Solution
(with correction
on step 5 3/16)

Minitab Exercise 4 Binomial Probabilities

Rev ~~01/14~~
12/14

In this session you will

- Generate random data of the binomial distribution
- Better understand the concept of binomial distributions
- Calculate cumulative probabilities using binomial distribution
- Approximate binomial probabilities using the normal distribution

Open a new Minitab project.

Step 1 Generating binomial random data.

Assume we have an automatic spot welding machine that welds hundreds of spots each day. We will inspect a number of these welds each day. Our probability of getting a bad weld is 20% (this is a really bad process). We want to know how many bad welds we can expect to get if we take samples of size 5 or samples of size 100.

In statistical terms, we say we are going to generate two sets of data taken from the same population (all the welds), where the probability of success (bad weld) is p=0.20. In one we will take samples of size 5. In the other we will take samples of size 100

Label columns C1 and C2 as **n5** and **n100**.

> **Calc>Random Data>Binomial**
> Number of rows to generate: **1000**
> Store in column: **n5**
> Number of trials: **5**
> Event probability: **0.2**
> Click OK.

What appears in the worksheet is a column of data with numbers between 0 and 5. These numbers are the numbers of successes (defective welds) that were in the sample of size 5.

Now generate 1000 rows of data for n100 with 100 trials, keeping the event probability at 0.2.

Step 2. Generating a percent histogram

Now we want to take a closer look at the distribution of n5. Let's look at the shape of the distribution with a histogram.

> **Graph>Histogram**
> select **Simple**
> Graph variable: **n5**
> Choose **Scale>Y scale type>Percent** Click OK twice.

The graph will show the five possible outcomes of the trials (x=0 successes, 1 success, 2, 3, 4, or 5* successes on the x axis. On the y axis, you can read the probability of obtaining x successes, given a sample size of n=5 and probability of success of 20%. * Explanation to come as to why 5 may not be in the data.

Step 3 Understanding Binomial Distribution

Since your "experiment" was repeated 1000 times, each of the bars on the histogram should be close to the theoretical probability.

Recall from the lecture for the binomial distribution

The probability of exactly x occurrences in n trials in an event has a constant probability of occurrence $p(x) = \binom{n}{x} p^x(1-p)^{n-x}$ where $p =$ probability of success and $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ $n =$ the number of trials with $n! = (n)(n-1)(n-2)(n-2)(\ldots\ldots.)(1)$ $x =$ the number successes

Work the following by hand – show your work.

Calculate the theoretical probability in our problem where p=.2 and n=5 for x=0 successes.

$P(0) = \binom{5}{0} 0.2^0 (1-0.2)^{5-0}$

$= (1)(1)(0.8)^5$

$= (1)(1)(.32768) = 32.77\%$

Does your calculation look close to the graph percentage? yes

What is the probability getting 5 bad welds in your sample?

$P(5) = \binom{5}{5} 0.2^5 (1-0.2)^{5-5}$

$= (1)(.00032)(1) = 0.00032 = 0.032\%$

You probability did not get any observations of x=5 in your data, this should explain why.

What is the probability of getting less than three bad welds? (Add the probabilities of getting 0, 1 and 2 bad welds - looking for P(x≤2).)

$P(0) = .32768$

$P(1) = \binom{5}{1} 0.2^1 (0.8)^4 = (5)(.2)(.4096) = .4096$

$P(2) = \binom{5}{2} 0.2^2 (0.8)^3 = (10)(0.04)(0.512) = .2048$

$$\begin{array}{lr} P(0) & .32768 \\ +\ P(1) & .4096 \\ +\ P(2) & +.2048 \\ \hline P(x<3) & .9421 \end{array} \Rightarrow 94.21\%$$

Step 4. Generating binomial probabilities with Minitab

Check your work above with Minitab

To find the theoretical probability in our problem where p=.2 and n=5 for x=0 successes:

> **Calc> Probability Distributions>Binomial**
> Choose **Probability**
> Number of trials: 5
> Event probability: 0.2
> Input constant: 0
> Click OK

Verify that your answer matches your previous calculation. Follow the same steps to check you answer for x=5 by changing the input constant to 5.

