Analysis of Variance

Mahesh Ganupooru & Rammohan Rayicherla
ICON Clinical Research

03 Dec 2016

- Introduction
- Assumptions
- Other aspects
- Programming
- Output

Introduction

Analysis of Variance (ANOVA) is a statistical method used to test differences between two or more means.

Why "Analysis of Variance" why not "Analysis of Means"?

Mean alone can't give perfect picture of the sample data. It is always accompanied by variance/SD.

Inferences about means are made by analyzing variance.

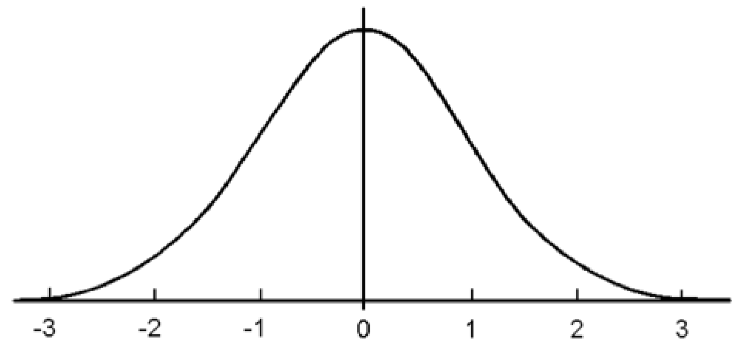
It gives the overall comparison results than pairwise comparisons like TRT1 vs. TRT2 or TRT2 vs TRT3

Generalized form of the t-test for two groups.

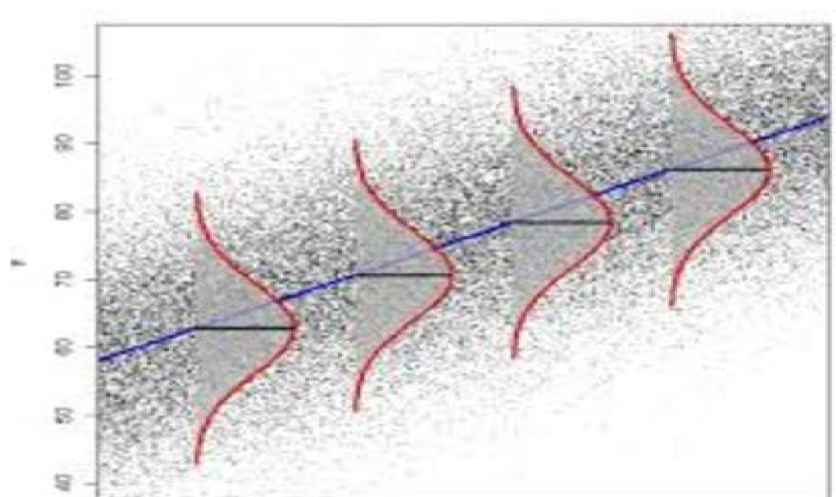
Assumptions of ANOVA

ANOVA can be applied in terms of Linear model based on the following assumptions about the nature of the outcome variable.

1. Independence of observations
2. Normality of the outcome variable



3. Homoscedasticity i.e. equality of variances among groups



Types of effects

Random-effects: Where the values of an independent variables observed are random. For example Baseline value

Fixed-effects: Where the values of an independent variables fixed or predetermined. For example Treatment

Mixed-effects model: A mixed-effects model contains independent variables of both fixed and random-effects types

Characteristics of ANOVA

ANOVA's results are independent of constant bias and scaling errors as well as the units used in expressing observations i.e. Adding or multiplying a constant to all observations does not alter significance.

Test Statistic:

ANOVA uses F-test statistic, which is based on the sum of squares

Types of Data

Balanced Data: when all possible groups of independent variables have equal number of size

Example for Balanced data Table of Treatment by Center							
Number of subjects	*Center*						
Treatment	*01*	*02*	*03*	*04*	*05*	*06*	*Total*
Test	6	6	6	6	6	6	36
Reference	6	6	6	6	6	6	36
Placebo	6	6	6	6	6	6	36
Total	18	18	18	18	18	18	108

Types of Data

Unbalanced Data: when all possible groups of independent variables have unequal number of size

Example for unbalanced data Table of Treatment by Center							
Number of subjects	*Center*						
Treatment	*01*	*02*	*03*	*04*	*05*	*06*	*Total*
Test	6	5	4	7	8	6	36
Reference	5	4	0	7	8	7	31
Placebo	6	4	2	6	8	6	32
Total	17	13	6	20	24	19	99

Types of Data & Models

Sequential Sum of Squares (Type - I): Only suitable for balanced data

Hierarchical or Partially Sequential Sum of Squares (Type - II): Cannot be used for models with interaction terms

Marginal or Orthogonal Sum of Squares (Type - III): Most suitable for unbalanced data with no missing cells in the design

Goodnight or Balanced Sum of Squares (Type - IV): A variation of type III, but specifically developed for designs with missing cells

Programming in SAS®

PROC ANOVA: SAS® recommends to use PROC ANOVA for the analysis of balanced data only, with the some exceptions.
It is faster and uses less resources than PROC GLM for balanced data.

SYNTAX:

```
PROC ANOVA DATA=data1;
   CLASS  TRT CENTER;
   MODEL  RSLTN= TRT CENTER;
RUN;
```

Programming in SAS®

PROC GLM: It is much advanced and produces more useful results.
Suitable for unbalanced data.
Analyzes both Balanced and Unbalanced data.

SYNTAX:

```
PROC GLM DATA=data1;
   CLASS  TRT CENTER;
   MODEL  RSLTN= TRT CENTER TRT*CENTER;
   LSMEANS TRT;
RUN;
```

Output / Mock-shell

Protocol No.: PG-1001-2016 Page X of Y

Table 14.2.1.1. Statistical Analysis of Change from Baseline in Average Index Knee Pain at Week 4
Full Analysis Set

					Difference				
							95% Confidence Interval		
Treatment Group / Comparisons	N	n	LS Mean	SE	Difference	SE of Difference	Lower	Upper	p-value
Placebo	xxx	xxx	xx.xx	xx.xx					
PG-1001 10 mg BID	xxx	xxx	xx.xx	xx.xx					
PG-1001 20 mg BID	xxx	xxx	xx.xx	xx.xx					
PG-1001 10 mg vs. Placebo					xx.xx	xx.xx	xx.xx	xx.xx	0.xxxx
PG-1001 20 mg vs. Placebo					xx.xx	xx.xx	xx.xx	xx.xx	0.xxxx
PG-1001 10 mg vs. PG-1001 20 mg					xx.xx	xx.xx	xx.xx	xx.xx	0.xxxx

Abbreviations: LS Mean = Lease Square Mean; SE = Standard Error.
N=total number of unique patients in the longitudinal model. n=number of patients assessed at each week.
Results are based on analysis of variance model with treatment and country as fixed effects and baseline score as a covariate.

Source Listing: 16.2.1
Program ID: Study xxxxxx.SAS. Source Data: xxxx File ID: xxxxxx.HTM. Transfer Date: DDMMMYYYY Runtime ID: DDMMMYYYY HH:MM


THANK YOU!