Review Session 5

API-201, 10.15.21
Sophie Hill

Agenda

- Review midterm Q2 sensitivity analysis
- Sampling distributions
- Build a hypothesis test worksheet in Excel!

Midterm Q2

Midterm Q2

Calculate the range of vaccination rates that would indicate that getting the vaccine is effective at preventing the spread of COVID.

Midterm Q2

Calculate the range of vaccination rates that would indicate that getting the vaccine is effective at preventing the spread of COVID.

We originally assumed that the vaccination rate was 95%. Now we are going to let that vary, so let's call it X.

Midterm Q2

Calculate the range of vaccination rates that would indicate that getting the vaccine is effective at preventing the spread of COVID.

We originally assumed that the vaccination rate was 95%. Now we are going to let that vary, so let’s call it X.

What does it mean for the vaccine to be effective? It means your risk of infection should be lower if you have the vaccine than if you don’t.

Midterm Q2

Calculate the range of vaccination rates that would indicate that getting the vaccine is effective at preventing the spread of COVID.

We originally assumed that the vaccination rate was 95%. Now we are going to let that vary, so let's call it X.

What does it mean for the vaccine to be effective? It means your risk of infection should be lower if you have the vaccine than if you don't.

i.e. , $P(\text{COVID} \mid \text{vax}) < P(\text{COVID} \mid \text{unvax})$

Midterm Q2

Calculate the range of vaccination rates that would indicate that getting the vaccine is effective at preventing the spread of COVID.

We originally assumed that the vaccination rate was 95%. Now we are going to let that vary, so let's call it *X*.

What does it mean for the vaccine to be effective? It means your risk of infection should be lower if you have the vaccine than if you don't.

i.e. , $P(\text{COVID} \mid \text{vax}) < P(\text{COVID} \mid \text{unvax})$

Let's set this to be equal to find the threshold, and then check whether the range will be above or below that threshold.

Midterm Q2

Calculate the range of vaccination rates that would indicate that getting the vaccine is effective at preventing the spread of COVID.

We originally assumed that the vaccination rate was 95%. Now we are going to let that vary, so let's call it *X*.

What does it mean for the vaccine to be effective? It means your risk of infection should be lower if you have the vaccine than if you don't.

i.e. , $P(\text{COVID} \mid \text{vax}) = P(\text{COVID} \mid \text{unvax})$

Let's set this to be equal to find the threshold, and then check whether the range will be above or below that threshold.

Midterm Q2

$$P(\text{COVID} \mid \text{vax}) = P(\text{COVID} \mid \text{unvax})$$

$$\frac{P(\text{COVID \& vax})}{P(\text{vax})} = \frac{P(\text{COVID \& unvax})}{P(\text{unvax})}$$

$$\frac{P(\text{COVID \& vax})}{X} = \frac{P(\text{COVID \& unvax})}{(1 - X)}$$

Midterm Q2

$$P(\mathrm{COVID} \mid \mathrm{vax}) = P(\mathrm{COVID} \mid \mathrm{unvax})$$

$$\frac{P(\mathrm{COVID}\ \&\ \mathrm{vax})}{P(\mathrm{vax})} = \frac{P(\mathrm{COVID}\ \&\ \mathrm{unvax})}{P(\mathrm{unvax})}$$

$$\frac{P(\mathrm{COVID}\ \&\ \mathrm{vax})}{X} = \frac{P(\mathrm{COVID}\ \&\ \mathrm{unvax})}{(1-X)}$$

$$\frac{P(\mathrm{vax} \mid \mathrm{COVID}) \cdot P(\mathrm{COVID})}{X} = \frac{P(\mathrm{unvax} \mid \mathrm{COVID}) \cdot P(\mathrm{COVID})}{(1-X)}$$

Midterm Q2

$$P(\text{COVID} \mid \text{vax}) = P(\text{COVID} \mid \text{unvax})$$

$$\frac{P(\text{COVID } \& \text{ vax})}{P(\text{vax})} = \frac{P(\text{COVID } \& \text{ unvax})}{P(\text{unvax})}$$

$$\frac{P(\text{COVID } \& \text{ vax})}{X} = \frac{P(\text{COVID } \& \text{ unvax})}{(1 - X)}$$

$$\frac{P(\text{vax} \mid \text{COVID}) \cdot P(\text{COVID})}{X} = \frac{P(\text{unvax} \mid \text{COVID}) \cdot P(\text{COVID})}{(1 - X)}$$

$$\frac{P(\text{vax} \mid \text{COVID})}{X} = \frac{P(\text{unvax} \mid \text{COVID})}{(1 - X)}$$

Midterm Q2

$$\frac{P(\text{vax} \mid \text{COVID})}{X} = \frac{P(\text{unvax} \mid \text{COVID})}{(1 - X)}$$

$$\frac{0.74}{X} = \frac{0.26}{(1 - X)}$$

$$0.74 - 0.74X = 0.26X$$

$$0.74 = X$$

Midterm Q2

We ended up with a value, 74%, that is the same as P(vax | COVID). How come?

Midterm Q2

We ended up with a value, 74%, that is the same as P(vax | COVID). How come?

If the vaccine is effective, then the vaccinated should be a *smaller* proportion of the infected than the population... i.e., the vaccinated should be a *larger* proportion of the population than the infected.

