Chemometrics

MA4605

Week 4. Lecture 8. ANOVA

September 27, 2011

ANalysis Of VAriance

ANOVA is a method of testing the equality of 3 or more population means by analysing the sample variances.
ANOVA is used in applications such as:

- compare the mean concentration of an analyte by several different methods.

ANalysis Of VAriance

ANOVA is a method of testing the equality of 3 or more population means by analysing the sample variances.
ANOVA is used in applications such as:

- compare the mean concentration of an analyte by several different methods.
- compare the mean titration results obtained by several different experimenters using same apparatus.

ANalysis Of VAriance

ANOVA is a method of testing the equality of 3 or more population means by analysing the sample variances.
ANOVA is used in applications such as:

- compare the mean concentration of an analyte by several different methods.
- compare the mean titration results obtained by several different experimenters using same apparatus.

Example Three different groups of people such as

- smoker
- nonsmoker(exposed to tobacco)
- nonsmoker(nonexposed)

are measured for nicotine. We can test whether on average they have the same level of nicotine.

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

$H_a : \mu_1 \neq \mu_2 \neq \mu_3$

ANOVA can be

- One-way ANOVA. It is used when the sample data are separated into groups according to 1 characteristic or factor.
- Two-way ANOVA. it is used to compare population means when the sample is separated into groups by 2 factors.

One way ANOVA tests the claim that the means are equal for k independent groups

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

H_a: not all the means are equal (at least one mean is different).

ANOVA can be

- One-way ANOVA. It is used when the sample data are separated into groups according to 1 characteristic or factor.
- Two-way ANOVA. it is used to compare population means when the sample is separated into groups by 2 factors.

One way ANOVA tests the claim that the means are equal for k independent groups

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

H_a: not all the means are equal (at least one mean is different).

Assumptions

- Each population from which a sample is taken is assumed to be normal.

ANOVA can be

- One-way ANOVA. It is used when the sample data are separated into groups according to 1 characteristic or factor.
- Two-way ANOVA. it is used to compare population means when the sample is separated into groups by 2 factors.

One way ANOVA tests the claim that the means are equal for k independent groups

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

H_a: not all the means are equal (at least one mean is different).

Assumptions

- Each population from which a sample is taken is assumed to be normal.
- Each sample is randomly selected and independent.

ANOVA can be

- One-way ANOVA. It is used when the sample data are separated into groups according to 1 characteristic or factor.
- Two-way ANOVA. it is used to compare population means when the sample is separated into groups by 2 factors.

One way ANOVA tests the claim that the means are equal for k independent groups

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

H_a: not all the means are equal (at least one mean is different).

Assumptions

- Each population from which a sample is taken is assumed to be normal.
- Each sample is randomly selected and independent.
- The populations have approximately equal variances(standard deviations).

Notations

- x_{ij} is the i^{th} observation of group j

Notations

- x_{ij} is the i^{th} observation of group j
- $\overline{x}_j$ is the mean of group j.
- s_j^2 is the variance of group j.

Notations

- x_{ij} is the i^{th} observation of group j
- $\overline{x}_j$ is the mean of group j.
- s_j^2 is the variance of group j.
- $\overline{\overline{x}}$ is the overall mean for all the observations
- n_j is the number of observations in group j.

Example

The results obtained in an investigation into the stability of of a fluorescent reagent stored under three different conditions. The values for the fluorescence signals are:

	Group1	Group2	Group3
	23	27	24
	23	29	26
	20	25	24
	21	23	
		24	
sample size n	4	5	3
sample means	21.75	25.6	24.67
sample variance	2.25	5.8	1.33

$\bar{\bar{x}}$=24.08

ANOVA distinguishes between two sources of variation: between groups and within groups.
To distinguish between groups, the variability between groups must be greater than the variability within groups.
This way we decide if the groups are significantly different.

Total variation = Total Sum of Squares= TSS
$=\sum_{i=1}^{n}(x_i - \overline{\overline{x}})^2$
$=(23-24.08)^2+(23-24.08)^2+...+(24-24.08)^2$
$=66.92$
Degrees of freedom= n-1=12-1=11

Number of groups is k=3.
Variation between groups = Sum of Squares Between groups
= SSB =$\sum_{j=1}^{k} n_j(\overline{x}_j - \overline{\overline{x}})^2$
=$4(21.75 - 24.08)^2 + 5(25.6 - 24.08)^2 + 3(24.67 - 24.08)^2$
=34.3
Degrees of freedom= Number of groups-1=k-1=3-1=2

Variation within groups = Sum of Squares Within groups
= Sum of Squared Errors = SSE
$=\sum_{j=1}^{k}(n_j - 1)s_j^2$
=3(2.25)+5(5.8)+2(1.33)
=32.62
Degrees of freedom$=\sum_{j=1}^{k}(n_j - 1)$=3+4+2=9

Rules

- Total variation= Variation between groups = Variation within groups
  66.92 = 34.3+ 32.62
- Total df= Between groups df+ Within groups df
  11= 2+9

Calculate the test statistic

- Mean Square Between groups= MSB = $\frac{SSB}{df\ \ between} = \frac{34.3}{2} = 17.15$
- Mean Square Within groups= MSE = $\frac{SSE}{df\ \ within} = \frac{32.62}{9} = 3.62$
- Test Statistics= F= $\frac{MSB}{MSE} = \frac{17.15}{3.62} = 4.7$
- F $> F_{2,9} = 4.26$, hence we reject the null hypothesis.