To find P(x≤X), the probability of x or fewer successes, use the **cumulative probability**.

For the problem: What is the probability of getting less than three bad welds? (Add the probabilities of getting 0, 1 and 2 bad welds, looking for P(x≤2). You can check your answer as follows:

> **Calc> Probability Distributions>Binomial**
> Choose **Cumulative probability**
> Number of trials: 5
> Event probability: 0.2
> Input constant: 2
> Click OK

The **inverse cumulative probability** is used when you want to find a value of x for a given cumulative probability. For example, to determine the number of defects were would get 90% of the time:

> **Calc> Probability Distributions>Binomial**
> Choose **Inverse cumulative probability**
> Number of trials: 5
> Event probability: 0.2
> Input constant: 0.9
> Click OK

Because binomial is a discrete distribution, it gives us the two values of x between which the 90% would fall.

Step 5. Binomial Practice

Work the following problems by hand (show all work), then you can check your work in Minitab.

Ten percent of the items in a large lot are defective. A sample of 8 items is drawn from the lot.

$p = 0.10$

$n = 8$

a) What is the probability that none of the sampled items are defective?

$P(0) = \binom{8}{0} .10^0 (1-.10)^8$

$= (1)(1)(.4305) = .4305$

b) What is the probability that one or more of the items are defective?

$P(x > 0) = 1 - P(0) = 1 - .4305 = .5697$

c) What is the probability that exactly 2 of the items are defective?

$P(2) = \binom{8}{2} .10^2 (.90)^6$

$= (28)(.01)(.53144) = 0.1488$

d) What is the probability that fewer than two items are defective?

$P(x < 2) = P(0) + P(1)$

$= .4035 + .3826 = .8131$

$P(1) = \binom{8}{1} .10^1 (.9)^7$

$= (8)(.10)(.47830) = .3826$

Step 6. Binomial problems with Minitab

You can use Minitab to help work the following problems. Write the inputs, answers, and any computations that Mintab did not do for you in the space. (The type of information needed is given in part a, provide similar information for the other parts.)

a) Among the working engineers in the United States, approximately 7% are women. If 8 engineers are selected at random from this occupation, use the binomial distribution to find out the probability that

Exactly 1 is female. *n*= 8 *p*= .07 *x*= 1 *P(x=1)* = 0.3369

At least 3 are female. (This is one minus the probability that there are two or less in the room.)

n= 8 *p*= .07 *x*= 2 *P(X≥x)* = 1- *P(X≤x)* = 0.0147

$P(x \leq 2) = .9853$ (from Minitab)

$1 - .9853 = .0147$

b) A microchip plant purchases large lots of silicon wafers. 200 are taken from the lot and inspected. If the number of defective wafers is greater than 14, the lot is rejected. Find the approximate probability of rejection of the lot if the proportion of the defective wafers in the lot is 5%

$n=200 \quad p=.05 \qquad P(\text{rejection}) = P(x>14) = 1-P(x \le 14)$

$P(x \le 14) = .9219$ (from Minitab)

$1-.9219 = .07813$

c) The vendor in b) wants the inspection criteria changed so that ~~there~~ their lots are accepted at least 95% of the time (the probability of x or less defectives is greater than 0.95). If the sample size is still 200 and the proportion defective is still 5%, how many defective parts in the lot does the vendor want set as the acceptable limit?

$n=200 \quad p=.05 \qquad P(X < x) = .95$ (use inverse probability)

$P(x < 14) = .9219$

$P(x < 15) = .9556$ ← to accept at least 95% of time reject where $x > 15$

Step 7. Large Sample sizes

Recall our original problem: Assume we have an automatic spot welding machine that welds hundreds of spots each day. We will inspect a number of these welds each day. Our probability of getting a bad weld is 20%. $p=.2$

Use Minitab to determine:

$n=100 \qquad x=18$

With a sample of size n= 100, what is the probability that we will get exactly 18 defects in the sample. .0909

With the sample size of n=100, what is the probability that the number of defects will be 18 or less? .3621

$n=100 \qquad x \le 18$

Recall from the lecture:

> For a process following a binomial distribution, if repeated samples of size n are taken from the process where the probability of success is p, the average number of successes in the sample, and the variance in the number of success are as follows
>
> Mean: $\mu = np$ Variance: $\sigma^2 = np(1-p)$

For the process where p=.20 and we take a sample size n=100, what are the theoretical mean and standard deviation of the number of successes we would observe.

