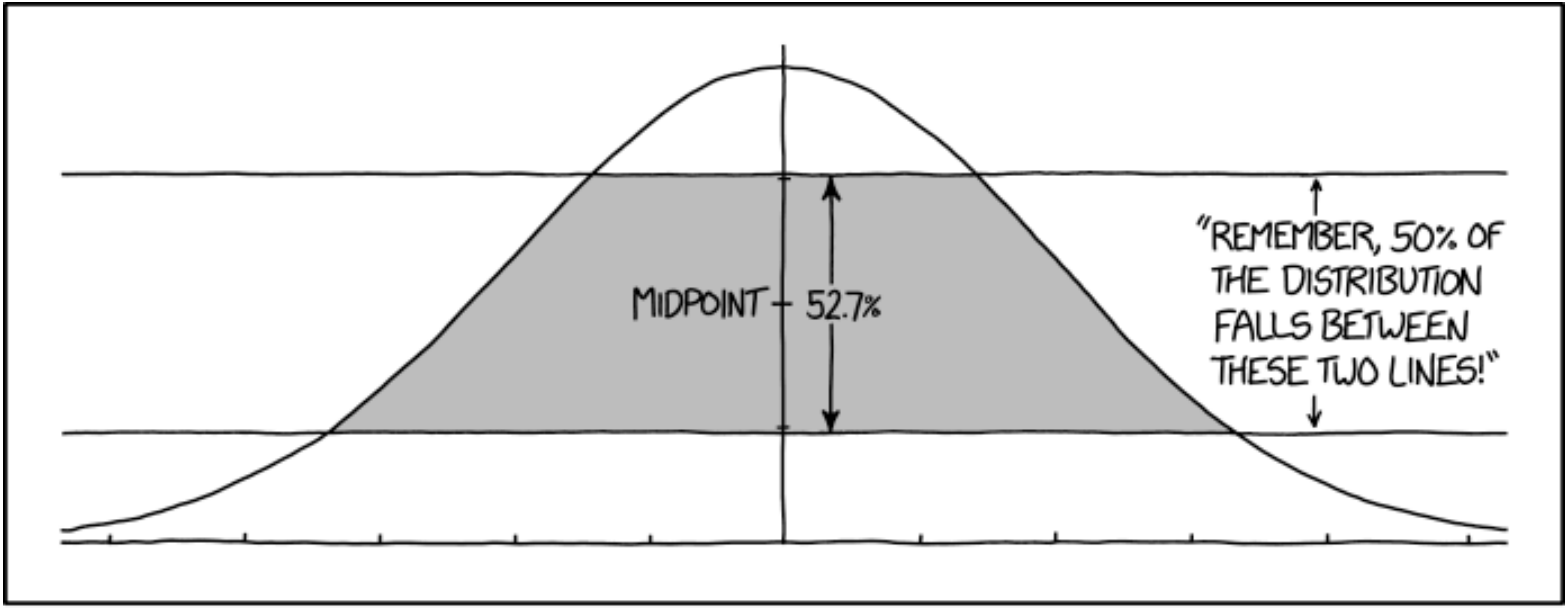
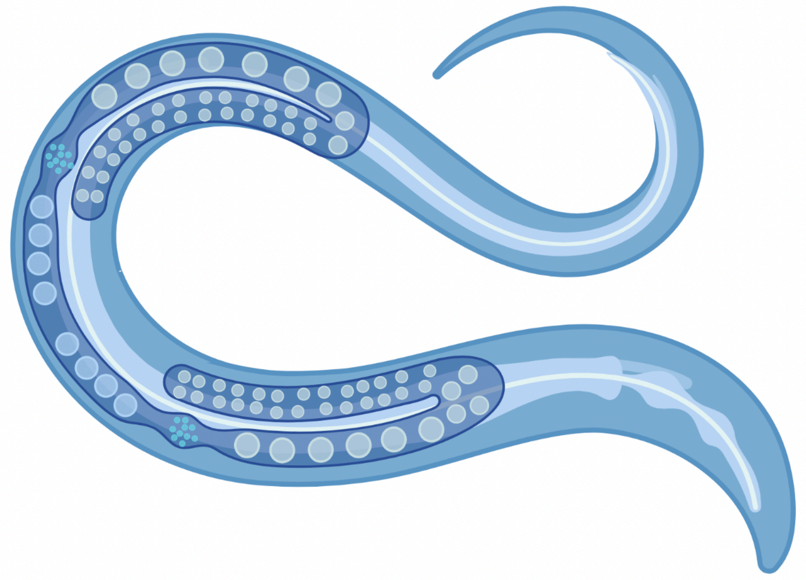
Lecture 04

9.30.21



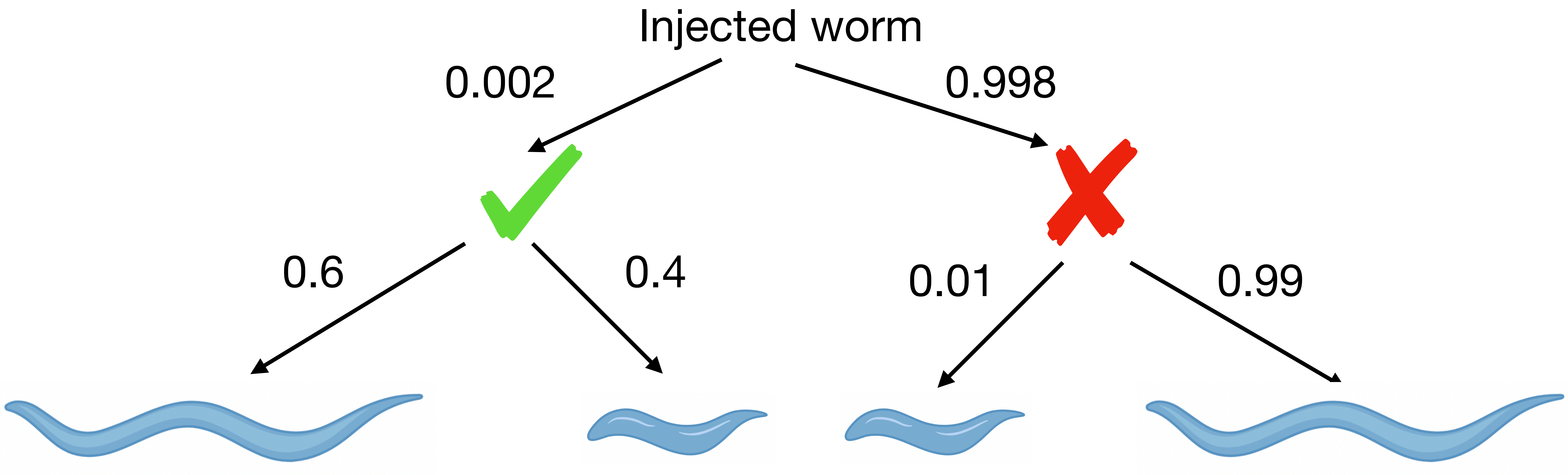
HOW TO ANNOY A STATISTICIAN



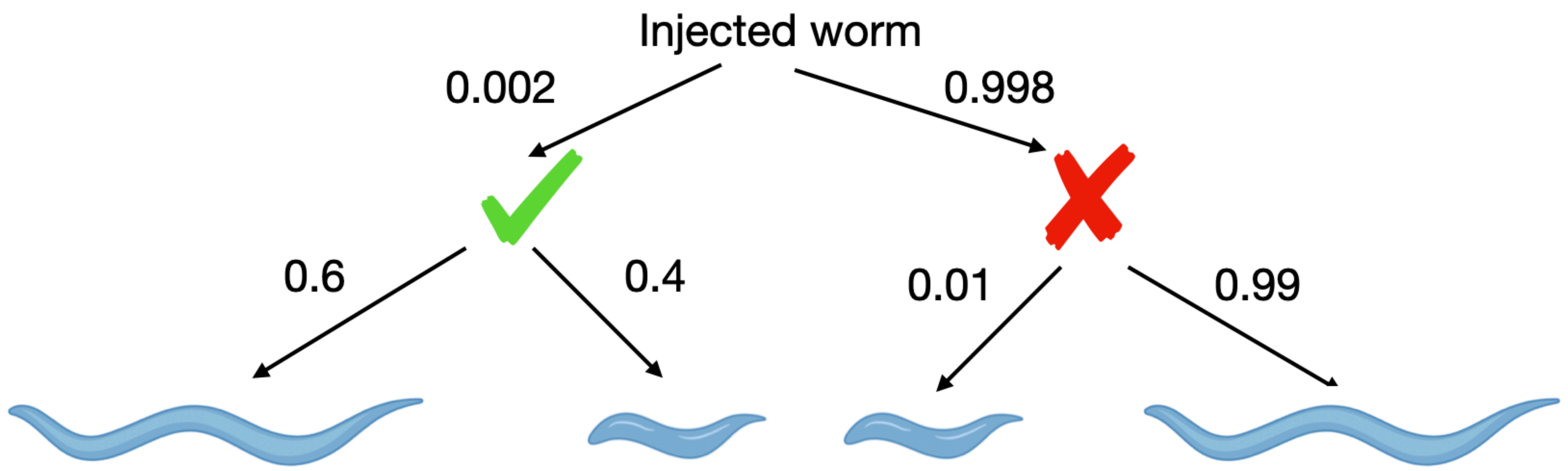
Refresher Quiz

When performing CRISPR in *C. elegans*, you must inject the adult worm with sgRNA and screen its 300 progeny for potential edits. Let's say the probability of getting an edit is 0.002. This would mean you must genotype ~500 worms to see one edit. In many labs, it is common to do a co-injection with a visible marker (i.e. *dpy*). Say the probability of seeing *dpy* in an edited worm is 0.4 and the probability of seeing *dpy* in a non-edited worm is 0.01. If you have a *dpy* worm, what is the probability that that worm is also edited?

When performing CRISPR in *C. elegans*, you must inject the adult worm with sgRNA and screen its 300 progeny for potential edits. Let's say the probability of getting an edit is 0.002. This would mean you must genotype ~500 worms to see one edit. In many labs, it is common to do a co-injection with a visible marker (i.e. *dpy*). Say the probability of seeing *dpy* in an edited worm is 0.4 and the probability of seeing *dpy* in a non-edited worm is 0.01. **If you have a *dpy* worm, what is the probability that that worm is also edited?**

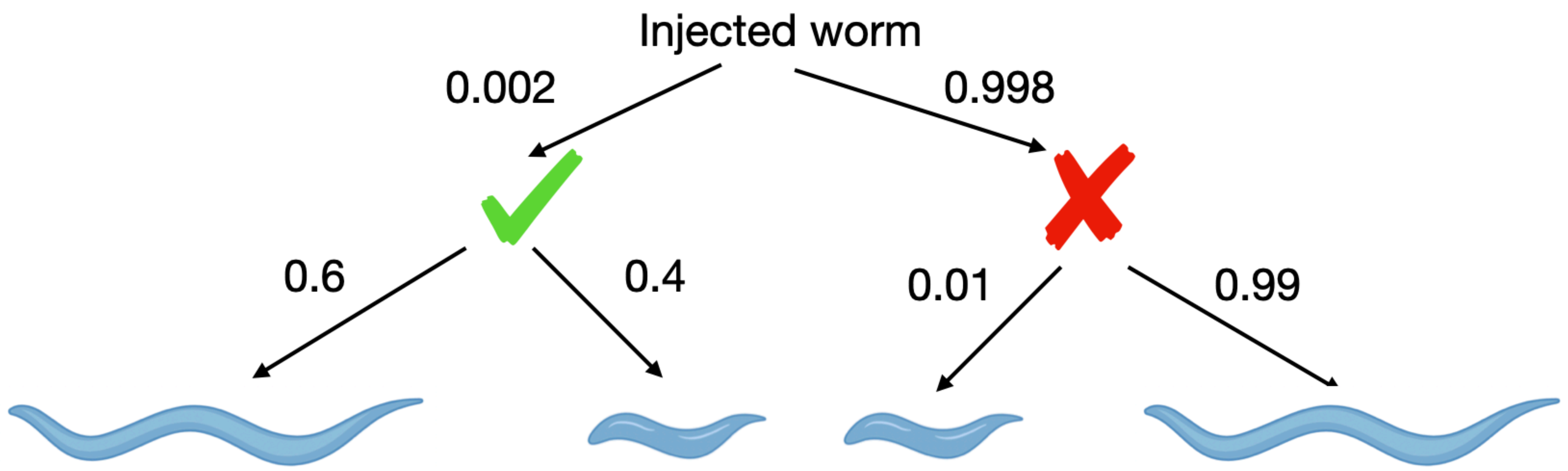


When performing CRISPR in *C. elegans*, you must inject the adult worm with sgRNA and screen its 300 progeny for potential edits. Let's say the probability of getting an edit is 0.002. This would mean you must genotype ~500 worms to see one edit. In many labs, it is common to do a co-injection with a visible marker (i.e. *dpy*). Say the probability of seeing *dpy* in an edited worm is 0.4 and the probability of seeing *dpy* in a non-edited worm is 0.01. If you have a *dpy* worm, what is the probability that that worm is also edited?



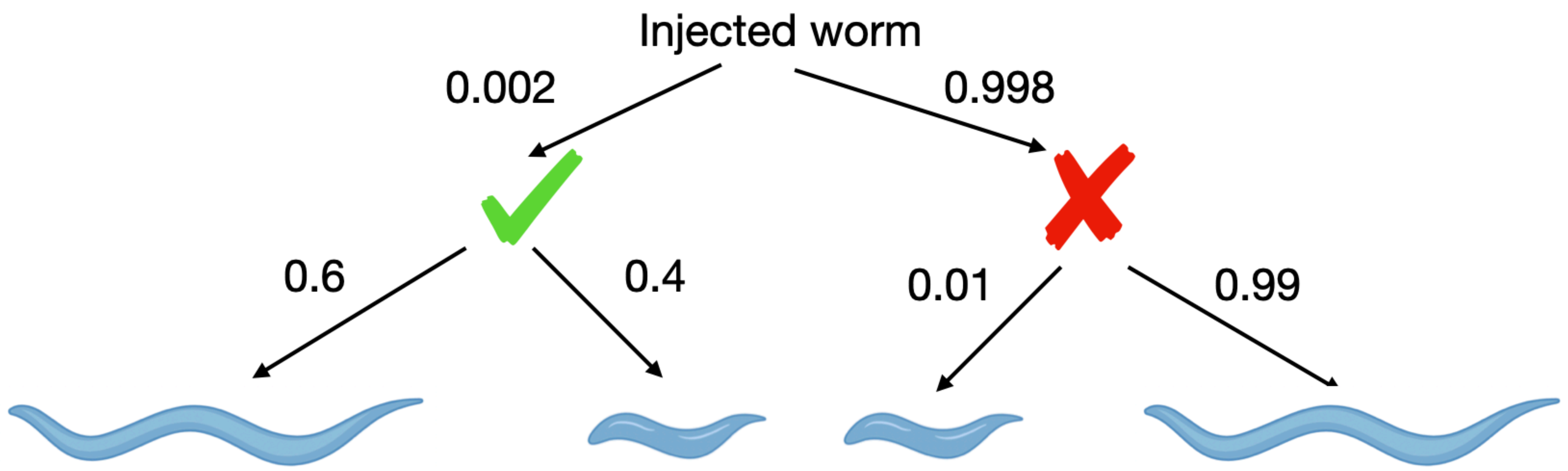
$$P(edit \mid dpy) = \frac{P(dpy \mid edit)P(edit)}{P(dpy)}$$

When performing CRISPR in *C. elegans*, you must inject the adult worm with sgRNA and screen its 300 progeny for potential edits. Let's say the probability of getting an edit is 0.002. This would mean you must genotype ~500 worms to see one edit. In many labs, it is common to do a co-injection with a visible marker (i.e. *dpy*). Say the probability of seeing *dpy* in an edited worm is 0.4 and the probability of seeing *dpy* in a non-edited worm is 0.01. If you have a *dpy* worm, what is the probability that that worm is also edited?



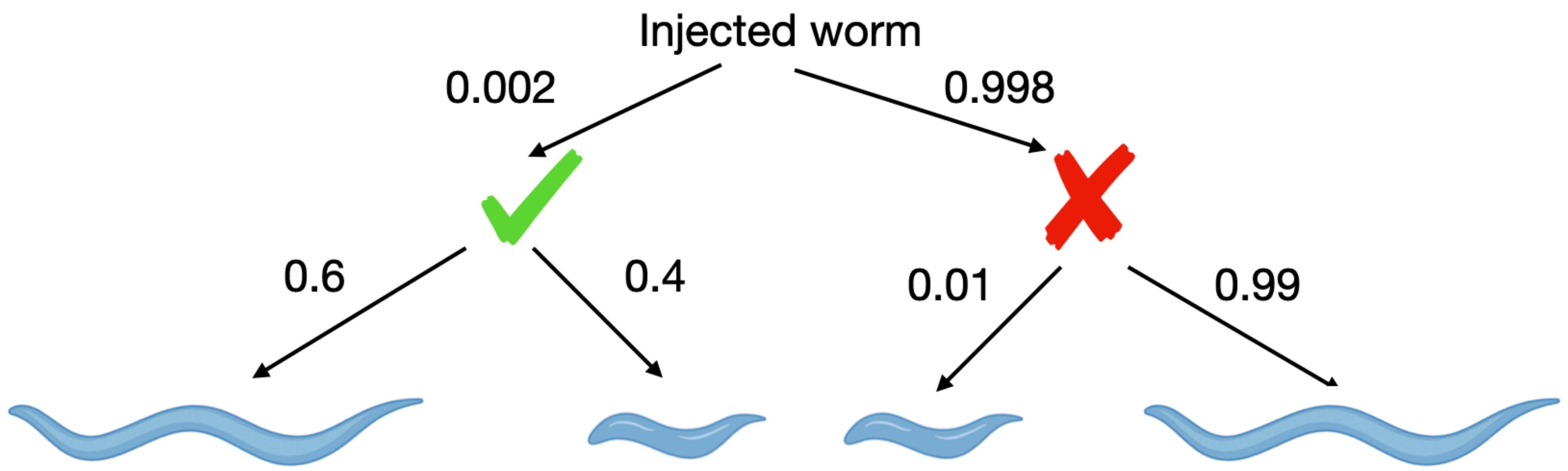
$$P(edit \mid dpy) = \frac{(0.4)P(edit)}{P(dpy)}$$

When performing CRISPR in *C. elegans*, you must inject the adult worm with sgRNA and screen its 300 progeny for potential edits. Let's say the probability of getting an edit is 0.002. This would mean you must genotype ~500 worms to see one edit. In many labs, it is common to do a co-injection with a visible marker (i.e. *dpy*). Say the probability of seeing *dpy* in an edited worm is 0.4 and the probability of seeing *dpy* in a non-edited worm is 0.01. If you have a *dpy* worm, what is the probability that that worm is also edited?



