Hypothesis tests and confidence intervals

Johan A. Elkink
School of Politics & International Relations
University College Dublin

23 October 2017

Outline

Margin of error

From the previous two lectures:

$$z_i = \frac{x_i - \bar{x}}{\sigma_x}$$

The z-distribution is a normal distribution with mean 0 and variance 1.

$$S.E. = \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

The **margin of error** m is:

$$m = z^c \cdot \sigma_{\bar{x}} = z^c \cdot \frac{\sigma_x}{\sqrt{n}}$$

z^c is a relevant critical value of z, for example, to get a 95% confidence interval, $z^c = 1.96$.

Confidence interval

$$CI = \bar{x} \pm m = \bar{x} \pm z^c \frac{\sigma_x}{\sqrt{n}}$$

Note that the confidence interval only takes into account the **sampling error**.

All other errors are ignored, such as **measurement error** or **non-response bias**.

Example

A sample of 100 respondents shows an average sympathy score for politician A of 35 (on a 0-100 scale), with a variance of 10. What is the 95% confidence interval?

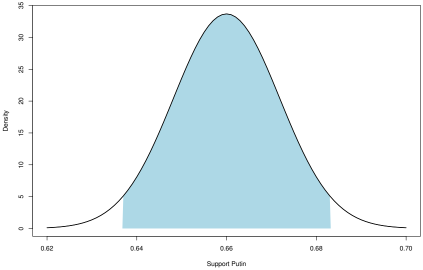
$$CI = \bar{x} \pm z^c \frac{\sigma_x}{\sqrt{n}} = 35 \pm 1.96 \cdot \frac{\sqrt{10}}{\sqrt{100}} = 35 \pm 0.62$$

Thus the confidence interval is $[34.38; 35.62]$.

CI of a proportion: example

Say, we want to look at the voter support for Vladimir Putin. 66% of 1600 respondents in a survey say they will vote for Putin.

$$SE(\hat{p}) = \sigma_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/(n-1)} = \sqrt{(.66 \cdot .34)/1599} = .012$$
$$CI_{95\%} = [\hat{p} - 1.96\sigma_{\hat{p}}; \hat{p} + 1.96\sigma_{\hat{p}}] = [.637; .683]$$



Exercise

(Verzani 7.6) A student wishes to find the proportion of left-handed people. She surveys 100 fellow students and finds that only 5 are left-handed. Does a 95% confidence interval contain the value of $p = \frac{1}{10}$?

Outline

Hypotheses

The research hypothesis, or alternative hypothesis, is always contrasted to the **null hypothesis**.

Null hypothesis (H_0): states the assumption that there is no effect or difference (depending on the research question).

Alternative hypothesis (H_1 or H_a): states the research hypothesis which we will test *assuming the null hypothesis*.

Because we need to have some estimate of the **sampling distribution** to determine how likely our outcome is. For this estimate, we use the null hypothesis.

Null hypothesis: example

Null hypothesis: men have the same thermometer score for Bertie Ahern as do women.

Alternative hypothesis: women have a higher thermometer score for Bertie Ahern than do men.

$$H_0 : \mu_{women} = \mu_{men}$$

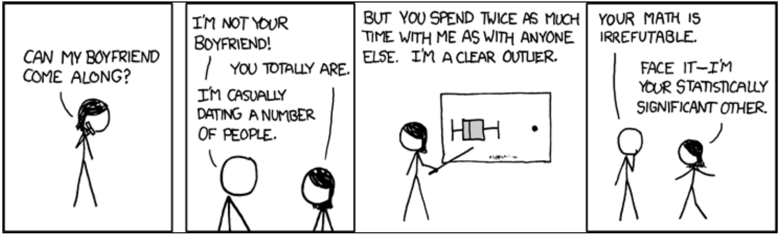
$$H_1 : \mu_{women} > \mu_{men}$$

Hypothesis testing

Given the sampling distribution, depending on the distribution and test, we can calculate a **test statistic**.

We can then calculate the probability of observing the observed value or more extreme, in the direction of the alternative hypothesis, given the assumptions of the null hypothesis.

This is the ***p*-value**.



Hypothesis testing: p-values

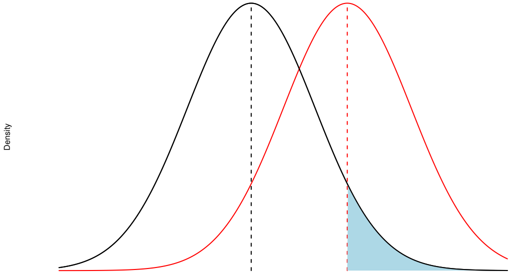


Figure: Calculating p-values

Hypothesis testing: old-fashioned

We used to calculate the **critical value** given the test statistic and the sampling distribution where we would **reject the null hypothesis**.

This critical value you can find in tables for the specific probability distribution.

The **α-level** refers to the maximum acceptable probability that you reject a null hypothesis that is correct.

Since Fisher (1923), typical alpha values:

$\alpha = .05$: most common in the social sciences

$\alpha = .01$: more common in pharmaceutical research

$\alpha = .10$: less picky social scientists

Note that you can reject a null hypothesis, but you can never “accept” a null hypothesis.

