z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.10-.07}{\sqrt{\frac{.07(1-.07)}{500}}} = \frac{.03}{.011411} = \mathbf{2.63}$

When **$\hat{p}$ = .06**, $z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.06-.07}{\sqrt{\frac{.07(1-.07)}{500}}} = \frac{-.01}{.011411} = -0.88$

Consulting Table A:

When z = 2.63, the Left Area = .9957.

When z = –0.88, the Left Area = .1894.

The area between $\hat{p}$ = .06 and .10 (X = 30 and 50) is .9957 – .1894 = **.8063.**

Solution to Question 10(b)

The approximate probability between 30 and 50 orders, inclusive, are damaged is .8063.

---

[*] I wouldn't even do the $np \geq 10$ and $n(1-p) \geq 10$ check. You can take it to the bank: when n is huge, the normal approximation to the binomial, $\hat{p}$-bell curve, method is the correct approach.

11. A recent article claimed 40% of 12-year old American children are at least 5 pounds overweight. Assuming this is true, what is the probability a random sample of 600 children finds no more than 220 12-year old American children are overweight?

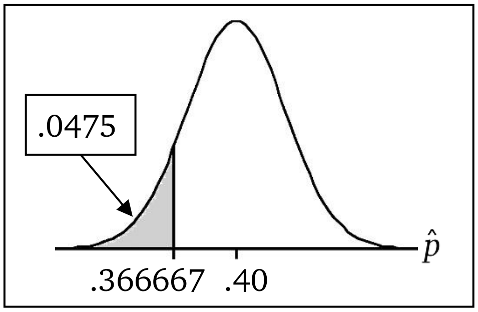
(A) 0.0470 (B) 0.0475 (C) 0.9525 (D) 0.9530 (E) none of the above

We have been given a *percentage* which means we have been given a p. This is a **binomial distribution**. Here we are told $\boldsymbol{p = 40\% = .40}$ American children are overweight. We have a binomial probability problem where **n is huge ($n = 600$)**. This is a **$\hat{p}$-bell curve** problem.

So $\boldsymbol{n = 600}$ and we want the count $\boldsymbol{x}$ to be no more than **220** are overweight, which is to say 220 or LESS are overweight. Compute the sample proportion, $\hat{p}$.

When $\boldsymbol{x = 220}$, $\hat{p} = \frac{220}{600} = .366667$. So .366667 or LESS are overweight. Find the probability that $\hat{p}$ is .366667, $P(\hat{p} \le .36667)$.

Draw a $\hat{p}$-bell curve centred at $\boldsymbol{p = .40}$ and shade in the region to the left of $\hat{p} = .366667$. Change the $\hat{p}$ score into a z score



When $\hat{p} = .366667$: $$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.366667 - .40}{\sqrt{\frac{.40(1-.40)}{600}}} = \frac{-.033333}{.02} = -1.67$$

Consulting Table A, the area to the left of $z = -1.67$ is **.0475**.

Solution to Question 11

The approximate probability no more than 220 children are overweight is .0475.
The correct answer is (B).

12. The table below shows the faculty membership for full-time undergraduates at a certain university.

Undergrad's Faculty	Arts	Science	Commerce	Nursing	Other
Percentage	35%	25%	22%	12%	6%

(a) If we select a random sample of 10 full-time undergraduates, what is the probability that no more than one of them belongs to the Nursing faculty?

(b) If we select a random sample of 12 full-time undergraduates, what is the probability that exactly 6 of them belong to the Science faculty?

(c) If we randomly select a sample of 100 full-time undergraduates, what is the probability that less than 40% of them belong to the Arts faculty?

(d) If we randomly select a sample of 200 full-time undergraduates, what is the probability that at least 30% of them belong to the Commerce faculty?

(e) If we randomly select a sample of 30 full-time undergraduates, what is the mean of the sample proportion who belong to the Nursing faculty? What is the variance?

(f) If we randomly select a sample of 50 full-time undergraduates, what is the mean amount of them who belong to the Science faculty? What is the variance?

