Unit 2: Probability and distributions

4. Binomial distribution

Sta 101 - Fall 2015

Duke University, Department of Statistical Science

Dr. Çetinkaya-Rundel

Slides posted at *http://bit.ly/sta101_f15*

1. Housekeeping

2. Main ideas

1. Binomial distribution is used for calculating the probability of exact number of successes for a given number of trials

2. Expected value and standard deviation of the binomial can be calculated using its parameters n and p

3. Shape of the binomial distribution approaches normal when the S-F rule is met

3. Summary

- Project 1 questions?
- PS2 due Friday night, PA2 due Sunday night
- RA3 on Monday

1. Housekeeping

2. Main ideas

1. Binomial distribution is used for calculating the probability of exact number of successes for a given number of trials
2. Expected value and standard deviation of the binomial can be calculated using its parameters n and p
3. Shape of the binomial distribution approaches normal when the S-F rule is met

3. Summary

1. Housekeeping

2. Main ideas

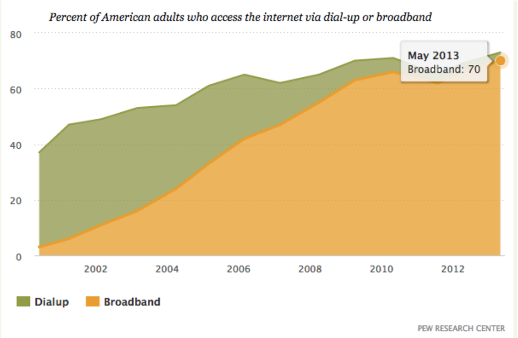
    1. Binomial distribution is used for calculating the probability of exact number of successes for a given number of trials

    2. Expected value and standard deviation of the binomial can be calculated using its parameters n and p

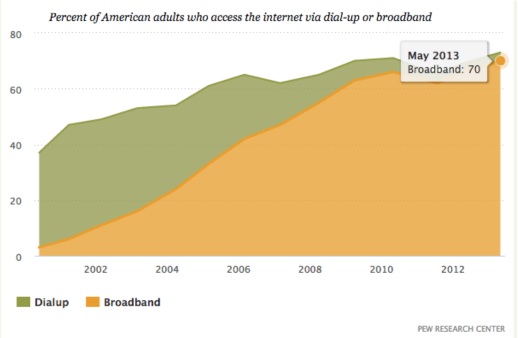
    3. Shape of the binomial distribution approaches normal when the S-F rule is met

3. Summary

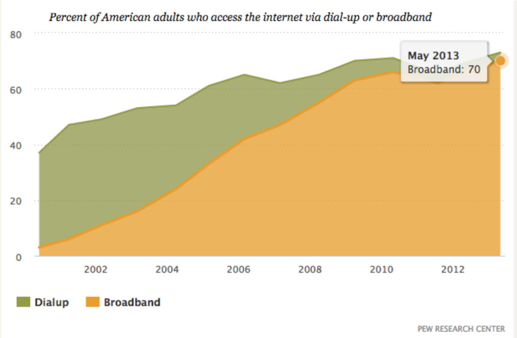
High-speed broadband connection at home in the US



High-speed broadband connection at home in the US

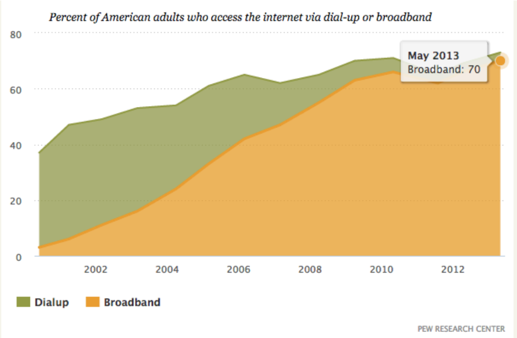


- Each person in the poll be thought of as a *trial*



- Each person in the poll be thought of as a *trial*
- A person is labeled a *success* if s/he has high-speed broadband connection at home, *failure* if not

High-speed broadband connection at home in the US



- Each person in the poll be thought of as a *trial*
- A person is labeled a *success* if s/he has high-speed broadband connection at home, *failure* if not
- Since 70% have high-speed broadband connection at home, *probability of success* is $p = 0.70$

Suppose we randomly select three individuals from the US, what is the probability that exactly 1 has high-speed broadband connection at home?

