Unit 2: Probability and distributions

3. Normal and binomial distributions

Sta 101 – Spring 2019

Duke University, Department of Statistical Science

Dr. Ellison

Slides posted at
https://www2.stat.duke.edu/courses/Spring19/sta101.001/

Outline

1. Housekeeping

2. Main ideas
   1. General Probability Distribution Properties: 🔍 👫 Two types of probability distributions: discrete and continuous
   2. Normal Distribution Properties:
      1. 🔍 Normal distribution is unimodal, symmetric, and follows the 68-95-99.7 rule
      2. 💡 ⚙ 🔍 👫 Z scores serve as a ruler for any distribution
   3. Binomial Distribution Properties:
      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
   4. Relationship between Binomial and Normal Distribution:
      1. 👫 ✋ 💡 💻 Shape of the binomial distribution approaches normal when the S-F rule is met

---

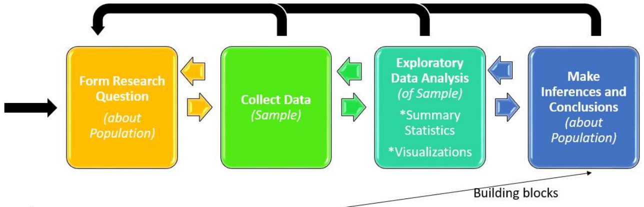
Announcements

Coming up…

- ▶ Lab Assignment 3 is due **Thursday just before your lab section time.**
- ▶ Problem Set 2 due **Wednesday 2/6 11:55 pm**
- ▶ Performance Assessment 2 due **Sunday 2/10 11:55 pm** (opens Wednesday)
- ▶ Readiness Assessment 3 next **Wednesday 2/6 in class!**

Outline

Course Overview

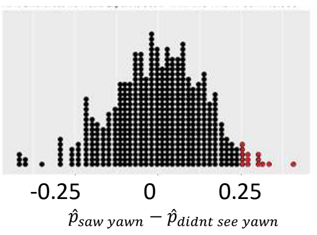


- **Mathematics behind statistics**
  - *Unit 2 - Probability & distributions:*
    - Basics of probability and chance processes
    - Bayesian perspective in statistical inference
    - The normal and binomial distributions.

Outline

Making an Inference

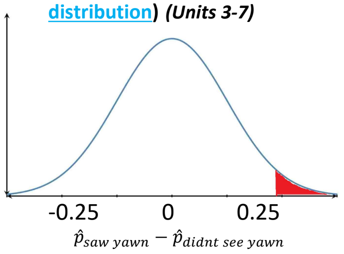
Randomization Distribution *(Unit 1)*



Vs.

Sampling Distribution

(**Special kind of normal distribution**) ***(Units 3-7)***



Outline

1. Housekeeping

2. Main ideas
   1. General Probability Distribution Properties: Two types of probability distributions: discrete and continuous
   2. Normal Distribution Properties:
      1. Normal distribution is unimodal, symmetric, and follows the 68-95-99.7 rule
      2. Z scores serve as a ruler for any distribution
   3. Binomial Distribution Properties:
      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
   4. Relationship between Binomial and Normal Distribution:
      1. Shape of the binomial distribution approaches normal when the S-F rule is met

---

Outline

Discrete Probability Distributions

vs.

Continuous Probability Distributions

Outline

How can we describe the probabilities for a set of events that are discrete?

Events	Probabilities
0 out of 5 randomly selected adults use Twitter	
1 out of 5 randomly selected adults use Twitter	
2 out of 5 randomly selected adults use Twitter	
3 out of 5 randomly selected adults use Twitter	
4 out of 5 randomly selected adults use Twitter	
5 out of 5 randomly selected adults use Twitter	

Events	Probabilities
Donald Trump is president.	
Barack Obama is president.	
George W. Bush is president.	
Anyone else is president.	

Outline

How can we describe the probabilities for a set of events that are discrete?

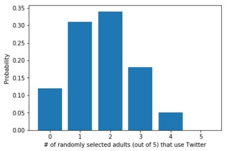
Equation

$$P\binom{k \text{ out of } n \text{ RS adults}}{\text{use Twitter}} = \binom{n}{k} p^k (1-p)^{n-k}$$

Table

Events	Probabilities
0 out of 5 randomly selected adults use Twitter	0.12
1 out of 5 randomly selected adults use Twitter	0.31
2 out of 5 randomly selected adults use Twitter	0.34
3 out of 5 randomly selected adults use Twitter	0.18
4 out of 5 randomly selected adults use Twitter	0.05
5 out of 5 randomly selected adults use Twitter	0

Histogram



Distribution Shorthand, if a random variable follows probabilities that follow a well-known distribution.
Ex: $X \sim Bin(n = 5, p = 0.34)$ means X follows a **Binomial Distribution** with n=5 trials and probability of a trial success is p=0.34

Two types of probability distributions: discrete and continuous

- A *discrete probability distribution* lists all possible events and the probabilities with which they occur
  - The events listed must be disjoint
  - Each probability must be between 0 and 1
  - The probabilities must total 1

  Example: Binomial distribution

Two types of probability distributions: discrete and continuous

- A *discrete probability distribution* lists all possible events and the probabilities with which they occur
  - The events listed must be disjoint
  - Each probability must be between 0 and 1
  - The probabilities must total 1

  Example: Binomial distribution

Question: Using the table below, what is the probability that at least 1 out of 100 randomly selected adults use Snapchat?

