Statistics 2
Chapter 6
Analysis of Variance

Ankara University, Faculty of Political Science, Department of Economics, Onur Özsoy

4/4/2018

Chapter 6
Analysis of Variance

GOALS
When you have completed this chapter, you will be able to:

ONE
Discuss the general idea of analysis of variance.

TWO
List the characteristics of the *F* distribution.

THREE
Conduct a test of hypothesis to determine if two sample variances came from the same or equal populations.

FOUR
Set up and organize data into a one-way and a two-way ANOVA table.

Chapter 6 *continued*
Analysis of Variance

GOALS

When you have completed this chapter, you will be able to:

FIVE

Define the terms treatment and blocks.

SIX

Conduct a test of hypothesis among three or more treatment means.

SEVEN

Develop confidence intervals for the difference between treatment means.

EIGHT

Conduct a test of hypothesis to determine if there is a difference among block means.

Characteristics of F-Distribution

- There is a "family" of *F* Distributions.

- Each member of the family is determined by two parameters: the numerator degrees of freedom and the denominator degrees of freedom.

- *F* cannot be negative, and it is a continuous distribution.

- The *F* distribution is positively skewed.

- Its values range from 0 *to* ∞ . As $F \rightarrow \infty$ the curve approaches the X-axis.

Test for Equal Variances

- For the two tail test, the test statistic is given by:

$$F = \frac{S_1^{\,2}}{S_2^{\,2}}$$

- $S_1^2 \; and \; S_2^2$ are the sample variances for the two samples. The null hypothesis is rejected if the computed test statistic is greater than the critical (table) value with confidence level $\alpha/2$ and numerator and denominator degrees of freedom.

EXAMPLE 1

- Colin, a stockbroker at Critical Securities, reported that the mean rate of return on a sample of 10 software stocks was 12.6 percent with a standard deviation of 3.9 percent. The mean rate of return on a sample of 8 utility stocks was 10.9 percent with a standard deviation of 3.5 percent. At the .05 significance level, can Colin conclude that there is more variation in the software stocks?

EXAMPLE 1 *continued*

- Step 1: $H_0 : \sigma_s \leq \sigma_u \qquad H_1 : \sigma_s > \sigma_u$
- Step 2: H_0 is rejected if F>3.68, αdf=(9,7), =.05 $F = (3.9)^2 / (3.5)^2 = 1.2416$
- Step 3:
- Step 4: H_0 is not rejected. There is insufficient evidence to claim more variation in the software stock.

Underlying Assumptions for ANOVA

- The *F* distribution is also used for testing the equality of more than two means using a technique called analysis of variance (ANOVA). ANOVA requires the following conditions:

  › The populations being sampled are normally distributed.

  › The populations have equal standard deviations.

  › The samples are randomly selected and are independent.

Analysis of Variance Procedure

- **The Null Hypothesis:** the population means are the same.

- **The Alternative Hypothesis:** at least one of the means is different.

- **The Test Statistic:** F=(between sample variance)/(within sample variance).

- **Decision rule:** For a given significance level α, reject the null hypothesis if *F* (computed) is greater than *F* (table) with numerator and denominator degrees of freedom.

NOTE

- If there are k populations being sampled, then the df (numerator) = k-1

- If there are a total of N sample points, then df (denominator) = N-k

- The test statistic is computed by: $F = [(SST)/$

- $(k-1)]/[(SSE)/(N-k)]$.

- SST represents the treatment sum of squares.

- SSE represents the error sum of squares.

- Let T_c represent the column totals, n_c represent the number of observations in each column, and ΣX represent the sum of all the observations.

