STAT/SOC/CSSS 221
Statistical Concepts and Methods for the Social Sciences

Random Variables

Christopher Adolph

Department of Political Science

and

Center for Statistics and the Social Sciences

University of Washington, Seattle

Aside on mathematical notation

$x \sim \text{Normal}(\mu, \sigma)$ We read this as "x is distributed Normally with mean μ and standard devation σ"

Note: more advanced texts than ours use the variance σ^2 in place of σ in the above statement

Plan for today

We've talked about probability,

but how does it relate to social science?

Through *random variables* ,

which follow specific *probability distributions* ,

of which the best known is the *Normal distribution*

Sample space for a coin flipping example

Suppose we toss a coin twice, and record the results.

We can use a set to record this complex event.

For example, we might see a head and a tail, or H, T.

Sample space for a coin flipping example

Suppose we toss a coin twice, and record the results.

We can use a set to record this complex event.

For example, we might see a head and a tail, or H, T.

The *sample space*, or universe of possible results is in this case a set of sets:

$$\Omega = \{\{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}\}$$

Note that our sample space has separate entries for every *ordering* of heads or tails we could see.

Gets complicated fast

Sample space for two dice

Now suppose that we roll two dice. The sample space is:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Sample space for two dice

Now suppose that we roll two dice. The sample space is:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

The sum of the dice rolls for each event:

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

Note that many sums repeat

Sample spaces and complex events

The sum of the dice rolls for each event

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

What are the odds?

Outcome	2	3	4	5	6	7
Frequency						
Probability						

Outcome	8	9	10	11	12
Frequency					
Probability					

Sample spaces and complex events

The sum of the dice rolls for each event

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

What are the odds?

Outcome	2	3	4	5	6	7
Frequency	1	2	3	4	5	6
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$

Outcome	8	9	10	11	12
Frequency	5	4	3	2	1
Probability	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Sample spaces and complex events

The sum of the dice rolls for each event

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

Each event (a, b) is equally likely. But each sum, $a + b$, is not.

Outcome	2	3	4	5	6	7
Frequency	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$
Probability	0.028	0.056	0.115	0.111	0.139	0.167

Outcome	8	9	10	11	12
Frequency	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
Probability	0.139	0.111	0.083	0.056	0.028

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status
presidential popularity	each citizen's opinion
...	...

How can we reduce the space to something manageable?

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status
presidential popularity	each citizen's opinion
...	...

How can we reduce the space to something manageable?

$\rightarrow$ map the sample space Ω to one or more *random variables*:

Ω for coin flips $\rightarrow$ $X = \#\text{ of heads}$

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status
presidential popularity	each citizen's opinion
...	...

How can we reduce the space to something manageable?

$\rightarrow$ map the sample space Ω to one or more *random variables*:

Ω for coin flips $\rightarrow$ $X = \#\text{ of heads}$

Ω for military casualties $\rightarrow$ $D = \#\text{ of deaths}$

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status
presidential popularity	each citizen's opinion
...	...

How can we reduce the space to something manageable?

$\rightarrow$ map the sample space Ω to one or more *random variables*:

Ω for coin flips $\rightarrow$ $X = \#$ of heads

Ω for military casualties $\rightarrow$ $D = \#$ of deaths

Ω for education outcomes $\rightarrow$ $Y = \#$ of students passing

Ω for presidential popularity $\rightarrow$ $S = \#$ support pres

The trouble with sample spaces

For most processes we could study, the sample space of all events is huge:

Process	Events in sample space is all combinations of:
coin flips	each coin's result
military casualties	each soldier's status
education outcomes	each student's passing status
presidential popularity	each citizen's opinion
...	...

How can we reduce the space to something manageable?

$\rightarrow$ map the sample space Ω to one or more *random variables*:

Ω for coin flips $\rightarrow$ $X = \#$ of heads

Ω for military casualties $\rightarrow$ $D = \#$ of deaths

Ω for education outcomes $\rightarrow$ $Y = \#$ of students passing

Ω for presidential popularity $\rightarrow$ $S = \#$ support pres

This mapping can produce discrete or continous variables,
and each will have a different distribution of probabilities

Probability for random variables

Consider the random variable $X = \#\text{ of heads in } M \text{ coin flips}$

Five things we'd like to know about the theoretical distribution of X:

$\Pr(X)$ How do we summarize the random distribution of X?

Probability for random variables

Consider the random variable $X = \#\text{ of heads in } M \text{ coin flips}$

Five things we'd like to know about the theoretical distribution of X:

$\Pr(X)$ How do we summarize the random distribution of X?

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?

Probability for random variables

Consider the random variable $X = \#\text{ of heads in } M \text{ coin flips}$

Five things we'd like to know about the theoretical distribution of X:

$\Pr(X)$ How do we summarize the random distribution of X?

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?

$\Pr(X \leq x)$ What is the probability that X is *at least* equal to some specific value, like $x = 1$?

Probability for random variables

Consider the random variable $X = \#\text{ of heads in } M \text{ coin flips}$

Five things we'd like to know about the theoretical distribution of X:

$\Pr(X)$ How do we summarize the random distribution of X?

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?

$\Pr(X \leq x)$ What is the probability that X is *at least* equal to some specific value, like $x = 1$?

$\mathrm{E}(X)$ What is the expected number of heads we will see on average?

Probability for random variables

Consider the random variable $X = \#\text{ of heads in } M \text{ coin flips}$

Five things we'd like to know about the theoretical distribution of X:

$\Pr(X)$ How do we summarize the random distribution of X?

$\Pr(X = x)$ What is the probability X is some specific value like $x = 1$?

$\Pr(X \leq x)$ What is the probability that X is *at least* equal to some specific value, like $x = 1$?

$\mathrm{E}(X)$ What is the expected number of heads we will see on average?

$\mathrm{sd}(X)$ On average, how much do we expect a given random outcome to differ from the expected result?

Understanding distributions by random simulation

The easiest way to understand probability distributions is by simulation

1. Using a coin, deck of cards, or a computer, we generate random events
2. We repeat step 1 *many* times, and record each result
3. We look at the distribution of results across these simulations to understand the underlying probability distribution of the random event

Coin flipping simulation

Let's create a simulation

We'll imagine that in tomorrow morning's section,
you each flip a coin, and your TA record the total sum of heads

This seems silly: coins aren't social phenomena,
so why are we studying them in a social statistics class?

Coins and (interesting) binary variables

Coins are just an example of a random binary variable

There are lots of interesting random binary variables:

- whether you vote Democrat or Republican in November
- whether you go to college
- whether commit a crime

Each of these is a stochastic variable consisting of a random part and a deterministic part

Understanding the random part well will help us isolate the deterministic part and study it

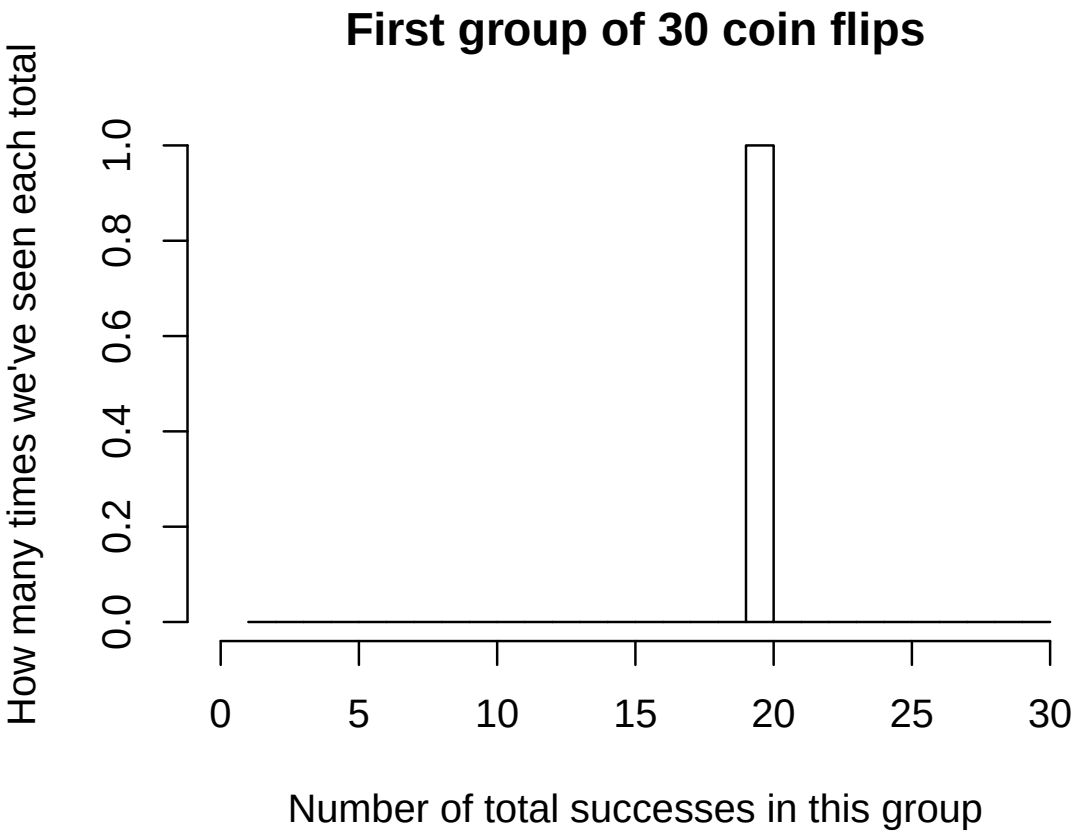
Coins and (interesting) binary variables

The sum of a binary variable across a group is often very interesting

We will look at the total number of students in your section who got “heads” in your coin flip

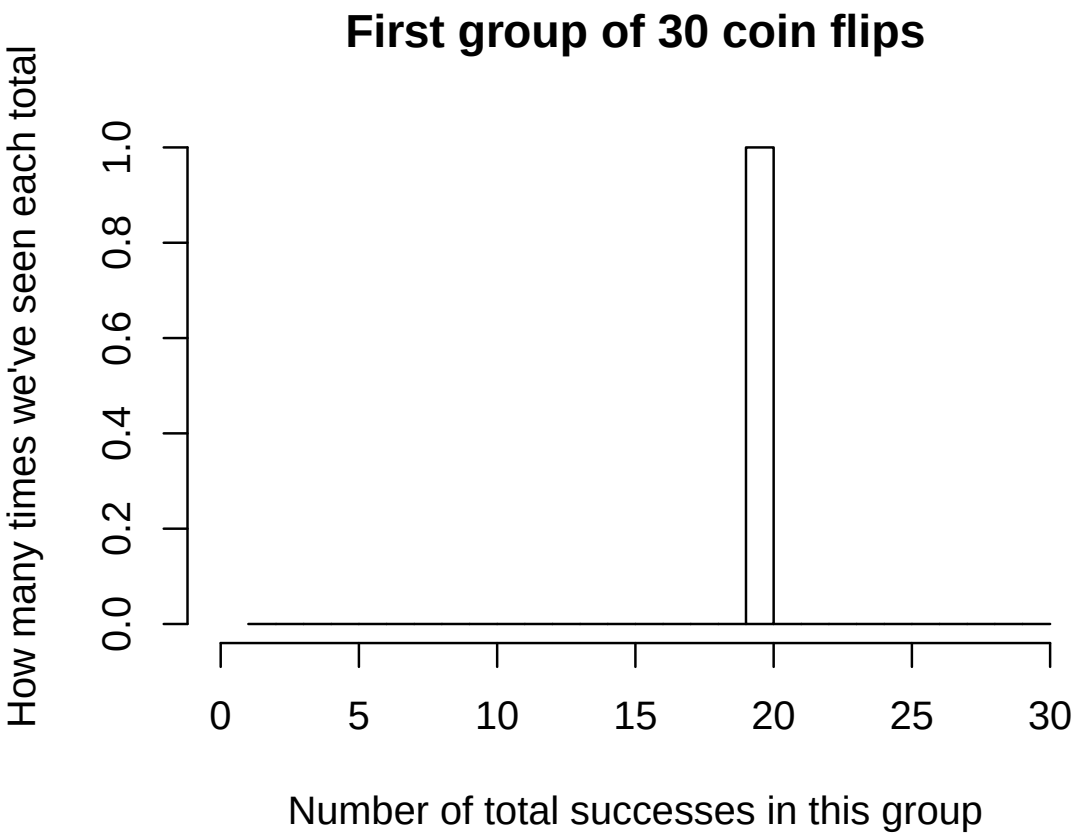
In actual social science, we might look at:

- the number of Democratic votes in a county
- the number of people in a high school class that go to college
- the number of people in a neighborhood who committed a crime



Each person in your section flips a coin, and we see 20 heads:

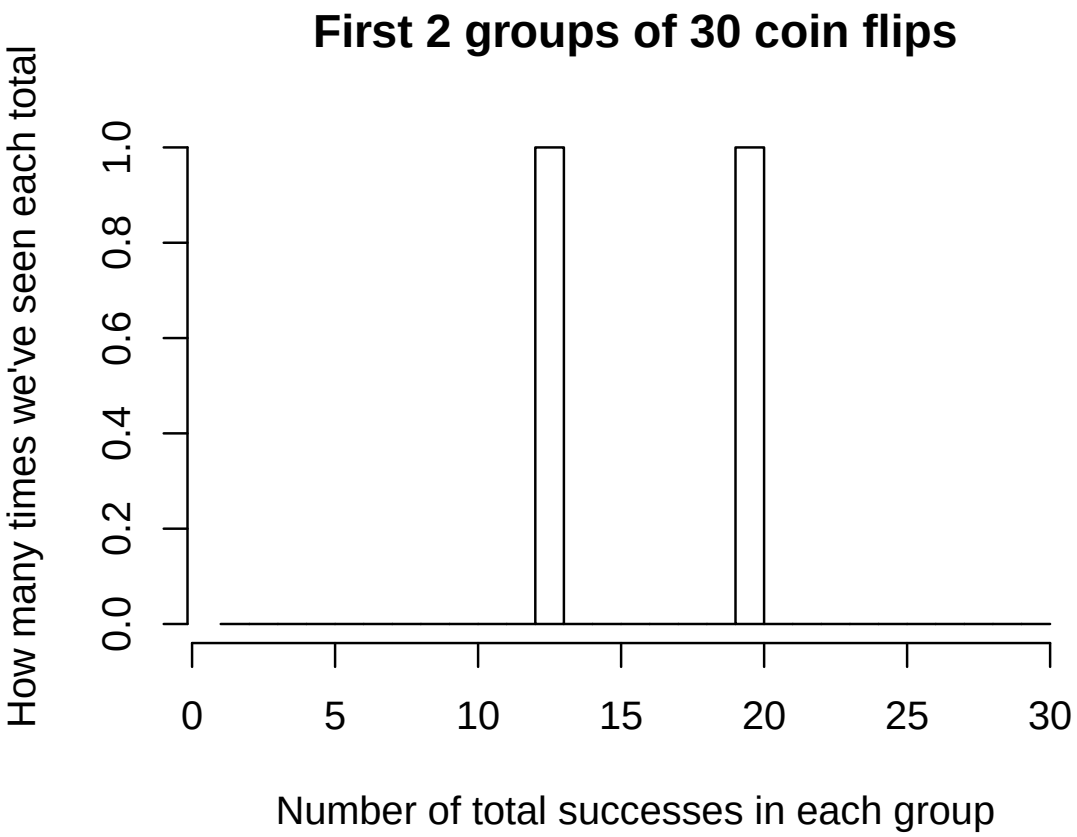
H T H T H T T H H H H
T T H H H H T T H H H
H H H H T H H T



Each person in your section flips a coin, and we see 20 heads

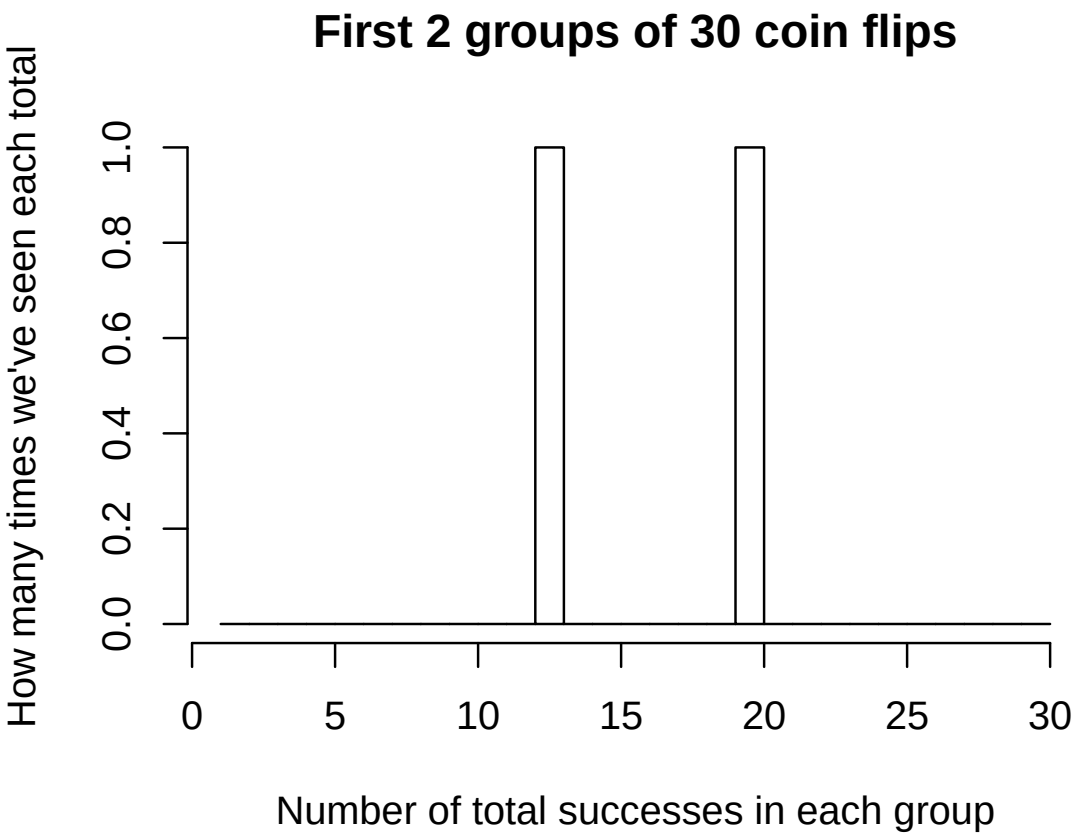
We record this in the histogram at the left.

Let's repeat the experiment, and see what we get this time



Each person in your section flips another coin, and we see 13 heads:

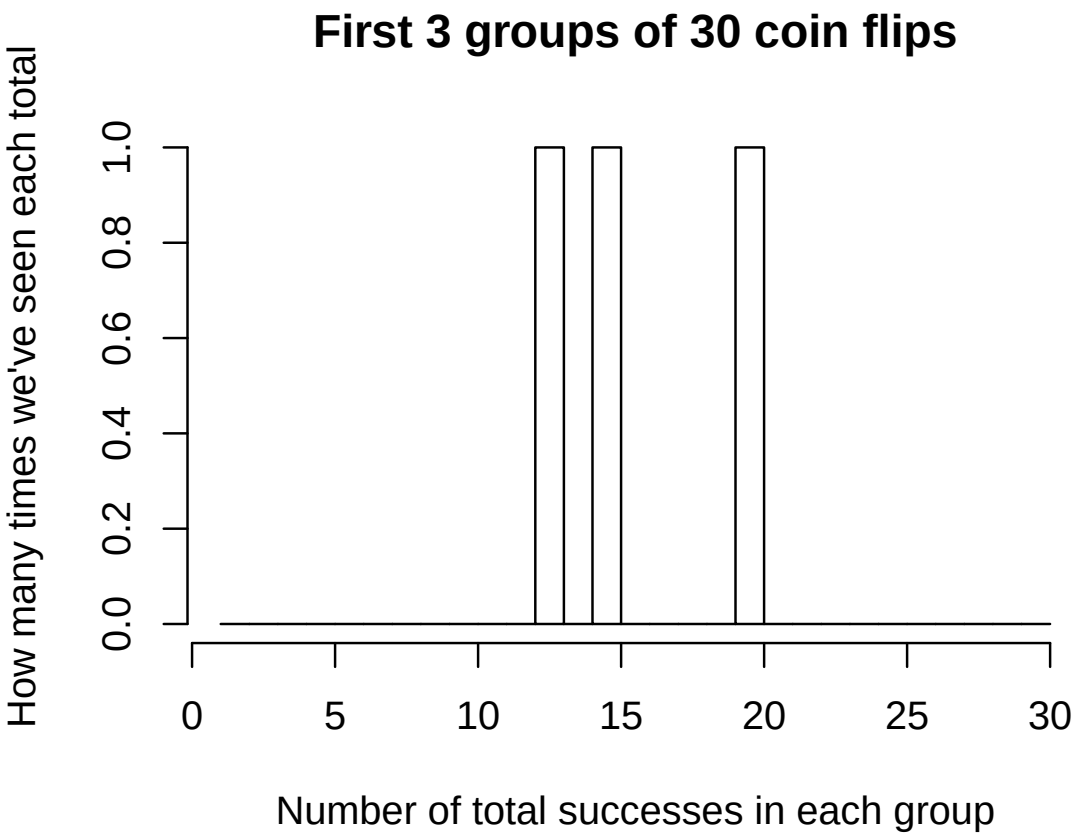
H T T H T T H H T H H
H T T H T H T T T T T
H T H H T T H T



Each person in your section flips another coin, and we see 13 heads:

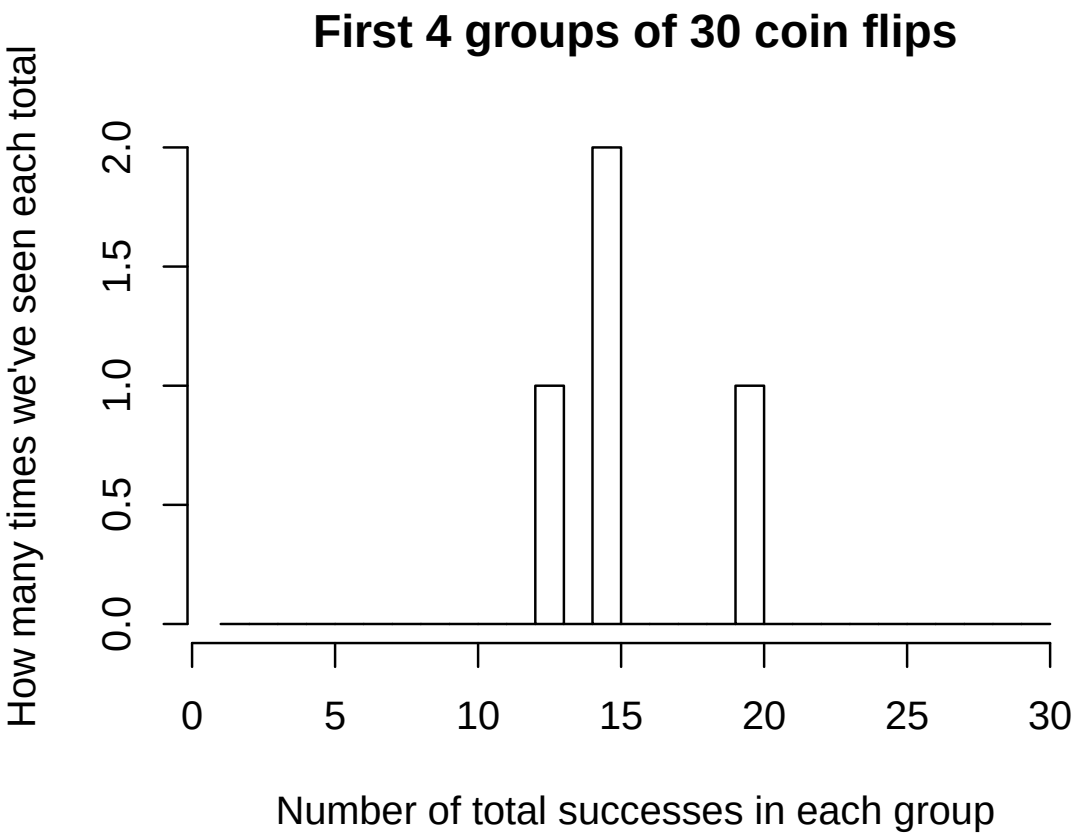
Notice that we've added a second result to the histogram

We can keep going. Watch the vertical axis carefully!



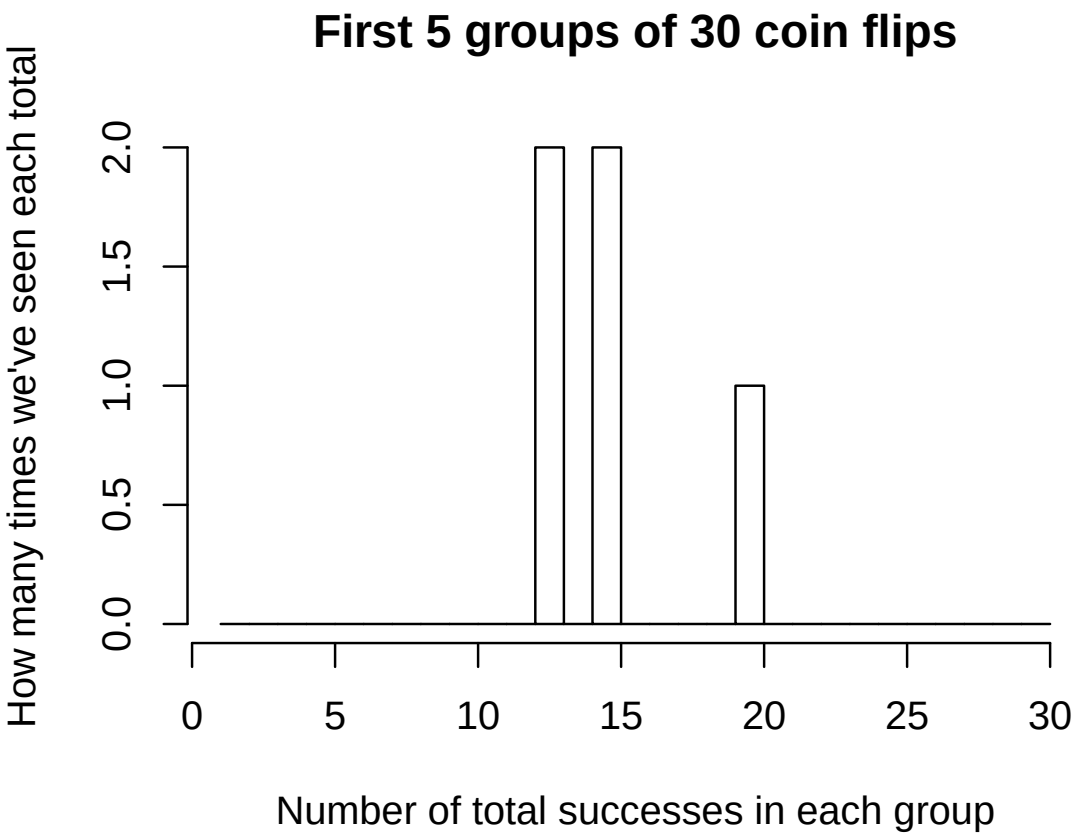
Each person in your section flips another coin, and we see 15 heads:

H T H H H T H H T T T
H H T T T T H H H T T
H T H T T T H H



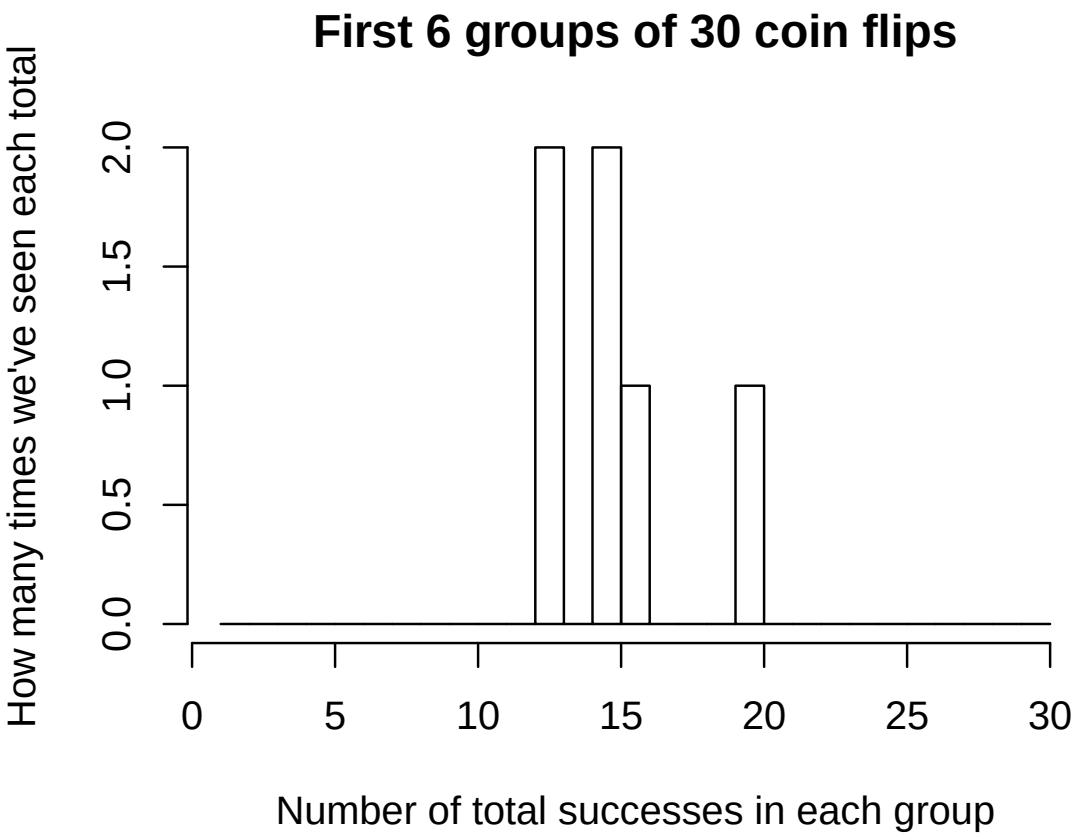
Each person in your section flips another coin, and we see 15 heads:

T H H T H H T H T H H
T H T H T T T T T H T
T H H H H T H T



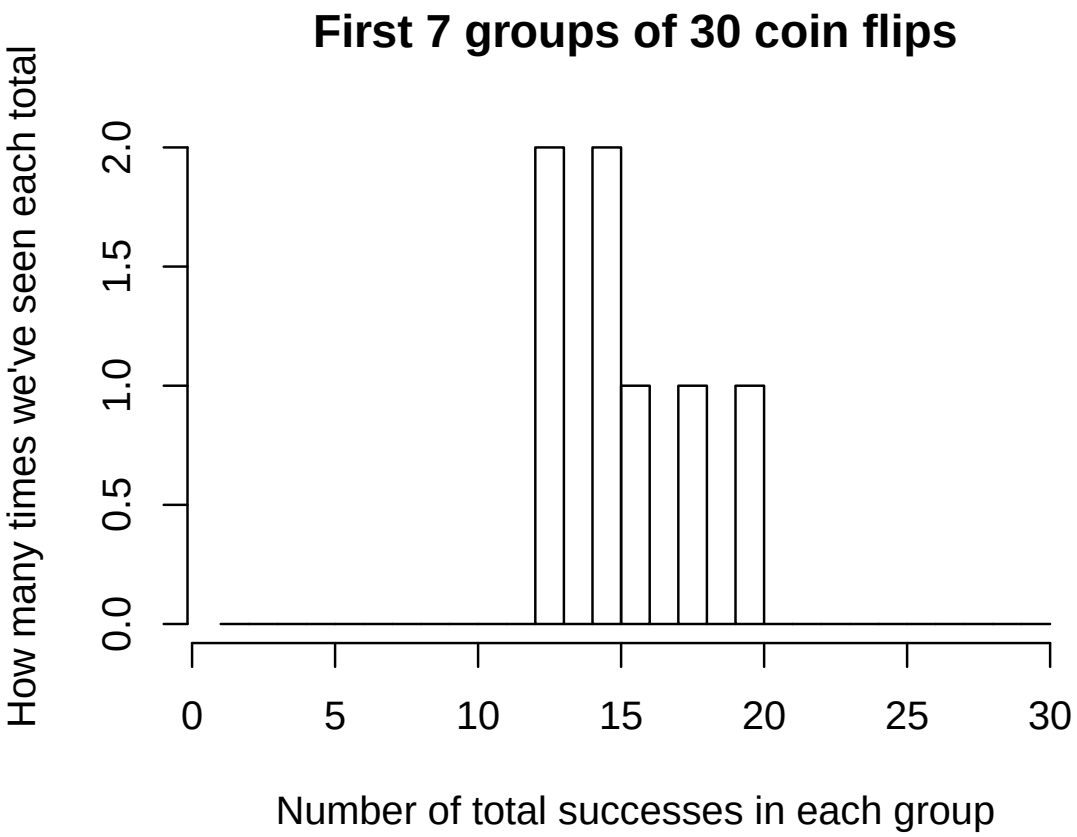
Each person in your section flips another coin, and we see 13 heads:

H H T T H T T T T T H
T H H T H H T T H T H
H T T T H T H T



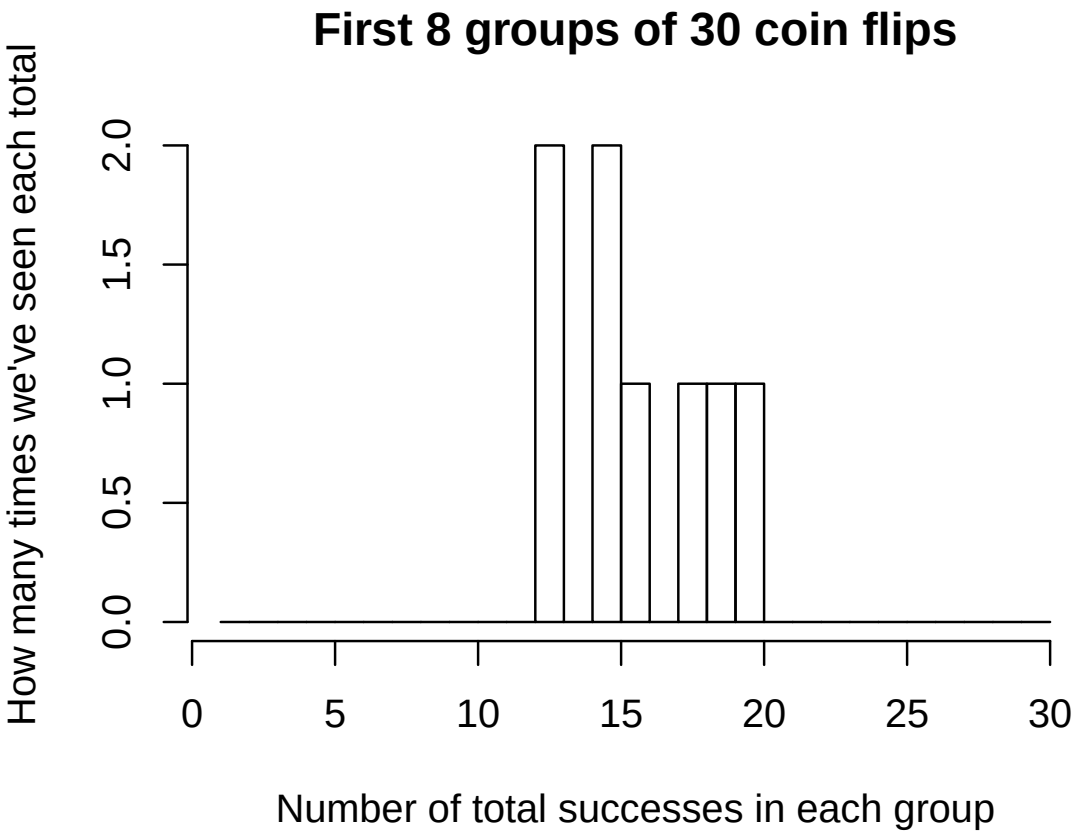
Each person in your section flips another coin, and we see 16 heads:

T H H T T H T T H H H
H H T T H T T H H H H
T H T T H H T T



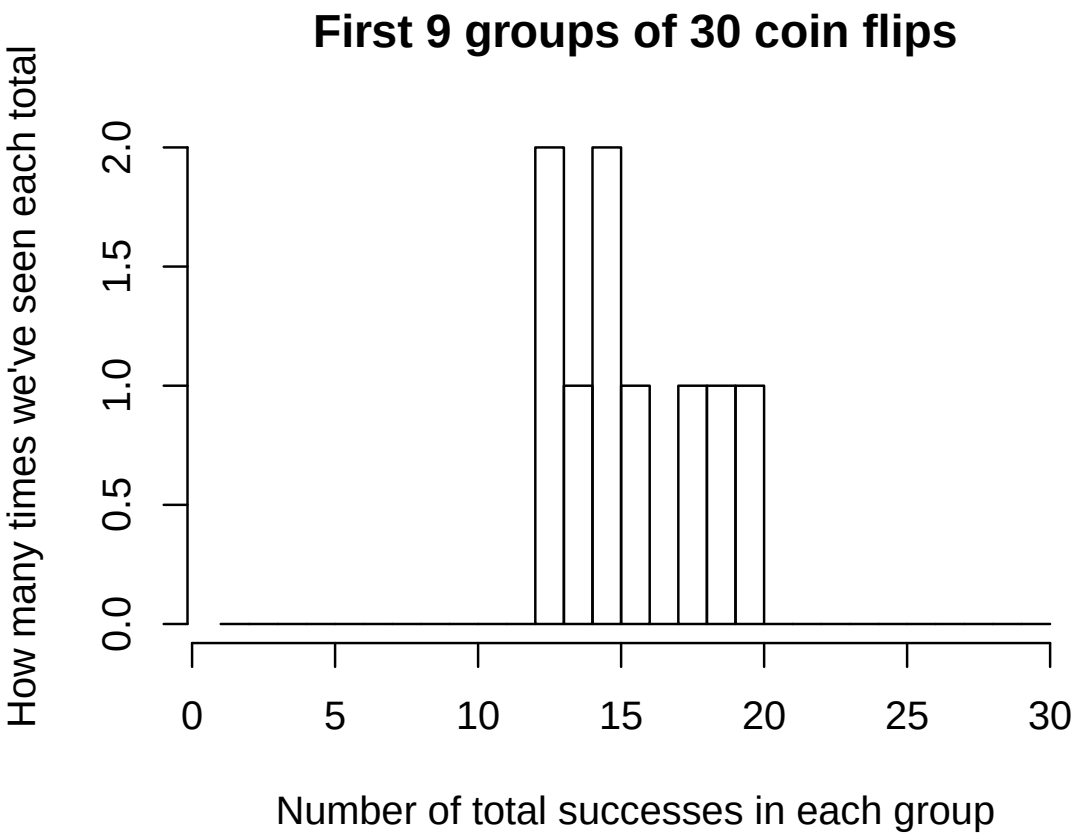
Each person in your section flips another coin, and we see 18 heads:

H H T T H H H H H H T
T H H T H H T T H T H
H H T H T T H T



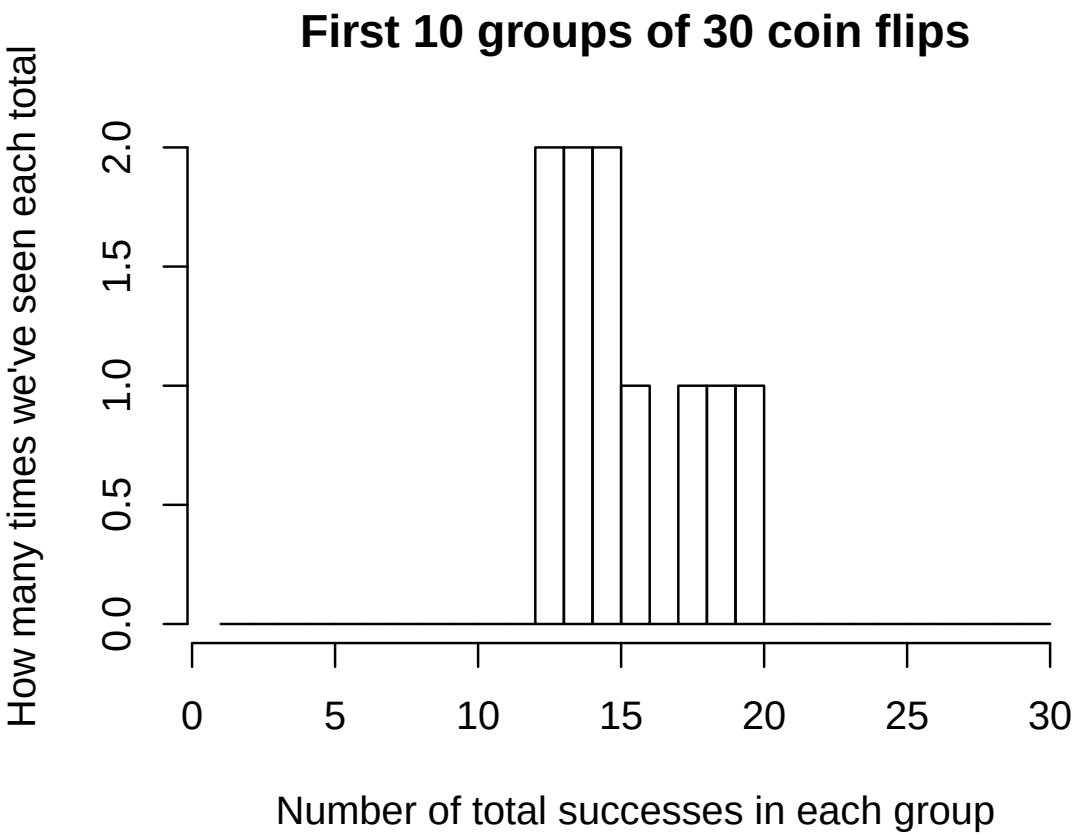
Each person in your section flips another coin, and we see 19 heads:

H T H H H H H H H H
H T H T H T T H H H H
H T T T T H T T H



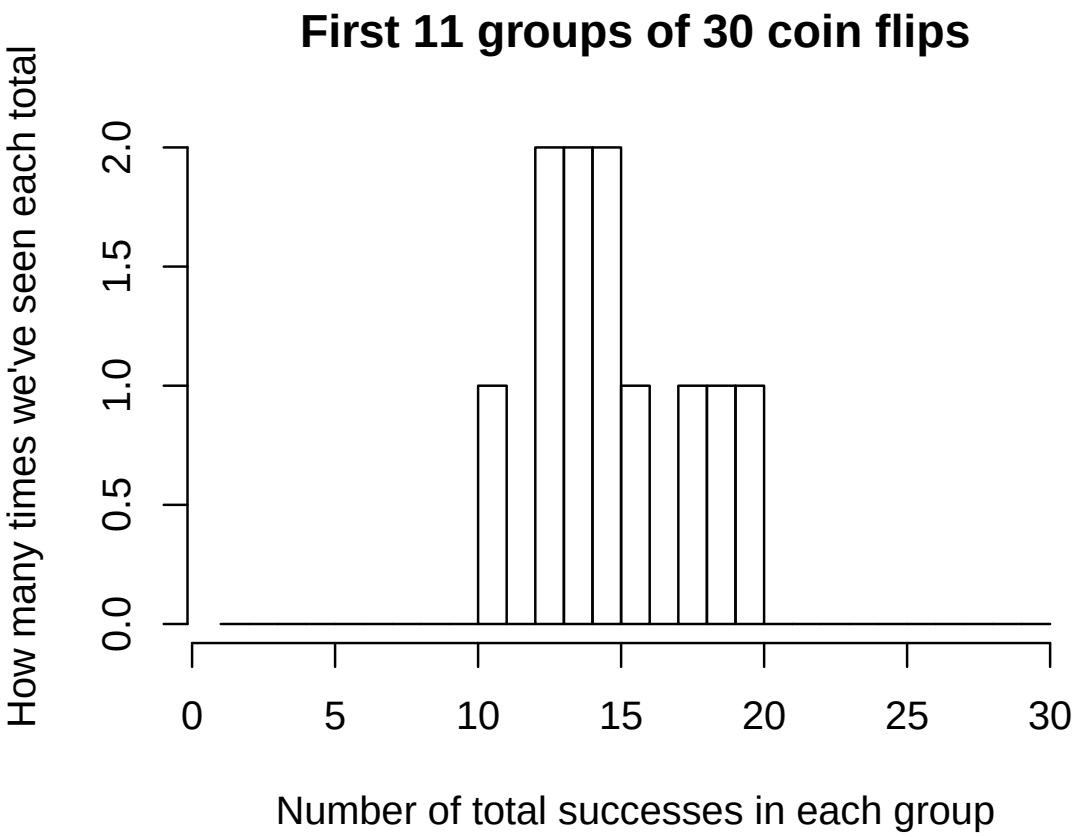
Each person in your section flips another coin, and we see 14 heads:

H H T H H T H T T T H
T H T H T T T T H T H
H T H T H H T T



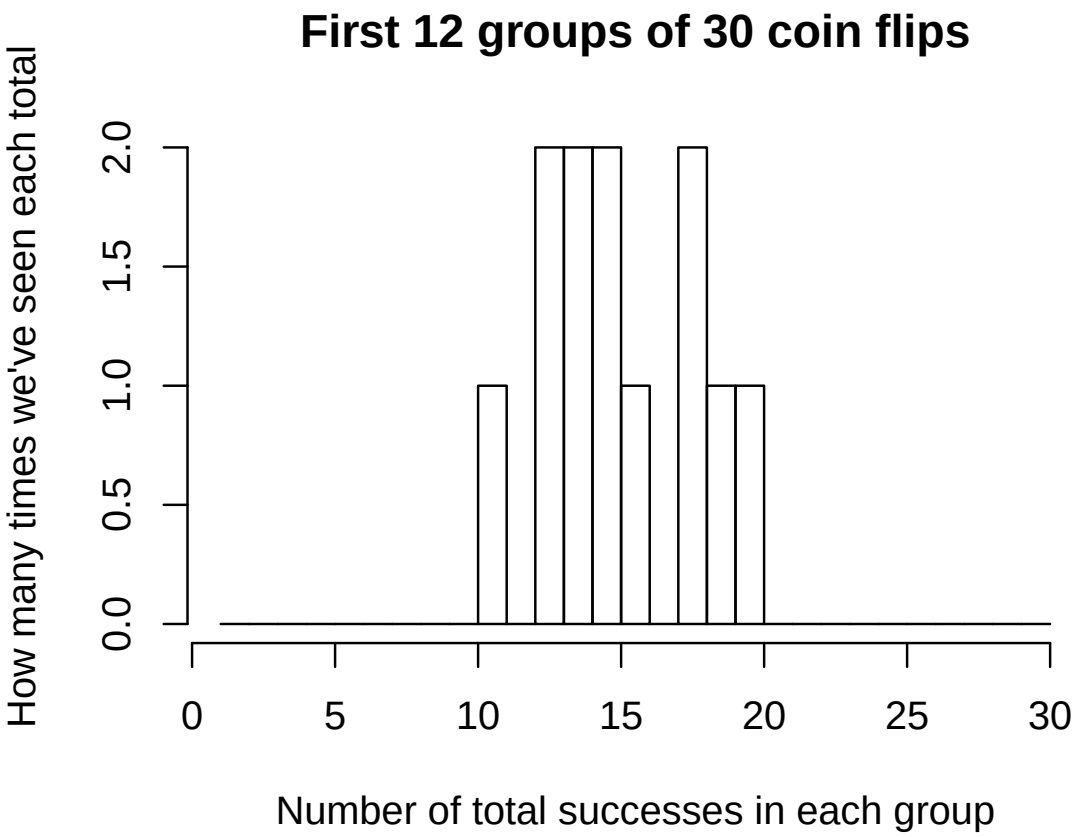
Each person in your section flips another coin, and we see 14 heads:

T T H H T H T H T T T
H H T H T H T T T H T
T T H H H T H H



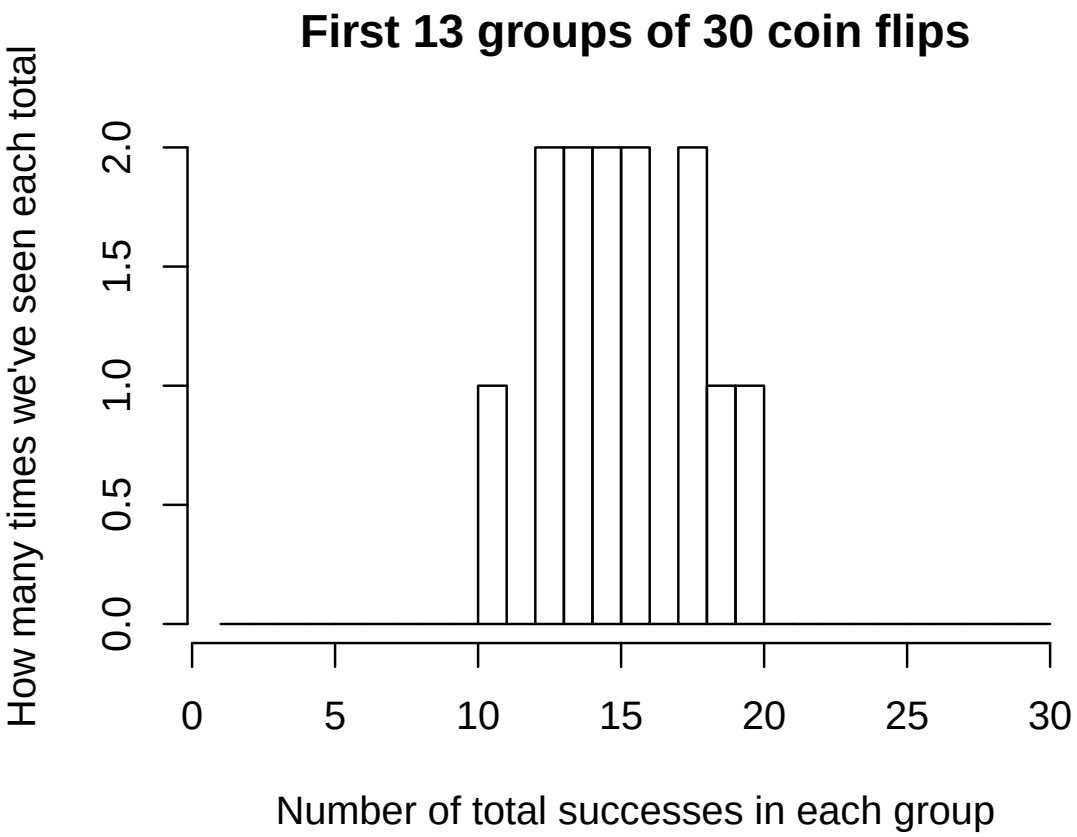
Each person in your section flips another coin, and we see 11 heads:

T T T T T H T H T T H
H T H H H T T H T T H
T H T T T T H T



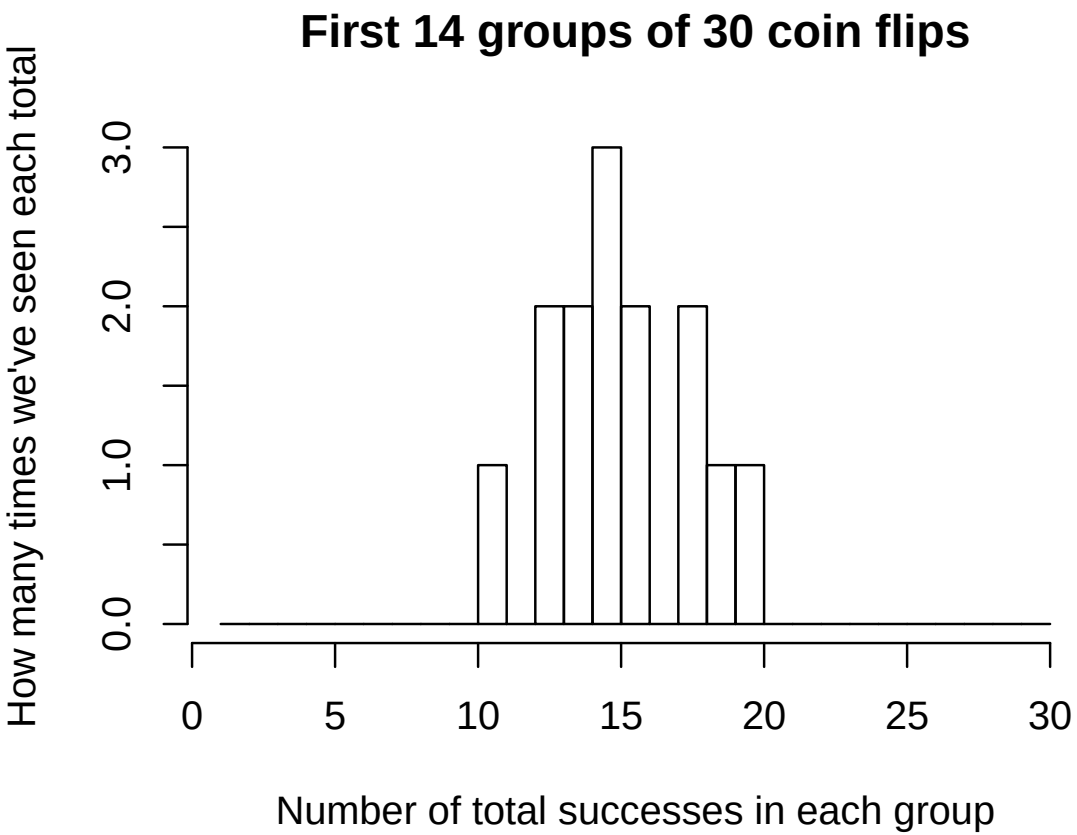
Each person in your section flips another coin, and we see 18 heads:

T H T T H T H H T T T
T H H H T H T H H T H
H H H H H T H H



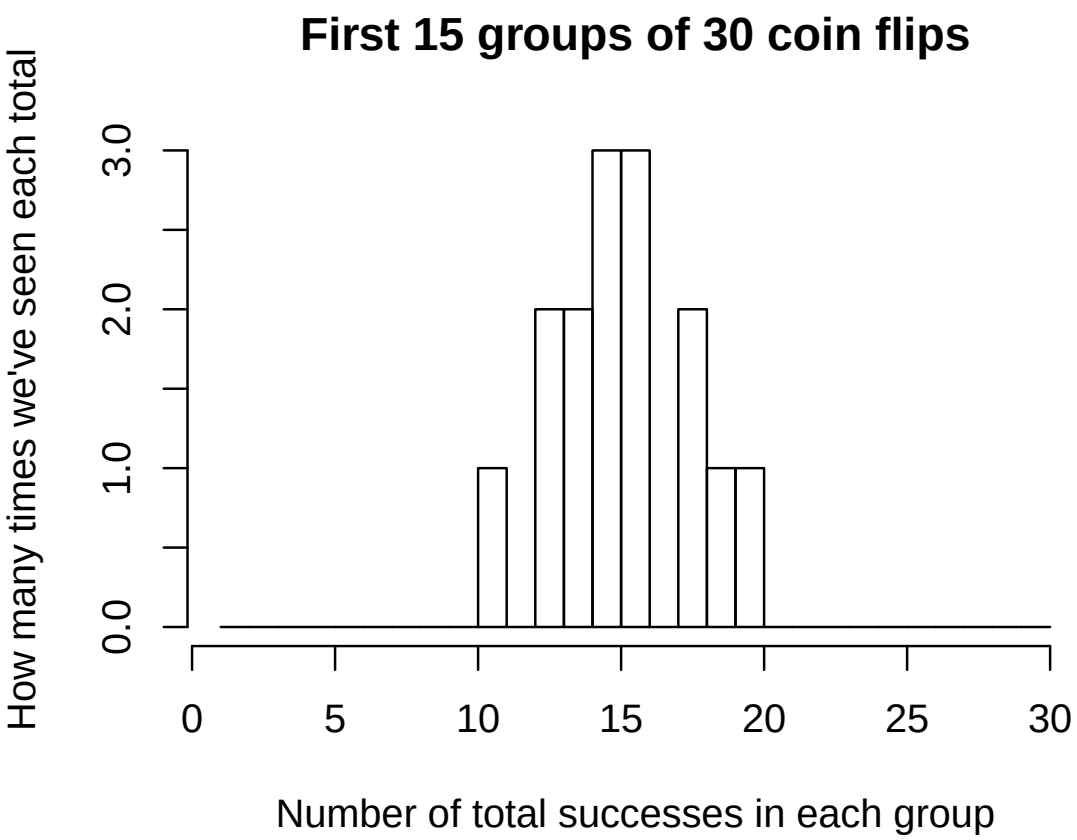
Each person in your section flips another coin, and we see 16 heads:

T H T H H T T T H H T
T H T H T T H H H T T
H H H H T T H H



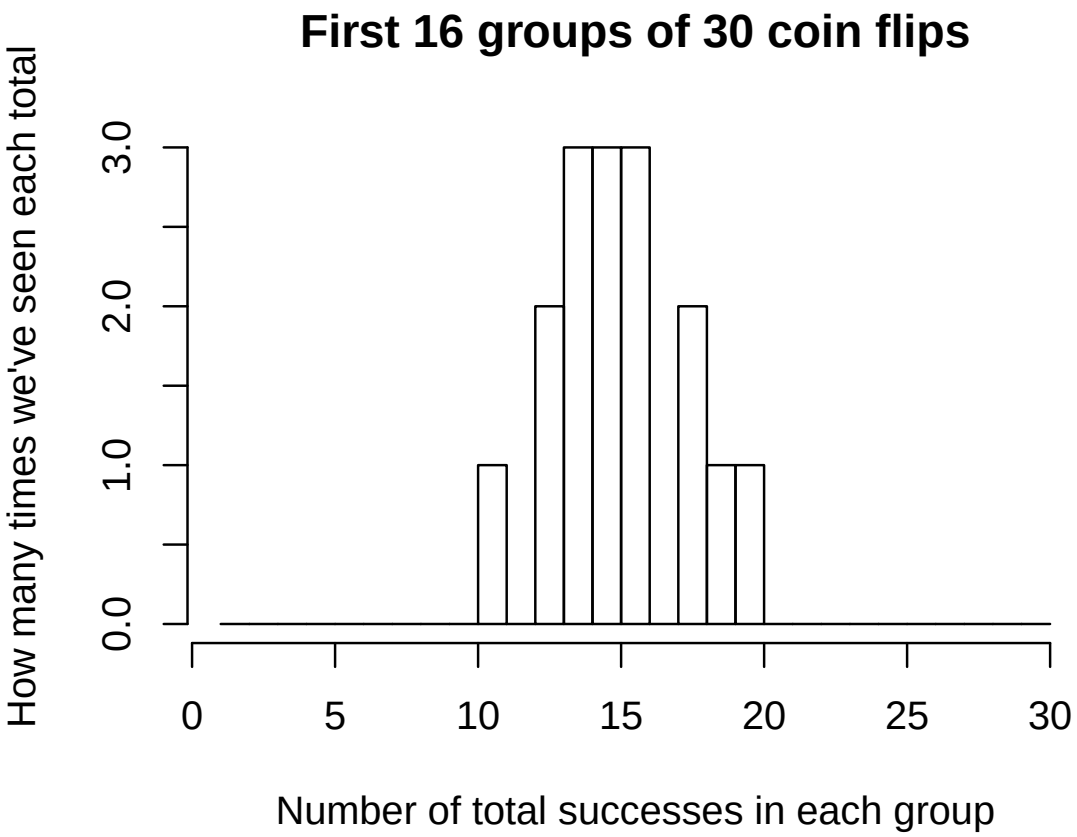
Each person in your section flips another coin, and we see 15 heads:

T T H H H T H H H T T
T T H T H T H T H H T
T T T H T H H H



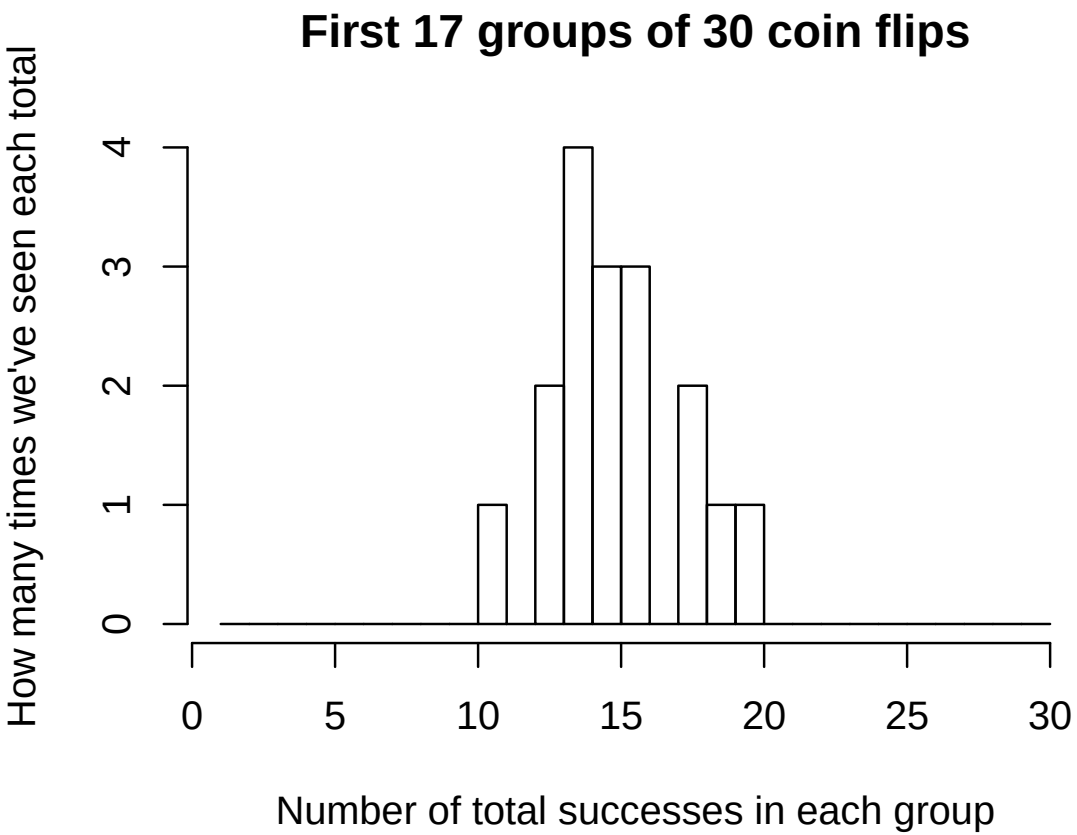
Each person in your section flips another coin, and we see 16 heads:

H H H T T H T T H T H
T H T T H T H H H T T
H T H H H T H T



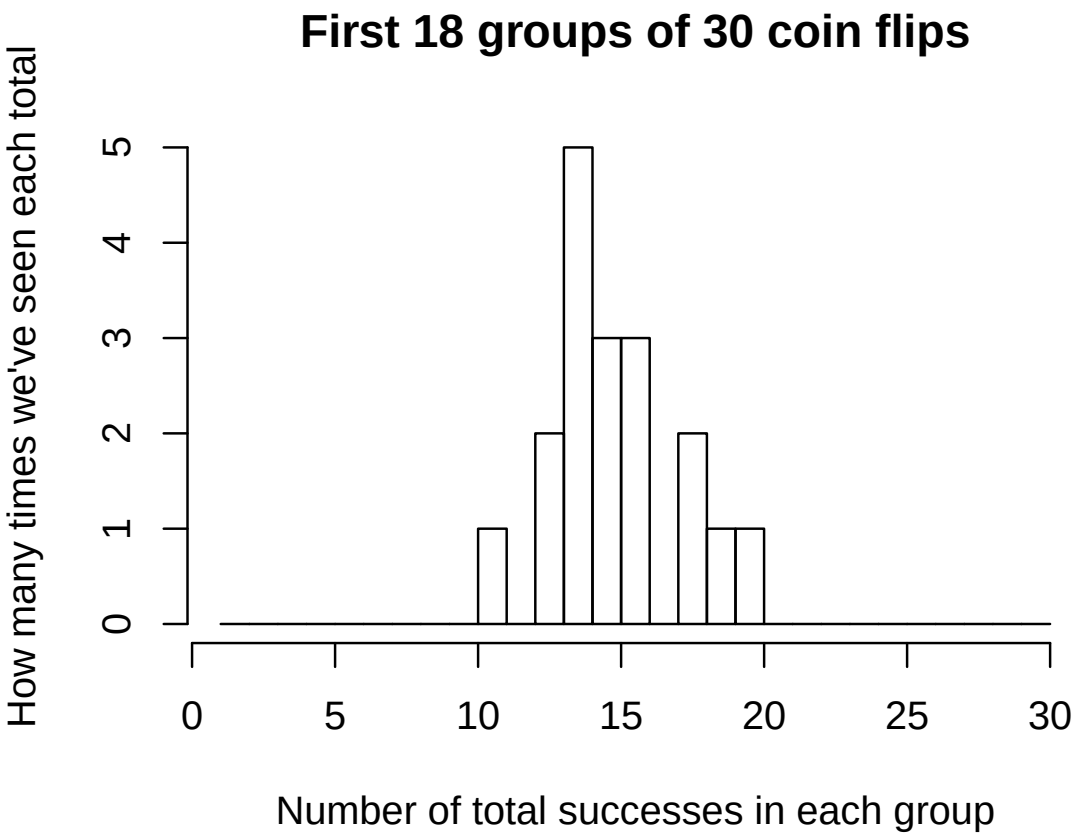
Each person in your section flips another coin, and we see 14 heads:

H H T H T H T T T T T
H H H H H T T T H H H
T T T H T T H T



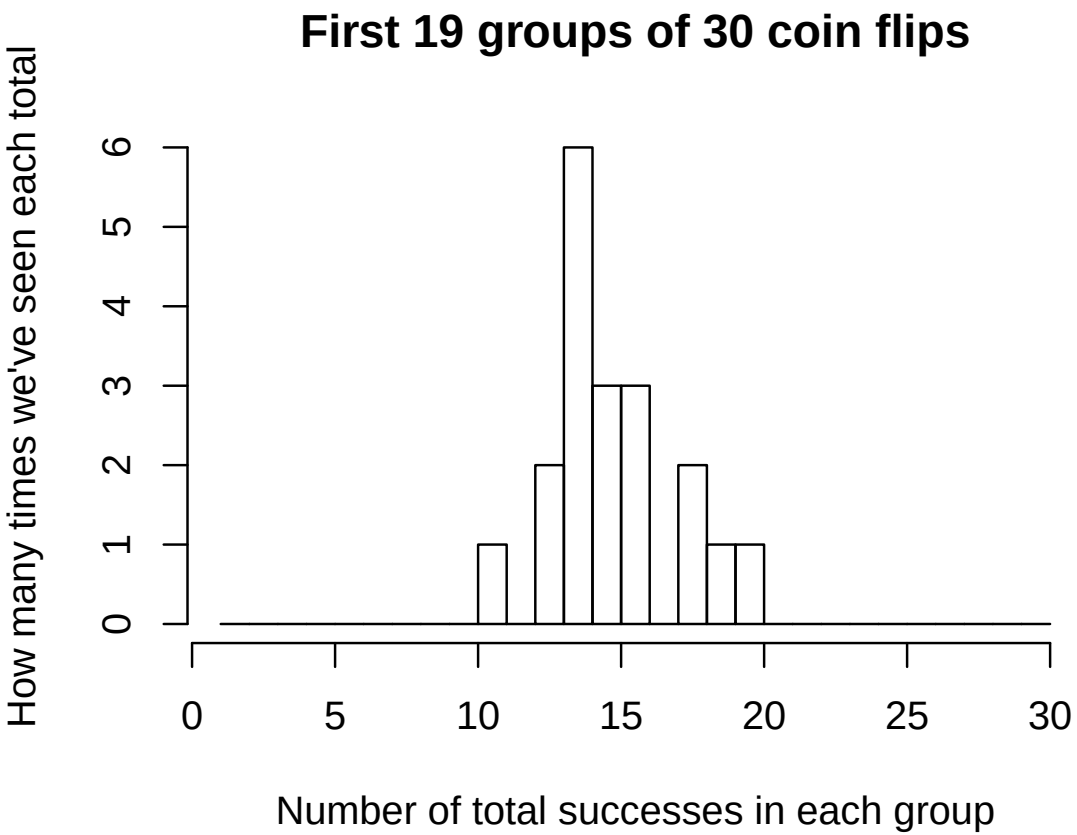
Each person in your section flips another coin, and we see 14 heads:

H T H H T T H T T T T
T H H T T T H H H T H
T H H H T T H T



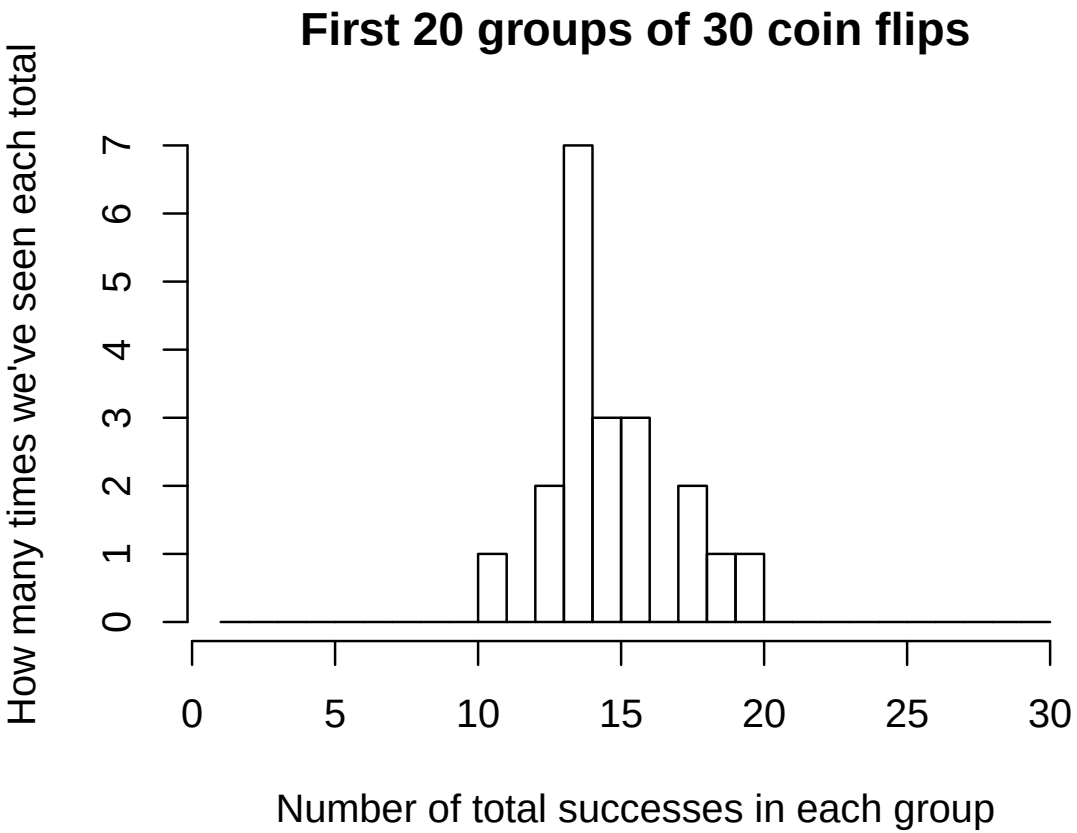
Each person in your section flips another coin, and we see 14 heads:

H T H H T T T H H H T
H H T H H T T H T T T
T H T T T H T H



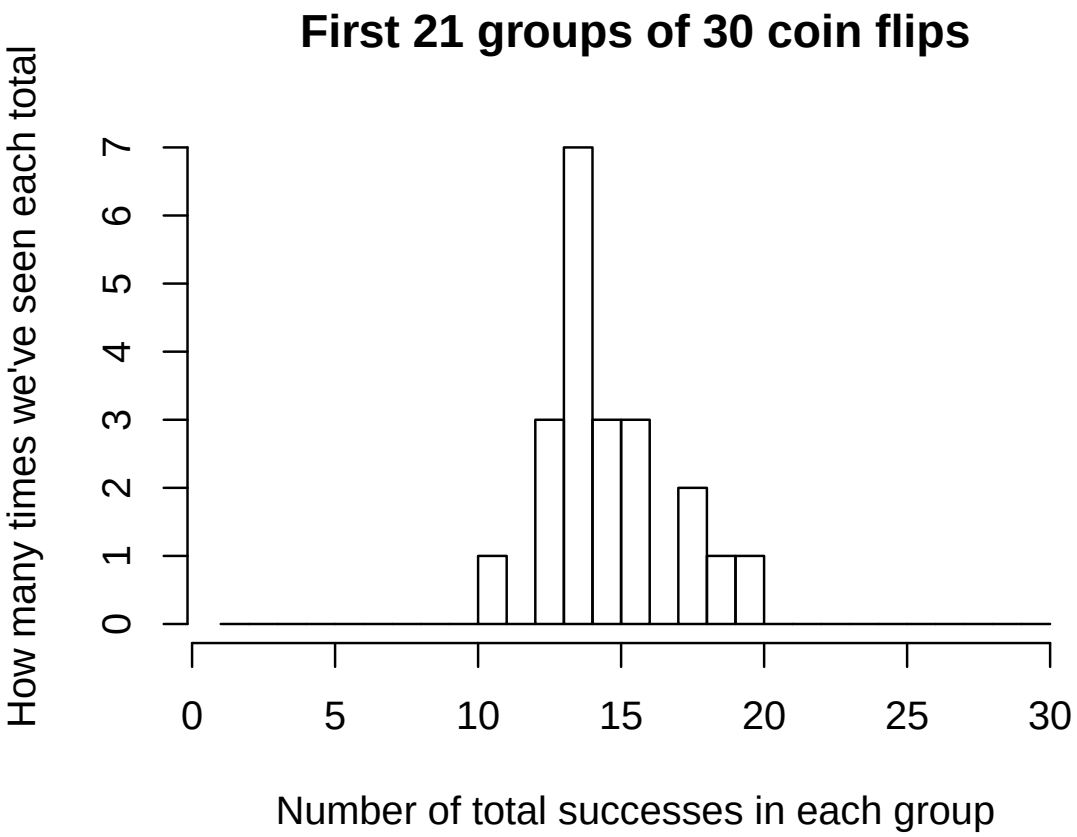
Each person in your section flips another coin, and we see 14 heads:

H T T H H T T T H H T
T T T H H T H H H T H
T H H H T T T T



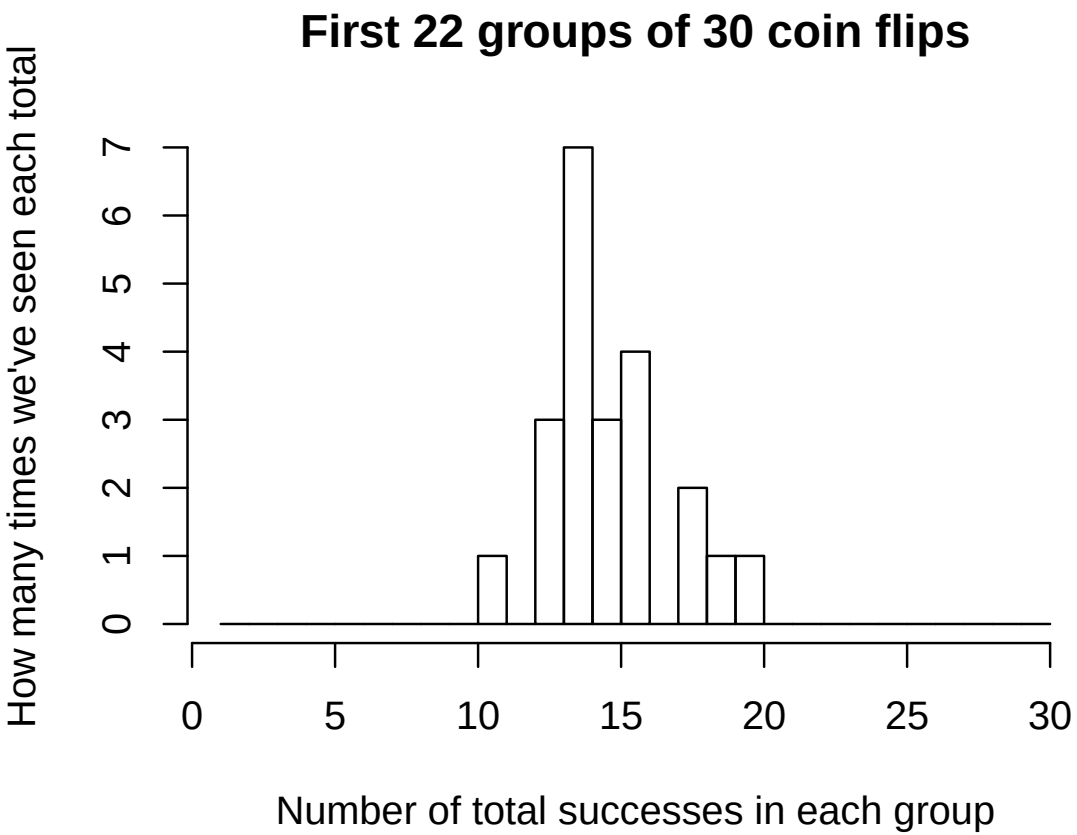
Each person in your section flips another coin, and we see 14 heads:

