One sample z-test

Gaston Sanchez

Tests of Significance

(FPP chapter 26)

Can the result be explained by chance or is another explanation necessary?

2 competing ideas

2 competing ideas (hypotheses)

Just chance

Null hypothesis

Other explanation

Alternative hypothesis

Main ingredients for a hypothesis test

1. State Null and Alternative hypotheses
2. Determine and calculate Test Statistic
3. Compute P-value
4. Make a Conclusion

Null Hypothesis

Results obtained due to random variation

Results can be explained by chance

Chance alone is a reasonable explanation

Alternative Hypothesis

Alternative to the null

Another explanation is necessary

Results can’t be explained by chance

Chance alone is not a reasonable explanation

Test Statistic

Measures the difference between the data and what is expected on the null hypothesis

Quantifies the differences between observed data end expected results (under null)

P-value

P-value is the chance of getting data like we got, or more extreme, given the null hypothesis is true.

The smaller the p-value, the stronger the evidence against the null hypothesis.

Conclusion

Either:

Fail to reject null hypothesis

OR

Reject null hypothesis

One sample z-test

One sample z-test

The z-test can be used for:

- Sums of draws from a box
- Proportion of 1's in draws from a box
- Average of draws

Use it for more than 25 draws, and when the SD of box is **known**

One sample z-test

Test statistic:

$$Z = \frac{\text{Observed} - \text{EV}}{\text{SE}}$$

How many SEs away an observed value is from its expected value (*computed under the null hypothesis!*)

Example (hypothetical)

Manufacturer: net weight 275 grams

EV = 275 g

SD = 20 g

Simple Random Sample of 25 bags of chips

266	254	248	249	297
261	293	261	266	279
222	212	282	281	265
240	284	253	274	243
272	279	261	273	295

Avg = 264.4 g **vs** EV = 275 g

Sample statistic *population parameter*

Can the difference be explained by chance?

Simple Random Sample of 25 bags of chips

266	254	248	249	297
261	293	261	266	279
222	212	282	281	265
240	284	253	274	243
272	279	261	273	295

draws = 25

Avg = 264.4 g *vs* EV = 275 g

Sample statistic *population parameter*

Is the observed difference due to chance?
Or is the avg weight much lower than the claimed 275 g?

Two Hypotheses

Null hypothesis:

- Results can be explained by chance
- Avg weight is 275 g

Alternative hypothesis:

- Not just chance
- Avg weight is less than 275 g

A test of significance gets at the question of whether an observed difference is real, or just chance variation.

z-statistic

$$z = \frac{\text{Observed - Expected}}{\text{SE}}$$

Quantifies the difference between observed data end expected results (under null)