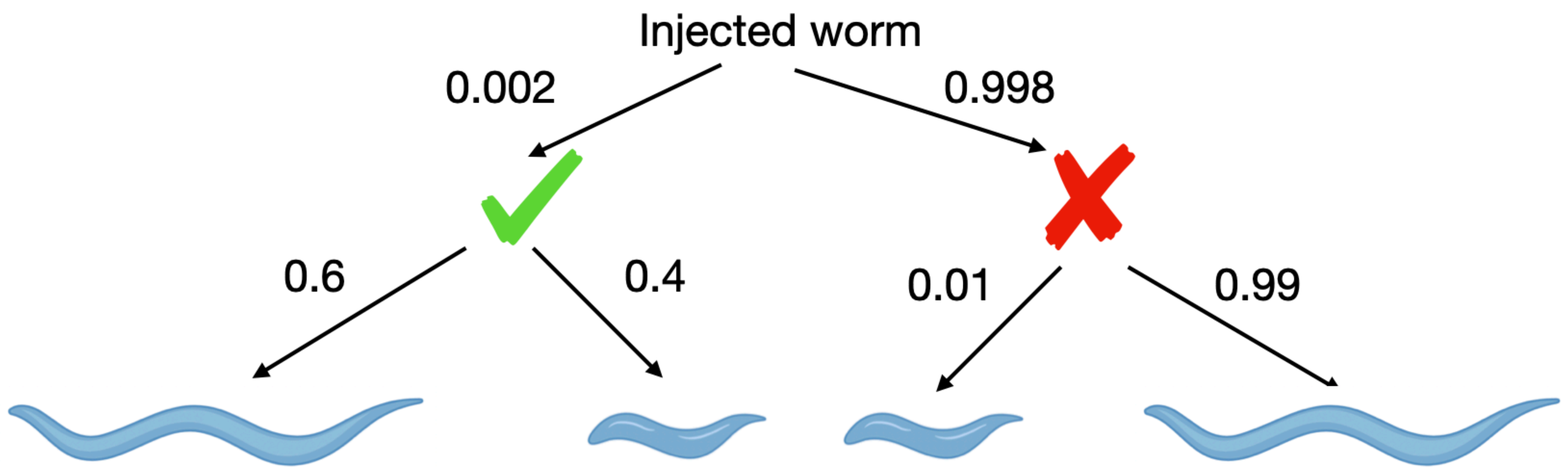
$$P(edit\,|\,dpy) = \frac{(0.4)(0.002)}{P(dpy)}$$

When performing CRISPR in *C. elegans*, you must inject the adult worm with sgRNA and screen its 300 progeny for potential edits. Let's say the probability of getting an edit is 0.002. This would mean you must genotype ~500 worms to see one edit. In many labs, it is common to do a co-injection with a visible marker (i.e. *dpy*). Say the probability of seeing *dpy* in an edited worm is 0.4 and the probability of seeing *dpy* in a non-edited worm is 0.01. If you have a *dpy* worm, what is the probability that that worm is also edited?



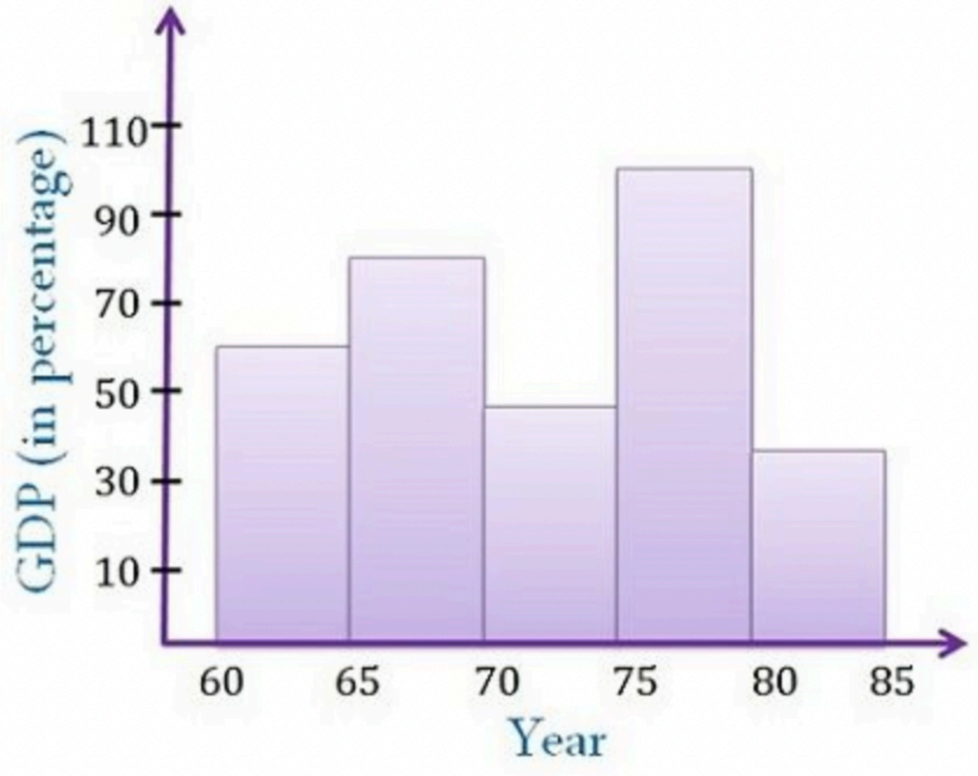
$$P(edit \mid dpy) = \frac{(0.4)(0.002)}{(0.002)(0.4) + (0.998)(0.01)}$$

When performing CRISPR in *C. elegans*, you must inject the adult worm with sgRNA and screen its 300 progeny for potential edits. Let's say the probability of getting an edit is 0.002. This would mean you must genotype ~500 worms to see one edit. In many labs, it is common to do a co-injection with a visible marker (i.e. *dpy*). Say the probability of seeing *dpy* in an edited worm is 0.4 and the probability of seeing *dpy* in a non-edited worm is 0.01. If you have a *dpy* worm, what is the probability that that worm is also edited?



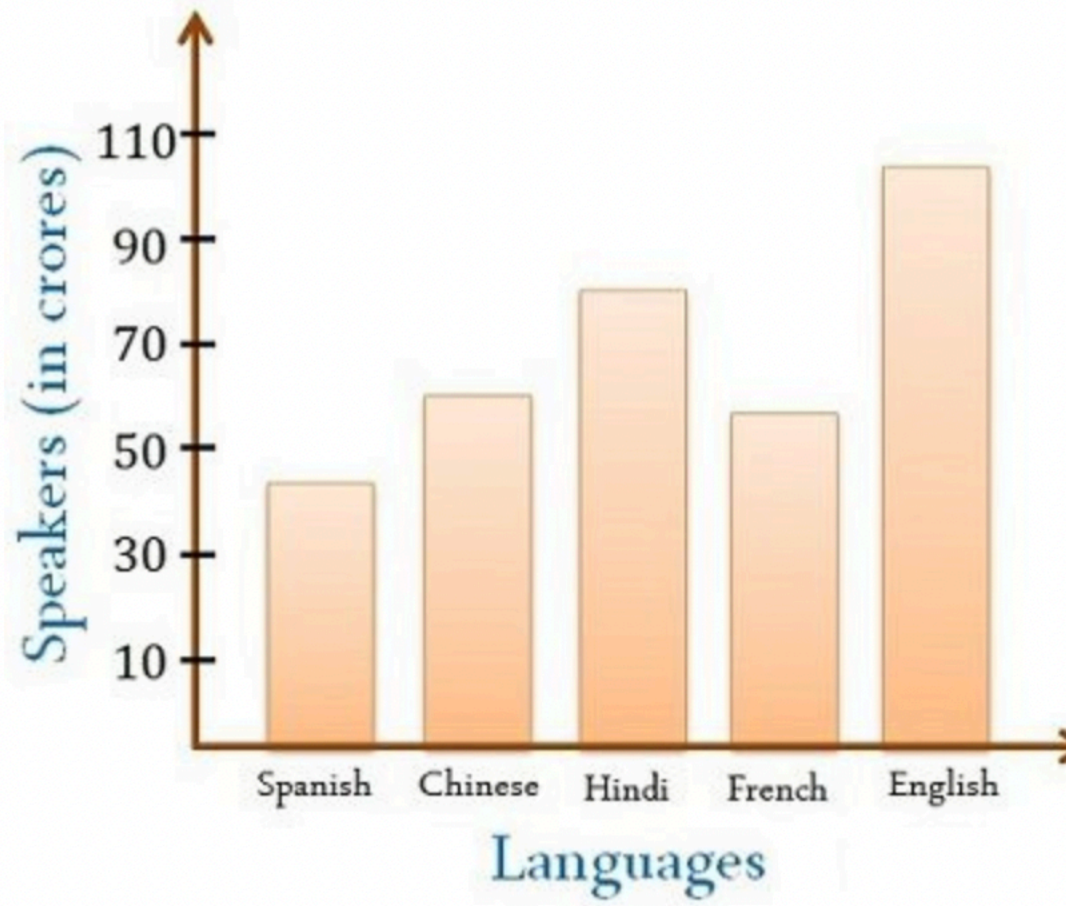
$$P(edit\,|\,dpy) = \frac{0.0008}{0.01078}$$

$$= 0.074$$



Histogram

BASIS FOR COMPARISON	HISTOGRAM	BAR GRAPH
Meaning	Histogram refers to a graphical representation, that displays data by way of bars to show the frequency of numerical data.	Bar graph is a pictorial representation of data that uses bars to compare different categories of data.
Indicates	Distribution of non-discrete variables	Comparison of discrete variables
Presents	Quantitative data	Categorical data
Spaces	Bars touch each other, hence there are no spaces between bars	Bars do not touch each other, hence there are spaces between bars.
Elements	Elements are grouped together, so that they are considered as ranges.	Elements are taken as individual entities.
Can bars be reordered?	No	Yes



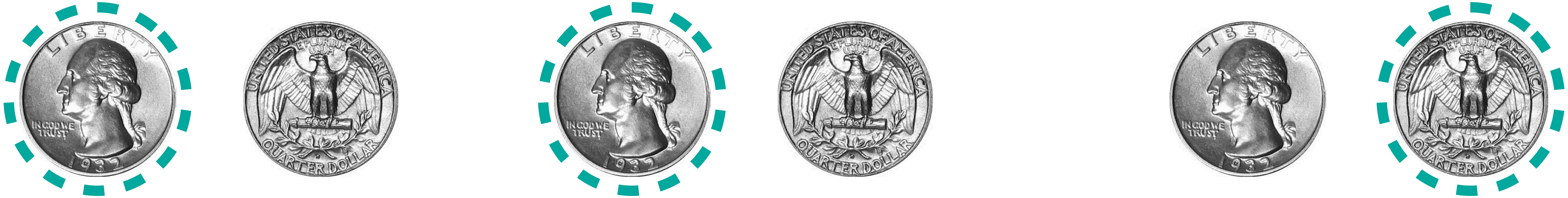
Bar Graph

Identifying binomial random variables

You flip a fair coin 3 times, what is the probability of getting “heads” 2 times?

Identifying binomial random variables

You flip a fair coin 3 times, what is the probability of getting “heads” 2 times?



Coin #1 Coin #2 Coin #3

0.5 x 0.5 x 0.5

Identifying binomial random variables

You flip a fair coin 3 times, what is the probability of getting “heads” 2 times?

Coin #1

Coin #2

Coin #3

0.5 x 0.5 x 0.5

Identifying binomial random variables

You flip a fair coin 3 times, what is the probability of getting "heads" 2 times?

Coin #1

Coin #2

Coin #3

0.5 x 0.5 x 0.5

Identifying binomial random variables

You flip a fair coin 3 times, what is the probability of getting “heads” 2 times?

Coin #1

Coin #2

Coin #3

(0.5 x 0.5 x 0.5) + (0.5 x 0.5 x 0.5) + (0.5 x 0.5 x 0.5) = 0.375

Identifying binomial random variables

You flip a fair coin 100 times, what is the probability of getting "heads" 70 times?

"Binomial random variable"

Identifying binomial random variables

- **B**inary outcomes - *there are two possible outcomes for each trial*
- **I**ndependent trials - *the outcomes of each trial are independent of each other*
- **n** is fixed - *the number of trials, n, is fixed in advance*
- **S**ame value of p - *the probability of success on a single trial is the same for all trials*

B - I - n - S

Binary - Independent - n - Same p

Identifying binomial random variables

You flip a fair coin 100 times, what is the probability of getting "heads" 70 times?

Binary ← Success or Failure (Heads or Tails)

Each toss is independent → **Independent**

n ← Flipping 100 times

Always 50% chance heads → **Same p**

B - I - n - S

Binary - Independent - n - Same p

Identifying binomial random variables

You randomly draw seven cards from a standard deck without replacement. What is the probability of getting 2 aces?

Success or Failure
(Ace or Not ace)

Binary

If you draw 1 ace your chance of another ace changes!

Independent

n

Draw 7 cards

Same p

✓ ✗ ✓ ✗

B - I - n - S

Binary - Independent - n - Same p

Identifying binomial random variables

A certain drug treatment cures 90% of cases of hookworm in children. Suppose that 20 children suffering from hookworm are to be treated, find the probability than exactly 18 will be cured.

Binary ← Success or Failure (Cured or not cured)

Each child is independent of another child → **Independent**

n ← Treat 20 children

90% → **Same p**

B - I - n - S

Binary - Independent - n - Same p

Identifying binomial random variables

Suppose we draw a random sample of five individuals from a population of cats where 37% are mutant. What is the probability that we draw a sample of two mutants?

		Binary	← Success or Failure (mutant or not mutant)
Yes	→	**Independent**	
		n	← 5 cats
0.37	→	**Same p**	

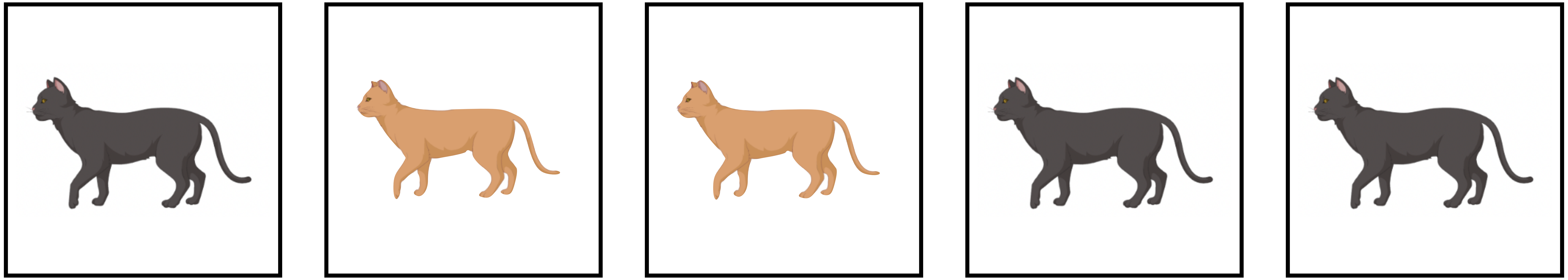
✓ ✓ ✓ ✓

B - I - n - S

Binary - Independent - n - Same p

Probability and the binomial distribution

Suppose we draw a random sample of five individuals from a population of cats where 37% are mutant. What is the probability that we draw a sample of two mutants?

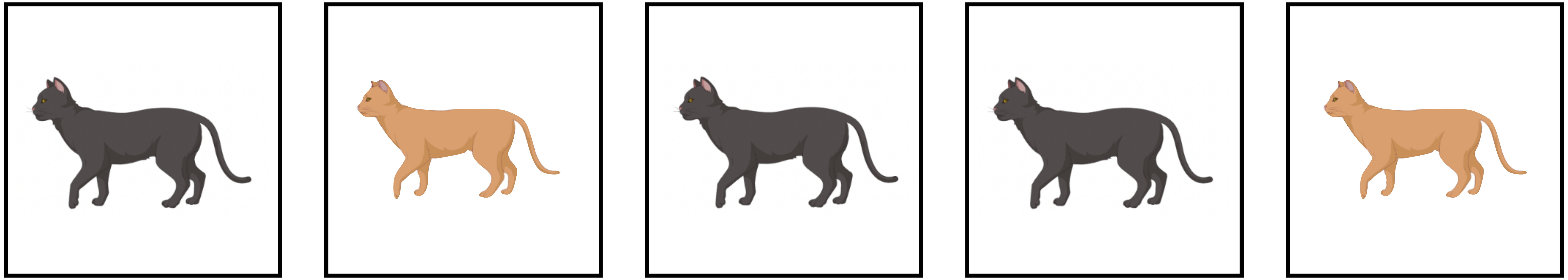


$$(1 - 0.37) \quad \text{x} \quad 0.37 \quad \text{x} \quad 0.37 \quad \text{x} \quad (1 - 0.37) \quad \text{x} \quad (1 - 0.37)$$

$$= (0.37)^2(0.63)^3$$

Probability and the binomial distribution

Suppose we draw a random sample of five individuals from a population of cats where 37% are mutant. What is the probability that we draw a sample of two mutants?

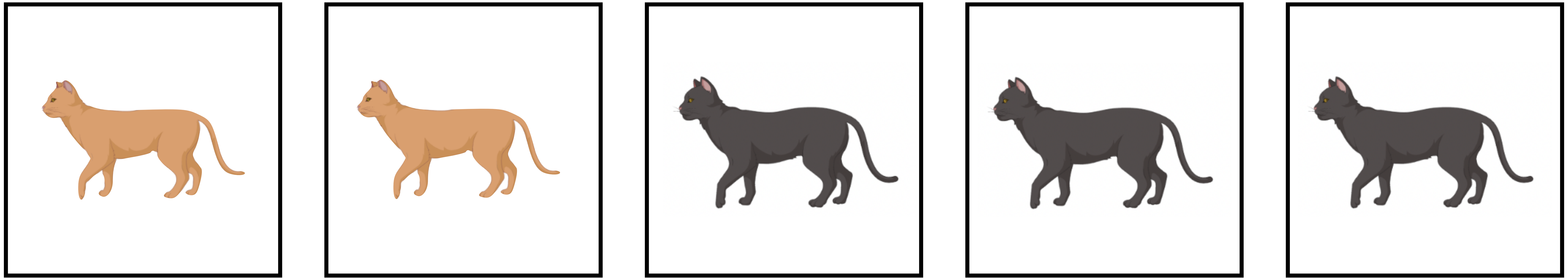


$$(1 - 0.37) \quad \times \quad 0.37 \quad \times \quad (1 - 0.37) \quad \times \quad (1 - 0.37) \quad \times \quad 0.37$$

$$= (0.37)^2(0.63)^3$$

Probability and the binomial distribution

Suppose we draw a random sample of five individuals from a population of cats where 37% are mutant. What is the probability that we draw a sample of two mutants?



$$0.37 \times 0.37 \times (1 - 0.37) \times (1 - 0.37) \times (1 - 0.37)$$

$= (0.37)^2(0.63)^3$ **x (number of possibilities)**

Probability and the binomial distribution

Suppose we draw a random sample of five individuals from a population of cats where 37% are mutant. What is the probability that we draw a sample of two mutants?

$= (0.37)^2(0.63)^3$ **x (number of possibilities)**

```
> choose(5, 2)
```

$_5C_2$

"5 choose 2"

$$\frac{5!}{2!(5-2)!} \qquad \frac{5*4*\cancel{3!}}{2!*\cancel{3!}}$$

$= 10$

Probability and the binomial distribution

Suppose we draw a random sample of five individuals from a population of cats where 37% are mutant. What is the probability that we draw a sample of two mutants?

$$= (0.37)^2(0.63)^3 \ \mathbf{x\ 10} = 0.34$$

```
> choose(5, 2)
```

$${}_5C_2$$

"5 choose 2"

$$\frac{5!}{2!(5-2)!} \qquad \frac{5*4*\cancel{3!}}{2!*\cancel{3!}}$$

$$= 10$$

Binomial distribution formula

Suppose we draw a random sample of five individuals from a population of cats where 37% are mutant. What is the probability that we draw a sample of two mutants?

