Assessing Discrete Data: Binomial Distribution

Dr. Ab Mosca (they/them)

Slides based off slides courtesy of OpenIntro and John McGreedy of Johns Hopkins University

Scenario

If a person randomly guesses the answers to 10 multiple choice questions, we can ask questions like:

- What is the probability that they get none right?
- What is the probability that they get all ten right?
- What is the probability that they get at least three right?
- How many do they get right on average?

https://www3.nd.edu/~dgalvin1/10120/10120_S16/Topic19_8p6_Galvin.pdf

Scenario

If a person randomly guesses the answers to 10 multiple choice questions, we can ask questions like:

- What is the probability that they get none right?
- What is the probability that they get all ten right?
- What is the probability that they get at least three right?
- How many do they get right on average?

These are examples of Bernoulli Experiments

https://www3.nd.edu/~dgalvin1/10120/10120_S16/Topic19_8p6_Galvin.pdf

Bernoulli Experiments

- The experiment consists of trials, repeated *n* times.
- The outcome of each trial is a *success* or a *failure*; and these are the only two possible outcomes for each trial.
- The probability of success remains the same for each trial. We use *p* for the probability of success (on each trial) and *q* = *1* - *p* for the probability of failure.
- The trials are independent (the outcome of previous trials has no influence on the outcome of the next trial).
- We are interested in the random variable X where X = the number of successes. Note the possible values of X are 0, 1, 2, 3, …, n.

https://www3.nd.edu/~dgalvin1/10120/10120_S16/Topic19_8p6_Galvin.pdf

Experiment

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

a) What is a trial?

Experiment

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

a) What is a trial?
   - **one guess**

Experiment

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

a) What is a trial?
   - **one guess**

b) What is success and what is failure?

Experiment

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

a) What is a trial?
   - **one guess**

b) What is success and what is failure?
   - **correct answer == success, wrong answer == failure**

Experiment

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

a) What is a trial?
   - **one guess**

b) What is success and what is failure?
   - **correct answer == success, wrong answer == failure**

c) What is the probability of success (p)?

Experiment

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

a) What is a trial?
   - **one guess**

b) What is success and what is failure?
   - **correct answer == success, wrong answer == failure**

c) What is the probability of success (p)?
   - **1 / 4 = 0.25**

Experiment

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

a) What is a trial?
   - **one guess**

b) What is success and what is failure?
   - **correct answer == success, wrong answer == failure**

c) What is the probability of success (p)?
   - **1 / 4 = 0.25**

d) What is the probability of failure (q)?

Experiment

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

a) What is a trial?
   - **one guess**

b) What is success and what is failure?
   - **correct answer == success, wrong answer == failure**

c) What is the probability of success (p)?
   - **1 / 4 = 0.25**

d) What is the probability of failure (q)?
   - **1 – p = 1 – 0.25 = 0.75**

Experiment

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

a) What is a trial?
   - **one guess**

b) What is success and what is failure?
   - **correct answer == success, wrong answer == failure**

c) What is the probability of success (p)?
   - **1 / 4 = 0.25**

d) What is the probability of failure (q)?
   - **1 – p = 1 – 0.25 = 0.75**

e) What values can X take?

Experiment

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

a) What is a trial?
   - **one guess**

b) What is success and what is failure?
   - **correct answer == success, wrong answer == failure**

c) What is the probability of success (p)?
   - **1 / 4 = 0.25**

d) What is the probability of failure (q)?
   - **1 – p = 1 – 0.25 = 0.75**

e) What values can X take?
   - **0, 1, 2, . . . , 10**

Another Experiment

A person flips a coin 12 times. We record the number of heads, and ask what is the probability of getting 3 heads?

a) What is a trial?
   - ○

b) What is success and what is failure?

c) What is the probability of success (p)?

d) What is the probability of failure (q)?

e) What values can X take?
   - ○

Another Experiment

A person flips a coin 12 times. We record the number of heads, and ask what is the probability of getting 3 heads?

a) What is a trial?
   - **one coin flip**

b) What is success and what is failure?
   - **heads == success, tails == failure**

c) What is the probability of success (p)?
   - **1 / 2 = 0.5**

d) What is the probability of failure (q)?
   - **1 – p = 1 – 0.5 = 0.5**

e) What values can X take?
   - **0, 1, 2, . . . , 12**

Another Experiment

A person rolls a die 5 times. We record the number of times 1 is rolled, and ask what is the probability of rolling 1 three times?

a) What is a trial?

b) What is success and what is failure?

c) What is the probability of success (p)?

d) What is the probability of failure (q)?

e) What values can X take?