Events	Probabilities
0 out of 100 randomly selected adults use Snapchat	0.13
1 out of 100 randomly selected adults use Snapchat	0.27
2 out of 100 randomly selected adults use Snapchat	0.27
…	…
99 out of 100 randomly selected adults use Snapchat	≈0
100 out of 100 randomly selected adults use Snapcha	≈0

Two types of probability distributions: discrete and continuous

- A *discrete probability distribution* lists all possible events and the probabilities with which they occur
  - The events listed must be disjoint
  - Each probability must be between 0 and 1
  - The probabilities must total 1

  Example: Binomial distribution

Question: Using the table below, what is the probability that at least 1 out of 100 randomly selected adults use Snapchat?

Events	Probabilities
0 out of 100 randomly selected adults use Snapchat	0.13
1 out of 100 randomly selected adults use Snapchat	0.27
2 out of 100 randomly selected adults use Snapchat	0.27
…	…
99 out of 100 randomly selected adults use Snapchat	≈0
100 out of 100 randomly selected adults use Snapcha	≈0

Know: Sum of all probabilities is 1

Want: Sum of these probabilities

Two types of probability distributions: discrete and continuous

- A *discrete probability distribution* lists all possible events and the probabilities with which they occur
  - The events listed must be disjoint
  - Each probability must be between 0 and 1
  - The probabilities must total 1

Example: Binomial distribution

<u>Question</u>: Using the table below, what is the probability that <u>at least</u> 1 out of 100 randomly selected adults use Snapchat?

Events	Probabilities
0 out of 100 randomly selected adults use Snapchat	0.13
1 out of 100 randomly selected adults use Snapchat	0.27
2 out of 100 randomly selected adults use Snapchat	0.27
...	...
99 out of 100 randomly selected adults use Snapchat	≈0
100 out of 100 randomly selected adults use Snapcha	≈0

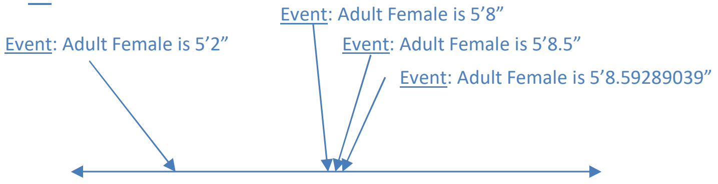
<u>Answer</u>:

$$P\begin{pmatrix}at\ least\ 1\ out\ 100\ randomly\\ selected\ adults\ use\ Snapchat\end{pmatrix} = 1 - P\begin{pmatrix}0\ out\ 100\ randomly\\ selected\ adults\ use\ Snapchat\end{pmatrix} = 1 - 0.13 = \mathbf{0.87}$$

Outline

How do we describe the probabilities for a set of events that are <u>continuous</u>?

<u>Ex:</u>



<u>Event</u>: Adult Female is 5’2”

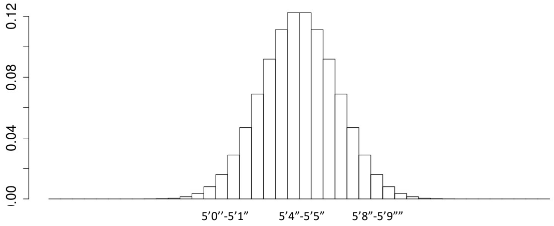
<u>Event</u>: Adult Female is 5’8”

<u>Event</u>: Adult Female is 5’8.5”

<u>Event</u>: Adult Female is 5’8.59289039”

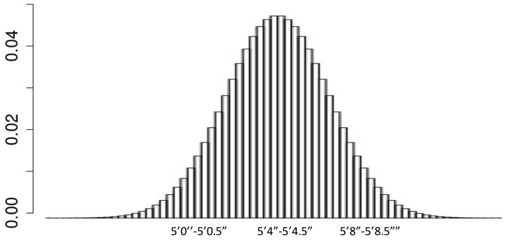
<u>Events</u> = Heights of Adult Females

