Seminar in Statistics and methodology

Wednesday, 9 April 2008

ANOVA

Eleonora Rossi
e.rossi@rug.nl

ANOVA
Analysis of Variance

ANOVA

Why not several t-tests? The Familiwise errorrate

Analysis of Variance (ANOVA)

- ANOVA: Analysis of Variance
- ANOVA compares 3 or more independent variables (groups)

  G_1, G_2, G_3….G_i
- Suitable for numeric and ordinal data
- When we compared two groups we were using t-test

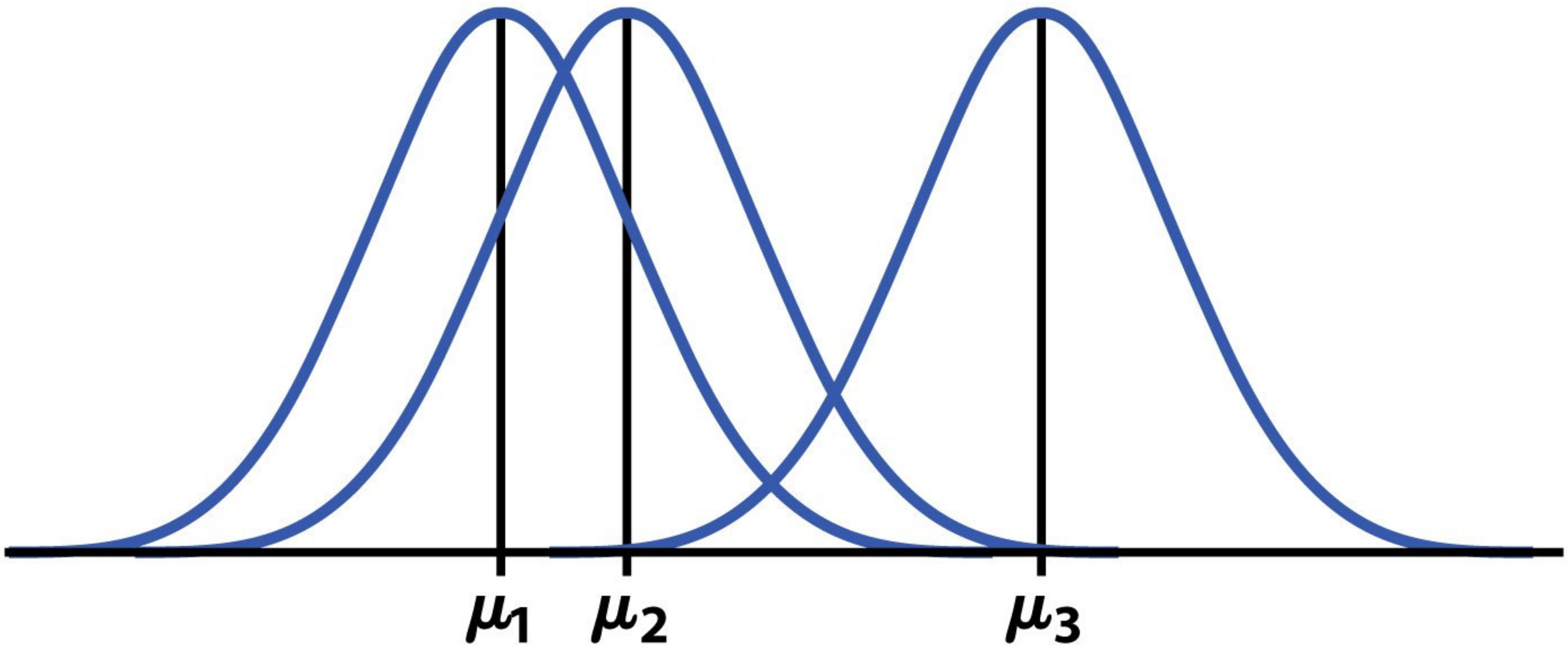


Figure 12-6
Introduction to the Practice of Statistics, Fifth Edition

Why not several t-tests?

