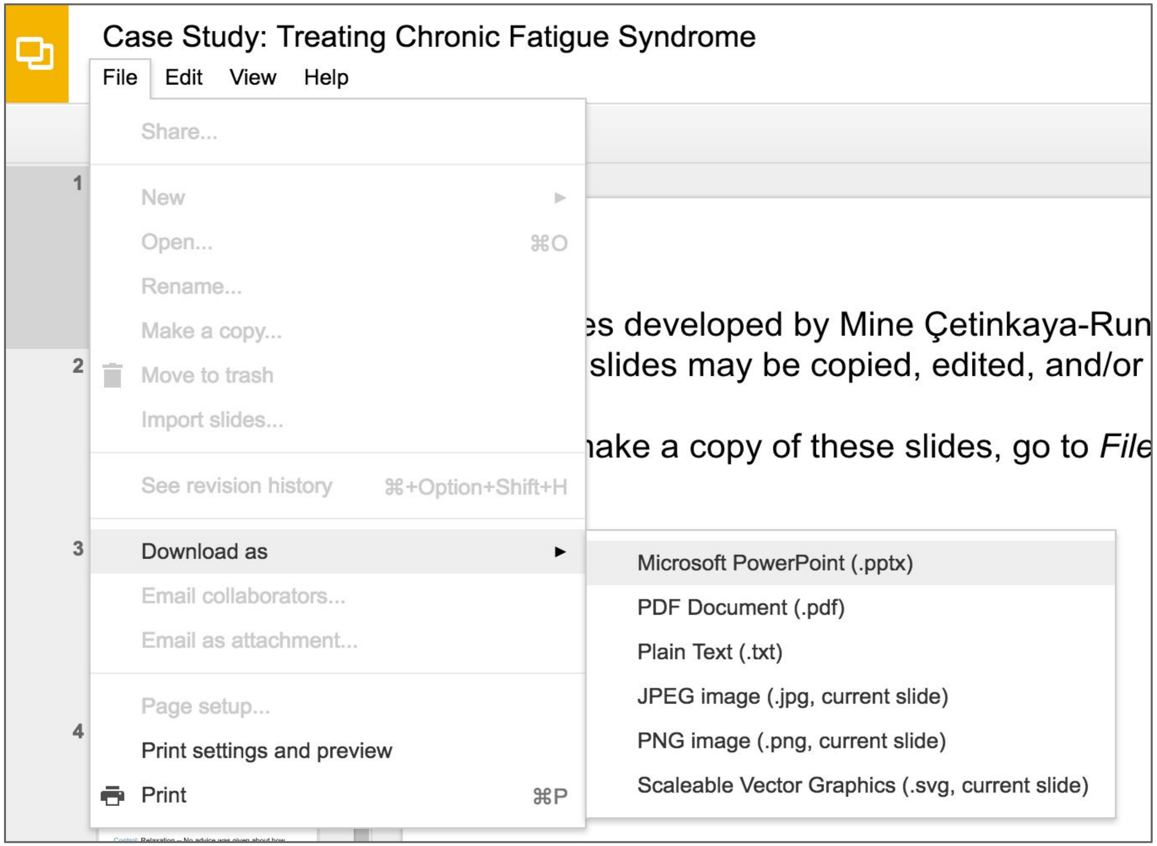
Slides developed by Mine Çetinkaya-Rundel of OpenIntro
Translated from LaTeX to Google Slides by Curry W. Hilton of OpenIntro.

To make a copy of these slides, go to *File > Download as > [option]*, as shown below. Or if you are logged into a Google account, you can choose *Make a copy...* to create your own version in Google Drive.



Binomial distribution

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of “exactly 1 of them refuses to administer the shock”:

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of “exactly 1 of them refuses to administer the shock”:

Scenario 1: $\underset{\text{(A) } \textit{refuse}}{\underline{0.35}} \times \underset{\text{(B) shock}}{\underline{0.65}} \times \underset{\text{(C) shock}}{\underline{0.65}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of “exactly 1 of them refuses to administer the shock”:

Scenario 1: $\underset{\text{(A) } \textit{refuse}}{\underline{0.35}} \times \underset{\text{(B) shock}}{\underline{0.65}} \times \underset{\text{(C) shock}}{\underline{0.65}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$

Scenario 2: $\underset{\text{(A) shock}}{\underline{0.65}} \times \underset{\text{(B) } \textit{refuse}}{\underline{0.35}} \times \underset{\text{(C) shock}}{\underline{0.65}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of “exactly 1 of them refuses to administer the shock”:

Scenario 1: $$\underset{\text{(A) } \textit{refuse}}{\underline{0.35}} \times \underset{\text{(B) shock}}{\underline{0.65}} \times \underset{\text{(C) shock}}{\underline{0.65}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$$

Scenario 2: $$\underset{\text{(A) shock}}{\underline{0.65}} \times \underset{\text{(B) } \textit{refuse}}{\underline{0.35}} \times \underset{\text{(C) shock}}{\underline{0.65}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$$

Scenario 3: $$\underset{\text{(A) shock}}{\underline{0.65}} \times \underset{\text{(B) shock}}{\underline{0.65}} \times \underset{\text{(C) } \textit{refuse}}{\underline{0.35}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$$