$= (0.37)^2(0.63)^3$ **x 10** $= 0.34$

$$_{n}C_{j} \cdot p^{j} \cdot (1-p)^{n-j}$$

n = number of trials   j = number of successes   p = probability of success

Probability and the binomial distribution

Suppose we draw a random sample of five individuals from a population of cats where 37% are mutant. What is the probability for <u>each possible outcome</u> (i.e. one mutant, two, three, four, five)?

$$ {}_nC_j \cdot p^j \cdot (1-p)^{n-j} $$

Mutants	Nonmutants	Probability
0	5	
1	4	
2	3	0.34
3	2	
4	1	
5	0	

$$ {}_5C_3 \ 0.37^3 \ (1-0.37)^{5-3} $$

$$ {}_5C_3 \ 0.37^3 \ 0.63^2 $$

$$ 10 \ \ 0.37^3 \ 0.63^2 $$

Probability and the binomial distribution

Suppose we draw a random sample of five individuals from a population of cats where 37% are mutant. What is the probability for <u>each possible outcome</u> (i.e. one mutant, two, three, four, five)?

$$_nC_j \cdot p^j \cdot (1-p)^{n-j}$$

Mutants	Nonmutants	Probability
0	5	
1	4	
2	3	0.34
3	2	0.2
4	1	
5	0	

$$_5C_3 \;\; 0.37^3 \;\; (1-0.37)^{5-3}$$

$$_5C_3 \;\; 0.37^3 \;\; 0.63^2$$

$$10 \;\; 0.37^3 \;\; 0.63^2$$

$$0.20$$

Probability and the binomial distribution

Suppose we draw a random sample of five individuals from a population of cats where 37% are mutant. What is the probability for <u>each possible outcome</u> (i.e. one mutant, two, three, four, five)?

$$ {}_{n}C_{j} \cdot p^{j} \cdot (1 - p)^{n-j} $$

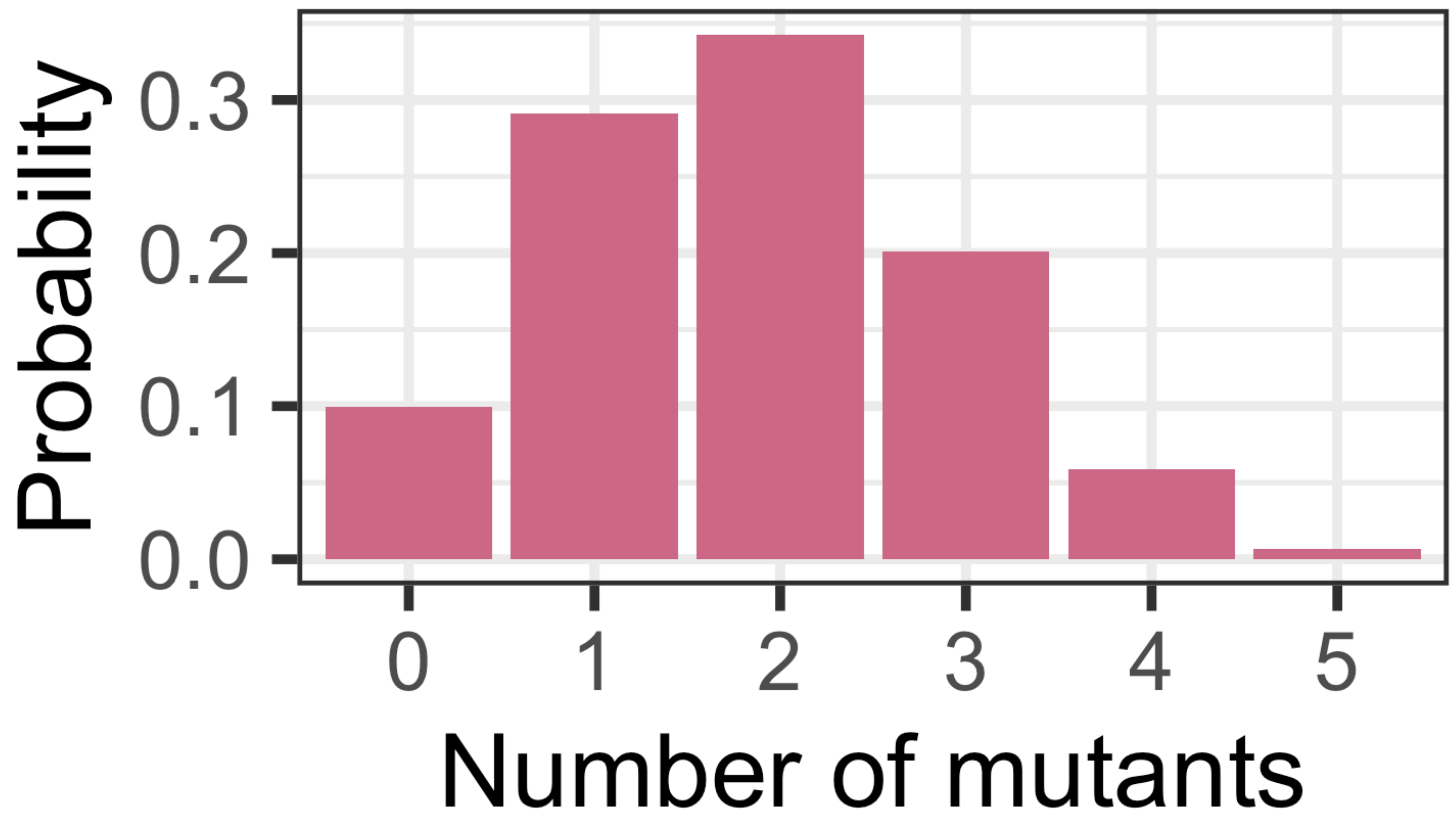
Mutants	Nonmutants	Probability
0	5	0.1
1	4	0.29
2	3	0.34
3	2	0.2
4	1	0.06
5	0	0.01

sum(probabilities) = 1

Plotting a binomial distribution

Suppose we draw a random sample of five individuals from a population of cats where 37% are mutant. What is the probability for <u>each possible outcome</u> (i.e. one mutant, two, three, four, five)?

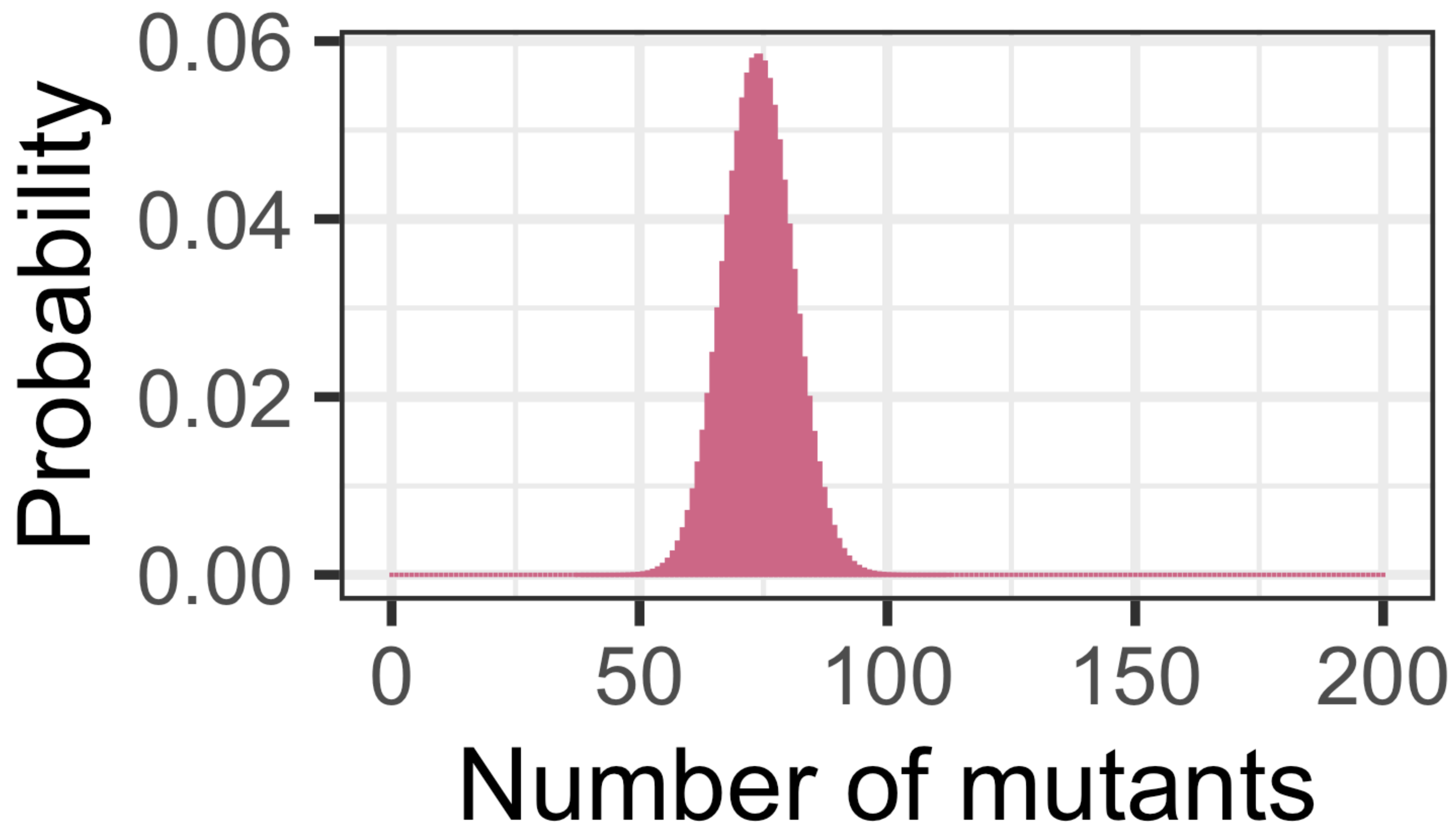
Mutants	Nonmutants	Probability
0	5	0.1
1	4	0.29
2	3	0.34
3	2	0.2
4	1	0.06
5	0	0.01



Plotting a binomial distribution

Suppose we draw a random sample of 200 individuals from a population of cats where 37% are mutant. What is the probability for <u>each possible outcome</u> (i.e. one mutant, two, three, four, 200)?

Mutants	Nonmutants	Probability
0	200	7.30E-41
50	150	9.20E-05
100	100	5.10E-05
125	75	1.50E-03
150	50	7.14E-28
200	0	4.30E-87



Unlimited binomial distribution = normal distribution!

Identifying binomial random variables

The probability of having blood type A is 0.4. There are 50 people in this room, what is the probability that 30 have type A?

Binary ← Success or Failure (A or Not A)

My type doesn't influence your type → **Independent**

n ← 50 people

0.4 → **Same p**

B - I - n - S

Binary - Independent - n - Same p

Identifying binomial random variables

The probability of having blood type A is 0.4. There are 50 people in this room, what is the probability that 30 have type A?

$$ {}_{n}C_{j} \cdot p^{j} \cdot (1-p)^{n-j} $$

$$ {}_{50}C_{30}\ (0.4)^{30}\ (1-0.4)^{50-30} $$

```
> choose(50, 30)
```

$$ {}_{50}C_{30}\ (0.4)^{30}\ (0.6)^{20} = 0.00198 $$

$$ = 0.198\,\% $$

Identifying binomial random variables

The probability of having blood type A is 0.4. There are 50 people in this room, what is the probability that 30 have type A?

```
> dbinom(x, size, prob)
```

```
> dbinom(j, n, p)
```

```
> dbinom(30, 50, 0.4)
```

```
> 0.001986621
```

Mean and standard deviation of a binomial

For a binomial random variable, the mean (i.e., the average number of successes) is equal to n*p

Toss a fair coin 10 times, we expect to get 5 heads (0.5 * 10)

Mean and standard deviation of a binomial

The probability of having blood type A is 0.4. There are 50 people in this room, what is the expected number of people with blood type A in this room?

mean = n * p

mean = 50 * 0.4

mean = 20

Mean and standard deviation of a binomial

The probability of having blood type A is 0.4. There are 50 people in this room, what is the expected number of people with blood type A in this room?

mean = n * p

mean = 50 * 0.4

mean = 20

$$sd = \sqrt{(np(1-p))}$$

$$sd = \sqrt{(50 * 0.4 * (1-0.4))}$$

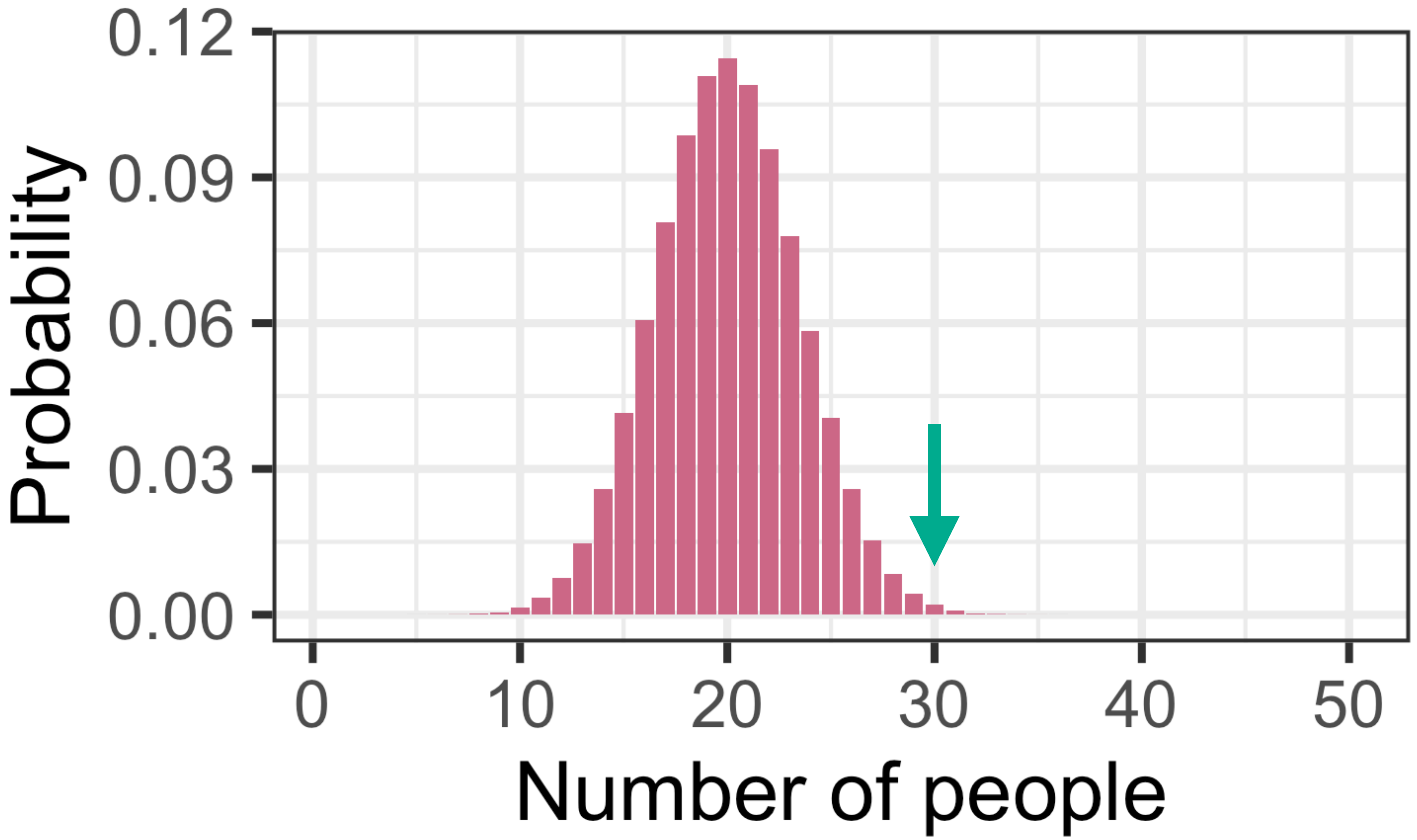
sd = 3.5

Mean and standard deviation of a binomial

The probability of having blood type A is 0.4. There are 50 people in this room, what is the expected number of people with blood type A in this room?

mean = 20

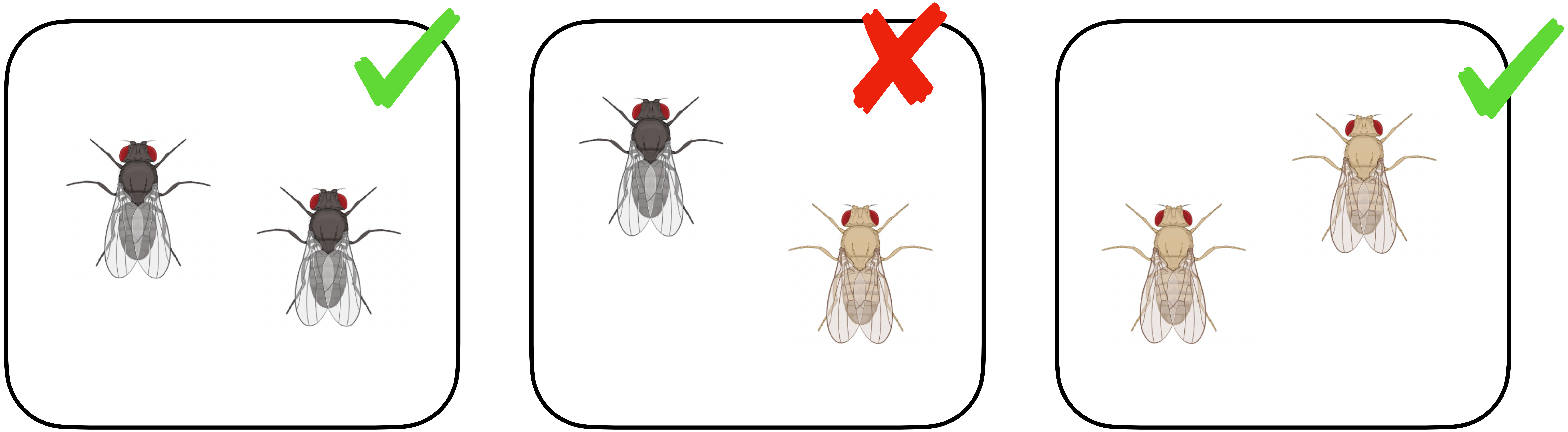
sd = 3.5



Identifying binomial random variables

In a large *Drosophila* population, 30% of the flies are black and 70% are gray. Suppose two flies are randomly chosen, what is the probability that both flies are the same color?

```
> dbinom(j, n, p)
```



```
dbinom(2, 2, 0.3)   +   dbinom(2, 2, 0.7)
```

Consider a family with five children, and suppose that the chance of an individual child catching chickenpox during a given year is 0.10. What is the probability that at least two of the children get the chickenpox?

Chickenpox is contagious and we are looking at a family!! These are not independent.

Binary ← Success or Failure (Disease or no)

Independent

n ← 5 children

Same p

✓ ✗ ✓ ✗

B - I - n - S

Binary - Independent - n - Same p

How could we adjust the question to make it a binomial variable?

Consider a random sample of five children, and suppose that the chance of an individual child catching chickenpox during a given year is 0.10. What is the probability that at least two of the children get the chickenpox?