- Why not to run several t-tests?
- Imagine we have a design with three groups that have to be compared:
  - G1, G2, G3
- We will have to run several separate t-tests (one to compare G1 with G2, one to compare G1 with G3, and one to compare G2 with G3)
- For every test we use a general α-level of 0.05

α-level

- α-level=0.05
  - 5% possibility to make Type I error, i.e. rejecting H_0, when H_0 is actually true.
  - 95% possibilities not too make type I error.
  - Our scope is too reduce the possibilities to have Type I error
- If we were to run 3 separate t-tests to compare G1, G2 and G3, each with a α-level of 0.05, the overall possibility not to make Type I error would be 0.857 [i.e. $(0.95)^3$]
- Therefore subtracting that from the overall possibility not to make Type 1 error (1=100%)
  - 1-0.857=0.14
  - We have 14% of possibilities to make Type 1 error.
  - 14% >> than the usual 5%
  - We can't be happy with that!

F_{ER}: The familiwise errorrate

$$F_{ER}: 1 - (0.95)^n$$

- Where n is the number of tests that have to be carried out
- The larger the number of tests that have to be carried out the larger the possibility to have Type I error
- Example with 4 groups
  - $1 - (0.95)^6 = 0.27$ →
  - 27% of possibilities to make type I error!!

ANOVA: general concepts

- The familywise errorrate is the reason why ANOVA is used instead of single t-tests
- ANOVA is an *omnibus* test (recall Latin! Omnibus=Adjective 2nd declination neutral plural= all)

↓

- ANOVA tell us the *overall* difference among the groups but it does not say anything about possible differences among groups

ANOVA: The Hypotheses

HYPOTHESES FOR ONE-WAY ANOVA

The **null and alternative hypotheses** for one-way ANOVA are

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

$$H_a: \text{not all of the } \mu_i \text{ are equal}$$

Definition, pg 730
Introduction to the Practice of Statistics, Fifth Edition

Analysis of Variance (ANOVA)

- H_0: all groups come from the same population; have the same means!
- We draw a SRS from each population and we use data to test H_0
- Do all groups have the same population mean? We need to compare the sample means

↓

We compare the variation among (<u>between</u>) the means of several groups with the variation <u>within</u> groups

ANOVA as a regression model

Taking the general equation to predict data

Outcome= Model + Error

It is possible to derive a simple regression equation

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

With which it is possible to predict one variable from a single predictor

ANOVA as a regression model

- ANOVA can be seen as a *multiple regression analysis*, given that it is possible to predict one variable (y) from several predictors

$$Y_i = b_0 + b_2 X_i + b_1 X_i + \varepsilon_i$$

In regression analysis, R^2 (coefficient of determination) expresses the ratio between the variance explained by the model (SS_M) and the variance expressed by the data (SS_T),

$$R^2 = SS_M / SS_T$$

ANOVA as a regression model

- Similarly, the ANOVA analysis can be seen as a comparison between a *systematic* variance (given by the model) and a *unsystematic* variance (given by the data). This is measured with the F-ratio

$$F = MS_M / MS_R$$

The model (ANOVA)

Outcome= Model + Error

or

DATA= FIT + RESIDUALS

↓

In terms of variance

$$TOTAL = BETWEEN_{group} + WITHIN_{group}$$

ANOVA: starting from the Sums of Squares Model

1. TOTAL SUMS OF SQUARES SS_T
   Finds the variation within the data
   $SS_T = \sum (X_i - X_{grand\ mean})^2$
   Where X_i is the value for each observed datapoint
   df (n-1)

2. Model sums of squares SS_M
   Explains how much variation the regression model can explain
   $SS_M = \sum n (X_k - X_{grand\ mean})^2$
   Where n is the number of participants in the group, and X_k is the mean for each group
   df (K-1)

3. Residual sums of squares SS_R
   Explains how much variation the regression model can not explain
   $SS_R = SS_T - SS_M$ ($SS_M = SS_T + SS_R$)
   $df_R = df_T - df_M$

SUMS OF SQUARES, DEGREES OF FREEDOM, AND MEAN SQUARES

Sums of squares represent variation present in the data. They are calculated by summing squared deviations. In one-way ANOVA there are three **sources of variation:** groups, error, and total. The sums of squares are related by the formula

$$\text{SST} = \text{SSG} + \text{SSE}$$

$$SS_T = SS_M + SS_R$$

Thus, the total variation is composed of two parts, one due to groups and one due to error.