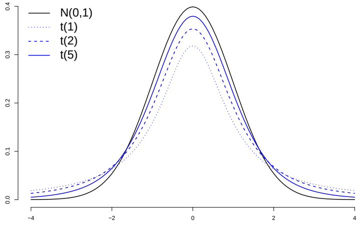
You either reject or you do not.

t-distribution

Note that the z-test assumes that σ_x^2 is known.

The t-distribution is used instead of the z-distribution when the variance σ_x^2 is estimated.

This distribution converges to the z-distribution as n gets larger.



The **Student t-distribution** is invented in 1908 by William S. Gosset, a Guinness employee, who published under pseudonym Student.

Significance test for a mean

Is the mean statistically significantly different from zero?

$$H_0 : \mu = 0$$
$$H_1 : \mu \neq 0$$

$$t = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{x}}} = \frac{\bar{x} - 0}{\hat{\sigma}_x / \sqrt{n}}$$

$$t^c_{(\alpha=.05)} = \pm 1.96 \quad \text{(for large } n\text{)}$$

Outline

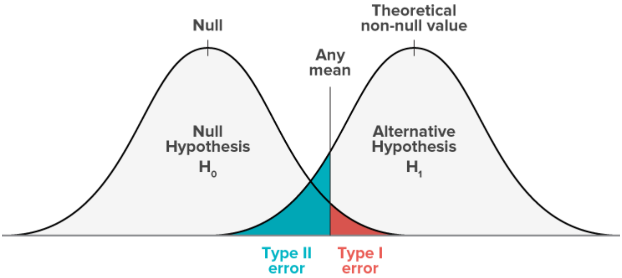
Type-I and Type-II errors

	H_0 is true	H_0 is not true
Do not reject H_0	**Correct decision** probability $1 - \alpha$	**Type II error** probability β
Reject H_0	**Type I error** probability α *significance level*	**Correct decision** probability $1 - \beta$ *power*

Type-I and Type-II errors

Type-I error: null hypothesis is falsely rejected.

Type-II error: null hypothesis should have been rejected, but was not.



We usually try to be **conservative** and prefer the risk of Type-II errors over that of Type-I errors.

http://grasshopper.com/blog/the-errors-of-ab-testing-your-conclusions-can-make-things-worse/

Power

Power: probability of rejecting a hypothesis that is false

$$1 - P(\text{Type II error})$$

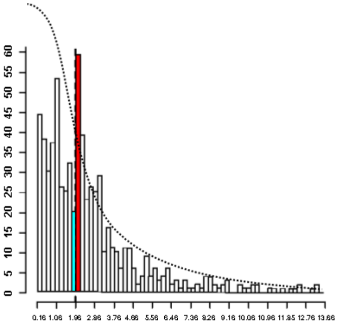
The power of a test increases when:

- the true value is further from the null hypothesis value;
- the variance is lower;
- the sample size is larger.

(Davidson and MacKinnon, 1999, 126)

Publication bias

"Another interesting example (...) is the propensity for published studies to contain a disproportionally large number of Type I errors; studies with statistically significant results tend to get published, whereas those with insignificant results do not." (Kennedy, 2008, 61)



(Gerber and Malhotra, 2008)

Optimal sample size

Recall, the margin of error m is:

$$m = z^c \cdot \sigma_{\bar{x}} = z^c \cdot \frac{\sigma_x}{\sqrt{n}}$$

Given a desired margin of error m, we can calculate the minimum required sample size as:

$$n^* = \left(\frac{z^c \sigma_x}{m} \right)^2$$

Exercise

Suppose a test evaluating students' motivation, attitude towards study, and study habits, produces a score on a 0-200 scale. The mean score for Irish students is 115.

A teacher expects older students to score higher and takes a sample of 25 student who are 30 or older. He finds a mean score $\bar{x} = 127.8$ and a variance of 30. State the hypotheses and calculate the p-value for a one-sided t-test.

(Moore, McCabe and Craig, 2012, 376)

Exercise

The owner of a car which has a board computer that calculates the efficiency of the engine in km/l performs, in addition, manual calculation by dividing the distance driven by the liters used each time she fills her tank. After 20 times, she has the following differences between her estimates and those of her car:

5.0	6.5	-0.6	1.7	3.7	4.5	8.0	2.2	4.9	3.0
4.4	0.1	3.0	1.1	1.1	5.0	2.1	3.7	-0.6	-4.2

State the hypotheses and perform a t-test whether her estimates differ significantly from those of her car.

(Moore, McCabe and Craig, 2012, 377)

Davidson, Russell and James G. MacKinnon. 1999. *Econometric theory and methods*. Oxford: Oxford University Press.

Fisher, Ronald A. 1923. "Statistical tests of agreement between observation and hypothesis." *Economica* (8):139–147.

Gerber, Alan and Neil Malhotra. 2008. "Do statistical reporting standards affect what is published? Publication bias in two leading political science journals." *Quarterly Journal of Political Science* 3(3):313–326.

Kennedy, Peter. 2008. *A guide to econometrics*. 6th ed. Malden, MA: Blackwell.

Moore, David S., George P. McCabe and Bruce A. Craig. 2012. *Introduction to the practice of statistics*. 7th international edition ed. New York: W.H. Freeman.