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of "exactly 1 of them refuses to administer the shock":

Scenario 1: $\underbrace{0.35}_{\text{(A) refuse}} \times \underbrace{0.65}_{\text{(B) shock}} \times \underbrace{0.65}_{\text{(C) shock}} \times \underbrace{0.65}_{\text{(D) shock}} = 0.0961$

Scenario 2: $\underbrace{0.65}_{\text{(A) shock}} \times \underbrace{0.35}_{\text{(B) refuse}} \times \underbrace{0.65}_{\text{(C) shock}} \times \underbrace{0.65}_{\text{(D) shock}} = 0.0961$

Scenario 3: $\underbrace{0.65}_{\text{(A) shock}} \times \underbrace{0.65}_{\text{(B) shock}} \times \underbrace{0.35}_{\text{(C) refuse}} \times \underbrace{0.65}_{\text{(D) shock}} = 0.0961$

Scenario 4: $\underbrace{0.65}_{\text{(A) shock}} \times \underbrace{0.65}_{\text{(B) shock}} \times \underbrace{0.65}_{\text{(C) shock}} \times \underbrace{0.35}_{\text{(D) refuse}} = 0.0961$

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of “exactly 1 of them refuses to administer the shock”:

Scenario 1: $\underset{\text{(A) } \textit{refuse}}{\underline{0.35}} \times \underset{\text{(B) shock}}{\underline{0.65}} \times \underset{\text{(C) shock}}{\underline{0.65}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$

Scenario 2: $\underset{\text{(A) shock}}{\underline{0.65}} \times \underset{\text{(B) } \textit{refuse}}{\underline{0.35}} \times \underset{\text{(C) shock}}{\underline{0.65}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$

Scenario 3: $\underset{\text{(A) shock}}{\underline{0.65}} \times \underset{\text{(B) shock}}{\underline{0.65}} \times \underset{\text{(C) } \textit{refuse}}{\underline{0.35}} \times \underset{\text{(D) shock}}{\underline{0.65}} = 0.0961$

Scenario 4: $\underset{\text{(A) shock}}{\underline{0.65}} \times \underset{\text{(B) shock}}{\underline{0.65}} \times \underset{\text{(C) shock}}{\underline{0.65}} \times \underset{\text{(D) } \textit{refuse}}{\underline{0.35}} = 0.0961$

The probability of exactly one 1 of 4 people refusing to administer the shock is the sum of all of these probabilities.

$$0.0961 + 0.0961 + 0.0961 + 0.0961 = 4 \times 0.0961 = 0.3844$$

Binomial distribution

The question from the prior slide asked for the probability of given number of successes, k, in a given number of trials, n, (k = 1 success in n = 4 trials), and we calculated this probability as

# of scenarios x P(single scenario)

Binomial distribution

The question from the prior slide asked for the probability of given number of successes, k, in a given number of trials, n, (k = 1 success in n = 4 trials), and we calculated this probability as

# of scenarios x P(single scenario)

- # of scenarios: there is a less tedious way to figure this out, we'll get to that shortly...

Binomial distribution

The question from the prior slide asked for the probability of given number of successes, *k*, in a given number of trials, *n*, (*k* = 1 success in *n* = 4 trials), and we calculated this probability as

# of scenarios x P(single scenario)

- # of scenarios: there is a less tedious way to figure this out, we'll get to that shortly...
- P(single scenario) = $p^k(1-p)^{n-k}$
  where p is the probability of success to the power of number of successes, probability of failure to the power of number of failures

Binomial distribution

The question from the prior slide asked for the probability of given number of successes, *k*, in a given number of trials, *n*, (*k* = 1 success in *n* = 4 trials), and we calculated this probability as

# of scenarios x P(single scenario)

- # of scenarios: there is a less tedious way to figure this out, we'll get to that shortly...
- P(single scenario) = $p^k(1-p)^{n-k}$
  where p is the probability of success to the power of number of successes, probability of failure to the power of number of failures

The *Binomial distribution* describes the probability of having exactly *k* successes in *n* independent Bernoulli trials with probability of success *p*.

Computing the # of scenarios

Earlier we wrote out all possible scenarios that fit the condition of exactly one person refusing to administer the shock. If n was larger and/or k was different than 1, for example, $n = 9$ and $k = 2$:

Computing the # of scenarios

Earlier we wrote out all possible scenarios that fit the condition of exactly one person refusing to administer the shock. If n was larger and/or k was different than 1, for example, $n = 9$ and $k = 2$:

*RR*SSSSSSS

Computing the # of scenarios

Earlier we wrote out all possible scenarios that fit the condition of exactly one person refusing to administer the shock. If *n* was larger and/or *k* was different than 1, for example, *n* = 9 and *k* = 2:

*RR*SSSSSSS
S*RR*SSSSSS

Computing the # of scenarios

Earlier we wrote out all possible scenarios that fit the condition of exactly one person refusing to administer the shock. If *n* was larger and/or *k* was different than 1, for example, *n* = 9 and *k* = 2:

*RR*SSSSSSS
S*RR*SSSSSS
SS*RR*SSSSS
...
SS*R*SS*R*SSS
...
SSSSSSS*RR*

writing out all possible scenarios would be incredibly tedious and prone to errors.

