Quantitative Analysis and Empirical Methods

Hypothesis testing

Jan Rovny

Sciences Po, Paris, CEE / LIEPP

Introduction

- Hypotheses
- Procedure of hypothesis testing
- Two-tailed and one-tailed tests
- Statistical tests with categorical variables

A hypothesis

- A testable statement about relationships between characteristics
- Since Karl Popper, scientific inquiry is not expected to *prove* facts, but rather to *falsify* or confirm theoretical postulates.
- The logic we take when testing hypotheses in statistical methods is thus a ‘negative’ logic:
- Each hypothesis has a logical opposite which we call the *null* hypothesis and denote it H_0.
- In statistics we often set up a null hypothesis which we seek to reject. If we reject the null, then the hypothesis of interest is supported by our analysis.

Example from last lecture

- $X \sim N(5, 16)$
- Is 7.5 significantly different from the mean of X?

Example from last lecture

- $X \sim N(5, 16)$
- Is 7.5 significantly different from the mean of X?
  - $H_1 : 7.5 \neq \bar{X}$
  - $H_0 : 7.5 = \bar{X}$

Example from last lecture

- $X \sim N(5, 16)$
- Is 7.5 significantly different from the mean of X?
  - $H_1 : 7.5 \neq \bar{X}$
  - $H_0 : 7.5 = \bar{X}$
- 7.5 in terms of Z-scores: $Z = \frac{7.5-5}{4} = .625$
- .625 is clearly within the $[-1.96, +1.96]$ interval, thus 7.5 is too close to $\bar{X}$. We *fail to reject* the null hypothesis.
- That means that H_0 stands and H_1 is not supported. 7.5 is not significantly different from $\bar{X}$.

Hypothesis Testing Procedure

1. State a null and alternative hypothesis: $H_0 : \mu = \mu_0$, $H_a : \mu \neq \mu_0$
2. Select a level of significance of interest: $\alpha = .05$ (we want to be 95% sure.)
3. Determine the sampling distribution of the test statistic. (If we are dealing with a means test and we know σ, we use the standard normal distribution and its Z statistic, if we are dealing with a means test and we don't know σ we use Student's t distribution and the T statistic.)
4. Calculate the test statistic (for z: $z = \frac{X - \mu}{\sigma}$)
5. Find the critical value in the appropriate statistical table
6. Make a conclusion about the null hypothesis (reject or fail to reject)

Test of Statistical Significance

Do men and women view gay marriage differently?

- A feeling thermometer on gay marriage 0=fully oppose; 100=fully support
- Poll: Women $\bar{X} = 51, s = 4$; men $\bar{X} = 46, s = 8$
- Difference: $51 - 46 = 5$;
- N=100 women, 100 men

Does the sample difference reflect the population difference or just sampling error?

1. Stating the hypotheses

- H_a: There is a difference in women's and men's feelings toward gay marriage in the population
- H_0: There is *NO* difference in women's and men's feelings toward gay marriage in the population.

2. Deciding the significance level

- Two possible errors we can commit in statistics:
  - Type I error: finding a relationship where there is none (false positive)
  - Type II error: finding no relationship where there is one (false negative)
- Usually select significance level $\alpha = 0.05$ (or 5%)
  - Rejecting H_0 will commit Type I error (false positive) no more than 5 times in 100
  - Rejecting H_0 only if the observation (the difference of 5 between women and men) could have occurred by chance fewer than 5 times out of 100.

3. The sampling distribution

- Comparing two means – CLT – normal distribution
- T or Z? $N < 1000$, so prefer T

4. The test statistic

- As before we take the observed or expected value and subtract our null from it:
    - $T = \frac{H_a - H_0}{se_{diff}}$
- But need to calculate the s.e. of the difference
    - $se_{diff} = \sqrt{se_1^2 + se_2^2} = \sqrt{se_{women}^2 + se_{men}^2}$
    - $se_w = \frac{s}{\sqrt{N}} = \frac{4}{\sqrt{100}} = 0.4$; $se_m = \frac{s}{\sqrt{N}} = \frac{8}{\sqrt{100}} = 0.8$
    - $se_{diff} = \sqrt{se_1^2 + se_2^2} = \sqrt{0.4^2 + 0.8^2} = 0.894$
- Back to T:
    - $T = \frac{H_a - H_0}{se_{diff}} = \frac{diff - 0}{se_{diff}} = \frac{5 - 0}{0.894} = 5.593$

5. and 6. Critical value and Conclusion

How likely are we to get a T value of 5.593 if H_0 were true?

- Same as asking: What is the probability of scoring 5.593 on the T-distribution? (df=n1+n2-2)
- T-table
- The cutoff at the 0.05 significance level is about 1.984, so it is extremely unlikely to get 5.593 by chance.
- Conclusion:
  - Reject H_0.
  - The difference of 5 is statistically significant. There is a significant difference between women's and men's feelings towards gay marriage. Women are significantly more in support.