Outline

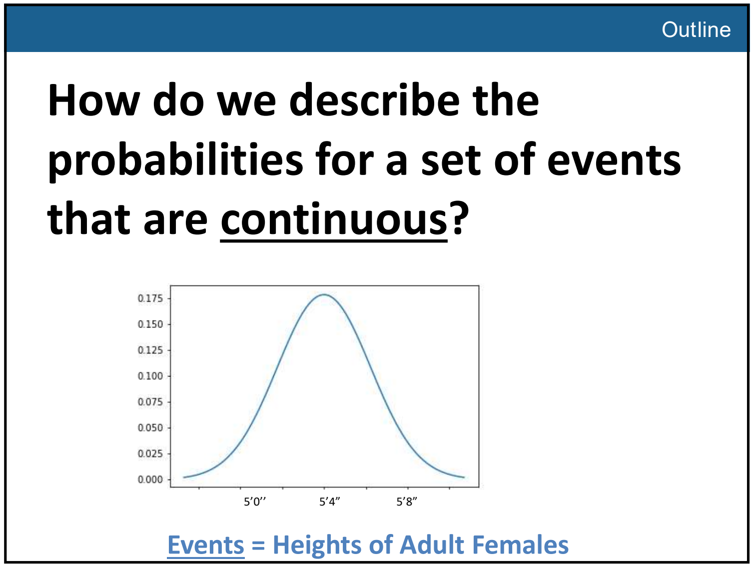


<u>Events</u> = Heights of Adult Females

Outline

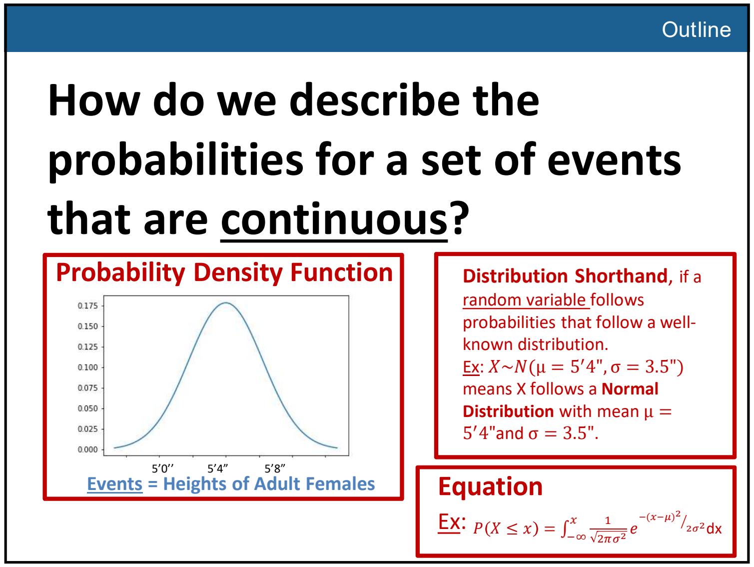


<u>Events</u> = Heights of Adult Females



How do we describe the probabilities for a set of events that are continuous?

Events = Heights of Adult Females



How do we describe the probabilities for a set of events that are continuous?

Probability Density Function

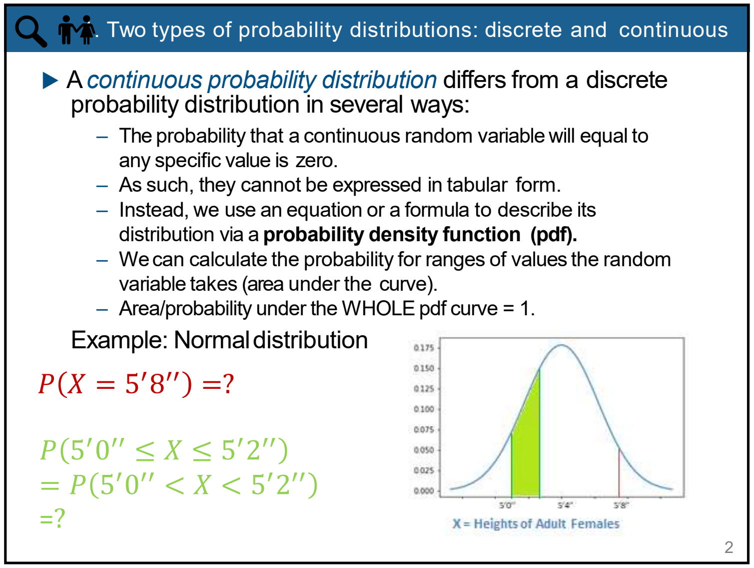
Events = Heights of Adult Females

Distribution Shorthand, if a random variable follows probabilities that follow a well-known distribution.

Ex: $X \sim N(\mu = 5'4", \sigma = 3.5")$ means X follows a **Normal Distribution** with mean $\mu = 5'4"$ and $\sigma = 3.5"$.

Equation

Ex: $P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx$



Two types of probability distributions: discrete and continuous

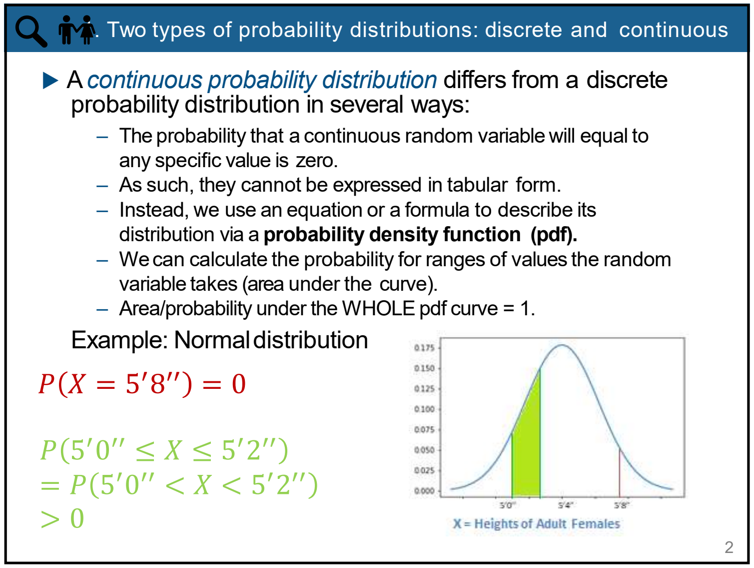
- A *continuous probability distribution* differs from a discrete probability distribution in several ways:
  - The probability that a continuous random variable will equal to any specific value is zero.
  - As such, they cannot be expressed in tabular form.
  - Instead, we use an equation or a formula to describe its distribution via a **probability density function (pdf).**
  - We can calculate the probability for ranges of values the random variable takes (area under the curve).
  - Area/probability under the WHOLE pdf curve = 1.

Example: Normal distribution

$P(X = 5'8'') = ?$

$P(5'0'' \le X \le 5'2'')$
$= P(5'0'' < X < 5'2'')$
$= ?$

X = Heights of Adult Females



Two types of probability distributions: discrete and continuous

- A *continuous probability distribution* differs from a discrete probability distribution in several ways:
  - The probability that a continuous random variable will equal to any specific value is zero.
  - As such, they cannot be expressed in tabular form.
  - Instead, we use an equation or a formula to describe its distribution via a **probability density function (pdf).**
  - We can calculate the probability for ranges of values the random variable takes (area under the curve).
  - Area/probability under the WHOLE pdf curve = 1.

