PSYCH-UH 1004Q: Statistics for Psychology

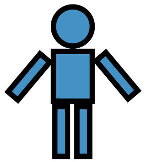
Class 5: Is a score/result extreme? - z-scores and the sampling distribution of the mean

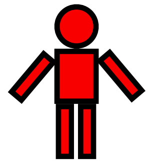
Prof. Jon Sprouse
Psychology

The z-score transformation

(The foundation of all statistical tests in this course.)

An example to get a feel for the z-score transformation: two students, two courses

 =70%

 =80%

Which student performed better on their respective exam?

In absolute terms, the red student. They got a higher percentage of the questions right.

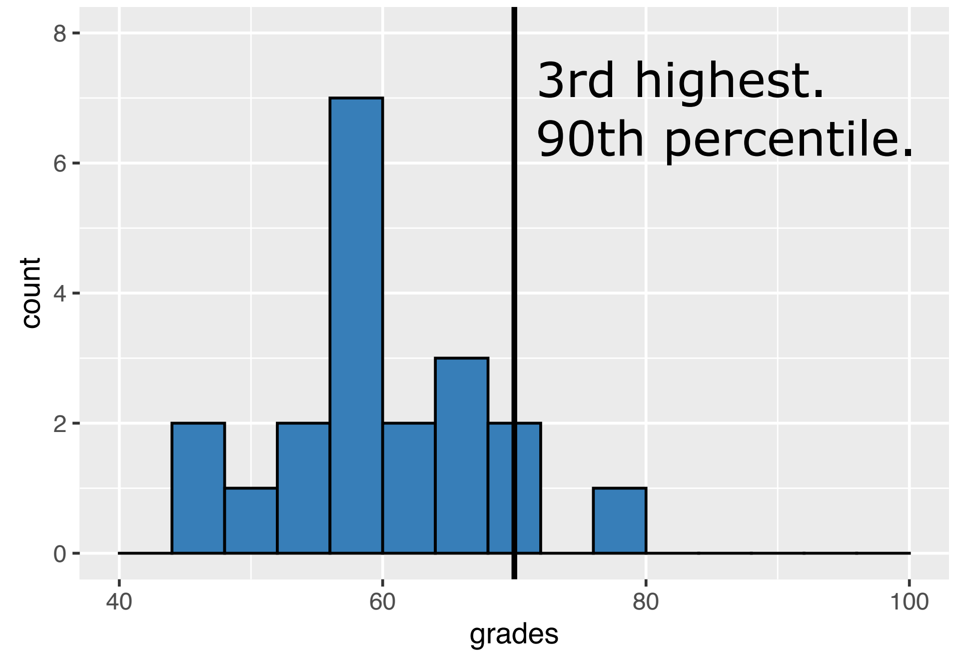
But we don't know how they performed in relative terms - that is, compared to the rest of the students in each of their courses. (Maybe the professor of the blue course is a harsher grader.)

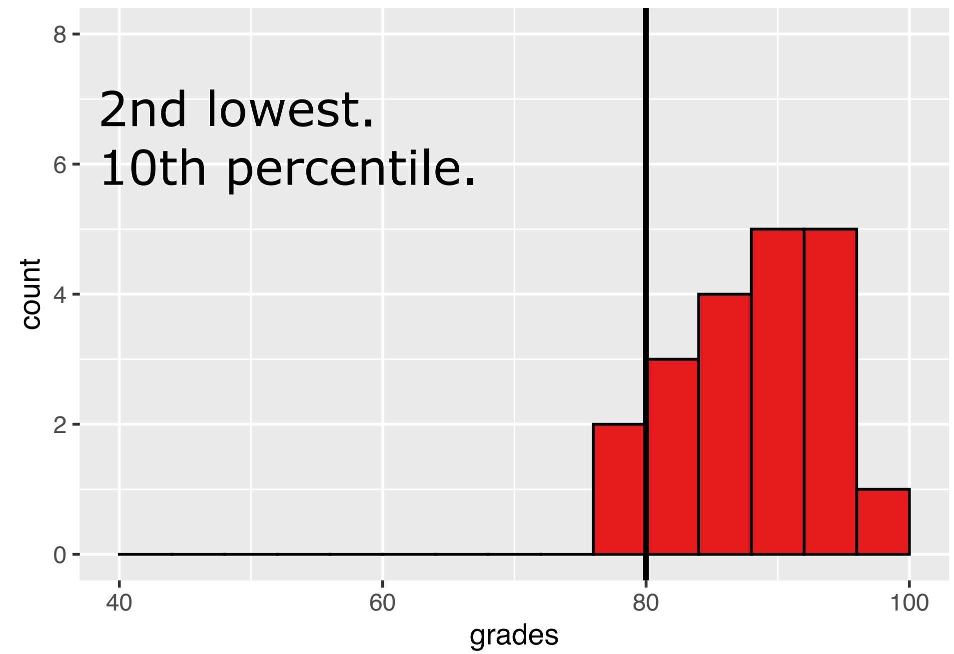
We already have the tools to answer this question. We can look at the distributions of grades in each course, and calculate the percentile ranks of their exam scores.

Finding percentile rank for two distributions

student	grade
1	78
2	72
3	**70**
4	67
5	65
6	65
7	64
8	62
9	59
10	59
11	59
12	58
13	57
14	57
15	57
16	56
17	55
18	49
19	46
20	45

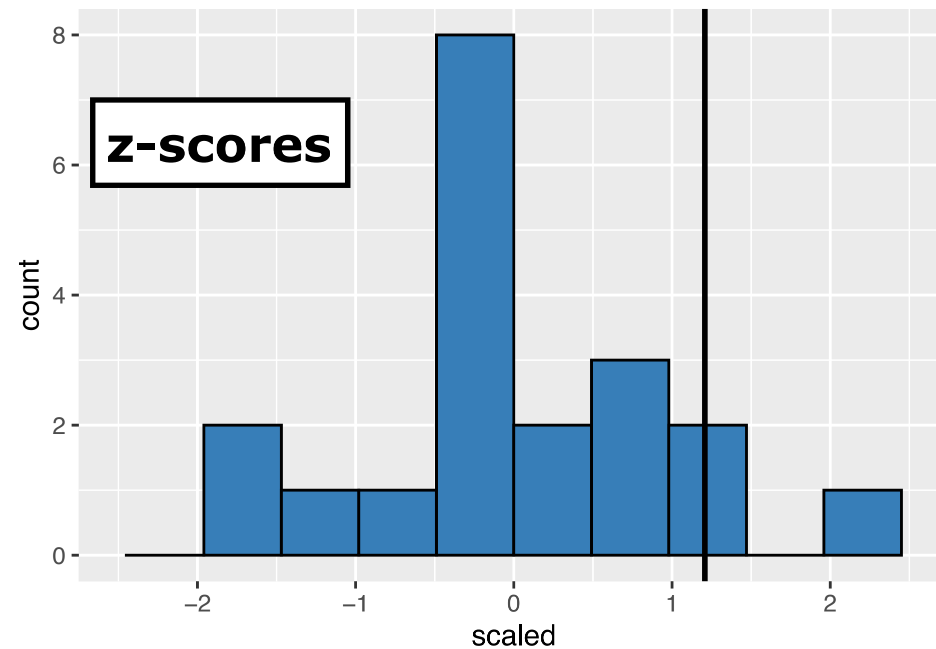
student	grade
1	97
2	96
3	95
4	95
5	94
6	94
7	92
8	91
9	90
10	90
11	89
12	87
13	87
14	86
15	86
16	82
17	82
18	81
19	**80**
20	79



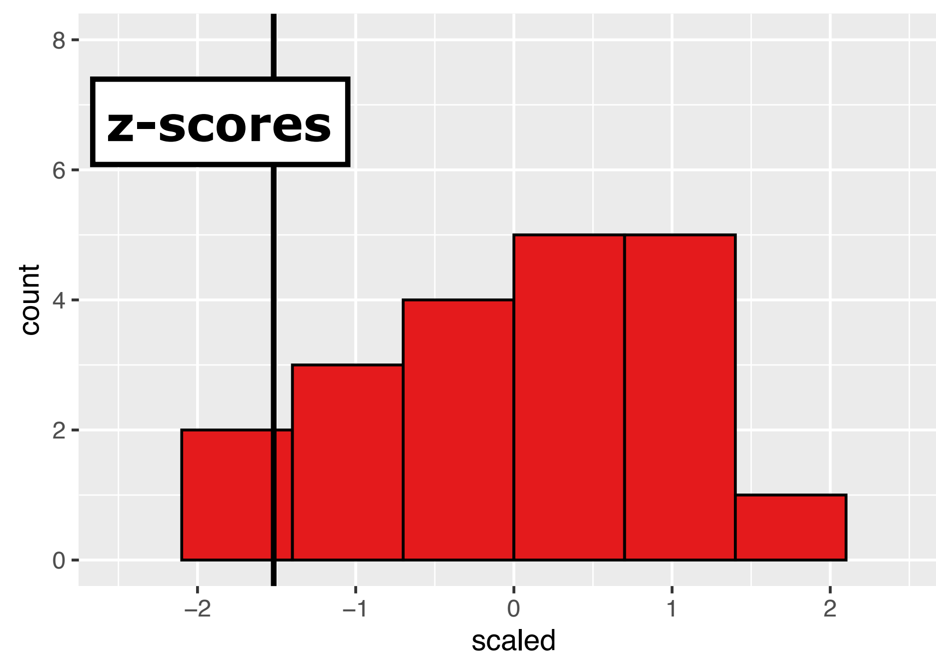


Standardizing the scores

With computers and the R language, it is not too difficult to sort scores, calculate percentile ranks, and plot the distributions for two different data sets.



However, we could also take a different approach that leads to a number of conceptual advantages (which we will discuss later). We can **standardize the scores**.



Standardizing means to put two different distributions on the same scale. The trick is to figure out how we can do that. The answer is to use our two descriptive statistics - **mean** and **standard deviation**.