At a glance, the table we are given in this question should lead us to expect this is just a **discrete probability distribution** question, similar to **Lesson 5, question 1**. We may have expected them to ask things like, "What is the probability a randomly selected person belongs to the Arts faculty?" But, a glance at each part of this question reveals that each time they give us a value of *n* (10 undergraduates, 12 undergraduates, etc.) This is a series of **binomial** questions.

Each individual cell in a discrete probability distribution is potentially the percentage *p* for an associated binomial problem for a given sample size *n*.

12. The table below shows the faculty membership for full-time undergraduates at a certain university.

Undergrad's Faculty	Arts	Science	Commerce	Nursing	Other
Percentage	35%	25%	22%	12%	6%

(a) If we select a random sample of 10 full-time undergraduates, what is the probability that no more than one of them belongs to the Nursing faculty?

This is a **Binomial Distribution**. Two outcomes: an undergraduate either belongs to the Nursing faculty, or they do not. There is a $p = 12\% = .12$ chance that they belong to Nursing, and we have a random sample of $n = 10$ undergraduates. Since **n is small**, this is just a **classic binomial probability problem**. $X = 0, 1, 2, \dots 10$ undergrads. We want "no more than one" to belong to Nursing, that is one undergrad or less. **Find $P(X \le 1)$**.

$n = 10$ undergraduates selected

$p = 12\% = .12$ belong to Nursing

X = the *number* of undergraduates in Nursing = $\underset{k=0,\,1}{\boxed{0, 1,}}$ 2, 3, . . . 10

Find $P(X = 0 \text{ or } 1)$.

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$\left.\begin{aligned} k=0: \quad P(X=0)&=\binom{10}{0}\times(.12)^0\times(1-.12)^{10}=\mathbf{.2}785 \\ k=1: \quad P(X=1)&=\binom{10}{1}\times(.12)^1\times(1-.12)^9=.3798 \end{aligned}\right\}\text{ add these together}$$

$P(X = 0 \text{ or } 1) = .2785 + .3798 = \mathbf{.6583}$.

Solution to Question 12(a)

There is a .6583 probability that no more than one undergrad belongs to Nursing.

12. The table below shows the faculty membership for full-time undergraduates at a certain university.

Undergrad's Faculty	Arts	Science	Commerce	Nursing	Other
Percentage	35%	25%	22%	12%	6%

(b) If we select a random sample of 12 full-time undergraduates, what is the probability that exactly 6 of them belong to the Science faculty?

This is a **Binomial Distribution**. Two outcomes: an undergraduate either belongs to the Science faculty, or they do not. There is a $p = 25\% = .25$ chance that they belong to Science, and we have a random sample of $n = 12$ undergraduates. Since **n is small**, this is just a **classic binomial probability problem**. $X = 0, 1, 2, \ldots 12$ undergraduates. We want "exactly 6" to belong to Science. **Find $P(X = 6)$.**

$n = 12$ undergraduates selected

$p = 25\% = .25$ belong to Science

X = the *number* of undergraduates in Science = $0, 1, 2, \ldots 5, \underset{k=6}{\boxed{6,}}\ 7 \ldots 12$

Find **$P(X = 6)$**.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$k = 6: \quad P(X = 6) = \binom{12}{6} \times (.25)^6 \times (1-.25)^6 = .0401$$

$P(X = 6) = .0401$.

Solution to Question 12(b)

There is a .0401 probability that exactly 6 undergraduates belongs to Science.

12. The table below shows the faculty membership for full-time undergraduates at a certain university.

Undergrad's Faculty	Arts	Science	Commerce	Nursing	Other
Percentage	35%	25%	22%	12%	6%

(c) If we randomly select a sample of 100 full-time undergraduates, what is the probability that less than 40% of them belong to the Arts faculty?