Computing the # of scenarios

Choose function

The *choose function* is useful for calculating the number of ways to choose k successes in n trials.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Computing the # of scenarios

Choose function

The *choose function* is useful for calculating the number of ways to choose *k* successes in *n* trials.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$k = 1, n = 4: \binom{4}{1} = \frac{4!}{1!(4-1)!} = \frac{4\times3\times2\times1}{1\times(3\times2\times1)} = 4$$

Computing the # of scenarios

Choose function

The *choose function* is useful for calculating the number of ways to choose *k* successes in *n* trials.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$k = 1, n = 4: \binom{4}{1} = \frac{4!}{1!(4-1)!} = \frac{4\times3\times2\times1}{1\times(3\times2\times1)} = 4$$

$$k = 2, n = 9: \binom{9}{2} = \frac{9!}{2!(9-2)!} = \frac{9\times8\times7!}{2\times1\times7!} = \frac{72}{2} = 36$$

Note: You can also use R for these calculations:

```
> choose(9,2)
[1] 36
```

Practice

Which of the following is false?

(a) There are n ways of getting 1 success in n trials, $\binom{n}{1} = n$.

(b) There is only 1 way of getting n successes in n trials, $\binom{n}{n} = 1$.

(c) There is only 1 way of getting n failures in n trials, $\binom{n}{0} = 1$.

(d) There are $n - 1$ ways of getting $n - 1$ successes in n trials, $\binom{n}{n-1} = n - 1$.

Practice

Which of the following is false?

(a) There are n ways of getting 1 success in n trials, $\binom{n}{1} = n$.

(b) There is only 1 way of getting n successes in n trials, $\binom{n}{n} = 1$.

(c) There is only 1 way of getting n failures in n trials, $\binom{n}{0} = 1$.

(d) *There are $n-1$ ways of getting $n-1$ successes in n trials, $\binom{n}{n-1} = n-1$.*

Binomial distribution (cont.)

Binomial probabilities

If p represents probability of success, (1-p) represents probability of failure, n represents number of independent trials, and k represents number of successes

$$P(k\ successes\ in\ n\ trials) = \binom{n}{k} p^k\ (1-p)^{(n-k)}$$

Practice

Which of the following is not a condition that needs to be met for the binomial distribution to be applicable?

(a) the trials must be independent
(b) the number of trials, n, must be fixed
(c) each trial outcome must be classified as a *success* or a *failure*
(d) the number of desired successes, k, must be greater than the number of trials
(e) the probability of success, p, must be the same for each trial

Practice

Which of the following is not a condition that needs to be met for the binomial distribution to be applicable?

(a) the trials must be independent
(b) the number of trials, n, must be fixed
(c) each trial outcome must be classified as a *success* or a *failure*
(d) the number of desired successes, k, must be greater than the number of trials
(e) the probability of success, p, must be the same for each trial

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

(a) pretty high
(b) pretty low

Gallup: http://www.gallup.com/poll/160061/obesity-rate-stable-2012.aspx, January 23, 2013.

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

(a) pretty high
(b) pretty low

Gallup: http://www.gallup.com/poll/160061/obesity-rate-stable-2012.aspx, January 23, 2013.

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

(a) $0.262^8 \times 0.738^2$

(b) $\binom{8}{10} \times 0.262^8 \times 0.738^2$

(c) $\binom{10}{8} \times 0.262^8 \times 0.738^2$

(d) $\binom{10}{8} \times 0.262^2 \times 0.738^8$

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

(a) $0.262^8 \times 0.738^2$

(b) $\binom{8}{10} \times 0.262^8 \times 0.738^2$

(c) $\binom{10}{8} \times 0.262^8 \times 0.738^2 = 45 \times 0.262^8 \times 0.738^2 = 0.0005$

(d) $\binom{10}{8} \times 0.262^2 \times 0.738^8$

The birthday problem

What is the probability that 2 randomly chosen people share a birthday?

The birthday problem

What is the probability that 2 randomly chosen people share a birthday?

Pretty low, $1 / 365 \approx 0.0027$

The birthday problem

What is the probability that 2 randomly chosen people share a birthday?

Pretty low, $1 / 365 \approx 0.0027$

What is the probability that at least 2 people out of 366 people share a birthday?

The birthday problem

What is the probability that 2 randomly chosen people share a birthday?

Pretty low, $1 / 365 \approx 0.0027$

What is the probability that at least 2 people out of 366 people share a birthday?

Exactly 1! (Excluding the possibility of a leap year birthday.)

The birthday problem (cont.)

What is the probability that at least 2 people (1 match) out of 121 people share a birthday?

The birthday problem (cont.)

What is the probability that at least 2 people (1 match) out of 121 people share a birthday?

Somewhat complicated to calculate, but we can think of it as the complement of the probability that there are no matches in 121 people.

$$P(no\ matches) = 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right)$$

The birthday problem (cont.)

What is the probability that at least 2 people (1 match) out of 121 people share a birthday?