Success or Failure (Disease or no) → **Binary**

Yes! Random sample → **Independent**

5 children → **n**

0.1 → **Same p**

B - I - n - S

Binary - Independent - n - Same p

How could we adjust the question to make it a binomial variable?

Consider a random sample of five children, and suppose that the chance of an individual child catching chickenpox during a given year is 0.10. What is the probability that at least two of the children get the chickenpox?

Pr{at least 2 children} = Pr{2 children} + Pr{3 children} + Pr{4 children} + Pr{5 children}

`1-pbinom(1, 5, 0.1)`

Pr{at least 2 children} = 1 - (Pr{0 children} + Pr{1 child})

`dbinom(0, 5, 0.1)`

`dbinom(1, 5, 0.1)`

R

Pr{at least 2 children} = 0.08

(Mean = np = 0.5)

A note on R distribution functions

"R" functions

Generate a vector of random variables

```
> rbinom(10, 100, 0.37)
36 32 40 29 37 36 39 31 32 36
```

"P" functions

Returns cumulative probability of a given value

```
> pbinom(5, 10, 0.37)
0.8794974
```

"D" functions

Returns probability of a given value

```
> dbinom(5, 10, 0.37)
0.1734251
```

"Q" functions

Returns value that generates a given prob.

```
> qbinom(0.879, 10, 0.37)
5
```

Fitting a binomial distribution to data

How can you tell if your collected data follows a binomial distribution?

Germinated	Not germinated	Observed frequency
0	5	17
1	4	53
2	3	94
3	2	79
4	1	33
5	0	4

1. Estimate probability of germination

$$\Pr(\text{germ}) = \frac{\text{Total \# germ}}{\text{Total \#}}$$

$$\Pr(\text{germ}) = \frac{\text{Total \# germ}}{(17+53+94+79+33+4) * 5}$$

Fitting a binomial distribution to data

How can you tell if your collected data follows a binomial distribution?

Germinated	Not germinated	Observed frequency
0	5	17
1	4	53
2	3	94
3	2	79
4	1	33
5	0	4

1. Estimate probability of germination

$$Pr(germ) = \frac{\text{Total \# germ}}{\text{Total \#}}$$

$$Pr(germ) = \frac{\text{Total \# germ}}{1400}$$

Fitting a binomial distribution to data

How can you tell if your collected data follows a binomial distribution?

Germinated	Not germinated	Observed frequency
0	5	17
1	4	53
2	3	94
3	2	79
4	1	33
5	0	4

1. Estimate probability of germination

$$\Pr(\text{germ}) = \frac{\text{Total \# germ}}{\text{Total \#}}$$

$$\Pr(\text{germ}) = \frac{0(17) + 1(53) + 2(94) + 3(79) + 4(33) + 5(4)}{1400}$$

Fitting a binomial distribution to data

How can you tell if your collected data follows a binomial distribution?

Germinated	Not germinated	Observed frequency
0	5	17
1	4	53
2	3	94
3	2	79
4	1	33
5	0	4

1. Estimate probability of germination

$$\Pr(\text{germ}) = \frac{\text{Total \# germ}}{\text{Total \#}}$$

$$\Pr(\text{germ}) = \frac{630}{1400} = 0.45$$

Fitting a binomial distribution to data

How can you tell if your collected data follows a binomial distribution?

Germinated	Not germinated	Observed frequency
0	5	17
1	4	53
2	3	94
3	2	79
4	1	33
5	0	4

1. Estimate probability of germination

P(germ) = 0.45

2. Compute binomial probabilities

```
dbinom(0, 5, 0.45)

    = 0.0503
```

Fitting a binomial distribution to data

How can you tell if your collected data follows a binomial distribution?

Germinated	Not germinated	Observed frequency	Probability
0	5	17	0.0503
1	4	53	
2	3	94	
3	2	79	
4	1	33	
5	0	4	

1. Estimate probability of germination

P(germ) = 0.45

2. Compute binomial probabilities

```
dbinom(0, 5, 0.45)

     = 0.0503
```

Fitting a binomial distribution to data

How can you tell if your collected data follows a binomial distribution?

Germinated	Not germinated	Observed frequency	Probability
0	5	17	0.0503
1	4	53	0.205
2	3	94	0.336
3	2	79	0.275
4	1	33	0.112
5	0	4	0.018

1. Estimate probability of germination

P(germ) = 0.45

2. Compute binomial probabilities

```
dbinom(0, 5, 0.45)

    = 0.0503
```

Fitting a binomial distribution to data

How can you tell if your collected data follows a binomial distribution?

Germ.	Not germ.	Observed frequency	Prob	Expected frequency
0	5	17	0.0503	
1	4	53	0.205	
2	3	94	0.336	
3	2	79	0.275	
4	1	33	0.112	
5	0	4	0.018	
		280		

1. Estimate probability of germination

P(germ) = 0.45

2. Compute binomial probabilities

3. Compute expected frequencies

expected = n * p

expected = 280 * 0.0503 = 14.09

Fitting a binomial distribution to data

How can you tell if your collected data follows a binomial distribution?

Germ.	Not germ.	Observed frequency	Prob	Expected frequency
0	5	17	0.0503	14.09
1	4	53	0.205	57.64
2	3	94	0.336	94.33
3	2	79	0.275	77.18
4	1	33	0.112	31.57
5	0	4	0.018	5.16
		280		

1. Estimate probability of germination

P(germ) = 0.45

2. Compute binomial probabilities

3. Compute expected frequencies

expected = n * p

expected = 280 * 0.0503 = 14.09

Fitting a binomial distribution to data

How can you tell if your collected data follows a binomial distribution?

Germ.	Not germ.	Observed frequency	Prob	Expected frequency
0	5	17	0.0503	14.09
1	4	53	0.205	57.64
2	3	94	0.336	94.33
3	2	79	0.275	77.18
4	1	33	0.112	31.57
5	0	4	0.018	5.16

1. Estimate probability of germination

P(germ) = 0.45

2. Compute binomial probabilities

3. Compute expected frequencies

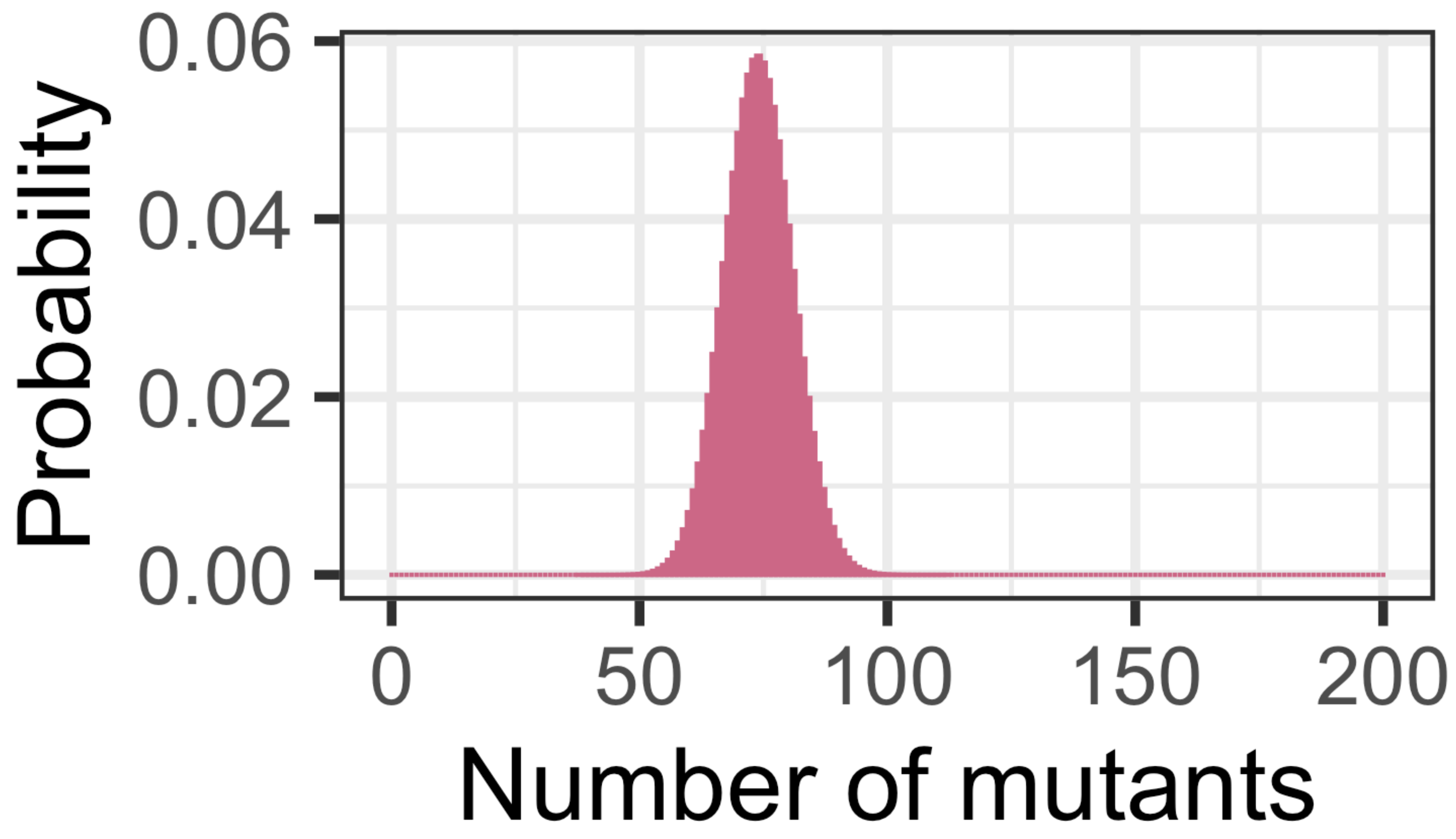
expected = n * p

4. Compare observed and expected frequencies ✓

The normal distribution

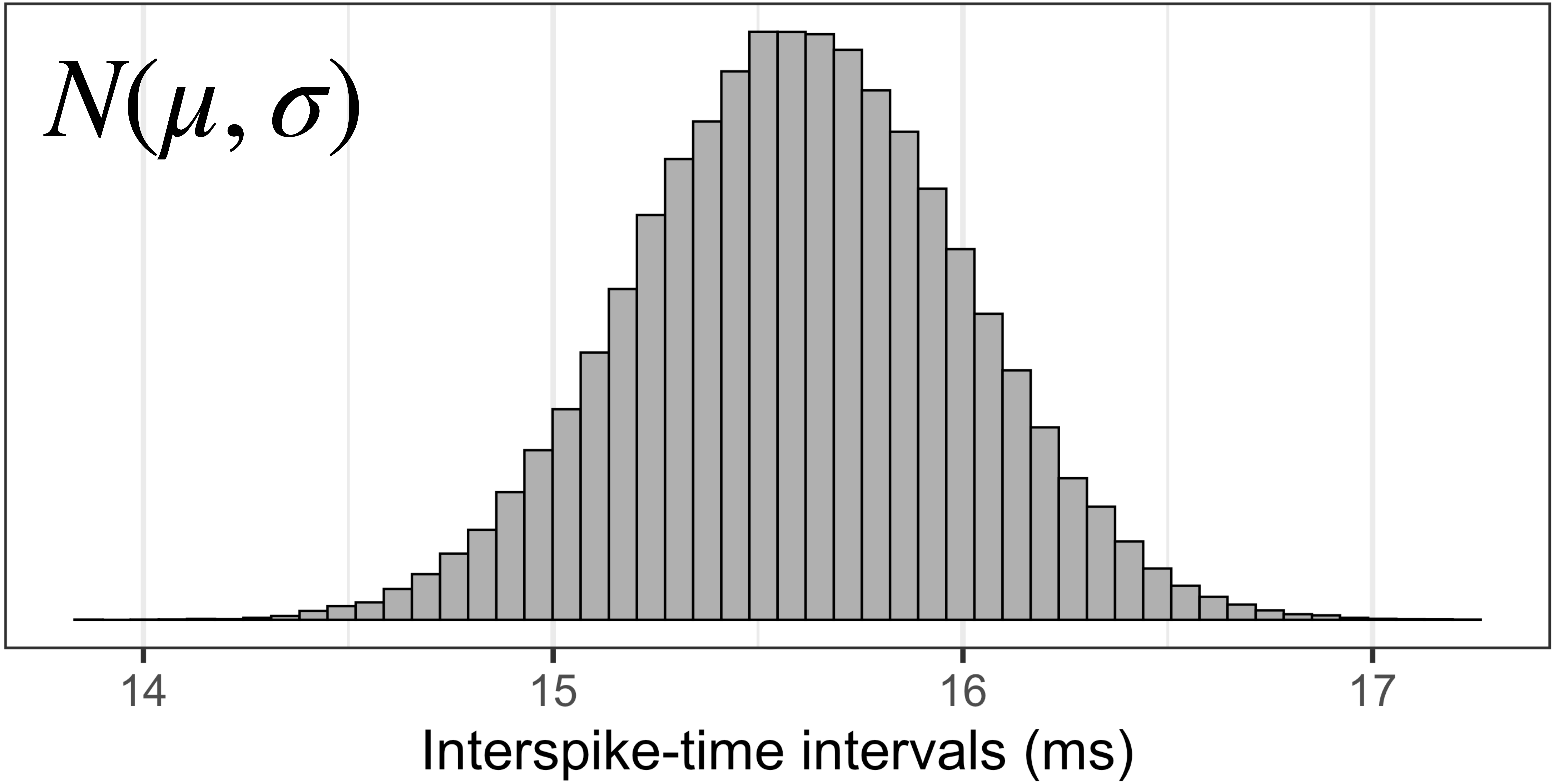
Suppose we draw a random sample of 200 individuals from a population of cats where 37% are mutant. What is the probability for <u>each possible outcome</u> (i.e. one mutant, two, three, four, 200)?

Mutants	Nonmutants	Probability
0	200	7.30E-41
50	150	9.20E-05
100	100	5.10E-05
125	75	1.50E-03
150	50	7.14E-28
200	0	4.30E-87

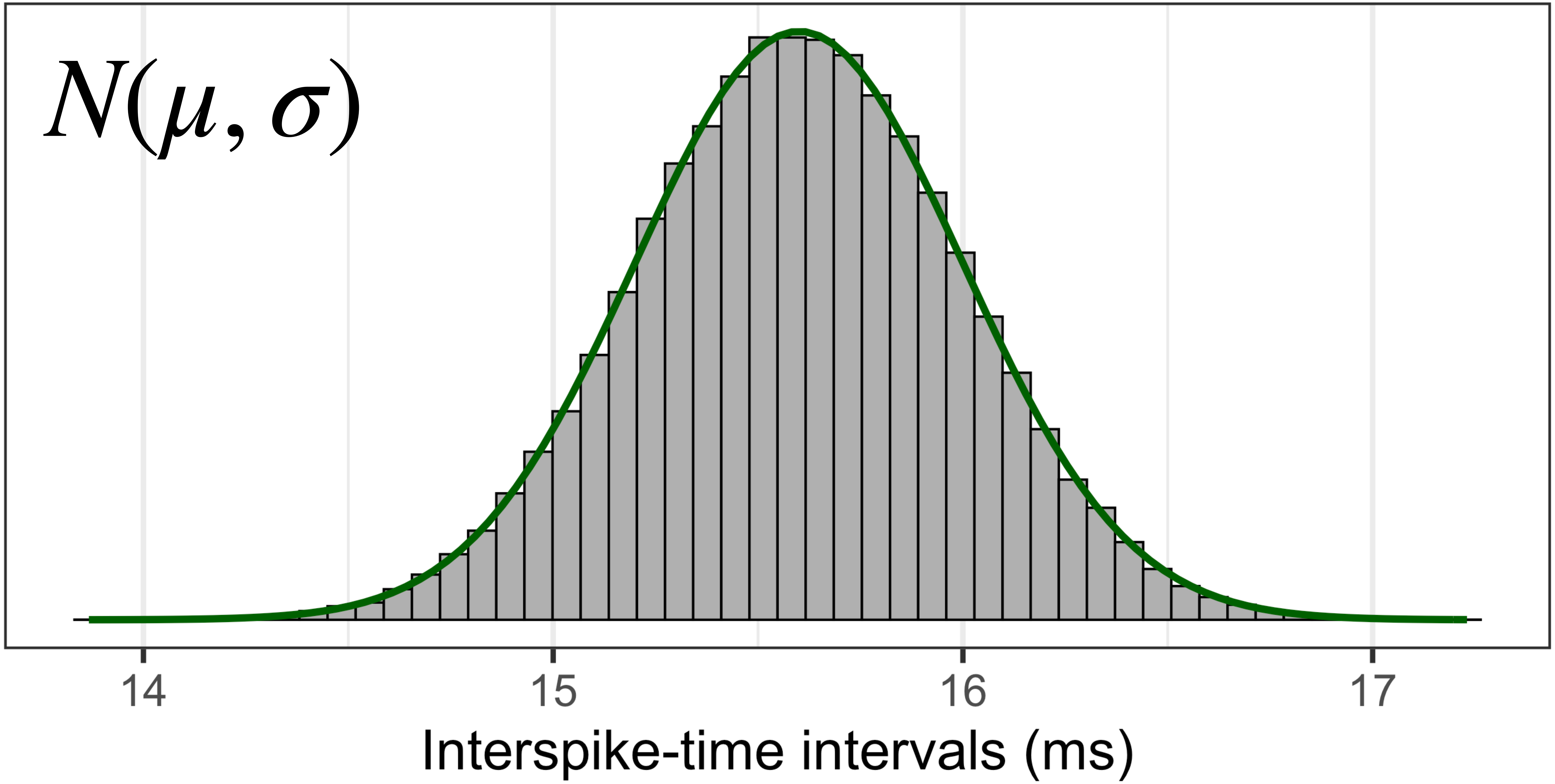


Unlimited binomial distribution = normal distribution!