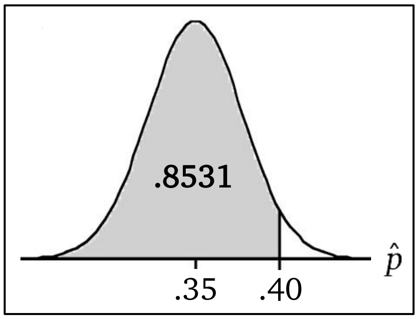
This is a **Binomial Distribution**. Two outcomes: an undergraduate either belongs to the Arts faculty, or they do not. There is a $p = 35\% = .35$ chance that they belong to Arts, and we have a random sample of $n = 100$ undergraduates. Since **n is huge**,[*] this must be a **$\hat{p}$-bell curve probability problem**. We want "less than 40%" to belong to Arts. That is the percentage of the sample, or the sample proportion, $\hat{p}$! **Find $P(\hat{p} < .40)$**.

$n = 100$ undergraduates selected

$p = 35\% = .35$ belong to Arts

Find $P(\hat{p} < .40)$

Draw a **$\hat{p}$-bell curve centred at $p = .35$**. Mark .40 on the curve and shade the area to the left. That area is the probability we are after. Compute the z score using $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$.



$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.40 - .35}{\sqrt{\frac{.35(1-.35)}{100}}} = \frac{.05}{.047697} = 1.048... = \mathbf{1.05}$$

When **$z = 1.05$, the Left Area = .8531**, according to Table A. $P(\hat{p} < .40) = .8531$.

Solution to Question 12(c)

There is a .8531 probability that less than 40% of undergraduates belong to Arts.

---

[*] Technically, we should use the Rule of Thumb to check the normal approximation is valid, but you can bet on it when n is huge. Here, $np = 100 \times .35 = 35$, and $n(1-p) = 100 \times .65 = 65$. Both are at least 10.

12. The table below shows the faculty membership for full-time undergraduates at a certain university.

Undergrad's Faculty	Arts	Science	Commerce	Nursing	Other
Percentage	35%	25%	22%	12%	6%

(d) If we randomly select a sample of 200 full-time undergraduates, what is the probability that at least 30% of them belong to the Commerce faculty?

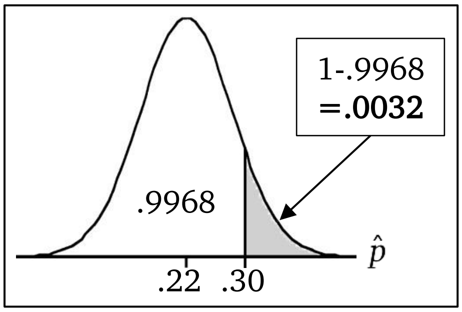
This is a **Binomial Distribution**. Two outcomes: an undergraduate either belongs to the Commerce faculty, or they do not. There is a $p = 22\% = .22$ chance that they belong to Commerce, and we have a random sample of $n = 200$ undergraduates. Since **n is huge**, this must be a **$\hat{p}$-bell curve probability problem**. We want "at least 30%" to belong to Commerce (30% or more). That is the percentage of the sample, or the sample proportion, $\hat{p}$! **Find $P(\hat{p} \geq .30)$.**

$n = 200$ undergraduates selected

$p = 22\% = .22$ belong to Commerce

Find $P(\hat{p} \geq .30)$

Draw a **$\hat{p}$-bell curve centred at $p = .22$**. Mark .30 on the curve and shade the area to the right. That area is the probability we are after. Compute the z score using $z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}}$.



$$z = \frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} = \frac{.30-.22}{\sqrt{\dfrac{.22(1-.22)}{200}}} = \frac{.08}{.029292} = 2.73$$

When **$z = 2.73$, the Left Area = .9968**, according to Table A.

$$P(\hat{p} \geq .30) = 1 - .9968 = .0032.$$

Solution to Question 12(d)

There is a .0032 probability that at least 30% of undergrads belong to Commerce.