Example: Normal distribution

$P(X = 5'8'') = 0$

$P(5'0'' \le X \le 5'2'')$
$= P(5'0'' < X < 5'2'')$
> 0

X = Heights of Adult Females

Outline

1. Housekeeping

2. Main ideas
   1. General Probability Distribution Properties: Two types of probability distributions: discrete and continuous
   2. Normal Distribution Properties:
      1. Normal distribution is unimodal, symmetric, and follows the 68-95-99.7 rule
      2. Z scores serve as a ruler for any distribution
   3. Binomial Distribution Properties:
      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
   4. Relationship between Binomial and Normal Distribution:
      1. Shape of the binomial distribution approaches normal when the S-F rule is met

Outline

What is a useful property that normal distributions have (that most others don't)?

Clicker question

Speeds of cars on a highway are normally distributed with mean 65 miles / hour. The minimum speed recorded is 48 miles / hour and the maximum speed recorded is 83 miles / hour. Which of the following is most likely to be the standard deviation of the distribution?

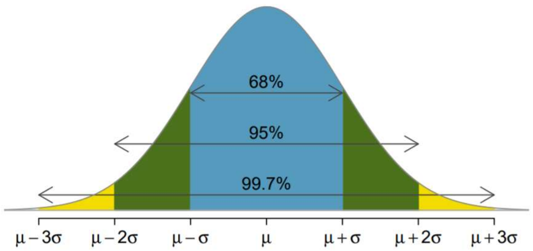
(a) -5
(b) 5
(c) 10
(d) 15
(e) 30

Clicker question

Speeds of cars on a highway are **normally distributed** with mean 65 miles / hour. The minimum speed recorded is 48 miles / hour and the maximum speed recorded is 83 miles / hour. Which of the following is most likely to be the standard deviation of the distribution?

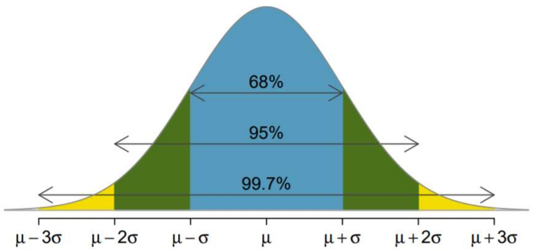
(a) -5 → *SD cannot be negative*
(b) *5* → 65 ± (3 × 5) = (50, 80)
(c) 10 → 65 ± (3 × 10) = (35, 95)
(d) 15 → 65 ± (3 × 15) = (20, 110)
(e) 30 → 65 ± (3 × 30) = (−25, 155)

Normal Distribution Follows the **68-95-99.7 rule**



- What percent of the observations are greater than μ+2σ?
- What percent of the observations are greater than μ+2.5σ?

Normal Distribution Follows the **68-95-99.7 rule**



- What percent of the observations are greater than μ+2σ? (about 2.5%)
- What percent of the observations are greater than μ+2.5σ? (need to use z-tables to figure out)

---

1. Housekeeping

2. Main ideas
    1. General Probability Distribution Properties: 🔍 👫 Two types of probability distributions: discrete and continuous
    2. Normal Distribution Properties:
        1. 🔍 Normal distribution is unimodal, symmetric, and follows the 68-95-99.7 rule
        2. 💡 ⚙ 🔍 👫 Z scores serve as a ruler for any distribution
    3. Binomial Distribution Properties:
        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
    4. Relationship between Binomial and Normal Distribution:
        1. 👫 ✋ 💡 💻 Shape of the binomial distribution approaches normal when the S-F rule is met

---

How can we determine if an observation is unusual?

3. Z scores serve as a ruler for any distribution

How can we determine if it would be unusual for an adult woman in North Carolina to be 96” (8 ft) tall?

How can we determine if it would be unusual for an adult alien woman to be 103 metreloots tall, assuming the distribution of heights of adult alien women is approximately normal?

---

3. Z scores serve as a ruler for any distribution

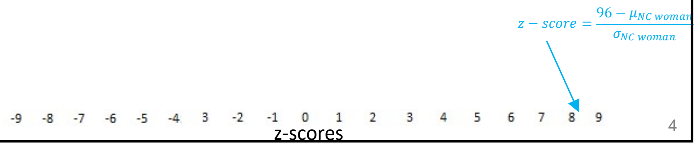
A **Z score** creates a common scale so you can assess data without worrying about the specific units in which it was measured. $Z = \frac{obs - mean}{SD}$

---

3. Z scores serve as a ruler for any distribution

A **Z score** creates a common scale so you can assess data without worrying about the specific units in which it was measured. $Z = \frac{obs - mean}{SD}$

How can we determine if it would be unusual for an adult woman in North Carolina to be 96” (8 ft) tall?



$$z - score = \frac{96 - \mu_{NC\ woman}}{\sigma_{NC\ woman}}$$

---

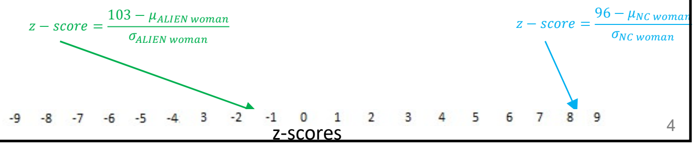
3. Z scores serve as a ruler for any distribution

A **Z score** creates a common scale so you can assess data without worrying about the specific units in which it was measured. $Z = \frac{obs - mean}{SD}$

How can we determine if it would be unusual for an adult woman in North Carolina to be 96” (8 ft) tall?

How can we determine if it would be unusual for an adult alien woman to be 103 metreloots tall, assuming the distribution of heights of adult alien women is approximately normal?