$\mu = (100)(.20) = 20 \qquad \sigma = \sqrt{(100)(.2)(1-.8)} = \sqrt{16} = 4$

In step one, we simulated an experiment where we took samples of size n=100 many times.
For the n100 column, calculate the descriptive statistics and histogram with normal curve. (**Stat>Basic Statistics>Display Descriptive Statistics**, and use **Graphs**-Histogram of data with normal curve)

Look in the session window to find your descriptive statistics for the n100 data. (will vary based on your random data)

What is the mean? 19.862 Standard Deviation? 3.917 (to 3 decimal places)

How does this compare to your theoretical calculations? fairly close

Step 8. Normal Approximation for Large Sample sizes

Step 8

Look at the graph with the normal curve. Recall that we are dealing with discrete data – each bar represents the number of times we got a particular count of successes in the experiment of taking 100 observations with the proportion of successes in the population is 20%.

Just from looking at the graph, how well do you think the normal curve approximates the distribution?

The distribution looks normal – good approximation.

> Where n is large and p is small you can use a normal approximation to the binomial. The rule of thumb is that np and n(1-p) must both be greater than 5.

Recall that on the graph, the bars are centered on the value that they represent. The normal curve is continuous. To do a normal approximation, you want the portion of the normal curve that is comparable to the binomial data. So you must look not to the center of the bar, but edge to edge so that you have an interval to evaluate.

To determine the probability of getting 18 defects using the normal distribution, we must use the interval from 17.5 to 18.5.

Using the normal probability distribution with μ = 20 and σ = 4 to estimate the binomial, what percent of the time would you expect to observe 18 defects?

Calc>Probability Distributions>Normal
Select **Cumulative probability**
Mean: **20**
Standard deviation: **4**
Input constant: **18.5** OK

Remember to get an interval on the normal, take the difference between two cumulatives. So also run the cumulative probability for 17.5 then take the difference.

P(x<18.5) =	0.3538	compare this to the binomial P(x≤18) = .3621
P(x<17.5)=	0.2660	
P(17.5<x<18.5)=	0.0978	compare to binomial P(x=18) = .0909

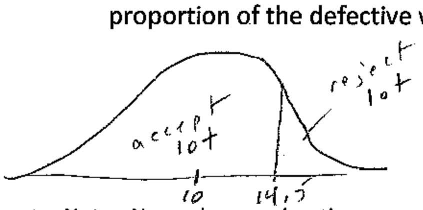
Are the above normal approximations reasonable?

yes – the difference is very small

Step 9. One more problem

Work the following problem using the normal distribution to estimate the binomial. Compare your answer to the problem in Step 6

A microchip plant purchases large lots of silicon wafers. 200 are taken from the lot and inspected. If the number of defective wafers is greater than 14, the lot is rejected. Find the approximate probability of rejection of the lot if the proportion of the defective wafers in the lot is 5%. $np = (200)(.05) = 10$ $\sigma = \sqrt{200(.05)(1-.05)} = 3.082$



