Comparing Samples: the logic of hypothesis testing

Research Methods for Political Science

Thomas Chadefaux

Trinity College Dublin

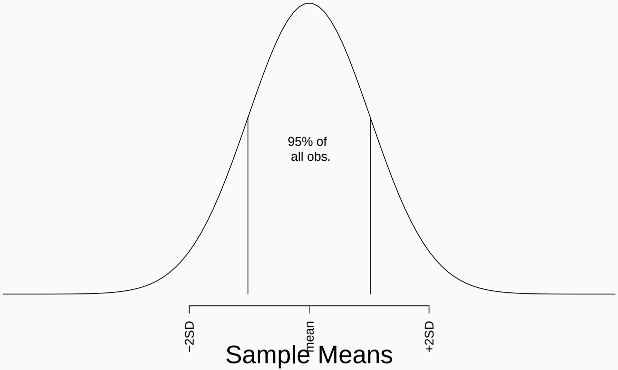
Motivation

Reminder: The sampling distribution

The sampling distribution (see last lecture) can be understand as all the likely outcomes of a sample. You could have large sample means, small ones, etc.

I.e., if you were to take lots of samples from the population, BY CHANCE, you may observe small or large sample means.

Reminder: The sampling distribution

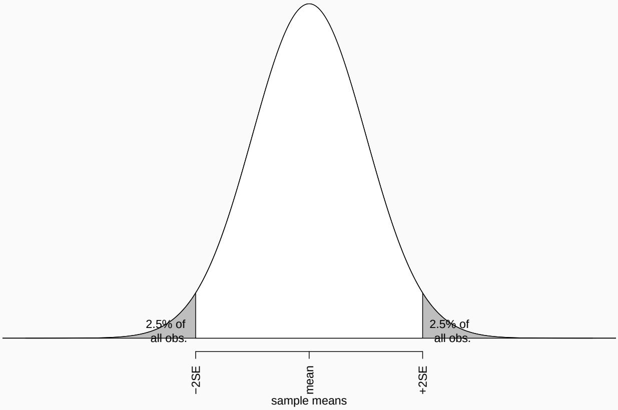


Hypothesis testing

Using the sampling distribution, we'll say that a particular sample mean is likely to have been drawn from a particular population, or unlikely (and hence that they are different).

As a heads up, we'll conclude that a particular sample is unlikely to have come from a particular population if it's mean is particularly extreme, i.e., if it falls in the extremes of the sampling distribution.

Hypothesis testing



Hypothesis tests: an example

An example

We want to test the hypothesis that university students do not have a better or worse IQ than the national average of 100.

We take a sample of 100 students' IQ, and find that it has mean 101.3, and standard deviation 15.

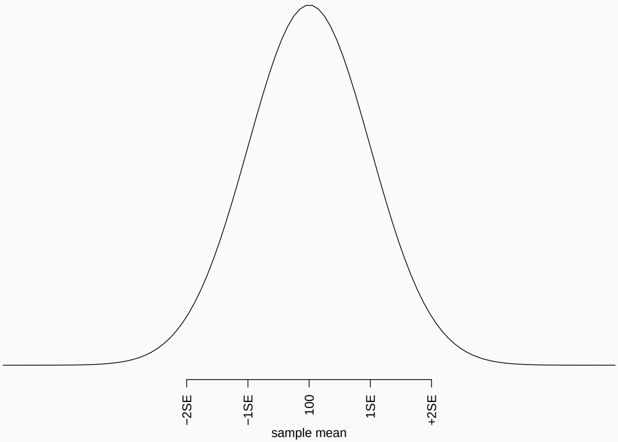
Can we conclude that university students are smarter than the average?

An example

To answer the question, start with a thought experiment. Suppose we took a random sample of size 100, but we draw that sample from the general public (i.e., not just students).

What would we expect the mean of that sample to look like?