H T T H H T T T T T H
T T H T H H T T H H H
T H T H H T H T



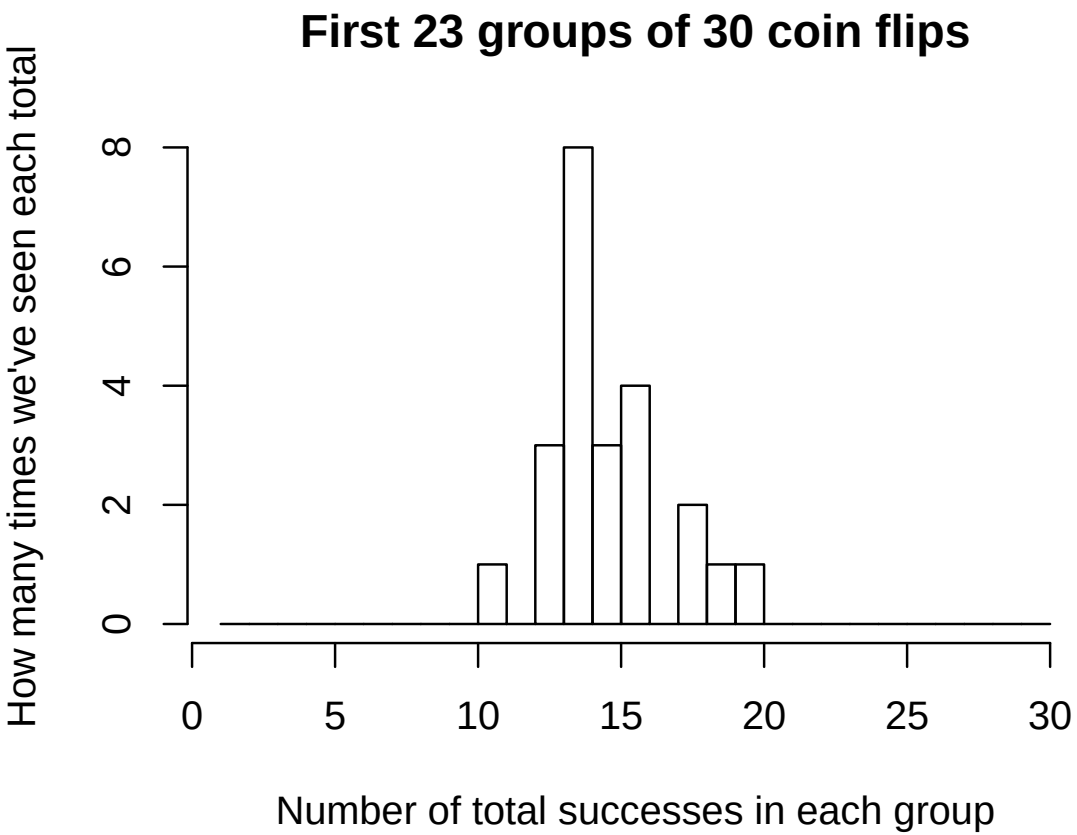
Each person in your section flips another coin, and we see 13 heads:

T T T H H T T T H T T
T T T T H T H T H H H
T H H H T H H T



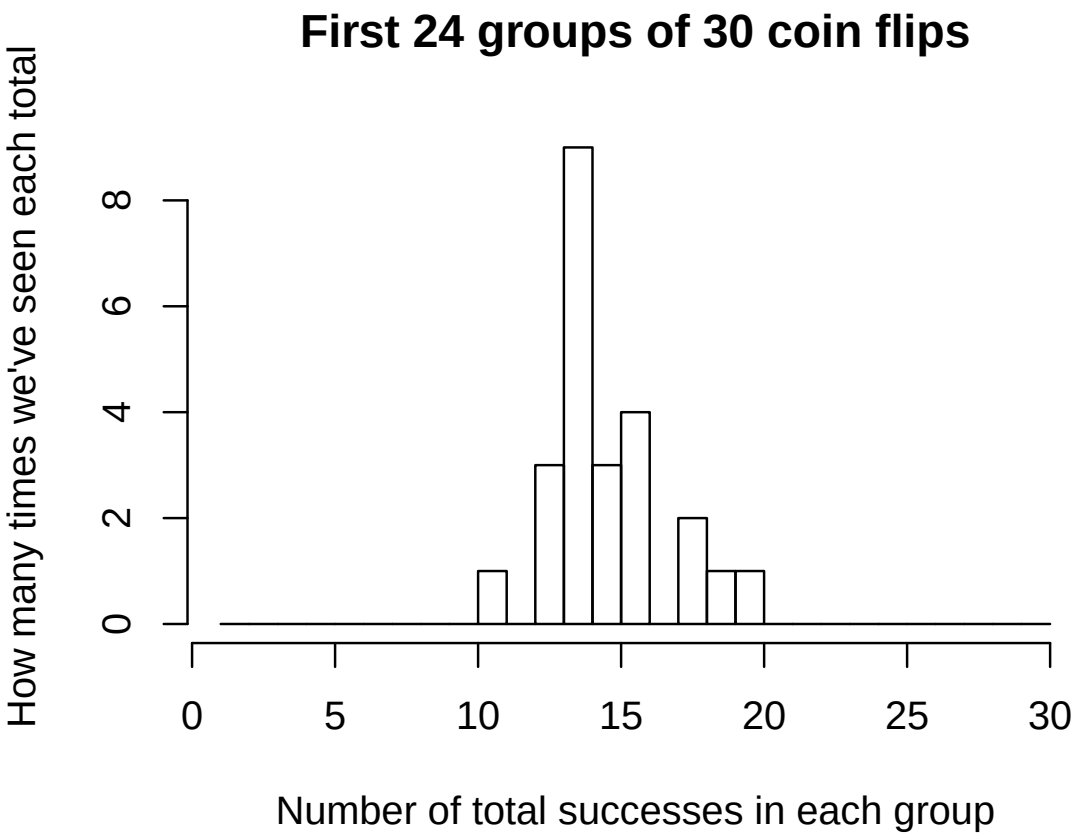
Each person in your section flips another coin, and we see 16 heads:

T H H T H T H T H H T
H H T H H T H T T H T
H T T H T T H H



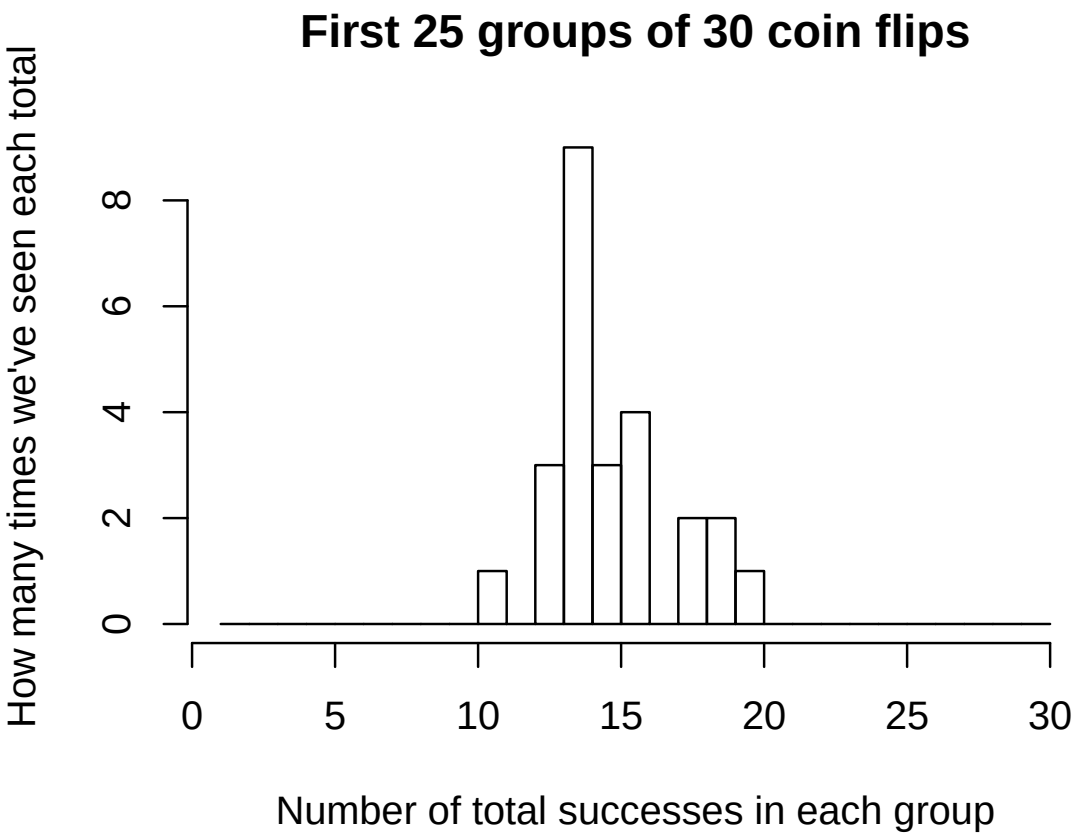
Each person in your section flips another coin, and we see 14 heads:

T T H H H T T H T T H
T H T H H T T H T H T
H T H T T T H H



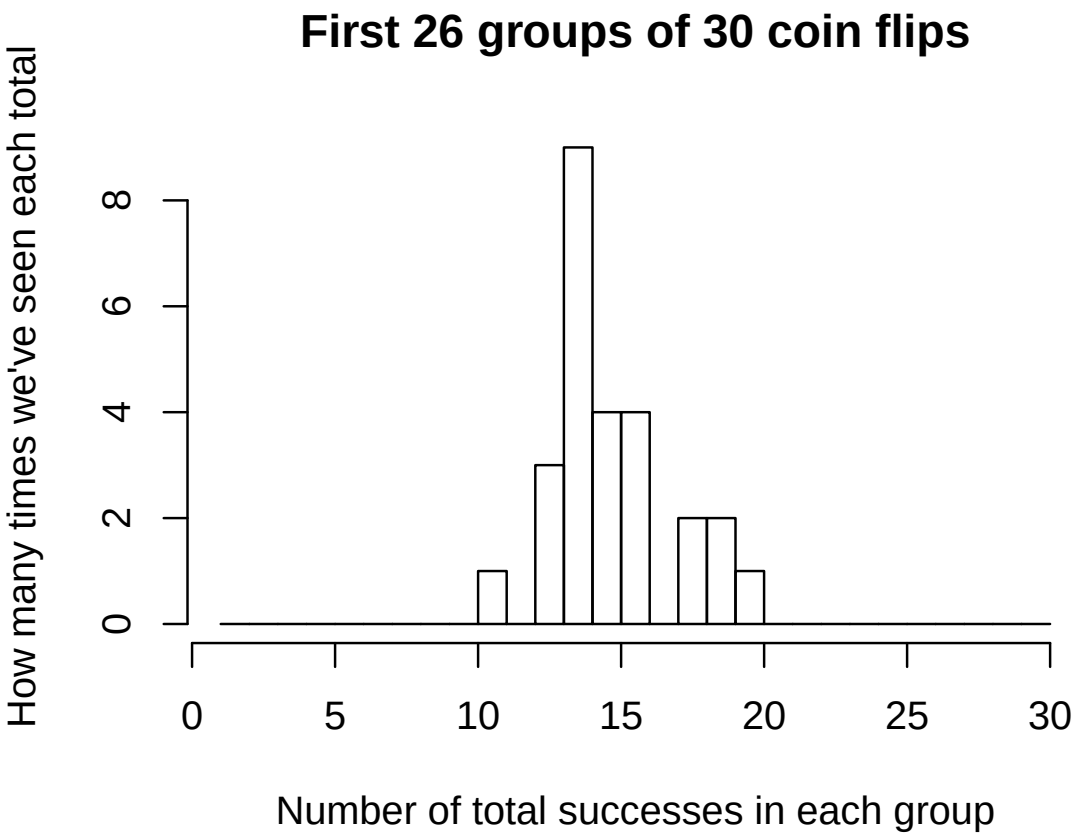
Each person in your section flips another coin, and we see 14 heads:

T H T H T H T H H H T
H T T H T T H T T H T
H H T T T H T H



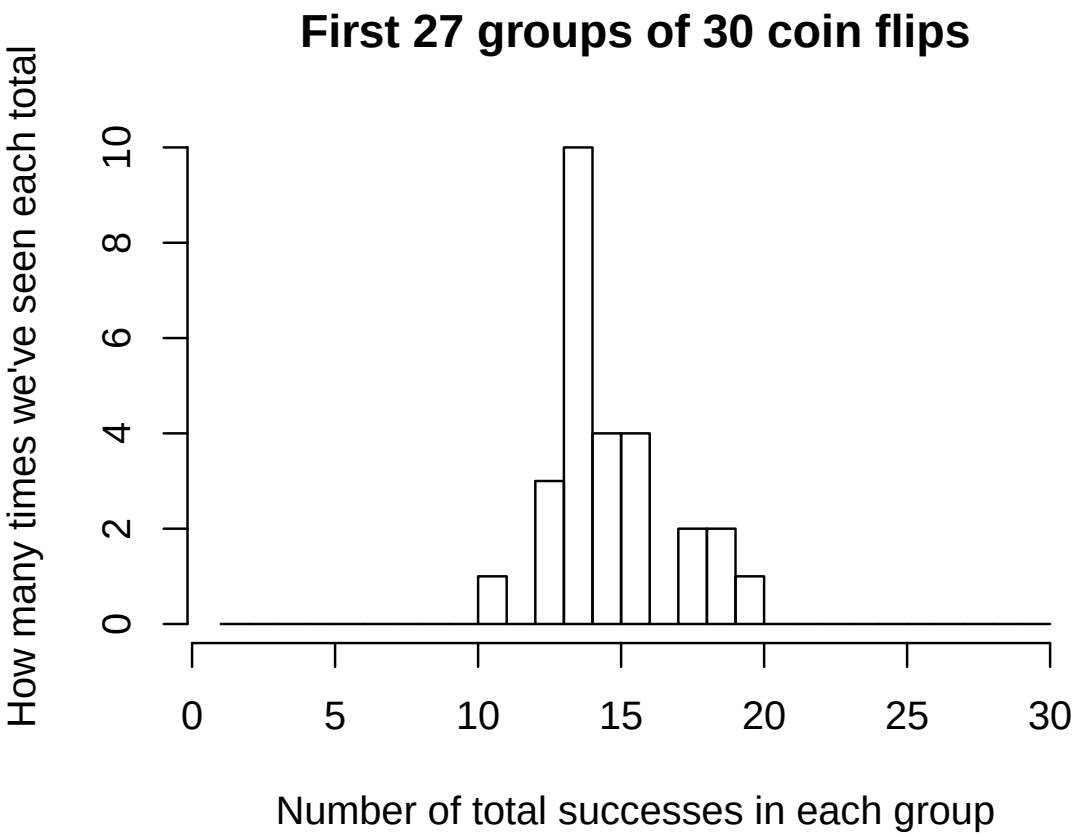
Each person in your section flips another coin, and we see 19 heads:

T T H T H T T H H T H
H T H T T H H T H H H
H T H H H H H H



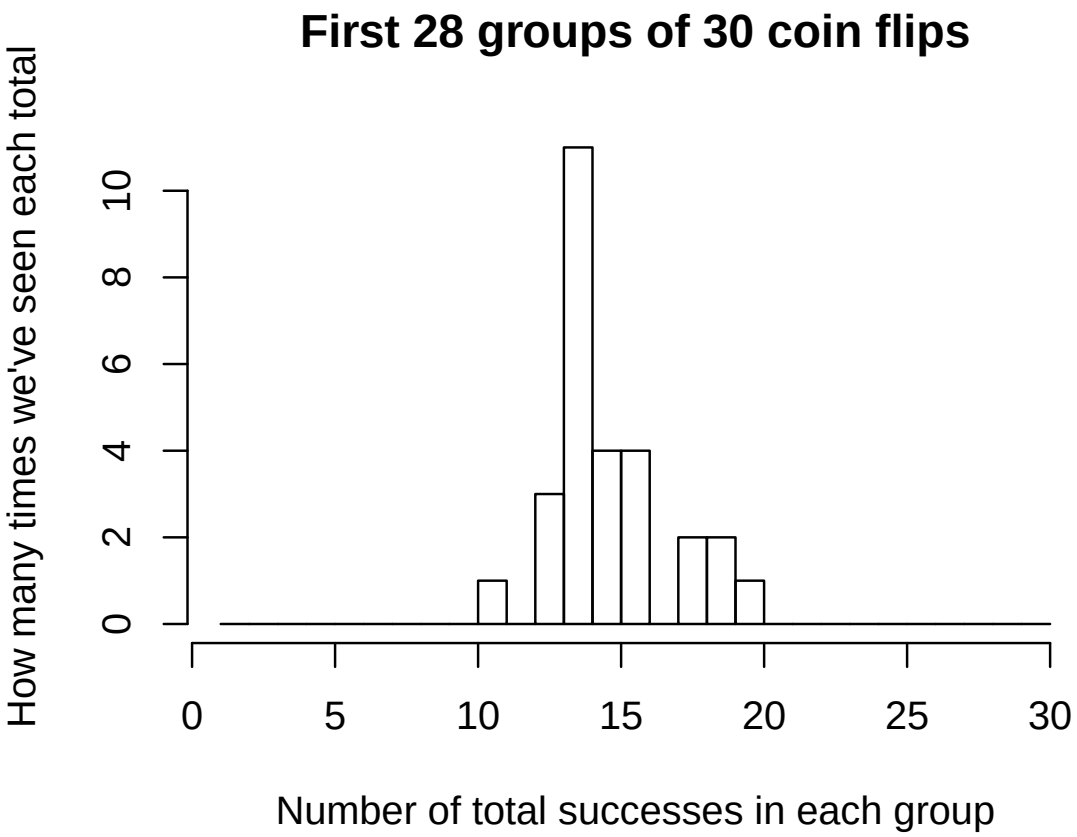
Each person in your section flips another coin, and we see 15 heads:

T H H H T H T H H H T
T H T H H H H T T T T
T T T T H H H T



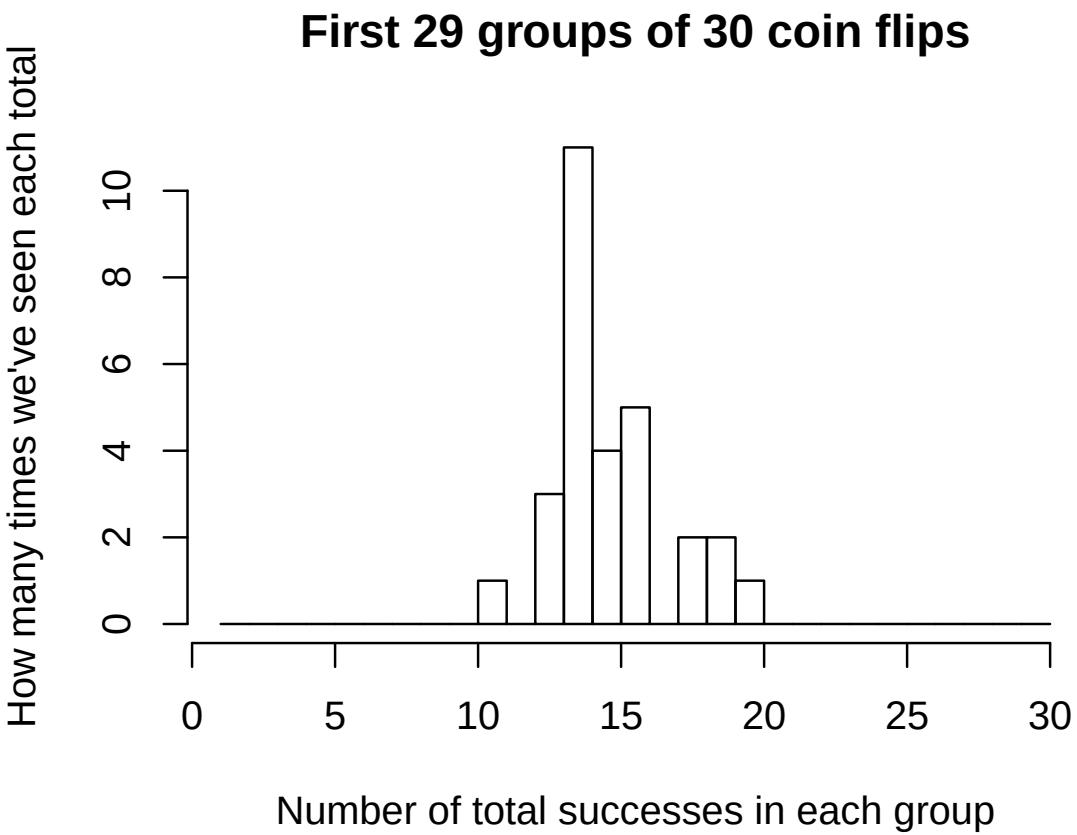
Each person in your section flips another coin, and we see 14 heads:

T T H H H T H H H T H
T T T H H T H T T H H
H T H T T T T T



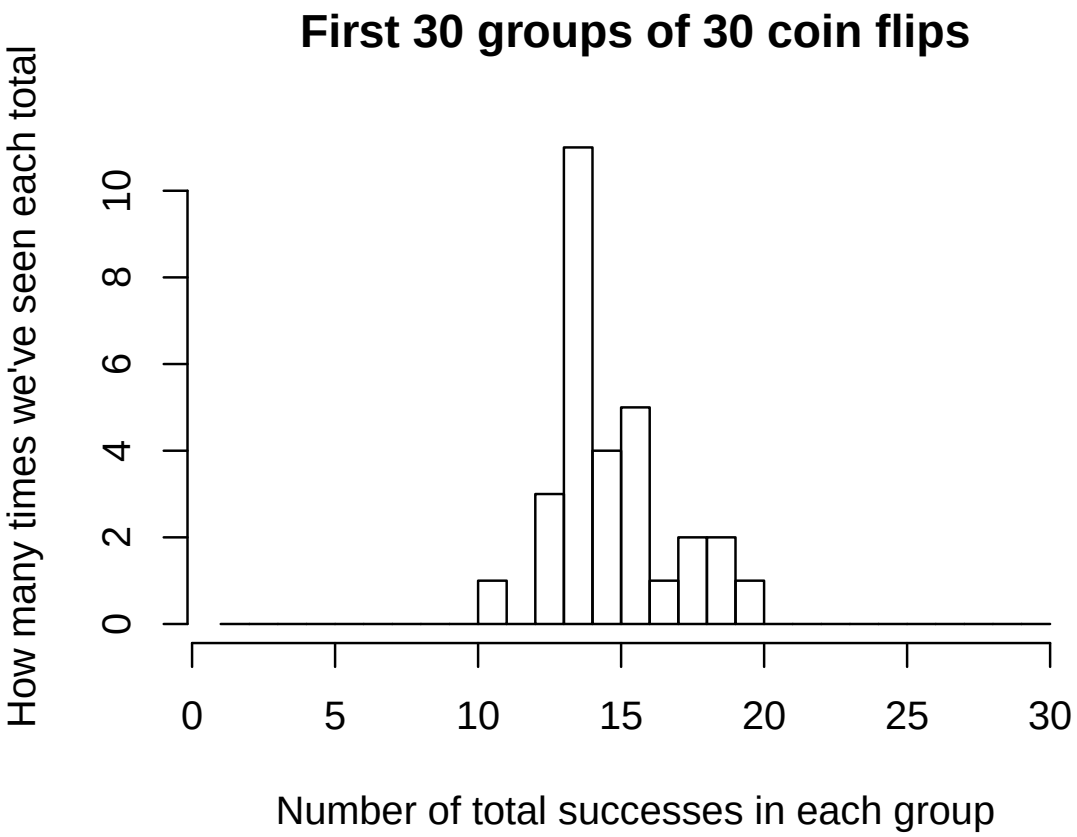
Each person in your section flips another coin, and we see 14 heads:

T H T T T H T T H T T
H H H T H H H T H T H
H T H T T T H T



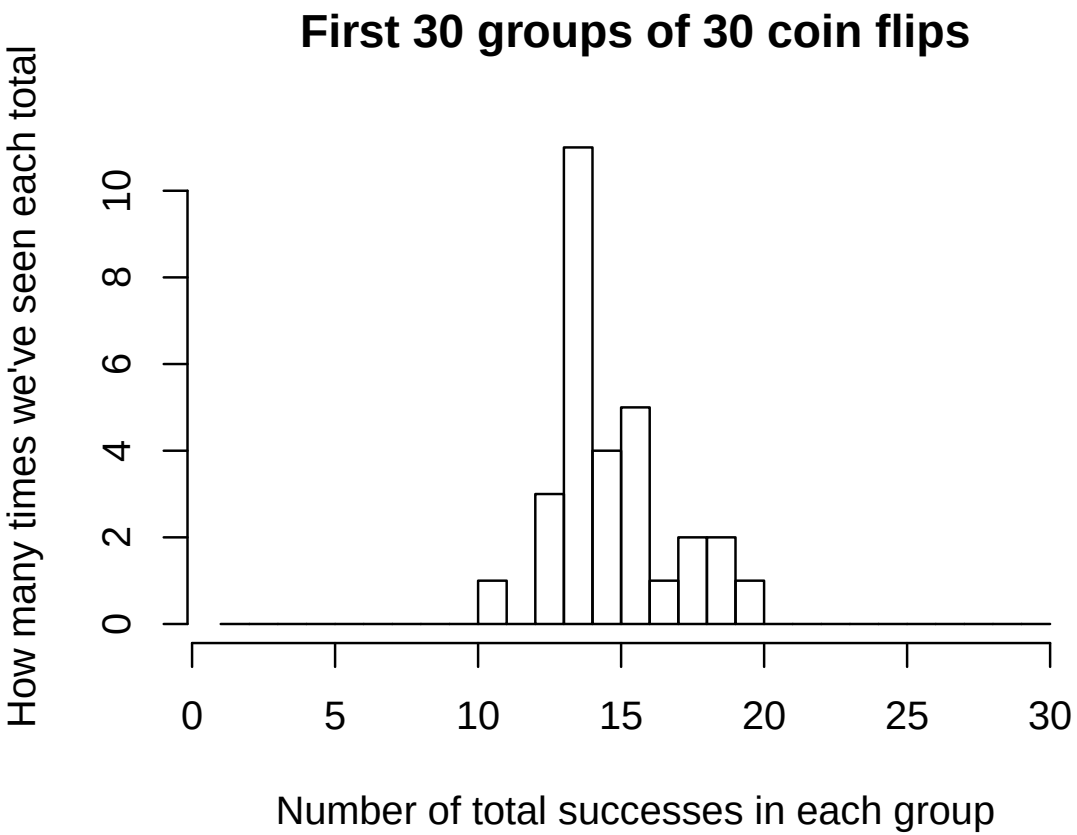
Each person in your section flips another coin, and we see 16 heads:

T H H T T T H H T H T
H T T H H T H T H H T
H H T H H H T T



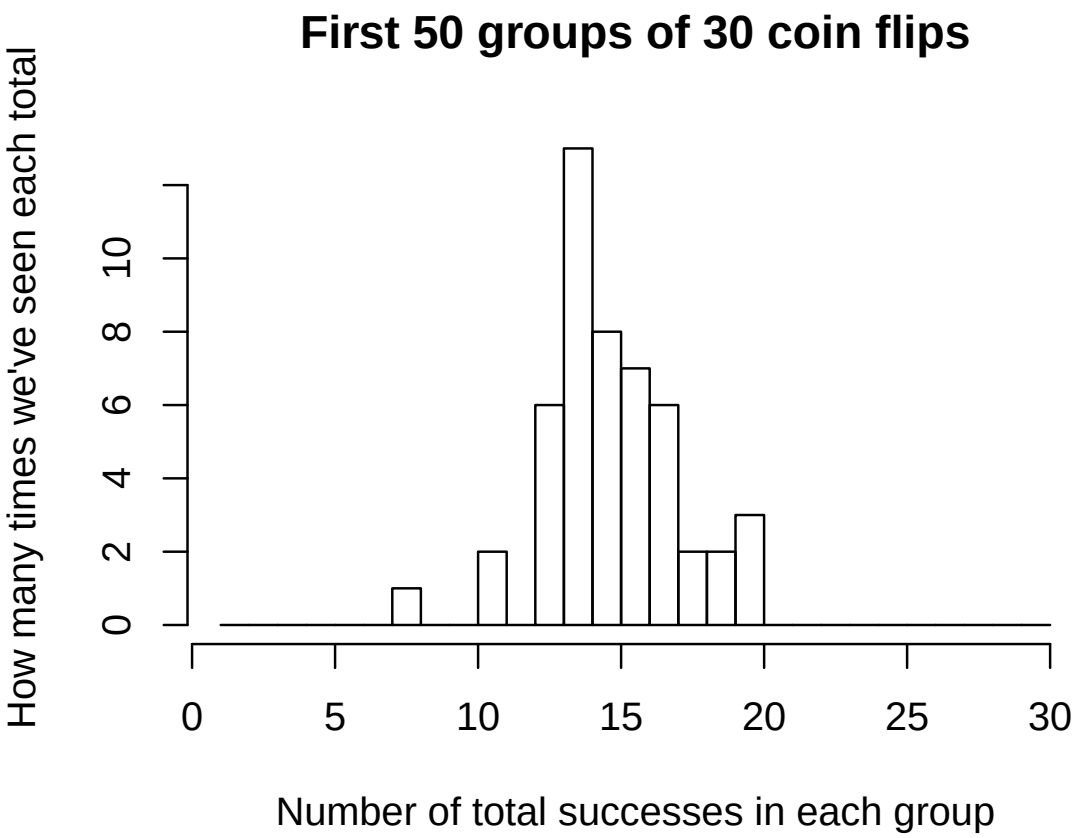
Each person in your section flips another coin, and we see 17 heads:

H T T H T H H T H T T
T H T H H T H H H T T
H H T H H H H T



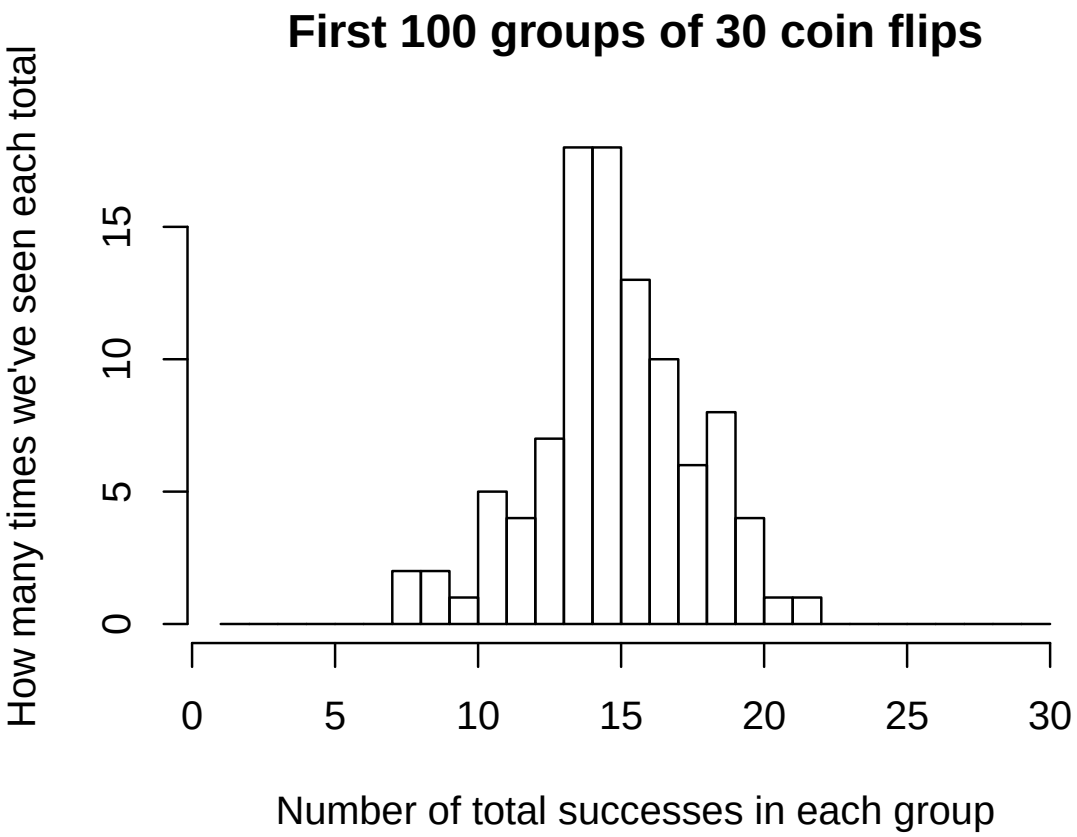
We've now spent the whole section flipping coins

A pattern is starting to emerge, but we need more samples



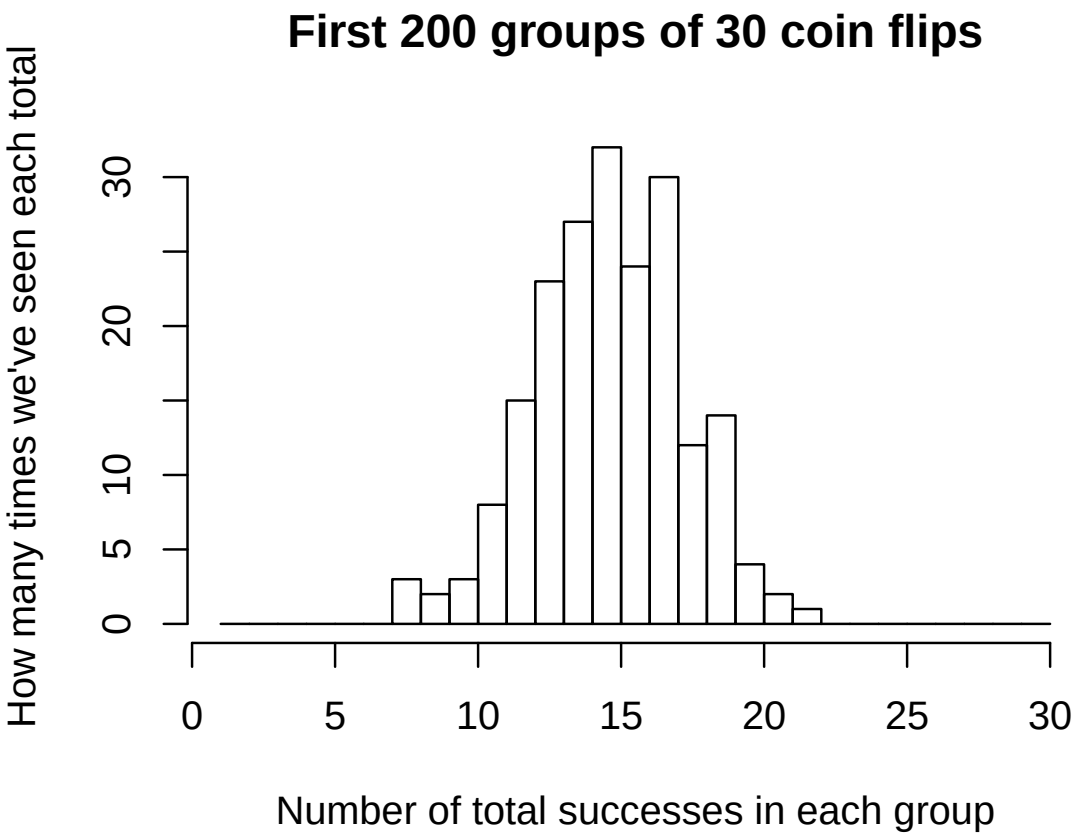
We could spend the next section flipping coins, with the following totals on each flip:

17 20 15 13 17 20 11
17 14 13 8 15 17 16 17
15 17 16 13 14 15



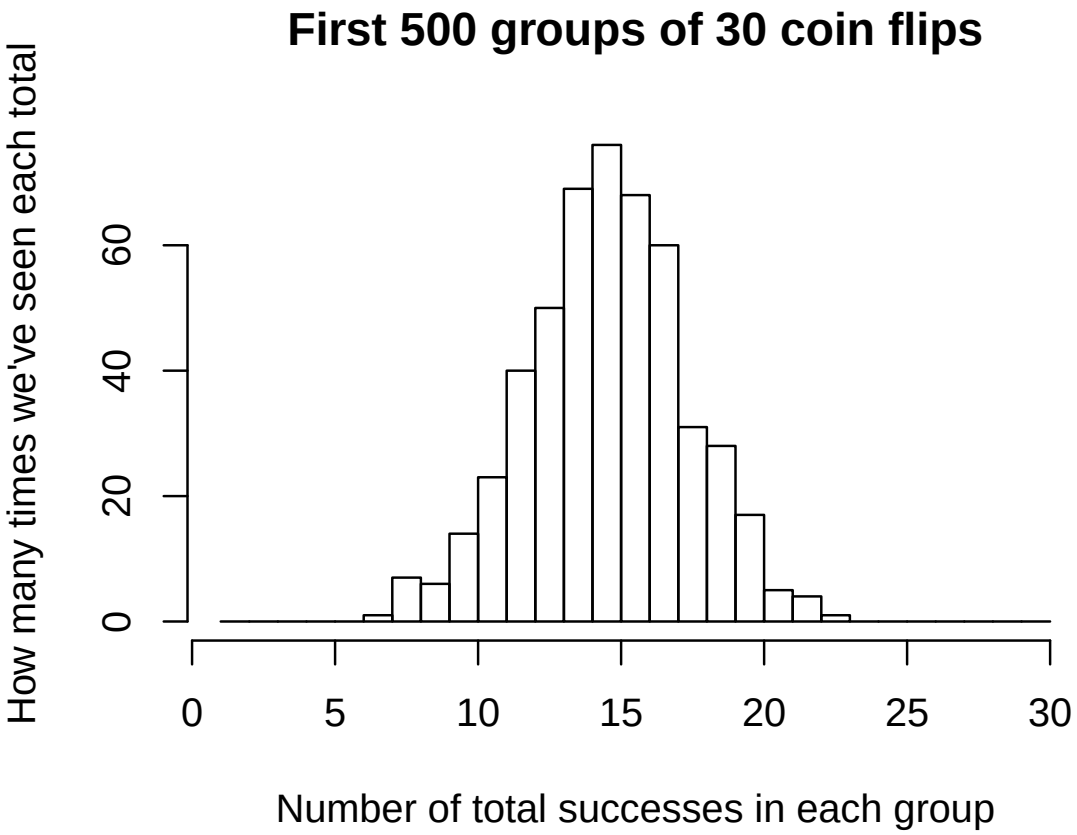
Or we could spend the whole week

15 9 19 15 11 16 15 16
17 15 11 19 19 18 12
15 16 16 14 14 10 9 14
11 15 17 17 8 15 19 12
20 21 14 15 16 18 19
14 16 19 15 13 17 22
15 12 18 12 15 18



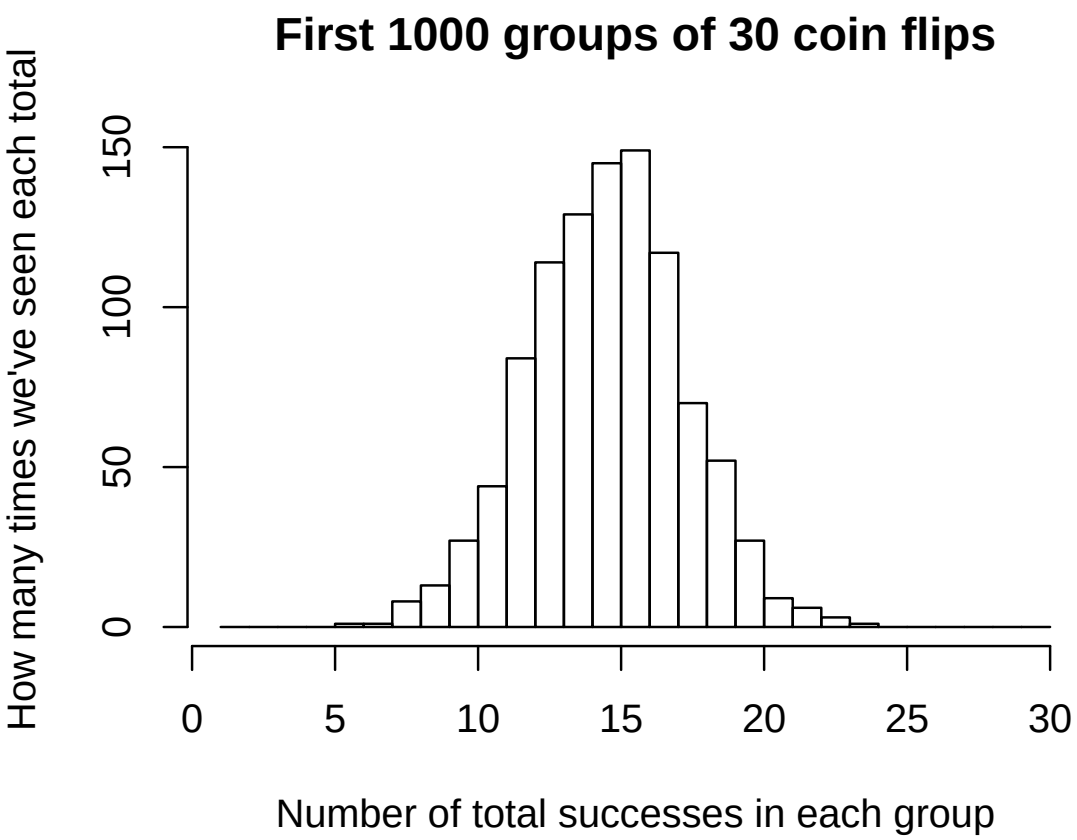
Two weeks?