5. and 6. Critical value and Conclusion

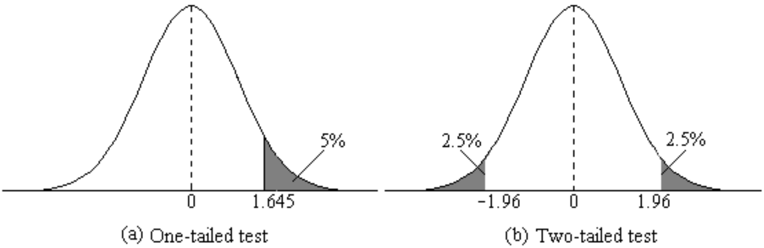
Alternatively, using confidence intervals:

- A 95% confidence interval around the difference (5) would be
  - $X \pm t * se = 5 \pm 1.984 * 0.894 = 5 \pm 1.774$
  - The 95% confidence interval is [3.226; 6.774]
- Conclusion:
  - 95 times out of a 100, the sample difference in women's and men's feelings on gay marriage will lie between 3.226 and 6.774.
  - We are thus confident (at the 0.05 level) that there is a true difference between their opinion in the population.

Two-Tailed v. One-Tailed Tests

- Until now, we have been doing our tests as if we had no expectation about the direction in which we expect 0 to lie.
- As a result, when we were testing whether our observed value is significantly different from, say, 0, we looked at both ends (or tails) of the distribution of our statistic of interest. This was a **two-tailed test**.
- In reality, we often have theoretical expectations about the direction where 0 lies.
- If we find a value of, say, 5 (such as in our example), and question whether it is significantly different from 0, why should we look for 0 on the right tail? It will not be there.
- Consequently, when testing whether a statistic is significantly different from 0, we would expect 0 to be on one particular side of the distribution. Here we can do a **one-tailed test**.

Two-Tailed v. One-Tailed Tests



Testing relationships between categorical variables

- We want to test how cases are dispersed across the dependent variable
- $H_0 =$ every category of the IV should have the same distribution as the total, i.e. the IV does not matter.

Party ID and career crosstabulation

		Law	Politics	Business	Education	Total
Republican	N	6	2	5	1	14
	%	42.9	14.3	35.7	7.1	100
Democrat	N	10	10	2	2	24
	%	41.7	41.7	8.3	8.3	100
Other	N	6	5	7	3	21
	%	28.6	23.8	33.3	14.3	100
Total	N	22	17	14	6	59
	%	37.3	28.8	23.7	10.2	100

χ^2 Test

- To test H_0, we use the χ^2 (read chi-squared) test
- This test compares each observed frequency (fo) with the expected (total) frequency (fe)
  - E.g. if H_0 is correct, 37.3% of the 14 republicans (=5.22) should want to go to into law; and 28.8% of the 14 Republicans (=4.03) should want to go into politics
  - Test: sum the squared differences, divide by the expected frequency: $\chi^2 = \sum_{i=1}^{N} \frac{(fo_i - fe_i)^2}{fe_i}$; N=number of cells (12)

Party ID and career crosstabulation

		Law	Politics	Business	Education	Total
Republican	N	6	2	5	1	14
	%	42.9	14.3	35.7	7.1	100
Democrat	N	10	10	2	2	24
	%	41.7	41.7	8.3	8.3	100
Other	N	6	5	7	3	21
	%	28.6	23.8	33.3	14.3	100
Total	N	22	17	14	6	59
	%	37.3	28.8	23.7	10.2	100

χ^2 Test

- The χ^2 test: $\chi^2 = \sum_{i=1}^{N} \frac{(fo_i - fe_i)^2}{fe_i} = (6-5.2)^2/5.2 + (2-4.0)^2/4.0 + ... = 7.87$
- Apply this value to the χ^2 distribution with appropriate degrees of freedom
- Df=(number of rows - 1)*(number of columns - 1) = (3-1)*(4-1)=6

Party ID and career crosstabulation

		Law	Politics	Business	Education	Total
Republican	N	6	2	5	1	14
	exp N	5.2	4.0	3.3	1.4	14
	%	42.9	14.3	35.7	7.1	100
Democrat	N	10	10	2	2	24
	exp N	8.9	6.9	5.7	2.4	24
	%	41.7	41.7	8.3	8.3	100
Other	N	6	5	7	3	21
	exp N	7.8	6.1	5.0	2.1	21
	%	28.6	23.8	33.3	14.3	100
Total	N	22	17	14	6	59
	%	37.3	28.8	23.7	10.2	100

χ^2 Test

- Our value of χ^2 is 7.78
- What is the critical value of χ^2 at the 0.05 confidence level with 6 df? Chi2-table
- The answer is 12.592. Our χ^2 is smaller than the critical value, so it is possible that 7.87 could occur more than 5 times out of 100 by random chance.
- We fail to reject H_0; there is no statistically significant difference between party ID and career choice.