12. The table below shows the faculty membership for full-time undergraduates at a certain university.

Undergrad's Faculty	Arts	Science	Commerce	Nursing	Other
Percentage	35%	25%	22%	12%	6%

(e) If we randomly select a sample of 30 full-time undergraduates, what is the mean of the sample proportion who belong to the Nursing faculty? What is the variance?

This is a **Binomial Distribution**. Two outcomes: an undergraduate either belongs to the Nursing faculty, or they do not. There is a $\boldsymbol{p} = \mathbf{12\% = .12}$ chance that they belong to Nursing, and we have a random sample of $\boldsymbol{n} = \mathbf{30}$ undergraduates. It is irrelevant whether ***n* is small or huge** because we are not asked to compute a probability. We are merely asked to compute the **mean and variance**.

Remember: There are two different means we can be asked to compute in a binomial setting. The mean *number*, μ_X, or the mean proportion, $\mu_{\hat{p}}$. Here, they clearly ask for "the mean of the sample proportion" and the variance (the square of the standard deviation). We want $\mu_{\hat{p}}$ and $\sigma^2_{\hat{p}}$.

$\boldsymbol{n} = \mathbf{30}$ undergraduates selected

$\boldsymbol{p} = \mathbf{12\% = .12}$ belong to Nursing

The **mean** is $\boldsymbol{\mu_{\hat{p}} = p = .12}$.

The **variance** is $\boldsymbol{\sigma^2_{\hat{p}} = \dfrac{p(1-p)}{n} = \dfrac{.12(1-.12)}{30} = .00352}$.[*]

Solution to Question 12(e)

The mean and variance of the sample proportion who are in the Nursing faculty from a sample of 30 undergraduates are .12 and .00352, respectively.

---

[*] Since the standard deviation of $\hat{p}$ is $\sqrt{\dfrac{p(1-p)}{n}}$, the variance is just the square of that. If we square that formula, that removes the square root. Thus, the variance of $\hat{p}$ is $\sigma^2_{\hat{p}} = \dfrac{p(1-p)}{n}$

12. The table below shows the faculty membership for full-time undergraduates at a certain university.

Undergrad's Faculty	Arts	Science	Commerce	Nursing	Other
Percentage	35%	25%	22%	12%	6%

(f) If we randomly select a sample of 50 full-time undergraduates, what is the mean amount of them who belong to the Science faculty? What is the variance?

This is a **Binomial Distribution**. Two outcomes: an undergraduate either belongs to the Science faculty, or they do not. There is a $\boldsymbol{p} = \mathbf{25\%} = \mathbf{.25}$ chance that they belong to Science, and we have a random sample of $\boldsymbol{n} = \mathbf{50}$ undergraduates. It is irrelevant whether **n is small or huge** because we are not asked to compute a probability. We are merely asked to compute the **mean and variance**.

Remember: There are two different means we can be asked to compute in a binomial setting. The mean *number*, μ_X, or the mean proportion, $\mu_{\hat{p}}$. Here, they ask for "the mean *amount* of them" and the variance (the square of the standard deviation). There is no mention of a sample proportion. We want μ_X, the mean *number* of undergrads and σ_X^2.

$\boldsymbol{n} = \mathbf{50}$ undergraduates selected

$\boldsymbol{p} = \mathbf{25\%} = \mathbf{.25}$ belong to Science

The **mean** is $\boldsymbol{\mu_X} = \boldsymbol{np} = 50 \times .25 = \mathbf{12.5}$.

The **variance** is $\boldsymbol{\sigma_X^2 = np(1-p) = 12.5(1-.25) = 9.375}$.[*]

Solution to Question 12(f)

The mean and variance of the amount of a sample of 50 undergraduates who are in the Science faculty are 12.5 and 9.375, respectively.

---

[*] Since the standard deviation of X is $\sqrt{np(1-p)}$, the variance is just the square of that. If we square that formula, that removes the square root. Thus, the variance of X is $\sigma_X^2 = np(1-p)$.