z-statistic

$$z = \frac{264.4 - 275}{SE}$$

$$SD = 20$$

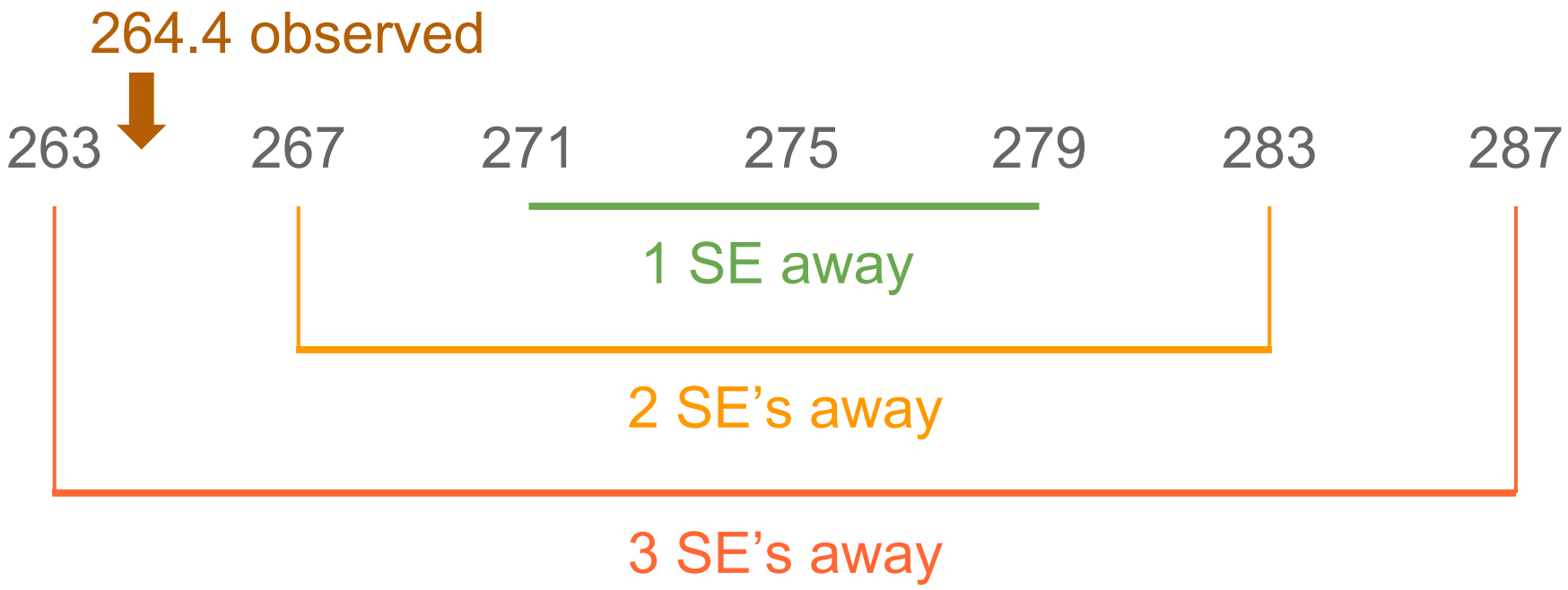
$$SE = 20 / \sqrt{25} = 4$$

z-statistic

$$z = \frac{264.4 - 275}{4} = -2.65$$

Assuming that the null hypothesis is true, the observed value of the sample average (264.4 g) is 2.65 SEs below its EV

How far from Expected Value?



P-value

$$z = \frac{264.4 - 275}{4} = -2.65$$

The chance of getting a sample average 2.65 SEs or more below its expected value is about **0.004** *(from the normal table)*