Transforming into z-scores

The z-score transformation is a very common type of standardization. Here is the equation for it. You apply this to every score in a distribution (X), one at a time, to derive a new z-score:

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

(I wrote this using population parameters to match the book)

The z-score transformation does two things:

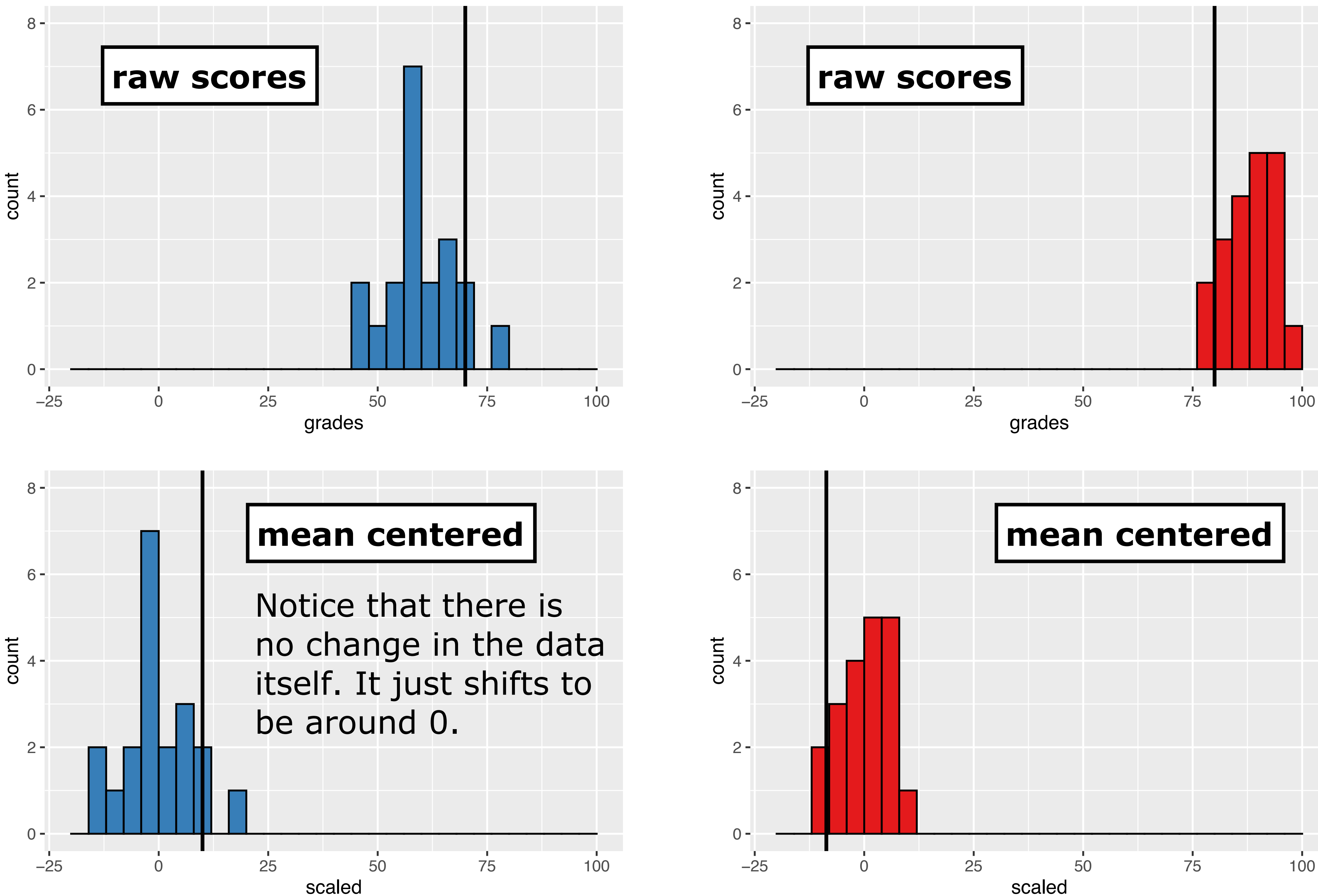
1. It subtracts the mean (μ) from each score (X). This is called mean centering because the <u>mean of the distribution is shifted to 0</u> so that all of the scores are symmetric around 0 (negative/positive).

2. It divides the difference by the standard deviation (σ). This is called scaling because it <u>changes the scale</u> to multiples of the standard deviation.

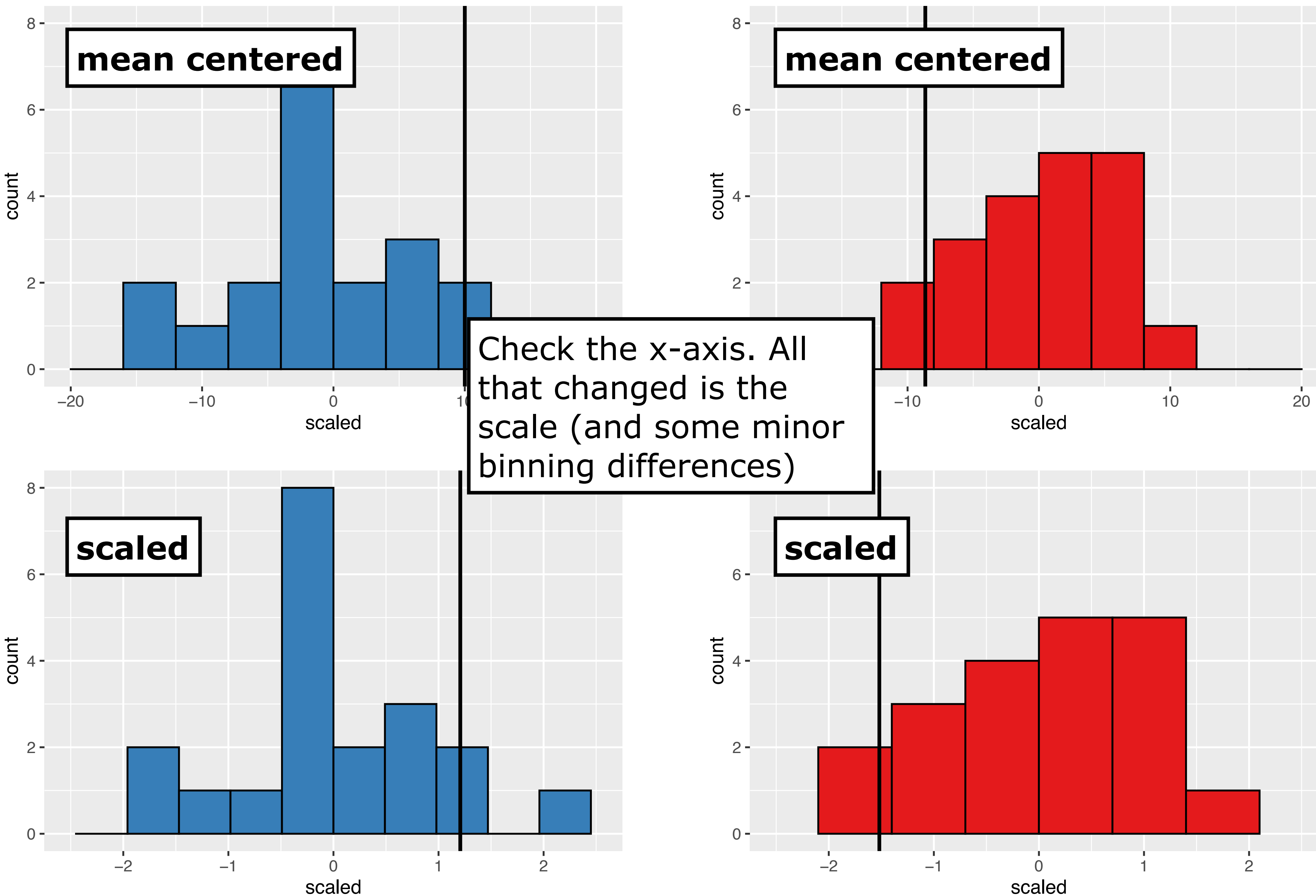
Let's look at each of these steps in isolation to see their effect.

Step 1: mean centering (X - μ)

(These wide x-axes are just for illustration - don’t do this!)



Step 2: scaling (/σ)



The arithmetic of the z-transformation

student	grade (μ=60)	centering ($x-\mu$)	scaled $(x-\mu)/\sigma$
1	78	18	2.17
2	72	12	1.45
3	**70**	**10**	**1.21**
4	67	7	0.84
5	65	5	0.60
6	65	5	0.60
7	64	4	0.48
8	62	2	0.24
9	59	-1	-0.12
10	59	-1	-0.12
11	59	-1	-0.12
12	58	-2	-0.24
13	57	-3	-0.36
14	57	-3	-0.36
15	57	-3	-0.36
16	56	-4	-0.48
17	55	-5	-0.60
18	49	-11	-1.33
19	46	-14	-1.69
20	45	-15	-1.81

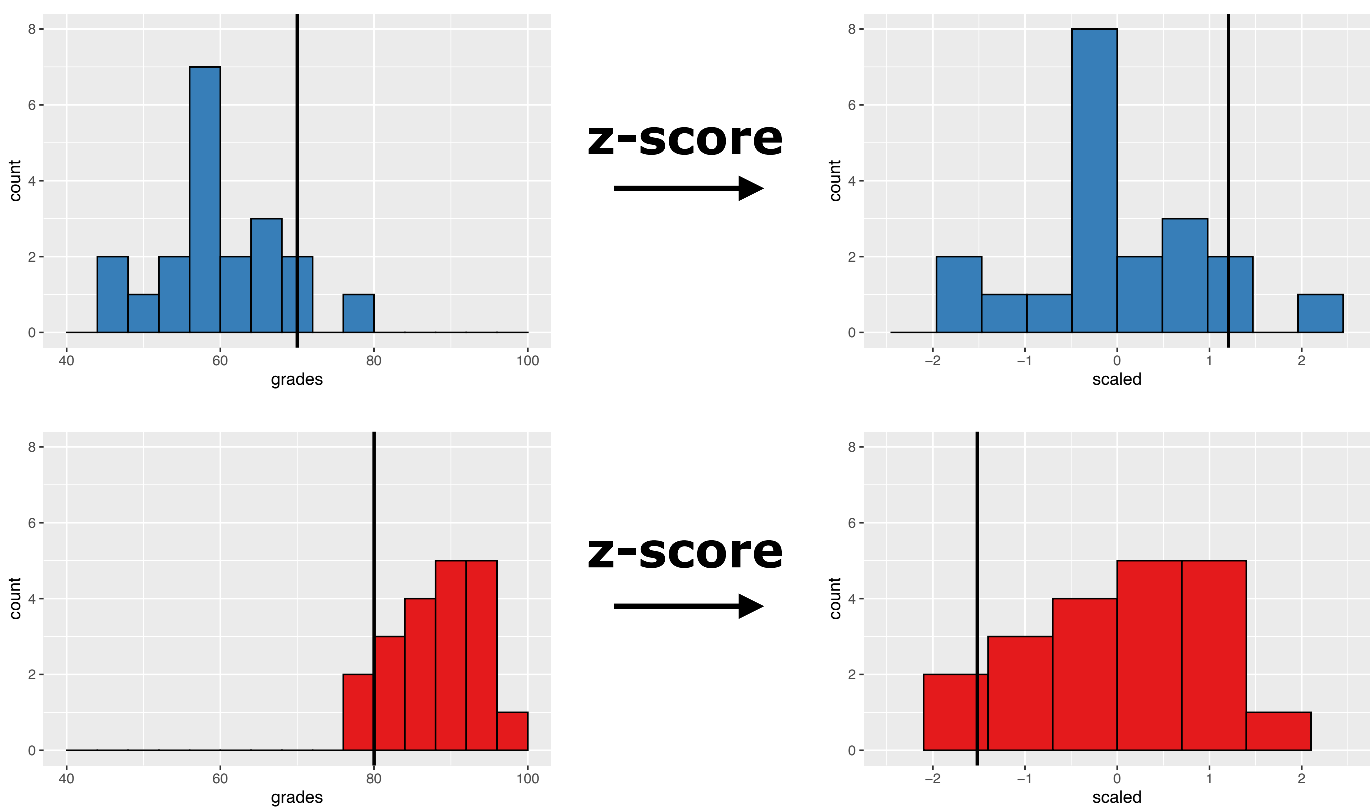
student	grade (μ=88.65)	centering ($x-\mu$)	scaled $(x-\mu)/\sigma$
1	97	8.35	1.47
2	96	7.35	1.29
3	95	6.35	1.11
4	95	6.35	1.11
5	94	5.35	0.94
6	94	5.35	0.94
7	92	3.35	0.58
8	91	2.35	0.41
9	90	1.35	0.24
10	90	1.35	0.24
11	89	-0.35	0.06
12	87	-1.65	-0.29
13	87	-1.65	-0.29
14	86	-2.65	-0.47
15	86	-2.65	-0.47
16	82	-6.65	-1.17
17	82	-6.65	-1.17
18	81	-7.65	-1.34
19	**80**	**-8.65**	**-1.52**
20	79	-9.65	-1.69

The z-score transformation does not change the shape of the distribution

The z-score transformation is a linear transformation. That is a precise mathematical term, but for our purposes, we can say that it preserves the relative structure of a data set.