Somewhat complicated to calculate, but we can think of it as the complement of the probability that there are no matches in 121 people.

$$\begin{aligned} P(no\ matches) &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\ &= \frac{365 \times 364 \times \cdots \times 245}{365^{121}} \end{aligned}$$

The birthday problem (cont.)

What is the probability that at least 2 people (1 match) out of 121 people share a birthday?

Somewhat complicated to calculate, but we can think of it as the complement of the probability that there are no matches in 121 people.

$$
\begin{aligned}
P(no\ matches) &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\
&= \frac{365 \times 364 \times \cdots \times 245}{365^{121}} \\
&= \frac{365!}{365^{121} \times (365 - 121)!}
\end{aligned}
$$

The birthday problem (cont.)

What is the probability that at least 2 people (1 match) out of 121 people share a birthday?

Somewhat complicated to calculate, but we can think of it as the complement of the probability that there are no matches in 121 people.

$$
\begin{aligned}
P(no\ matches) &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\
&= \frac{365 \times 364 \times \cdots \times 245}{365^{121}} \\
&= \frac{365!}{365^{121} \times (365 - 121)!} \\
&= \frac{121! \times \binom{365}{121}}{365^{121}}
\end{aligned}
$$

The birthday problem (cont.)

What is the probability that at least 2 people (1 match) out of 121 people share a birthday?

Somewhat complicated to calculate, but we can think of it as the complement of the probability that there are no matches in 121 people.

$$
\begin{aligned}
P(\textit{no matches}) &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\
&= \frac{365 \times 364 \times \cdots \times 245}{365^{121}} \\
&= \frac{365!}{365^{121} \times (365 - 121)!} \\
&= \frac{121! \times \binom{365}{121}}{365^{121}} \approx 0
\end{aligned}
$$

The birthday problem (cont.)

What is the probability that at least 2 people (1 match) out of 121 people share a birthday?

Somewhat complicated to calculate, but we can think of it as the complement of the probability that there are no matches in 121 people.

$$
\begin{aligned}
P(no\ matches) &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{120}{365}\right) \\
&= \frac{365 \times 364 \times \cdots \times 245}{365^{121}} \\
&= \frac{365!}{365^{121} \times (365 - 121)!} \\
&= \frac{121! \times \binom{365}{121}}{365^{121}} \approx 0
\end{aligned}
$$

$$P(at\ least\ 1\ match) \approx 1$$

Expected value

A 2012 Gallup survey suggests that 26.2% of Americans are obese.

Among a random sample of 100 Americans, how many would you expect to be obese?

Expected value

A 2012 Gallup survey suggests that 26.2% of Americans are obese.

Among a random sample of 100 Americans, how many would you expect to be obese?

- Easy enough, 100 x 0.262 = 26.2.

Expected value

A 2012 Gallup survey suggests that 26.2% of Americans are obese.

Among a random sample of 100 Americans, how many would you expect to be obese?

- Easy enough, 100 x 0.262 = 26.2.
- Or more formally, $\mu = np$ = 100 x 0.262 = 26.2.

Expected value

A 2012 Gallup survey suggests that 26.2% of Americans are obese.

Among a random sample of 100 Americans, how many would you expect to be obese?

- Easy enough, 100 x 0.262 = 26.2.
- Or more formally, $\mu = np = 100 \times 0.262 = 26.2$.
- But this doesn't mean in every random sample of 100 people exactly 26.2 will be obese. In fact, that's not even possible. In some samples this value will be less, and in others more. How much would we expect this value to vary?

Expected value and its variability

Mean and standard deviation of binomial distribution

$$\mu = np \qquad \sigma = \sqrt{np(1-p)}$$

Expected value and its variability

Mean and standard deviation of binomial distribution

$$\mu = np \qquad\qquad \sigma = \sqrt{np(1-p)}$$

Going back to the obesity rate:

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.262 \times 0.738} \approx 4.4$$

Expected value and its variability

Mean and standard deviation of binomial distribution

$$\mu = np \qquad\qquad \sigma = \sqrt{np(1-p)}$$

Going back to the obesity rate:

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.262 \times 0.738} \approx 4.4$$

We would expect 26.2 out of 100 randomly sampled Americans to be obese, with a standard deviation of 4.4.

Note: Mean and standard deviation of a binomial might not always be whole numbers, and that is alright, these values represent what we would expect to see on average.

Unusual observations

Using the notion that *observations that are more than 2 standard deviations away from the mean are considered unusual* and the mean and the standard deviation we just computed, we can calculate a range for the plausible number of obese Americans in random samples of 100.

$$26.2 \pm (2 \times 4.4) \rightarrow (17.4, 35.0)$$

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

(a) Yes (b) No

	Excellent	Good	Only fair	Poor	Total excellent/ good
	%	%	%	%	%
Independent private school	31	47	13	2	78
Parochial or church-related schools	21	48	18	5	69
Charter schools	17	43	23	5	60
Home schooling	13	33	30	14	46
Public schools	5	32	42	19	37

Gallup, Aug. 9-12, 2012

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

(a) Yes (b) No

$$\mu = np = 1{,}000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1{,}000 \times 0.13 \times 0.87} \approx 10.6$$

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

(a) Yes (b) No

$$\mu = np = 1,000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1,000 \times 0.13 \times 0.87} \approx 10.6$$

Method 1: *Range of usual observations:* $130 \pm 2 \times 10.6 = (108.8, 151.2)$
100 is outside this range, so would be considered unusual.