18 11 15 17 17 11 21
18 15 15 12 17 12 16
15 12 17 13 14 13 17
13 19 12 16 11 17 12
14 13 19 18 18 15 13
14 13 19 12 17 8 14 13
16 13 13 16 15 15 12
17 17 17 16 15 14 14
19 15 13 17 17 17 16
17 14 13 17 12 18 13
14 18 17 15 15 13 15
17 16 19 16 13 19 10
16 10 15 16 17 13 18
12 13 17 14 15 12 17
12 16



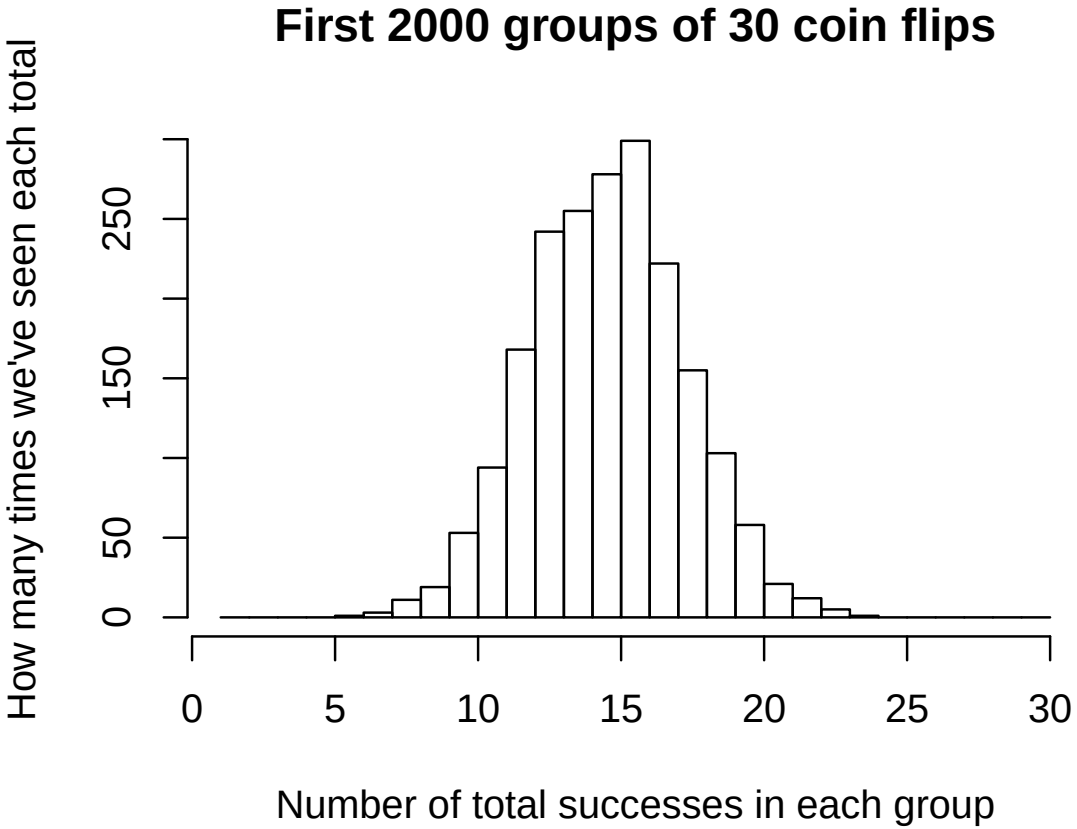
A month?

16 14 11 9 14 15 11 18
19 16 17 15 18 15 12
13 11 16 16 15 16 15
17 13 16 12 11 16 15
12 13 12 15 12 16 17
17 16 14 20 15 17 18
11 16 15 14 15 13 16
17 15 15 19 8 11 13 17
12 16 19 16 18 15 15
17 14 10 14 16 9 14 17
14 15 15 16 14 15 17
14 12 19 14 11 16 14
15 13 18 16 14 14 12
21 15 18 16 19 10 19
14 16 12 15 15 14 15
14 15 15 10 19 17 ...



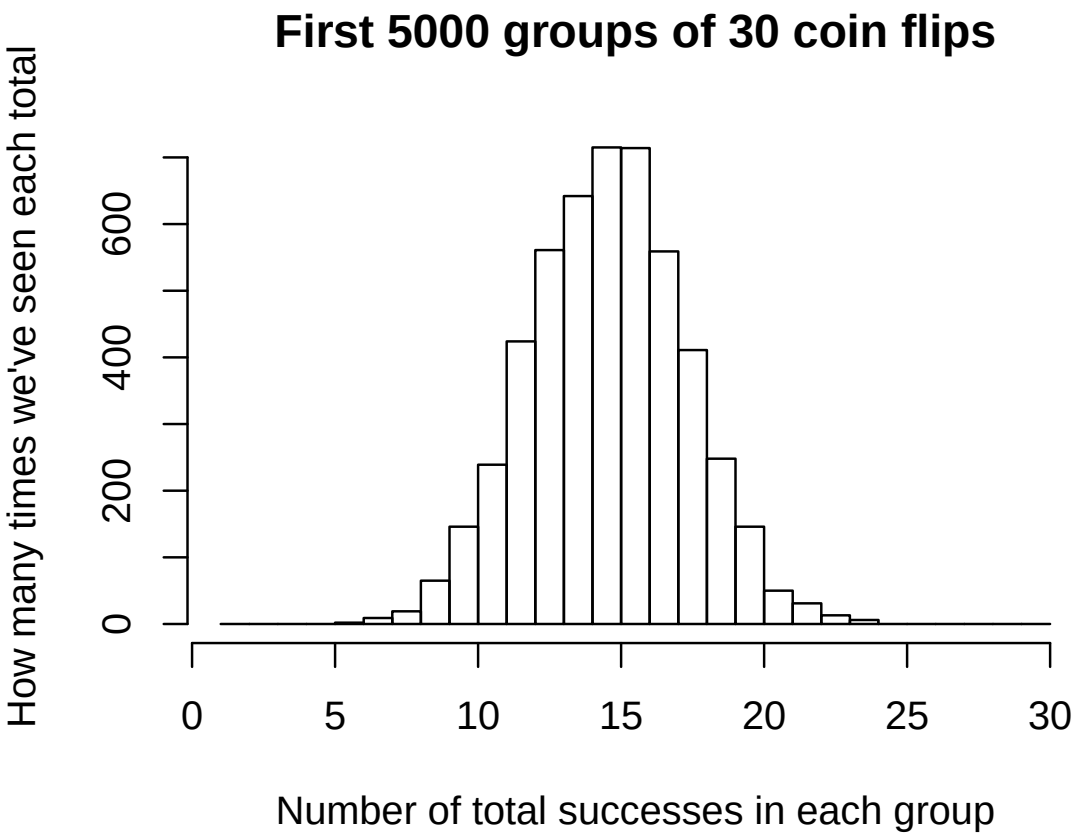
Obviously, we're not really flipping coins

22 19 14 19 18 16 15
16 19 12 13 15 15 13
19 15 19 20 15 18 11
14 8 14 14 13 15 12 18
17 9 11 16 16 13 14 16
18 16 17 10 13 17 19
17 14 15 14 13 11 14
15 16 16 17 13 14 18
20 10 14 15 16 10 13
23 16 12 17 16 21 18
17 19 13 19 16...

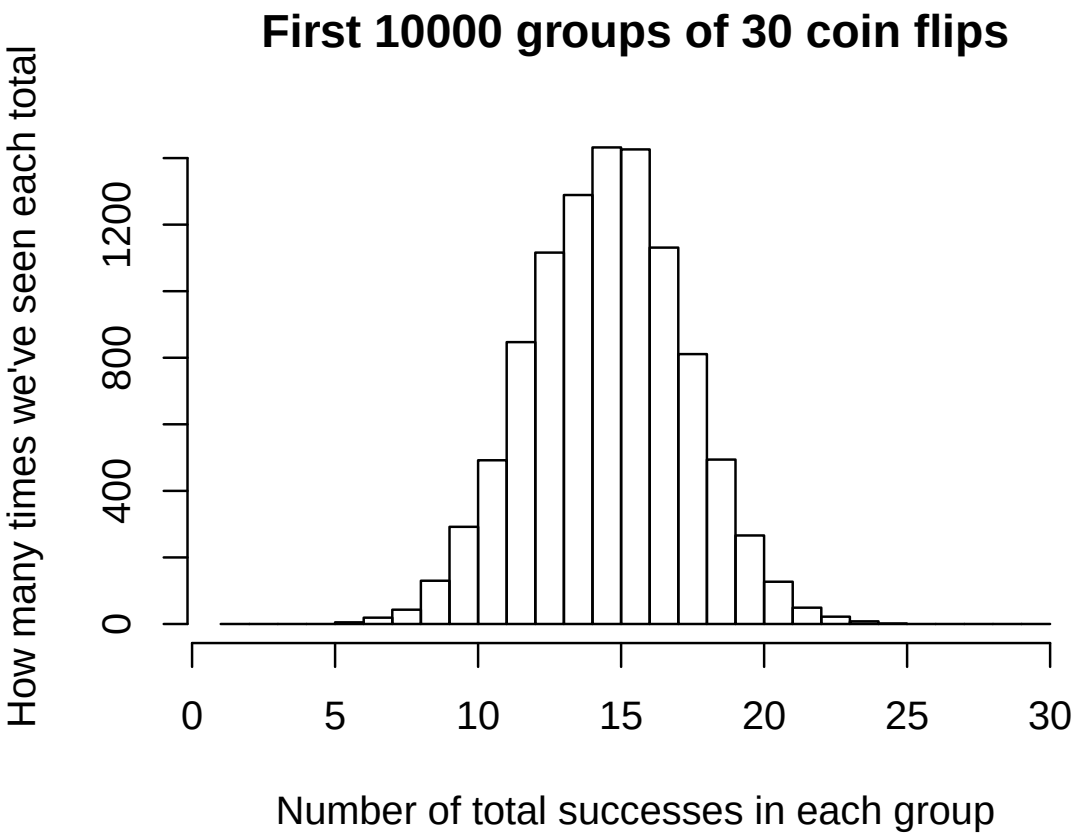


Instead, a computer program is “simulating” the process of flipping 180 coins, over and over

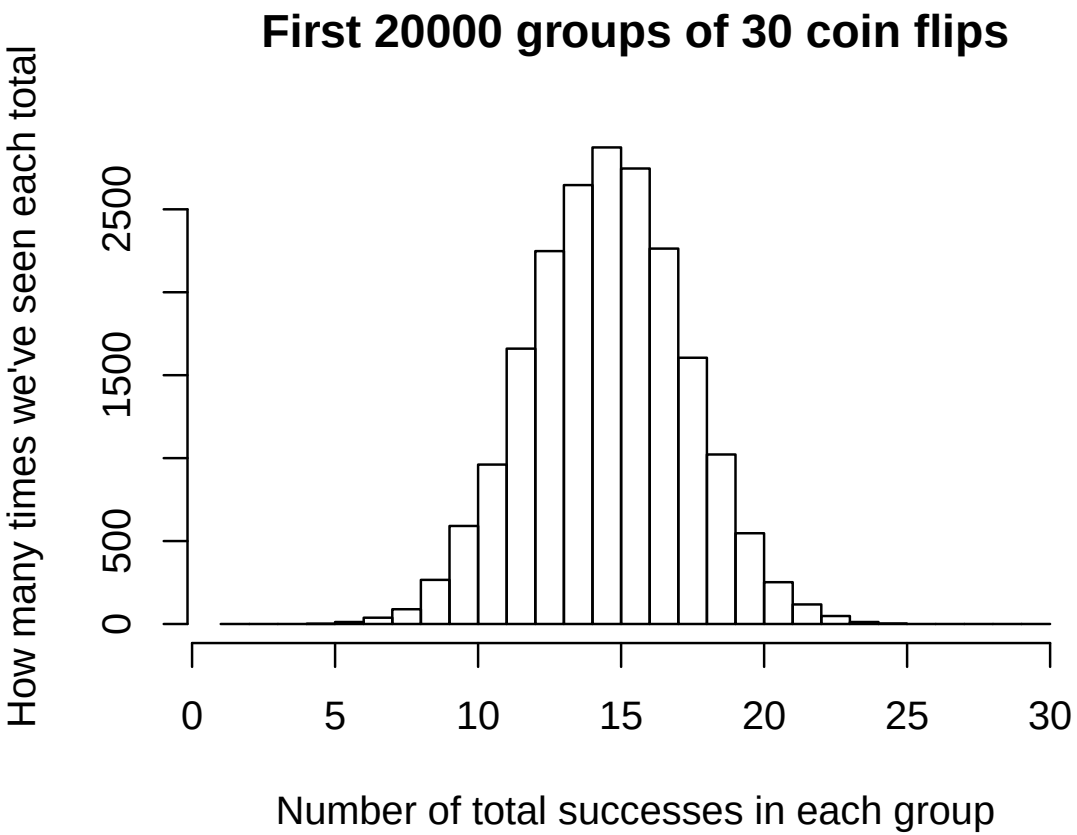
18 13 19 13 9 16 13 14
15 13 12 14 17 15 18
17 16 15 13 16 12 14
15 14 16 13 11 15 18
15 12 16 16 19 18 16
17 16 18 17 14 18 20
11 12 17 18 17 16 14
21 16 14 15 14 13 18
13 17 21 16 18 11 14
15 11 17 18 20 19 11
16 13 17 13 11 20 10...



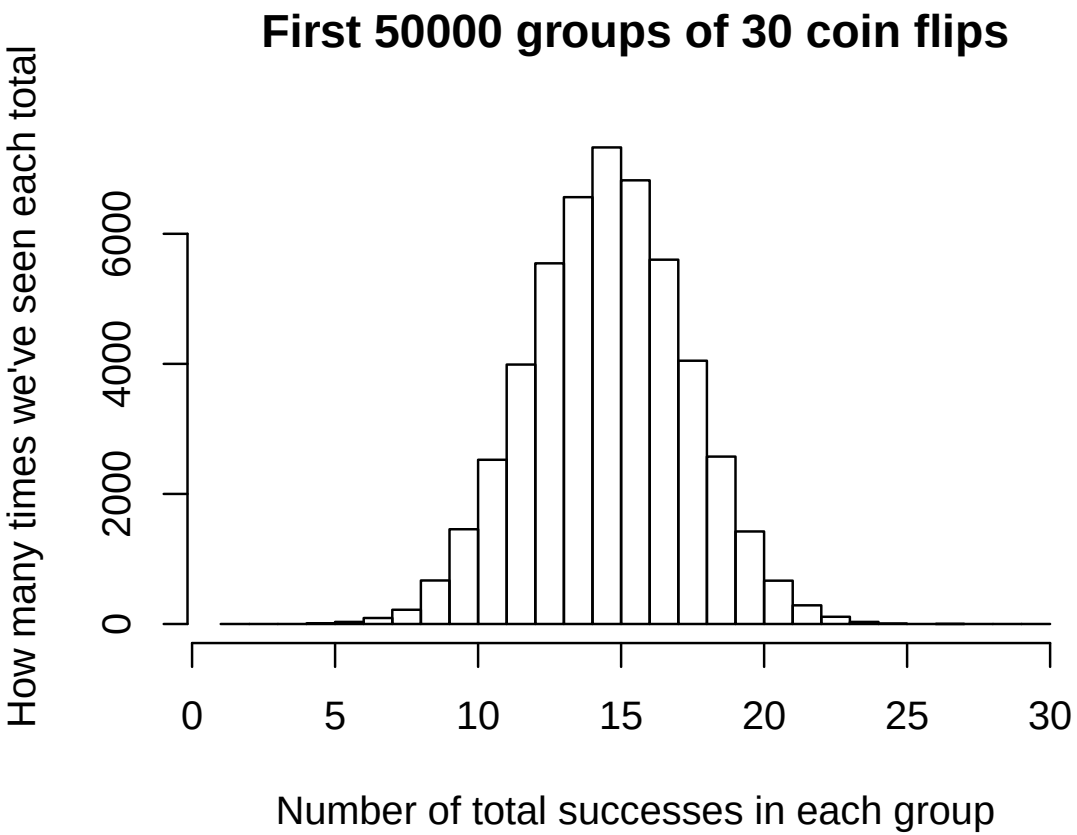
We're going up to 100,000 coin flips, to see if a pattern clarifies



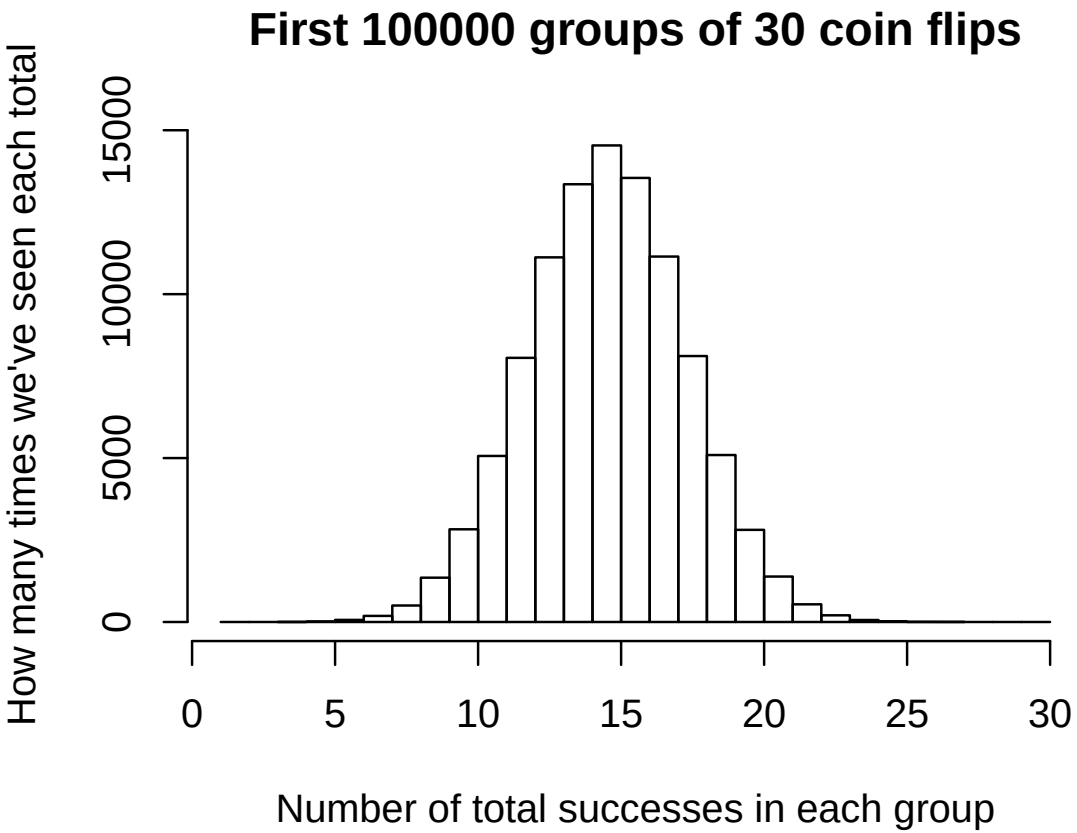
A computer can do this in less than 1 second



Notice a bell shape has emerged

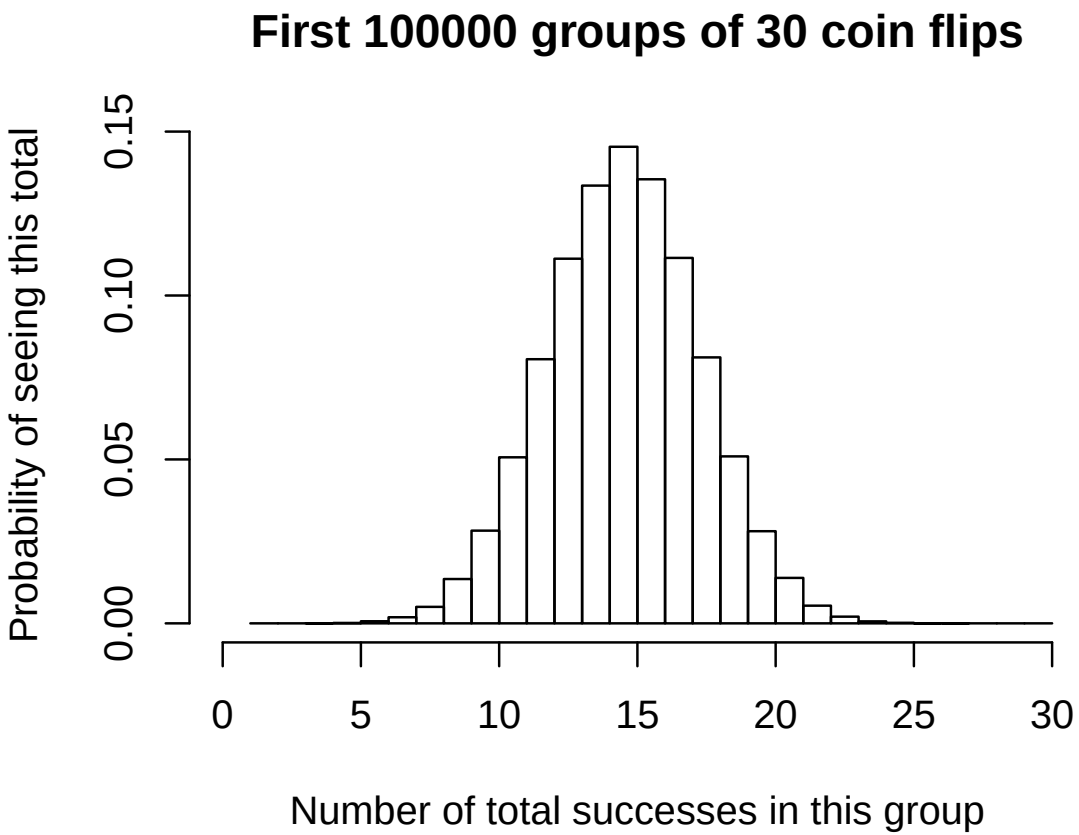


Or as close to a bell as we can get with only 30 possible outcomes



No matter how many more trials we add, the distribution of 30 coin flips will always converge to the pattern at the left

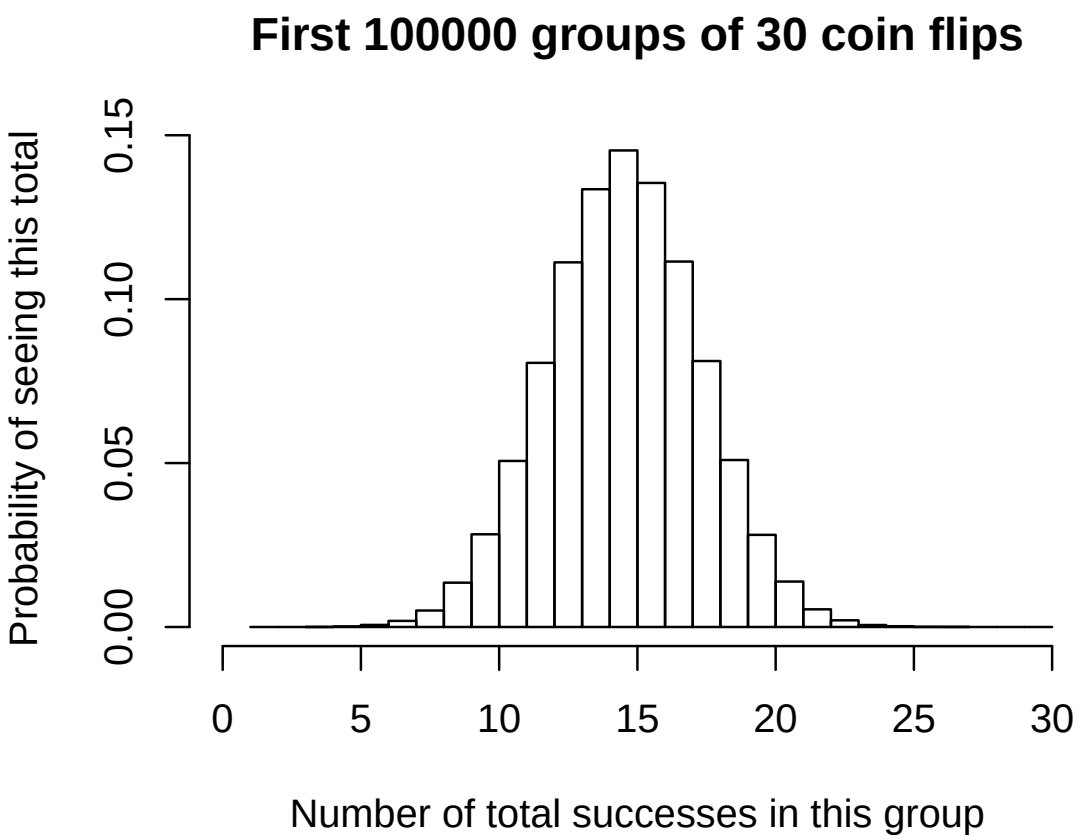
So what if we divide the frequency of each outcome by the total trials?



Now we have each total of heads as a proportion of the total trials

Based on this large sample, we can conclude these are the true probabilities of each sum from a random flip of 30 coins

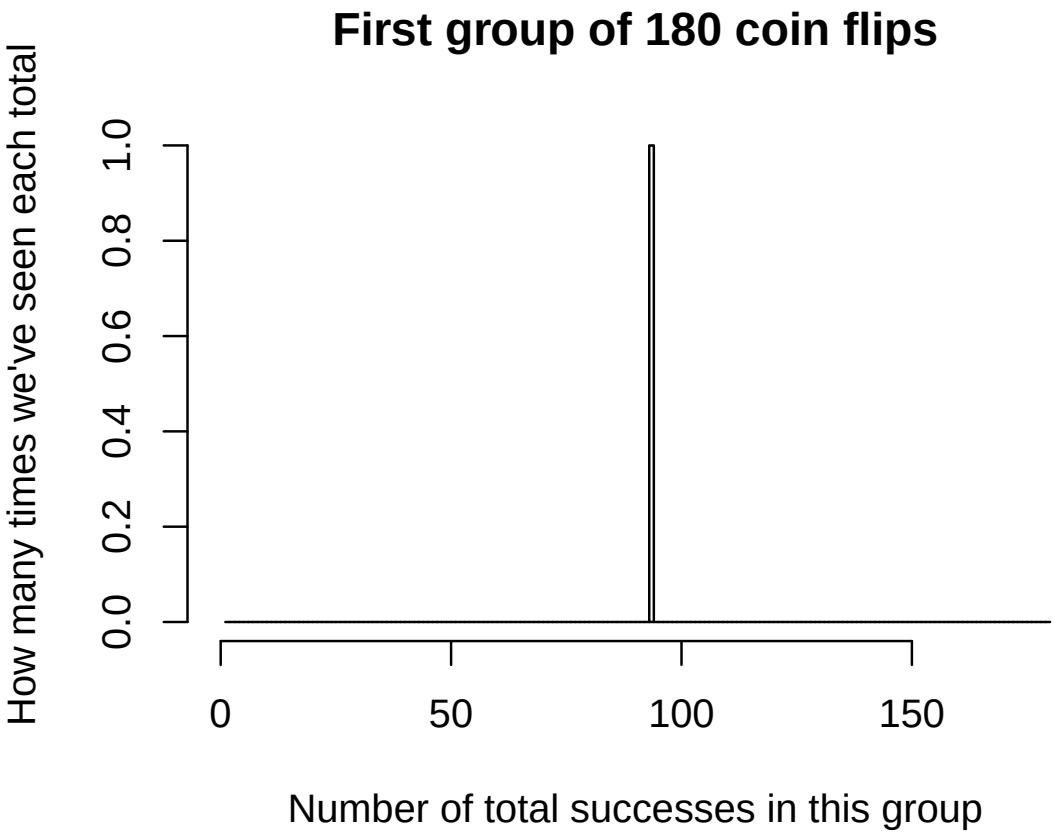
So the pattern at the left is the theoretical probability distribution of 30 coin flips



This probability distribution is known as a binomial.

It represents the probability of the sum of a finite number of binary variables

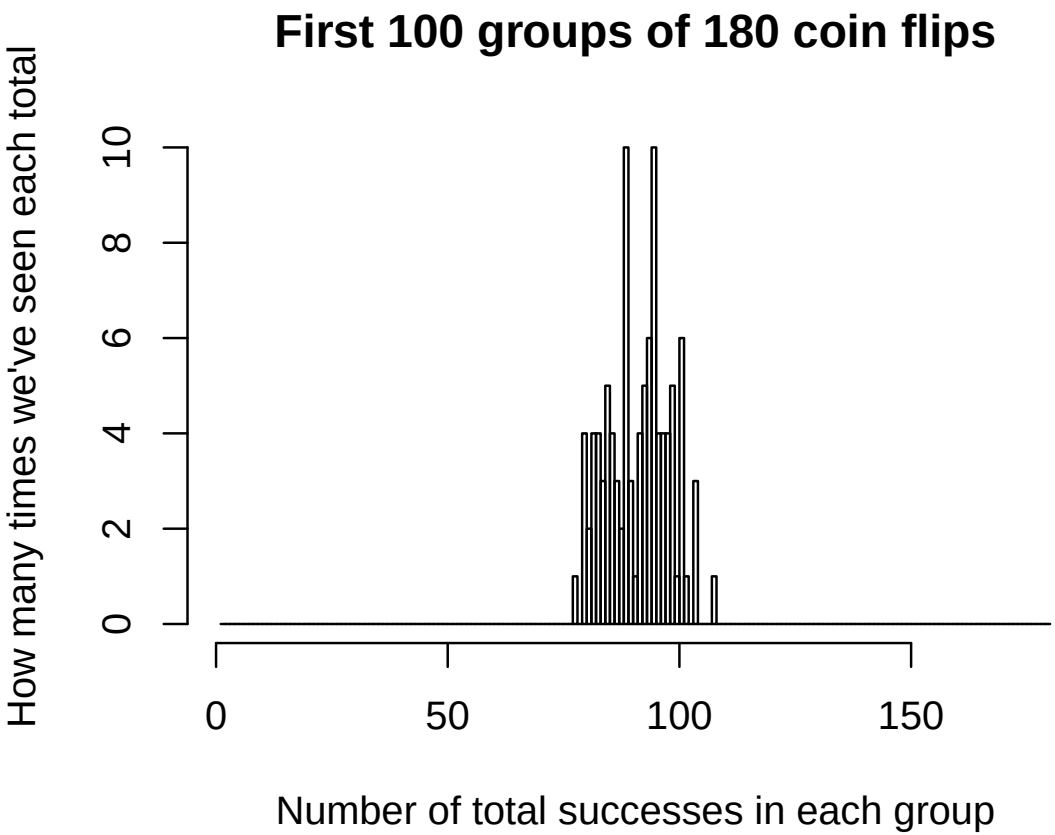
We won't go much further with the binomial, but it will help us understand another distribution



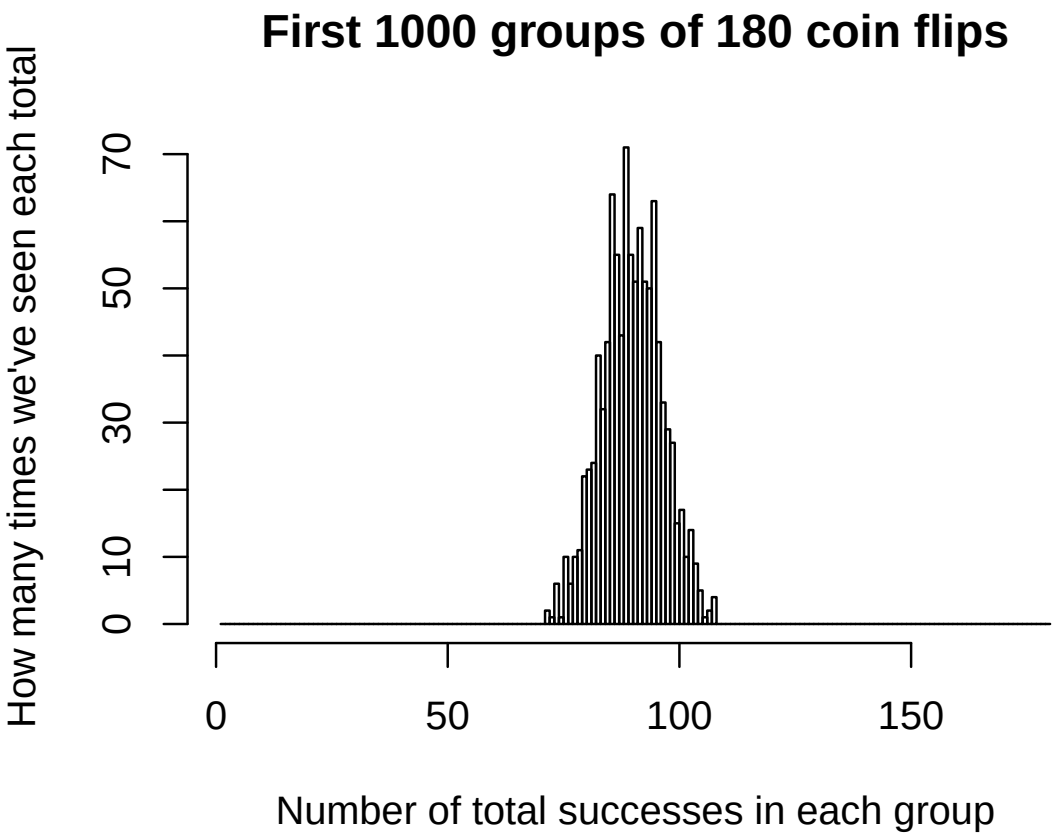
Suppose we ran our experiment in our lecture, so that there 180 coins to flip in each trial

On our first flip we might get 94 heads out of 180 trials

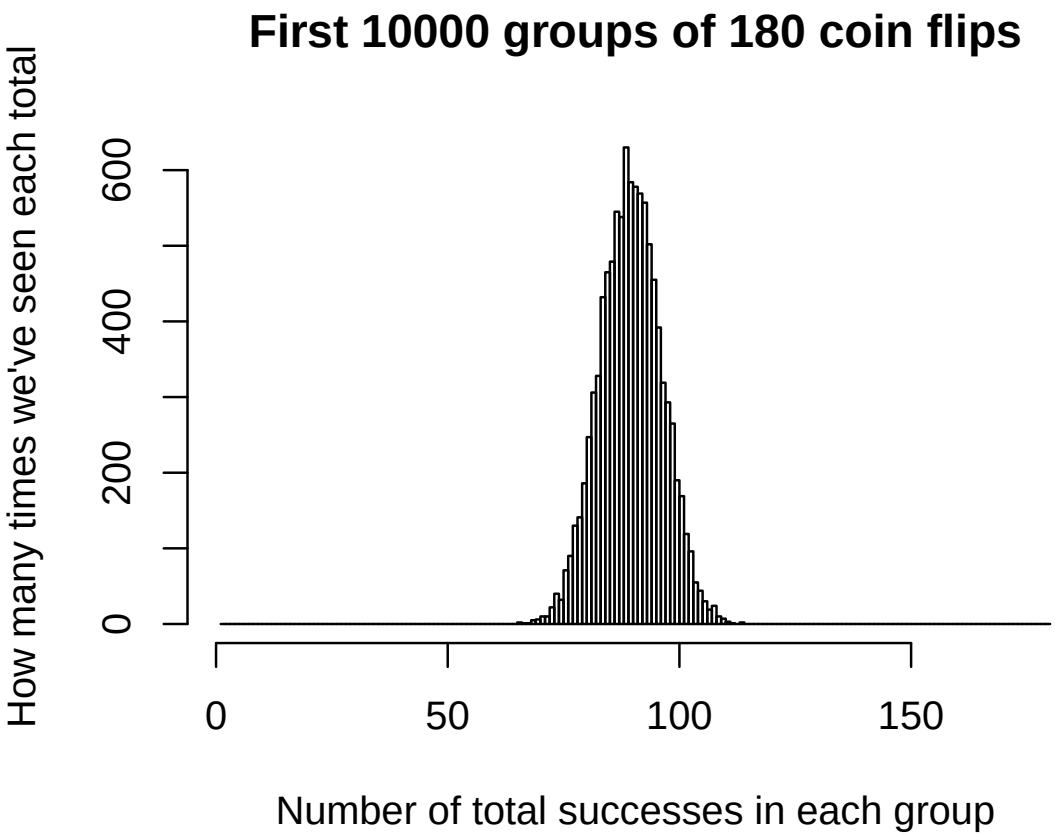
This time, let's skip ahead



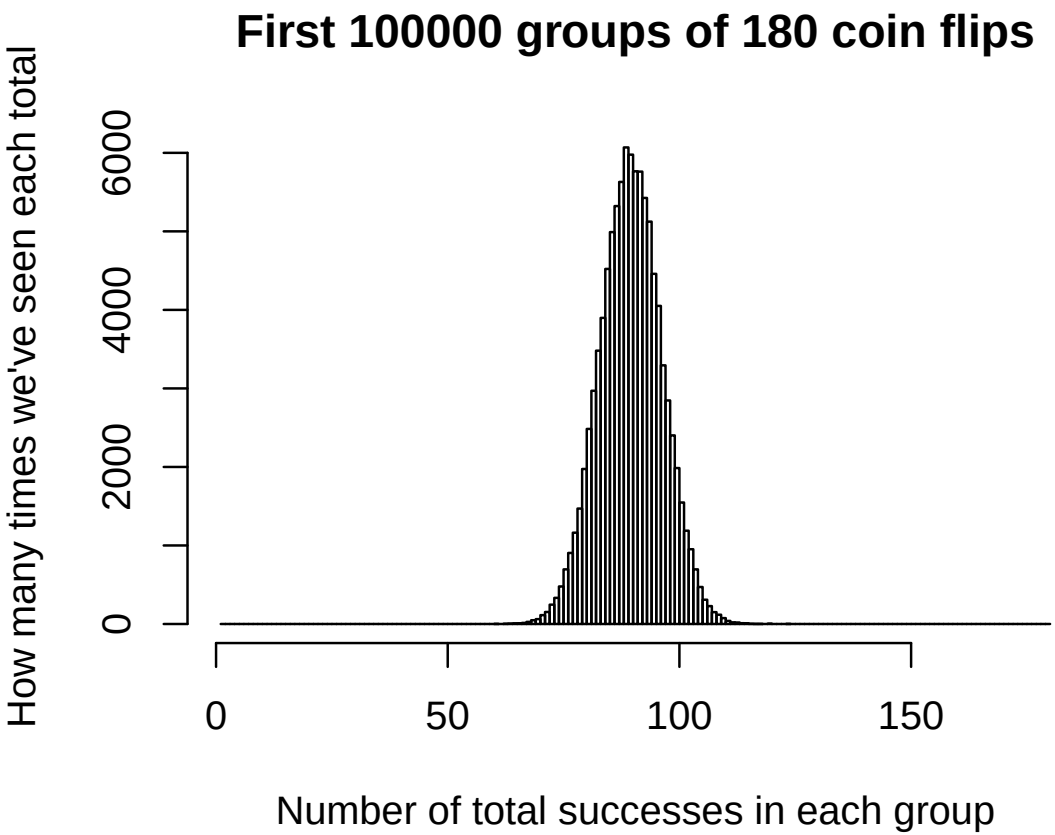
Here is our distribution after 100 flips...