An example



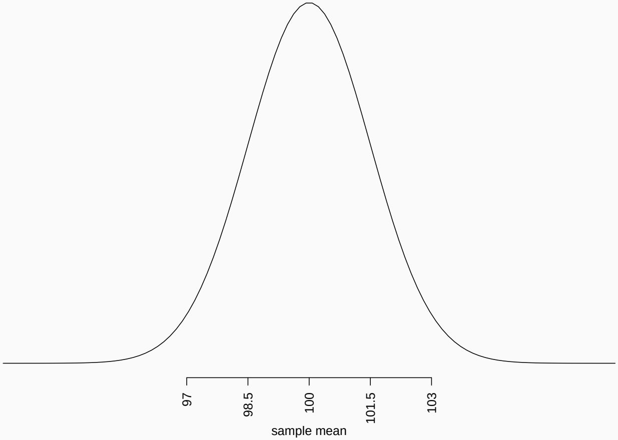
An example

On average, the mean of the sample would be 100, and the standard deviation of these sample means would be given by the standard error: We don't know that standard error, but we can estimate it on the basis of our sample:

$$SE = \frac{s}{\sqrt{n}} = 15/10 = 1.5$$

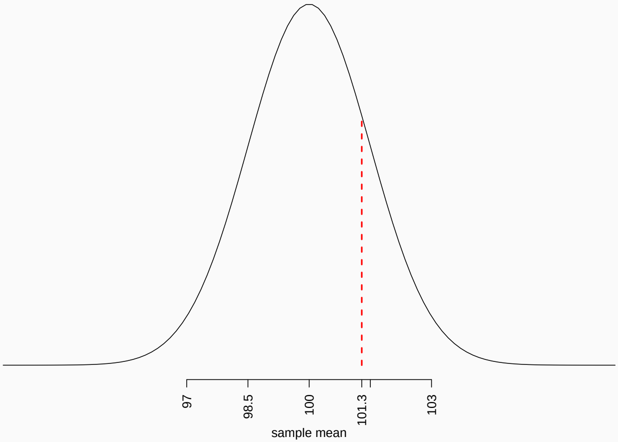
An example

So our sampling distribution would actually look like this:



An example

and the sample mean we actually observed is 101.3, so is here:



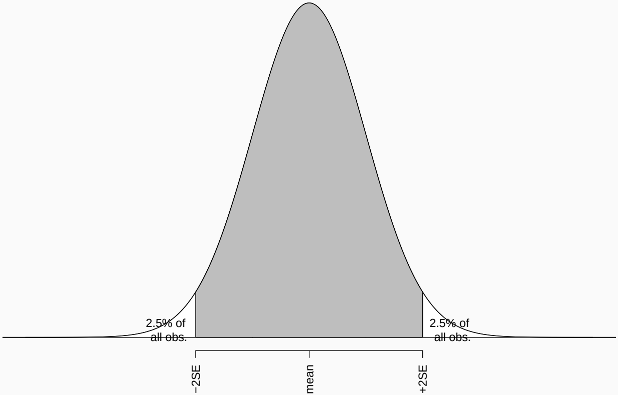
An example

So what would be the probability to observe a sample mean as large as 101.3 IF the sample was draw from the general public?

It would actually be fairly high, or at least not unlikely. It would be a fairly common outcome.

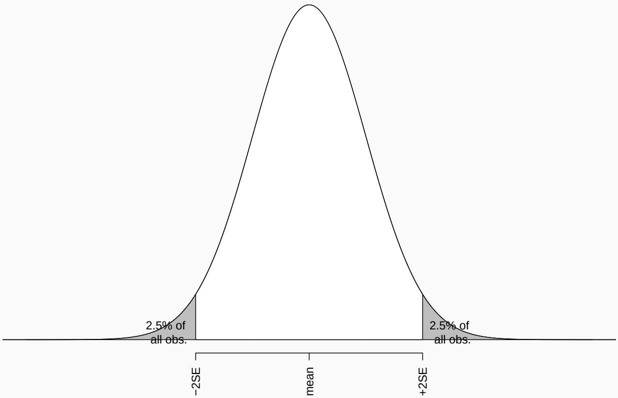
An example

A common outcome is one that would fall within these bounds:



An example

A rare outcome is one that would fall outside of these bounds (either a particular large sample mean, or a particularly low sample mean)



An example

In our example, we find that 101.3 is a common outcome for a sample mean drawn from the general public.