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

(a) Yes (b) No

$$\mu = np = 1,000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1,000 \times 0.13 \times 0.87} \approx 10.6$$

Method 1: *Range of usual observations:* $130 \pm 2 \times 10.6 = (108.8, 151.2)$
100 is outside this range, so would be considered unusual.

Method 2: *Z-score of observation:* $Z = \frac{x - mean}{SD} = \frac{100-130}{10.6} = -2.83$
100 is more than 2 SD below the mean, so would be considered unusual.

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-children.aspx

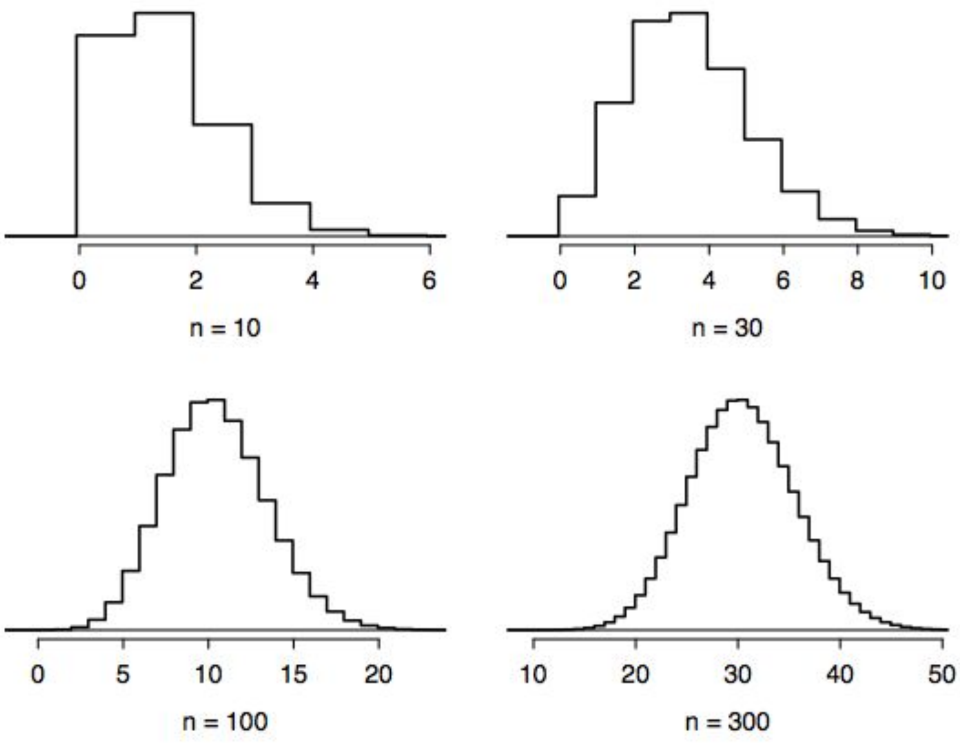
Shapes of binomial distributions

For this activity you will use a web applet. Go to http://socr.stat.ucla.edu/htmls/SOCR_Experiments.html and choose Binomial coin experiment in the drop down menu on the left.

- Set the number of trials to 20 and the probability of success to 0.15. Describe the shape of the distribution of number of successes.
- Keeping p constant at 0.15, determine the minimum sample size required to obtain a unimodal and symmetric distribution of number of successes. Please submit only one response per team.
- Further considerations:
    - What happens to the shape of the distribution as n stays constant and p changes?
    - What happens to the shape of the distribution as p stays constant and n changes?

Distributions of number of successes

Hollow histograms of samples from the binomial model where $p = 0.10$ and $n = 10$, 30, 100, and 300. What happens as n increases?



Low large is large enough?

The sample size is considered large enough if the expected number of successes and failures are both at least 10.

$$np \geq 10 \qquad \text{and} \qquad n(1 - p) \geq 10$$

Low large is large enough?

The sample size is considered large enough if the expected number of successes and failures are both at least 10.

$$np \geq 10 \qquad \text{and} \qquad n(1 - p) \geq 10$$

$$10 \times 0.13 \approx 1.3$$

$$10 \times (1 - 0.13) = 8.7$$

Practice

Below are four pairs of Binomial distribution parameters. Which distribution can be approximated by the normal distribution?

(a) $n = 100$, $p = 0.95$
(b) $n = 25$, $p = 0.45$
(c) $n = 150$, $p = 0.05$
(d) $n = 500$, $p = 0.015$

Practice

Below are four pairs of Binomial distribution parameters. Which distribution can be approximated by the normal distribution?