Another Experiment

A person rolls a die 5 times. We record the number of times 1 is rolled, and ask what is the probability of rolling 1 three times?

a) What is a trial?
   - **one roll**

b) What is success and what is failure?
   - **1 == success, all other numbers == failure**

c) What is the probability of success (p)?
   - **1 / 6 = 0.17**

d) What is the probability of failure (q)?
   - **1 – p = 1 – 0.17 = 0.83**

e) What values can X take?
   - **0, 1, 2, 3, 4, 5**

Bernoulli Experiments

Typically we are interested in the number of successes among all of our trials.

We call the number of successes k .

Bernoulli Experiments

Typically we are interested in the number of successes among all of our trials.

We call the number of successes k .

Ex. A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

Ex. A person flips a coin 12 times. We record the number of heads, and ask what is the probability of getting 3 heads?

Ex. A person rolls a die 5 times. We record the number of times 1 is rolled, and ask what is the probability of rolling 1 three times?

Binomial distribution

The questions from the prior slides asked for the probability of a given number of successes, k, in a given number of trials, n, we calculated this probability as

# of scenarios x P(single scenario)

Binomial distribution

The questions from the prior slides asked for the probability of a given number of successes, *k*, in a given number of trials, *n*, we calculated this probability as

# of scenarios x P(single scenario)

- # of scenarios – next slide

- P(single scenario) = $p^k(1-p)^{n-k}$
  where p is the probability of success to the power of number of successes

Binomial distribution

The questions from the prior slides asked for the probability of a given number of successes, *k*, in a given number of trials, *n*, we calculated this probability as

# of scenarios x P(single scenario)

- # of scenarios – next slide

- P(single scenario) = $p^k(1-p)^{n-k}$
  where p is the probability of success to the power of number of successes

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

Computing the # of scenarios

Combinations

A *combination* 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

Combinations

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

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

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

Computing the # of scenarios

Combinations

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

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

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

n

k

Computing the # of scenarios

Combinations

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

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

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

n

k

$$\binom{10}{8} = \frac{10!}{8!(10-8)!} = \frac{10*9*8*\cdots*2*1}{8*7*\cdots*2*1(2*1)} = \frac{10*9*\cancel{8*\cdots*2*1}}{\cancel{8*7*\cdots*2*1}(2*1)} = 45$$

Computing the # of scenarios

Combinations

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

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

use a calculator or type “10 choose 8” into google

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

n

k

$$\binom{10}{8} = \frac{10!}{8!(10-8)!} = \frac{10*9*8*\cdots*2*1}{8*7*\cdots*2*1(2*1)} = \frac{10*9*\cancel{8*\cdots*2*1}}{\cancel{8*7*\cdots*2*1}(2*1)} = 45$$

Computing the # of scenarios

Combinations

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

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

A person flips a coin 12 times. We record the number of heads, and ask what is the probability of getting 3 heads?

Computing the # of scenarios

Combinations

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

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

n

A person flips a coin 12 times. We record the number of heads, and ask what is the probability of getting 3 heads?

k

$$\binom{12}{3} = 220$$

Computing the # of scenarios

Combinations

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

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

A person rolls a die 5 times. We record the number of times 1 is rolled, and ask what is the probability of rolling 1 three times?

Computing the # of scenarios

Combinations

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

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

n

A person rolls a die 5 times. We record the number of times 1 is rolled, and ask what is the probability of rolling 1 three times?

k

$$\binom{5}{3} = 10$$

Practice

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

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

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

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) = # of scenarios x P(single scenario)

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

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

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

n

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

k

$\binom{10}{8} = 45$, p = 0.25, P(k = 8 in 10 trials) = ?