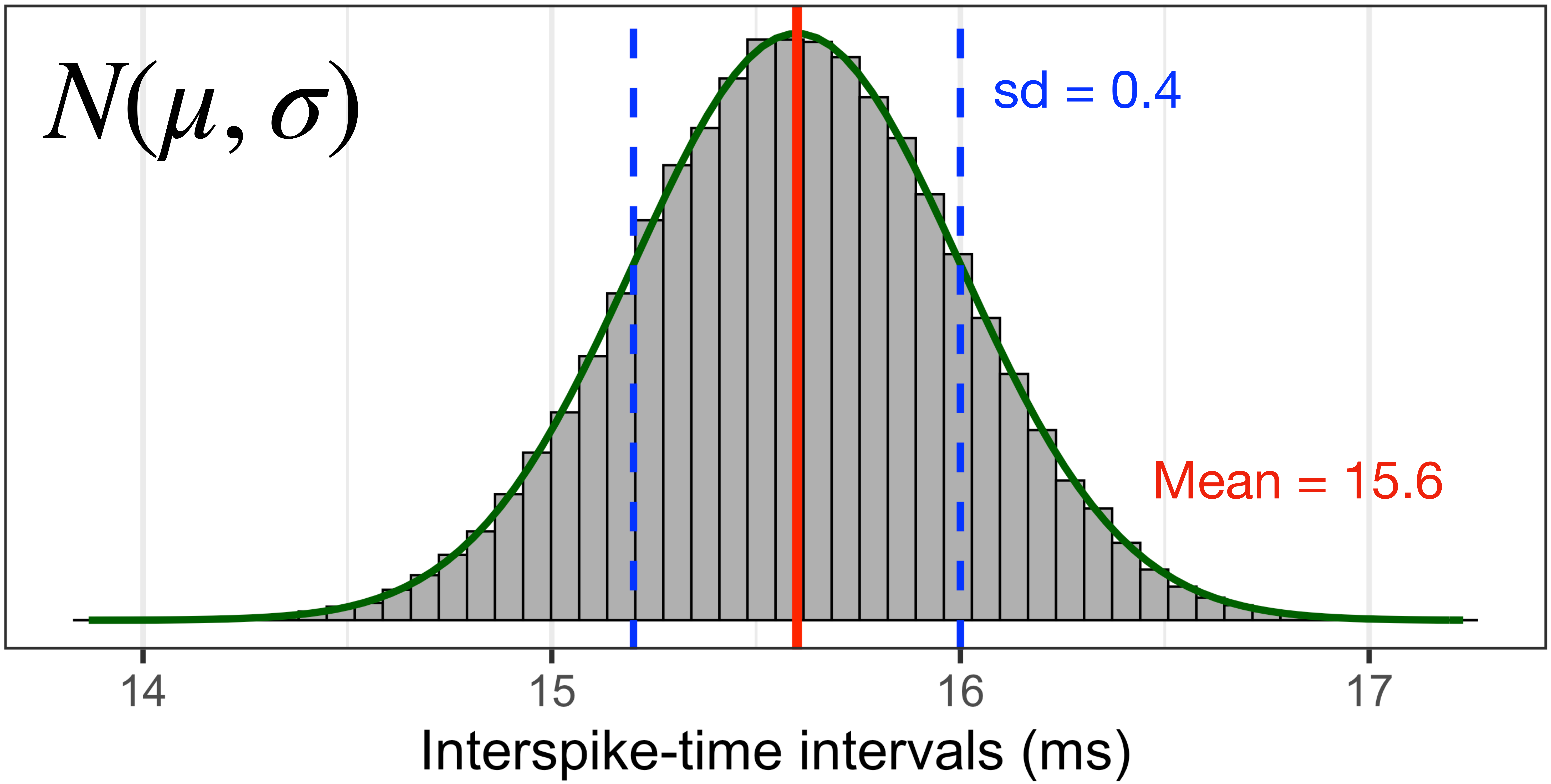
The normal distribution



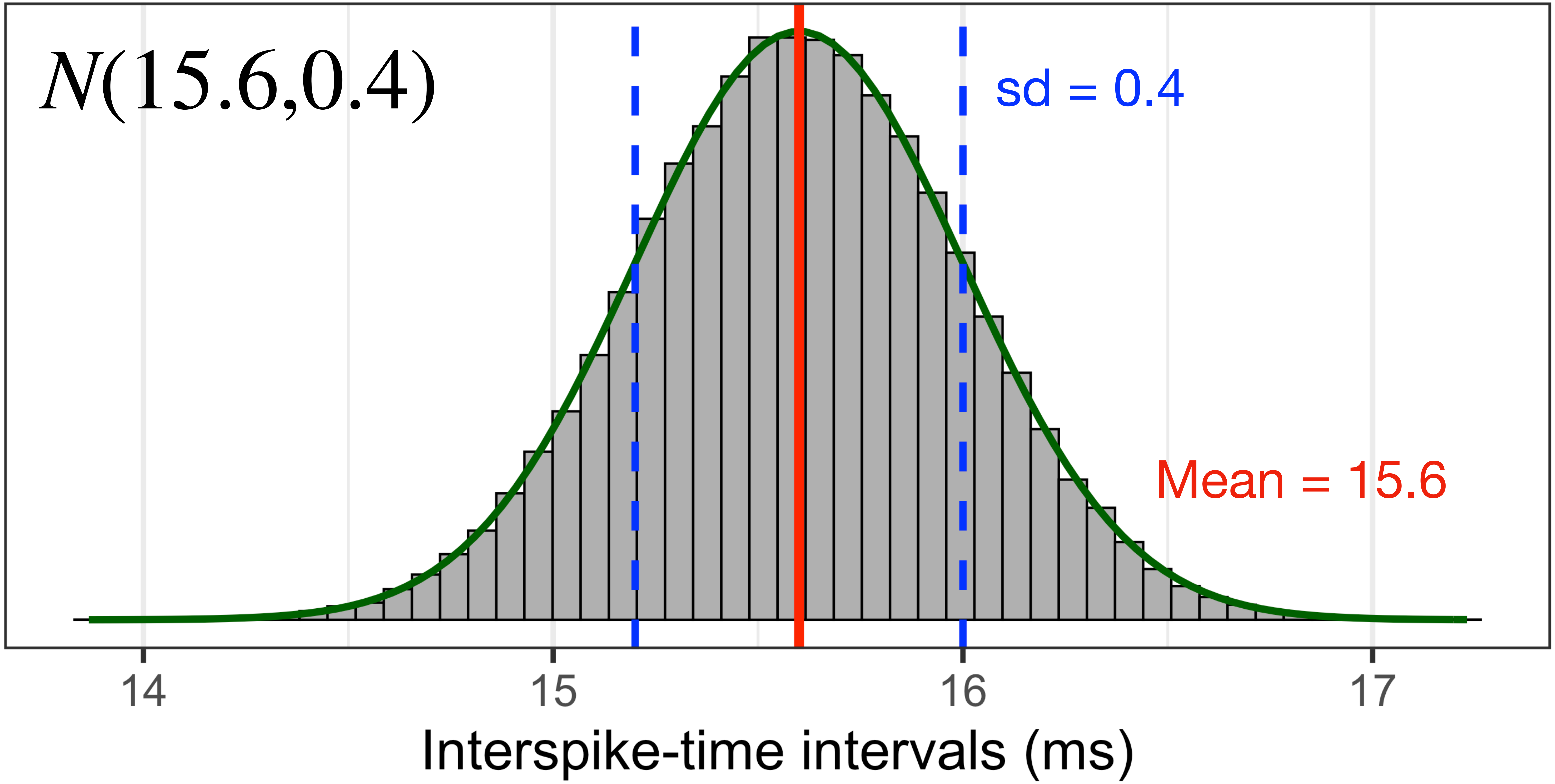
The normal distribution



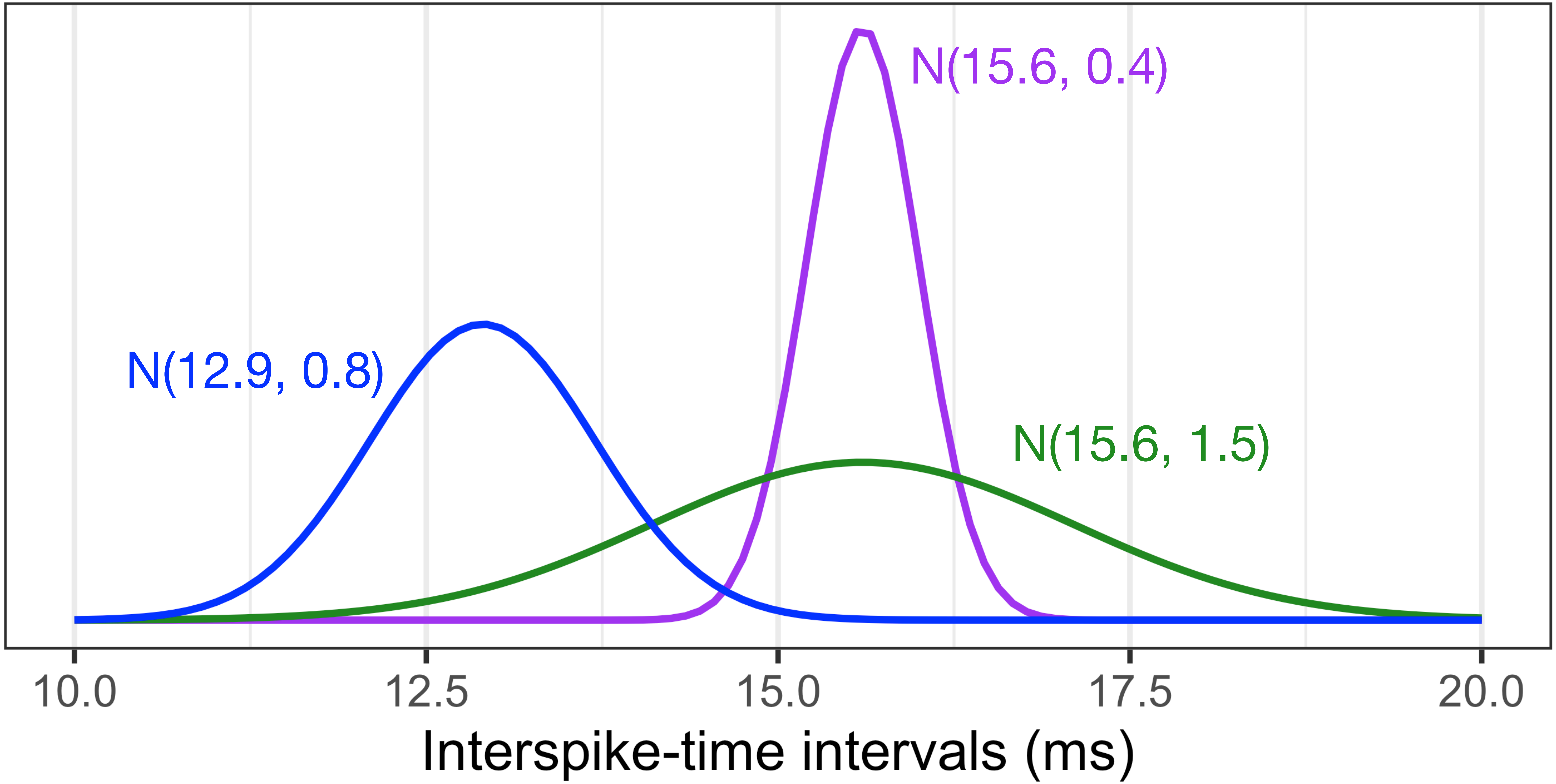
The normal distribution



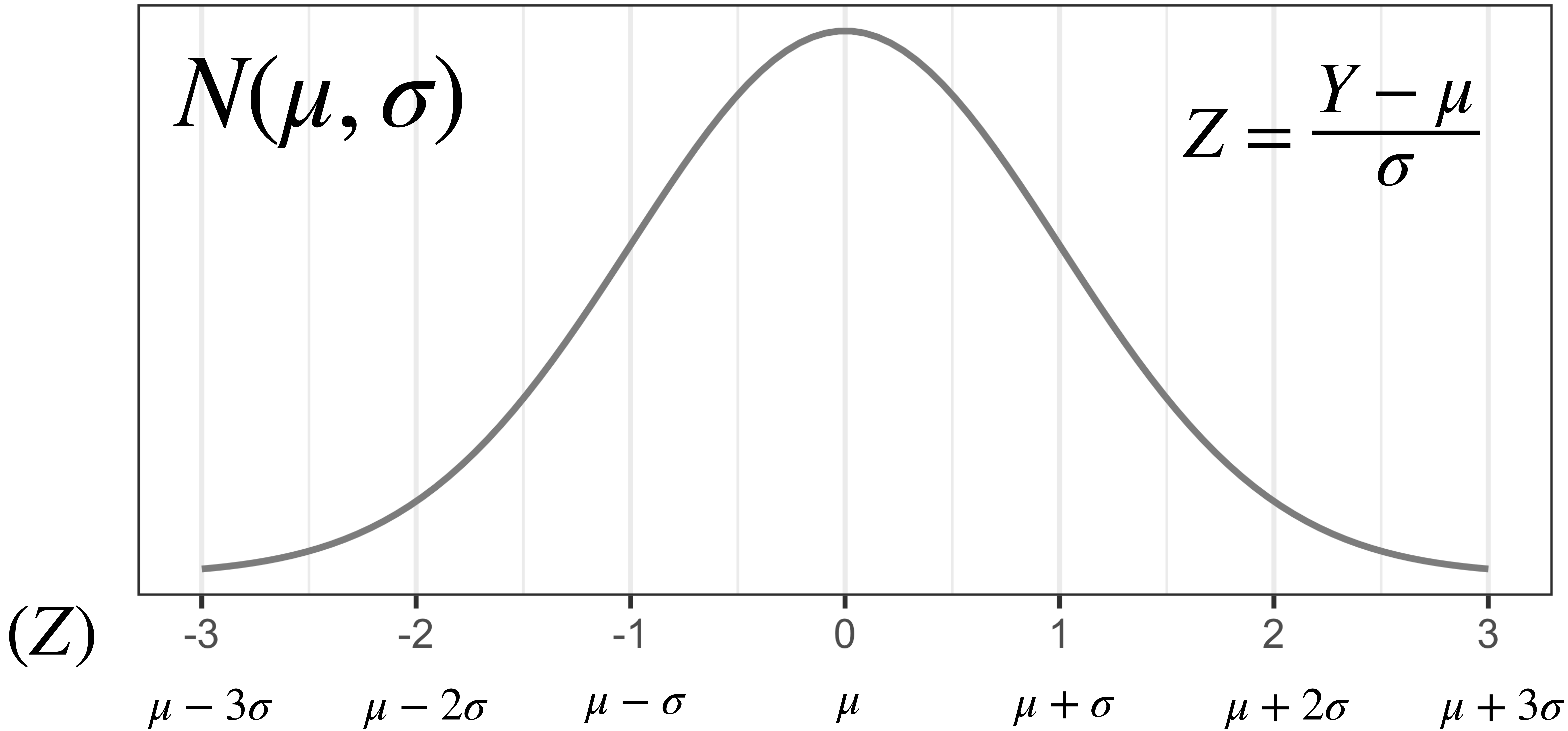
The normal distribution



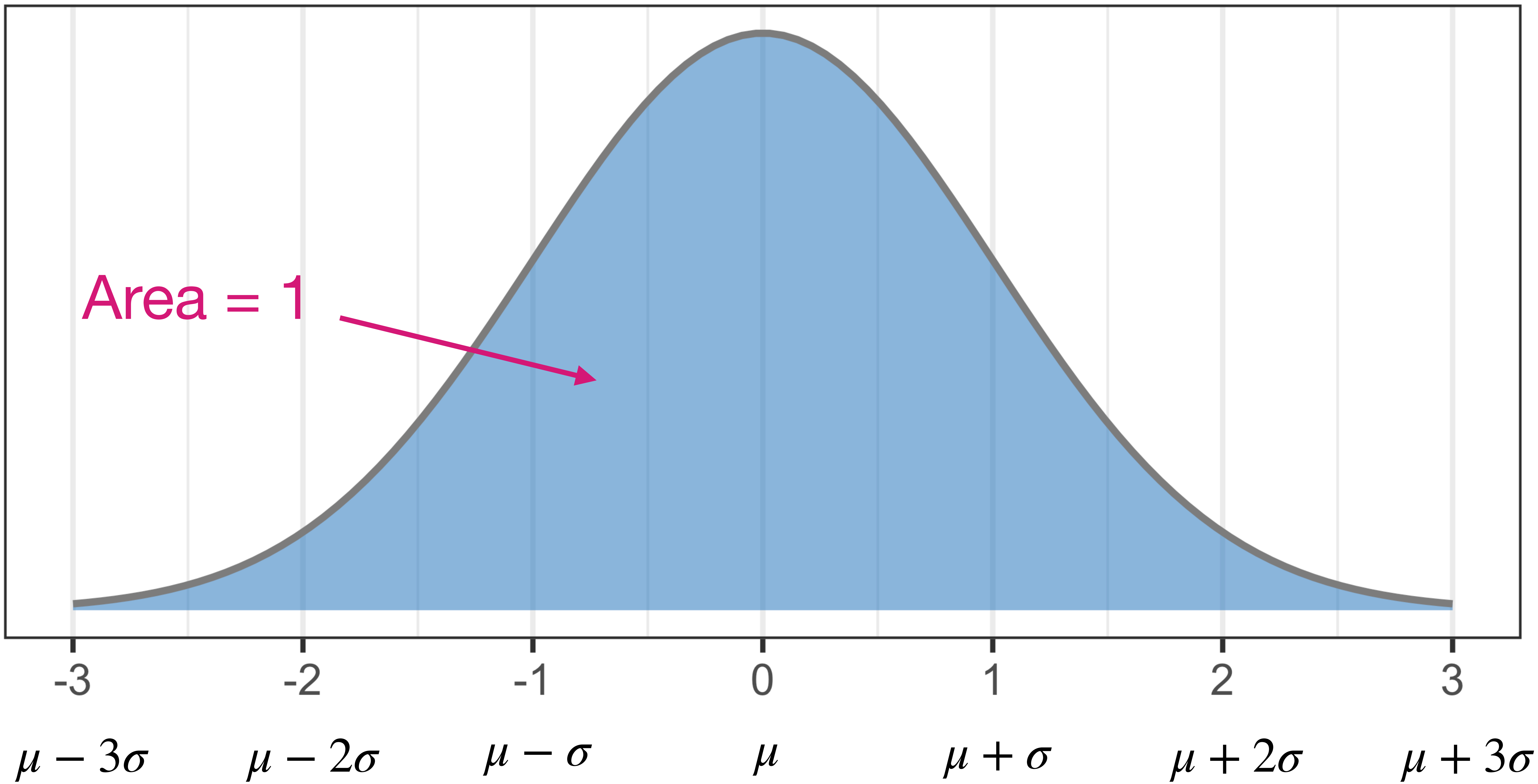
The normal distribution



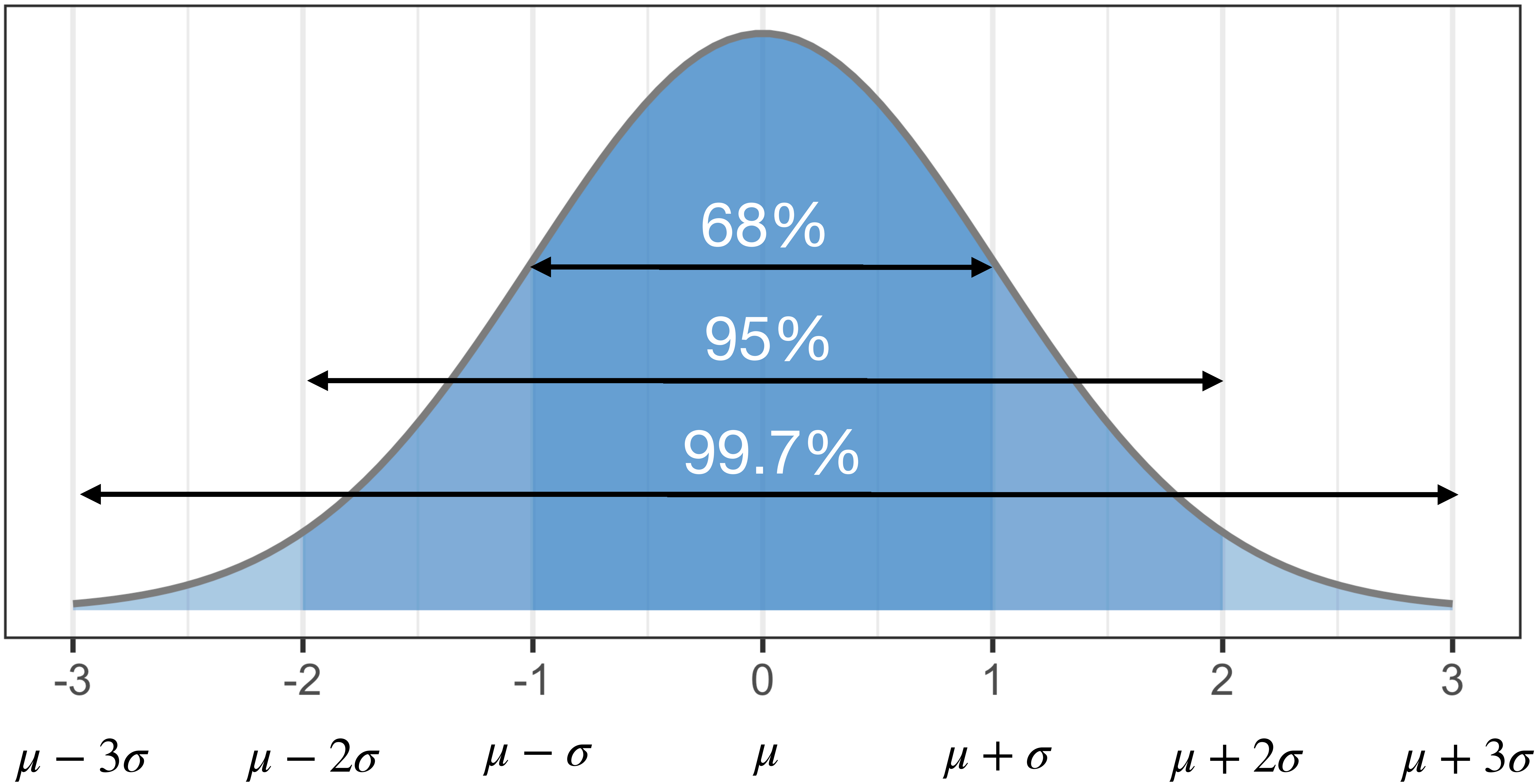
Z scale is a standardized normal distribution