(a) $n = 100$, $p = 0.95$
(b) $n = 25$, $p = 0.45 \longrightarrow 25 \times 0.45 = 11.25$, $25 \times 0.55 = 13.75$
(c) $n = 150$, $p = 0.05$
(d) $n = 500$, $p = 0.015$

An analysis of Facebook users

A recent study found that “Facebook users get more than they give”. For example:

1. 40% of Facebook users in our sample made a friend request, but 63% received at least one request
2. Users in our sample pressed the like button next to friends' content an average of 14 times, but had their content ``liked" an average of 20 times
3. Users sent 9 personal messages, but received 12
4. 12% of users tagged a friend in a photo, but 35% were themselves tagged in a photo

Any guesses for how this pattern can be explained?

http://www.pewinternet.org/Reports/2012/Facebook-users/Summary.aspx

An analysis of Facebook users

A recent study found that “Facebook users get more than they give”. For example:

1. 40% of Facebook users in our sample made a friend request, but 63% received at least one request
2. Users in our sample pressed the like button next to friends' content an average of 14 times, but had their content ``liked" an average of 20 times
3. Users sent 9 personal messages, but received 12
4. 12% of users tagged a friend in a photo, but 35% were themselves tagged in a photo

Any guesses for how this pattern can be explained?

Power users contribute much more content than the typical user.

http://www.pewinternet.org/Reports/2012/Facebook-users/Summary.aspx

Practice

This study also found that approximately 25% of Facebook users are considered power users. The same study found that the average Facebook user has 245 friends. What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users? Note any assumptions you must make.

We are given that $n = 245$, $p = 0.25$, and we are asked for the probability $P(K \geq 70)$. To proceed, we need independence, which we'll assume but could check if we had access to more Facebook data.

Practice

This study also found that approximately 25% of Facebook users are considered power users. The same study found that the average Facebook user has 245 friends. What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users? Note any assumptions you must make.

We are given that $n = 245$, $p = 0.25$, and we are asked for the probability $P(K \geq 70)$. To proceed, we need independence, which we'll assume but could check if we had access to more Facebook data.

$$
\begin{aligned}
P(X \geq 70) &= P(K = 70 \text{ or } K = 71 \text{ or } K = 72 \text{ or } \ldots \text{ or } K = 245) \\
&= P(K = 70) + P(K = 71) + P(K = 72) + \ldots + P(K = 245)
\end{aligned}
$$

Practice

This study also found that approximately 25% of Facebook users are considered power users. The same study found that the average Facebook user has 245 friends. What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users? Note any assumptions you must make.

We are given that $n = 245$, $p = 0.25$, and we are asked for the probability $P(K \geq 70)$. To proceed, we need independence, which we'll assume but could check if we had access to more Facebook data.

$$\begin{aligned} P(X \geq 70) &= P(K = 70 \text{ or } K = 71 \text{ or } K = 72 \text{ or } \ldots \text{ or } K = 245) \\ &= P(K = 70) + P(K = 71) + P(K = 72) + \ldots + P(K = 245) \end{aligned}$$

This seems like an awful lot of work...

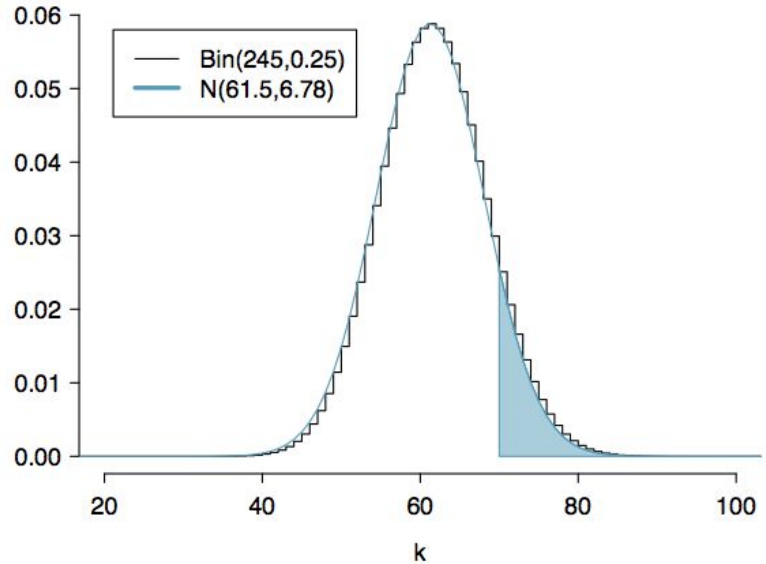
Normal approximation to the binomial

When the sample size is large enough, the binomial distribution with parameters n and p can be approximated by the normal model with parameters $\mu = np$ and $\sigma = \sqrt{np(1-p)}$.

- In the case of the Facebook power users, $n = 245$ and $p = 0.25$.

$$\mu = 245 \times 0.25 = 61.25 \qquad \sigma = \sqrt{245 \times 0.25 \times 0.75} = 6.78$$

- $Bin(n = 245, p = 0.25) \approx N(\mu = 61.25, \sigma = 6.78)$.

