ECON20003 Quantitative Methods 2

Tutorial 4 (Week 2 - Thursday)

Zheng Fan

The University of Melbourne

Introduction

Zheng Fan

- Ph.D candidate in Economics at Unimelb
- know more about me: zhengfan.site

If you need help,

- Consultation & Ed discussion board (your first priority)
- Email Dr. Xuan Vu for all subject matters
- Consult Stop 1 for special consideration
- Email: fan.z@unimelb.edu.au (last resort!)

Before posting any questions, make sure you have reviewed the materials on Canvas and questions on Ed discussion board!

Roadmap

Today, we will be using R-studio to perform some hypothesis testing.

- Comparing the Central Locations of Two Populations
- 1. paired-sample design
- 2. independent measures design

Paired-sample design

Assumptions:

1. The data is a random sample of independent pairs of observations
2. The variable of interest is quantitative and continuous.
3. The measurement scale is interval or ratio.
4. σ_D is unknown but the population of the differences is (approximately) normally distributed

Independent Measures Design

The two-independent-sample t test and the corresponding confidence interval estimator for the difference between two population means are based on the following assumptions:

1. The data consists of two independent random samples of independent observations (i.e. both the samples and the observations within each sample are independent).
2. The variable of interest is quantitative and continuous.
3. The measurement scale is interval or ratio.
4. σ_1 and σ_2 areunknown but the sampled populations are normally distributed (at least approximately).

Independent Measures Design: Setup

Consider two independent random samples:

$$X_{11}, \ldots, X_{1n_1} \sim (\mu_1, \sigma_1^2), \qquad X_{21}, \ldots, X_{2n_2} \sim (\mu_2, \sigma_2^2)$$

Parameter of interest:

$$\mu_1 - \mu_2$$

Estimator:

$$\bar{X}_1 - \bar{X}_2$$

Three inferential scenarios arise depending on assumptions about σ_1^2 and σ_2^2.

Scenario 1: Population Variances Known

Assumption:

$$\sigma_1^2,\ \sigma_2^2 \text{ known}$$

Standard error:

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Confidence interval:

$$(\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Test statistic:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

This scenario is mainly theoretical.

Scenario 2: Variances Unknown but Equal

Assumption:

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

Pooled variance estimator:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Standard error:

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Confidence interval:

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2,\ n_1+n_2-2} \cdot s_{\bar{X}_1 - \bar{X}_2}$$

Test statistic:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{s_{\bar{X}_1 - \bar{X}_2}} \sim t_{n_1+n_2-2}$$

Scenario 3: Variances Unknown and Unequal

Assumption:

$$\sigma_1^2 \neq \sigma_2^2$$

Standard error:

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Confidence interval:

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2,\ \nu}\, s_{\bar{X}_1 - \bar{X}_2}$$

Test statistic:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} \sim t_\nu$$

Welch–Satterthwaite degrees of freedom:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

This is the default Welch two sample t test.