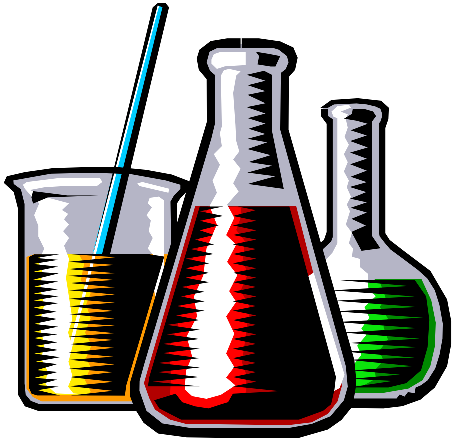
Econ 3790: Business and Economic Statistics

Instructor: Yogesh Uppal
Email: yuppal@ysu.edu

Chapter 13, Part A: Analysis of Variance and Experimental Design

- Introduction to Analysis of Variance
- Analysis of Variance: Testing for the Equality of *k* Population Means



Introduction to Analysis of Variance

- Analysis of Variance (ANOVA) can be used to test for the equality of three or more population means.

- We want to use the sample results to test the following hypotheses:

$$H_0\colon\ \mu_1 = \mu_2 = \mu_3 = \cdot\ \cdot\ \cdot = \mu_k$$

H_a: Not all population means are equal

Introduction to Analysis of Variance

H_0: $\mu_1 = \mu_2 = \mu_3 = \cdot\ \cdot\ \cdot = \mu_k$

H_a: Not all population means are equal

- If H_0 is rejected, we cannot conclude that *all* population means are different.
- Rejecting H_0 means that at least two population means have different values.

Assumptions for Analysis of Variance

- For each population, the response variable is normally distributed.
- The variance of the response variable, denoted σ^2, is the same for all of the populations.
- The observations must be independent.

Test for the Equality of k Population Means

- Hypotheses

H_0: $\mu_1 = \mu_2 = \mu_3 = \cdot\ \cdot\ \cdot = \mu_k$
H_a: Not all population means are equal

- Test Statistic

$$F = \text{MSTR}/\text{MSE}$$

Between-Treatments Estimate of Population Variance

- A between-treatment estimate of σ^2 is called the mean square treatment and is denoted MSTR.

$$MSTR = \frac{SSTR}{k-1}$$

Denominator represents the degrees of freedom

Numerator is the sum of squares due to treatments and is denoted SSTR

Within-Samples Estimate of Population Variance

- The estimate of σ^2 based on the variation of the sample observations within each sample is called the mean square error and is denoted by MSE.

$$MSE = \frac{SSE}{n_T - k}$$

Denominator represents the degrees of freedom associated with SSE

Numerator is the sum of squares due to error and is denoted SSE

Test for the Equality of k Population Means

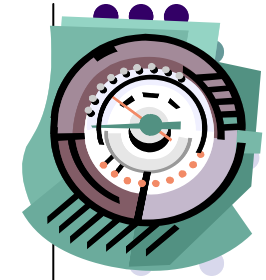
- k: # of subpopulations you are comparing.
- n_T: Total number of observations.

- Rejection Rule

$$\boxed{\text{Reject } H_0 \text{ if } F \geq F_\alpha}$$

where the value of F_α is based on an F distribution with k - 1 numerator d.f. and n_T - k denominator d.f.

Hypothesis Testing About the Variances of Two Populations



Selected Values from the *F* Distribution Table

Denominator Degrees of Freedom	Area in Upper Tail	Numerator Degrees of Freedom				
		7	8	9	10	15
8	.10	2.62	2.59	2.56	2.54	2.46
	.05	3.50	3.44	3.39	3.35	3.22
	.025	4.53	4.43	4.36	4.30	4.10
	.01	6.18	6.03	5.91	5.81	5.52
9	.10	2.51	2.47	2.44	2.42	2.34
	.05	3.29	3.23	3.18	3.14	3.01
	.025	4.20	4.10	4.03	3.96	3.77
	.01	5.61	5.47	5.35	5.26	4.96

Comparing the Variance Estimates: The *F* Test

- If the null hypothesis is true and the ANOVA assumptions are valid, the sampling distribution of MSTR/MSE is an *F* distribution with MSTR d.f. equal to k - 1 and MSE d.f. equal to n_T - k.
- If the means of the k populations are not equal, the value of MSTR/MSE will be inflated because MSTR overestimates σ^2.
- Hence, we will reject H_0 if the resulting value of MSTR/MSE appears to be too large to have been selected at random from the appropriate *F* distribution.

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F
Treatment	SSTR	$k - 1$	MSTR	MSTR/MSE
Error	SSE	$n_T - k$	MSE	
Total	SST	$n_T - 1$		

SST is partitioned into SSTR and SSE.

SST's degrees of freedom (d.f.) are partitioned into SSTR's d.f. and SSE's d.f.

ANOVA Table

SST divided by its degrees of freedom $n_T - 1$ is the overall sample variance that would be obtained if we treated the entire set of observations as one data set.

With the entire data set as one sample, the formula for computing the total sum of squares, SST, is:

$$SST = \sum_{j=1}^{k} \sum_{i=1}^{n_T} (x_{ij} - \bar{x})^2 = SSTR + SSE$$

$$SSTR = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2$$

$$SSE = \sum_{j=1}^{k} (n_j - 1) s_j^2$$

ANOVA Table

- ANOVA can be viewed as the process of partitioning the total sum of squares and the degrees of freedom into their corresponding sources: treatments and error.

- Dividing the sum of squares by the appropriate degrees of freedom provides the variance estimates and the F value used to test the hypothesis of equal population means.

Test for the Equality of k Population Means

- Example: Reed Manufacturing

Janet Reed would like to know if there is any significant difference in the mean number of hours worked per week for the department managers at her three manufacturing plants (in Buffalo, Pittsburgh, and Detroit).

Test for the Equality of k Population Means

- Example: Reed Manufacturing

A simple random sample of five managers from each of the three plants was taken and the number of hours worked by each manager for the previous week is shown on the next slide.

Conduct an F test using $\alpha = .05$.

Test for the Equality of *k* Population Means

Observation	Plant 1 Buffalo	Plant 2 Pittsburgh	Plant 3 Detroit
1	48	73	51
2	54	63	63
3	57	66	61
4	54	64	54
5	62	74	56
Sample Mean	55	68	57
Sample Variance	26.0	26.5	24.5

Test for the Equality of k Population Means

- p -Value and Critical Value Approaches

- 1. Develop the hypotheses.

H_0: $\mu_1 = \mu_2 = \mu_3$

H_a: Not all the means are equal

where:

μ_1 = mean number of hours worked per week by the managers at Plant 1

μ_2 = mean number of hours worked per week by the managers at Plant 2

μ_3 = mean number of hours worked per week by the managers at Plant 3

Test for the Equality of *k* Population Means

- Compute the test statistic using ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	*F*
Treatment	490	2	245	9.5
Error	308	12	25.67	
Total	798	14		

Test for the Equality of k Population Means

- p –Value Approach

- 4. Compute the critical value.

With 2 numerator d.f. and 12 denominator d.f., $F_\alpha = 3.89$.

- 5. Determine whether to reject H_0.

The F > F_α, so we reject H_0.

We have sufficient evidence to conclude that the mean number of hours worked per week by department managers is not the same at all 3 plant.