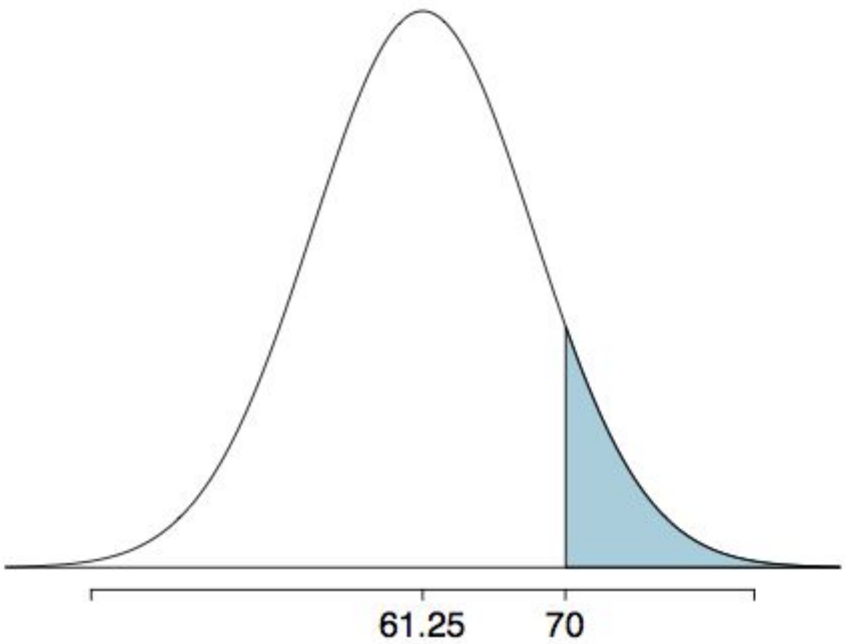


Practice

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?

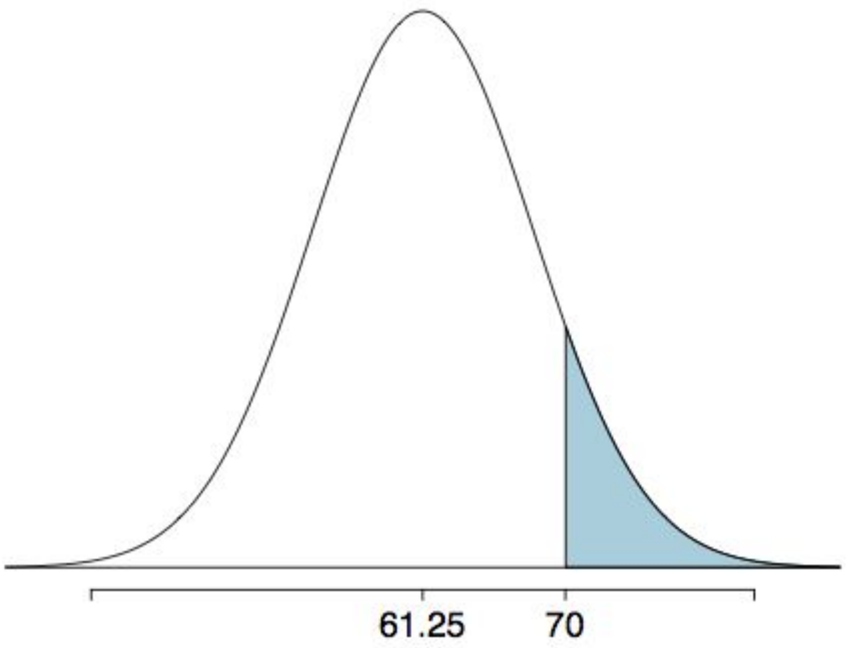
Practice

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?



Practice

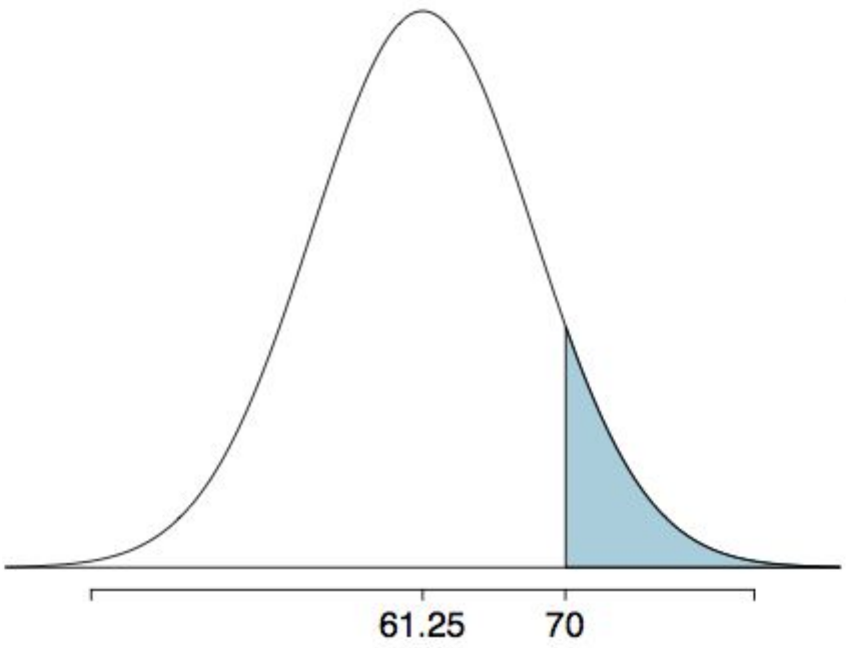
What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?



$$Z = \frac{obs - mean}{SD} = \frac{70 - 61.25}{6.78} = 1.29$$

Practice

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?



$$Z = \frac{obs - mean}{SD} = \frac{70 - 61.25}{6.78} = 1.29$$

$$P(Z > 1.29) = 1 - 0.9015 = 0.0985$$

The normal approximation breaks down on small intervals

- The normal approximation to the binomial distribution tends to perform poorly when estimating the probability of a small range of counts, even when the conditions are met.