Midterm Q2

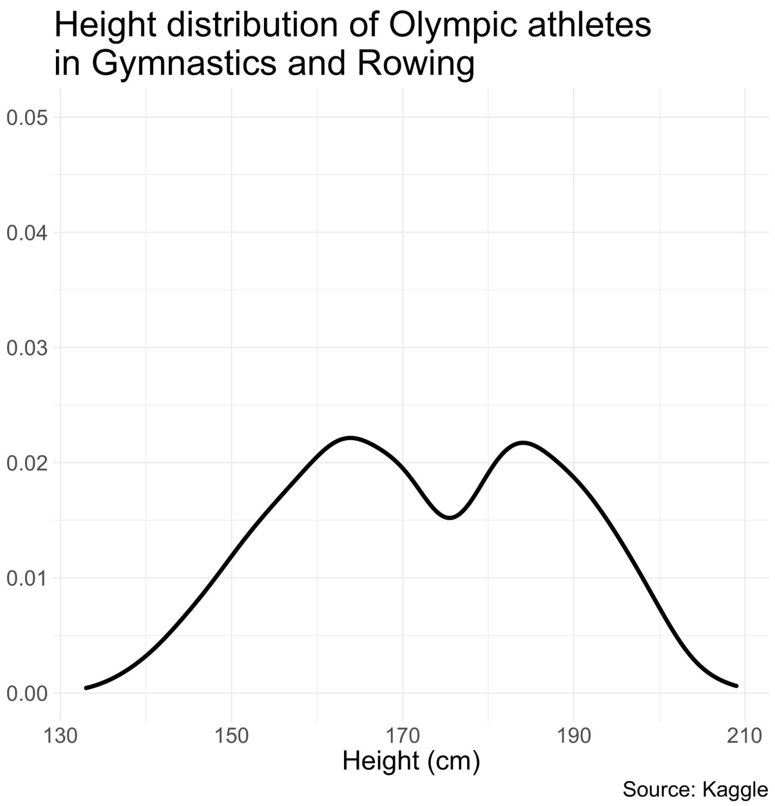
We ended up with a value, 74%, that is the same as P(vax | COVID). How come?

If the vaccine is effective, then the vaccinated should be a *smaller* proportion of the infected than the population… i.e., the vaccinated should be a *larger* proportion of the population than the infected.

So the vaccine is effective if the vaccination rate in the population is above 74% (and has no effect when it is equal to 74%).

Sampling distribution

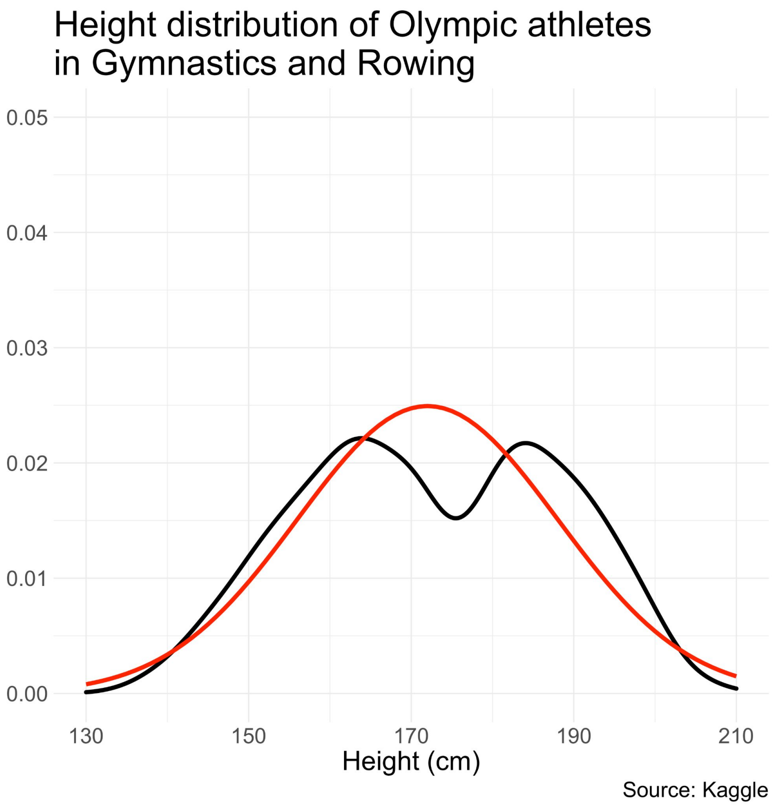
Distribution of height of Olympic athletes



Distribution of height of Olympic athletes

Not normally distributed!

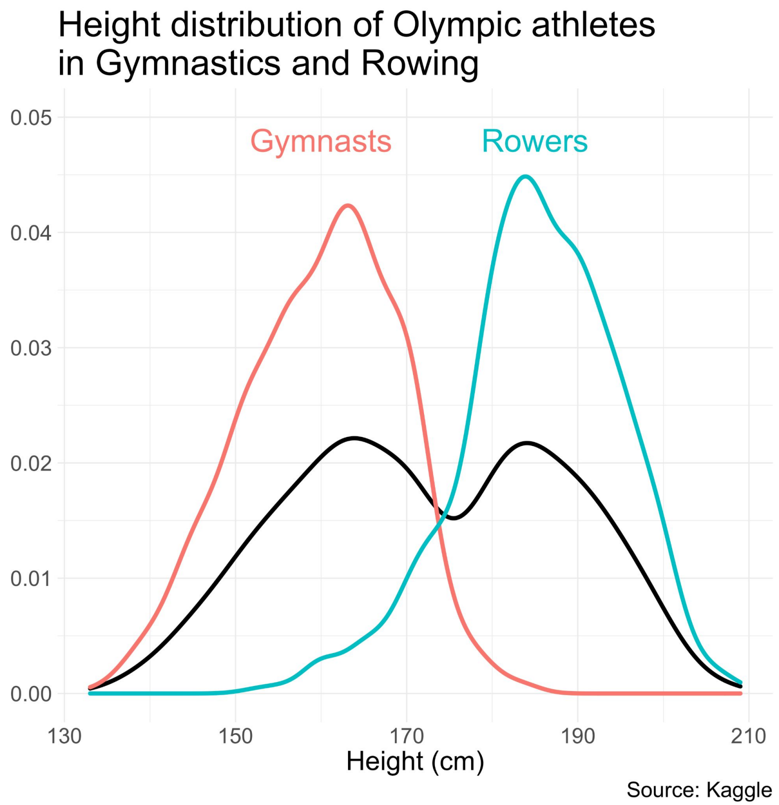
Why?



