Comparing means and proportions

Johan A. Elkink

University College Dublin

29 March 2012

1. Two-sample tests of means
2. Two-sample test of proportions
3. Tests on two-way tables

Outline

Two-sample t-test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

with degrees of freedom $min(n_1 - 1, n_2 - 1)$.

Two-sample t-test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

with degrees of freedom $min(n_1 - 1, n_2 - 1)$.

A somewhat more precise estimate of the degrees of freedom can be calculated and is used by SPSS.

Two-sample t-test: example

Imagine an experiment of extra teaching for half a class, with x representing the score on a test at the end.

Group	n	$\bar{x}$	s_x
Treatment	21	51.48	11.01
Control	23	41.52	17.15

Perform a t-test that the extra teaching significantly helped.

Two-sample t-test: example

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{51.48 - 41.52}{\sqrt{\frac{11.01^2}{21} + \frac{17.15^2}{23}}} = 2.31$$

The degrees of freedom are the lowest of $n_1 - 1$ and $n_2 - 1$, i.e. 20. We can look this up in a table to find $0.01 < p < 0.02$.

Confidence interval of the difference

The confidence interval follows then as:

$$CI = (\bar{x}_1 - \bar{x}_2) \pm t^c \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Exercise

Does increased calcium reduce blood pressure?

Treatment	n	$\bar{x}$	s
Calcium	10	5.000	8.743
Placebo	11	-0.273	5.901

(Moore, McCabe & Craig 2012: 444)

Exercise

Does increased calcium reduce blood pressure?

Treatment	n	$\bar{x}$	s
Calcium	10	5.000	8.743
Placebo	11	-0.273	5.901

Also calculate a 95% confidence interval.

(Moore, McCabe & Craig 2012: 444)

Exercise

A movie is rated by male and female visitors to a website.

Rating:	1	2	3	4	5
Female	5	44	48	183	188
Male	5	29	30	91	66

- Calculate $\bar{x}$ and s_x for each group.
- Perform a t-test whether the means differ.

(Moore, McCabe & Craig 2012: 450)

Exercise

Open the `bes_class_data.sav` file.

Is trust in politicians in general (**trustpol**) significantly lower than trust in parliament (**trustprl**)?

(Note that this is a paired sample - the same respondents on each variable.)

Exercise

Are voters who turned out to vote (**turnout**) more likely to trust politicians? (**trustpol**)?

(Note that this is not a paired sample.)

Outline

1. Two-sample tests of means
2. Two-sample test of proportions
3. Tests on two-way tables

Comparing proportions

Recall, for a dichotomous variable,

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = p(1-p).$$

Thus for two proportions we get:

$$t = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Exercise

Here is data from a study on binge drinking among college students:

	n_{sample}	$n_{drinking}$	$\hat{p} = \frac{n_{drinking}}{n_{sample}}$
Men	5,348	1,392	0.260
Women	8,471	1,748	0.206

Do men binge drink more than women in this sample?

Outline

Chi-squared test

To check for dependence between two categorical variables, we can make use of the chi-squared test (χ^2-test). Based on the margins of the table, we can determine the expected values of the cells under independence and then calculate whether the cells differ significantly:

$$\chi^2 = \sum \frac{(n_{observed} - n_{expected})^2}{n_{expected}}$$

Chi-squared test: *df*

The degrees of freedom is calculated as $df = (r-1)(c-1)$, with r the number of rows and c the number of columns of the two-way table.

Exercise

We will skip the manual calculation of χ^2. Executing the test in SPSS delivers a *p*-value just like with the *t*-test.

Open `bes_class_data.sav`. Perform a test for the relation between:

- **euvote** and **labour**.
- **turnout** and **labour**.