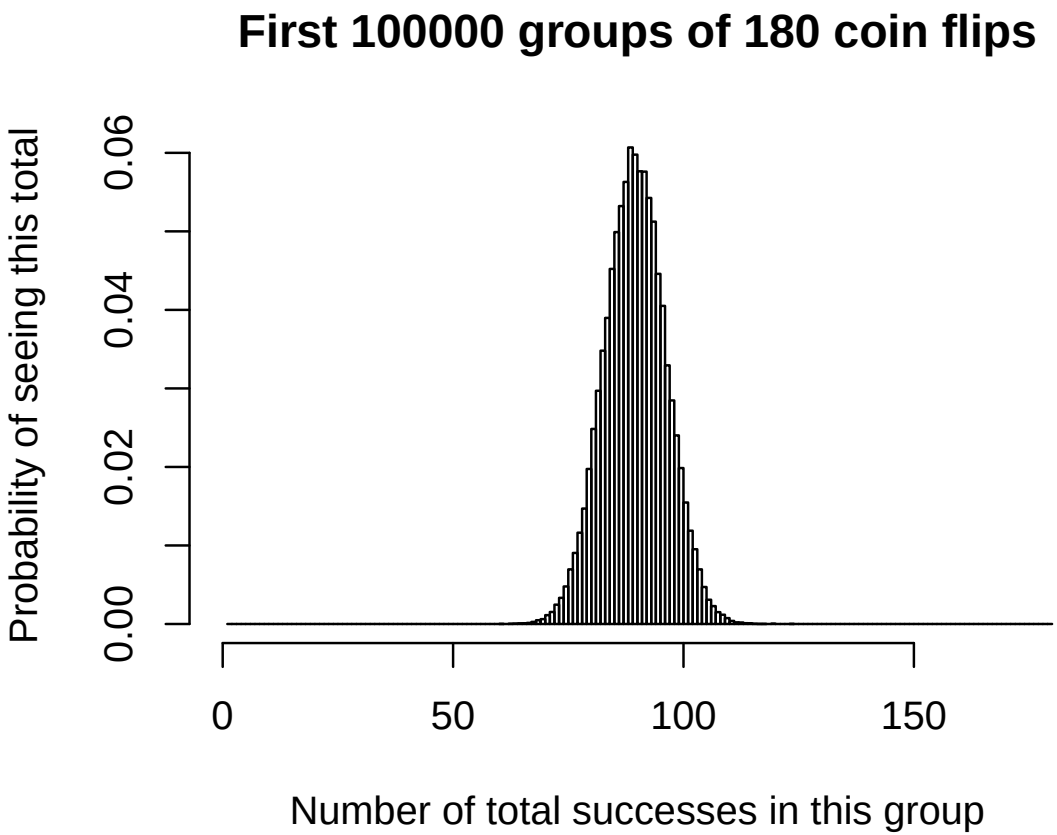
Here is our distribution after 1000 flips...



Here is our distribution after 10,000 flips...



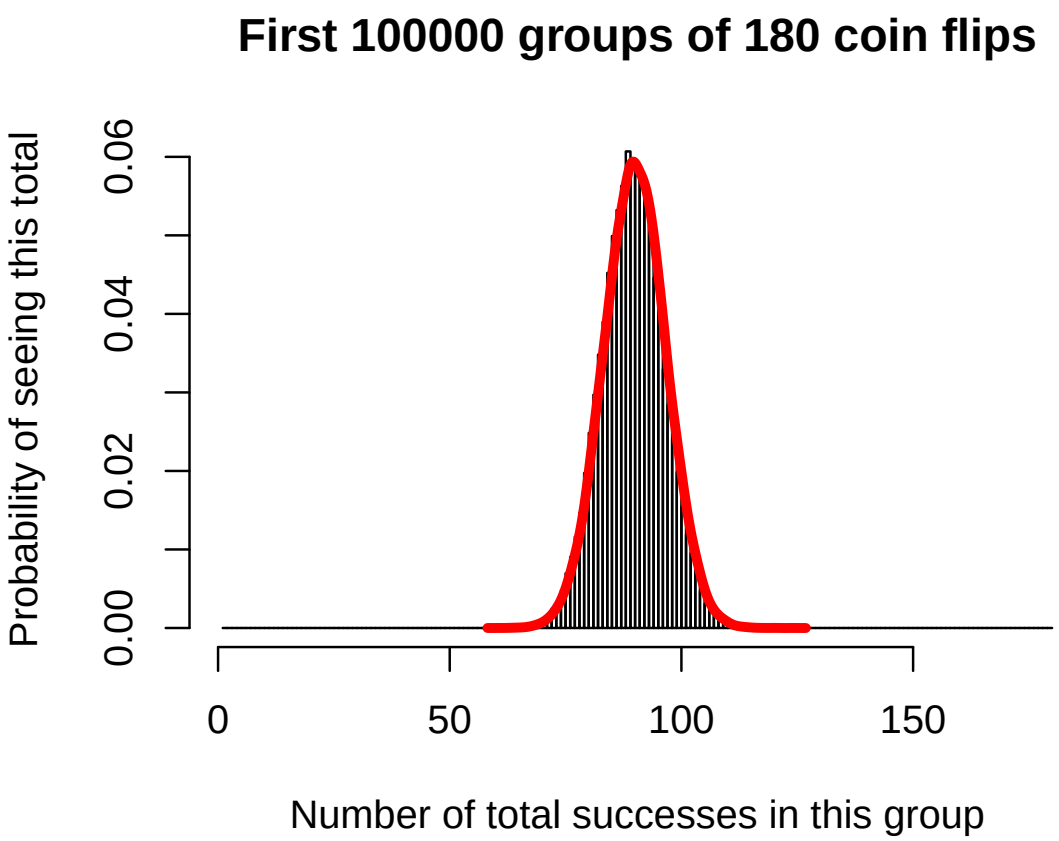
Here is our distribution after 100,000 flips...



Once again, we can divide by the total number of simulation runs to get probabilities

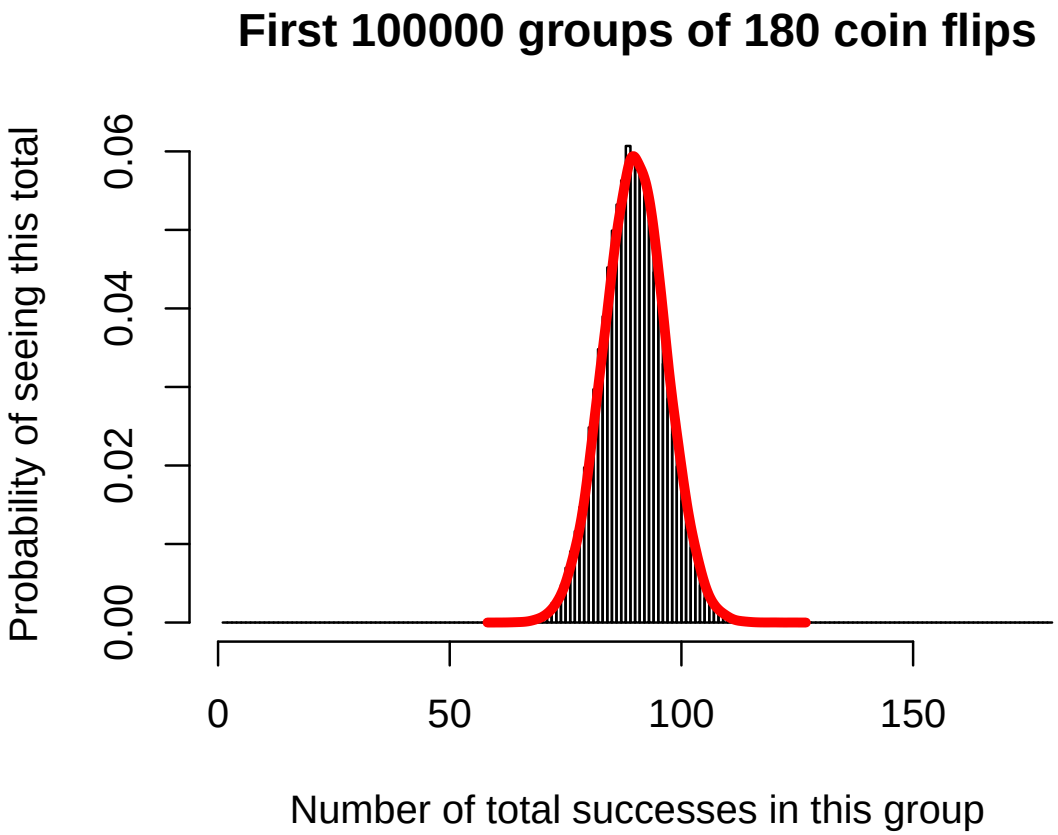
Notice that extreme values are vanishingly rare

And that the bell is smoothly traced out by the histogram's bars



If we trace out this smooth curve, we get the density in red

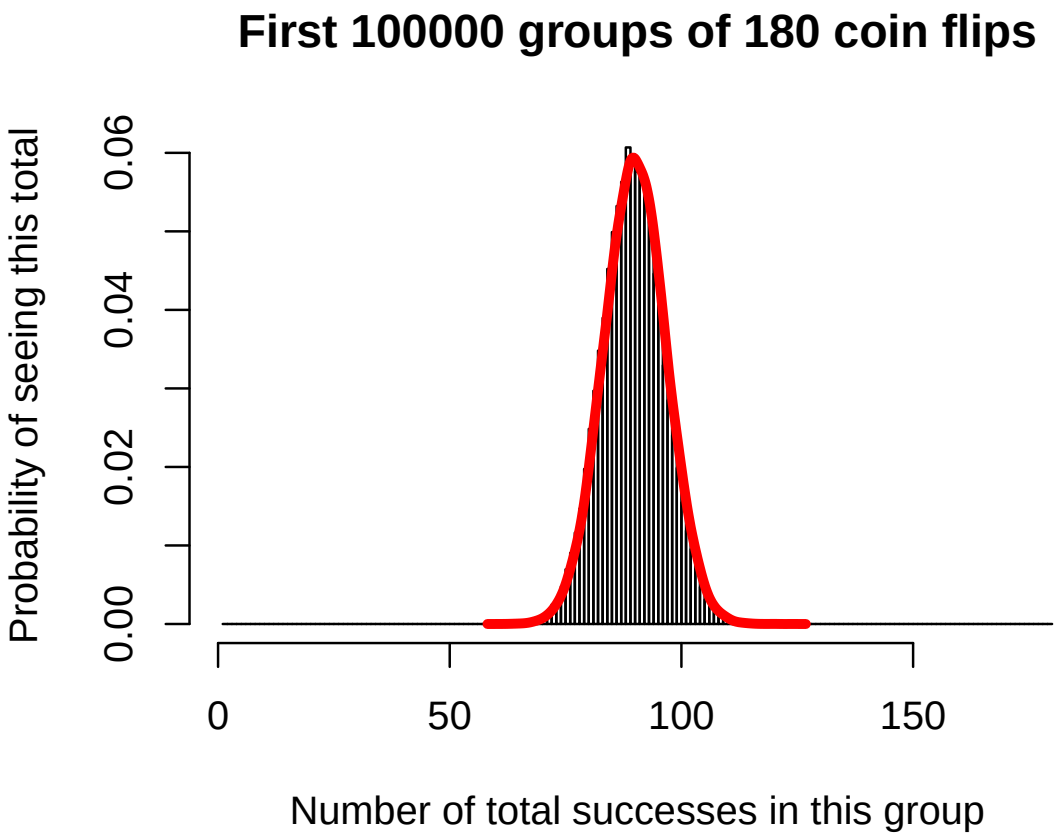
We've discovered something fundamental: if you add together many independent random variables, even as simple as coin flips, their sum approximates a bell curve



Central Limit Theorem

The more independent random variables we sum together, the more closely their sum approximates a Normal distribution.

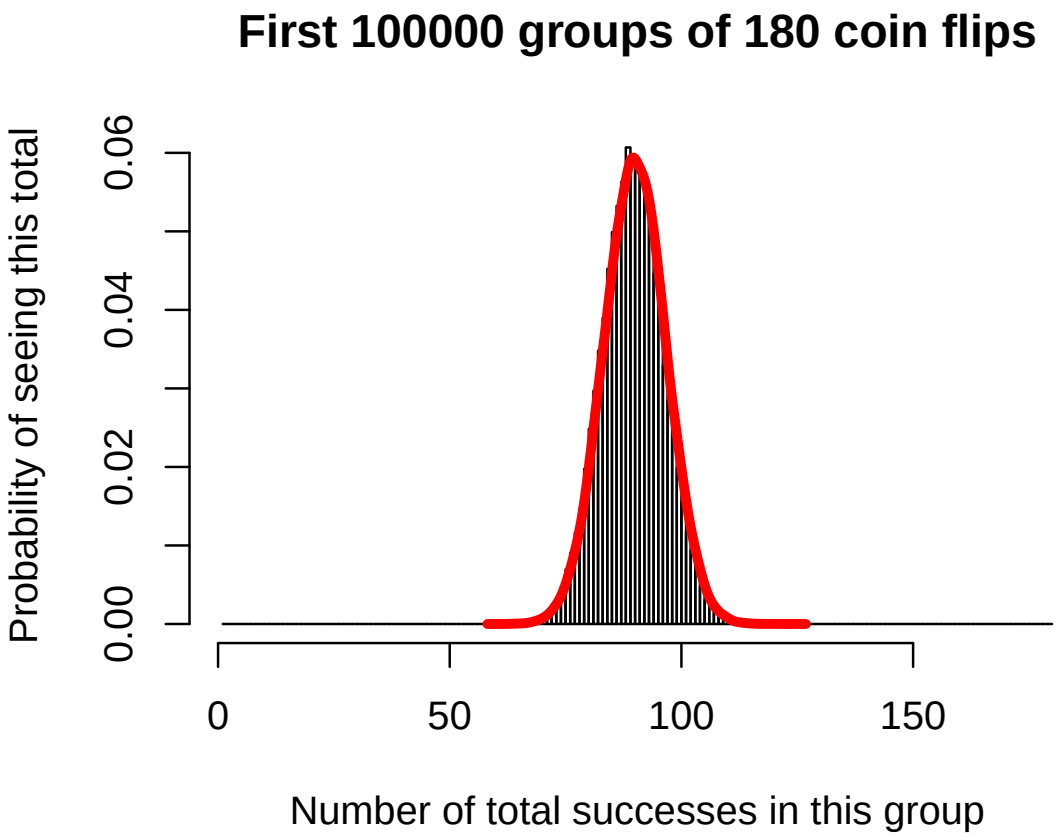
Example: if we ask everyone in America if they have a job, and add their responses together, we get the unemployment rate. The unemployment rate may be approximately Normally distributed



Central Limit Theorem

The more independent random variables we sum together, the more closely their sum approximates a Normal distribution.

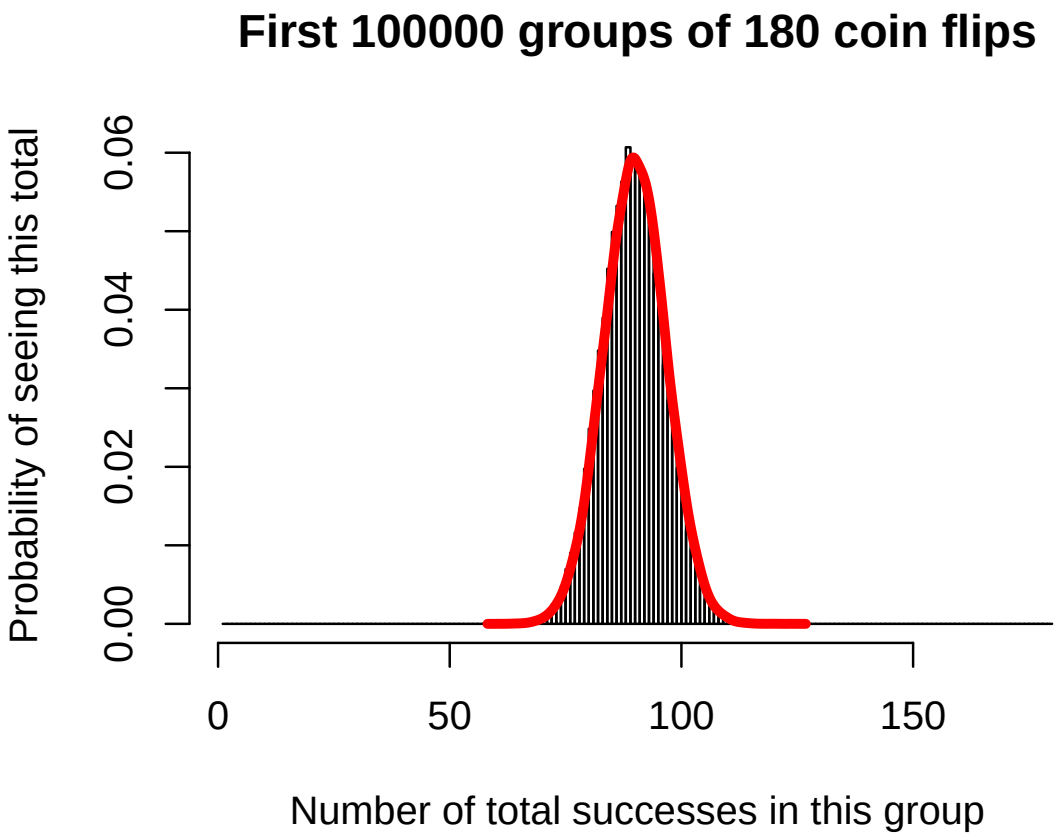
Notice that the Normal distribution only holds in this special case



Central Limit Theorem

The more independent random variables we sum together, the more closely their sum approximates a Normal distribution.

It is a distribution named Normal, not the "normal" distribution you see in the world



Central Limit Theorem

The more independent random variables we sum together, the more closely their sum approximates a Normal distribution.

It's also called the Gaussian distribution, and is just one possible distributions out of thousands known to statisticians

But it's very useful in intro statistics!

The Normal distribution

The Normal distribution describes the way *some* variables randomly vary

If a variable is Normally distributed, then its theoretical distribution can be summarized in just two numbers: its mean and standard deviation

In our coin flip example with 180 flips in each trial,
we observed a mean of 89.997 and a standard deviation of 6.68.

If the coin flips were Normally distributed,
we would say they followed a Normal(89.997, 6.68) distribution

The 68-95-99.7 Rule for Normal Random Variables

If a variable is Normal, then:

- 68% of observed cases will have values inside the mean ± 1 s.d.

  In our coin example, 68% of heads totals will lie in:

$$89.997 \pm 6.68 = [83.32, 96.68]$$

The 68-95-99.7 Rule for Normal Random Variables

If a variable is Normal, then:

- 68% of observed cases will have values inside the mean ± 1 s.d.

  In our coin example, 68% of heads totals will lie in:

$$89.997 \pm 6.68 = [83.32, 96.68]$$

- 95% of observed cases will have values inside the mean ± 2 s.d.

  In our coin example, 95% of heads totals will lie in:

$$89.997 \pm 2 \times 6.68 = [76.64, 103.36]$$

The 68-95-99.7 Rule for Normal Random Variables

If a variable is Normal, then:

- 68% of observed cases will have values inside the mean ± 1 s.d.

  In our coin example, 68% of heads totals will lie in:

$$89.997 \pm 6.68 = [83.32, 96.68]$$

- 95% of observed cases will have values inside the mean ± 2 s.d.

  In our coin example, 95% of heads totals will lie in:

$$89.997 \pm 2 \times 6.68 = [76.64, 103.36]$$

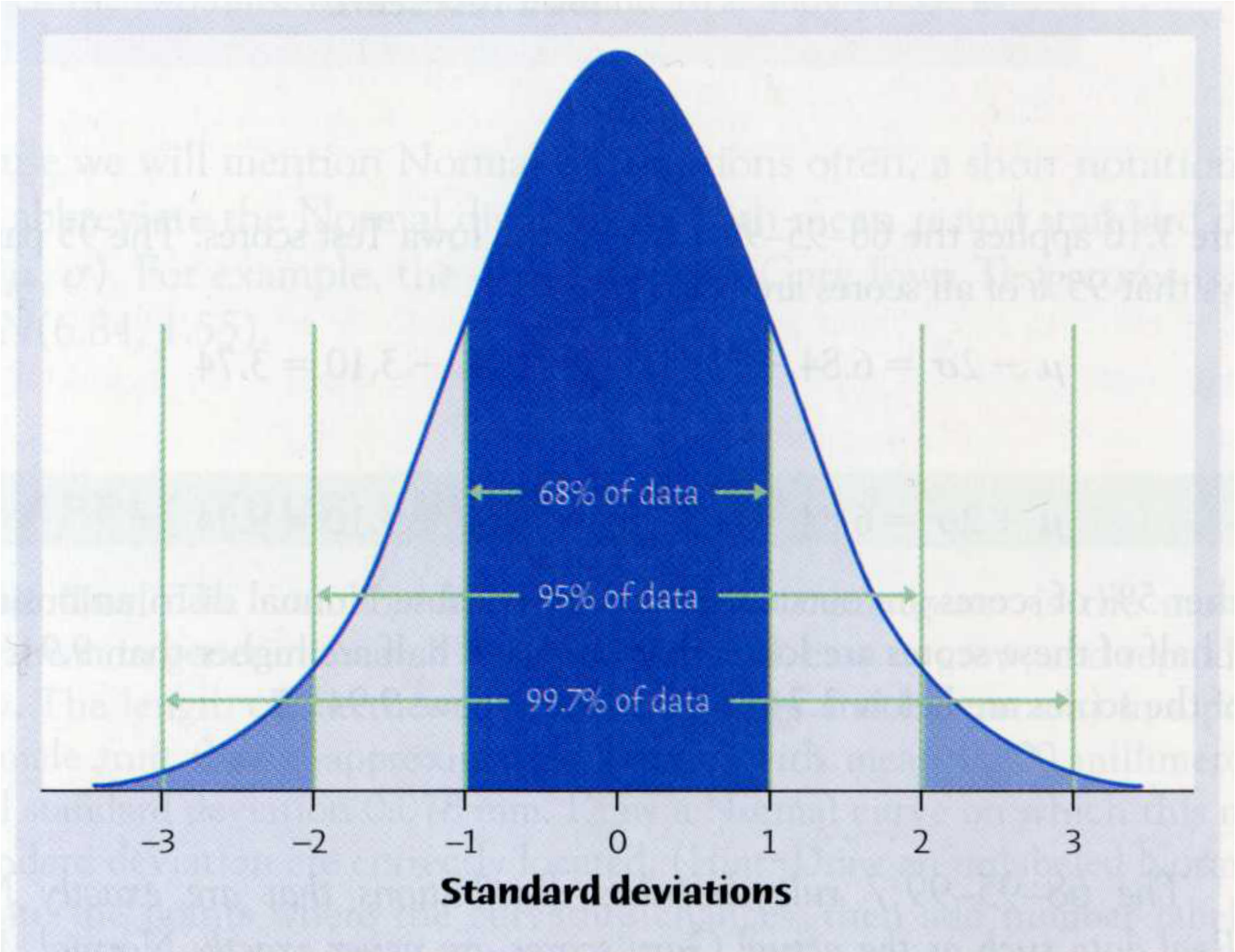
- 99.7% of observed cases will have values inside the mean ± 3 s.d.

  In our coin example, 99.7% of heads totals will lie in:

$$89.997 \pm 3 \times 6.68 = [69.96, 110.04]$$

Of course, our coin flips are only approximately Normal, and this holds only approximately here

The 68-95-99.7 Rule for Normal Random Variables



If a random variable follows the standard Normal distribution, a predictable amount of its values lie in each of these intervals

Continuous Random Variables

Technically, a sum of coin flips is a *discrete* variable (it only takes integer values)

But what about continuous random variables like:

1. Height of a child. We can measure height really precisely with the right equipment.
2. Time until the next bus. We can split a second finer and finer.
3. Exchange rate. How much the dollar is worth in euros.
4. Gross Domestic Product. Total value of all goods and services.

We can't even list all the outcomes of these variable, so we can't list the probability of each outcome.

In general, we won't see repitition of the same exact value ever. All frequencies are 0 or 1.

Example: Waiting for the train

Suppose your city has a subway that runs very regularly.
Every ten minutes there is a train.

Like most subway riders, you show up at the subway unaware of the scheduled time for the next train.

How long will your wait for the next train be in minutes?

Call your wait X:

X is continuous; you can chop it into tiny fractions of a second.

X is also rigidly bounded. It can't be less than 0, or more than 10.

Example: Waiting for the train

X lies somewhere between 0 and 10 minutes.

What is the probability that X is some particular value?

For exmaple, what is the probability that the train will arrive in exactly 3 minutes 25.00000000. . . seconds?

Example: Waiting for the train

X lies somewhere between 0 and 10 minutes.

What is the probability that X is some particular value?

For exmaple, what is the probability that the train will arrive in exactly 3 minutes 25.00000000. . . seconds?

Zero. That is,

$$\mathrm{P}(X = 3 : 25.0000000\ldots) \approx 0$$

Example: Waiting for the train

Why? There is an uncountable infinity of possible arrival times between 0 and 10 minutes.

If we split the total probability of train arrival ($= 1$) into an infinite number of pieces, each piece will be about 0.

In general, the probability that a continuous variable will take on an exact value is always 0.

(Note that we now refer to $\mathrm{P}(X)$ instead of $\Pr(X)$. We use a different notation for the probability of continuous variables.)

Example: Waiting for the train

We cannot talk about the probability of specific values of continuous distribution

Instead, focus on the probability that X lies in a specific interval.

For example, what is the probability that the train will arrive at or after 1 minute has passed, but before 5 minutes?

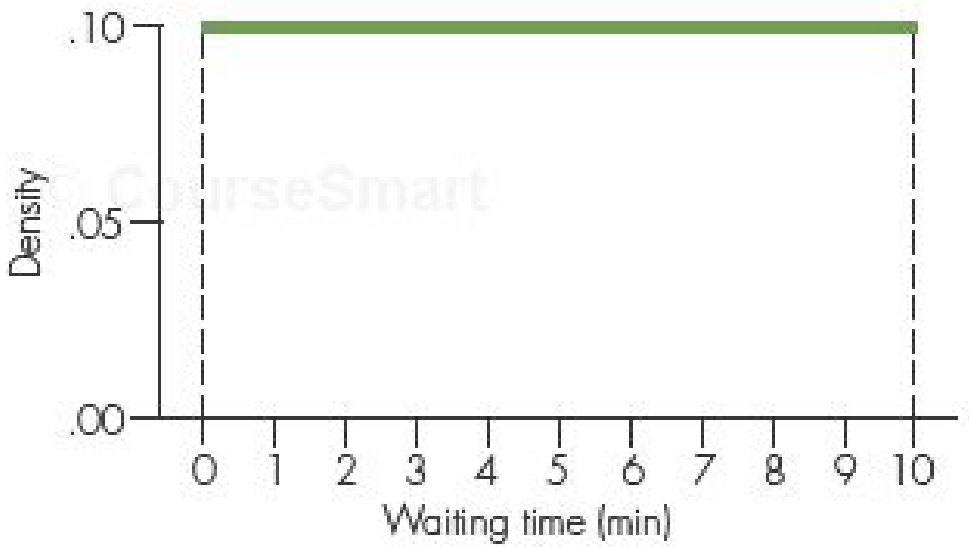
$$\mathrm{P}(1 \leq X < 5) > 0$$

Probabilities over intervals of continuous variables are positive, so we can calculate this. But we need to think a bit about the shape of the distribution

The Uniform Distribution

In the train example, there is no reason to consider the train more likely to arrive at any particular moment.

This is a rare case where all of the possible outcomes of a continuous variable are equally, or Uniformly, likely:



The Uniform Distribution

In our example, the probability the train will arrive between minute 1 and minute 5 is

$$\mathrm{P}(1 \leq X < 5) = \frac{5-1}{10-0} = 0.4$$

Because the uniform distribution is a rectangle, it's easy to calculate the areas that correspond to probabilities for an interval of X

Calculating probabilities for continuous distributions

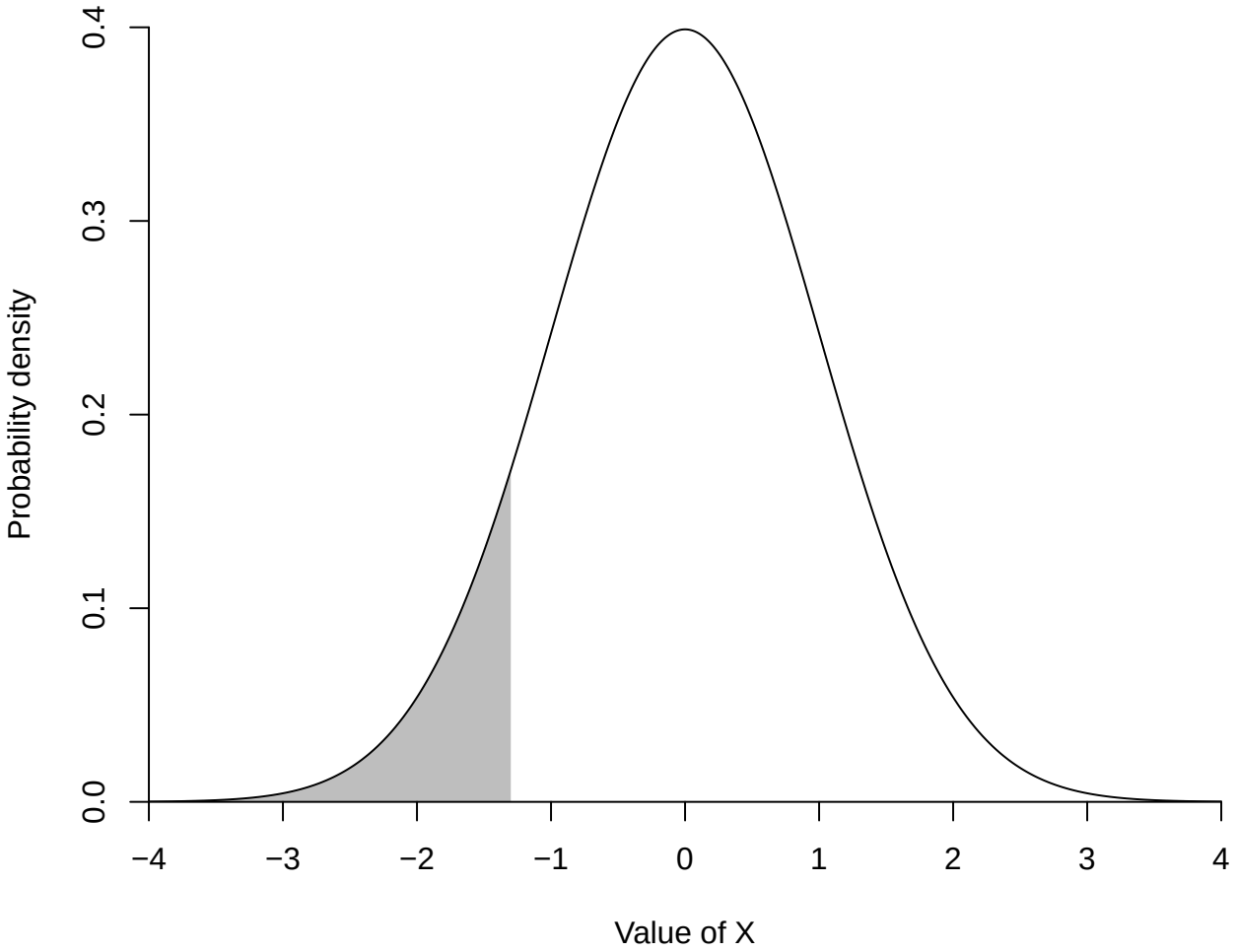
But most continuous distributions follow complex curves

We will need a computer to compute areas under these curves,
or a table to look them up

But to use a table to look up the area under the curve,
we still need to think about what quantity we want to calculate

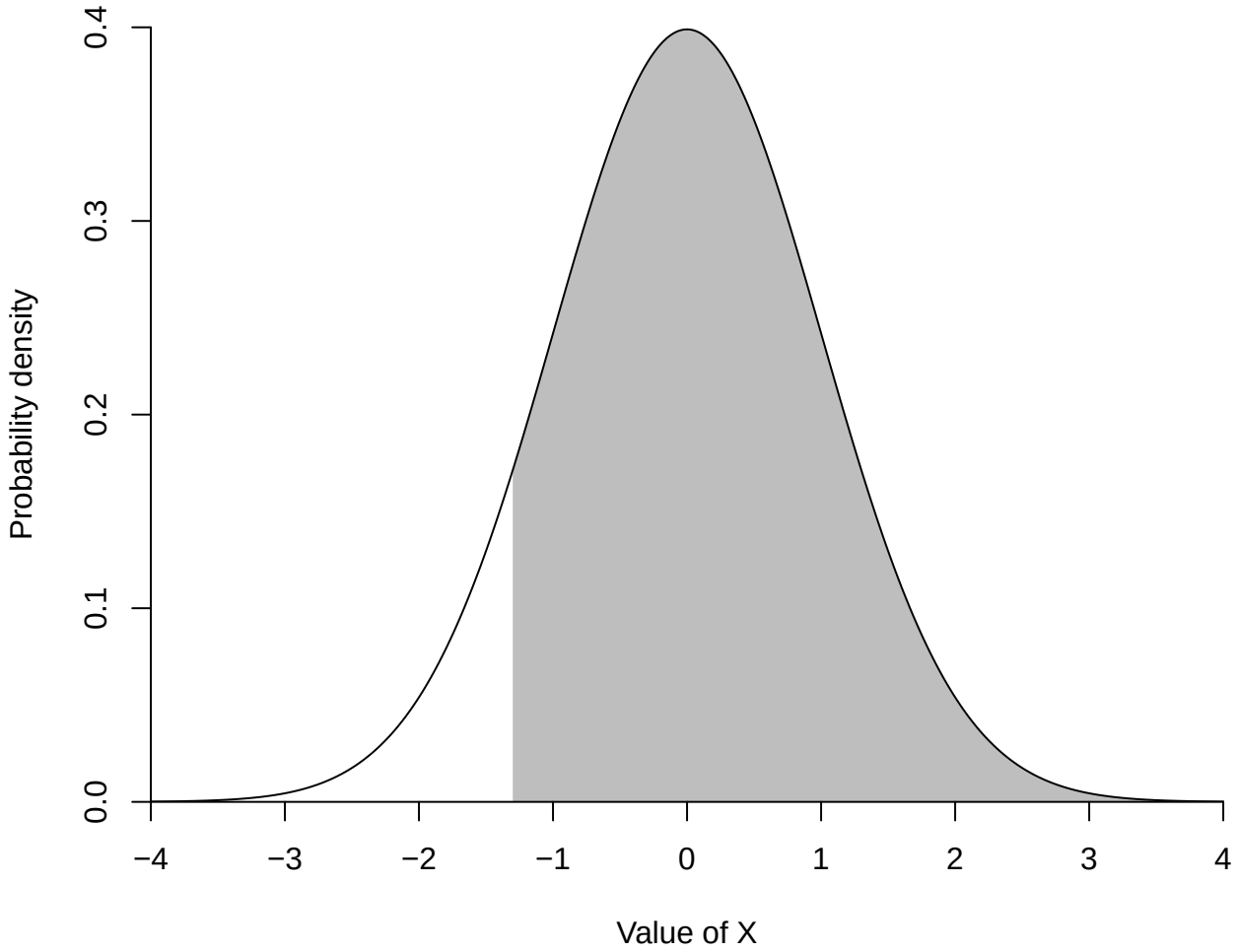
Three rules will help

And remember that the area under a probability density always totals 1



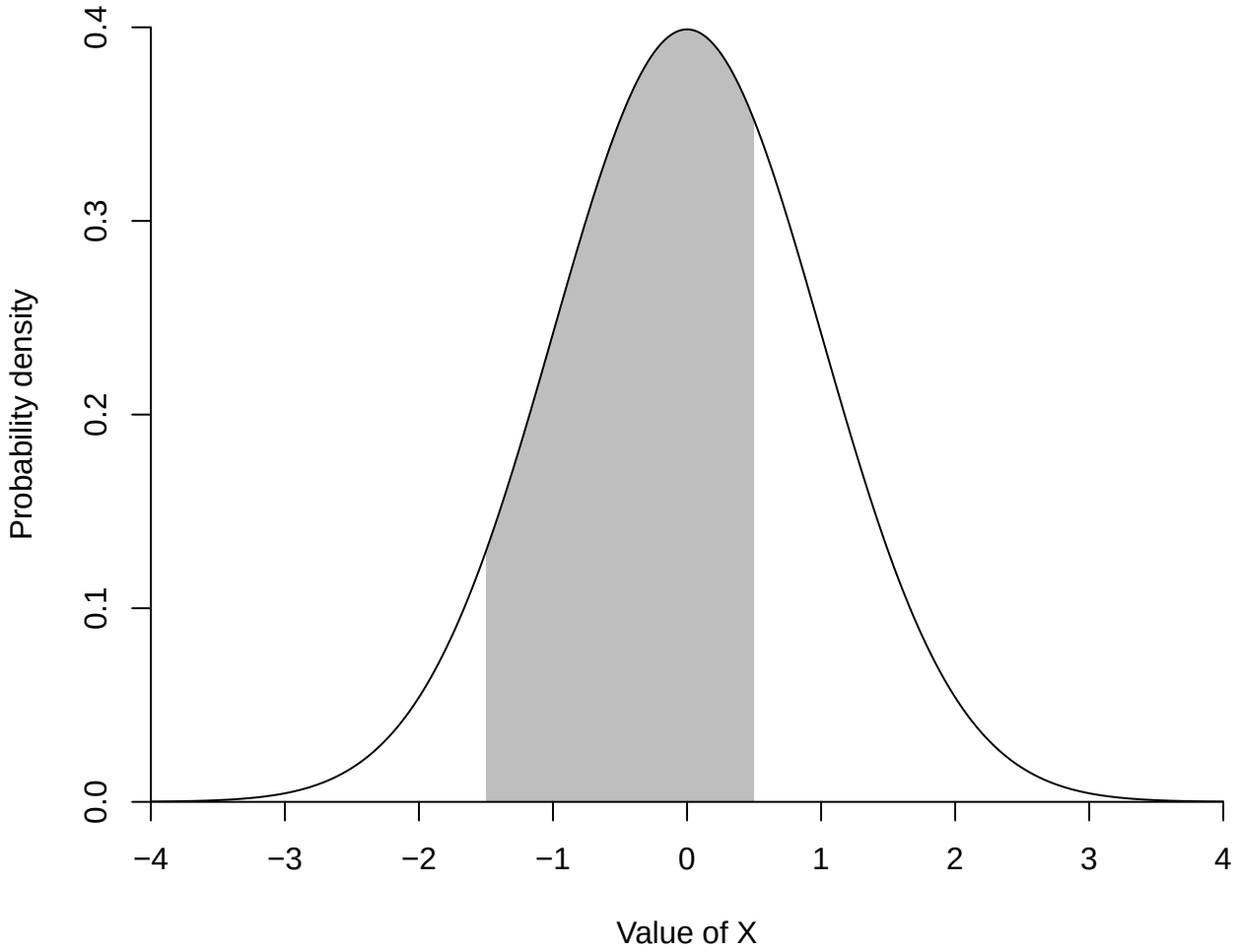
We can often use one tail to calculate the other tail