Distribution of height of Olympic athletes

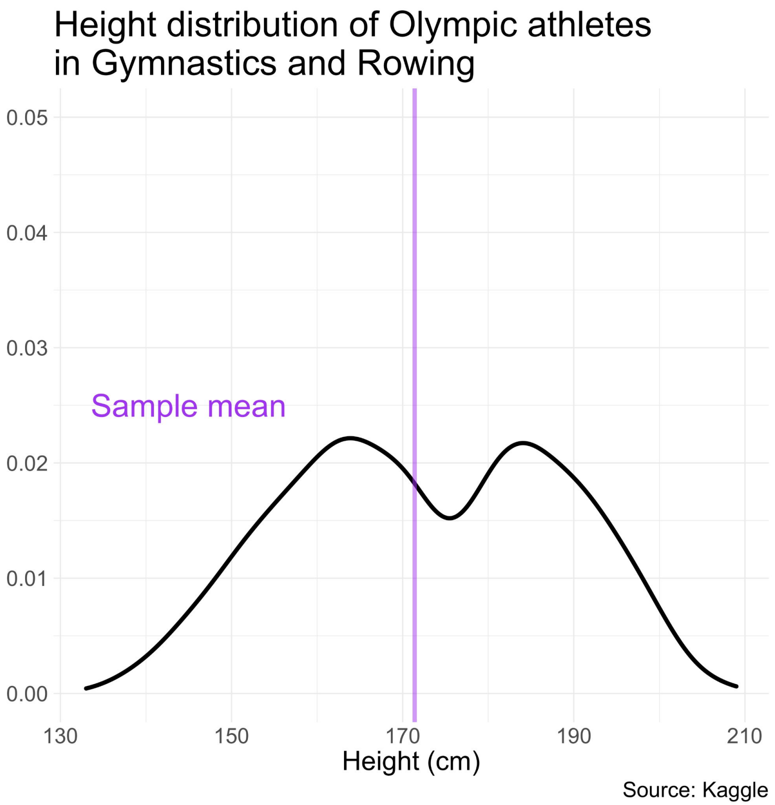
Not normally distributed!

Gymnasts tend to be much shorter than Rowers...



Distribution of height of Olympic athletes

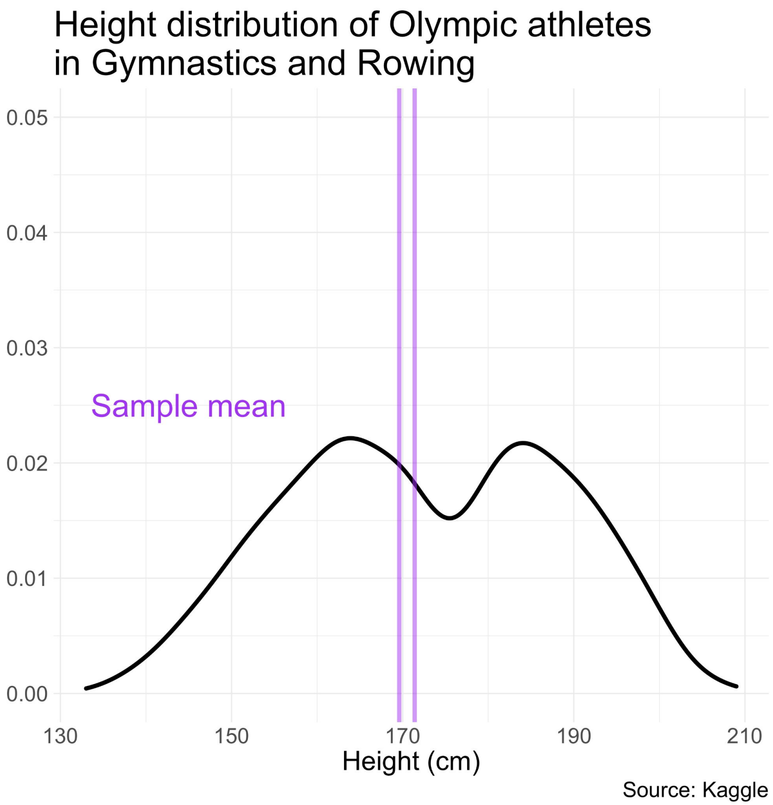
Now let's take a random sample of 10 athletes and calculate the sample mean.



Distribution of height of Olympic athletes

Now let's take a random sample of 10 athletes and calculate the sample mean…

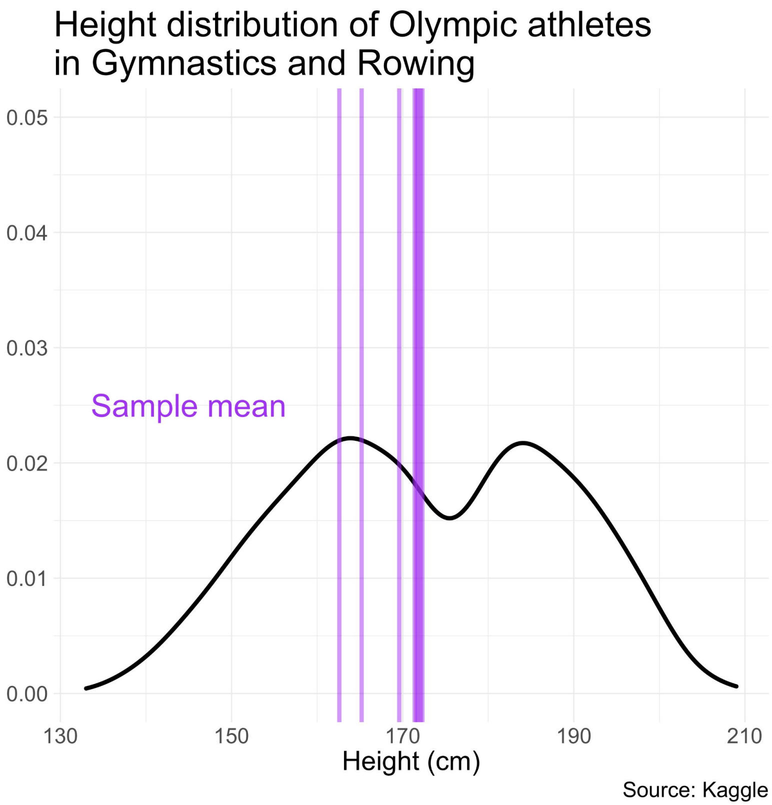
… and again.