SUMMARY OF KEY CONCEPTS IN LESSON 6

- **The parameters of a binomial distribution are n and p.**
- The binomial distribution is a **discrete** distribution where each trial is **independent**. If we have a fixed number of trials n and if the probability of "yes" is the same for each trial, p, the random variable X has a binomial distribution where X = the number of "yeses". We can say $\boldsymbol{X \sim B\,(n, p)}$.
    - If we are given a **percentage**, a **proportion**, or a **fraction** we are given a value of $\boldsymbol{p}$. We will immediately suspect we have a binomial distribution at that point. All we need is a value for n to clinch it.
    - If we are rolling dice, tossing coins, or guessing on a test, we have a binomial distribution where we are expected to know the value of p ourselves. Again, we must have a specific number of trials n or else the problem is not binomial.
- The binomial probability formula is $\boldsymbol{P\left(X=k\right)=\binom{n}{k}p^{k}\left(1-p\right)^{n-k}}$.
- If X has a **binomial** distribution, then there are two different means we can be asked to compute, and two different standard deviations. The mean *number* X, and its standard deviation; or, the mean of the sample *proportion* $\hat{p}$, and its standard deviation.
    - The **mean of X = μ_X = np** and the **standard deviation of X = σ_X = $\sqrt{np\left(1-p\right)}$.**
    - The **mean of $\hat{p}$ = $\mu_{\hat{p}}$ = p** and the **standard deviation of $\hat{p}$ = $\sigma_{\hat{p}}$ = $\sqrt{\dfrac{p\left(1-p\right)}{n}}$**.
- Our **Rule of Thumb** says: If $N \geq 10n$, and if $np \geq 10$ and $n(1-p) \geq 10$, then both X and $\hat{p}$ in a binomial distribution are approximately normal.
- If we want to find the probability the sample proportion $\hat{p}$ is above, below or between some given amount(s), we can bet our Rule of Thumb will tell us $\hat{p}$ is approximately normal, so we can use a **$\hat{p}$-bell curve** to compute the approximate probability**.**
    - The standardizing formula for $\hat{p}$ is $\boldsymbol{z=\dfrac{\hat{p}-\mu_{\hat{p}}}{\sigma_{\hat{p}}}=\dfrac{\hat{p}-p}{\sqrt{\dfrac{p\left(1-p\right)}{n}}}}$.

- **There are two cases of binomial probability problems.**
  - Case 1: **When n is small ($n \le 20$, let's say).**
    - That will be our classic binomial problem.
      - List what n equals.
      - List what p equals.
      - List $X = 0, 1, 2, 3, \ldots n$ and box in the value(s) you want, k.
      - Solve the probability using $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.
  - Case 2: **When n is huge ($n \ge 100$, let's say).**
    - That will certainly be a $\hat{p}$-bell curve problem.
      - List what n equals.
      - List what p equals.
      - List what $\hat{p}$ equals, or compute it knowing that $\hat{p} = \frac{x}{n}$
      - Draw a $\hat{p}$-bell curve centred at p, mark the value(s) of $\hat{p}$ on the horizontal axis, and shade the relevant area.
      - Use $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$ to compute the z score(s) and read Table A to determine the desired probability.

LECTURE PROBLEMS FOR LESSON 6

For your convenience, here are the questions I used as examples in this lesson. Do not make any marks or notes on these questions below. Especially, do not circle the correct choice in the multiple choice questions. You want to keep these questions untouched, so that you can look back at them without any hints. Instead, make any necessary notes, highlights, etc. in the lecture part above.

1. Thirty-five percent of the voters in the last election voted Liberal. If you randomly selected ten voters from the last election, what is the probability exactly four of them voted Liberal?
(See the solution on page 389.)

2. A die is rolled seven times.
(a) What is the probability we roll a three four times?
(A) 0.0156 **(B)** 0.2857 **(C)** 0.4286 **(D)** 0.5714 **(E)** 0.8988
(See the solution on page 390.)
(b) What is the probability you get at least one 5?
(A) 0.2791 **(B)** 0.6093 **(C)** 0.7209 **(D)** 0.3907 **(E)** 1
(See the solution on page 392.)