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

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

A person randomly guesses the answers to 10 multiple choice questions (each questions has 4 answers to chose from). We record the number of correct answers, and ask what is the probability of getting 8 answers correct?

n

k

$\binom{10}{8} = 45$, p = 0.25

P(k = 8 in 10 trials) = $\binom{10}{8} * 0.25^8 * (1-0.25)^{(10-8)}$

= 0.00039 -> 0.039%

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

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

n

A person flips a coin 12 times. We record the number of heads, and ask what is the probability of getting 3 heads?

k

$\binom{12}{3} = 220$, p = 0.5

P(k = 3 in 12 trials) = ?

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

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

n

A person flips a coin 12 times. We record the number of heads, and ask what is the probability of getting 3 heads?

k

$\binom{12}{3} = 220$, p = 0.5

P(k = 3 in 12 trials) = $\binom{12}{3} * 0.5^3 * (1-0.5)^{(12-3)}$

= 0.054 -> 5.4%

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

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

n

A person rolls a die 5 times. We record the number of times 1 is rolled, and ask what is the probability of rolling 1 three times?

k

$\binom{5}{3} = 10$, p = 0.17

P(k = 3 in 5 trials) = ?

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

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

n

A person rolls a die 5 times. We record the number of times 1 is rolled, and ask what is the probability of rolling 1 three times?

k

$\binom{5}{3} = 10$, p = 0.17

P(k = 3 in 5 trials) = $\binom{5}{3} * 0.17^3 * (1-0.17)^{(5-3)}$

= 0.034 -> 3.4%

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

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

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

Going back to the multiple choice problems:

$$\sigma = \sqrt{np(1-p)} = \sqrt{10 * 0.25(1-0.25)} = 1.37$$

Expected value and its variability

Mean and standard deviation of binomial distribution

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

Going back to the multiple choice problems:

$$\sigma = \sqrt{np(1-p)} = \sqrt{10 * 0.25(1-0.25)} = 1.37$$

We would expect 2.5 out of 10 problems to be correct, with a standard deviation of 1.37.

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 correct answers in samples of 10.

$$2.5 \pm (2 \times 1.37) \rightarrow (0, 5.24)$$

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

	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

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

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

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

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

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

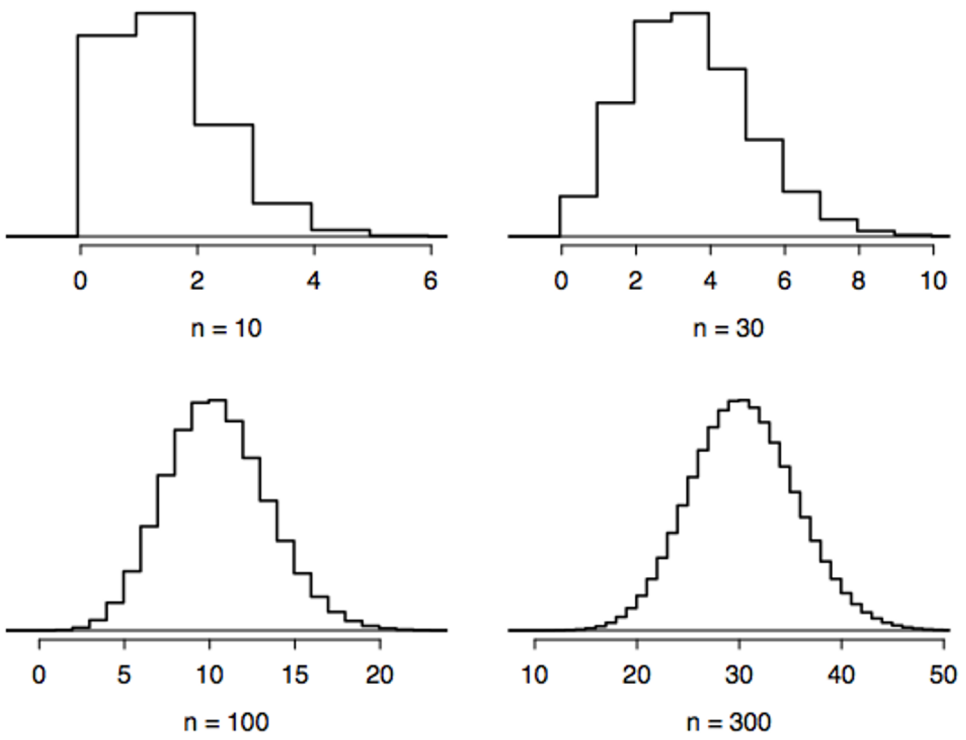
Method 1: *Range of usual observations:* $130 \pm 2 \times 10.6 = (108.8, 151.2)$
100 is outside this range, so would be considered unusual.