Let's call these people Anthony (A), Barry (B), Cam (C). Each one of the three scenarios below will satisfy the condition of "exactly 1 of them says Yes":

Suppose we randomly select three individuals from the US, what is the probability that exactly 1 has high-speed broadband connection at home?

Let's call these people Anthony (A), Barry (B), Cam (C). Each one of the three scenarios below will satisfy the condition of "exactly 1 of them says Yes":

Scenario 1: $\underset{\text{(A) } \textit{yes}}{\underline{0.70}} \times \underset{\text{(B) no}}{\underline{0.30}} \times \underset{\text{(C) no}}{\underline{0.30}} \approx 0.063$

Suppose we randomly select three individuals from the US, what is the probability that exactly 1 has high-speed broadband connection at home?

Let's call these people Anthony (A), Barry (B), Cam (C). Each one of the three scenarios below will satisfy the condition of "exactly 1 of them says Yes":

Scenario 1: $\underset{\text{(A) } \textit{yes}}{\underline{0.70}} \times \underset{\text{(B) no}}{\underline{0.30}} \times \underset{\text{(C) no}}{\underline{0.30}} \approx 0.063$

Scenario 2: $\underset{\text{(A) no}}{\underline{0.30}} \times \underset{\text{(B) } \textit{yes}}{\underline{0.70}} \times \underset{\text{(C) no}}{\underline{0.30}} \approx 0.063$

Suppose we randomly select three individuals from the US, what is the probability that exactly 1 has high-speed broadband connection at home?

Let's call these people Anthony (A), Barry (B), Cam (C). Each one of the three scenarios below will satisfy the condition of "exactly 1 of them says Yes":

Scenario 1: $\underset{\text{(A) } \textit{yes}}{\underline{0.70}} \times \underset{\text{(B) no}}{\underline{0.30}} \times \underset{\text{(C) no}}{\underline{0.30}} \approx 0.063$

Scenario 2: $\underset{\text{(A) no}}{\underline{0.30}} \times \underset{\text{(B) } \textit{yes}}{\underline{0.70}} \times \underset{\text{(C) no}}{\underline{0.30}} \approx 0.063$

Scenario 3: $\underset{\text{(A) no}}{\underline{0.30}} \times \underset{\text{(B) no}}{\underline{0.30}} \times \underset{\text{(C) } \textit{yes}}{\underline{0.70}} \approx 0.063$

Suppose we randomly select three individuals from the US, what is the probability that exactly 1 has high-speed broadband connection at home?

Let's call these people Anthony (A), Barry (B), Cam (C). Each one of the three scenarios below will satisfy the condition of "exactly 1 of them says Yes":

Scenario 1: $\underset{\text{(A) } \textit{yes}}{\underline{0.70}} \times \underset{\text{(B) no}}{\underline{0.30}} \times \underset{\text{(C) no}}{\underline{0.30}} \approx 0.063$

Scenario 2: $\underset{\text{(A) no}}{\underline{0.30}} \times \underset{\text{(B) } \textit{yes}}{\underline{0.70}} \times \underset{\text{(C) no}}{\underline{0.30}} \approx 0.063$

Scenario 3: $\underset{\text{(A) no}}{\underline{0.30}} \times \underset{\text{(B) no}}{\underline{0.30}} \times \underset{\text{(C) } \textit{yes}}{\underline{0.70}} \approx 0.063$

The probability of exactly one 1 of 3 people saying Yes is the sum of all of these probabilities.

$$0.063 + 0.063 + 0.063 = 3 \times 0.063 = 0.189$$

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 $n = 3$ trials), and we calculated this probability as

$$\# \ of \ scenarios \times P(single \ scenario)$$

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 $n = 3$ trials), and we calculated this probability as

$$\# \ of \ scenarios \times P(single \ scenario)$$

- $P(single \ scenario) = p^k \ (1 - p)^{(n-k)}$

  probability of success to the power of number of successes, probability of failure to the power of number of failures