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

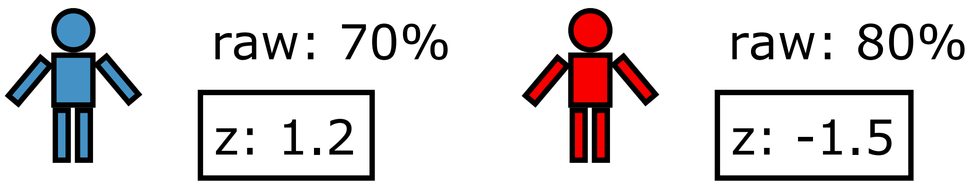
We can see this in our raw and z-score transformed plots. The shape of the distribution is the same, except for some minor binning differences that arise because of the scale change (I tried to minimize those).

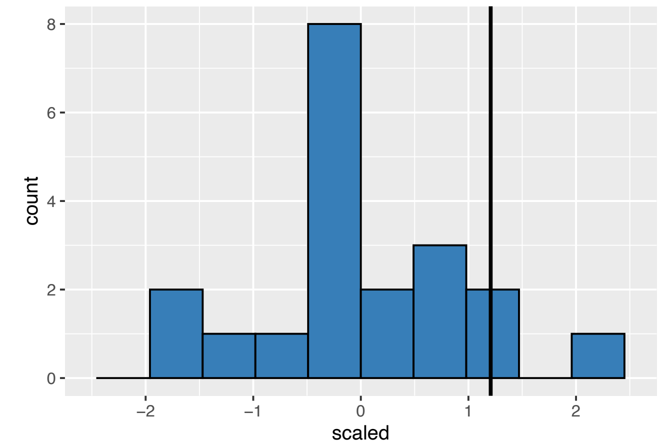


Crucially, both courses are now on the same scale. They have the same mean. And they use the same scale - standard deviations.

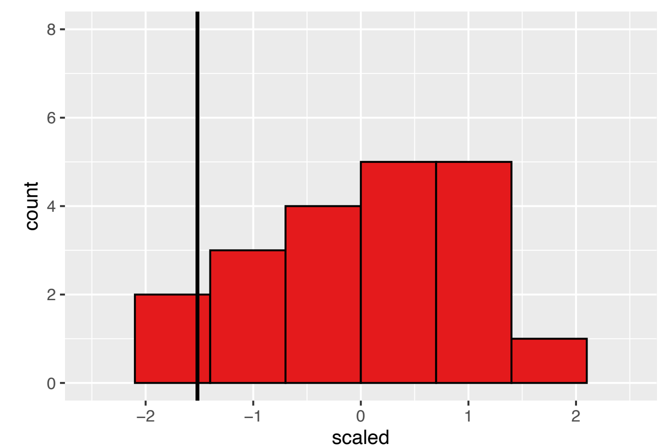
We can compare z-scores across distributions

The z-score transformation allows us to compare scores between the two distributions directly.





Z-scores work in a very specific way. The mean is always 0. The sign of the z-score tells you if the score is above (+) or below (-) the mean. The magnitude tells us how far above or below the mean - in number of standard deviations!

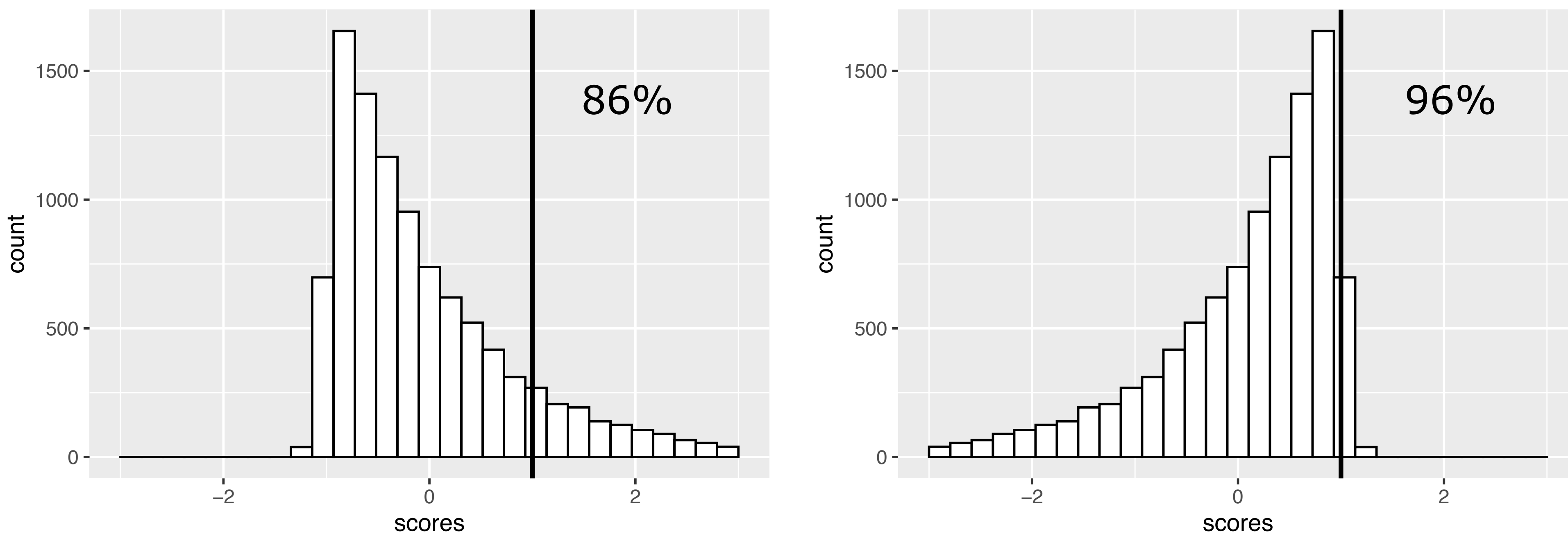


The student in the blue class scored 1.2 sd <u>above</u> the mean; the red scored 1.5 <u>below</u> the mean.

But there is one caveat - shape of the distribution

In order for the z-scores from two distributions to be comparable, the two distributions must have the same shape.

To show this, we can take two very skewed distributions, z-score transform them, and then look at the percentile rank of a z-score of +1.



This is an extreme example, but it illustrates the general point - if you want to compare two distributions using z-scores, be sure that they have (roughly) the same shape.

The power of z-scores combined with the normal distribution

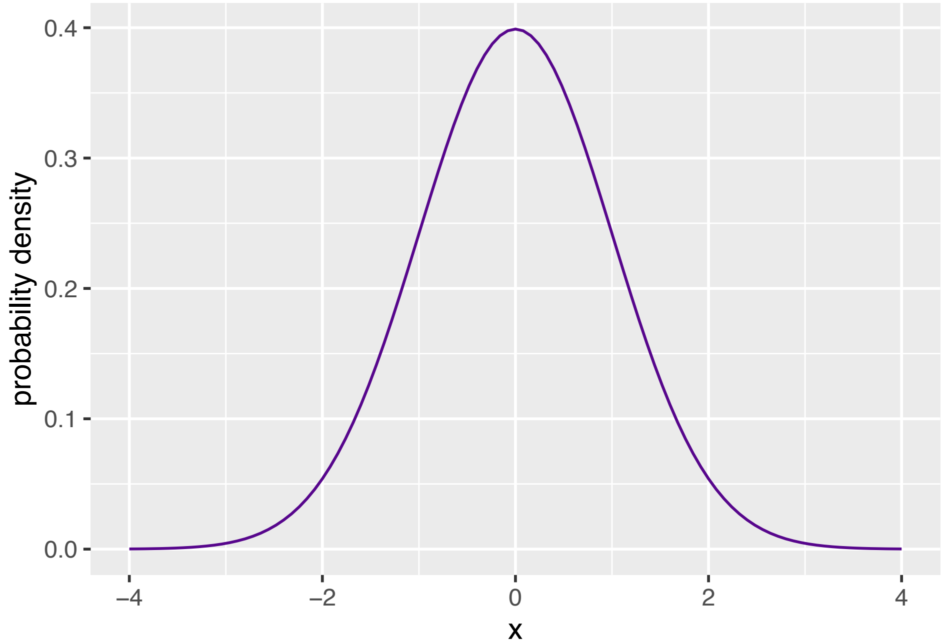
1. A first introduction to the normal distribution

(We will go into more depth another day.)

The normal distribution

In the real world, there are an infinite number of distributions that data can appear in. But mathematicians have identified several special distributions, named them, and explored their mathematical properties (binomial, beta, gamma, etc).

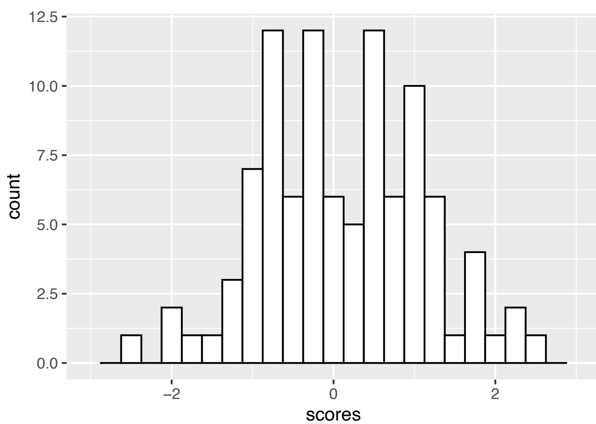
The most important distribution in frequentist statistics is called **the normal distribution**. It is also called the Gaussian distribution after German mathematician Carl Friedrich Gauss (1777-1855). And sometimes called the "bell curve" after its shape.



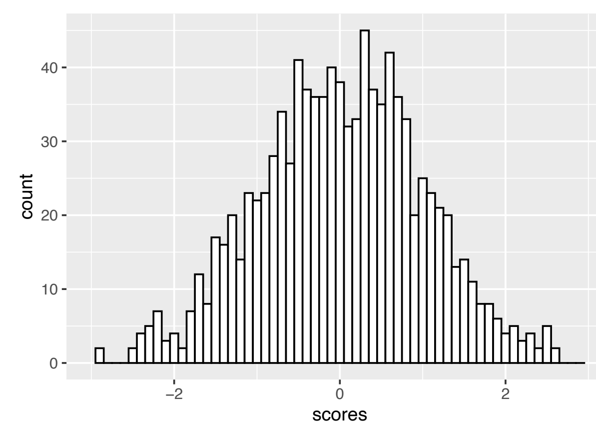
Why is it smooth?