Degrees of freedom are related to the deviations that are used in the sums of squares. The degrees of freedom are related in the same way as the sums of squares:

$$\text{DFT} = \text{DFG} + \text{DFE}$$

To calculate each **mean square,** divide the corresponding sum of squares by its degrees of freedom.

Definition, pg 734
Introduction to the Practice of Statistics, Fifth Edition

ANOVA: Mean squares

However, given that the SS_T and SS_M are summed values coming from a different numbers in the sum, usually having SS_T always a higher number of summed elements than SS_M

Mean squares are therefore used: SST and SSM are divided (weighted) by their df.

1. $MS_T = SS_T / df_T$
2. $MS_M = SS_M / df_M$

   It is the *average variation* that can be explained by the model (systematic variation). It is an estimate for the between group variance
3. $MS_R = SS_R / df_R$

   It is the average variation that can not be explained by the model (unsystematic variation). Estimate for the within group variance

ANOVA:
The F-ratio (or f-statistics)

It is the ratio (it always assumes values >0) between the amount of systematic variance and the amount of unsystematic variance

$$F = MS_M / MS_R$$

MS_M: Systematic variance. The variance explained by the model

MS_R: Unsystematic variance. The variance that is not explained by the model

If $F < 1 \rightarrow MS_R > MS_M$

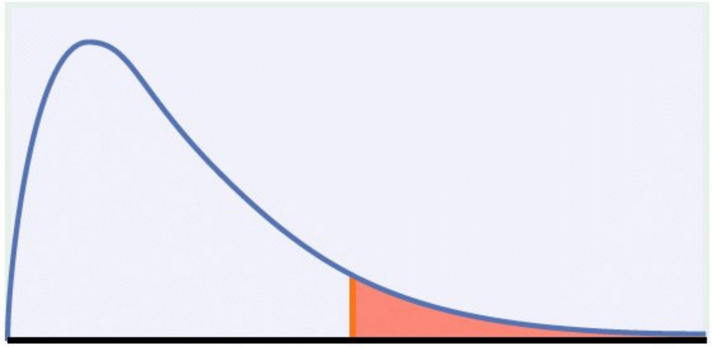
If $F > 1 \rightarrow MS_R < MS_M$

THE ANOVA F TEST

To test the null hypothesis in a one-way ANOVA, calculate the ***F* statistic**

$$F = \frac{\text{MSG}}{\text{MSE}}$$

F = MSM /MSR



When H_0 is true, the F statistic has the $F(I - 1, N - I)$ distribution. When H_a is true, the F statistic tends to be large. We reject H_0 in favor of H_a if the F statistic is sufficiently large.

The ***P*-value** of the F test is the probability that a random variable having the $F(I - 1, N - I)$ distribution is greater than or equal to the calculated value of the F statistic.

Definition, pg 735
Introduction to the Practice of Statistics, Fifth Edition

ANOVA: Assumptions 1

- Data should come from a normal distribution
- Variances should be fairly similar
  - The Levine's test for equality of variances (SPSS) tells us if the variances significantly differ or not
  - If sample sizes are equal ANOVA is quite robust to control for the difference in variances
  - If sample sizes are different ANOVA is not robust in case variances differ
- Data should be independent
- Data at least in an ordinal scale

Robustness of ANOVA

RULE FOR EXAMINING STANDARD DEVIATIONS IN ANOVA

If the largest standard deviation is less than twice the smallest standard deviation, we can use methods based on the assumption of equal standard deviations, and our results will still be approximately correct.[2]

Definition, pg 729a
Introduction to the Practice of Statistics, Fifth Edition

- Pretty robust against violations of normality and equal variances
- However, a combination of *unequal sample sizes* and *unequal variances* can sometimes cause larger type I error (significant result while it should not be)

ANOVA: Assumptions 2

- Data if one of the basic assumptions is violated (i.e. distribution is not normal, variances are not equal and group sizes are different)

⇩

- The non parametric version of the ANOVA test can be applied:

  KRUSKAL-WALLIS TEST

  Similar ideas about other non parametric tests (they use rankings. Ranking all data from smallest to highest and then assign a rank)