3. A student is writing a multiple-choice Statistics exam. Each question has 5 choices and only one choice is correct. There are a total of 20 questions on the exam. If the student is simply guessing on every single question:
(a) What is the probability he just barely passes the exam (gets exactly 50%)?
(A) 0.5000 **(B)** 0.1762 **(C)** 0.0026 **(D)** 0.0020 **(E)** 0.0002
(See the solution on page 393.)
(b) What is the probability he passes the exam?
(A) 0.5000 **(B)** 0.1762 **(C)** 0.0026 **(D)** 0.0020 **(E)** 0.0002
(See the solution on page 394.)

4. A seed company has determined its seeds have a 90% chance of germinating. If 20 seeds are planted what is the probability more than 18 will germinate?
(A) 0.270 **(B)** 0.285 **(C)** 0.392 **(D)** 0.608 **(E)** 0.715
(See the solution on page 395.)

5. It is known 75% of the executives at a major multinational corporation are male. In a random sample of 8 executives from this corporation, what is the probability 3 or 4 of them are female?
(A) 0.2942 **(B)** 0.0865 **(C)** 0.2076 **(D)** 0.0231 **(E)** 0.1096
(See the solution on page 396.)

6. An airline determines 97% of the people who booked a flight actually show up in time to take their seat. Assuming this is true, what is the probability, in a randomly selected sample of 12 people who independently booked various flights, no more than 10 of them showed up?

(See the solution on page 397.)

7. Determine whether the variable X has a binomial distribution in each of the following cases. If it does, explain why and determine the values of the parameters n and p (or define the parameters n and p if the actual value is unknown). If X does not have a binomial distribution, explain why not.

(a) You randomly select ten cards from a deck of playing cards, with replacement.

X = number of "Diamonds" selected.

(See the solution on page 399.)

(b) You randomly select ten cards from a deck of playing cards, with replacement.

X = number of "Queens" selected.

(See the solution on page 399.)

(c) You randomly select ten cards from a deck of playing cards, without replacement.

X = number of "Clubs" selected.

(See the solution on page 400.)

(d) You roll a fair die repeatedly.

X = number of rolls until you get a "6".

(See the solution on page 400.)

(e) You randomly select one resident from each of the 30 apartments in a specific building.

X = number of females selected.

(See the solution on page 401.)

(f) A boy asks the same girl to go on a date every day for a week.

X = number of times the girl agrees to go on a date with the boy.

(See the solution on page 401.)

(g) An airplane carrying 200 passengers has two emergency exits, one at the front of the plane and one over the wings. The plane makes an emergency landing.

X = number of passengers that exit the plane through the front emergency exit.

(See the solution on page 402.)

(h) I am looking at a specific row in the random digits table. There are 100 rows on the random digit table we are given, and each row has 40 digits.

X = number of times there is a "0" in the row I am looking at.

(See the solution on page 402.)

(i) Weights of free-range roasting chickens follow a normal distribution with mean 1.955 kg and standard deviation 0.325 kg. You select a random sample of 12 chickens.

X = number of chickens that weigh less than 1.5 kilograms.

(See the solution on page 403.)

(j) Every morning (Monday to Friday), I buy one cup of coffee at *Tim Horton's* during their *Roll Up The Rim to Win* contest.

X = number of times in one week I win a prize.

(See the solution on page 404.)

8. A newspaper reports that one in three drivers routinely exceed the speed limit. Assuming this is true, we select a random sample of 30 drivers.

(a) What is the probability exactly half of them exceed the speed limit?

(See the solution on page 405.)

(b) What is the mean number of drivers in a sample of this size who routinely exceed the speed limit, and what is the standard deviation?

(See the solution on page 406.)

9. In Big City only 35% of the voters in the last election were in favour of a one-time levy to cover the cost of sewer upgrades. During the current campaign a random sample of 575 voters will be selected. Assume the opinion has not changed since the last election.