- This approximation for intervals of values is usually improved if cutoff values are extended by 0.5 in both directions.

- The tip to add extra area when applying the normal approximation is most often useful when examining a range of observations. While it is possible to also apply this correction when computing a tail area, the benefit of the modification usually disappears since the total interval is typically quite wide.

Find more resources at openintro.org/os, including

- Slides
- Videos
- Statistical Software Labs
- Discussion Forums (free support for students and teachers)
- Learning Objectives

Teachers only content is also available for Verified Teachers, including

- Exercise solutions
- Sample exams
- Ability to request a free desk copy for a course
- Statistics Teachers email group

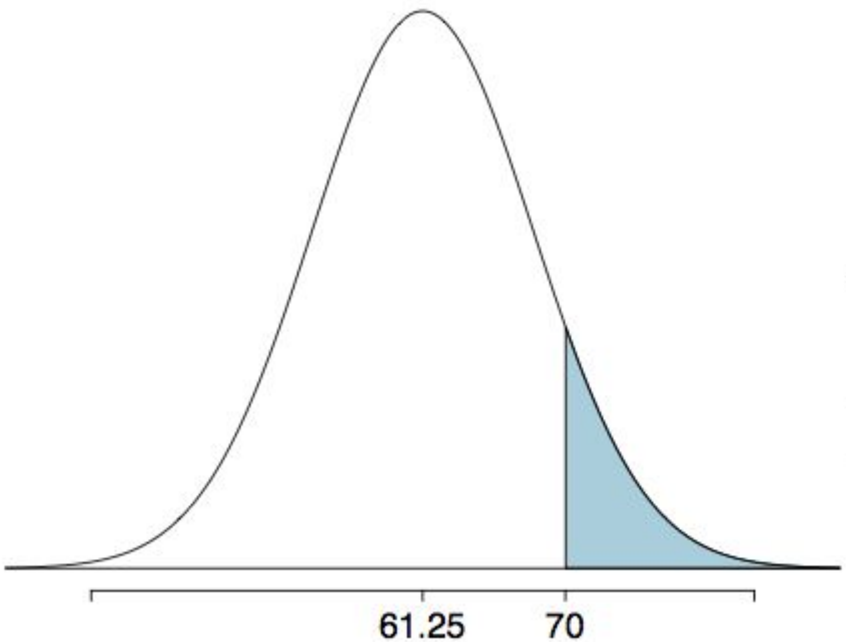
Questions? Contact us.

Appendix

Probability Tables

Practice

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?

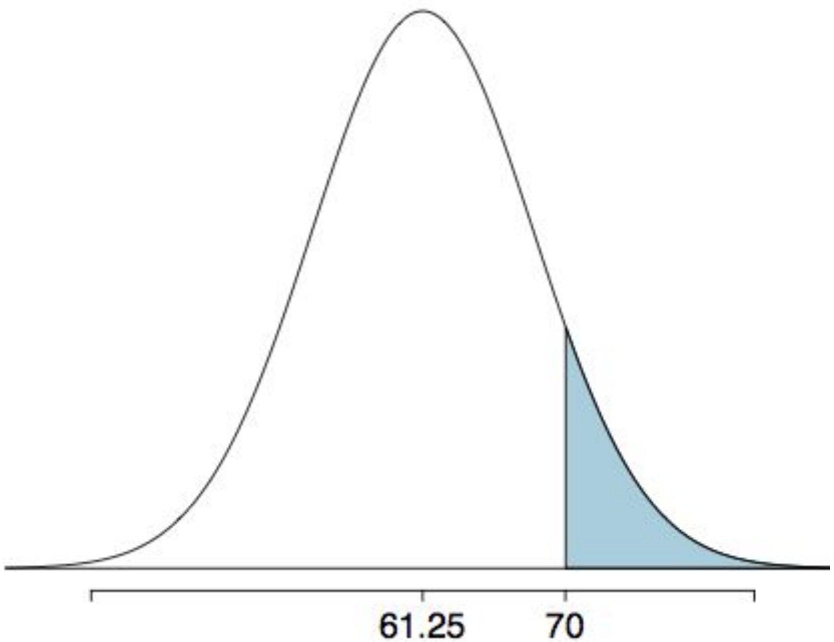


$$Z = \frac{obs - mean}{SD} = \frac{70 - 61.25}{6.78} = 1.29$$

	Second decimal place of Z				
Z	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015

Practice

What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?



$$Z = \frac{obs - mean}{SD} = \frac{70 - 61.25}{6.78} = 1.29$$

	Second decimal place of Z				
Z	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015

$$P(Z > 1.29) = 1 - 0.9015 = 0.0985$$