… but we still don’t know if these women’s heights are unusual… yet!



$$z - score = \frac{103 - \mu_{ALIEN\ woman}}{\sigma_{ALIEN\ woman}}$$

$$z - score = \frac{96 - \mu_{NC\ woman}}{\sigma_{NC\ woman}}$$

3. Z scores serve as a ruler for any distribution

All Distributions

$$Z = \frac{obs - mean}{SD}$$

- **Z score**: number of standard deviations the observation falls above or below the mean (for any distribution)

---

3. Z scores serve as a ruler for any distribution

All Distributions

$$Z = \frac{obs - mean}{SD}$$

Mean of the population the observation comes from

Standard deviation of the population the observation comes from

- **Z score**: number of standard deviations the observation falls above or below the mean (for any distribution)

---

3. Z scores serve as a ruler for any distribution

All Distributions

$$Z = \frac{obs - mean}{SD}$$

Mean of the population the observation comes from

Standard deviation of the population the observation comes from

- **Z score**: number of standard deviations the observation falls above or below the mean (for any distribution)
- Z score defined for distributions of any shape, but only when the distribution is normal can we use Z scores to calculate percentiles and assess if an observation is unusual.

---

3. Z scores serve as a ruler for any distribution

All Distributions

$$Z = \frac{obs - mean}{SD}$$

Mean of the population the observation comes from

Standard deviation of the population the observation comes from

- **Z score**: number of standard deviations the observation falls above or below the mean (for any distribution)
- Z score defined for distributions of any shape, but only when the distribution is normal can we use Z scores to calculate percentiles and assess if an observation is unusual.

Normal Distributions

- **Z distribution** (also called the *standardiZed normal* distribution, is a special case of the normal distribution where $\mu = 0$ and $\sigma = 1$. Distribution of z-scores of obs from normal distributions. $Z \sim N(\mu = 0, \sigma = 1)$

3. Z scores serve as a ruler for any distribution

$$Z = \frac{obs - mean}{SD}$$

Mean of the population the observation comes from

Standard deviation of the population the observation comes from

All Distributions

- **Z score**: number of standard deviations the observation falls above or below the mean (for any distribution)
- Z score defined for distributions of any shape, but only when the distribution is normal can we use Z scores to calculate percentiles and assess if an observation is unusual

Normal Distributions

- **Z distribution** (also called the *standardiZed normal* distribution, is a special case of the normal distribution where $\mu = 0$ and $\sigma = 1$. Distribution of z-scores of obs from normal distributions. $Z \sim N(\mu = 0, \sigma = 1)$
- Observations from normal distributions with $|Z| > 2$ are usually considered *unusual*

5

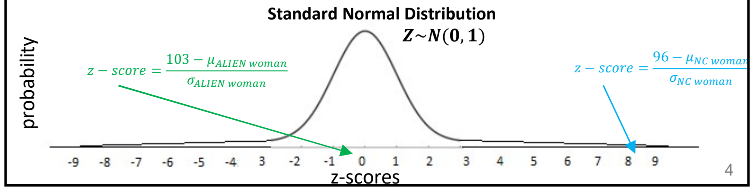
---

3. Z scores serve as a ruler for any distribution

A **Z score** creates a common scale so you can assess data without worrying about the specific units in which it was measured.

It is **unusual** for an adult woman in North Carolina to be 96” (8 ft) tall *(we know that adult heights are approx. normal)*

It is **not unusual** for an adult alien woman(?) to be 103 metreloots tall, assuming the distribution of heights of adult alien women is approximately normal.



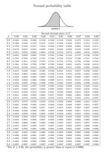
4

Question: What is the probability that the **z-score** of an **observation from a normal distribution** is between -1 and 1?

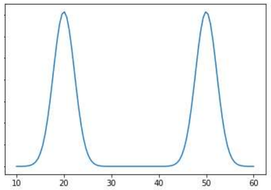
Question: What is the probability that the **z-score** of an **observation from a normal distribution** is between -1 and 1?

Answer: $P(-1 \leq Z \leq 1) = 0.68$

We CAN use the z-tables to find this probability, because the observation comes from a normal distribution.

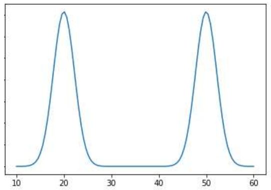


Question: What is the probability that the z-score of an observation from the distribution below is between -1 and 1?

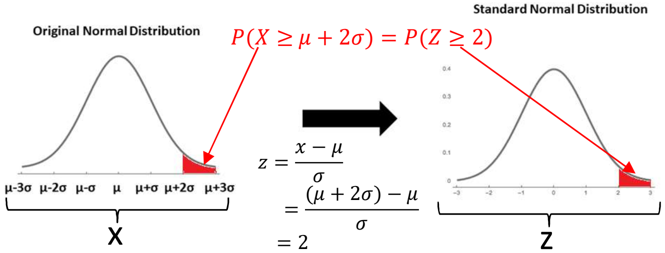


Question: What is the probability that the z-score of an observation from the distribution below is between -1 and 1?

Answer: $P(-1 \leq Z \leq 1) = ?$



We're **not given enough information**! We CAN'T use the z-table because the observation doesn't come from a normal distribution.

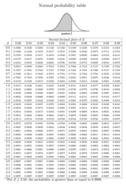


$P(X \geq \mu + 2\sigma) = P(Z \geq 2)$

$$z = \frac{x - \mu}{\sigma}$$

$$= \frac{(\mu + 2\sigma) - \mu}{\sigma}$$

$$= 2$$



Clicker question

Scores on a standardized test are normally distributed with a mean of 100 and a standard deviation of 20. If these scores are converted to standard normal Z scores, which of the following statements will be correct?

(a) The mean will equal 0, but the median cannot be determined.
(b) The mean of the standardized Z-scores will equal 100.
(c) The mean of the standardized Z-scores will equal 5.
(d) Both the mean and median score will equal 0.
(e) A score of 70 is considered unusually low on this test.