Distribution of height of Olympic athletes

Now let's take a random sample of 10 athletes and calculate the sample mean…

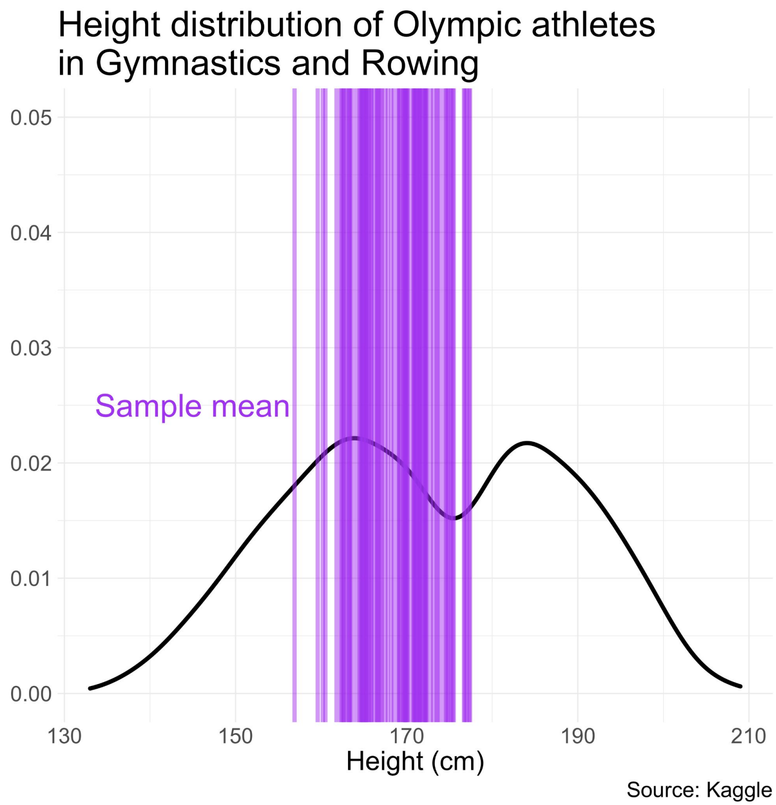
… and again.



Distribution of height of Olympic athletes

Now let's take a random sample of 10 athletes and calculate the sample mean…

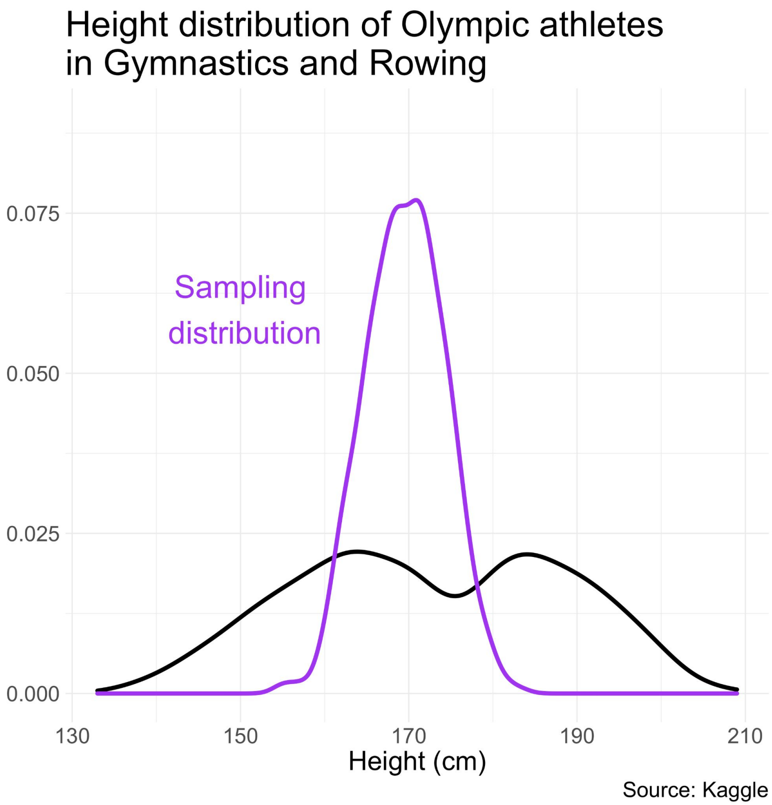
… and again.



Distribution of height of Olympic athletes

If we did this lots of times, we could plot the PDF of the sample means, i.e., the sampling distribution.