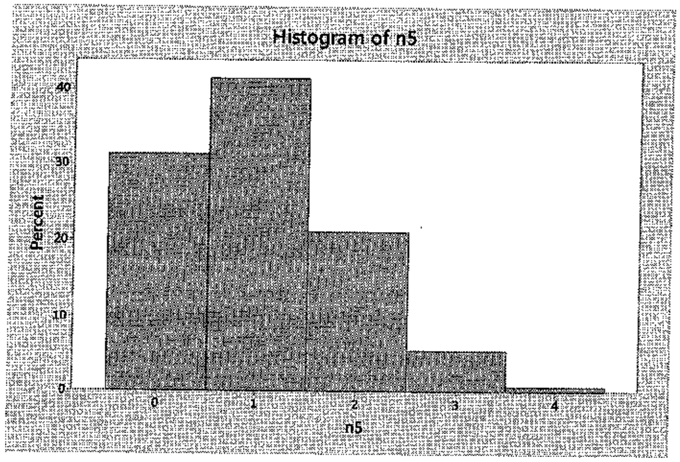
$z = \frac{14.5-10}{3.082} = 1.46$

from z table $P(z<1.46) = .9279$ $1-.9279 = .0721$

Note: Normal approximations are not that common any more with use of statistical software. But normal approximations for binomial are used in determining confidence intervals and in hypothesis testing, so it is important to understand the underlying assumptions demonstrated.

Step 2

Histogram of n5



Step 4

Probability Density Function

```
Binomial with n = 5 and p = 0.2

x  P( X = x )
0     0.32768
```

Cumulative Distribution Function

```
Binomial with n = 5 and p = 0.2

x  P( X ≤ x )
2     0.94208
```

Inverse Cumulative Distribution Function

```
Binomial with n = 5 and p = 0.2

x  P( X ≤ x )        x  P( X ≤ x )
1     0.73728        2     0.94208
```

Step 5

Probability Density Function

```
Binomial with n = 8 and p = 0.1

x  P( X = x )
0    0.430467
```

Probability Density Function

```
Binomial with n = 8 and p = 0.1

x  P( X = x )
2    0.148803
```

Cumulative Distribution Function

```
Binomial with n = 8 and p = 0.1

x  P( X ≤ x )
1    0.813105
```

Step 6

Probability Density Function

Binomial with n = 8 and p = 0.07

x	P(X = x)
1	0.336952

Cumulative Distribution Function

Binomial with n = 8 and p = 0.07

x	P(X ≤ x)
2	0.985301

Cumulative Distribution Function

Binomial with n = 200 and p = 0.05

x	P(X ≤ x)
14	0.921866

Inverse Cumulative Distribution Function

Binomial with n = 200 and p = 0.05

x	P(X ≤ x)	x	P(X ≤ x)
14	0.921866	15	0.955644

Step 7

Probability Density Function

Binomial with n = 100 and p = 0.2

x	P(X = x)
18	0.0908981

Cumulative Distribution Function

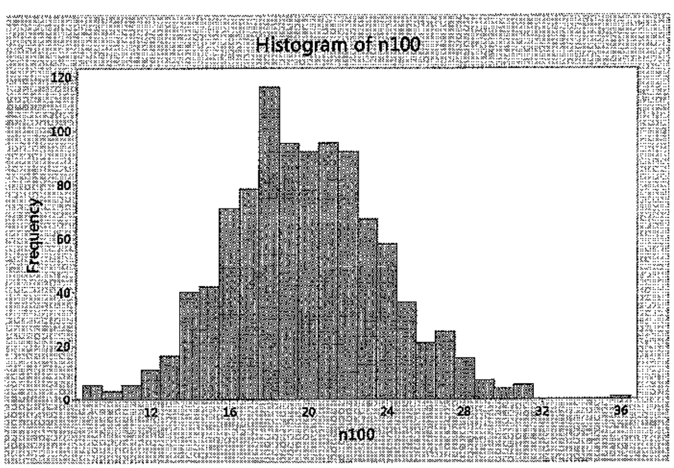
Binomial with n = 100 and p = 0.2

x	P(X ≤ x)
18	0.362087

Descriptive Statistics: n100

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
n100	1000	0	19.862	0.124	3.917	9.000	17.000	20.000	22.000	36.000

Histogram of n100



Step 8

Cumulative Distribution Function

Normal with mean = 20 and standard deviation = 4

x	$P(X \le x)$
18.5	0.353830

Cumulative Distribution Function

Normal with mean = 20 and standard deviation = 4

x	$P(X \le x)$
17.5	0.265986

Step 9

Cumulative Distribution Function

Normal with mean = 10 and standard deviation = 3.082

x	$P(X \le x)$
14.5	0.927867