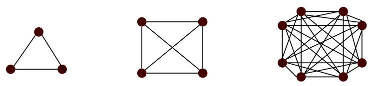
ANOVA

Analysis of Variance

- Analysis of variance:
  - One-way
  - Two-way
  - Repeated measures
  - Non-parametric

Comparing Means of Several Groups

- T-test compares means of two groups
- What if we have more?
- Perform multiple t-tests
  - Multiple testing issues



3 groups, 3 tests | 4 groups, 6 tests | 8 groups, 28 tests

in general: n groups give n(n-1)/2 tests

One Way Analysis of Variance

- Alternative is one way ANOVA
- Illustrate with example
  - Does blood group have an effect on blood pressure in mmHg

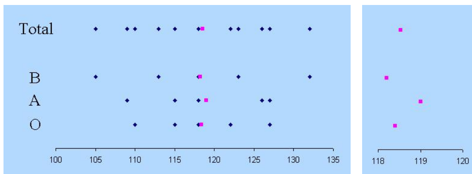
O	A	B
115	127	105
122	126	123
118	115	132
127	118	113
110	109	118

Outline

- Test if groups have common mean (H_0)
- Assume groups have a common variance
- Produce two estimates of variance
  - s^2_B and $s^2_{W,}$ (only s^2_B holds under H_0)
- Under H_0 $s^2_B = s^2_{W,}$ we expect $s^2_B/s^2_W = 1$
- Assume data are normally distributed
  - Distribution of s^2_B/s^2_W under H_0 is known

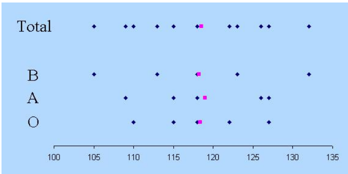
Example

- Are data scattered about common mean?
  - Do we need to account for group means
- In this case scattered about common mean



Example

- Groups have common mean
  - Variation of group means around over-all mean
  - Much less than data points around group mean



Recall – Sample Variance

- Calculated from formula:

$$\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

one over the degrees of freedom (d.f.)

sum of squared errors (sum of squares) – a measure of the variation in the sample

Partitioning the Variance

- Split total variation (sums of squares) in parts

$$\sum_{i=1}^{N}(x_i - \bar{x})^2 = \sum_{i=1}^{N}\left((x_i - m_i) + (m_i - \bar{x})\right)^2$$
$$= \sum_{i=1}^{N}(x_i - m_i)^2 + \sum_{i=1}^{N}(m_i - \bar{x})^2$$

sum of variations within each group

sum of variations between group means

where m_i is the appropriate group mean for observation x_i

Example – Total Sum of Squares

O	A	B
115	127	105
122	126	123
118	115	132
127	118	113
110	109	118

$\bar{x} = 118.53$

Variance of data = $\frac{1}{N\text{-}1}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \frac{1}{14}815.733$

Total SS (TSS)= 815.73

Example – Within Sums of Squares

O	A	B
115	127	105
122	126	123
118	115	132
127	118	113
110	109	118
118.4	119	118.2

O	A	B
11.56	64.00	174.24
12.96	49.00	23.04
0.16	16.00	190.44
73.96	1.00	27.04
70.56	100.00	0.04

squared differences from group mean

Within groups sum of squares (WSS) = $\sum_{i=1}^{N}(x_i - m_i)^2$ = 814

Example – Between Sums of Squares

O	A	B
115	127	105
122	126	123
118	115	132
127	118	113
110	109	118
118.4	119	118.2

$\bar{x} = 118.53$

O	A	B
0.018	0.218	0.111
0.018	0.218	0.111
0.018	0.218	0.111
0.018	0.218	0.111
0.018	0.218	0.111

squared differences group mean to mean

Between group means SS (BSS) = $\sum_{i=1}^{N}(m_i - \bar{x})^2$ = 1.73

Estimating Variance Between Groups

$$\frac{1}{n_O - 1}\sum_O (x_i - m_i)^2 \approx \sigma^2 \qquad \sum_O (x_i - m_i)^2 \approx (n_O - 1)\sigma^2$$

$$\frac{1}{n_A - 1}\sum_A (x_i - m_i)^2 \approx \sigma^2 \qquad \sum_A (x_i - m_i)^2 \approx (n_A - 1)\sigma^2$$

$$\frac{1}{n_B - 1}\sum_B (x_i - m_i)^2 \approx \sigma^2 \qquad \sum_B (x_i - m_i)^2 \approx (n_B - 1)\sigma^2$$

$$\text{WSS} = \sum_{i=1}^{N} (x_i - m_i)^2 \approx (N-k)\sigma^2$$

where k is the number of groups (=3 here)

Mean Sums of Squares

$$\text{WSS} = \sum_{i=1}^{N} (x_i - m_i)^2 \approx (N-k)\sigma^2$$

MSS = WSS/(N-k) is an estimate of the variance

Estimating Variance Between Groups

$$\text{BSS} = \sum_{i=1}^{N} (m_i - \bar{x})^2 \approx (k-1)\sigma^2$$

MSS = BSS/(k-1) is an estimate of the variance

The ANOVA Table

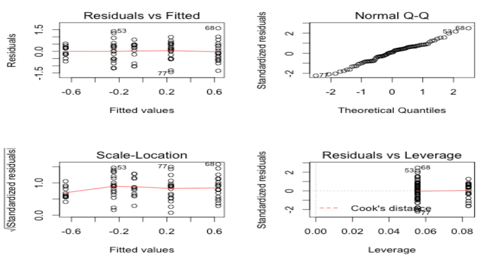
ANOVA

Blood Pressure

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	BSS	k-1	BSS/k-1	$\frac{\text{BSS/k-1}}{\text{WSS/N-k}}$	p-value
Within Groups	WSS	N-k	WSS/N-k		
Total	Total SS	N-1			

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	2	1.7	0.87	0.013	0.987
Residuals	12	814.0	67.83		

Assumptions



Summary

- Comparing means from several groups
- Assuming the data are normal
- Assuming groups have the same variance
- Constructing two estimates of that variance
- If their ratio is big reject idea of common means
- Analysing variance to test for difference in means

Post Hoc Tests

- Tukey
  - Compare all pairs of columns
- Bonferroni
  - Compare all pairs of columns or selected pairs
- Dunnett
  - Compare all columns to control column
- Trend test
  - Test for linear trend in means across columns