ANOVA: the SPSS output

$F = MS_M / MS_R$

ANOVA

SCI

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4662.233	2	2331.116	7.137	.001
Within Groups	191729.2	587	326.626		
Total	196391.4	589			

Figure 12-8
Introduction to the Practice of Statistics, Fifth Edition

- Uses *Sums of Squares* and *Degrees of freedom*
- DF_M= Groups -1
- DF_R= N-groups
- DFT= N-1

SPSS

- Makes it easy to divide total variation (sums of squares) into between group and within group variation:
- TOTAL = BETWEEN + WITHIN
- DATA (total) = FIT (between) + RESIDUAL (within)

ANOVA

SCI

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4662.233	2	2331.116	7.137	.001
Within Groups	191729.2	587	326.626		
Total	196391.4	589			

Figure 12-8
Introduction to the Practice of Statistics, Fifth Edition

ANOVA: What does the F-value tell us

- The F-value tells us if there is a difference among groups but in case there is a difference it does not tell us which differences are among them
- Running single t-tests analyze the single differences among groups is not a good method (as seen before)
- Two possibilities are then available

CONTRASTS

Post hoc analyses

Contrasts

- *A priori* or planned comparisons
- Say you have mean IQ scores of three groups:

  A: journalists; B: politicians; C: animal activists
- ANOVA only tells you whether A=B=C or not.
- To get more information, you could *a priori* specify what the data would look like
- For instance: A = B < C
- This can be tested, even when the H_0 of A=B=C cannot be rejected!

Post hoc analysis

- Otherwise Post-hoc comparisons:
- Only possible after the H_0 has been rejected: the effect of a factor *must be significant* before it is allowed to compare the different levels! That is not necessary for contrasts
- Frequently, the post-hoc tests are less powerful = less able to show significant results

Beyond ANOVA: CONTRASTS & Post hoc analysis

CONTRASTS

- Are a priori comparisons (i.e. the ideas about the contrasts should be thought even before seeing the F-value.
- We can think about them as a one-way analysis
  - We already have in mind some possible effects among the groups
  - Example: we want to test the efficacy of a new version of a medicament (Pressurin) to control high blood pressure

    G1= control group (treated with a placebo)

    G2= group treated with Pressurin

    G3= group treated with the old version Pressurin

    We expect (even before running the ANOVA analysis that the control group will not show an amelioration of the level of blood pressure and G2 and G3 will show an amelioration in the level of blood pressure.

There are always (k-1) contrasts

How to create CONTRASTS

Contrast Coefficients

Contrast	JOBC		
	1	2	3
1	-.5	-.5	1
2	1	-1	0

Figure 12-9a
Introduction to the Practice of Statistics, Fifth Edition

Contrast Tests

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
SCI	1	9.69	2.736	3.543	587	.000
	2	-.78	2.078	-.378	587	.706

Figure 12-9b
Introduction to the Practice of Statistics, Fifth Edition

Beyond ANOVA: CONTRASTS & Post hoc analysis

POST HOC ANALYSES (data mining or exploratory data analysis)

- There are no a priori ideas about what specific differences there will be.
- Are pair wise comparisons that compare all the different combinations among groups

- Post –hoc analyses control the familywise error by maintaining a general α- level across all comparisons = 0.05

Post-hoc analyses: Bonferroni correction

- Possible course of action: divide the acceptable level of significance (in general: 0.05) by the number of comparisons (for the previous example: 3)
  - For each comparison the α-level has to be lowered to: 0.017
- The Bonferroni correction is quite a conservative procedure, but if we want to have a firm control on the Type I error it is a good procedure to use.

Multiple Comparisons

Dependent Variable: SCI

Bonferroni

(I) JOBC	(J) JOBC	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-.78	2.078	1.000	-5.77	4.20
	3	-10.09*	2.671	.001	-16.50	-3.67
2	1	.78	2.078	1.000	-4.20	5.77
	3	-9.30*	3.161	.010	-16.89	-1.71
3	1	10.09*	2.671	.001	3.67	16.50
	2	9.30*	3.161	.010	1.71	16.89

*. The mean difference is significant at the .05 level.

Figure 12-10

Introduction to the Practice of Statistics, Fifth Edition