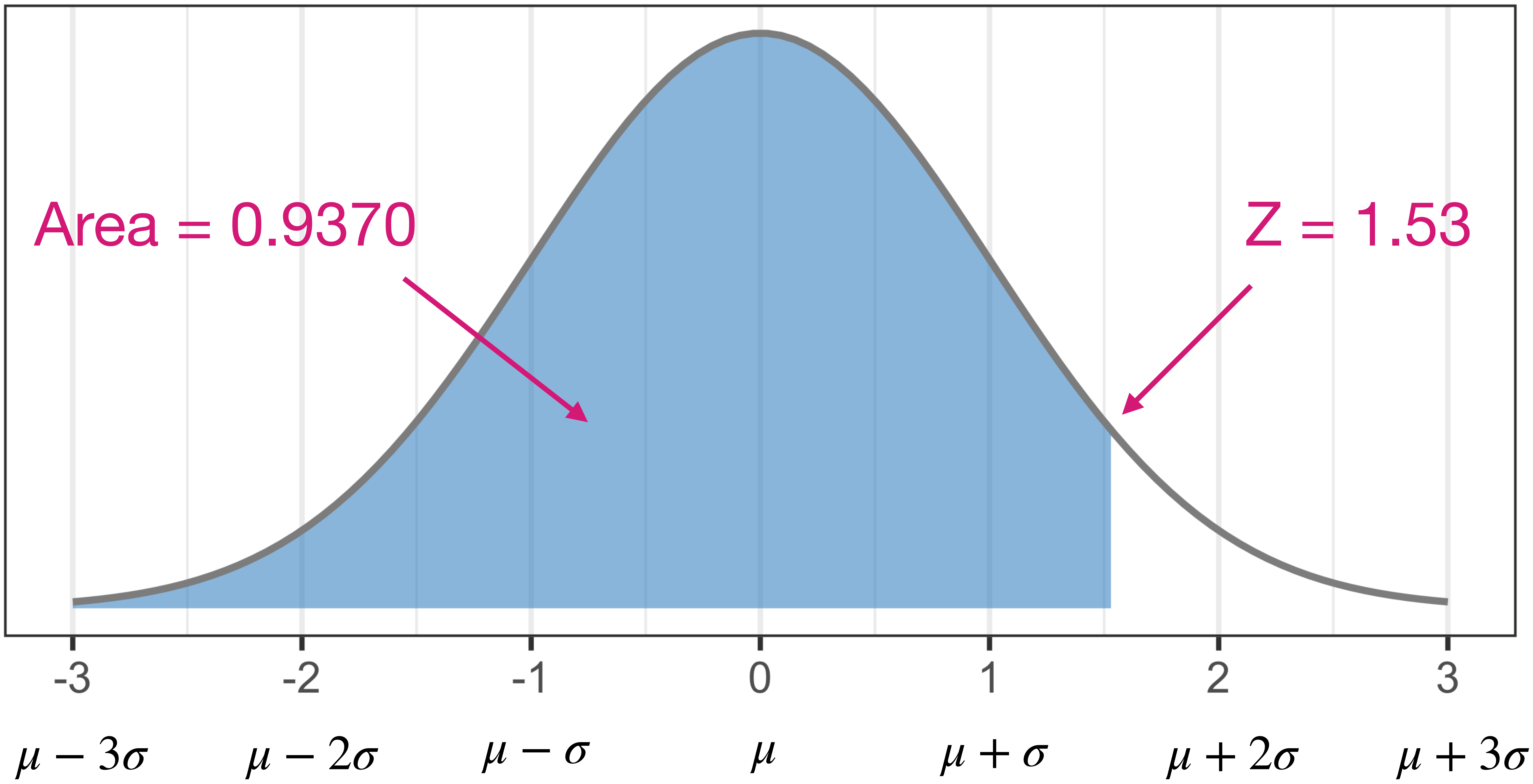
Z scale is a standardized normal distribution



Z scale is a standardized normal distribution



Z scale is a standardized normal distribution



Z = 1.53

Area = 0.9370

Tables of the Normal Distribution

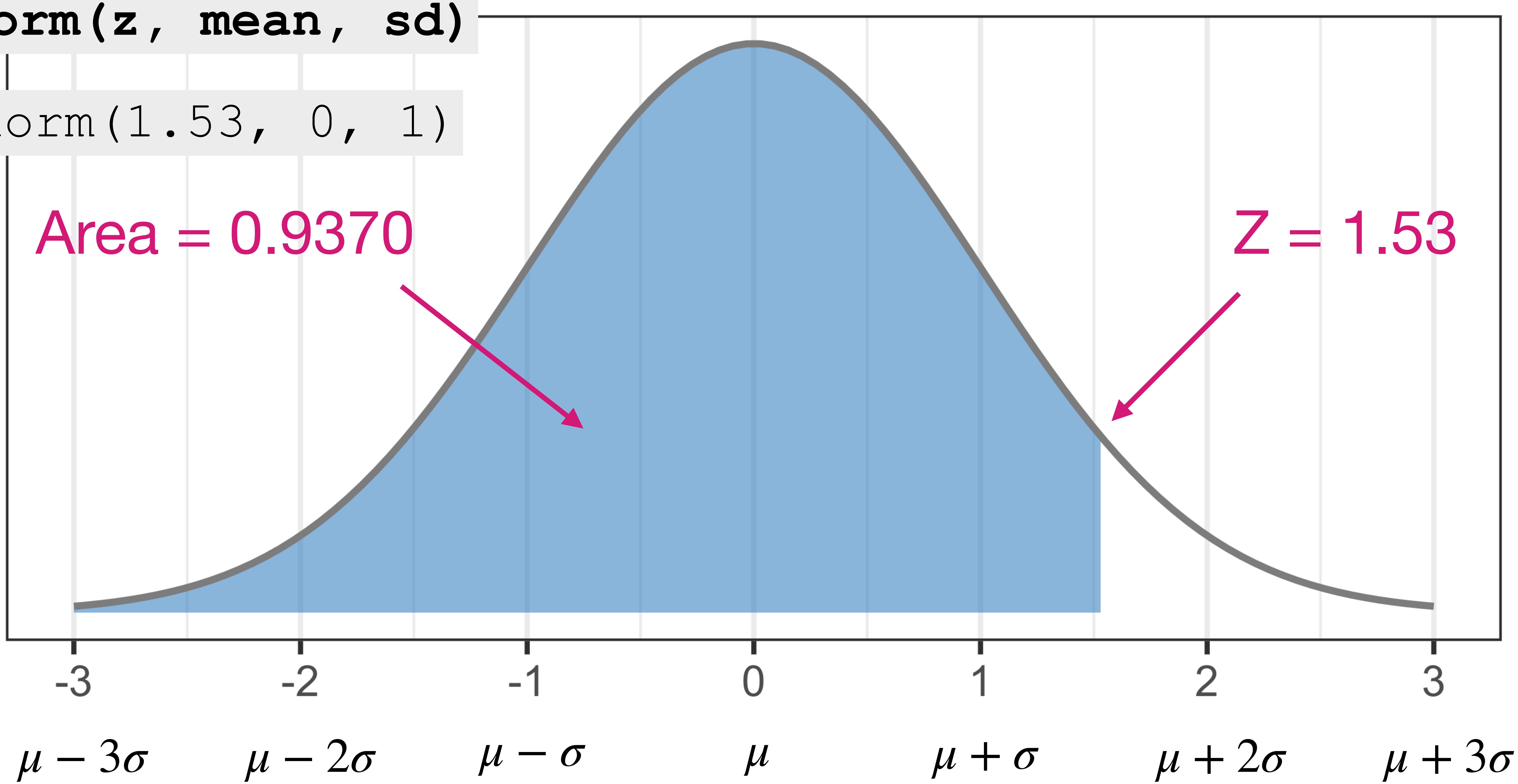
Probability Content from -oo to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Z scale is a standardized normal distribution

```
> pnorm(z, mean, sd)
```

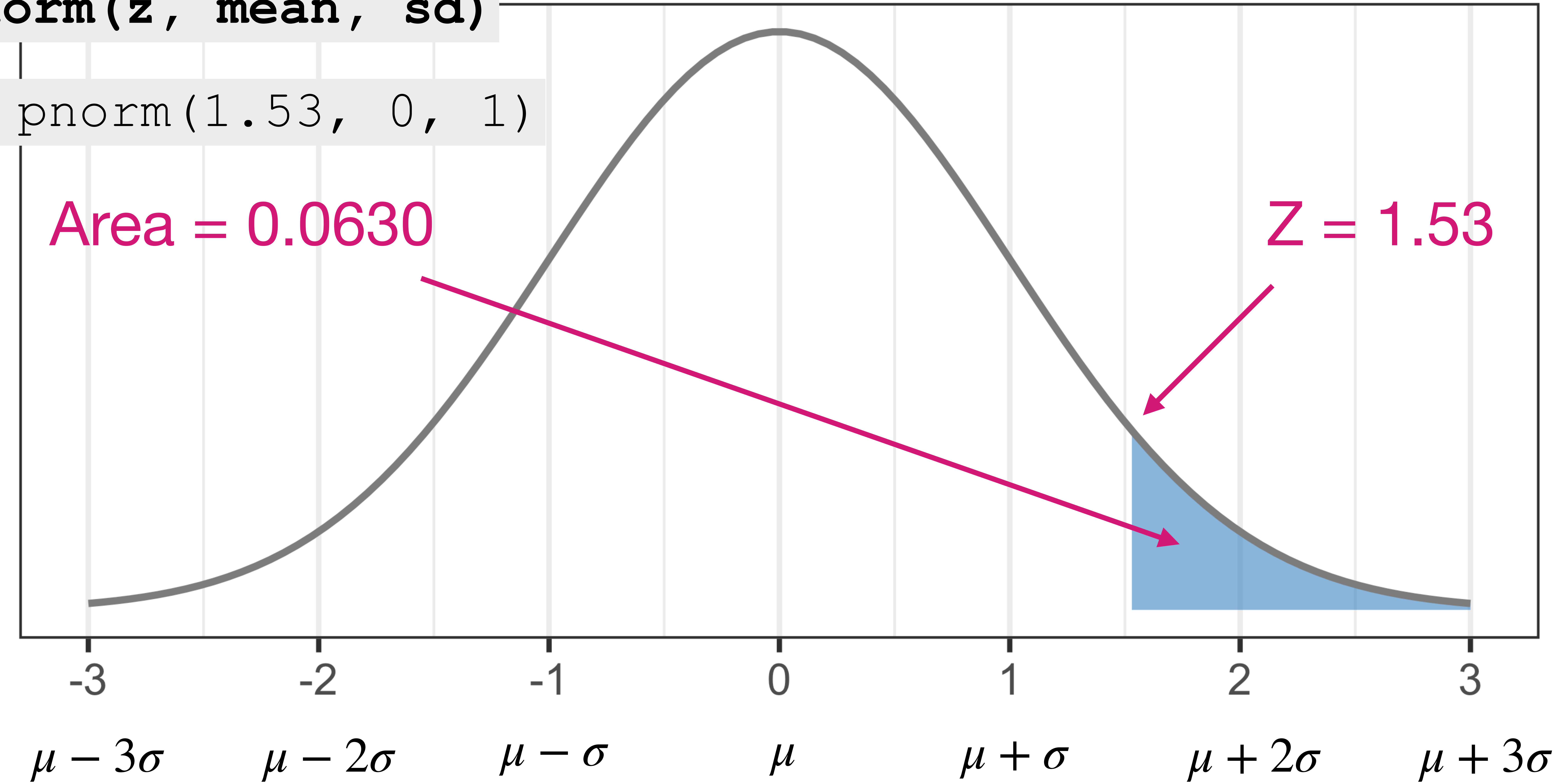
```
> pnorm(1.53, 0, 1)
```



Z scale is a standardized normal distribution

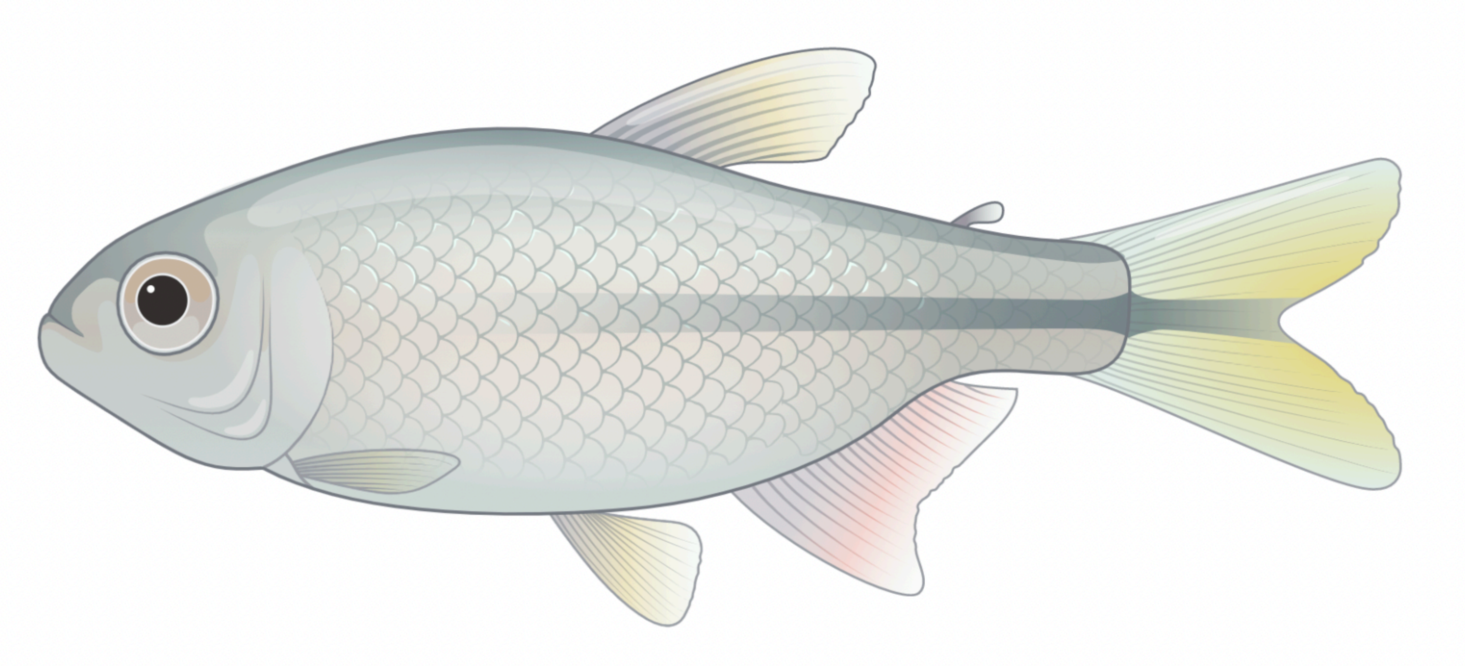
```
> pnorm(z, mean, sd)
```

```
> 1 - pnorm(1.53, 0, 1)
```

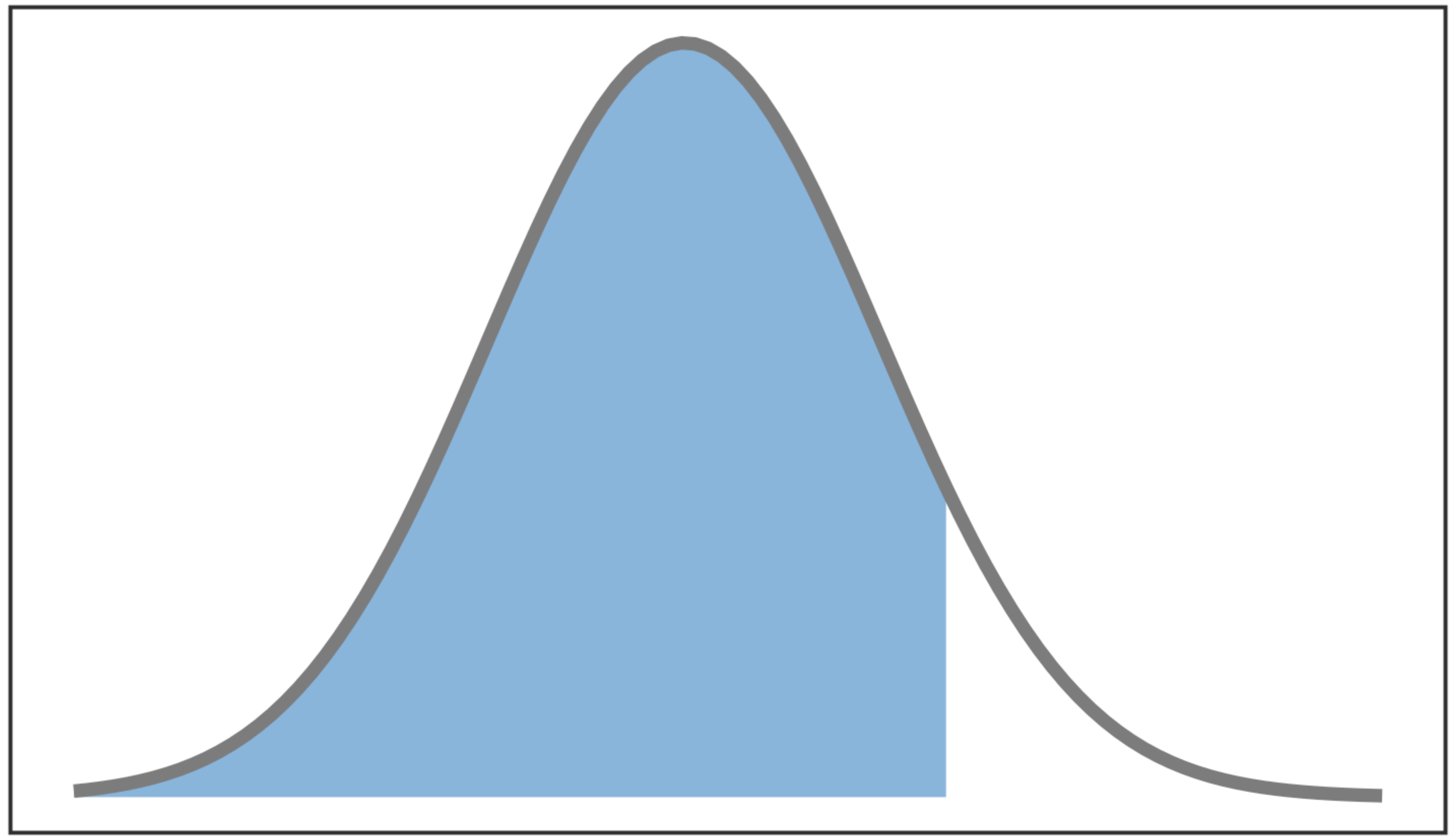


In a certain population of fish, individual lengths follow a normal distribution. The mean length of the fish is 54.0 mm, and the standard deviation is 4.5 mm. <u>What percentage of the fish are shorter than 60 mm?</u>