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

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?



How large is large enough?

The sample size is considered large enough to be approximated by the normal distribution if the expected number of successes and failures are both at least 10.

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

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) $n = 500$, $p = 0.015$

$np \geq 10$ and $n(1 - p) \geq 10$

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) $n = 25$, $p = 0.45 \rightarrow 25 \times 0.45 = 11.25$, $25 \times 0.55 = 13.75$

(c) $n = 150$, $p = 0.05$

(d) $n = 500$, $p = 0.015$

$np \geq 10$ and $n(1 - p) \geq 10$

Ex. 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

Ex. 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

Ex. Facebook

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.

Ex. Facebook

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.

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

Ex. Facebook

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.

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

This seems like an awful lot of work...

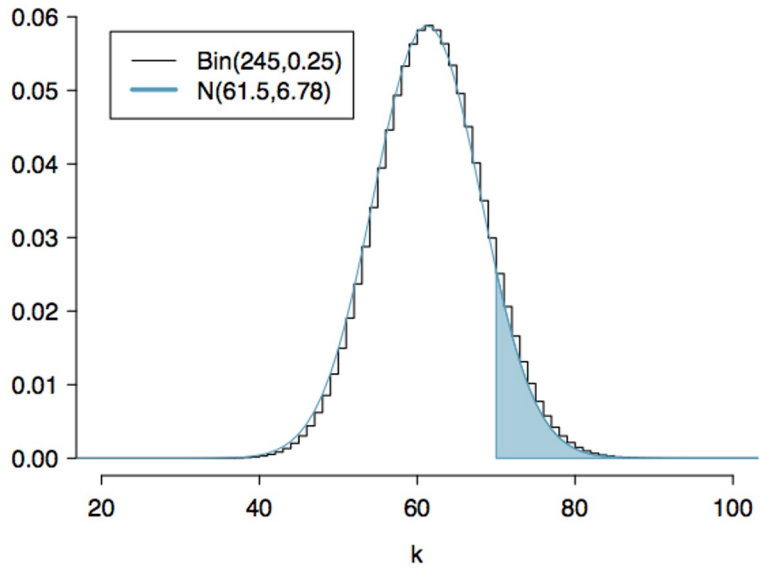
Normal approximation to the binomial

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 $p = 0.25$.