The normal distribution is a theoretical distribution (defined by math), not an empirical one (defined by actual data points). The idea is that this is what we'd find if collected a sample of infinite size, and plotted a histogram with infinitely small bin widths. We can see this by simulating larger and larger samples:

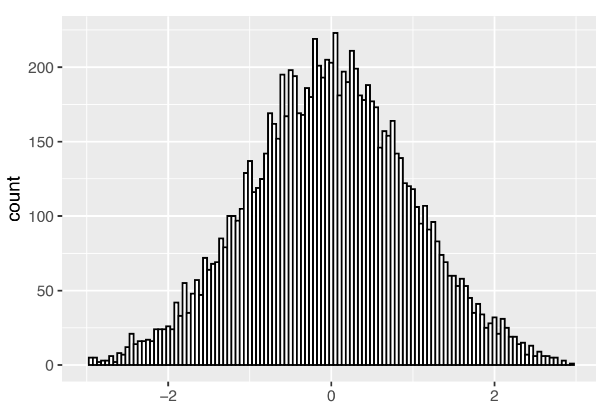
scores=100, bin=.25



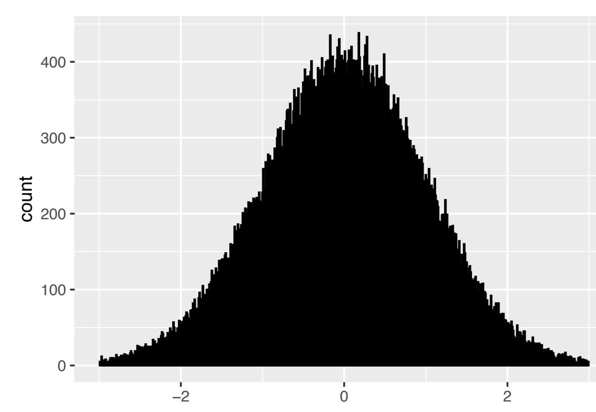
scores=1000, bin=.1



scores=10000, bin=.05



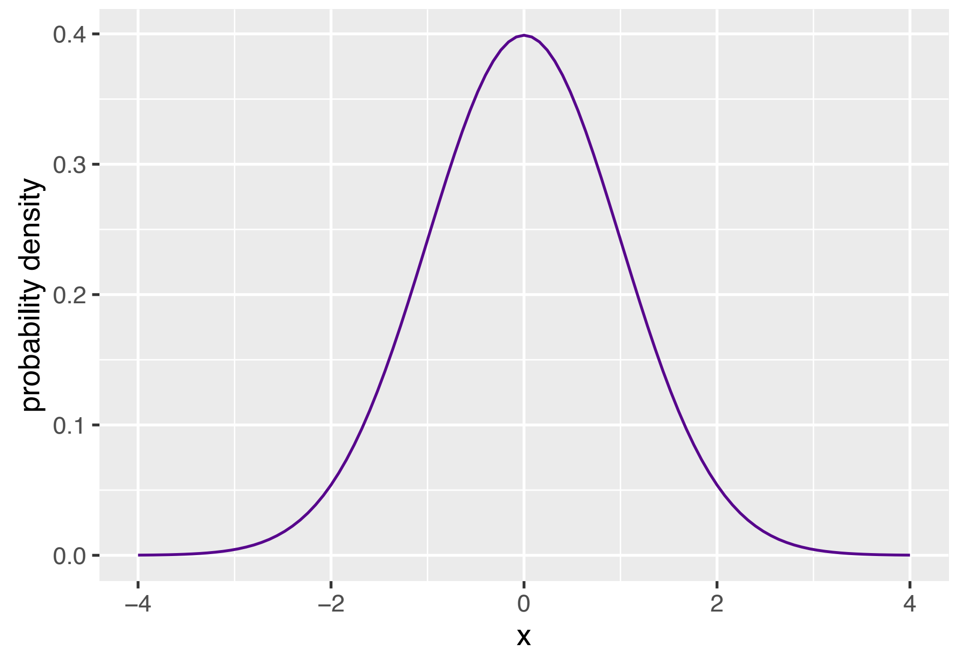
scores=100000, bin=.01



The equation for the normal distribution

All theoretical distributions are defined by an equation. The equation for the normal distribution is below. You do not have to memorize this. The equation for the normal distribution requires you to enter two numbers (called parameters!) in order for it to yield the distribution - those two represent the **mean** and **standard deviation** of the distribution!

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

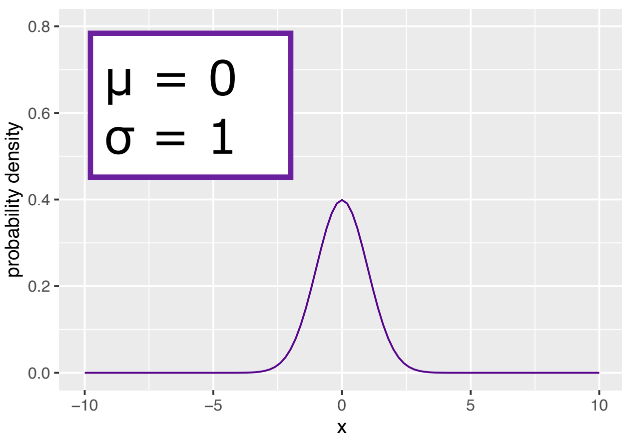


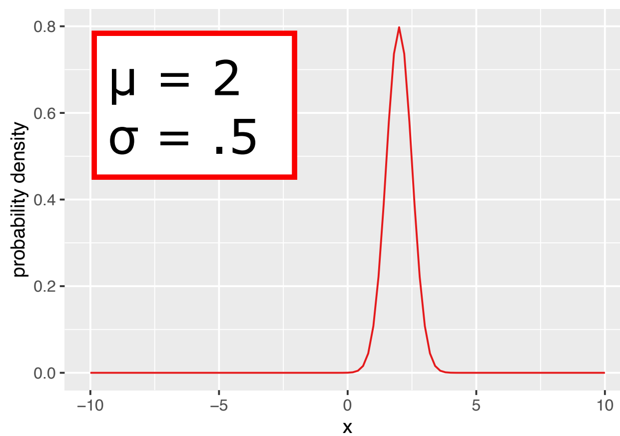
Caution: You will also notice that the y-axis is labeled "probability density". I have not introduced this idea yet. We've only been talking about frequency. Please think about it as related to frequency for now. We will discuss it in a couple of days.

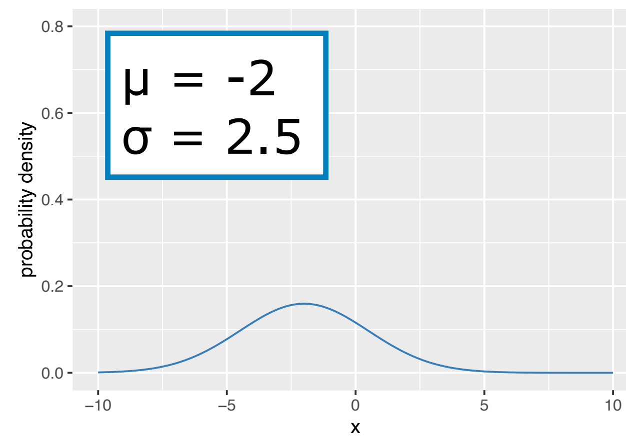
It is a family of distributions

By plugging in different means and standard deviations, we get different instances of the normal distribution. But they are all normal distributions!

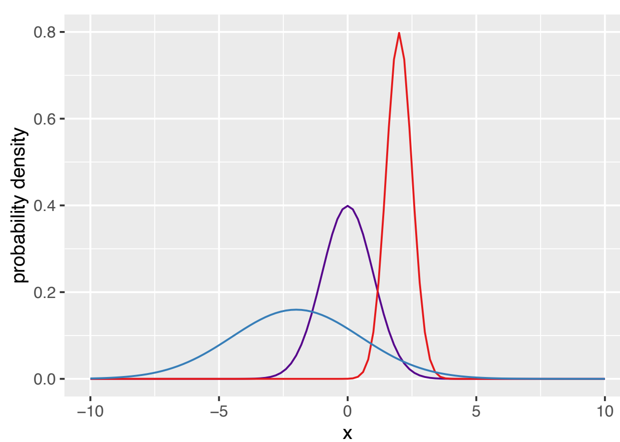
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$







Here they are plotted together to show the effects of the parameters.



This really shows how μ is "location" and σ is "spread".

They are all normal distributions!

Three reasons the normal distribution is important to frequentist statistics

1. It is very common.

A large number of natural and psychological phenomena are distributed normally. The saying is that if ancient civilizations knew about the normal distribution, they would have built monuments to it, because it is such a fundamental aspect of nature.

2. It arises under repeated sampling.

Our goal as scientists is to study populations, and we do that by sampling. As we will see later today, if you sample from a single population repeatedly, the distribution of the sample means will typically be a normal distribution (this is due to the Central Limit Theorem, which we won't see until later). So, the normal distribution is fundamental to frequentist statistics.

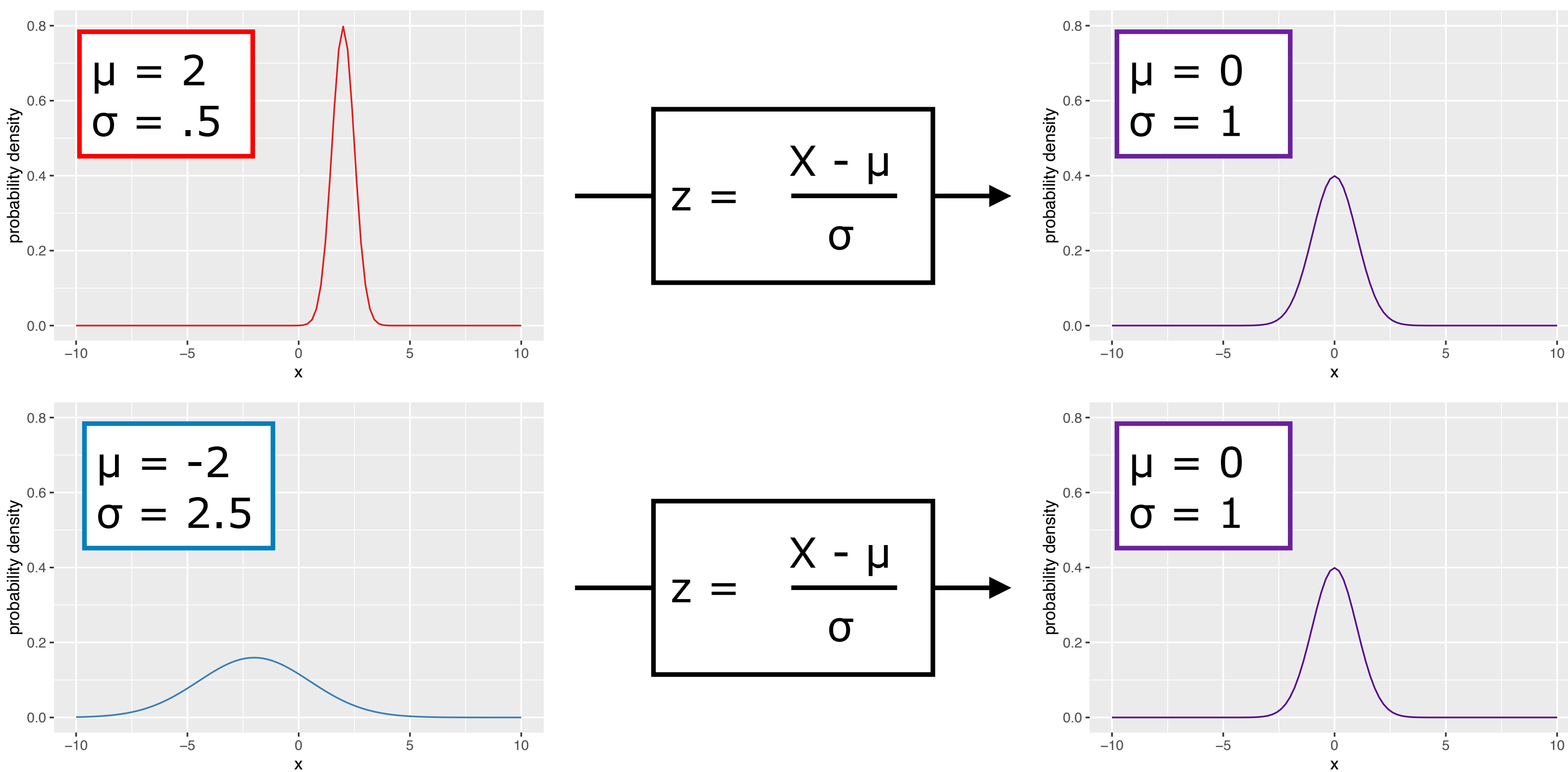
3. It has nice mathematical properties.