Even if you use R to calculate p-value, ALWAYS smart to make a quick sketch



$$Z = \frac{60 - 54}{4.5} = 1.33$$



```
> pnorm(z, mean, sd)
```

```
> pnorm(1.33, 0, 1)
```

```
> pnorm(1.33)
```

```
> 0.908
```

Default is mean = 0, sd = 1

$Z = 1.33$

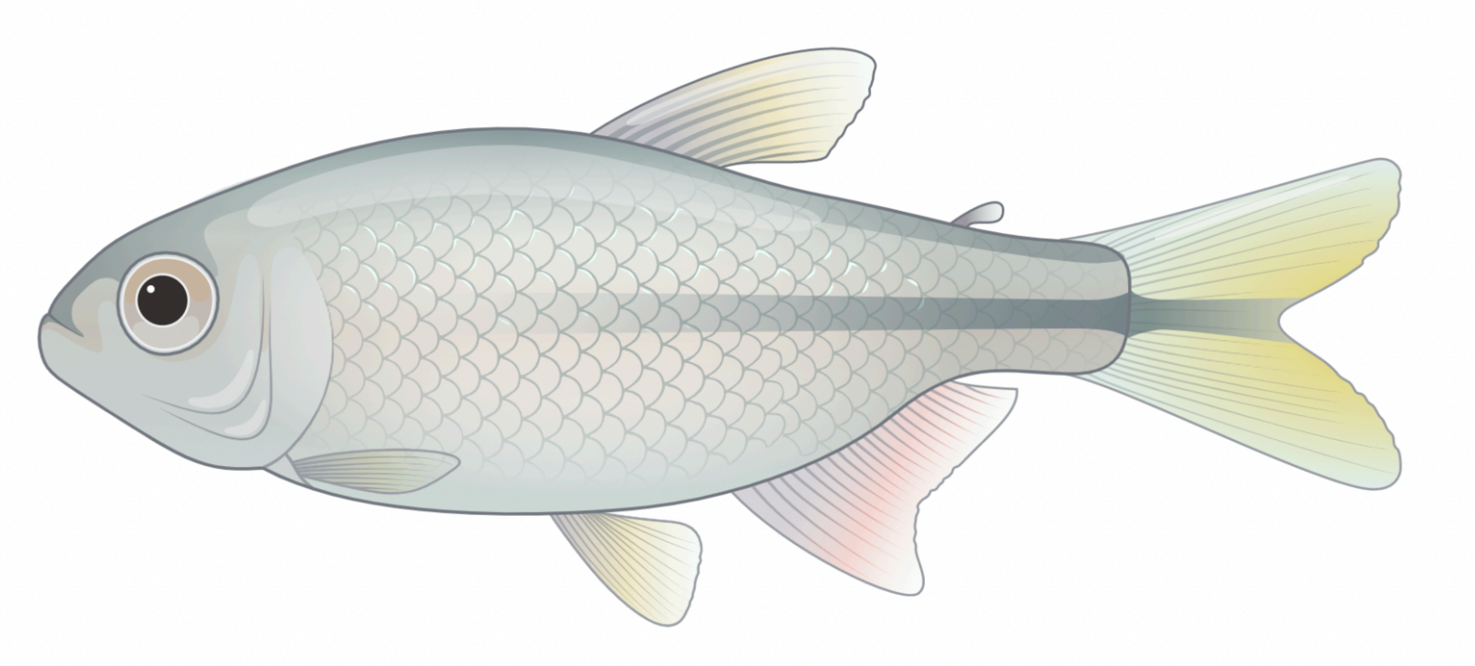
Area = 0.9082

Tables of the Normal Distribution

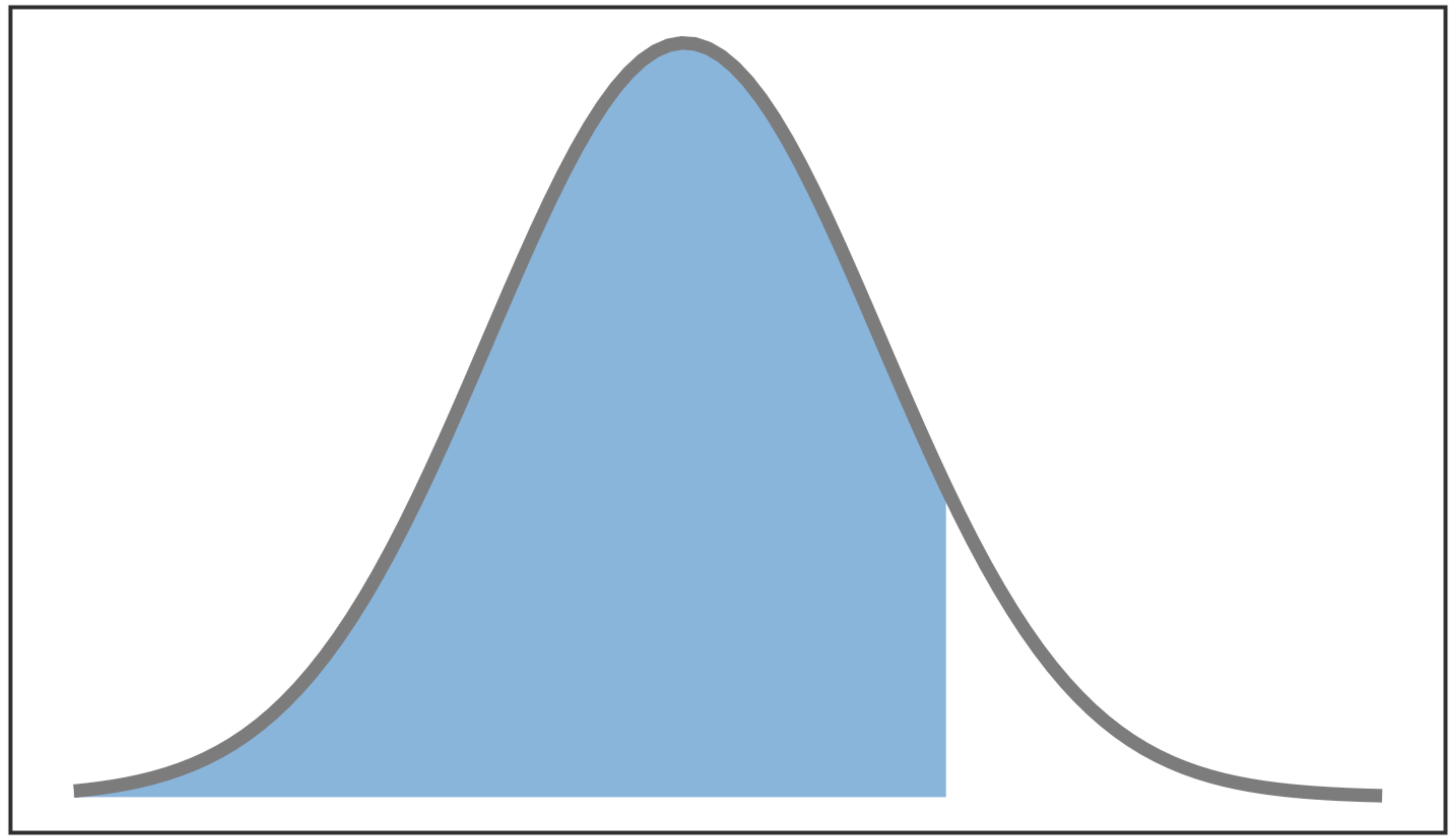
Probability Content from -oo to Z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	[illegible]	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

In a certain population of fish, individual lengths follow a normal distribution. The mean length of the fish is 54.0 mm, and the standard deviation is 4.5 mm. <u>What percentage of the fish are shorter than 60 mm?</u>



Even if you use R to calculate p-value, ALWAYS smart to make a quick sketch

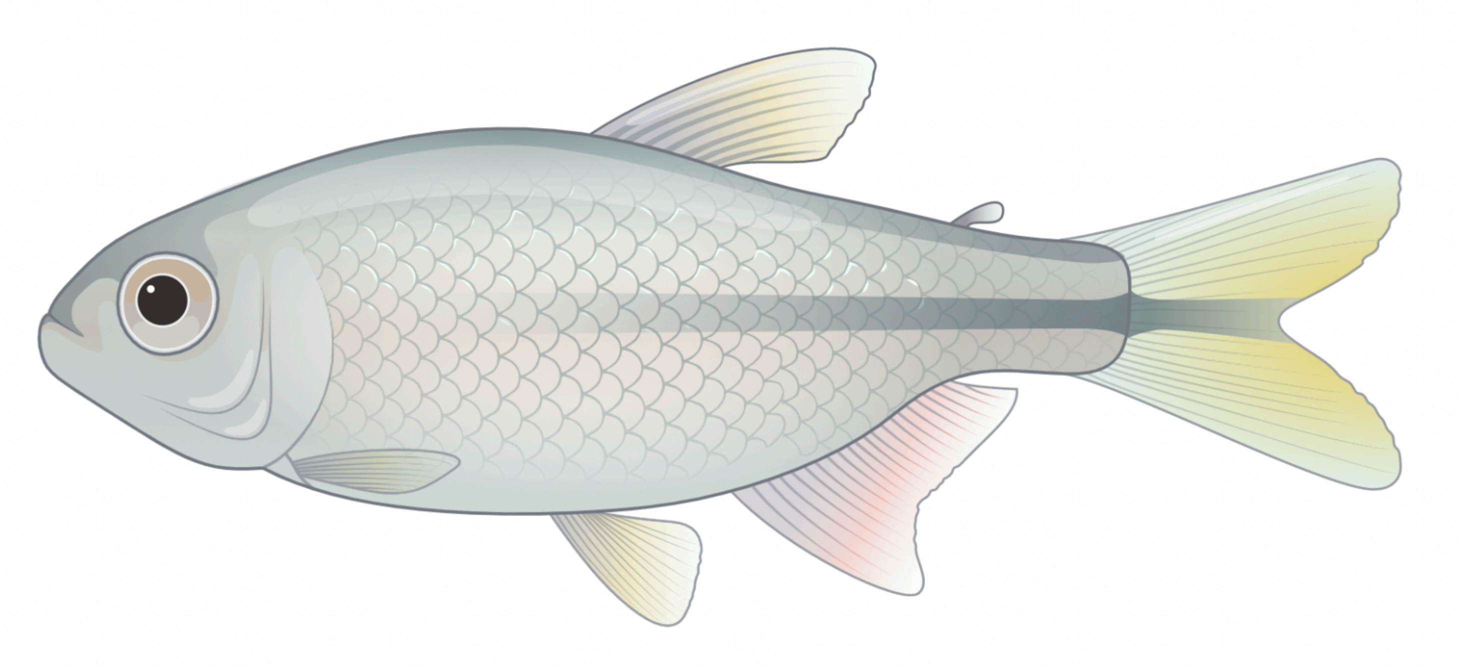


```
> pnorm(z, mean, sd)
```

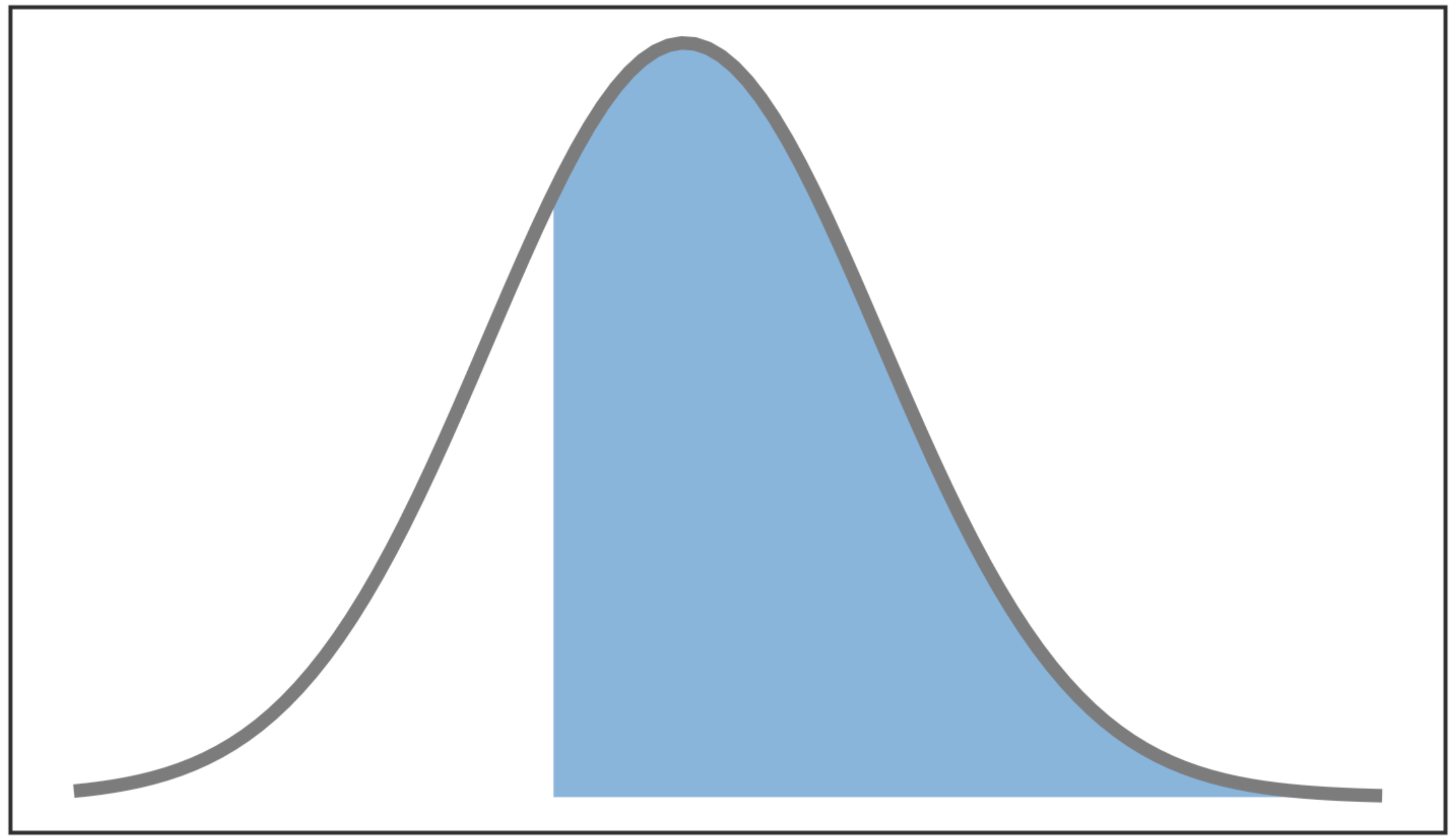
```
> pnorm(60, 54, 4.5)
```

```
> 0.908
```

In a certain population of fish, individual lengths follow a normal distribution. The mean length of the fish is 54.0 mm, and the standard deviation is 4.5 mm. <u>What percentage of the fish are longer than 51 mm?</u>