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 $n = 3$ trials), and we calculated this probability as

$$\# \ of \ scenarios \times P(single \ scenario)$$

- $P(single \ scenario) = p^k \, (1 - p)^{(n-k)}$

  probability of success to the power of number of successes, probability of failure to the power of number of failures

- number of scenarios: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

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 $n = 3$ trials), and we calculated this probability as

$$\#\ of\ scenarios \times P(single\ scenario)$$

- $P(single\ scenario) = p^k\ (1 - p)^{(n-k)}$

  probability of success to the power of number of successes, probability of failure to the power of number of failures

- number of scenarios: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

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

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

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

Note: You can also use R for the calculation of number of scenarios:

```
> choose(5,3)
```

```
[1] 10
```

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

Note: You can also use R for the calculation of number of scenarios:

```
> choose(5,3)
```

```
[1] 10
```

Note: And to compute probabilities

```
> dbinom(1, size = 3, prob = 0.7)
```

```
[1] 0.189
```

Clicker question

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, $\binom{n}{n-1} = n-1$.

Clicker question

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, $\binom{n}{n-1} = n-1$.*

Clicker question

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

Clicker question

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

Clicker question

According to the results of the Pew poll suggesting that 70% of Americans have high-speed broadband connection at home, is the probability of exactly 2 out of 15 randomly sampled Americans having such connection at home pretty high or pretty low?

(a) pretty high

(b) pretty low

Clicker question

According to the results of the Pew poll suggesting that 70% of Americans have high-speed broadband connection at home, is the probability of exactly 2 out of 15 randomly sampled Americans having such connection at home pretty high or pretty low?

(a) pretty high
(b) *pretty low*

Clicker question

According to the results of the Pew poll 70% of Americans have high-speed broadband connection at home, what is the probability that exactly 2 out of 15 randomly sampled Americans have such connection at home?

(a) $0.70^2 \times 0.30^{13}$

(b) $\binom{2}{15} \times 0.70^2 \times 0.30^{13}$

(c) $\binom{15}{2} \times 0.70^2 \times 0.30^{13}$

(d) $\binom{15}{2} \times 0.70^{13} \times 0.30^2$

Clicker question

According to the results of the Pew poll 70% of Americans have high-speed broadband connection at home, what is the probability that exactly 2 out of 15 randomly sampled Americans have such connection at home?

(a) $0.70^2 \times 0.30^{13}$

(b) $\binom{2}{15} \times 0.70^2 \times 0.30^{13}$

(c) $\binom{15}{2} \times 0.70^2 \times 0.30^{13}$
$= \frac{15!}{13! \times 2!} \times 0.70^2 \times 0.30^{13} = 105 \times 0.70^2 \times 0.30^{13} = 8.2e-06$

(d) $\binom{15}{2} \times 0.70^{13} \times 0.30^2$

1. Housekeeping

2. Main ideas

    1. Binomial distribution is used for calculating the probability of exact number of successes for a given number of trials

    2. Expected value and standard deviation of the binomial can be calculated using its parameters n and p

    3. Shape of the binomial distribution approaches normal when the S-F rule is met

3. Summary

According to the results of the Pew poll suggestion that 70% of Americans have high-speed broadband connection at home, among a random sample of 100 Americans, how many would you expect to have such connection at home?

According to the results of the Pew poll suggestion that 70% of Americans have high-speed broadband connection at home, among a random sample of 100 Americans, how many would you expect to have such connection at home?

- $100 \times 0.70 = 70$

According to the results of the Pew poll suggestion that 70% of Americans have high-speed broadband connection at home, among a random sample of 100 Americans, how many would you expect to have such connection at home?

- $100 \times 0.70 = 70$
  - Or more formally, $\mu = np = 100 \times 0.7 = 7$

According to the results of the Pew poll suggestion that 70% of Americans have high-speed broadband connection at home, among a random sample of 100 Americans, how many would you expect to have such connection at home?