It is a symmetric distribution. It is determined by only two parameters (mean and standard deviation). And, as we will see next, we can calculate percentile ranks (and quantiles more generally) easily using z-scores.

The Standard Normal Distribution

If you take any normal distribution (any mean and standard deviation), and z-score transform the scores, the result is a normal distribution with a mean of 0 and a standard deviation of 1. We call it the **Standard Normal Distribution**.

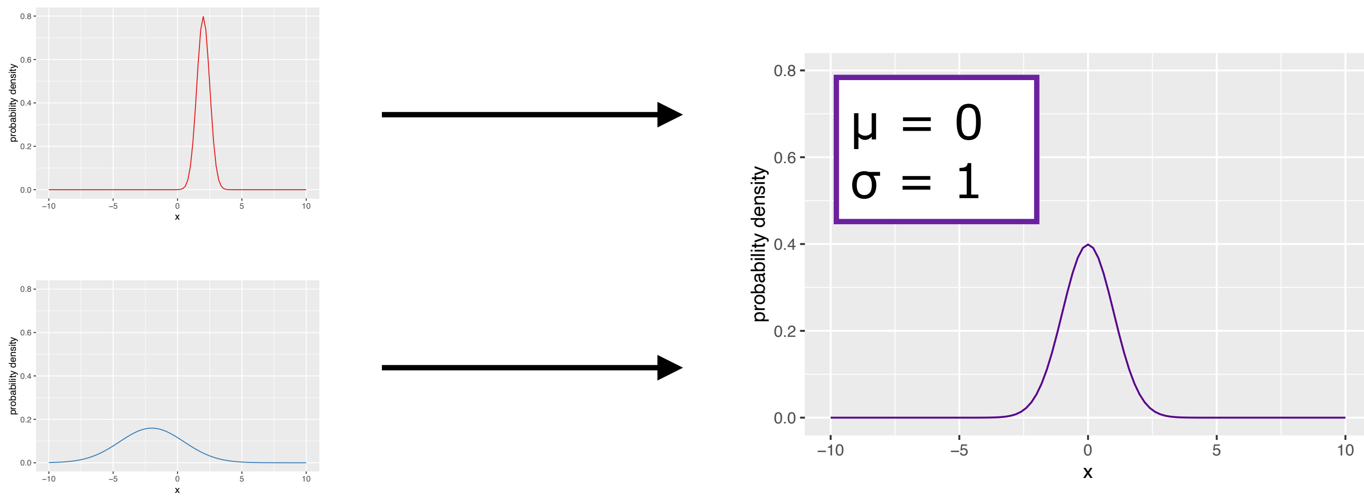
This follows from the properties of the z-score transformation: it always results in a mean of 0 and a standard deviation of 1.



2. Using z-scores and the standard normal distribution to quickly identify percentile ranks

This makes our lives easier, if we want

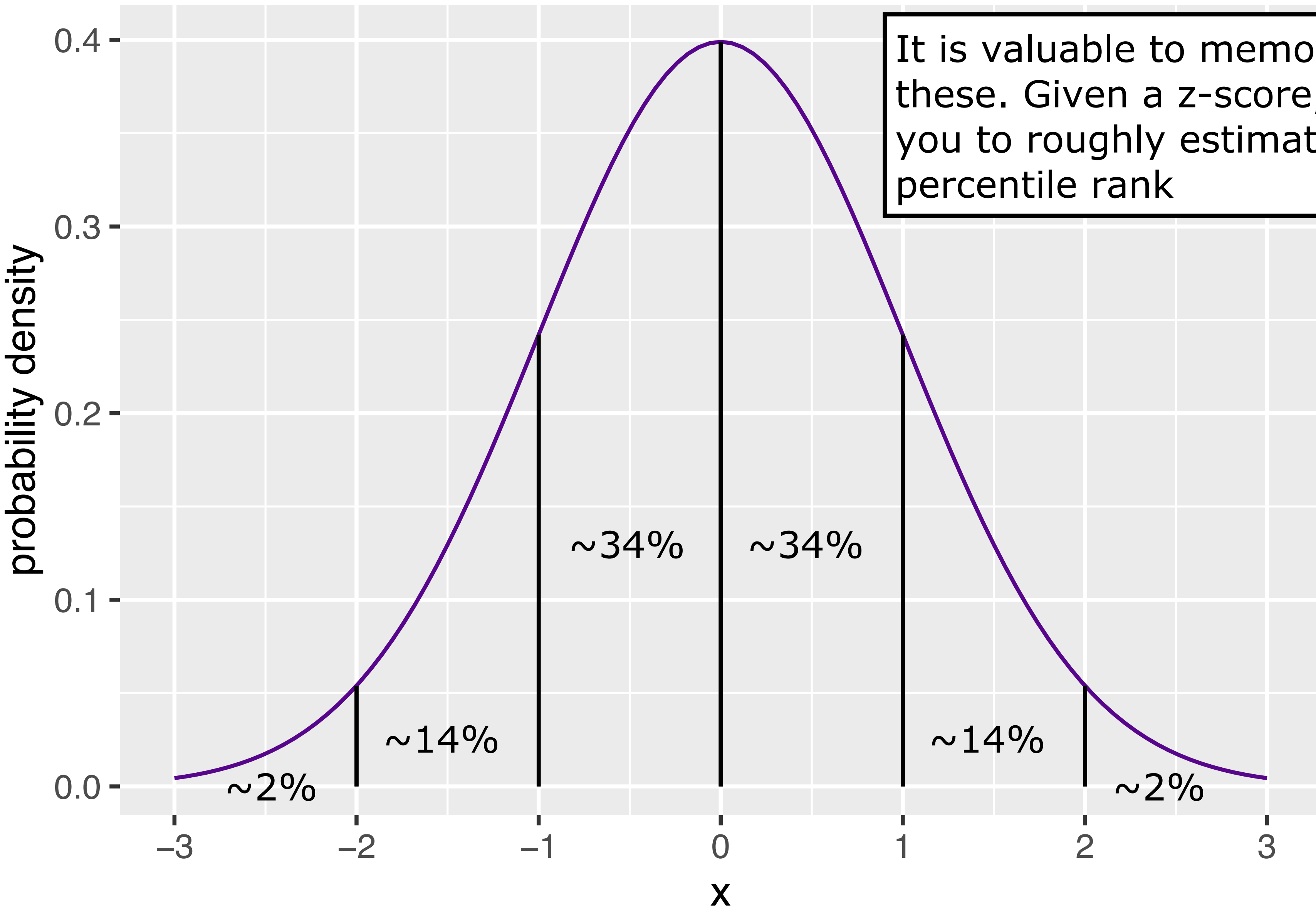
The fact that every normal distribution can be converted to a standard normal distribution (while preserving the structure of the distribution) means that, for all phenomena involving a normal distribution, we really only need to ever work with one distribution: the standard normal distribution.



This means that we can calculate all of the properties of this one distribution, print it in a textbook (like Appendix A of ours), and then use that for any study that we do with normally distributed phenomena.

This was a major advantage in the early 20th century when frequentist statistics was developed (because we didn't have computers to do the calculations). It is less of a benefit now, but there are conceptual advantages - you can evaluate studies quickly in your mind without doing any calculations.

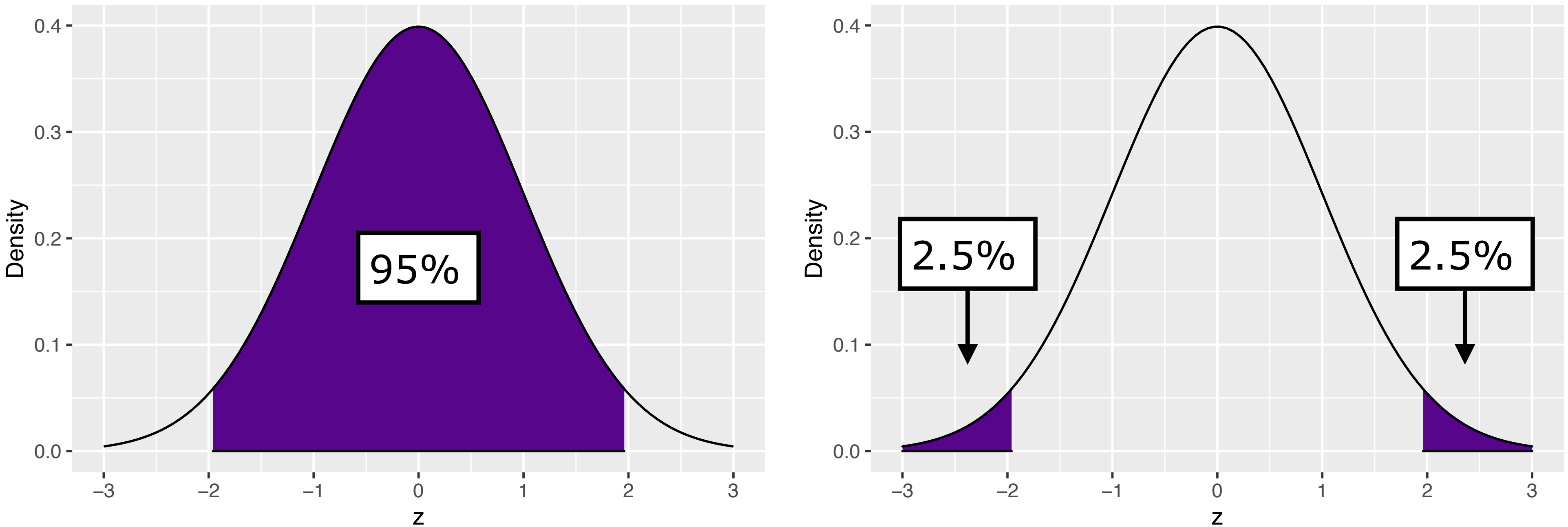
The percentage of scores at each z-score



It is valuable to memorize these. Given a z-score, it allows you to roughly estimate the percentile rank

A famous z-score: +/-1.96

Here is another z-score worth memorizing: +1.96 and its negative -1.96. Together, these two identify the 95% of scores in the middle of the distribution.



Another way to look at this is that +/- 1.96 z-scores identifies the scores in the two tails that together add up to be the most extreme 5% of scores (2.5% in each tail).

Using table A.1 in the textbook

Appendix A STATISTICAL TABLES

Note: All of the entries in the following tables were computed by the author, except where otherwise indicated.