(a) What is the mean and standard deviation of the number of voters sampled in favour of the levy?

(A) 0.35; 0.020 **(B)** 0.65; 0.020 **(C)** 201.25; 11.44
(D) 373.75; 11.44 **(E)** 350; 13.46

(See the solution on page 419.)

(b) What is the mean and standard deviation of the proportion of voters sampled in favour of the levy?

(A) 0.35; 0.020 **(B)** 0.65; 0.020 **(C)** 201.25; 11.44
(D) 373.75; 11.44 **(E)** 350; 13.46

(See the solution on page 420.)

(c) What is the probability the sample proportion will be between 30% and 40% in favour of the levy?

(See the solution on page 421.)

(d) What is the probability a random sample of 575 voters finds at least 225 in favour of the levy?

(See the solution on page 422.)

10. A mail-order company finds 7% of its orders tend to be damaged in shipment. If 500 orders are shipped:

(a) Compute the mean and standard deviation of the number of orders that would be damaged.

(See the solution on page 425.)

(b) Find the approximate probability between 30 and 50 orders (inclusive) will be damaged.

(See the solution on page 426.)

11. A recent article claimed 40% of 12-year old American children are at least 5 pounds overweight. Assuming this is true, what is the probability a random sample of 600 children finds no more than 220 12-year old American children are overweight?

(A) 0.0470 **(B)** 0.0475 **(C)** 0.9525 **(D)** 0.9530 **(E)** none of the above

(See the solution on page 427.)

12. The table below shows the faculty membership for full-time undergraduates at a certain university.

Undergrad's Faculty	Arts	Science	Commerce	Nursing	Other
Percentage	35%	25%	22%	12%	6%

(a) If we select a random sample of 10 full-time undergraduates, what is the probability that no more than one of them belongs to the Nursing faculty?
(See the solution on page 429.)

(b) If we select a random sample of 12 full-time undergraduates, what is the probability that exactly 6 of them belong to the Science faculty?
(See the solution on page 430.)

(c) If we randomly select a sample of 100 full-time undergraduates, what is the probability that less than 40% of them belong to the Arts faculty?
(See the solution on page 431.)

(d) If we randomly select a sample of 200 full-time undergraduates, what is the probability that at least 30% of them belong to the Commerce faculty?
(See the solution on page 432.)

(e) If we randomly select a sample of 30 full-time undergraduates, what is the mean of the sample proportion who belong to the Nursing faculty? What is the variance?
(See the solution on page 433.)

(f) If we randomly select a sample of 50 full-time undergraduates, what is the mean amount of them who belong to the Science faculty? What is the variance?
(See the solution on page 434.)

HOMEWORK FOR LESSON 6

- Study the lesson thoroughly until you can do all of the Lecture Problems from start to finish without any assistance. **I have collected the Lecture Problems together for your convenience starting on page 436 above.**
- I have provided a **Summary of Key Concepts** starting on page 428 above.
- **Do not try to learn the material by doing your hand-in assignments. Learn the lesson first, then use the hand-in assignments to test your understanding of the lesson.** Before each hand-in assignment, I will send you tips telling you what lesson you should be studying to prepare for the assignment. Make sure you sign up for **Grant's FREE Homework Help** at **grantstutoring.com** to receive these tips.
- If you have the ***Multiple-Choice Problems Set for Basic Statistical Analysis I (Stat 1000)*** by Smiley Cheng available in the Statistics section of the UM Book Store (and I do recommend you get this book for all the old midterm and final exams contained within it, if nothing else), then additional practise at many of the concepts taught in this lesson is available in:
    - **Section BIN: Binomial Distribution. The solutions to this section are provided in Appendix B of my book starting on page B-11 below.**
- **Have you signed up for Grant's FREE Homework Help yet?** Go to **grantstutoring.com** to sign up.
- **Grant's Final Exam Prep Seminar** has been scheduled by now. Go to **grantstutoring.com** to get all the info and register if you are interested.