- $100 \times 0.70 = 70$
  - Or more formally, $\mu = np = 100 \times 0.7 = 7$
- But this doesn't mean in every random sample of 100 Americans exactly 70 will have high-speed broadband connection at home. In some samples there will be fewer of those, and in others more. How much would we expect this value to vary?

According to the results of the Pew poll suggestion that 70% of Americans have high-speed broadband connection at home, among a random sample of 100 Americans, how many would you expect to have such connection at home?

- $100 \times 0.70 = 70$
  - Or more formally, $\mu = np = 100 \times 0.7 = 7$
- But this doesn't mean in every random sample of 100 Americans exactly 70 will have high-speed broadband connection at home. In some samples there will be fewer of those, and in others more. How much would we expect this value to vary?
  - $\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.70 \times 0.30} \approx 4.58$

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.

1. Housekeeping

2. Main ideas

1. Binomial distribution is used for calculating the probability of exact number of successes for a given number of trials

2. Expected value and standard deviation of the binomial can be calculated using its parameters n and p

3. Shape of the binomial distribution approaches normal when the S-F rule is met

3. Summary

https://gallery.shinyapps.io/dist_calc/

https://gallery.shinyapps.io/dist_calc/

You can use the normal distribution to approximate binomial probabilities when the sample size is large enough.

https://gallery.shinyapps.io/dist_calc/

You can use the normal distribution to approximate binomial probabilities when the sample size is large enough.

S-F rule: The sample size is considered large enough if the expected number of successes and failures are both at least 10

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

Clicker question

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

(a) $n = 25, p = 0.45$

(b) $n = 100, p = 0.95$

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

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

Clicker question

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

(a) $n = 25, p = 0.45$

(b) $n = 100, p = 0.95$

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

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

What is the probability that among a random sample of 1,000 Americans at least three-fourths have high-speed broadband connection at home?

What is the probability that among a random sample of 1,000 Americans at least three-fourths have high-speed broadband connection at home?

$$Binom(n = 1000, p = 0.7)$$

What is the probability that among a random sample of 1,000 Americans at least three-fourths have high-speed broadband connection at home?

$$Binom(n = 1000, p = 0.7)$$

$$P(K \geq 750) = P(K = 750) + P(K = 751) + P(K = 752) + \cdots + P(K = 1000)$$

What is the probability that among a random sample of 1,000 Americans at least three-fourths have high-speed broadband connection at home?

$$Binom(n = 1000, p = 0.7)$$

$$P(K \geq 750) = P(K = 750) + P(K = 751) + P(K = 752) + \cdots + P(K = 1000)$$

1. Using R:

```
> sum(dbinom(750:1000, size = 1000, prob = 0.7))
```

```
[1] 0.00026
```

What is the probability that among a random sample of 1,000 Americans at least three-fourths have high-speed broadband connection at home?

$$Binom(n = 1000, p = 0.7)$$

$$P(K \geq 750) = P(K = 750) + P(K = 751) + P(K = 752) + \cdots + P(K = 1000)$$

1. Using R:

```
> sum(dbinom(750:1000, size = 1000, prob = 0.7))
```

```
[1] 0.00026
```

2. Using the normal approximation to the binomial: Since we have at least expected successes $(1000 \times 0.7 = 700)$ and 10 expected failures $(1000 \times 0.3 = 300)$,

$$Binom(n = 1000, p = 0.7) \sim$$
$$N(\mu = 1000 \times 0.7, \sigma = \sqrt{1000 \times 0.7 \times 0.3})$$

Application exercise: 2.4 Binomial distribution

See course website for details.

1. Housekeeping

2. Main ideas

   1. Binomial distribution is used for calculating the probability of exact number of successes for a given number of trials

   2. Expected value and standard deviation of the binomial can be calculated using its parameters n and p

   3. Shape of the binomial distribution approaches normal when the S-F rule is met

3. Summary

1. Binomial distribution is used for calculating the probability of exact number of successes for a given number of trials
2. Expected value and standard deviation of the binomial can be calculated using its parameters n and p
3. Shape of the binomial distribution approaches normal when the S-F rule is met