Suppose we wanted to know $\mathrm{P}(X > -1.3)$, but all we knew was $\mathrm{P}(X < -1.3) = 0.097$, or the area in gray



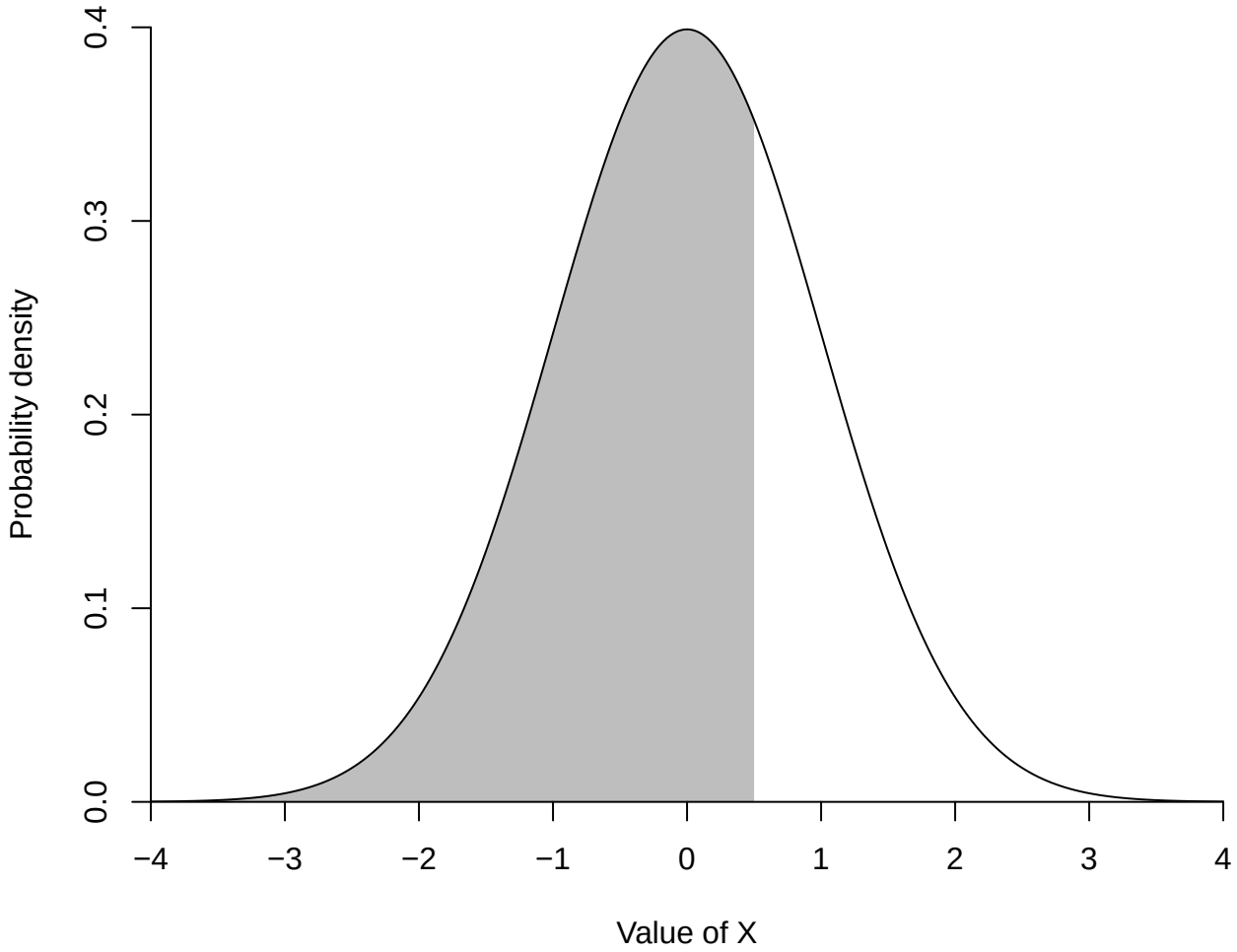
Rule 1:
$\mathrm{P}(X > a) = 1 - \mathrm{P}(X \leq a)$

So $\mathrm{P}(X > -1.3) = 1 - 0.097 = 0.903$, or the new area in gray



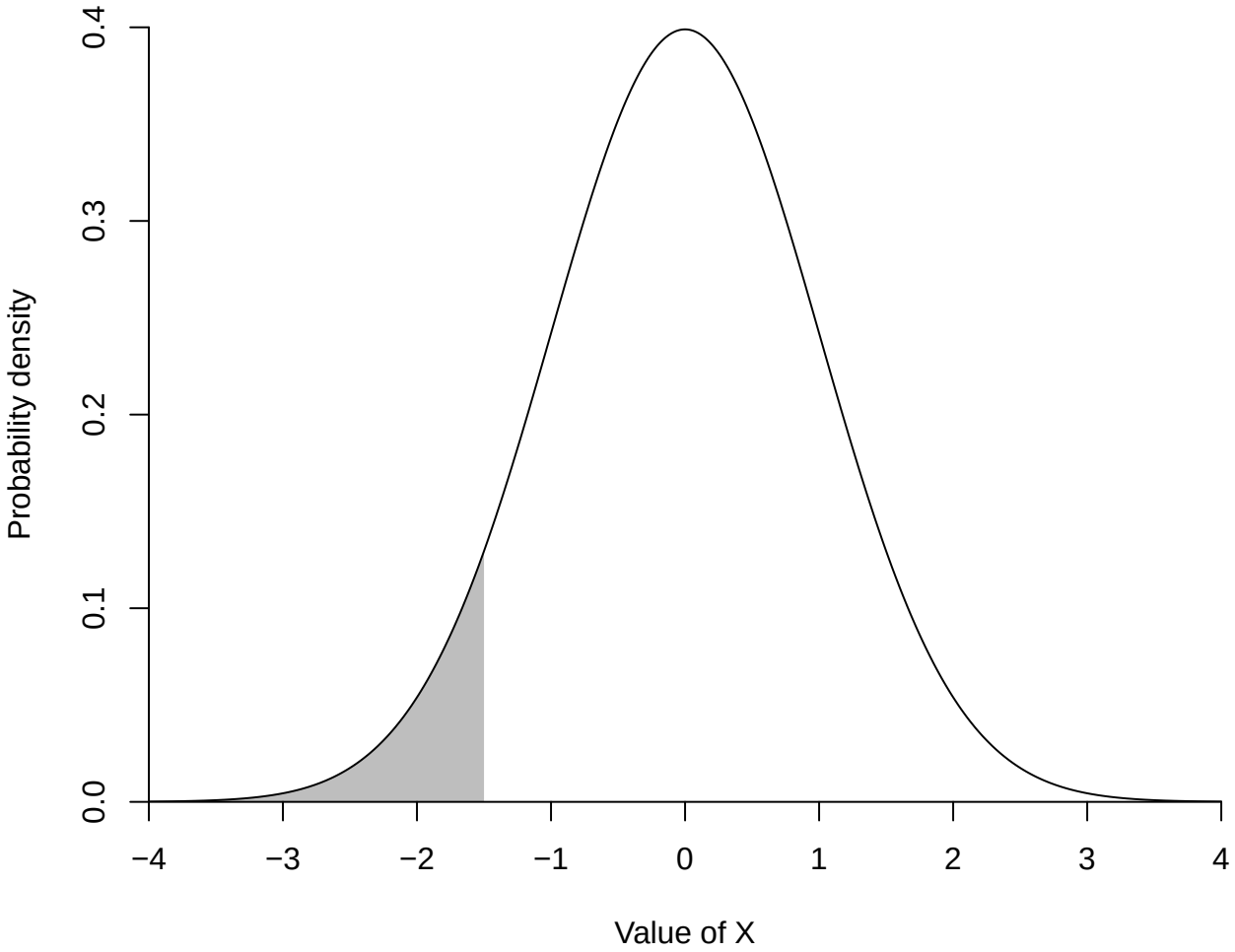
What if we wanted to know an area in the middle of the curve?

That is, what if we wanted to know $\mathrm{P}(a \leq X < b)$, or in this case, $\mathrm{P}(-1.5 < X \leq 0.5)$?

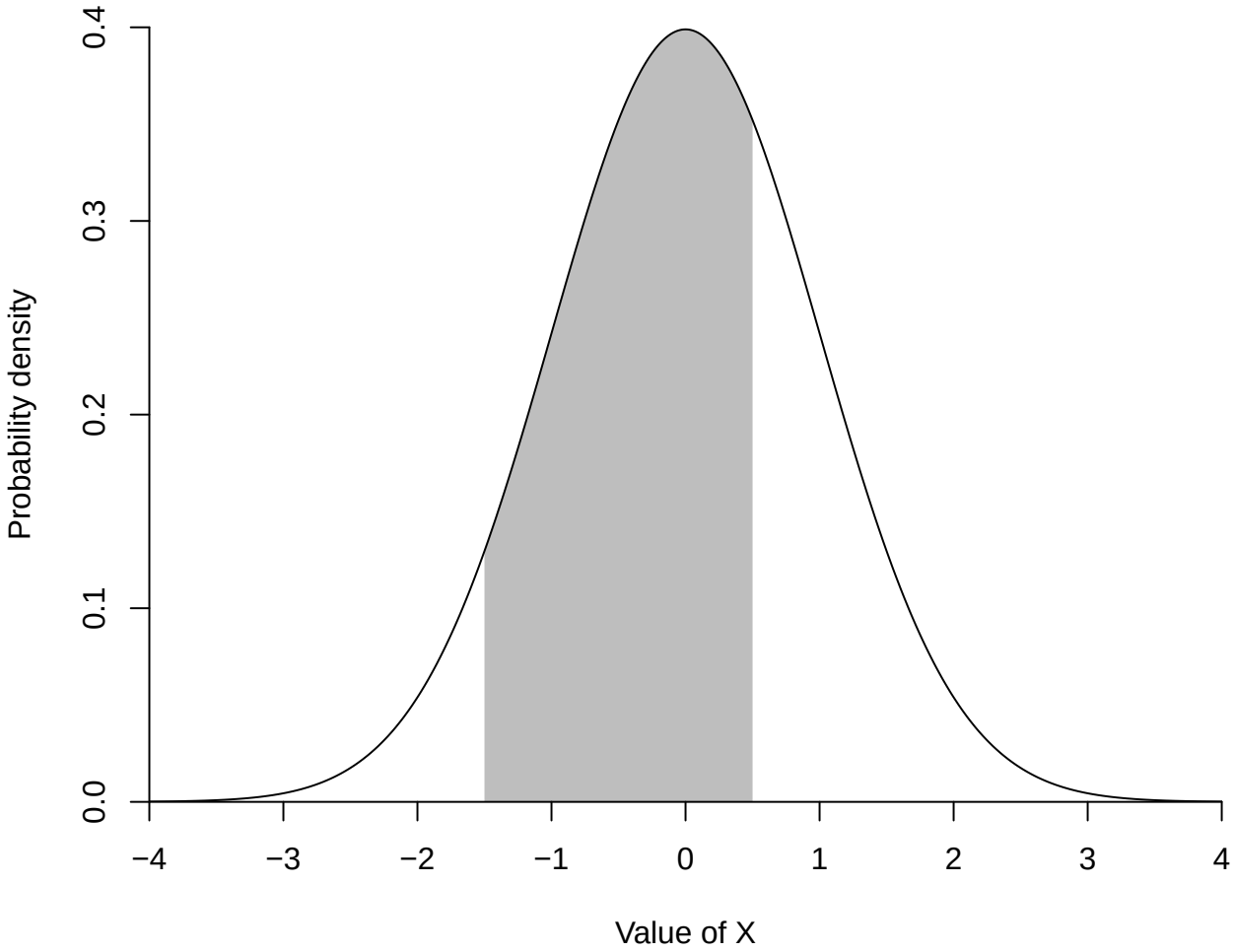


Rule 2: $\mathrm{P}(a < X \leq b) = \mathrm{P}(X \leq b) - \mathrm{P}(X \leq a)$

That is, start with the whole area from $-\infty$ to $b\ldots$

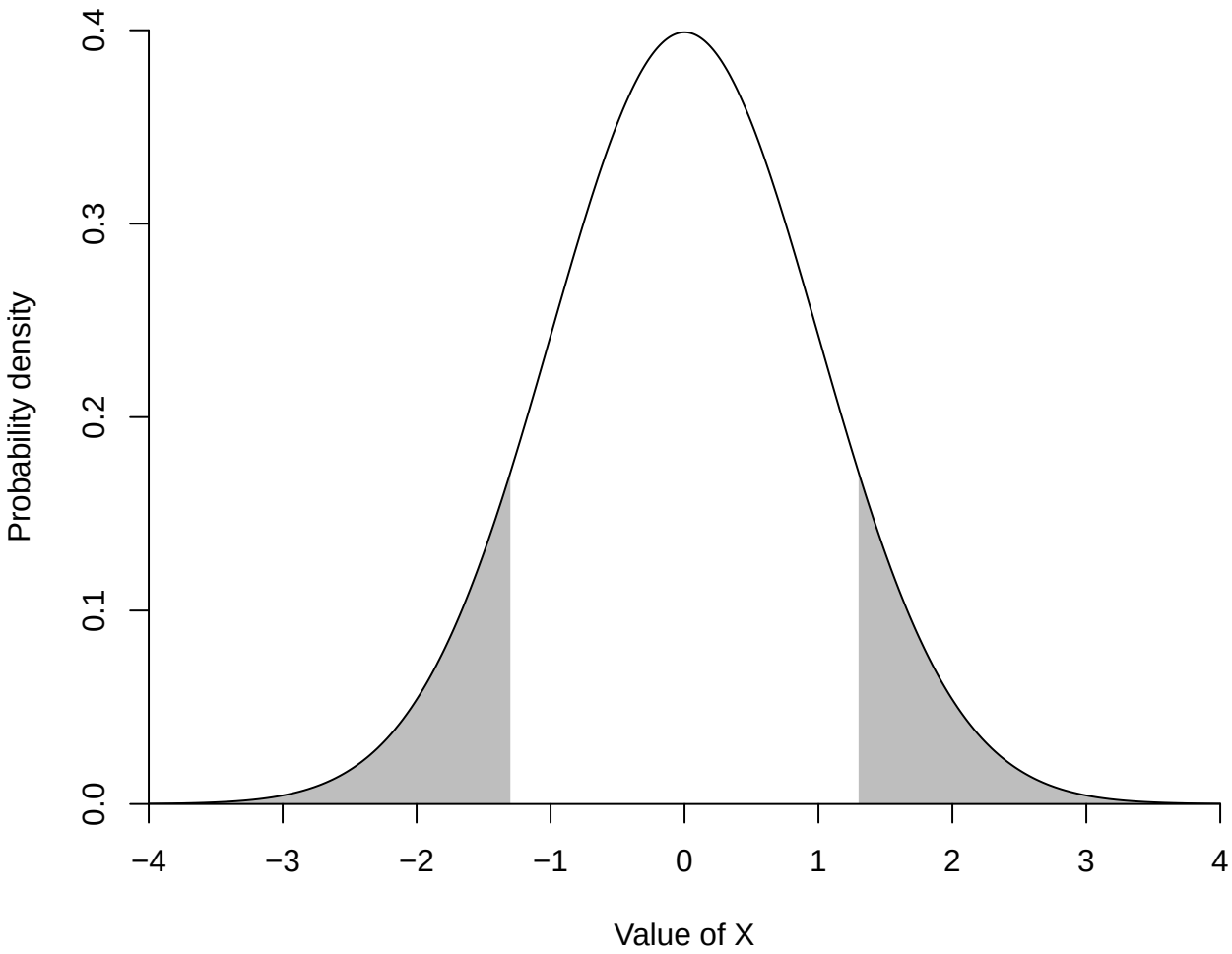


and subtract off the area from $-\infty$ to a



leaving just the part we want!

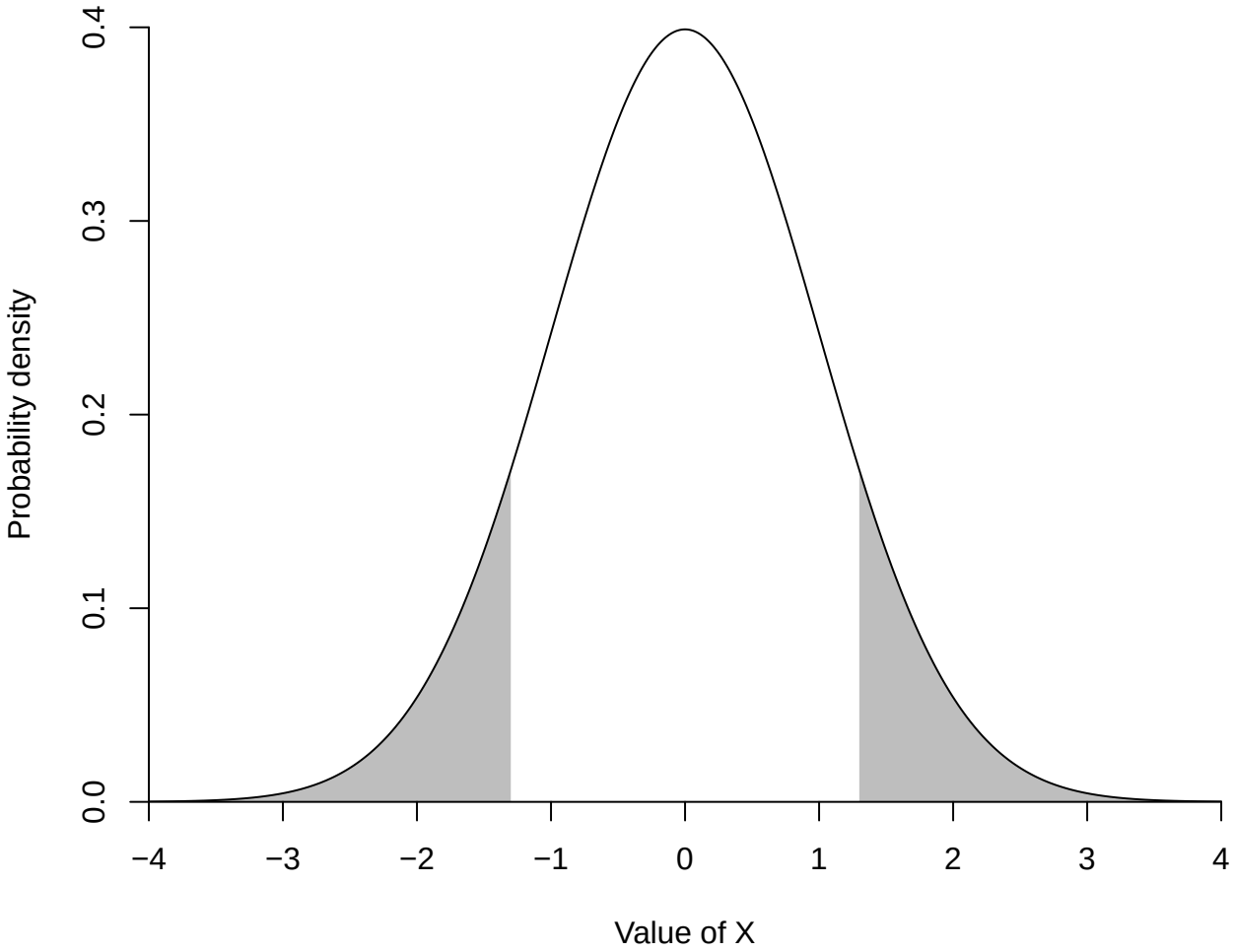
$$\mathrm{P}(-1.5 < X \leq 0.5)$$
$$\mathrm{P}(X \leq 0.5) - \mathrm{P}(X \leq -0.5)$$
$$0.691 - 0.067$$
$$0.625$$



The above rules always work. For symmetric distributions, we have a third rule as well

Here we have $\mathrm{P}(X \leq -1.3)$ and $\mathrm{P}(X \geq 1.3)$

The two areas at left are the same, and so must these probabilities be



Rule 3: $\mathrm{P}(X \leq \mu - d) = \mathrm{P}(X \geq \mu + d)$

For any symmetric distribution, tails equally far from the mean have the same area, and hence values as extreme as $\mu \pm d$ are equally likely

Comparing distributions with different moments

Normally distributed variables can have widely varying means μ and variances σ^2

This raises a question: if we compare two values from two different Normal distributions, how do we decide which is “more extreme”?

For example, which is more impressive?

1. A 90% on a test with a mean of 80% and a standard deviation of 6%

Comparing distributions with different moments

Normally distributed variables can have widely varying means μ and variances σ^2

This raises a question: if we compare two values from two different Normal distributions, how do we decide which is "more extreme"?

For example, which is more impressive?

1. A 90% on a test with a mean of 80% and a standard deviation of 6%
2. A 65% on a test with a mean of 30% and a standard deviation of 25%

Comparing distributions with different moments

Normally distributed variables can have widely varying means μ and variances σ^2

This raises a question: if we compare two values from two different Normal distributions, how do we decide which is "more extreme"?

For example, which is more impressive?

1. A 90% on a test with a mean of 80% and a standard deviation of 6%
2. A 65% on a test with a mean of 30% and a standard deviation of 25%
3. A 38% on a test with a mean of 25% and a standard deviation of 5%

z-scores

To solve this sort of problem, it helps to *standardize* a normal variable to have the same mean and variance

That is, we convert each score to a common metric, called a z-score, in which the mean is 0, and each unit is a standard deviation move away from zero

For random variable x with mean μ and variance σ^2, the z-score is:

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

Notice that while the original variable $X \sim \text{Normal}(\mu, \sigma)$,

the z-score is $Z \sim \text{Normal}(0, 1)$

z-scores: Example

So, which is more impressive?

1. A 90% on a test with a mean of 80% and a standard deviation of 6%

$$z = \frac{x - \mu}{\sigma} = \frac{0.9 - 0.8}{0.06} = 1.67$$

z-scores: Example

So, which is more impressive?

1. A 90% on a test with a mean of 80% and a standard deviation of 6%

$$z = \frac{x - \mu}{\sigma} = \frac{0.9 - 0.8}{0.06} = 1.67$$

2. A 65% on a test with a mean of 30% and a standard deviation of 25%

$$z = \frac{x - \mu}{\sigma} = \frac{0.65 - 0.3}{0.25} = 1.4$$

z-scores: Example

So, which is more impressive?

1. A 90% on a test with a mean of 80% and a standard deviation of 6%

$$z = \frac{x - \mu}{\sigma} = \frac{0.9 - 0.8}{0.06} = 1.67$$

2. A 65% on a test with a mean of 30% and a standard deviation of 25%

$$z = \frac{x - \mu}{\sigma} = \frac{0.65 - 0.3}{0.25} = 1.4$$

3. A 38% on a test with a mean of 25% and a standard deviation of 5%

$$z = \frac{x - \mu}{\sigma} = \frac{0.38 - 0.25}{0.05} = 2.6$$

z-scores: Example

So, which is more impressive?

1. A 90% on a test with a mean of 80% and a standard deviation of 6%

$$z = \frac{x - \mu}{\sigma} = \frac{0.9 - 0.8}{0.06} = 1.67$$

2. A 65% on a test with a mean of 30% and a standard deviation of 25%

$$z = \frac{x - \mu}{\sigma} = \frac{0.65 - 0.3}{0.25} = 1.4$$

3. A 38% on a test with a mean of 25% and a standard deviation of 5%

$$z = \frac{x - \mu}{\sigma} = \frac{0.38 - 0.25}{0.05} = 2.6$$

All three scores are impressive.
But the student with the 38% should be proudest.

z-scores and percentiles

What are the (theoretical) percentiles of the three test scores?

That is, what percentage of test-takers did student 1 beat?
student 2? student 3?

We can easily look up the percentile of a z-score (using Table A in your text)

For our example, for grade x

μ	σ	x	z	percentile
0.80	0.06	0.90	1.67	95th
0.30	0.25	0.65	1.40	92nd
0.25	0.05	0.38	2.60	99th

So all of these scores are actually *A*s.

z-scores and critical values

If we can go from z-scores to percentiles, we can also go from percentiles to z-scores

Suppose you took the third exam, with the mean of 25% and the standard deviation of 5%.

How well would you have to score to be in the top 20% of the class?

To answer this, we first need to find the z^*, or critical value, which marks the 80th percentile.

z-scores and critical values

we first need to find the z^*, or critical value, which marks the 80th percentile.

If we look this up in Table A, we find the desired $z^* \approx 0.84$

What actual test score does this correspond to?

Note that if $z = (x - \mu)/\sigma$, then

$$x^* \quad = \quad z^*\sigma + \mu$$

z-scores and critical values

we first need to find the z^*, or critical value, which marks the 80th percentile.

If we look this up in Table A, we find the desired $z^* \approx 0.84$

What actual test score does this correspond to?

Note that if $z = (x - \mu)/\sigma$, then

$$\begin{aligned} x^* &= z^*\sigma + \mu \\ &= 0.84 \times 0.05 + 0.25 \end{aligned}$$

z-scores and critical values

we first need to find the z^*, or critical value, which marks the 80th percentile.

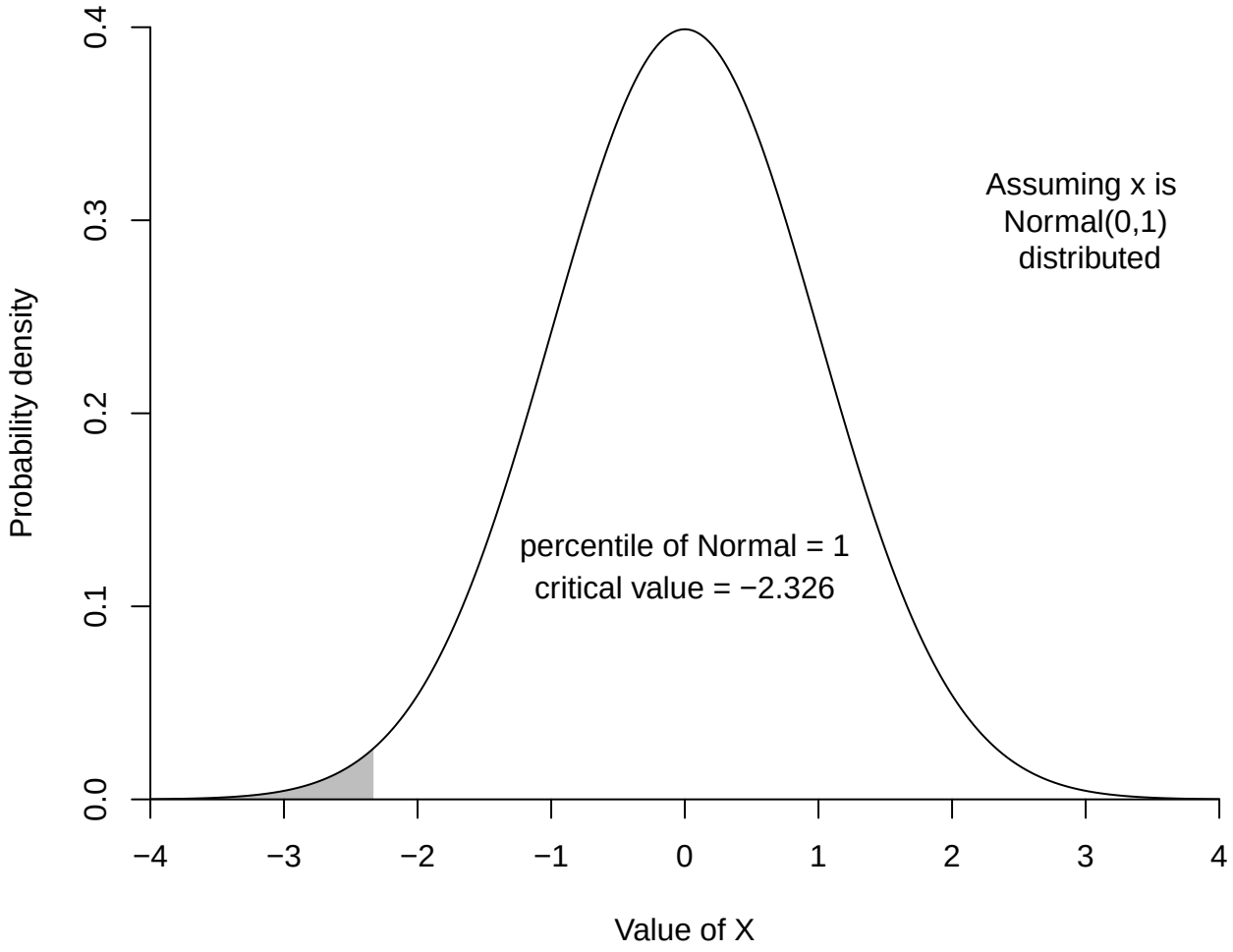
If we look this up in Table A, we find the desired $z^* \approx 0.84$

What actual test score does this correspond to?

Note that if $z = (x - \mu)/\sigma$, then

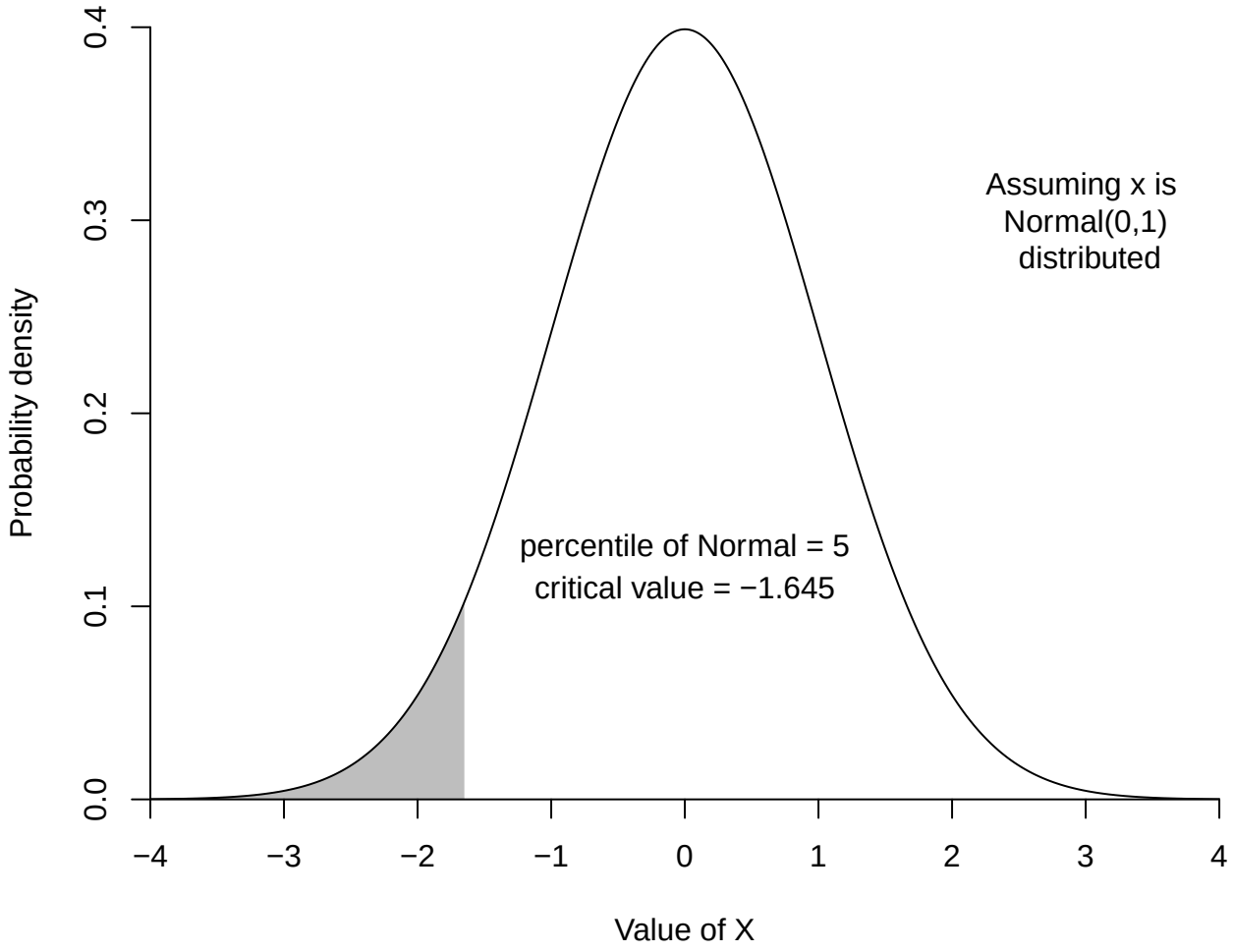
$$\begin{aligned} x^* &= z^*\sigma + \mu \\ &= 0.84 \times 0.05 + 0.25 \\ &= 29.2\% \end{aligned}$$

Upshot: if we know the theoretical distribution of a Normal variable, we can freely convert between the variable, z-scores, and percentiles



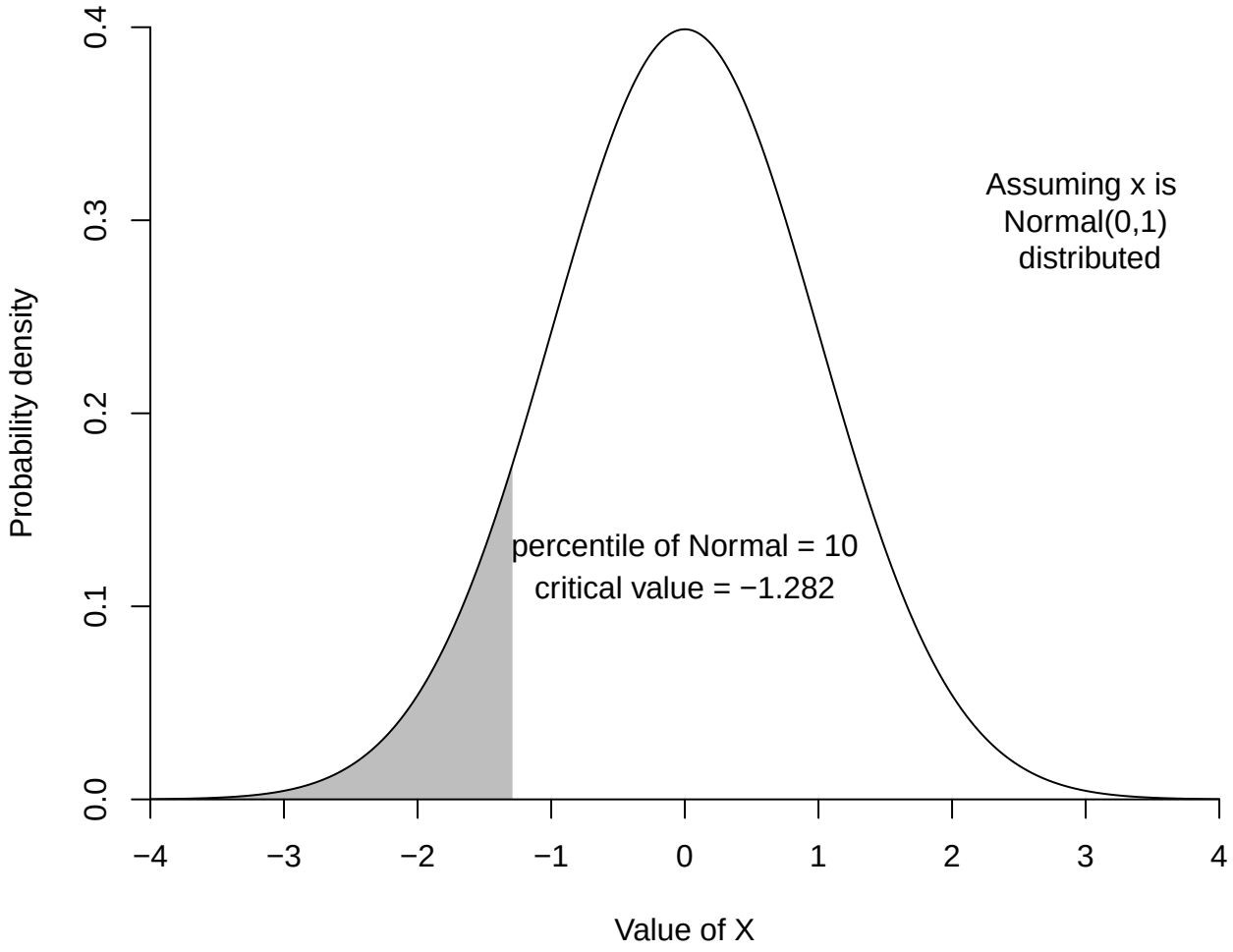
Let's take a quick survey of the percentiles and critical values of the standard Normal