Clicker question

Scores on a standardized test are **normally distributed** with a **mean of 100** and a **standard deviation of 20**. If these scores are converted to standard normal Z scores, which of the following statements will be correct?

(a) The mean will equal 0, but the median cannot be determined.
(b) The mean of the standardized Z-scores will equal 100.
(c) The mean of the standardized Z-scores will equal 5.
(d) Both the mean and median score will equal 0.
~~(e) A score of 70 is considered unusually low on this test.~~

$$z-score = \frac{70-100}{20} = -1.5 > -2$$ Not unusual

Outline

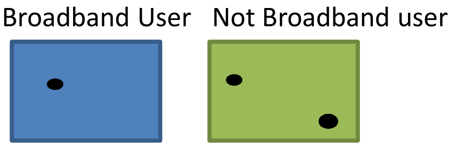
1. Housekeeping

2. Main ideas
    1. General Probability Distribution Properties: Two types of probability distributions: discrete and continuous
    2. Normal Distribution Properties:
        1. Normal distribution is unimodal, symmetric, and follows the 68-95-99.7 rule
        2. Z scores serve as a ruler for any distribution
    3. Binomial Distribution Properties:
        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
    4. Relationship between Binomial and Normal Distribution:
        1. Shape of the binomial distribution approaches normal when the S-F rule is met

---

Outline

How do we know if a random variable follows a binomial distribution? How do we use a binomial distribution to calculate probabilities?



Outline

Do we have enough information to fill out this probability table?

Events	Probabilities
0 out of 3 randomly selected adults have Broadband.	
1 out of 3 randomly selected adults have Broadband.	
2 out of 3 randomly selected adults have Broadband.	
3 out of 3 randomly selected adults have Broadband.	

Do we have enough information to fill out this probability table?

Events	Probabilities
0 out of 3 randomly selected adults have Broadband.	
1 out of 3 randomly selected adults have Broadband.	
2 out of 3 randomly selected adults have Broadband.	
3 out of 3 randomly selected adults have Broadband.	

Binomial Distribution Conditions Met:
- n=3 fixed trials
- Each trial is independent.
- Each trial can have one of two outcomes (success=having Broadband, failure=not having Broadband)
- ____________

No!

Suppose we know now that the probability of any one randomly selected adult having Broadband is p=0.7. Do we have enough information to fill out this probability table?

Events	Probabilities
0 out of 3 randomly selected adults have Broadband.	
1 out of 3 randomly selected adults have Broadband.	
2 out of 3 randomly selected adults have Broadband.	
3 out of 3 randomly selected adults have Broadband.	

Suppose we know now that the probability of any one randomly selected adult having Broadband is p=0.7. Do we have enough information to fill out this probability table?

Events	Probabilities
0 out of 3 randomly selected adults have Broadband.	
1 out of 3 randomly selected adults have Broadband.	
2 out of 3 randomly selected adults have Broadband.	
3 out of 3 randomly selected adults have Broadband.	

Binomial Distribution Conditions Met:
- n=3 fixed trials
- Each trial is independent.
- Each trial can have one of two outcomes (success=having Broadband, failure=not having Broadband)
- Probability of a given trial being a "success" (having Broadband) is the same (p=0.7).

Yes! We can use the **Binomial Distributio** to find these probabilities.

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

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

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

Binomial distribution (cont.)

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

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

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

$$0! = 1$$
$$1! = 1$$
$$2! = 2 \times 1$$
$$\ldots$$
$$5! = 5 \times 4 \times 3 \times 2 \times 1$$

Binomial distribution (cont.)

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

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

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

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

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

```
[1] 10
```

Note: And to compute probabilities

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

Outline

Suppose we know now that the probability of any one randomly selected adult having Broadband is p=0.7.

Events	**Probabilities**
0 out of 3 randomly selected adults have Broadband.	$\binom{3}{0} 0.7^0 (1-0.7)^{3-0}$
1 out of 3 randomly selected adults have Broadband.	$\binom{3}{1} 0.7^1 (1-0.7)^{3-1}$
2 out of 3 randomly selected adults have Broadband.	$\binom{3}{2} 0.7^2 (1-0.7)^{3-2}$
3 out of 3 randomly selected adults have Broadband.	$\binom{3}{3} 0.7^3 (1-0.7)^{3-3}$

Outline

Where does the Binomial Distribution probability equation come from?

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

Considering many scenarios

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":

Considering many scenarios

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

Considering many scenarios

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

Considering many scenarios

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

Considering many scenarios

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

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

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

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

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

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

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

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 $n = 3$ 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

---

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 $n = 3$ 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: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

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

$$P(k\ success\ in\ n\ trials) = (\#of\ scenarios)P(single\ scenario)$$

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

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

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

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 at least 2 out of 15 randomly sampled Americans have such connection at home?

a) $\binom{15}{2} 0.7^2 (1-0.7)^{15-2}$

b) $\binom{15}{0} 0.7^0 (1-0.7)^{15-0} + \binom{15}{1} 0.7^1 (1-0.7)^{15-1}$

c) $\binom{15}{2} 0.7^2 (1-0.7)^{15-2} + \binom{15}{3} 0.7^3 (1-0.7)^{15-3} + \ldots$
$+ \binom{15}{15} 0.7^{15} (1-0.7)^{15-15}$

d) $1 - \left(\binom{15}{0} 0.7^0 (1-0.7)^{15-0} + \binom{15}{1} 0.7^1 (1-0.7)^{15-1} \right)$

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 at least 2 out of 15 randomly sampled Americans have such connection at home?