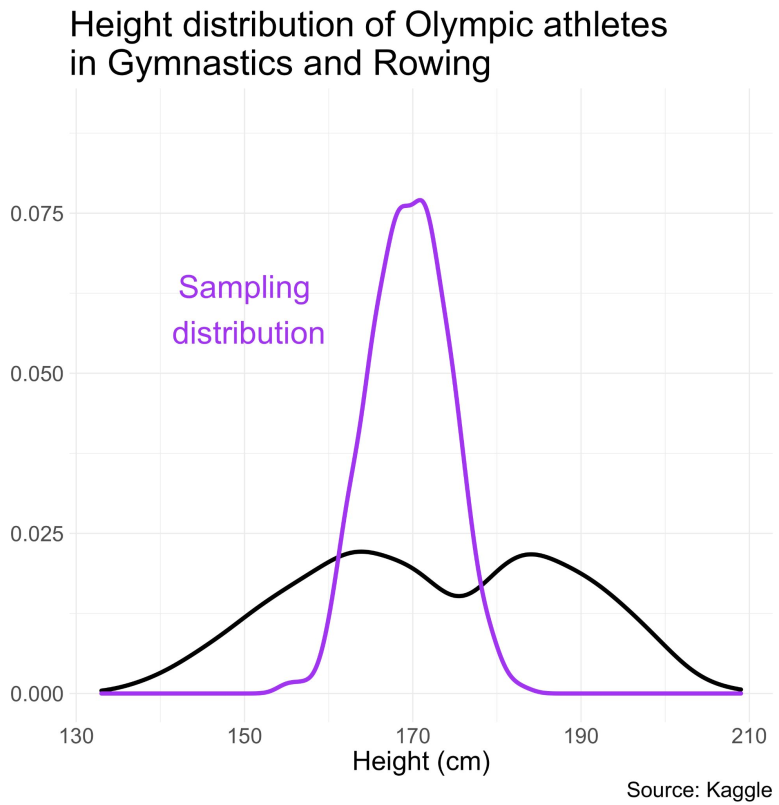
What do you notice about its shape, compared to the PDF of the raw data?



Distribution of height of Olympic athletes

Key takeaway:
The sampling distribution is approximately normal, even though the raw distribution is *not*.

This is the magic of the CLT!!



Hypothesis tests

Hypothesis tests

Conducting a hypothesis test for a sample proportion is actually pretty straightforward: once you've done one, you've done them all!

Hypothesis tests

Conducting a hypothesis test for a sample proportion is actually pretty straightforward: once you've done one, you've done them all!

Let's make a hypothesis test worksheet in Excel (link to template on Canvas).

HYPOTHESIS TESTING FOR PROPORTIONS

Inputs		
Null hypothesis	q =	0.50
Significance level	alpha =	0.05
Sample estimate	q-hat =	0.47
Sample size	n =	800

Outputs		
Does the CLT apply?		Yes
Standard error	SE =	0.0177
Z-score	Z =	-1.6971
p-value	p =	0.0897
Is p-value < alpha?		FALSE
Do we reject the null?		Fail to reject

FAQs

Why does the formula for the p-value look like that??

We told you that you can calculate a p-value from a Z-score in Excel using this formula:

$$= 2 * (1 - NORM.S.DIST(ABS(Z),TRUE))$$

Why does the formula for the p-value look like that??

We told you that you can calculate a p-value from a Z-score in Excel using this formula:

$$= 2 * (1 - NORM.S.DIST(ABS(Z),TRUE))$$

But what does it actually mean??

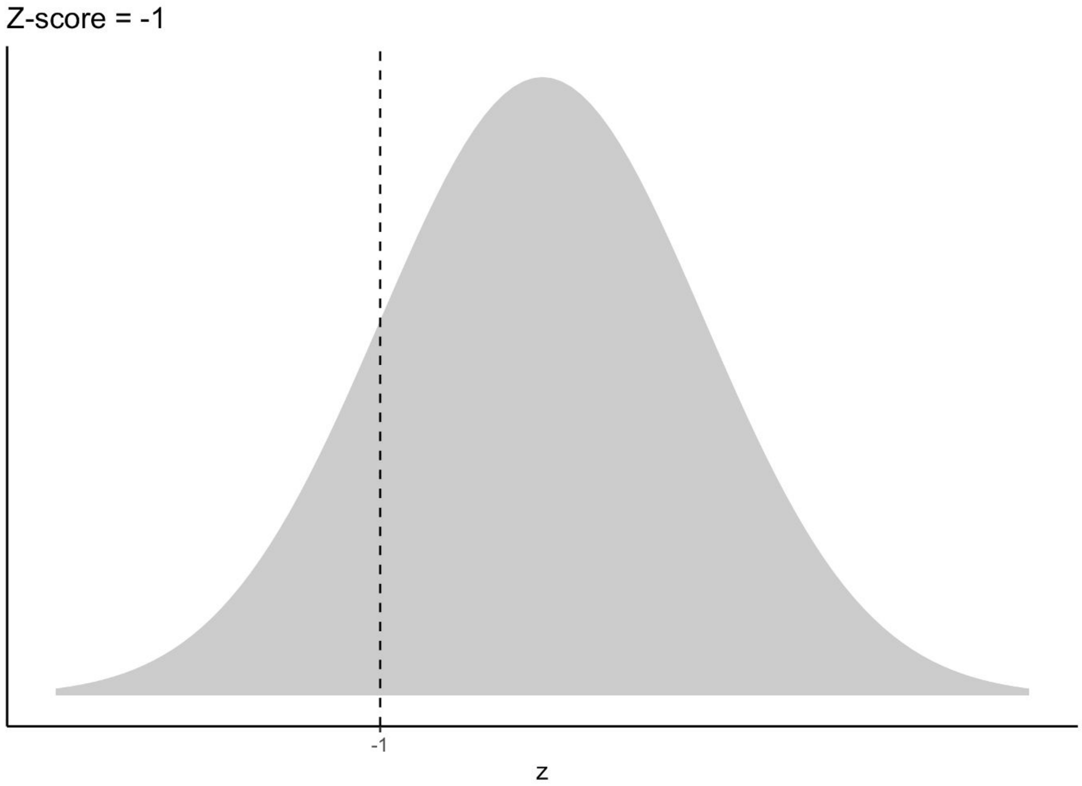
Why does the formula for the p-value look like that??