Here is the first 1% of the distribution



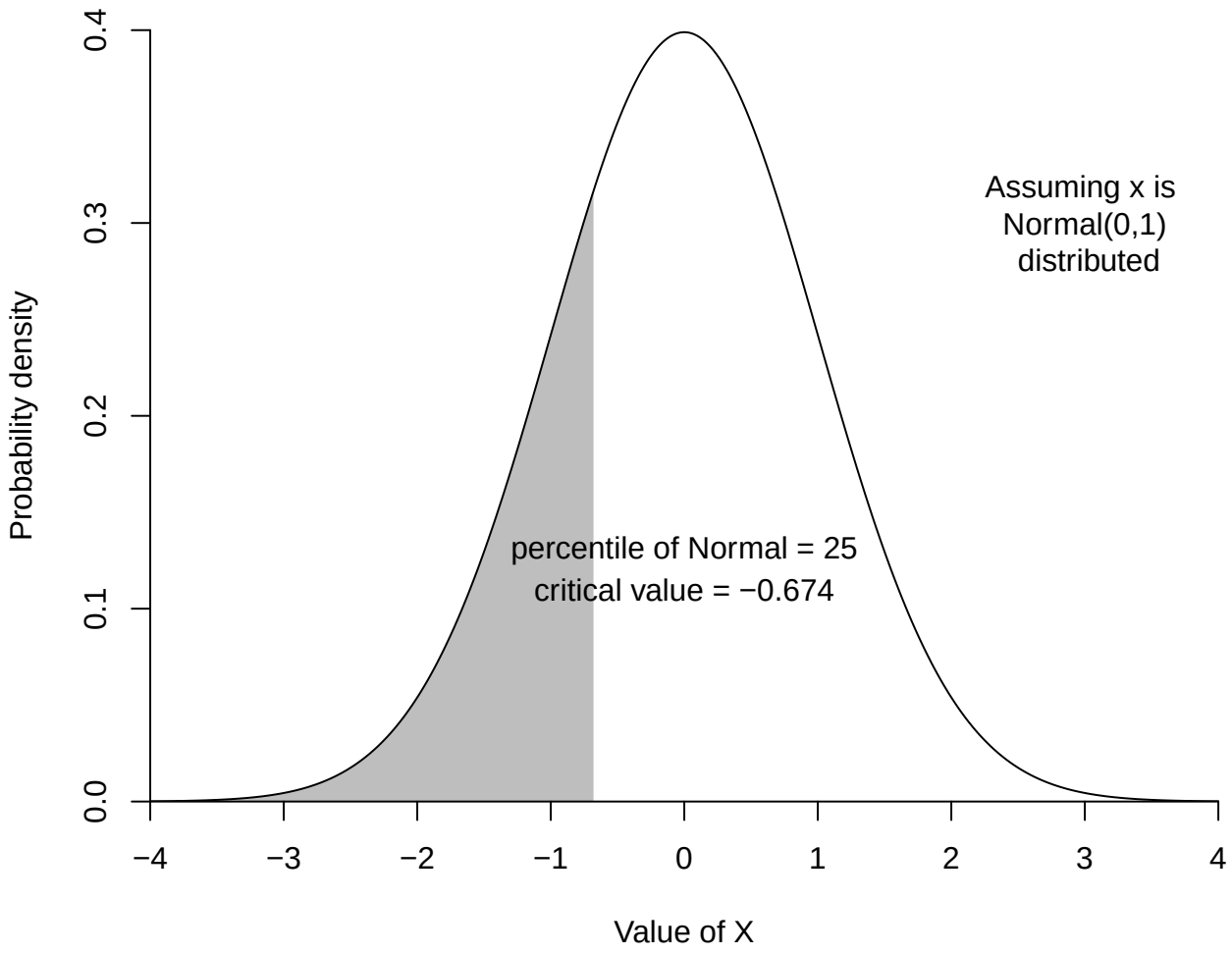
Let's take a quick survey of the percentiles and critical values of the standard Normal

Here is the first 5% of the distribution



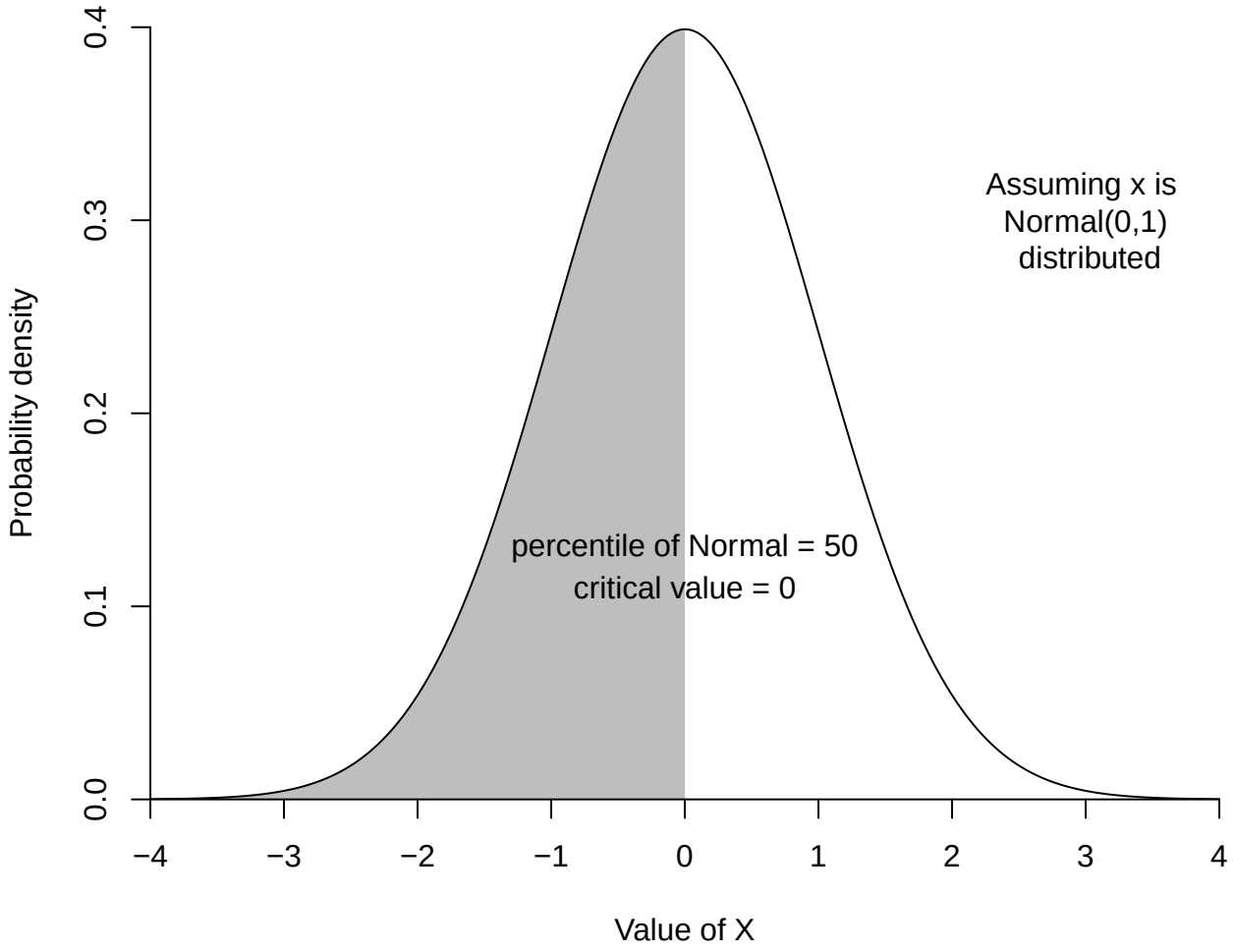
Let's take a quick survey of the percentiles and critical values of the standard Normal

Here is the first 10% of the distribution



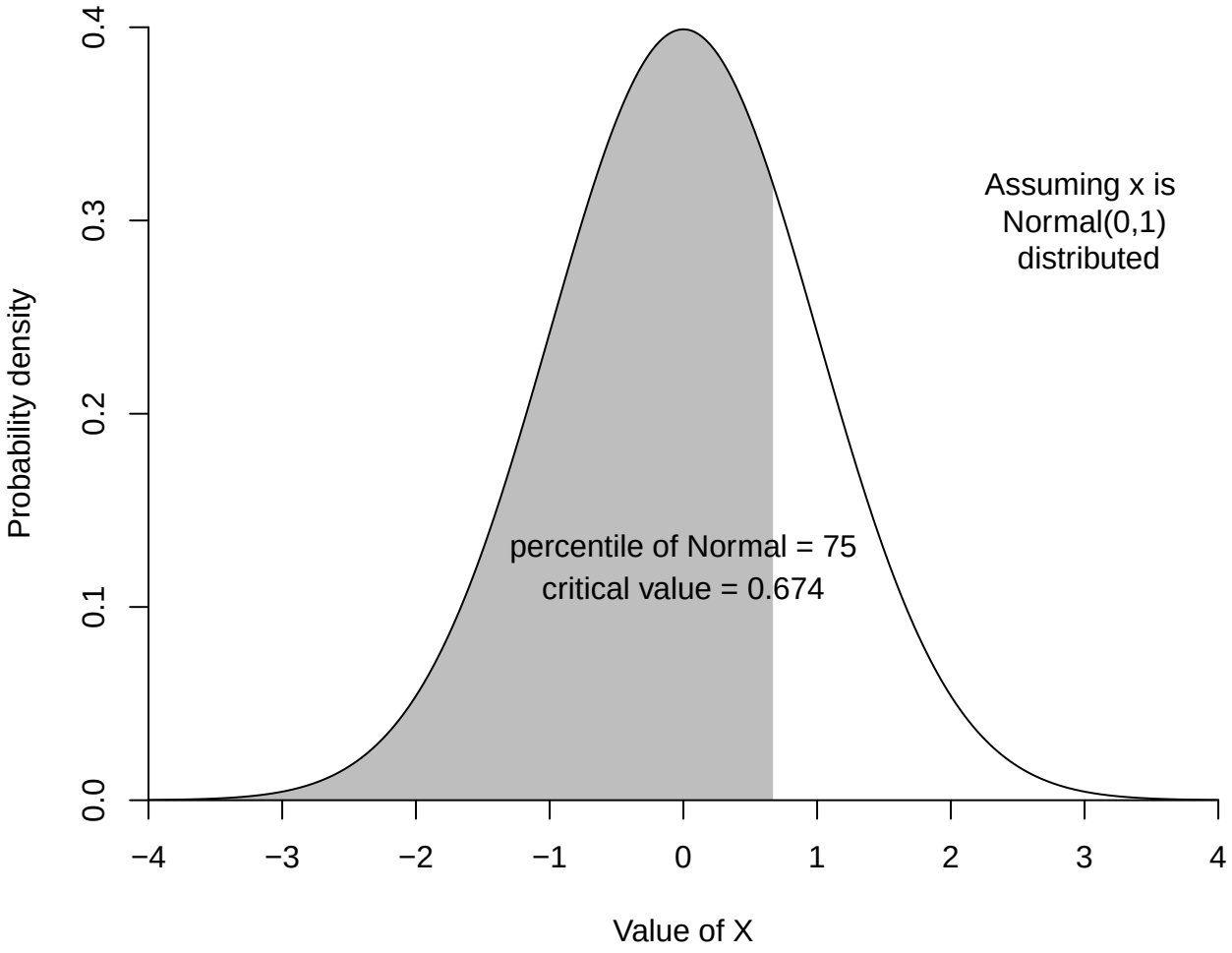
Let's take a quick survey of the percentiles and critical values of the standard Normal

Here is the first 25% of the distribution



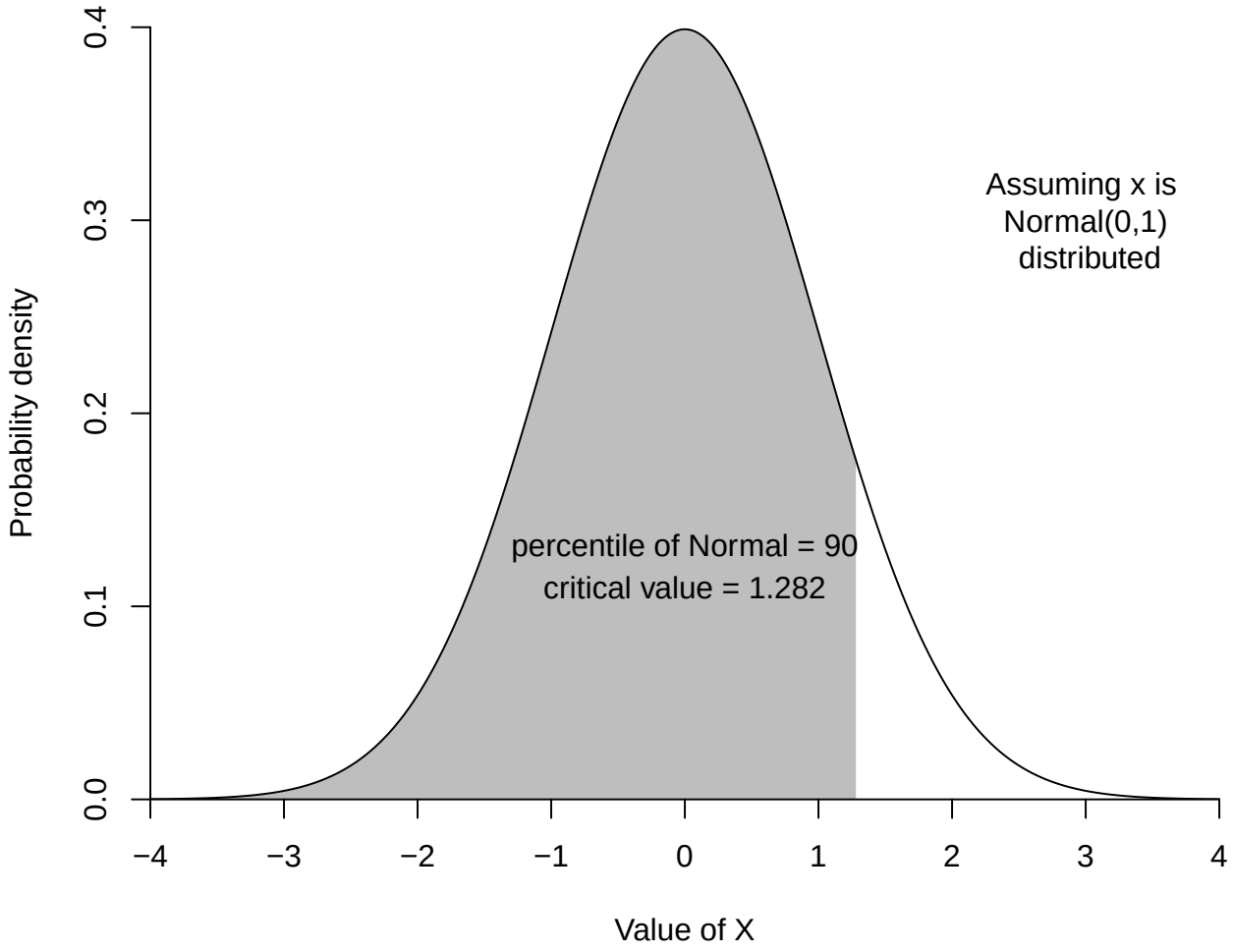
Let's take a quick survey of the percentiles and critical values of the standard Normal

Here is the first 50% of the distribution



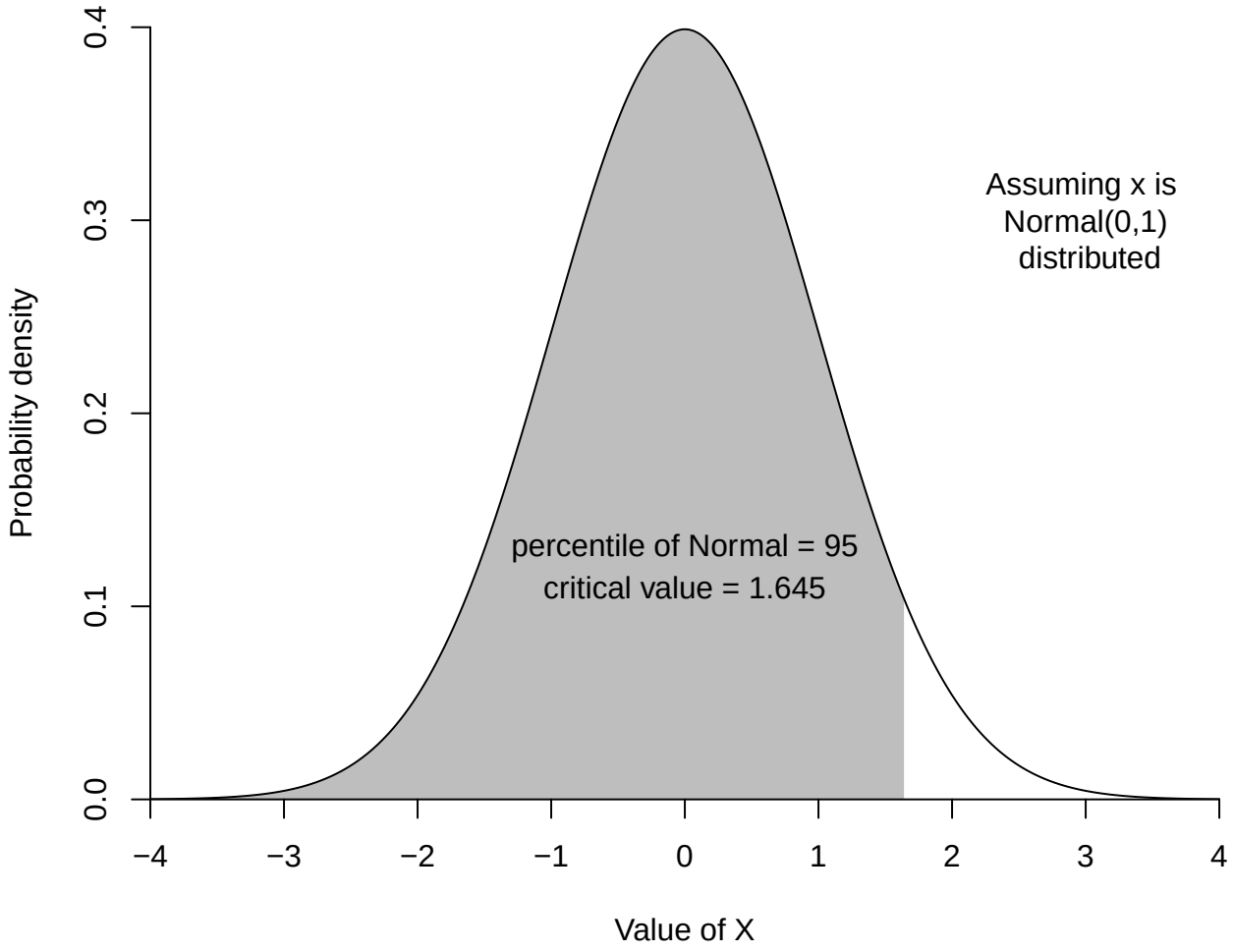
Let's take a quick survey of the percentiles and critical values of the standard Normal

Here is the first 75% of the distribution



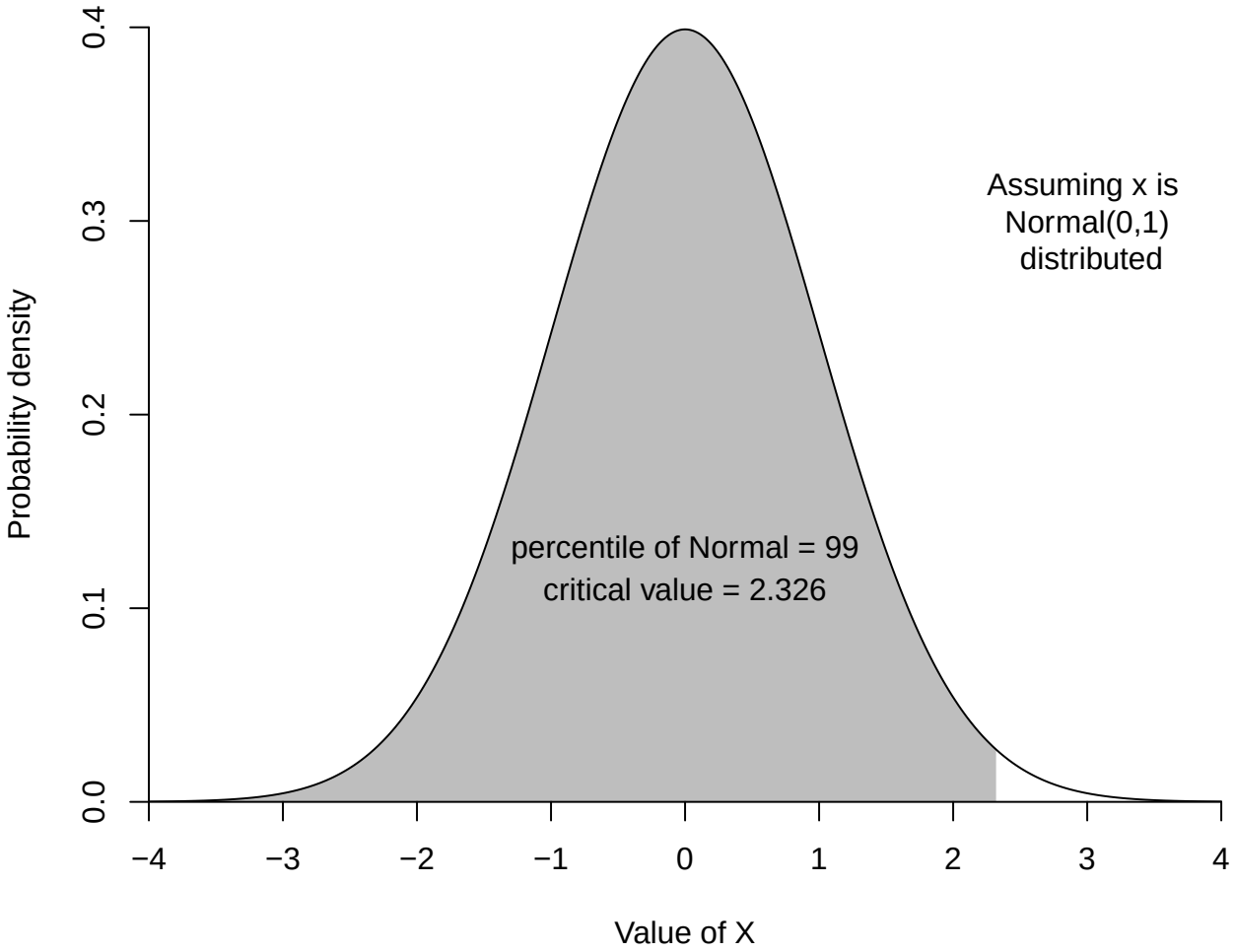
Let's take a quick survey of the percentiles and critical values of the standard Normal

Here is the first 90% of the distribution



Let's take a quick survey of the percentiles and critical values of the standard Normal

Here is the first 95% of the distribution



Let's take a quick survey of the percentiles and critical values of the standard Normal

Here is the first 99% of the distribution

Unemployment example

Let's apply this framework to a real world variable.

Unemployment example

Let's apply this framework to a real world variable.

Unemployment in the US in 2010 was 9.6%, but varied across states

Unemployment example

Let's apply this framework to a real world variable.

Unemployment in the US in 2010 was 9.6%, but varied across states

Suppose that the average state had 9.6% unemployment, but that the standard deviation across states is 2.2.

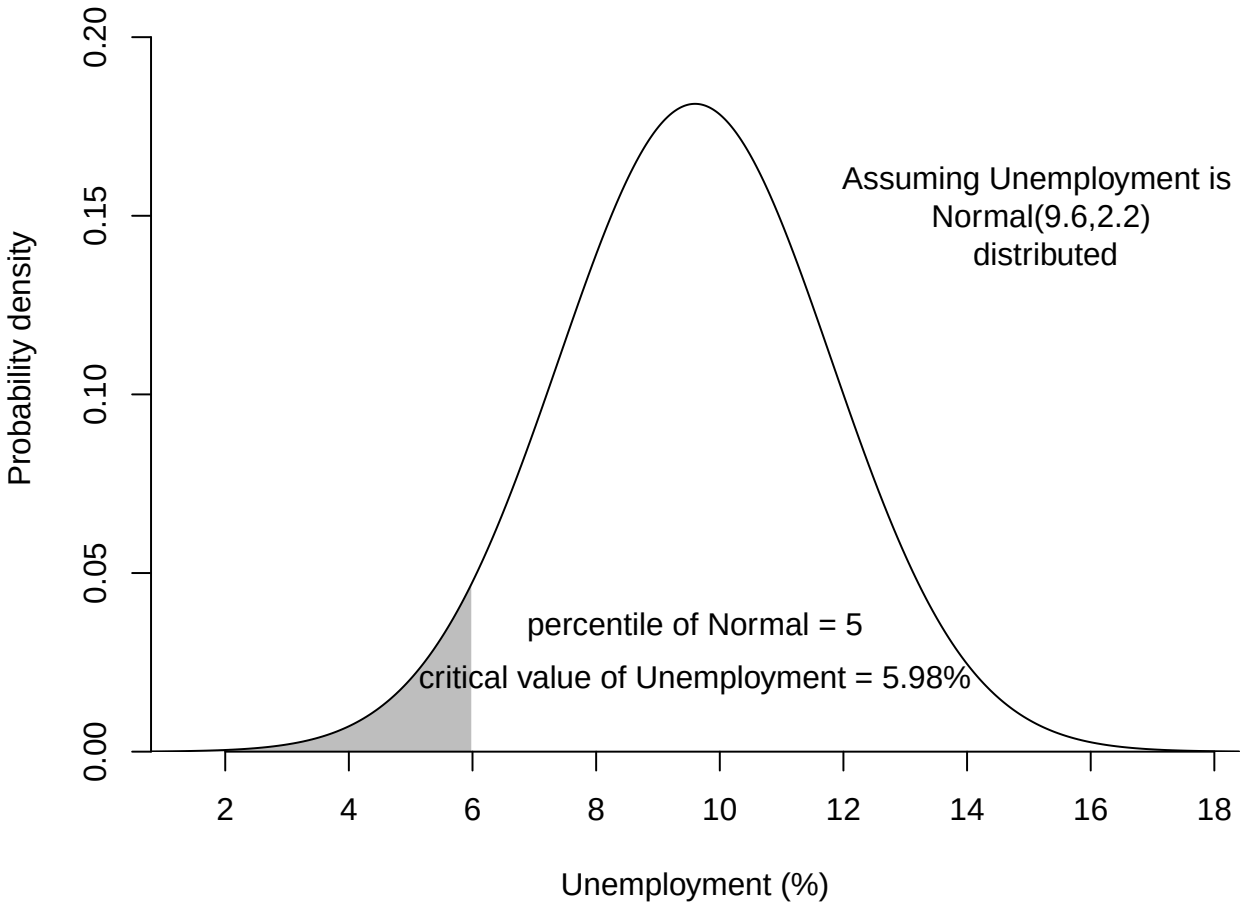
Unemployment example

Let's apply this framework to a real world variable.

Unemployment in the US in 2010 was 9.6%, but varied across states

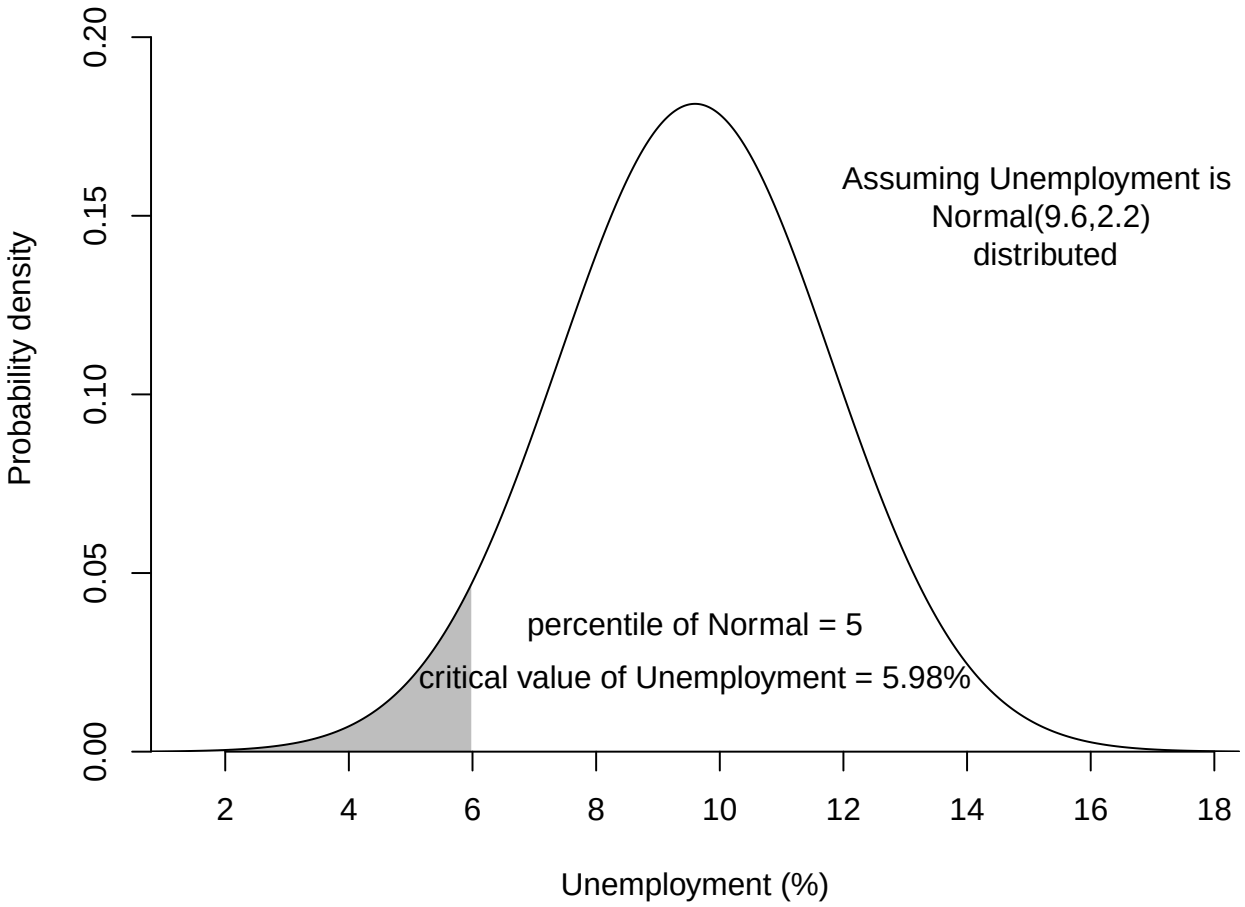
Suppose that the average state had 9.6% unemployment, but that the standard deviation across states is 2.2.

If we suppose unemployment is Normally distributed,
this leads to a Normal(9.6, 2.2) distribution



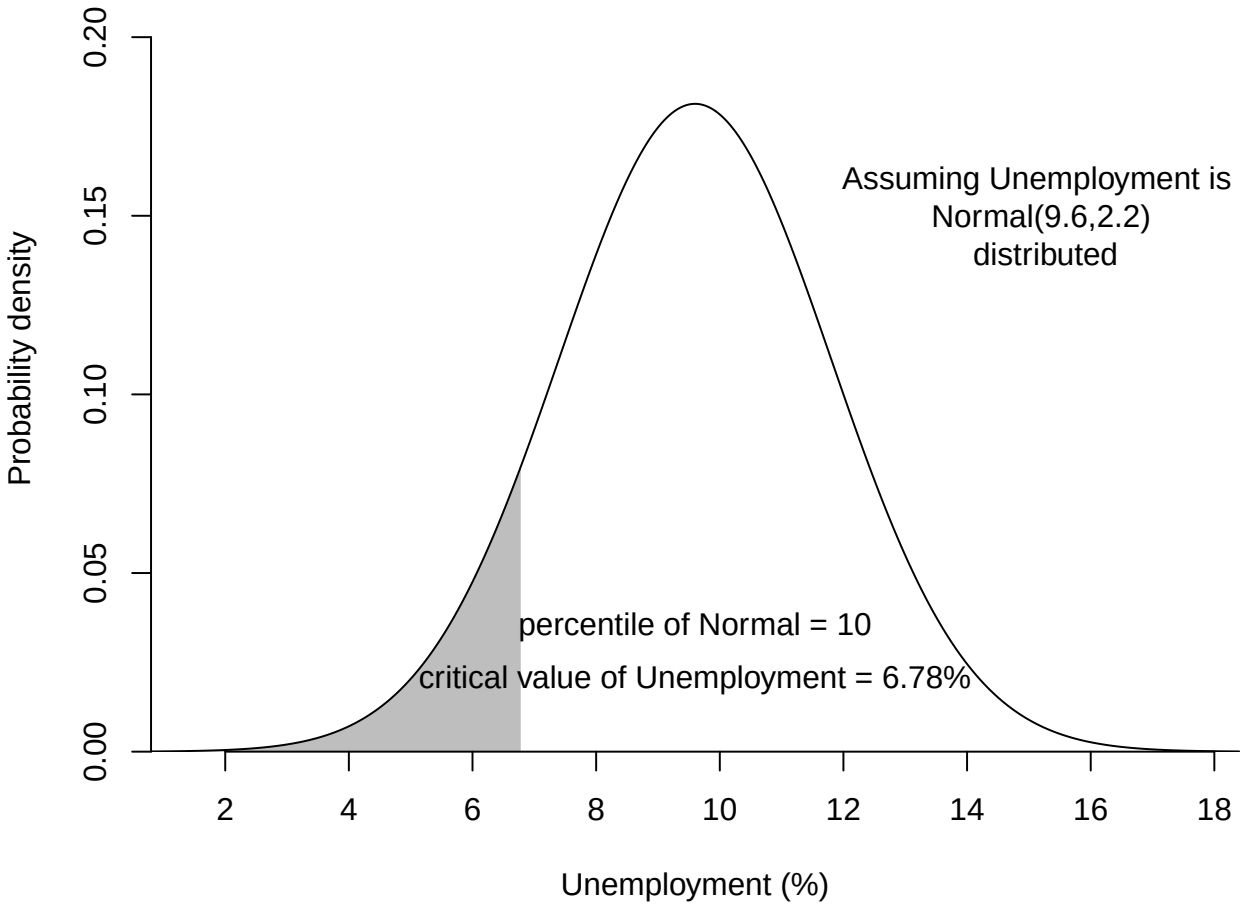
Using the given distribution, I have calculated, using Table A, the critical values for various percentiles

That is, I look up z^* at the requested percentile, then calculate $z^*\sigma + \mu$



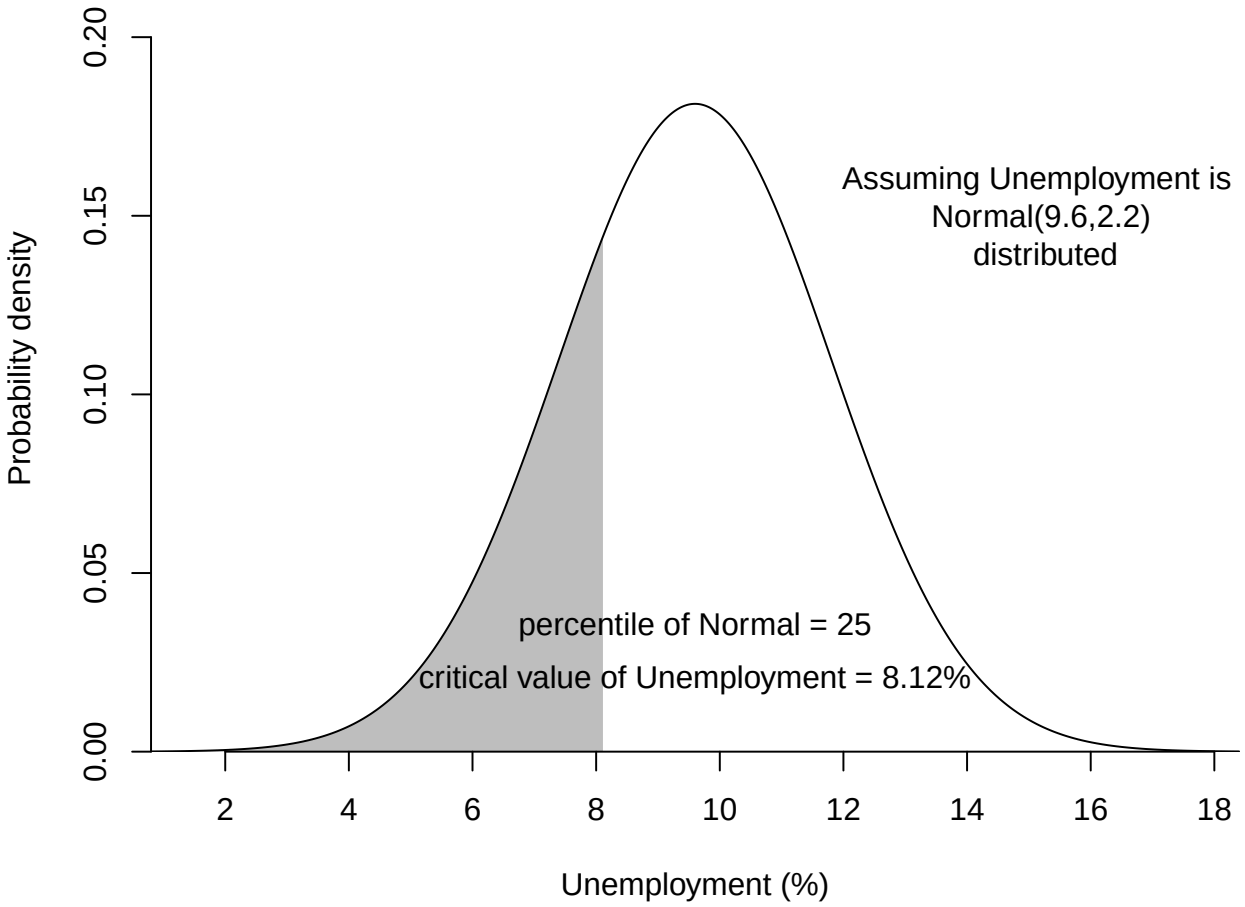
Using the given distribution, I have calculated, using Table A, the critical values for various percentiles