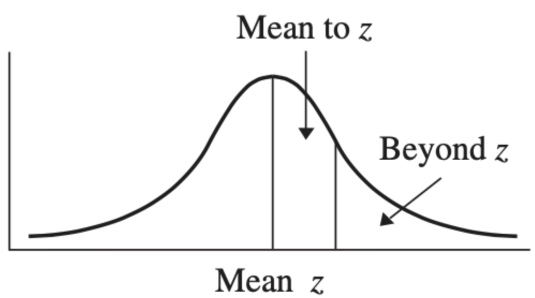


Table A.1
Areas Under the Standard Normal Distribution

z	Mean to z	Beyond z	z	Mean to z	Beyond z
.00	.0000	.5000	.41	.1591	.3409
.01	.0040	.4960	.42	.1628	.3372
.02	.0080	.4920	.43	.1664	.3336
.03	.0120	.4880	.44	.1700	.3300
.04	.0160	.4840	.45	.1736	.3264
.05	.0199	.4801	.46	.1772	.3228
.06	.0239	.4761	.47	.1808	.3192
.07	.0279	.4721	.48	.1844	.3156
.08	.0319	.4681	.49	.1879	.3121
.09	.0359	.4641	.50	.1915	.3085
.10	.0398	.4602	.51	.1950	.3050
.11	.0438	.4562	.52	.1985	.3015
.12	.0478	.4522	.53	.2019	.2981
.13	.0517	.4483	.54	.2054	.2946

You can use the values in this table plus basic arithmetic to calculate the percentage of scores either less than or greater than any given z.

Using R - the normal distribution functions

R has built-in functions for various distributions, including the normal distribution. This replaces the need for tables in books… with more flexibility.

The default options for these assume the standard normal distribution, but you can specify other normal distributions by changing the mean and sd arguments.

distribution function: pnorm(q, mean = 0, sd = 1)

If you enter a score (q), it will tell you the percentile rank (percentage of scores less than or equal to the z-score).

quantile function: qnorm(p, mean = 0, sd = 1)

If you enter percentile (p), it will tell you the score that divides the distribution at that percentile

random generator: rnorm(n, mean = 0, sd = 1)

This will randomly sample n scores from a normal distribution.

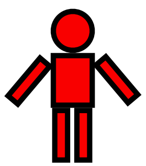
Evaluating a sample relative to a distribution of samples

(Up until now, we've been focusing on individuals. But evaluating samples is the core of what we do as scientists.)

Evaluating individuals

Up until now, we've been evaluating individual scores against distributions of individual scores — asking how common or rare an individual score is based on the distribution of scores (e.g., percentile rank). We know two ways to do it:

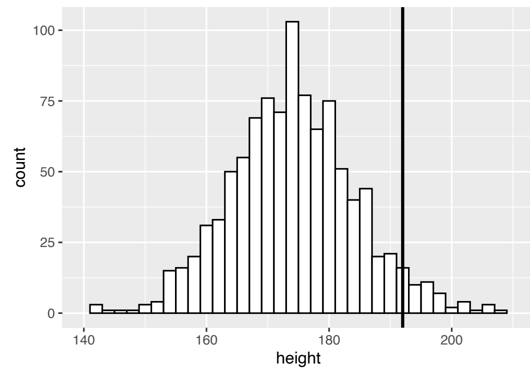
option 1: calculate directly



= 192cm

individual score
to evaluate

distribution of individual scores



use R functions

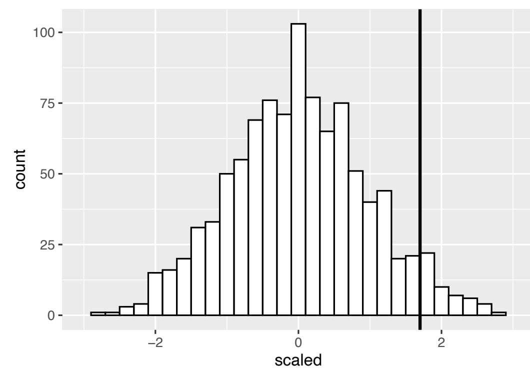
If you know the mean and standard deviation, you can feed them into pnorm().

pnorm(192, mean=175, sd=10) = .955

option 2: z-scores

$$z = \frac{192 - 175}{10} = 1.7$$

standard normal



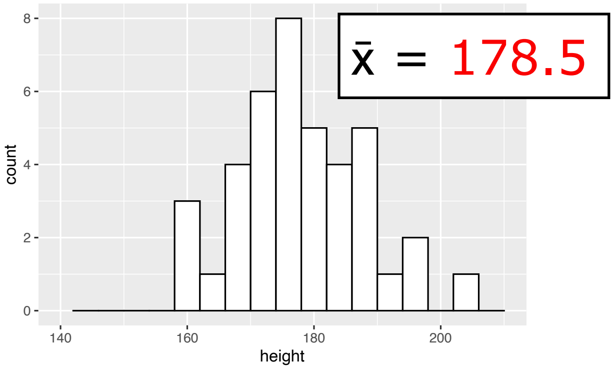
use table A.1

Look up the percentile rank of the z-score 1.7 = .955

Evaluating samples (i.e., groups)

In science, we rarely want to evaluate individuals. Instead, we want to evaluate the samples that we have collected. We want to ask how common or rare the sample measurement is. Let's develop methods to do that!

option 1: calculate directly



sample statistic to evaluate

distribution of sample means

?

use R functions

If you know the mean and standard deviation, you can feed them into pnorm().

pnorm(178.5, mean = **?**, sd = **?**)

option 2: z-scores

z = ?

standard normal

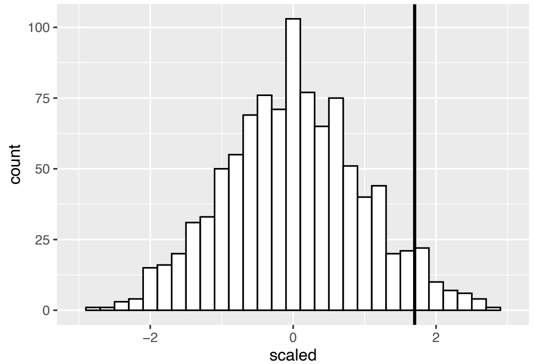


table A.1

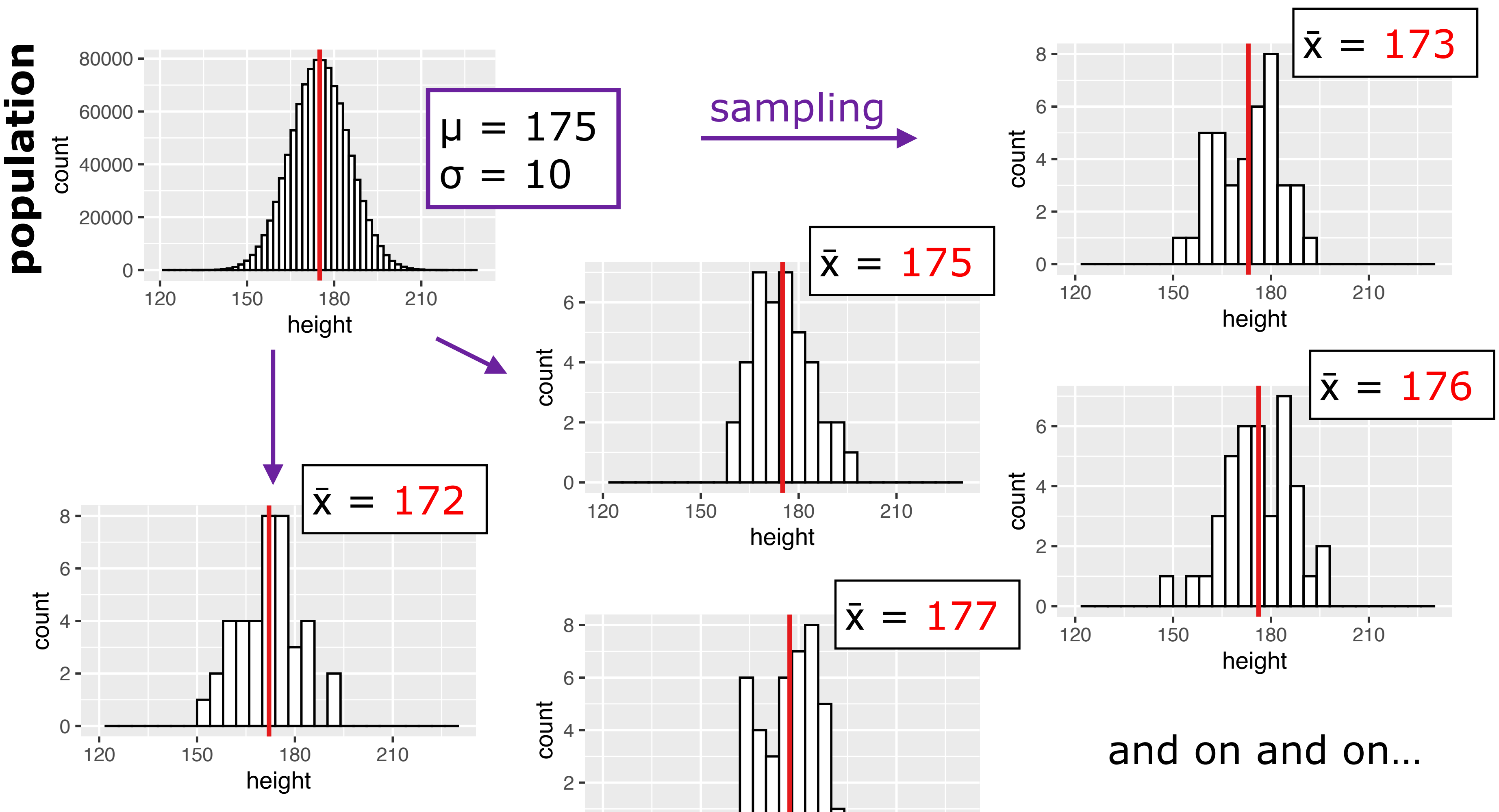
Look up the percentile rank of the z-score 1.7 = .955

The distribution of sample means

How would we create a **distribution of sample means**?

Step 1: Choose a sample size that we are interested in. e.g., 40

Step 2: Select a very large (ideally infinite) number of samples of that size from the population of individuals.



The distribution of sample means

How would we create a **distribution of sample means**?

Step 1: Choose a sample size that we are interested in. e.g., 40

Step 2: Select a very large (ideally infinite) number of samples of that size from the population of individuals.