We told you that you can calculate a p-value from a Z-score in Excel using this formula:

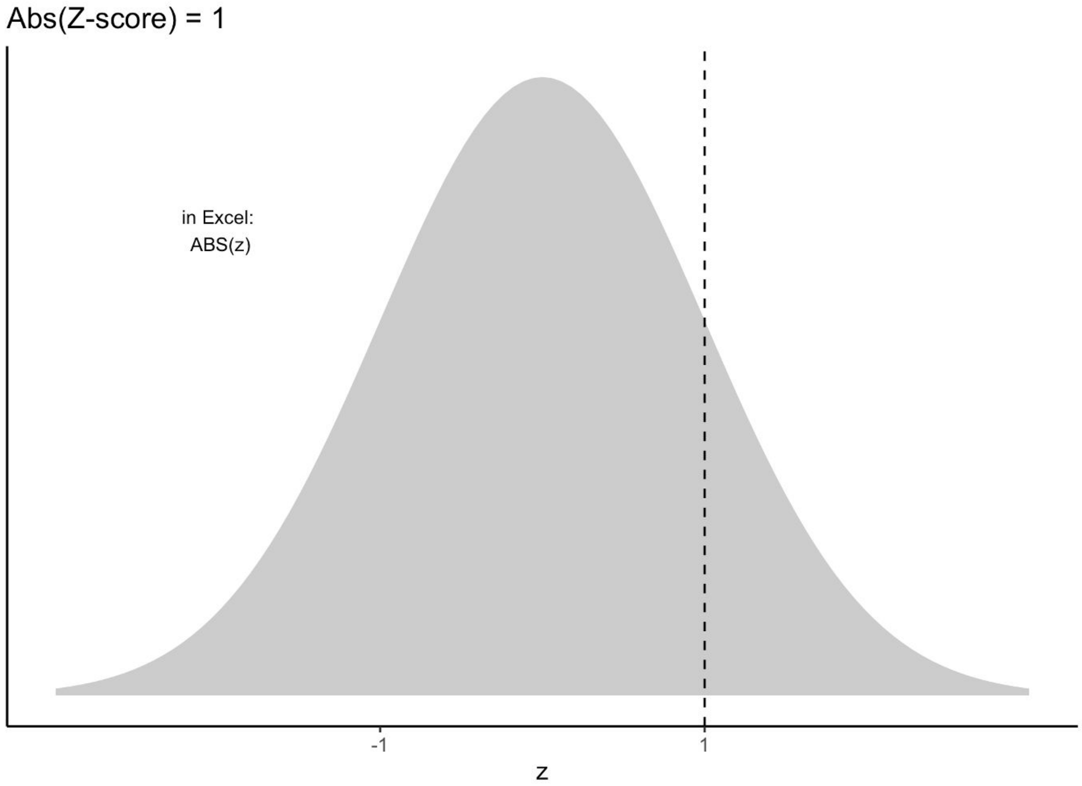
$$= 2 * (1 - NORM.S.DIST(ABS(Z), TRUE))$$

But what does it actually mean?? Let's break it down step-by-step.

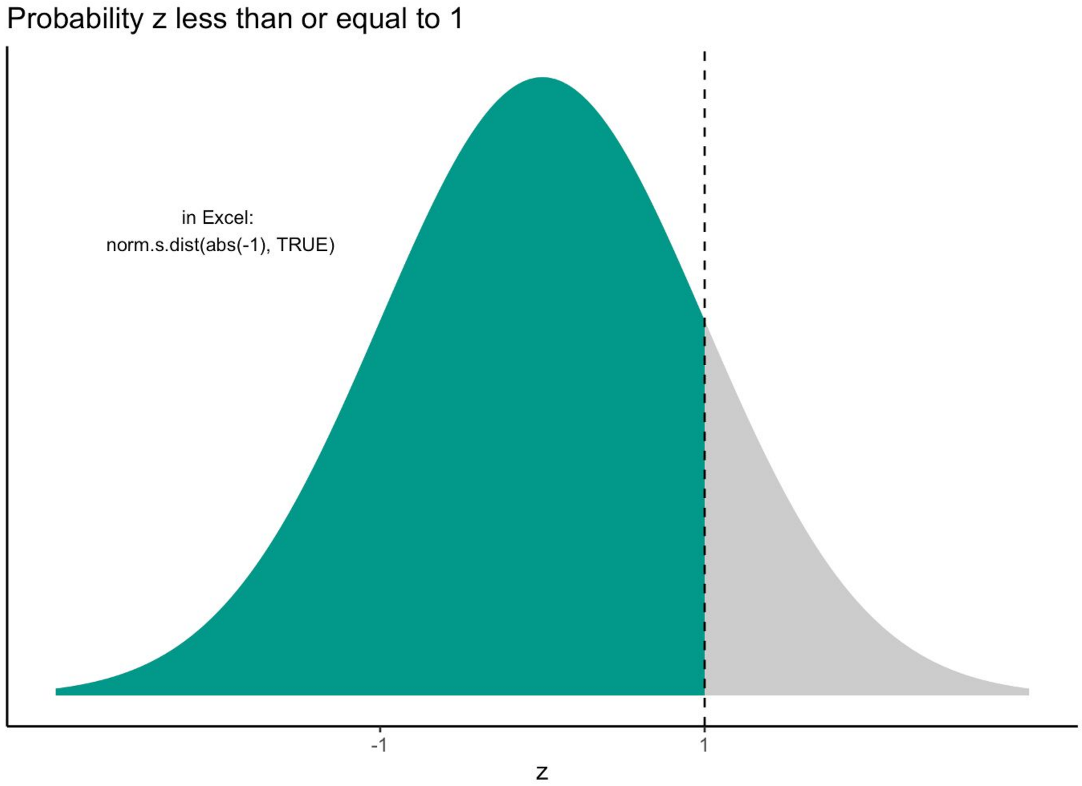
Why does the formula for the p-value look like that??