I find that 5% of states should be below 5.98% unemployment



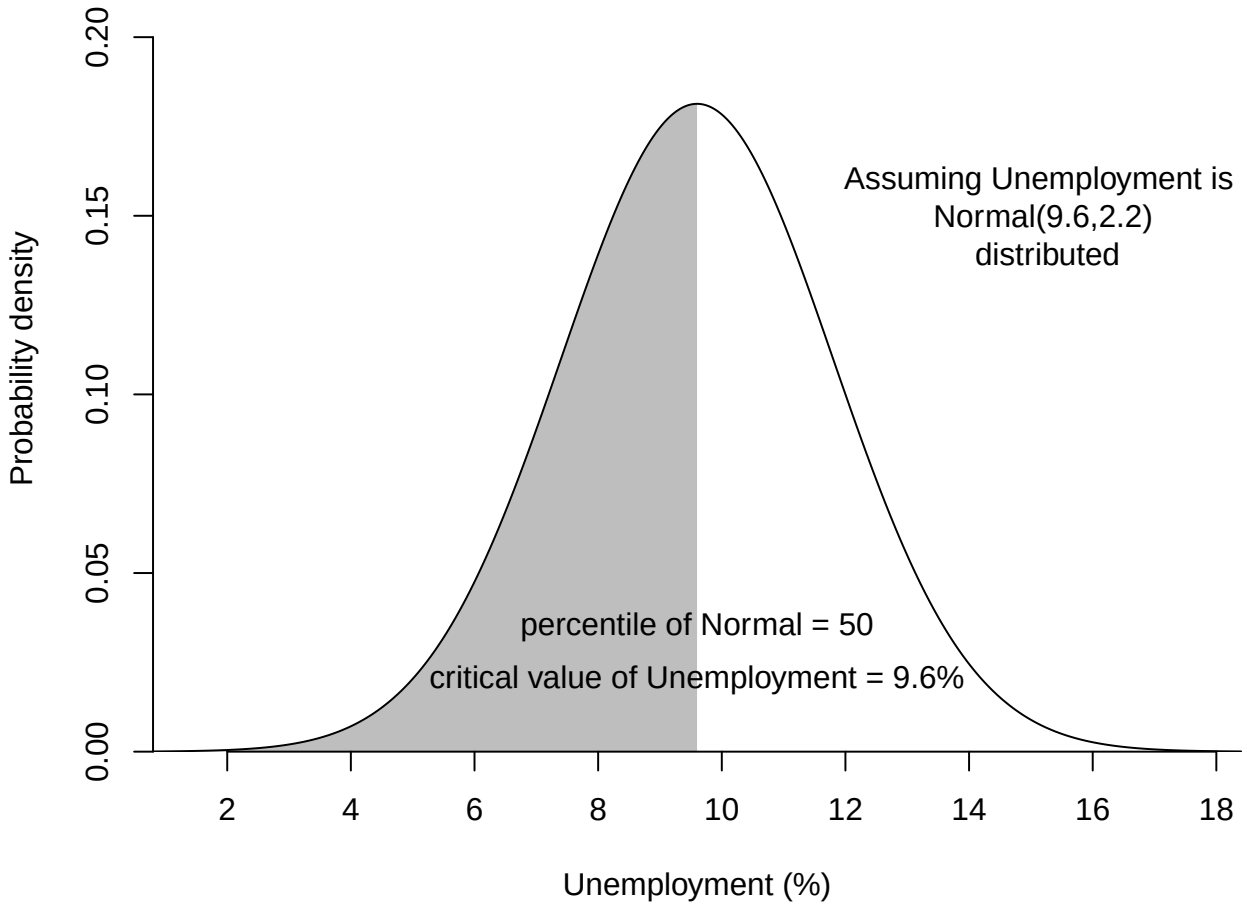
Using the given distribution, I have calculated, using Table A, the critical values for various percentiles

I find that 10% of states should be below 6.78% unemployment



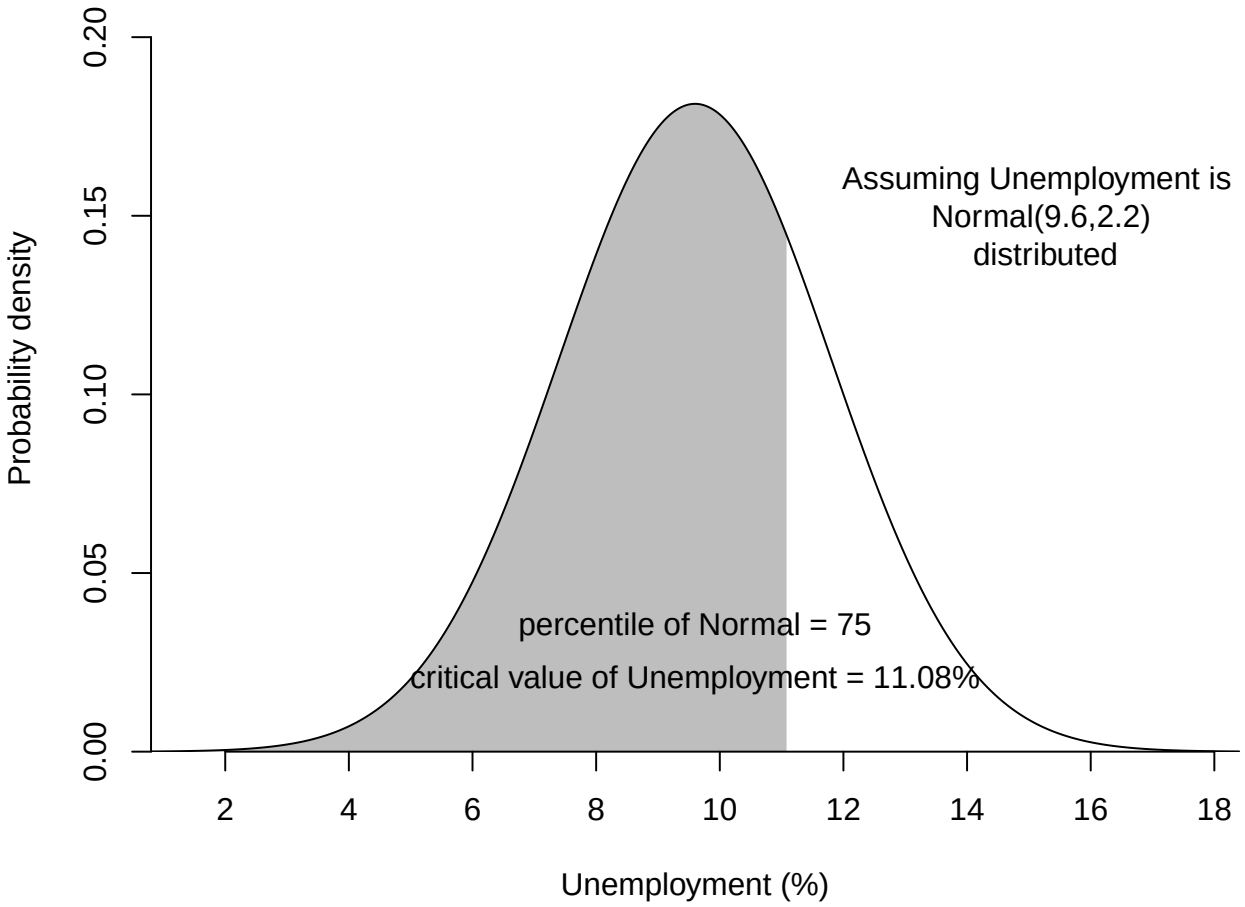
Using the given distribution, I have calculated, using Table A, the critical values for various percentiles

I find that 25% of states should be below 8.12% unemployment



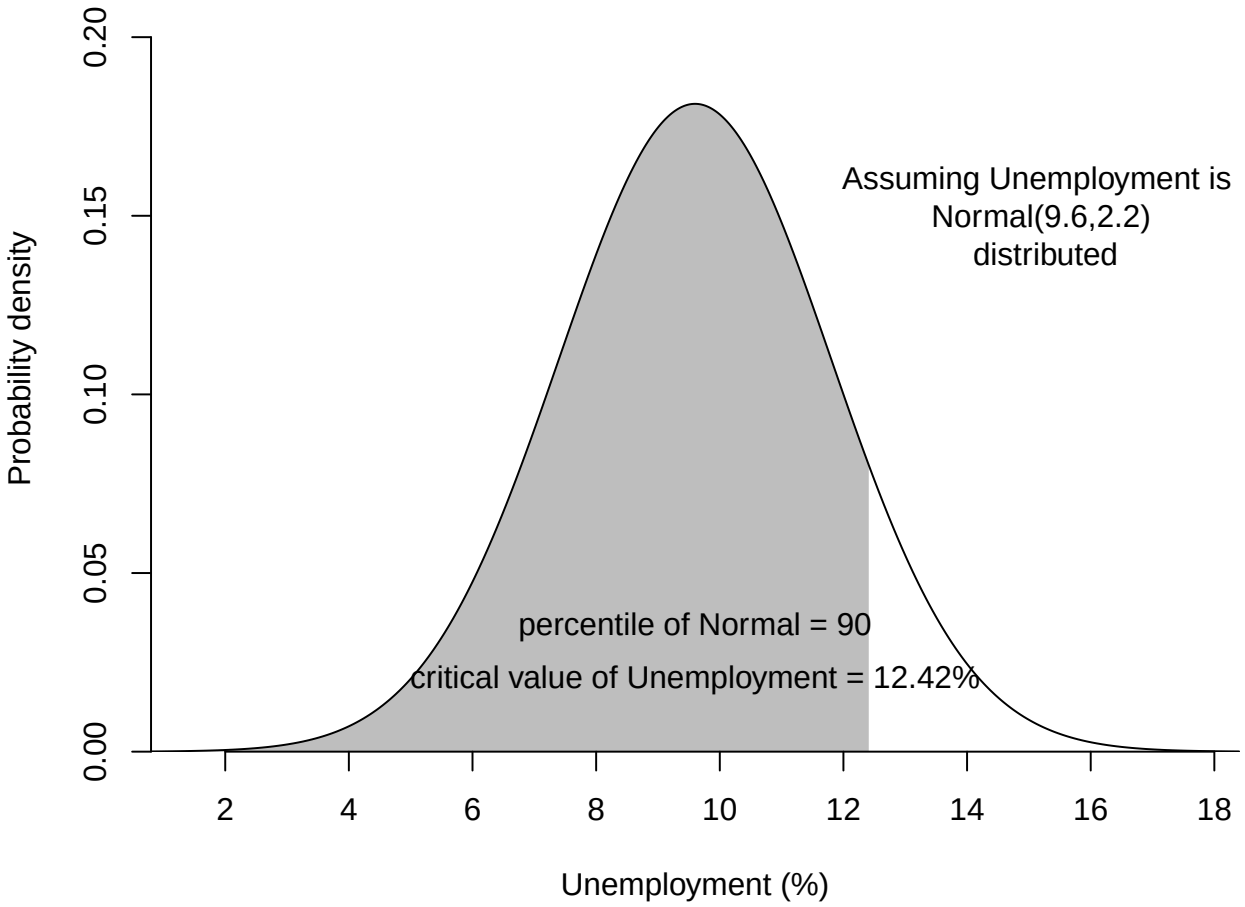
Using the given distribution, I have calculated, using Table A, the critical values for various percentiles

I find that 50% of states should be below 9.6% unemployment



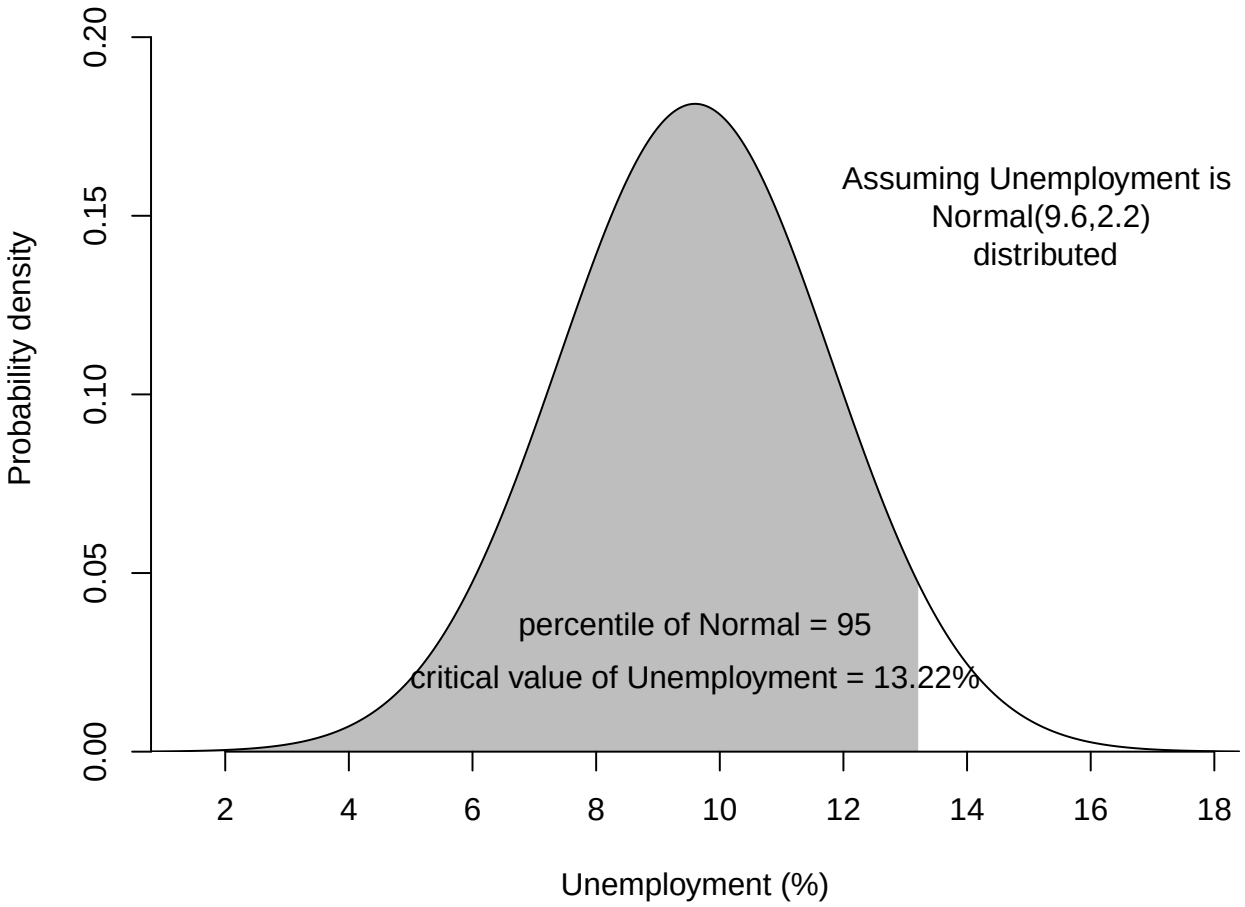
Using the given distribution, I have calculated, using Table A, the critical values for various percentiles

I find that 75% of states should be below 11.08% unemployment



Using the given distribution, I have calculated, using Table A, the critical values for various percentiles

I find that 90% of states should be below 12.42% unemployment



Using the given distribution, I have calculated, using Table A, the critical values for various percentiles

I find that 95% of states should be below 13.22% unemployment

z-scores and probability intervals

Suppose you wanted to summarize the range of most probable outcomes for a theoretical Normal distribution.

For example, if the mean male height in the US is 5 ft 10 in, and the standard deviation is 3 in, *and* height is Normally distributed,

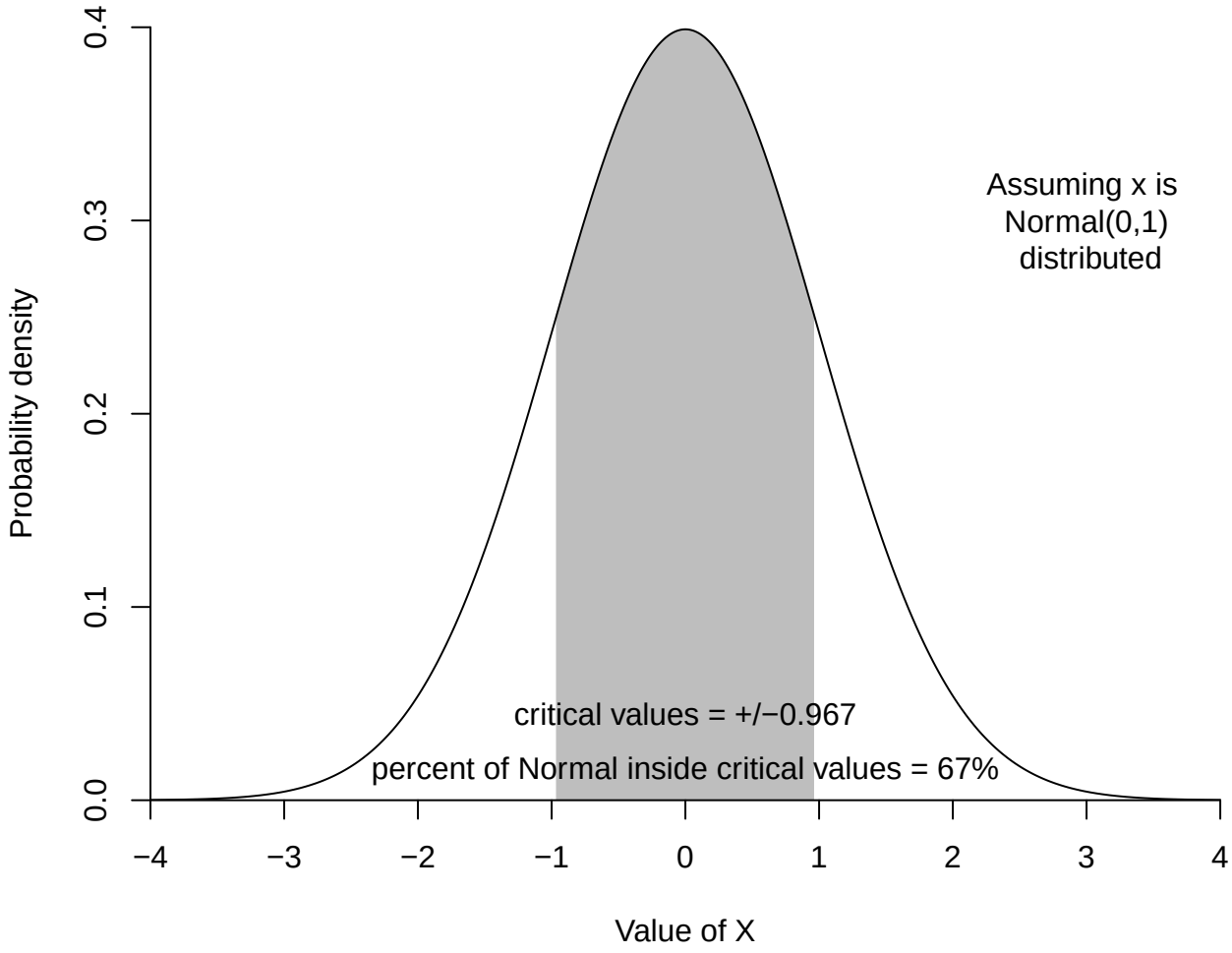
1. What critical values of height bound two-third of all men?
2. What critical values of height bound 95% of all men?
3. What critical values of height bound 99% of all men?

z-scores and probability intervals

What critical values of height bound 95% of all men?

Slightly tricky:

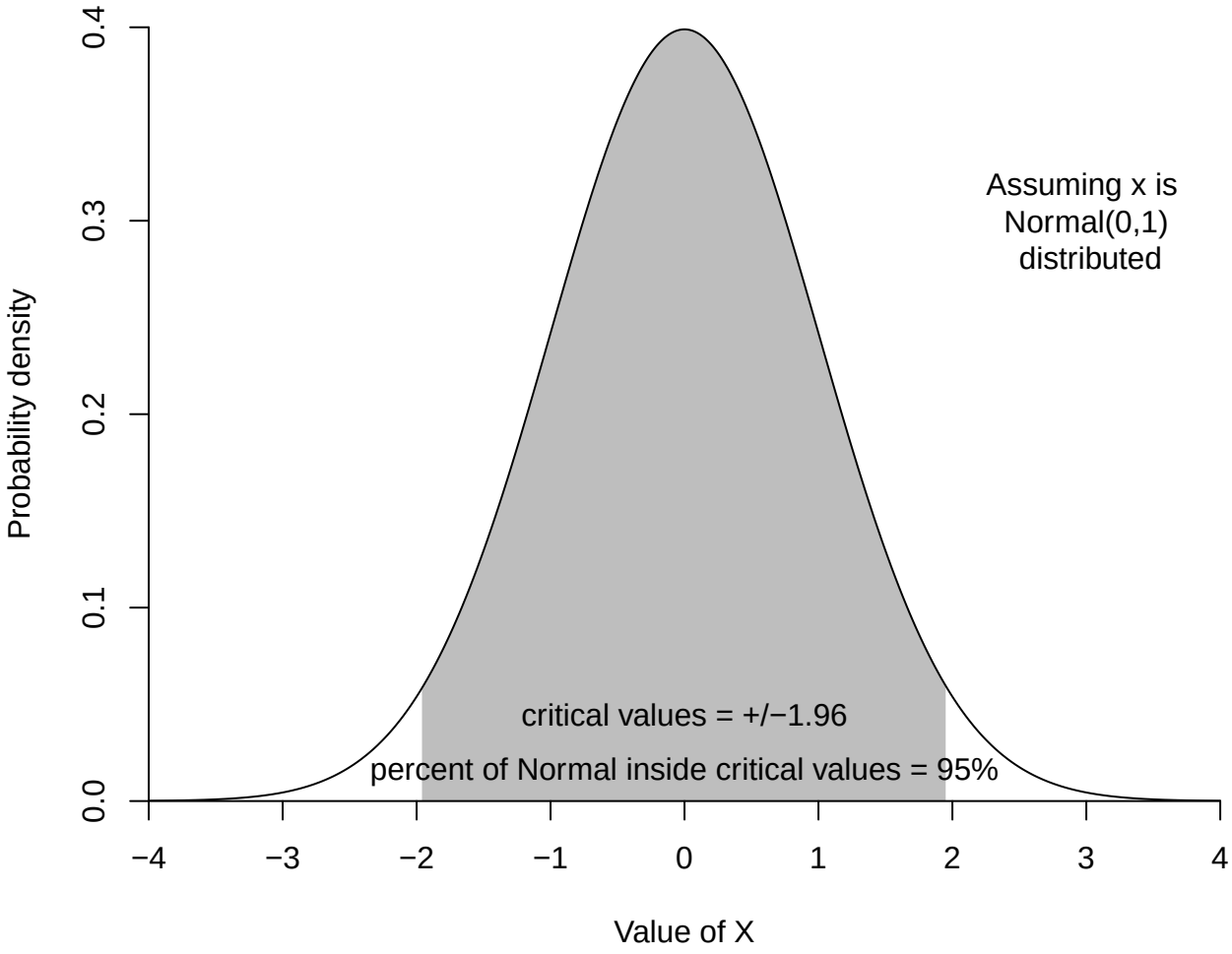
We need the critical values for the 2.5th and 97.5th percentiles (Why?)



The critical values for the 67%, 95%, and 99% intervals are memorable

Two-thirds of a Normal distribution lies within ≈ 1 standard deviation of the mean

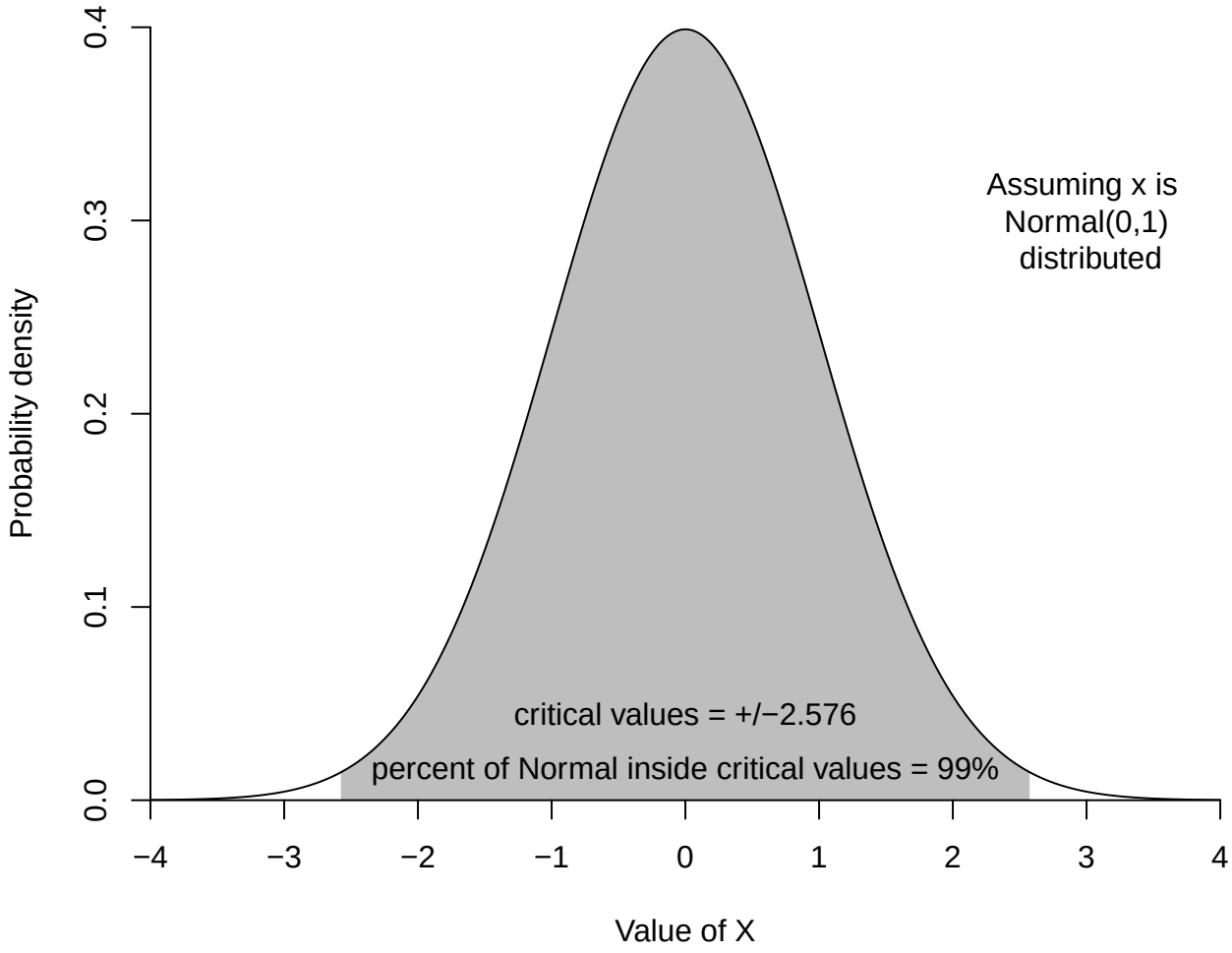
(Remember: z-scores are in standard deviation units!)



The critical values for the 67%, 95%, and 99% intervals are memorable

Two-thirds of a Normal distribution lies within ≈ 2 standard deviation of the mean

(Remember: z-scores are in standard deviation units!)



The critical values for the 67%, 95%, and 99% intervals are memorable

Two-thirds of a Normal distribution lies within ≈ 2.5 standard deviation of the mean

(Remember: z-scores are in standard deviation units!)

If the mean male height in the US is 5 ft 10 in, and the standard deviation is 3 in, *and* height is Normally distributed,

1. What critical values of height bound two-third of all men?
   $70 \text{ inches } \pm 3 \text{ inches } \times 0.967 \approx$ 5 ft 7 to 6 ft 1.

If the mean male height in the US is 5 ft 10 in, and the standard deviation is 3 in, *and* height is Normally distributed,

1. What critical values of height bound two-third of all men?
   $70 \text{ inches } \pm 3 \text{ inches } \times 0.967 \approx$ 5 ft 7 to 6 ft 1.

2. What critical values of height bound 95% of all men?
   $70 \text{ inches } \pm 3 \text{ inches } \times 1.96 \approx$ 5 ft 4 to 6 ft 4

If the mean male height in the US is 5 ft 10 in, and the standard deviation is 3 in, *and* height is Normally distributed,

1. What critical values of height bound two-third of all men?
   $70 \text{ inches } \pm 3 \text{ inches } \times 0.967 \approx$ 5 ft 7 to 6 ft 1.

2. What critical values of height bound 95% of all men?
   $70 \text{ inches } \pm 3 \text{ inches } \times 1.96 \approx$ 5 ft 4 to 6 ft 4

3. What critical values of height bound 99% of all men?
   $70 \text{ inches } \pm 3 \text{ inches } \times 2.576 \approx$ 5 ft 2 in to 6 ft 6 in

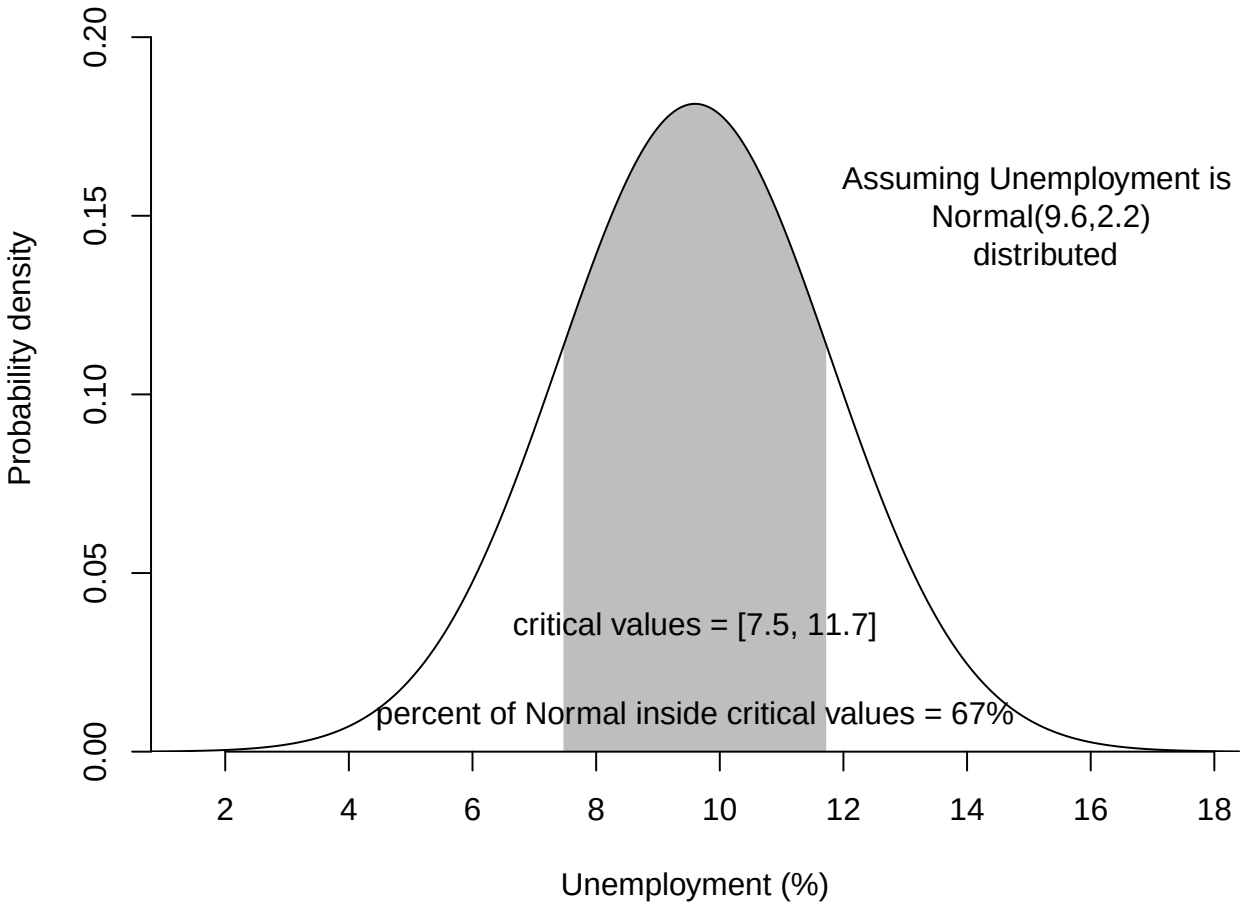
If the mean male height in the US is 5 ft 10 in, and the standard deviation is 3 in, *and* height is Normally distributed,

1. What critical values of height bound two-third of all men?
   $70 \text{ inches } \pm 3 \text{ inches } \times 0.967 \approx$ 5 ft 7 to 6 ft 1.

2. What critical values of height bound 95% of all men?
   $70 \text{ inches } \pm 3 \text{ inches } \times 1.96 \approx$ 5 ft 4 to 6 ft 4

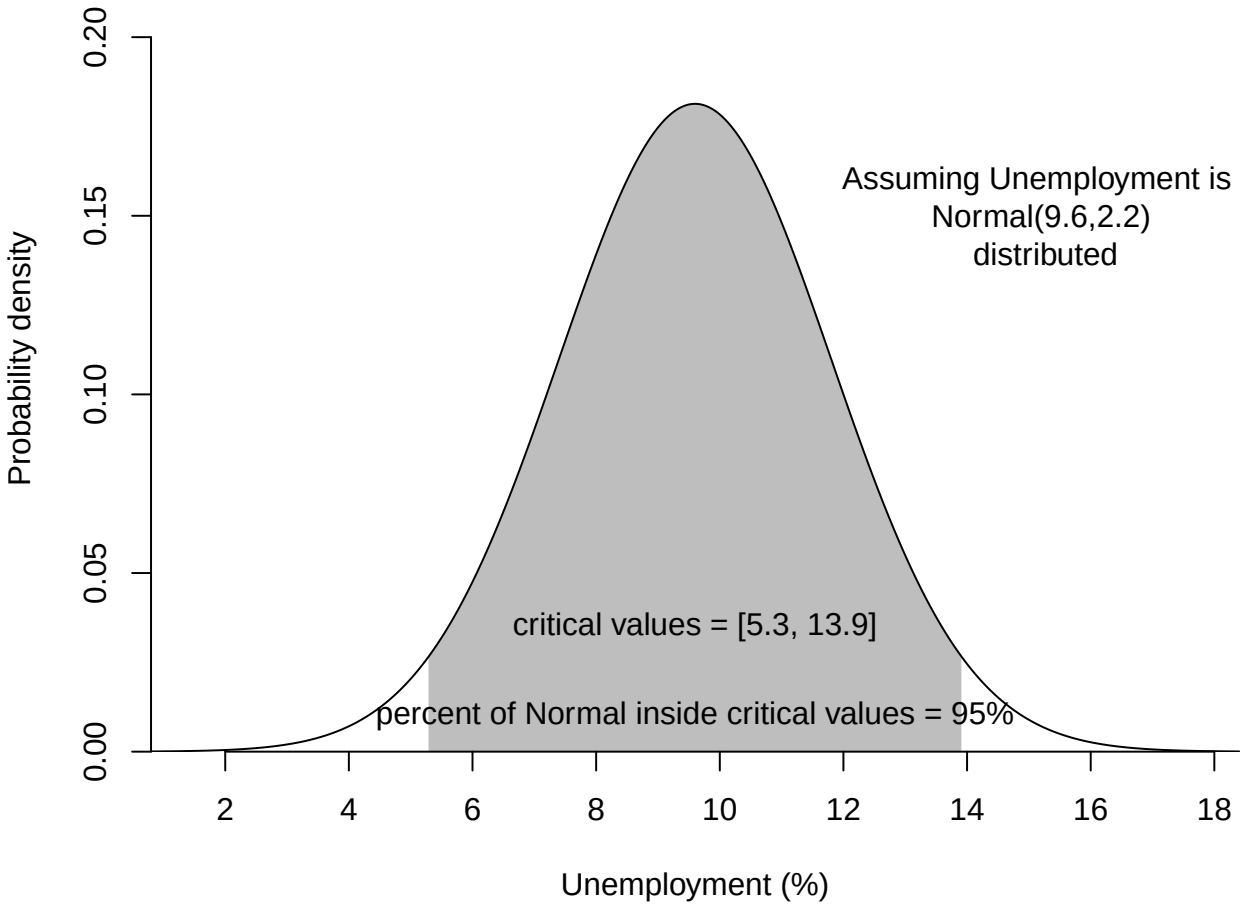
3. What critical values of height bound 99% of all men?
   $70 \text{ inches } \pm 3 \text{ inches } \times 2.576 \approx$ 5 ft 2 in to 6 ft 6 in

Warning! These statements hold only for variables that really are Normal. *Not* for all data.



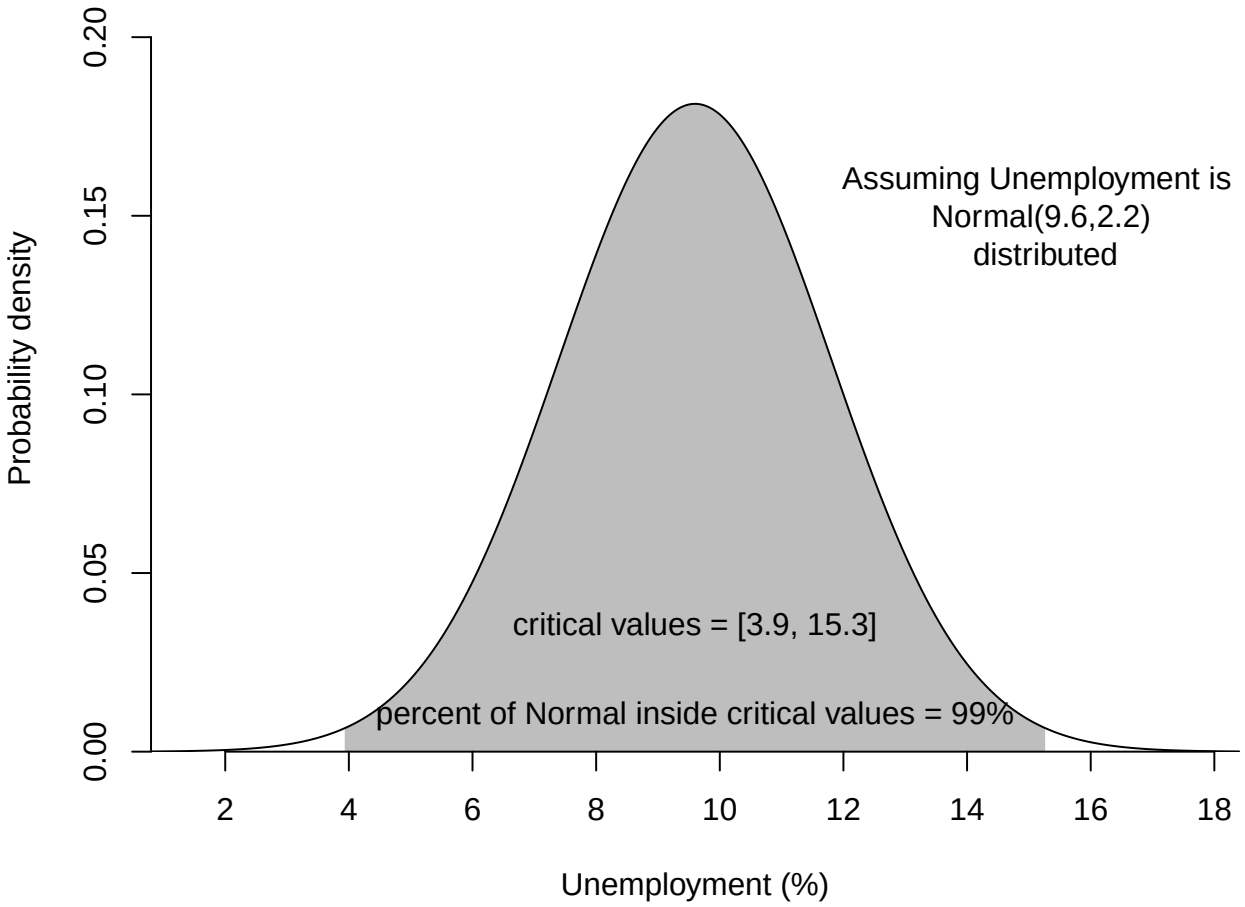
We can apply the same logic to the unemployment example

At left is the 67% interval



We can apply the same logic to the unemployment example

At left is the 95% interval



We can apply the same logic to the unemployment example

At left is the 99% interval