Step 3: Create a distribution from all of the sample means.

population **sampling distribution**

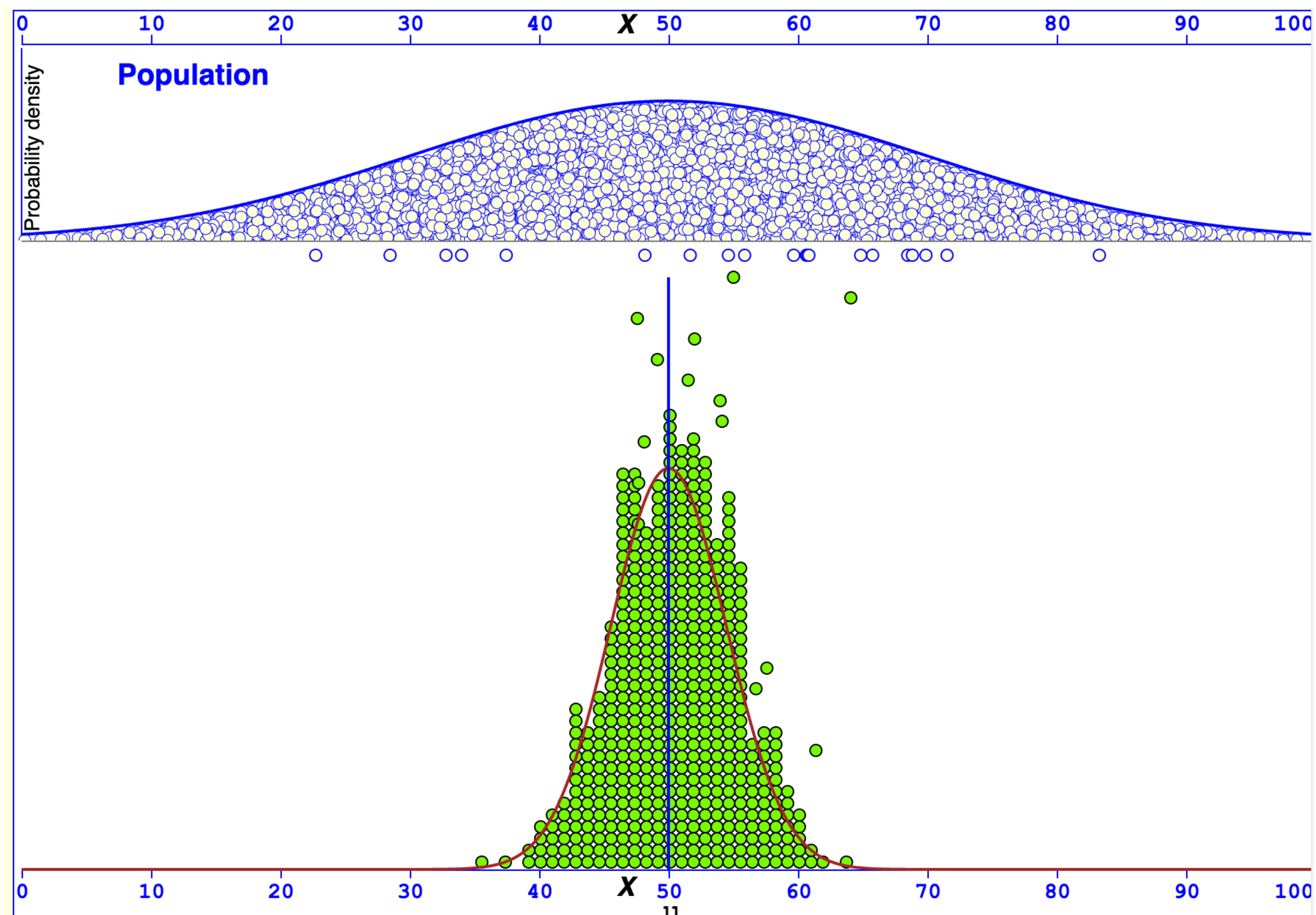
The result is a distribution of sample means. This distribution gets a special name. It is called **the sampling distribution of the mean**. (In principle, you can have sampling distributions of any statistic: mean, median, standard deviation, etc. But this one is the foundation of frequentist statistics.)

An animation of the sampling distribution of the mean

This is an interactive chart that simulates sampling repeatedly from a population, calculating a mean for each sample, and drawing the resulting distribution.

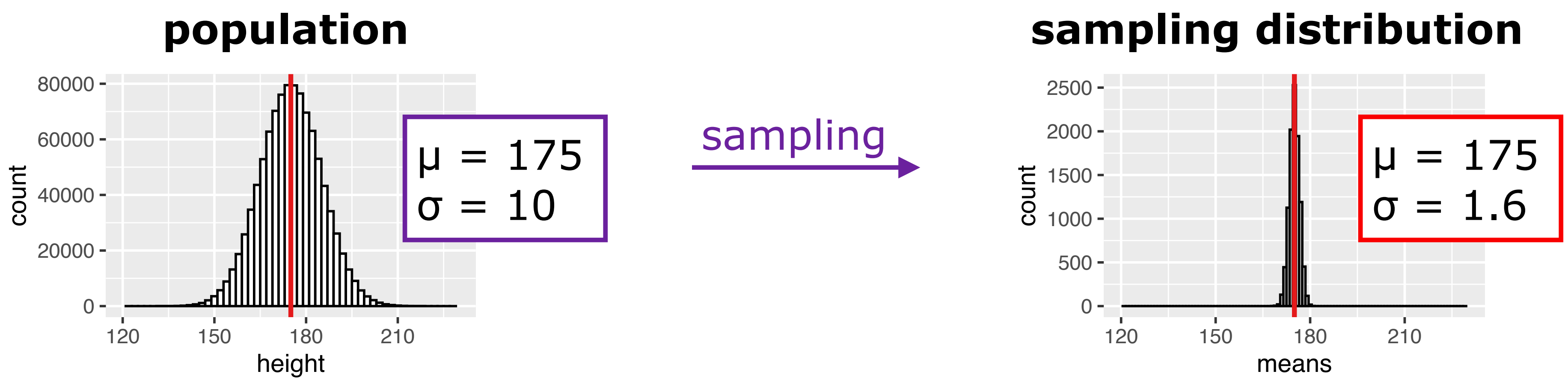
https://www.esci-dances.thenewstatistics.com/

This will allow us to watch the sampling distribution of the mean build up!



The mean of the sampling distribution of the mean

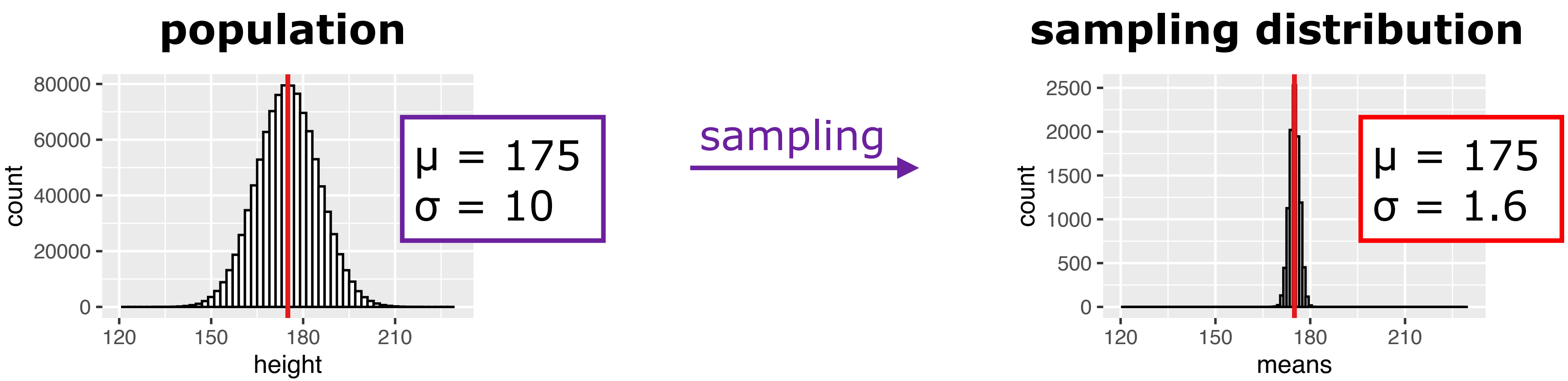
The mean of the sampling distribution of the mean is the same as the population mean.



This follows from the process of random sampling that is used to form the sampling distribution of the mean (and the fact that the mean is an unbiased estimator).

The standard deviation of the sampling distribution of the mean = standard error

The standard deviation of the <u>sampling distribution of the mean</u> is smaller than the standard deviation of the population. We can see this if we plot the two distributions on the same x-axis scale.



The standard deviation of the sampling distribution is an important quantity, so it gets its own name - **the standard error of the mean**, or just standard error for short. It also has its own symbol, which is the typical standard deviation (of a population) symbol with the subscript for sample means.

$\sigma_{\bar{x}}$

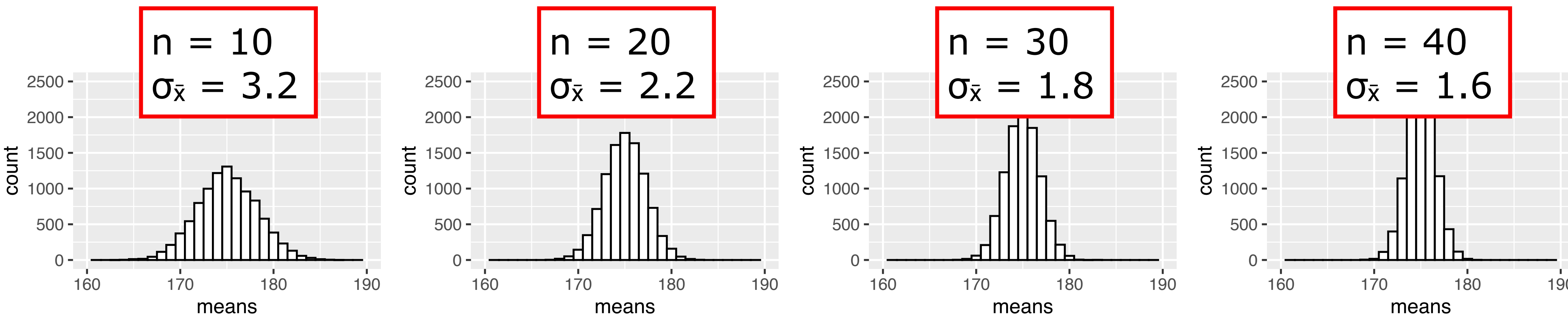
The standard error is proportional to the sample size

2

The size of the standard error of the mean is proportional to the size of the sample (n). Here is the equation:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

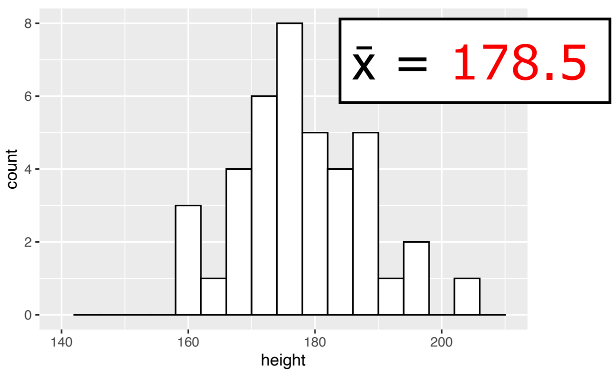
We can show this (inductively) by selecting different sample sizes, simulating the sampling distribution of the mean, and calculating the standard error of the simulated distributions:



Now we are ready to evaluate samples!

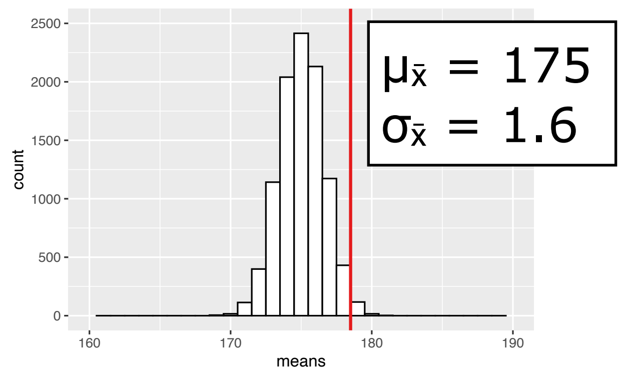
In science, we rarely want to evaluate individuals. Instead, we want to evaluate the samples that we have collected. We want to ask how common or rare the sample measurement is. Let's develop methods to do that!

option 1: calculate directly



sample statistic to evaluate

sampling distribution



use R functions

If you know the mean and standard error, you can feed them into pnorm().