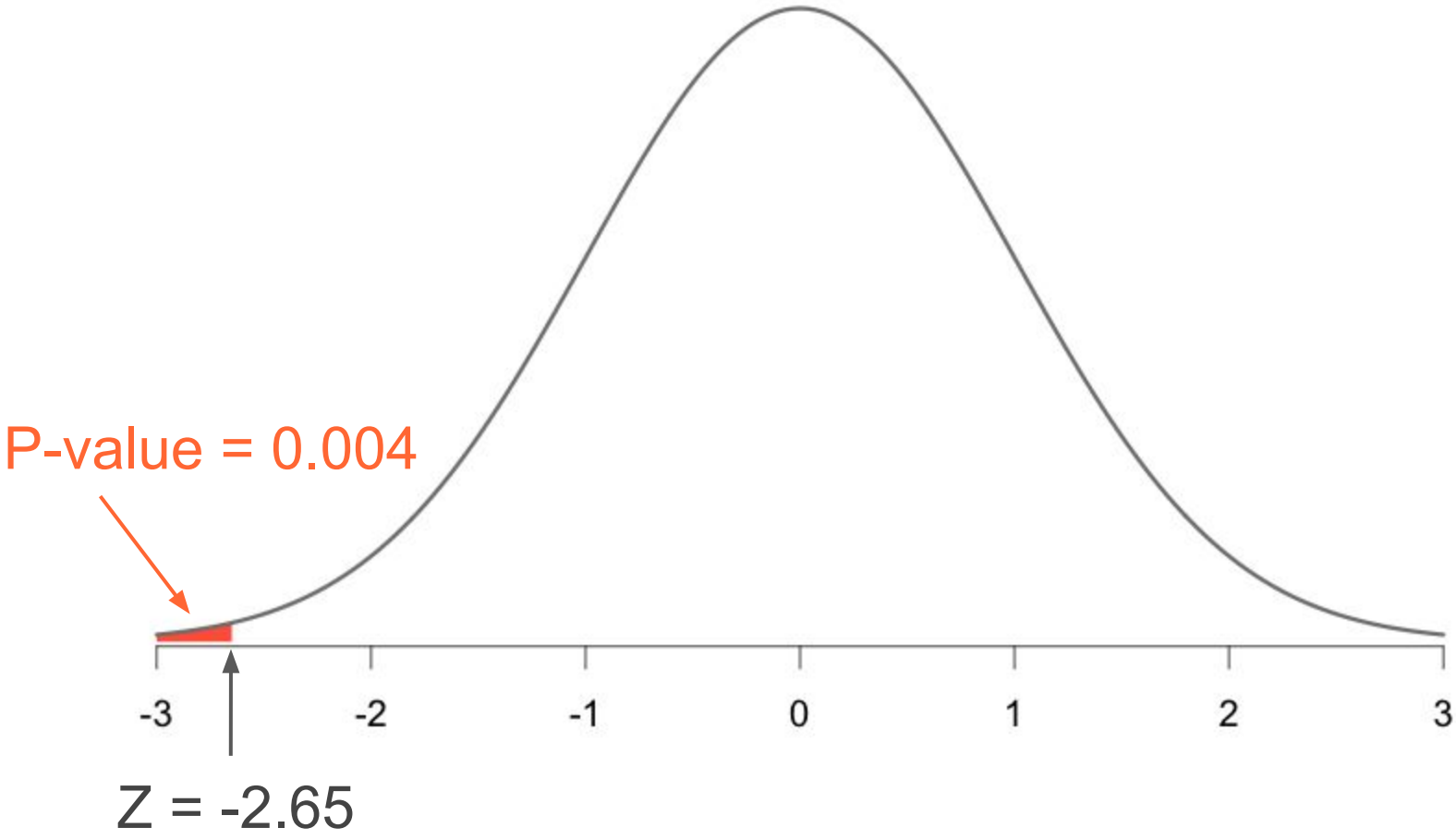
P-value

P-value

P-value is the chance of getting a big test statistic, assuming the null hypothesis to be right.

P-value is NOT the chance of the null hypothesis being right

How far from Expected Value?



Conclusion

- We reject the null hypothesis
- Not just chance
- Something else is going on (e.g. failure in manufacturing process, broken machines, scales not calibrated)

We are not finding what is the cause of the difference. We are just simply saying that there is a difference not reasonably explained by chance alone.

Making a test of significance

To make a test of significance you have to:

- Set up the null hypothesis (in terms of a box model for the data)
- Set up the alternative hypothesis
- Pick a test statistic, to measure the difference between the data and what is expected on the null hypothesis
- Compute the observed significance level P
- Make a conclusion

To make a test of significance you have to:

How small the observed significance level has to be before rejecting the null hypothesis?

If *P* is less than 5%, the result is called statistically significant

If *P* is less than 1%, the result is called highly significant

Another Example

From past year the national values:

Avg of MSAT = 519, SD of MSAT = 110

In this year, a SRS of 100 students who took MSAT in CA:

Sample Avg = 504, SD = 100

Did all CA students have a lower avg MSAT score or can results be explained by chance?

MSAT scores example

Null:

a) Difference can be exaplined by chance
b) Avg of all CA students = 519

Alternative:

a) Not just chance
b) Avg of all CA students is lower than 519

MSAT scores example

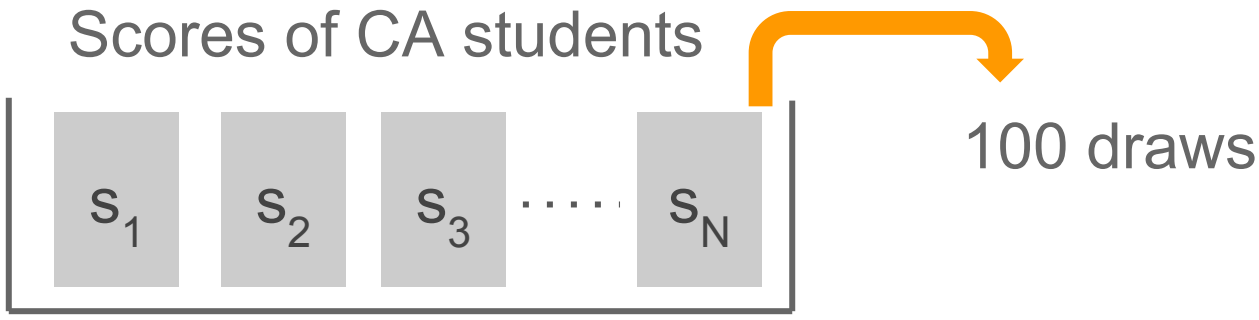
$$z = \frac{504 - 519}{SE}$$

SE avg = (SE sum) / # draws

SE sum = $\sqrt{100}$ (SD box)

SD box = ?

What is the box?



SD of box is unknown

But we can use SD of sample (Bootstrap method)

What SD should you use?

If you know SD of box or it is implied by the null hypothesis, then use it.