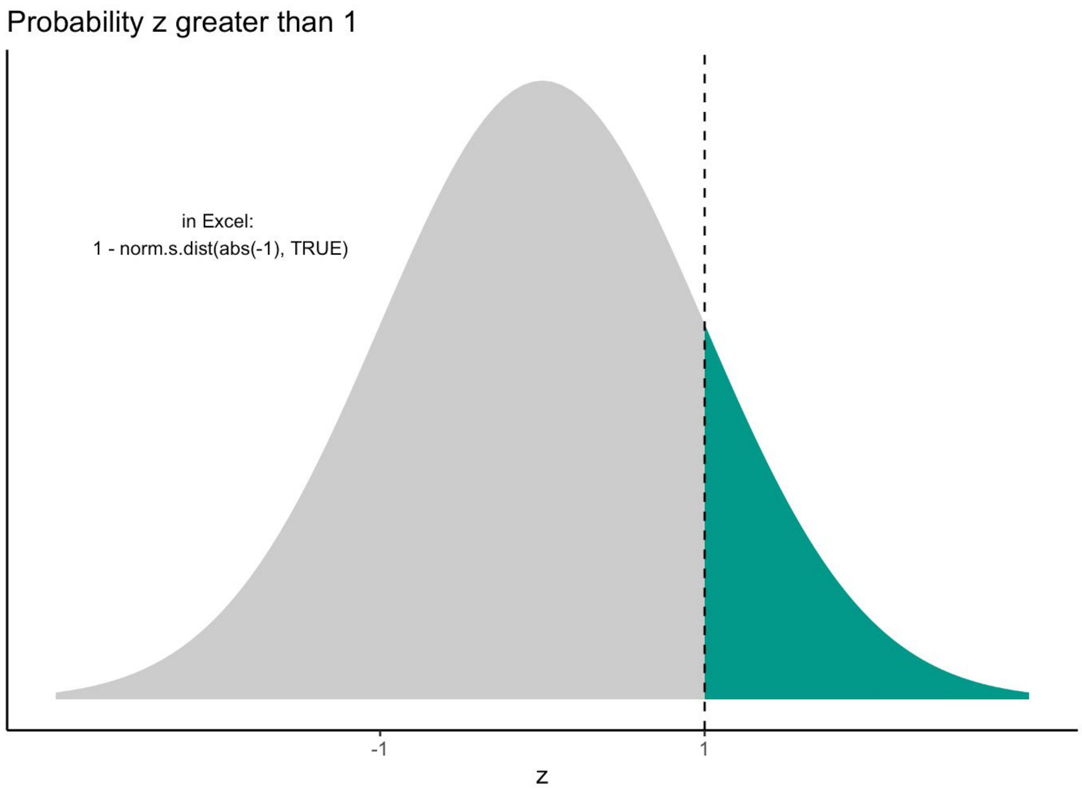
Why does the formula for the p-value look like that??



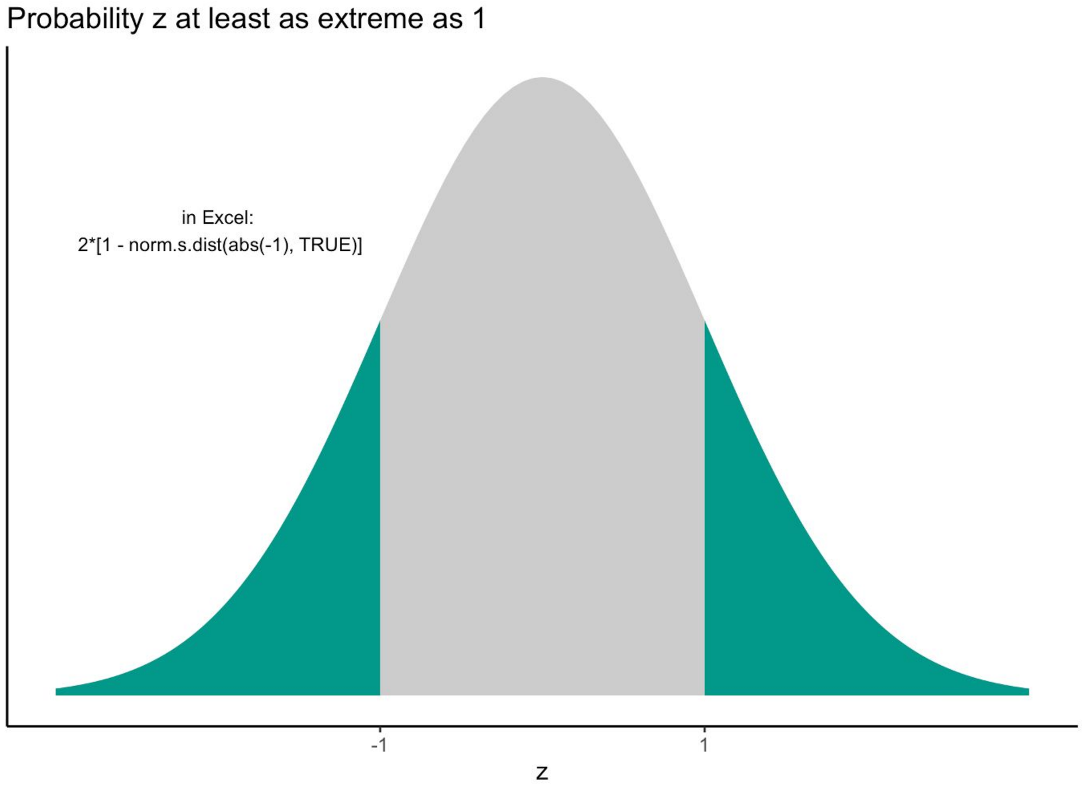
Why does the formula for the p-value look like that??