a) $\binom{15}{2}0.7^2(1-0.7)^{15-2}$

b) $\binom{15}{0}0.7^0(1-0.7)^{15-0}+\binom{15}{1}0.7^1(1-0.7)^{15-1}$

c) $\binom{15}{2}0.7^2(1-0.7)^{15-2}+\binom{15}{3}0.7^3(1-0.7)^{15-3}+\ldots+\binom{15}{15}0.7^{15}(1-0.7)^{15-15}$

This way takes less time!

d) $1-\left(\binom{15}{0}0.7^0(1-0.7)^{15-0}+\binom{15}{1}0.7^1(1-0.7)^{15-1}\right)$

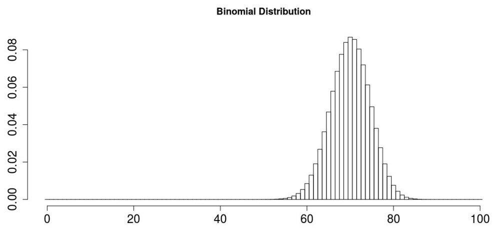
Outline

1. Housekeeping

2. Main ideas
   1. General Probability Distribution Properties: Two types of probability distributions: discrete and continuous
   2. Normal Distribution Properties:
      1. Normal distribution is unimodal, symmetric, and follows the 68-95-99.7 rule
      2. Z scores serve as a ruler for any distribution
   3. Binomial Distribution Properties:
      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
   4. Relationship between Binomial and Normal Distribution:
      1. Shape of the binomial distribution approaches normal when the S-F rule is met

Outline

What is the mean and standard deviation of a binomial distribution?



Expected value and standard deviation of binomial

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?

▶Mean of a Binomial Distribution :

$$\mu = np = 100 \times 0.7 = 70$$

▶Standard Deviation of a Binomial Distribution :

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.7 \times 0.3} = 4.58$$

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?

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.

Outline

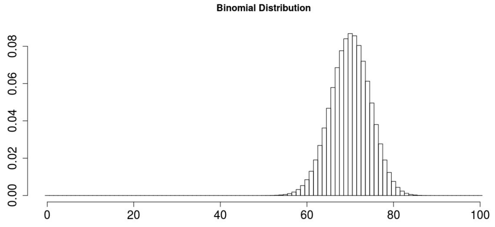
1. Housekeeping

2. Main ideas
  1. General Probability Distribution Properties: 🔍 👫 Two types of probability distributions: discrete and continuous
  2. Normal Distribution Properties:
    1. 🔍 Normal distribution is unimodal, symmetric, and follows the 68-95-99.7 rule
    2. 👤 ⚙ 🔍 👫 Z scores serve as a ruler for any distribution
  3. Binomial Distribution Properties:
    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
  4. Relationship between Binomial and Normal Distribution:
    1. 👫 ✋ 👤 💻 Shape of the binomial distribution approaches normal when the S-F rule is met

Outline

When can we approximate the binomial distribution with a normal distribution?

Why would we want to do this?



✋ 👤 Shape of the binomial distribution

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

✋ 👤 Shape of the binomial distribution

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

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

Shape of the binomial distribution

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

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

expected number of successes

expected number of failures

What is the probability that among a random sample of 1,000 Americans at least three-fourths have high-speed broadband connection at home? *(Remember probability of a randomly selected American having Broadband is p=0.70).*

What is the probability that among a random sample of 1,000 Americans at least three-fourths have high-speed broadband connection at home? *(Remember probability of a randomly selected American having Broadband is p=0.70).*

X = # of Americans in a random sample (of 1000) that have Broadband

Long and tedious way…

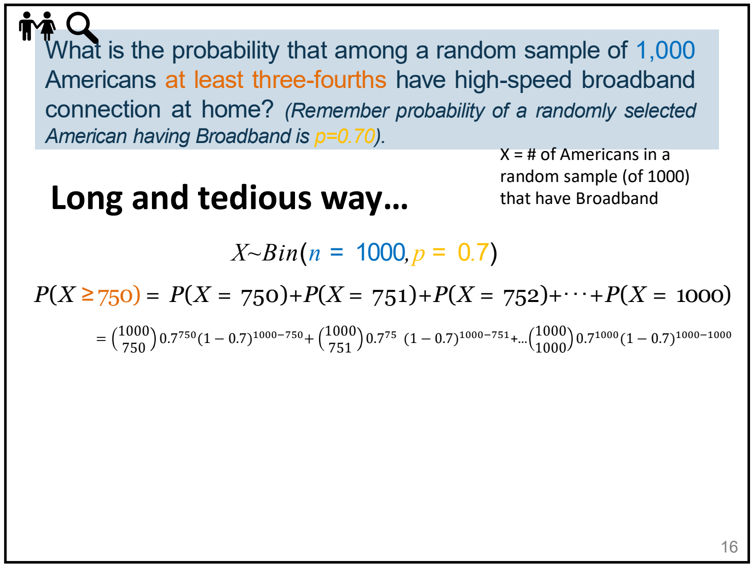
$$X \sim Binom(n = ?, p = ?)$$

What is the probability that among a random sample of 1,000 Americans at least three-fourths have high-speed broadband connection at home? *(Remember probability of a randomly selected American having Broadband is p=0.70).*

X = # of Americans in a random sample (of 1000) that have Broadband

Long and tedious way…

$$X \sim Bin(n = 1000, p = 0.7)$$



What is the probability that among a random sample of 1,000 Americans at least three-fourths have high-speed broadband connection at home? *(Remember probability of a randomly selected American having Broadband is p=0.70).*