If not, use sample SD to estimate SD of box (bootstrap method)

If you are confused, think about what the box is

MSAT scores example

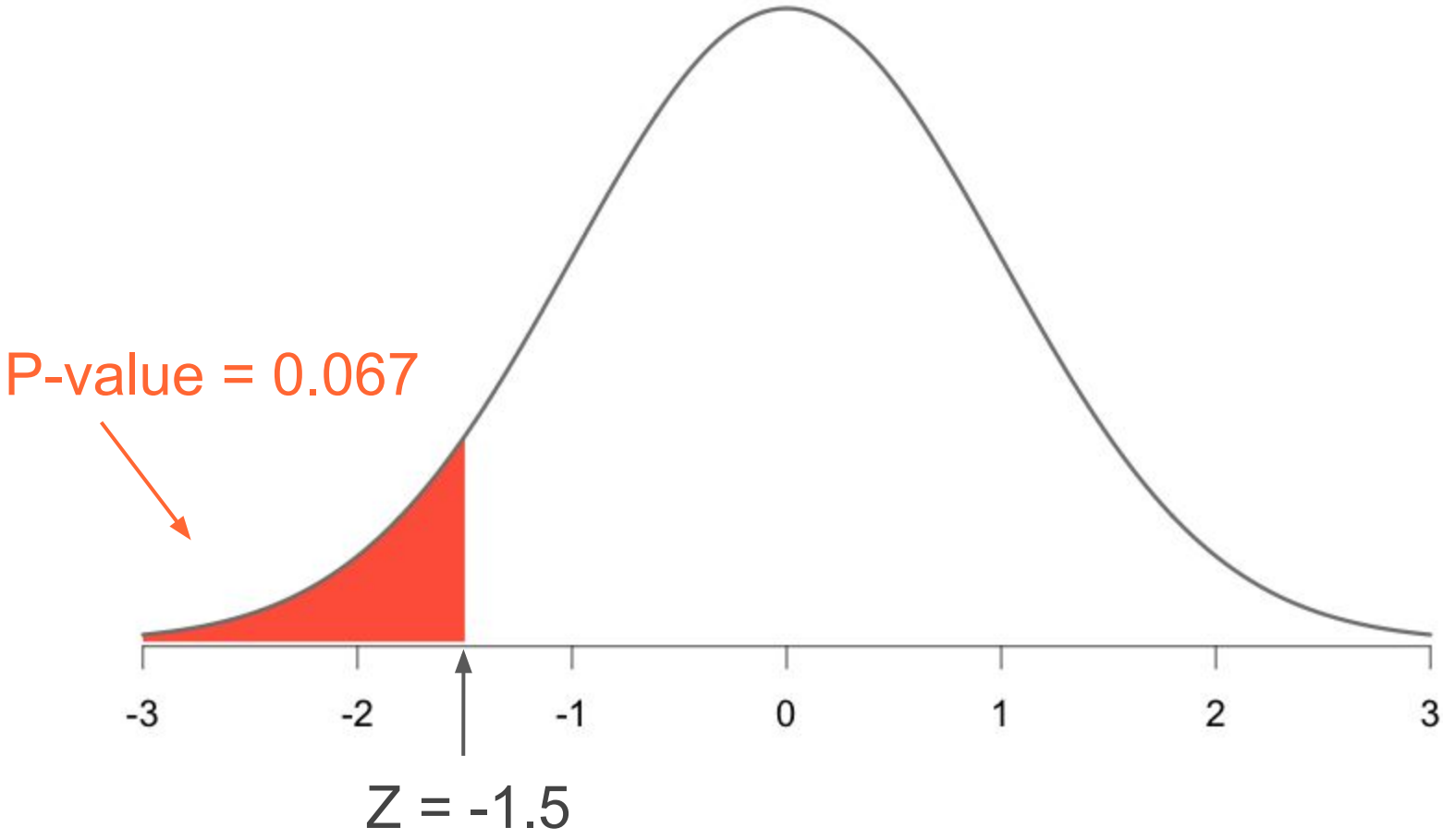
SD box = 100 (SD of sample)

SE sum = $\sqrt{100}$ (100) = 1000

SE avg = 1000 / 100 = 10

$$z = \frac{504 - 519}{10} = -1.5$$

MSAT scores example



MSAT scores example

Conclusion:

- The null hypothesis can't be rejected.
- The difference (504 -vs- 519) could be due to chance
- CA students have same avg MSAT as nation