ECON20003 Quantitative Methods 2

Tutorial 6 (Week 3 - 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!

Statistical Setup

- k independent groups indexed by $j = 1, \ldots, k$
- Each group has n_j observations x_{ij}
- Total sample size $n = \sum_{j=1}^{k} n_j$
- Hypotheses

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

$$H_A : \text{Not all population means are equal}$$

Illustration Data Table: MST Scores by Tutorial

	Tutorial 1	Tutorial 2	Tutorial 3
1	68	74	81
2	72	70	78
3	65	76	85
4	70	73	80

Key Sample Quantities

- Group mean

$$\bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$$

- Grand mean

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}$$

- ANOVA compares variation around $\bar{x}_j$ and variation around $\bar{x}$

Variance Decomposition Idea

Total variation can be split into two parts

$$\text{Total Variation} = \text{Between Group Variation} + \text{Within Group Variation}$$

- Between group variation captures differences across means
- Within group variation captures noise inside groups

ANOVA Variance Decomposition and Estimation

Sum of Squares: $SS = SST + SSE$

$$SS = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x})^2 \qquad SST = \sum_{j=1}^{k} n_j(\bar{x}_j-\bar{x})^2 \qquad SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2$$

Degrees of Freedom

$$df_{Total} = n-1 \qquad df_{Between} = k-1 \qquad df_{Within} = n-k$$

Mean Squares

$$MST = \frac{SST}{k-1} \qquad MSE = \frac{SSE}{n-k}$$

ANOVA Test Statistic

$$F = \frac{MST}{MSE}$$

- Under H_0, both estimate the same population variance
- Large values of F indicate evidence against H_0
- Sampling distribution

$$F \sim F_{k-1,\, n-k}$$

Exercise 1: Manual ANOVA Task

Task

Four groups are compared in terms of completion time. Each group has $n_j = 30$ observations and there are $k = 4$ groups.

Using R calculate summary statistics, manually compute the ANOVA test.

Useful formulas:

Mean: Group mean: $\bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$. Grand mean: $\bar{x} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}$

Sum of Squares: $SS = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2$, $SST = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2$, $SSE = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 = \sum_{j=1}^{k} (n_j - 1) s_j^2$

Degrees of Freedom: $df_{Total} = n - 1$, $df_{Between} = k - 1$, $df_{Within} = n - k$

Mean Squares: $MST = \frac{SST}{k-1}$ $MSE = \frac{SSE}{n-k}$

Test statistic: $F = \frac{MST}{MSE}$

Assumptions

The one way ANOVA F test is based on the following assumptions

1. The data consist of k independent random samples of independent observations drawn from k populations
2. The variable of interest is quantitative and continuous
3. The measurement scale is interval or ratio
4. Each population is normally distributed
5. All populations have the same variance (We can use test using leveneTest to test. If unequal, use Welch's F-test)

Exercise 2

If the sampled populations are clearly not normally distributed,

- Nonparametric counterpart of these tests is the **Kruskal-Wallis test**, a generalization of the Wilcoxon rank-sum test to two or more (sub-) populations.

With the following assumption

- The data set constitutes k independent random samples of independent observations drawn from k (sub-) populations.
- The variable of interest is quantitative and continuous.
- The measurement scale is at least **ordinal**.
- The sampled populations differ at most with respect to their central locations (i.e. medians)