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

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

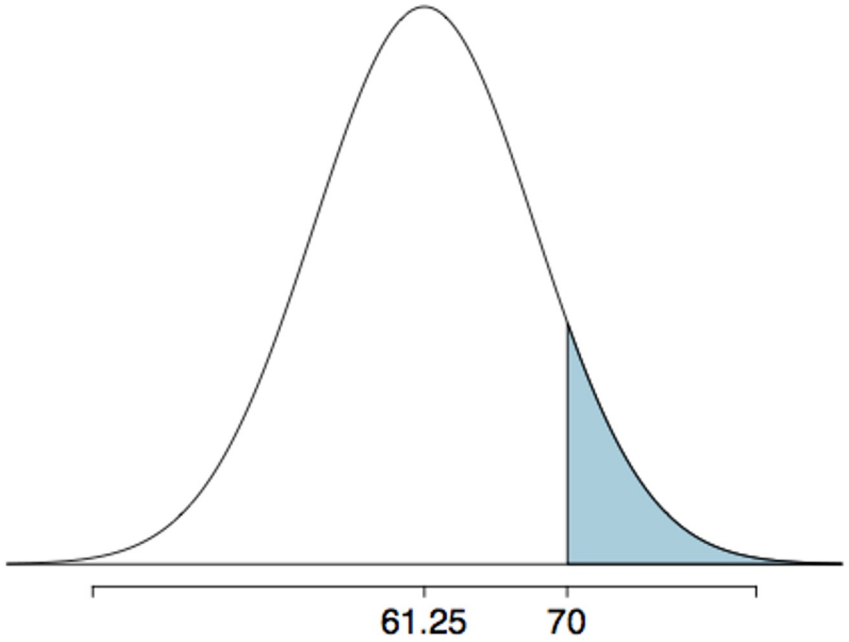


Practice

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

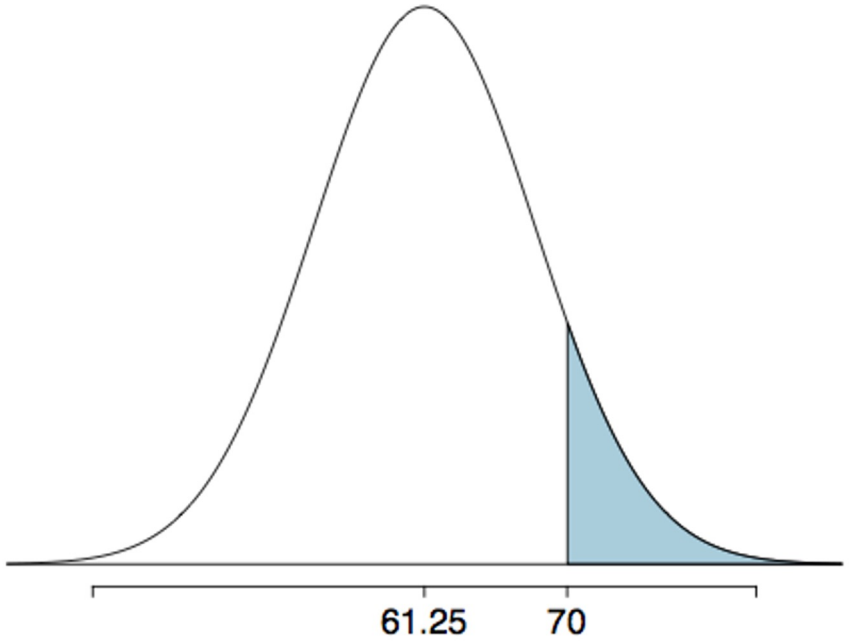
Practice

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



Practice

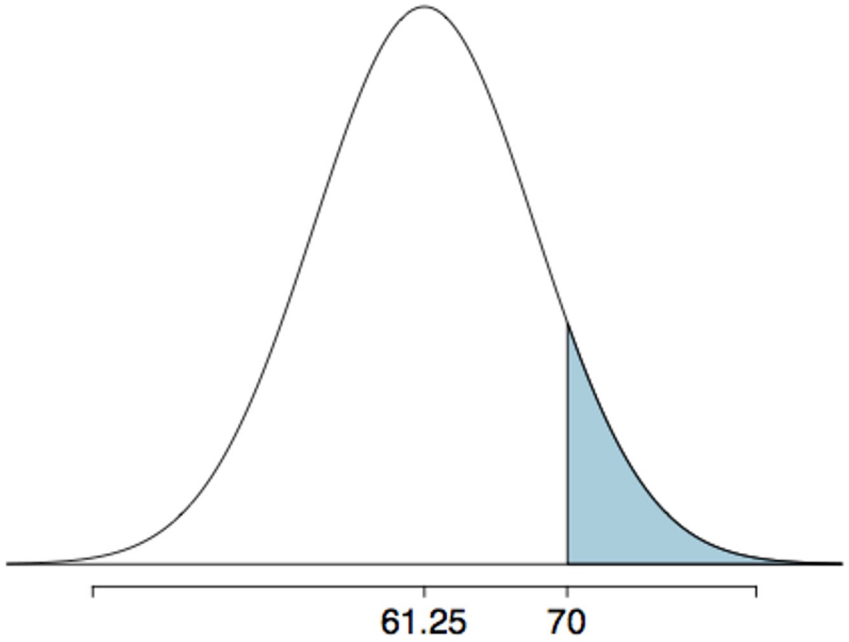
What is the probability that the average Facebook user with 245 friends has 70 or more friends who would be considered power users?



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

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.