Even if you use R to calculate p-value, ALWAYS smart to make a quick sketch

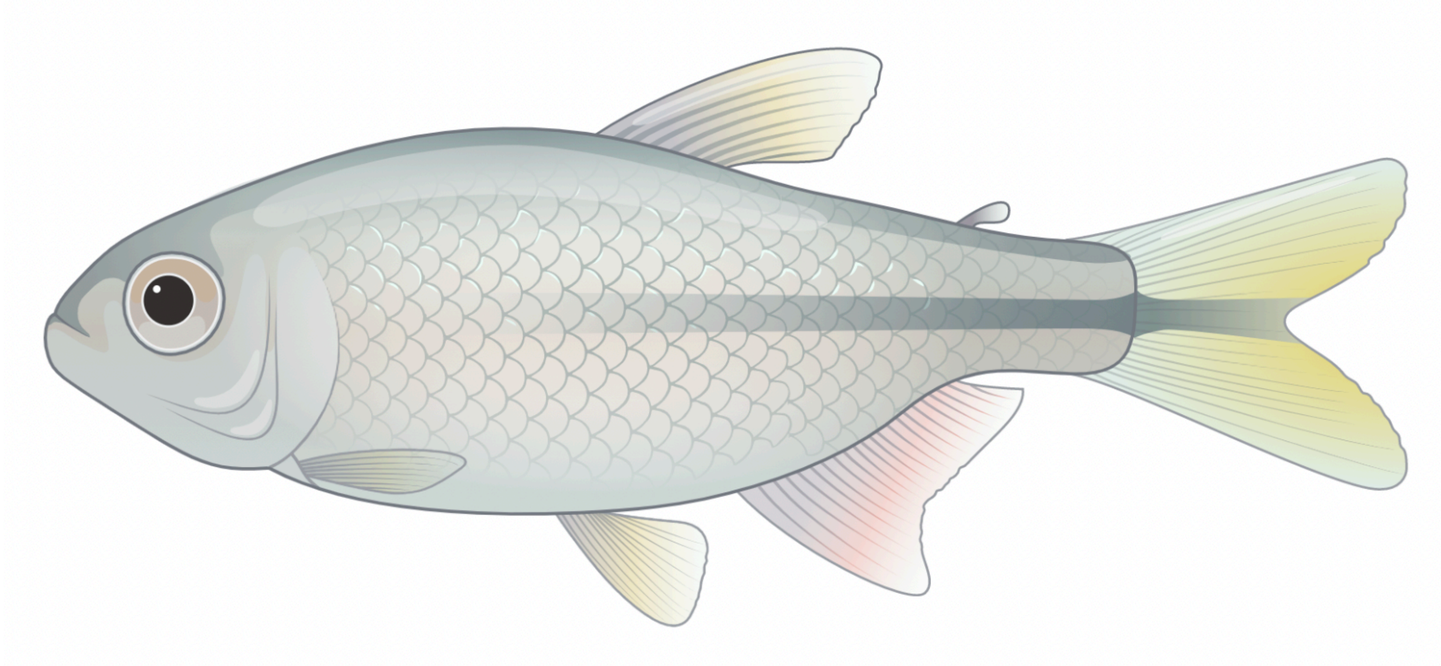


```
> pnorm(z, mean, sd)
```

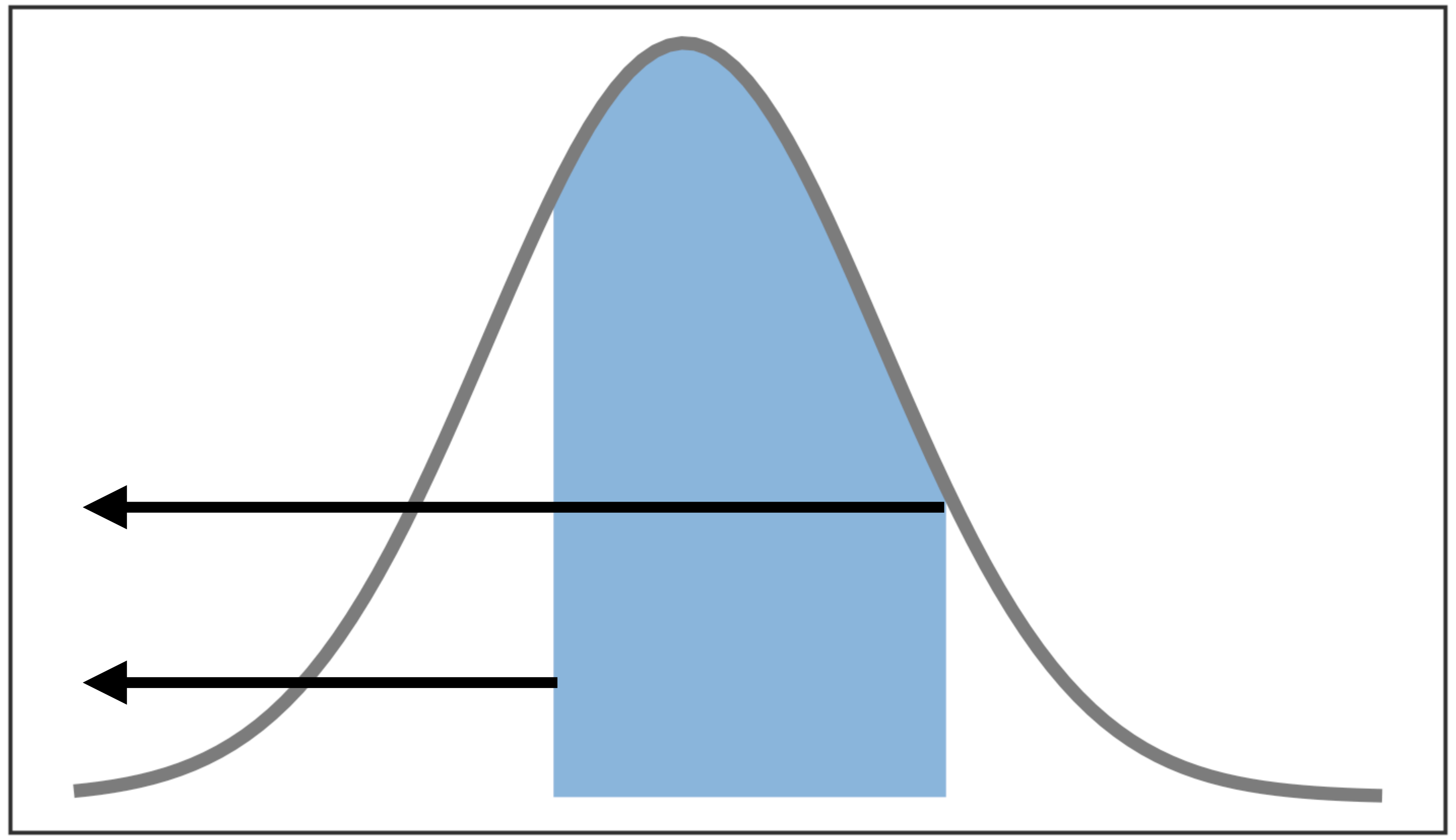
```
> 1 - pnorm(51, 54, 4.5)
```

```
> 0.747
```

In a certain population of fish, individual lengths follow a normal distribution. The mean length of the fish is 54.0 mm, and the standard deviation is 4.5 mm. <u>What percentage of the fish are between 51-60 mm?</u>



Even if you use R to calculate p-value, ALWAYS smart to make a quick sketch



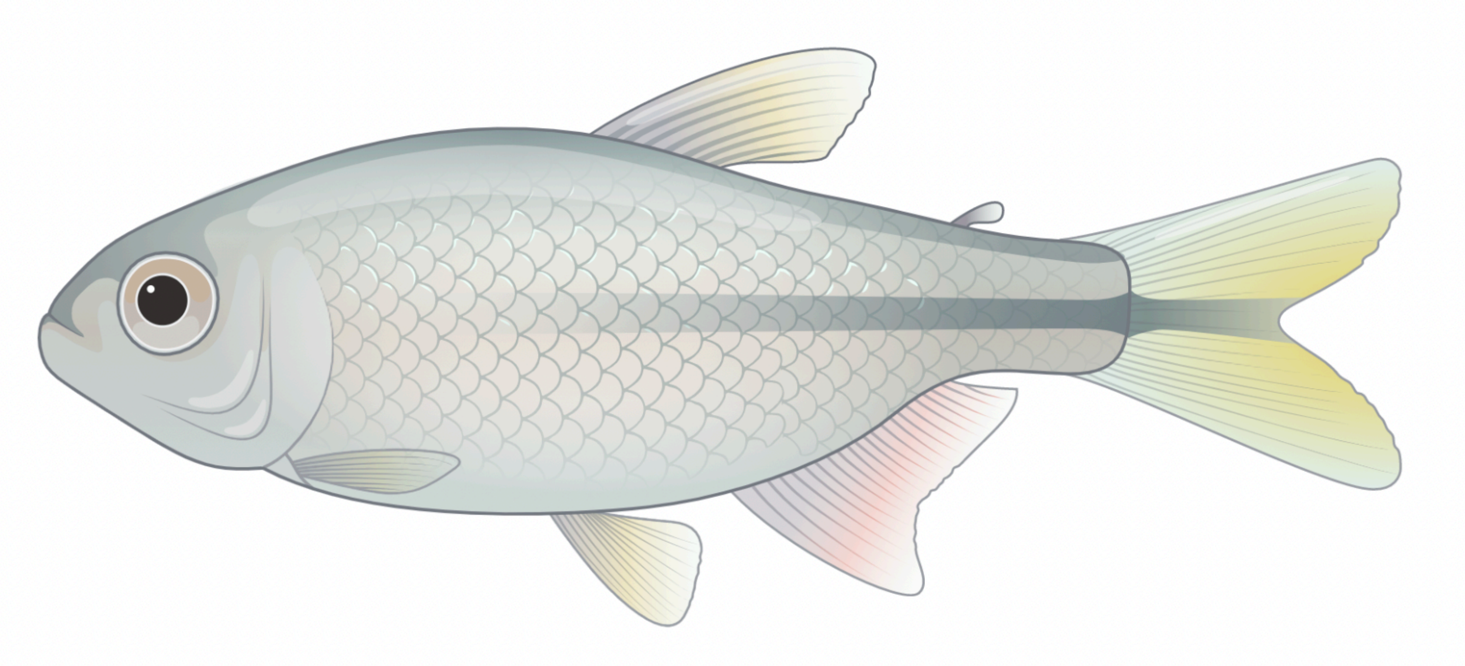
```
> pnorm(z, mean, sd)
```

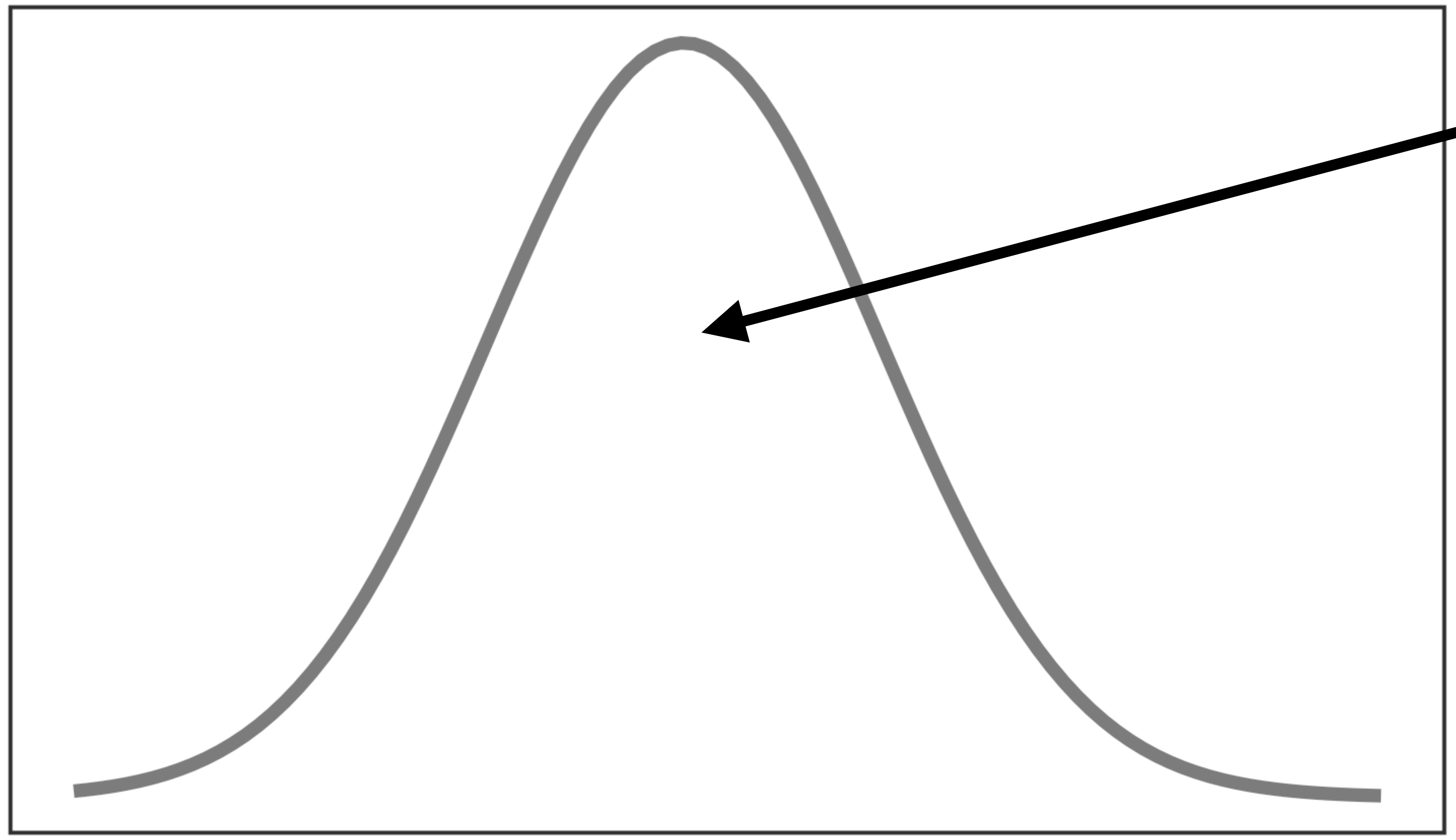
```
> pnorm(60, 54, 4.5) -
                pnorm(51, 54, 4.5)
```

```
> 0.656
```

In a certain population of fish, individual lengths follow a normal distribution. The mean length of the fish is 54.0 mm, and the standard deviation is 4.5 mm. <u>What percentage of the fish are exactly 53 mm?</u>

Even if you use R to calculate p-value, ALWAYS smart to make a quick sketch



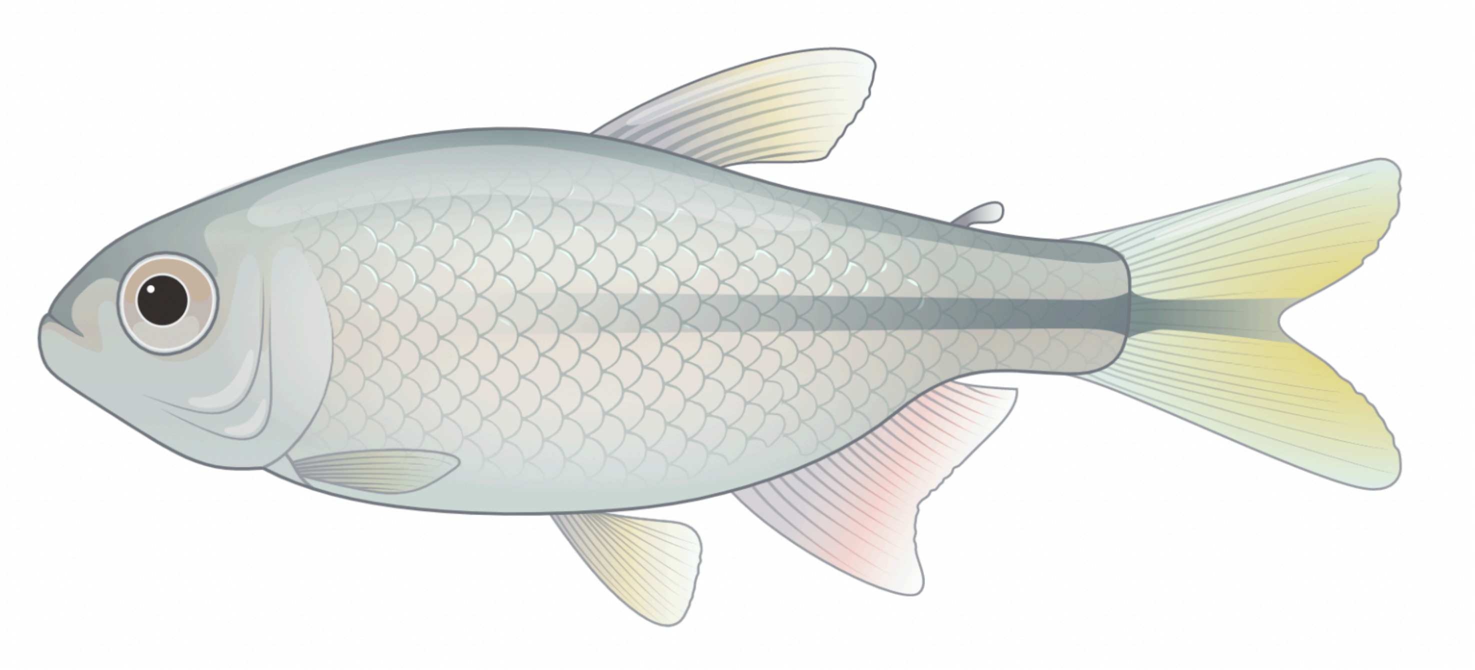


Invisible shaded area

0%! We need a range to calculate area under a curve…

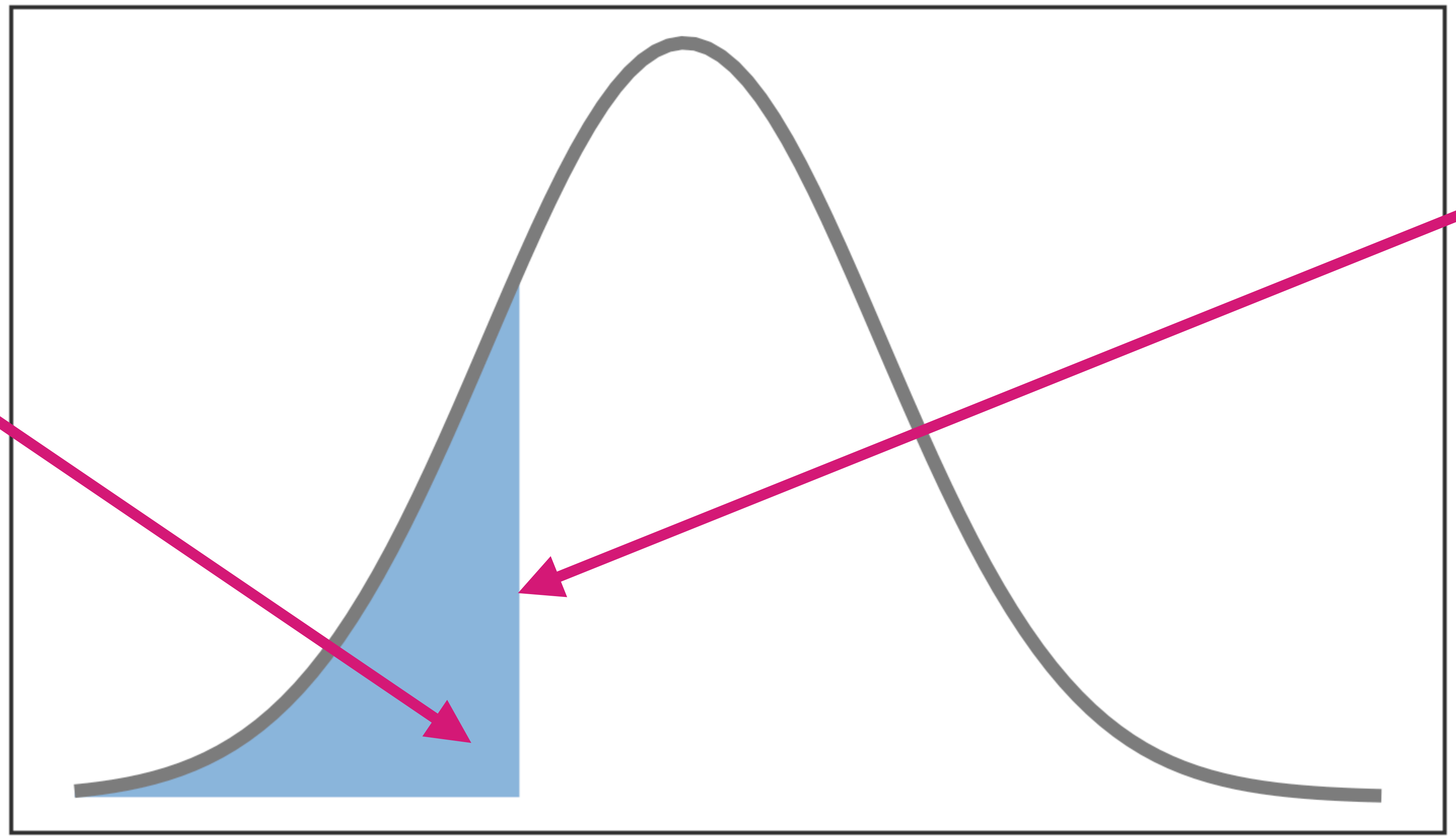
In a certain population of fish, individual lengths follow a normal distribution. The mean length of the fish is 54.0 mm, and the standard deviation is 4.5 mm. <u>What is the 20th percentile of fish length distributions?</u>

Even if you use R to calculate p-value, ALWAYS smart to make a quick sketch



20th percentile = 20% fish lengths < value

Area = 0.2



z = ?

```
> qnorm(p, mean, sd)
```

```
> qnorm(0.2, 54, 4.5)
```

```
> 50.2
```

Announcements

- Next class: Practical with R - **bring your laptop!!!**
  - If you are not familiar with R, please spend a little time with **swirl** before lecture on Tuesday — we won't have time to learn everything!
- Homework #2 due Tuesday @ 6pm
- TA office hours: Friday 1:30-2:30 and Tuesday 4:30-5:30 (Zoom)