Formulae

$$SS(total) = \Sigma(X^2) - \frac{(\Sigma X)^2}{n}$$

$$SST = \Sigma\left(\frac{T_c^2}{n_c}\right) - \frac{(\Sigma X)^2}{n}$$

$$SSE = SS(total) - SST$$

EXAMPLE 2

- Rosenbaum Restaurants specialize in meals for senior citizens and families. Katy Polsby, President, recently developed a new meat loaf dinner. Before making it a part of the regular menu she decides to test it in several of her restaurants. She would like to know if there is a difference in the mean number of dinners sold per day at the Sylvania, Perrysburg, and Point Place restaurants for sample of five days. At the .05 significance level can Katy conclude that there is a difference in the mean

EXAMPLE 2 *continued*

	S y l v a n i a	P e r r y s b u r g	P o i n t P l a c e	
	1 3	1 0	1 8	
	1 2	1 2	1 6	
	1 4	1 3	1 7	
	1 2	1 1	1 7	
			1 7	t o t a l
T $_c$	5 1	4 6	8 5	1 8 2
n $_c$	4	4	5	1 3
	6 5 3	5 3 4	1 4 4 7	2 6 3 4

EXAMPLE 2 *continued*

- From the table Katy determines SST=76.25, SSE=9.75, and the Test statistic: F=[76.25/2]/[9.75/10]=39.1026
- Step 1: $H_0: \mu_1 = \mu_2 = \mu_3$ $H_1:$ Not all the means are the same
- Step 2: H_0 is rejected if F>4.10
- Step 3: F=39.10
- Step 4: H_0 is rejected. There is a difference in the mean number of dinners sold

Inferences About Treatment Means

- When we reject the null hypothesis that the means are equal, we may want to know which treatment means differ.

- One of the simplest procedures is through the use of confidence intervals.

Confidence Interval for the Difference Between Two Means

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm t\sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

- where t is obtained from the t table with degrees of freedom (N-k).
- MSE = [SSE/(N-k)]

EXAMPLE 3

- From EXAMPLE 2 develop a 95% confidence interval for the difference in the mean number of meat loaf dinners sold in Point Place (pop #1) and Sylvania (pop #2). Can Katy conclude that there is a difference between the two restaurants?

$$(17 - 12.75) \pm 2.228 \sqrt{.975 \left(\frac{1}{4} + \frac{1}{5} \right)}$$

$$4.25 \pm 1.48 \Rightarrow (2.77, 5.73)$$

Two-Factor ANOVA

- For the two-factor ANOVA we test whether there is a significant difference between the *treatment effect* and whether there is a difference in the *blocking effect*.

- Let B_r be the block totals (r for rows)

- Let SSB represent the sum of squares for the blocks where:

$$SSB = \Sigma \left[\frac{B_r^2}{k} \right] - \frac{(\Sigma X)^2}{n}$$

EXAMPLE 4

⦿ The Bieber Manufacturing Co. operates 24 hours a day, five days a week. The workers rotate shifts each week. Todd Bieber, the owner, is interested in whether there is a difference in the number of units produced when the employees work on various shifts. A sample of five workers is selected and their output recorded on each shift. At the .05 significance level, can we conclude there is a difference in the mean

EXAMPLE 4 *continued*

Employee	Day Output	Evening Output	Night Output
McCartney	31	25	35
Neary	33	26	33
Schoen	28	24	30
Thompson	30	29	28
Wagner	28	26	27

EXAMPLE 4 *continued*

- TREATMENT EFFECT
  - › Step 1: $H_0 : \mu_1 = \mu_2 = \mu_3$ $H_1 :$ Not all means are equal.
  - › Step 2: H_0 is rejected if F>4.46, df=(2,8).
  - › Compute the various sum of squares: SS(total) = 139.73, SST = 62.53, SSB = 33.73, SSE = 43.47. df(block) = 4, df(treatment) = 2, df(error)=8.
  - › Step 3: F=[62.53/2]/[43.47/8]=5.75

EXAMPLE 4 *continued*

> › Step 4: H_0 is rejected. There is a difference in the average number of units produced for the different time periods.

- Block Effect:

> › Step 1: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 \qquad H_1 :$ Not all means are equal.

> › Step 2: H_0 is rejected if F>3.84, *df=(4,8)*

> › Step 3: F=[33.73/4]/[43.47/8]=1.55

> › Step 4: H_0 is not rejected since there is no significant difference in the average number of units produced for the different employees.