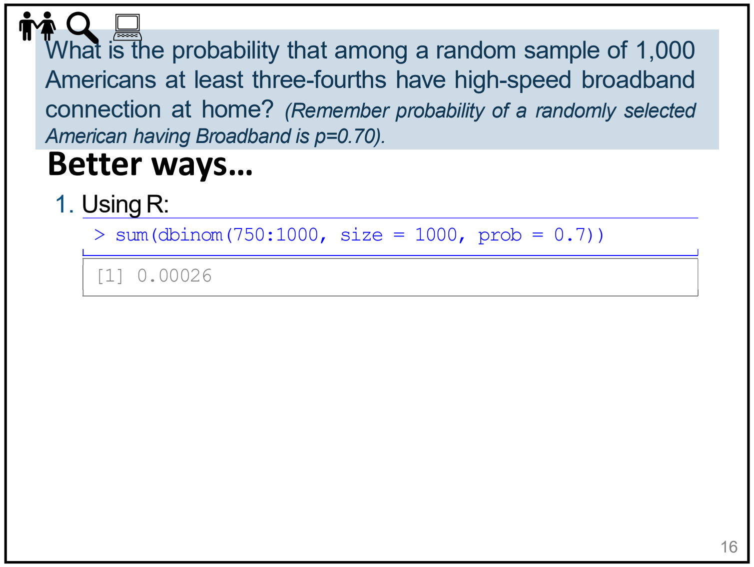
X = # of Americans in a random sample (of 1000) that have Broadband

Long and tedious way…

$$X \sim Bin(n = 1000, p = 0.7)$$

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

$$= \binom{1000}{750} 0.7^{750} (1-0.7)^{1000-750} + \binom{1000}{751} 0.7^{75} \ (1-0.7)^{1000-751} + \ldots \binom{1000}{1000} 0.7^{1000} (1-0.7)^{1000-1000}$$



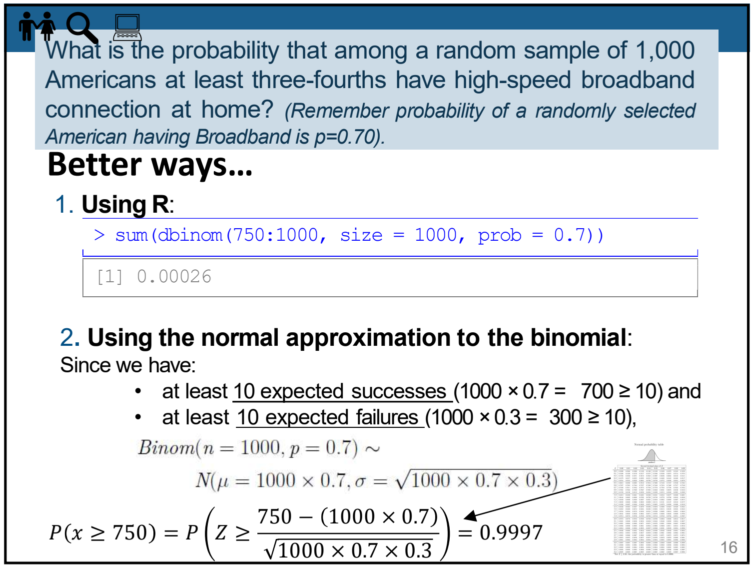
What is the probability that among a random sample of 1,000 Americans at least three-fourths have high-speed broadband connection at home? *(Remember probability of a randomly selected American having Broadband is p=0.70).*

Better ways…

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? *(Remember probability of a randomly selected American having Broadband is p=0.70).*

Better ways…

1. **Using R**:

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

```
[1] 0.00026
```

2. **Using the normal approximation to the binomial**:

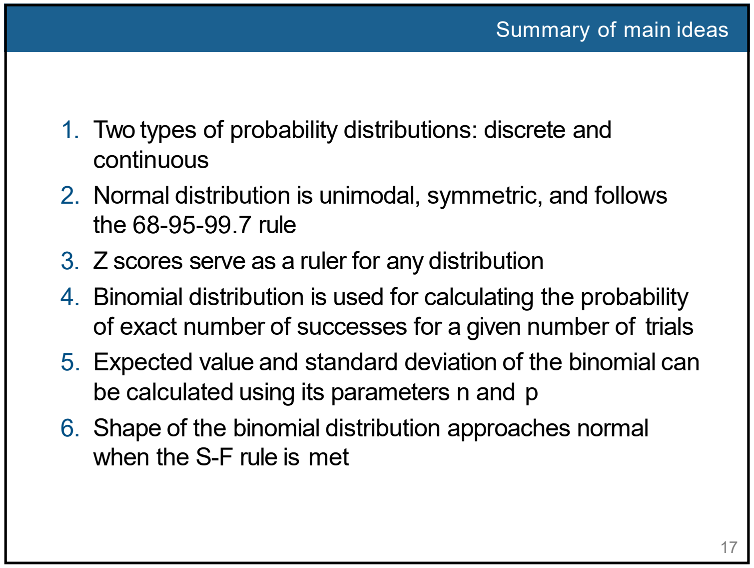
Since we have:

- at least 10 expected successes (1000 × 0.7 = 700 ≥ 10) and
- at least 10 expected failures (1000 × 0.3 = 300 ≥ 10),

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

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

$$P(x \geq 750) = P\left(Z \geq \frac{750 - (1000 \times 0.7)}{\sqrt{1000 \times 0.7 \times 0.3}}\right) = 0.9997$$



Summary of main ideas

1. Two types of probability distributions: discrete and continuous
2. Normal distribution is unimodal, symmetric, and follows the 68-95-99.7 rule
3. Z scores serve as a ruler for any distribution
4. Binomial distribution is used for calculating the probability of exact number of successes for a given number of trials
5. Expected value and standard deviation of the binomial can be calculated using its parameters n and p
6. Shape of the binomial distribution approaches normal when the S-F rule is met