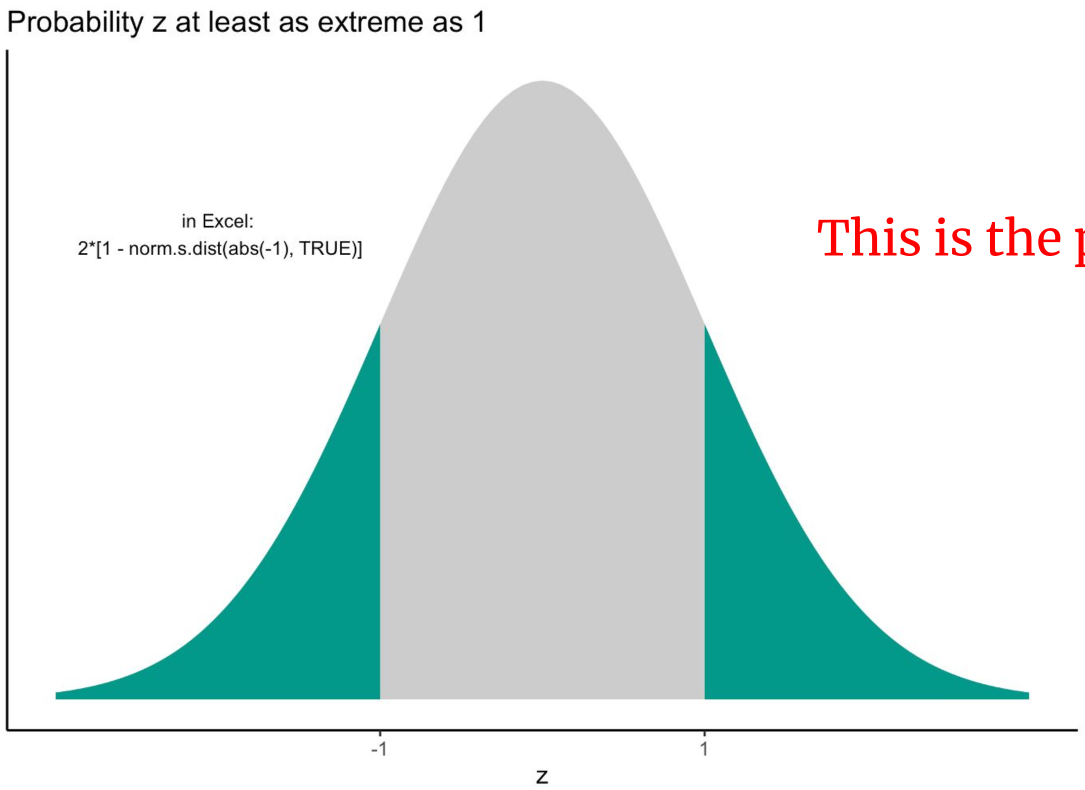
Why does the formula for the p-value look like that??



Why does the formula for the p-value look like that??



Why does the formula for the p-value look like that??



This is the p-value!!

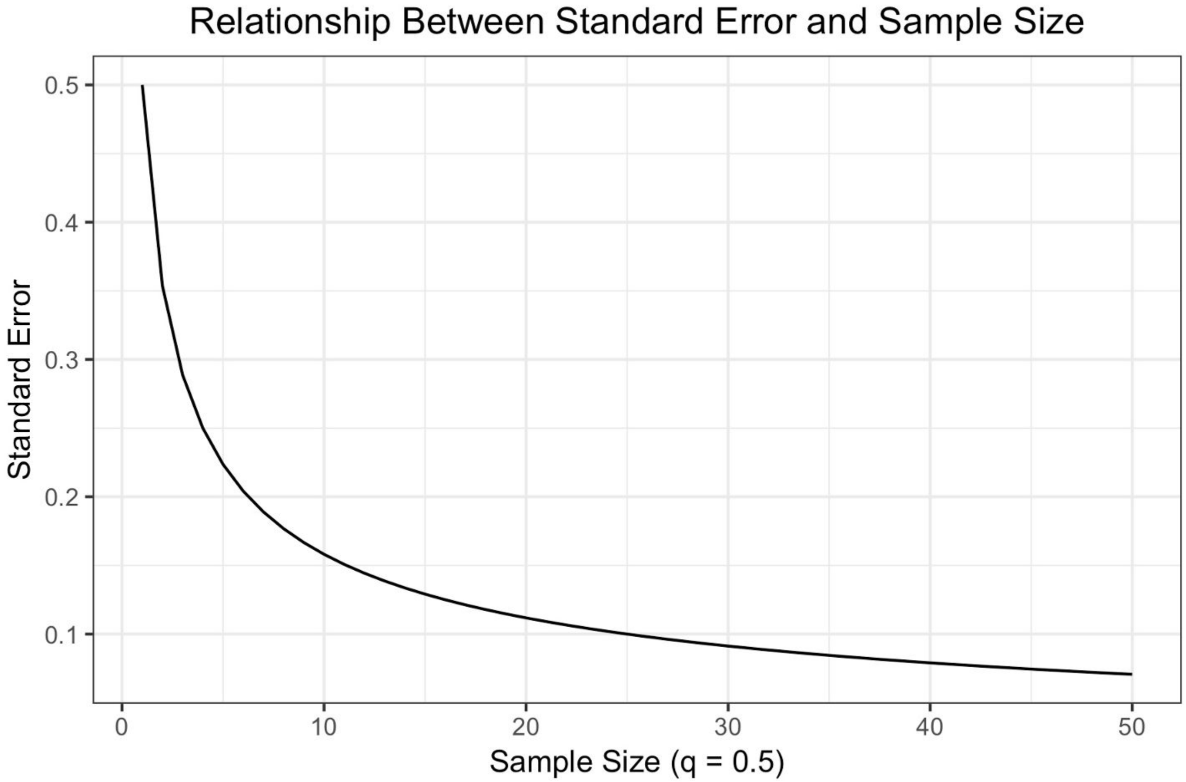
How does the proportion and sample size affect the SE?

$$SE = \sqrt{\frac{q * (1 - q)}{n}}$$

What happens to SE as n increases (holding q fixed)?

What happens to SE as q increases (holding n fixed)?

How does the proportion and sample size affect the SE?



How does the proportion and sample size affect the SE?