Therefore, we cannot conclude that students have a different IQ from the general public.

A slightly neater way

We saw that 101.3 is not a particularly large IQ mean. Another way to say the same thing is to say that 101.3 is only 0.85 standard errors away from the hypothetical mean of 100.

You'll remember that this is what we talked about when we discussed z-scores. z-scores are a measure of the number of standard deviations away a particular value is from the mean.

A slightly neater way

A neater way would therefore be to calculate the z-score first:

$$z = \frac{\text{raw score - mean}}{\text{standard deviation}}$$

Since we're dealing with sample means here, the standard deviation is called the standard error. So

$$z = \frac{\overline{X} - \mu_{\text{hyp}}}{SE}$$

In our example:

$$z = \frac{101.3 - 100}{1.5} = 0.866$$

An example

You'll also remember that observations that are within $\pm 2SD$ of the mean (2 standard errors here) are not rare, since they happen to take on such values 95% of the time.

Observations that are outside of $\mu_{\text{hyp}} \pm 2SD$ are rare.

Here we have $z = 0.866$, so our sample mean is less than 1 standard deviation away from the mean, and hence we cannot conclude that it is "rare".

An example

If the observation (the sample mean) is not rare, then we would be likely to observe it if the null hypothesis were true.

Therefore we do not reject the null hypothesis.

Step by step procedure

Stating the null hypothesis

Written H_0 (‘h-not’)

Above, our hypothesis was that there is nothing “special” about students’ IQ, i.e.:

$$H_0 : \mu = 100,$$

where μ is the population mean for students.

Stating the alternative hypothesis

Written H_1.

The alternative hypothesis (H1) asserts the opposite of the null hypothesis:

$$H_1 : \mu \neq 100,$$

Decision rule

Specifies precisely *when* H_0 should be rejected. I.e., how rare a sample mean do we want to observe for us to conclude that there is something special going on?

level of significance: To do this, you need to choose a *level of significance* (α). The level of significance specifies a degree of "rarity" that you'll require. E.g., a 5% significance level means that you'll reject the Null hypothesis if your z score would occur by chance in only 5% of cases.

Critical z-score. To this level of significance will correspond a "critical value". For example, a z-score higher than 2 (technically 1.96) would occur in only 5% of cases, so a level of significance of 5% will *impose* that you have a critical z-score of 2.

Calculation

Calculate your z-score (we did this above, so just to repeat).

$$z = (101.3 - 100)/1.5 = 0.866$$

Make a decision

Here z is smaller than your critical value of 2.

Practically, this means that this outcome is common, so you cannot reject the null hypothesis. We say that you "fail" to reject the null hypothesis.

Another example, all together

Another example

Hypothesis: TCD students are taller than the general public's average height of 170cm

Relevant information:

- Sample size: 900
- Sample mean: 175
- Sample standard deviation: 10

1. state your hypotheses

$$H_0 : \mu_{TCD} = 170$$

$$H_1 : \mu_{TCD} \neq 170$$

2. State your decision rule

- Level of significance 1%
- corresponding critical z-value: 2.32 (more on how to find those later)

2. Calculate standard error

$$SE = \frac{s}{\sqrt{n}} = \frac{10}{10} = 1$$

3. Calculate z score

$$z = \frac{\overline{X} - \mu_{\text{hyp}}}{SE} \tag{1}$$

$$= \frac{175 - 170}{1} \tag{2}$$

$$= 5 \tag{3}$$

4. Make a decision

Since $z = 5$ is larger than the critical z-value of 2.32, we can reject the null hypothesis at the 1% level of significance.