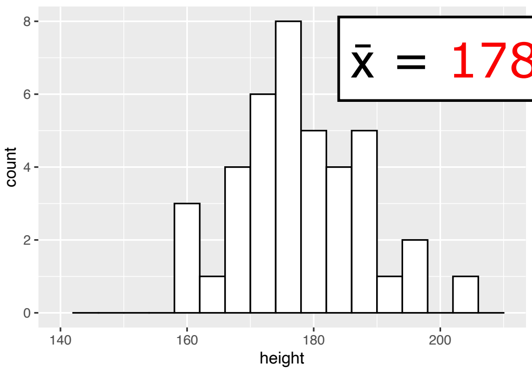
pnorm(178.5, mean = 175, sd = 1.6) = .986

Now we are ready to evaluate samples!

In science, we rarely want to evaluate individuals. Instead, we want to evaluate the samples that we have collected. We want to ask how common or rare the sample measurement is. Let's develop methods to do that!

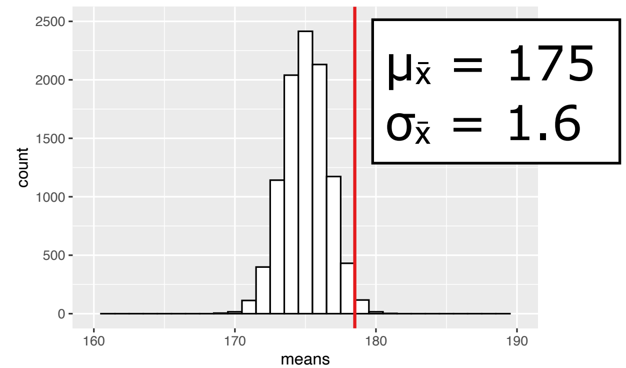
---

option 1: calculate directly



sample statistic to evaluate

sampling distribution



use R functions

If you know the mean and standard error, you can feed them into pnorm().

pnorm(178.5, mean = 175, sd = 1.6) = .986

---

option 2: z-scores

z = ?

standard normal

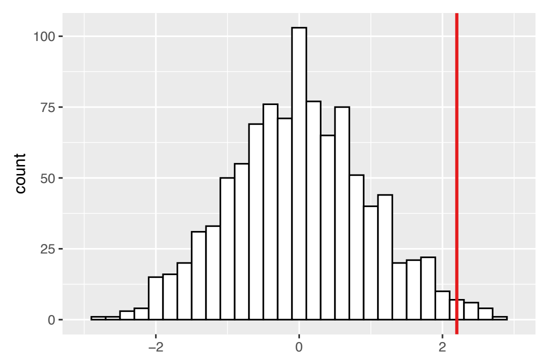


table A.1

Look up the percentile rank of the z-score 2.2 = .986

z-scores for groups (based on the sampling distribution of the mean) 2

If we want to use the z-score method for evaluating a sample mean, we need to adjust the z-score formula a bit. It will have the same logic - mean centering and scaling:

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

But notice that we made two changes.

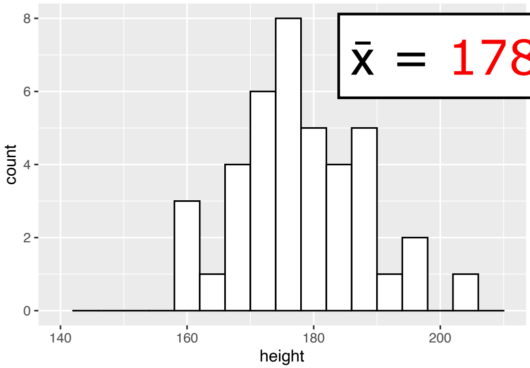
First, this is about the sample mean, so we substituted in the sample mean for the raw score.

Second, because we are working with samples that are part of the sampling distribution of the mean, the standard deviation is the standard deviation of the sampling distribution of the mean. In other words, the standard error.

Now we are ready to evaluate samples!

In science, we rarely want to evaluate individuals. Instead, we want to evaluate the samples that we have collected. We want to ask how common or rare the sample measurement is. Let's develop methods to do that!

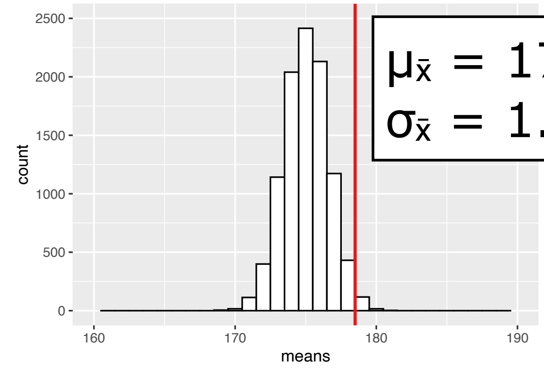
option 1: calculate directly



$\bar{x}$ = 178.5

sample statistic
to evaluate

sampling distribution



$\mu_{\bar{x}}$ = 175
$\sigma_{\bar{x}}$ = 1.6

use R functions

If you know the mean and standard error, you can feed them into pnorm().

pnorm(178.5, mean = 175, sd = 1.6) = .986

option 2: z-scores

$$z = \frac{178.5 - 175}{1.6} = 2.2$$

standard normal

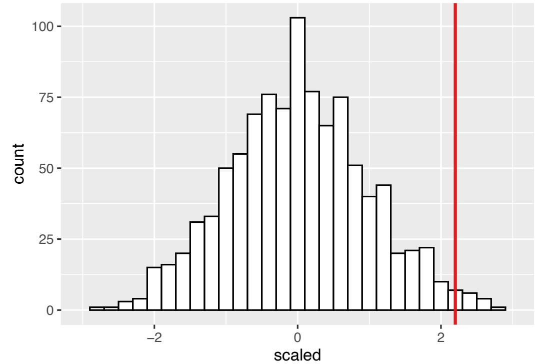


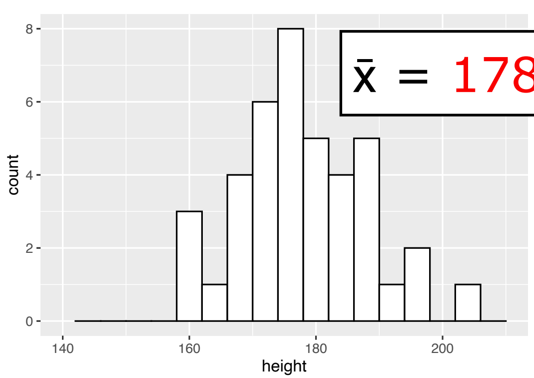
table A.1

Look up the percentile rank of the z-score 2.2 = .986

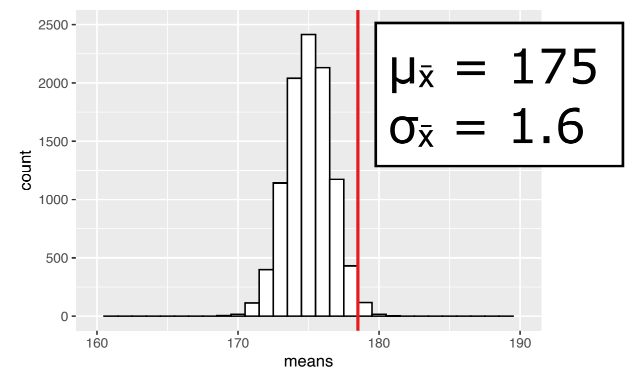
What did we learn?

We learned that a sample with a mean of 178.5cm that was drawn from a distribution with a mean of 175cm and a standard deviation of 10cm is more extreme than 98.6% of possible samples.

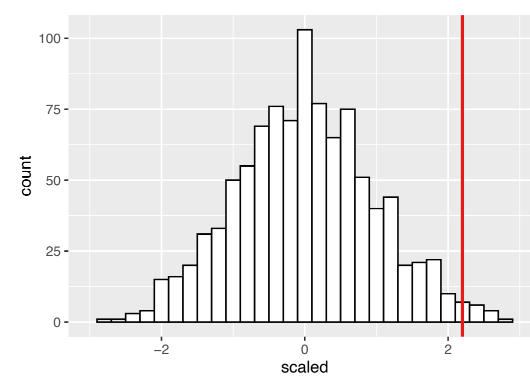
sample to test



sampling distribution



standard normal



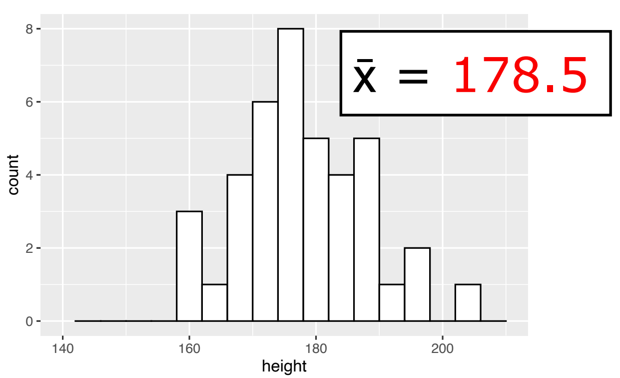
$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = 2.2$$

What did we learn?

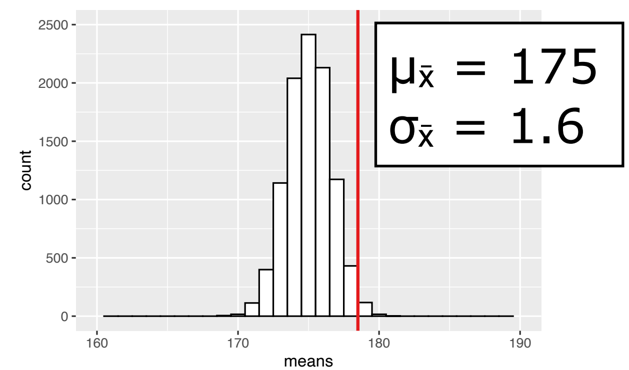
More generally, we learned two options for asking the following question: **Is the sample that we obtained extreme relative to a given population?**

This is our first step toward inferential statistics. The question we ask in inferential statistics is a version of this question. And the methods we will use are built on these methods!

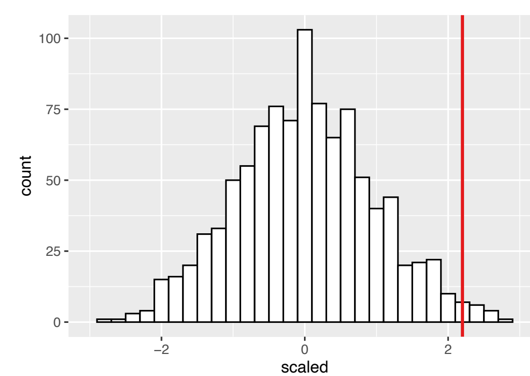
sample to test



sampling distribution



standard normal



$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = 2.2$$