# ANOVA – Analysis of Variance

What is ANOVA and why do we use it?

- Can test hypotheses about mean differences between *more than 2 samples.*
- Can also make inferences about the effects of several different IVs, each with several different levels.
- Uses sample data to draw inferences about populations
- Looks at the ratio of treatment effect (differences between groups) to error (difference between individual scores and their group means)
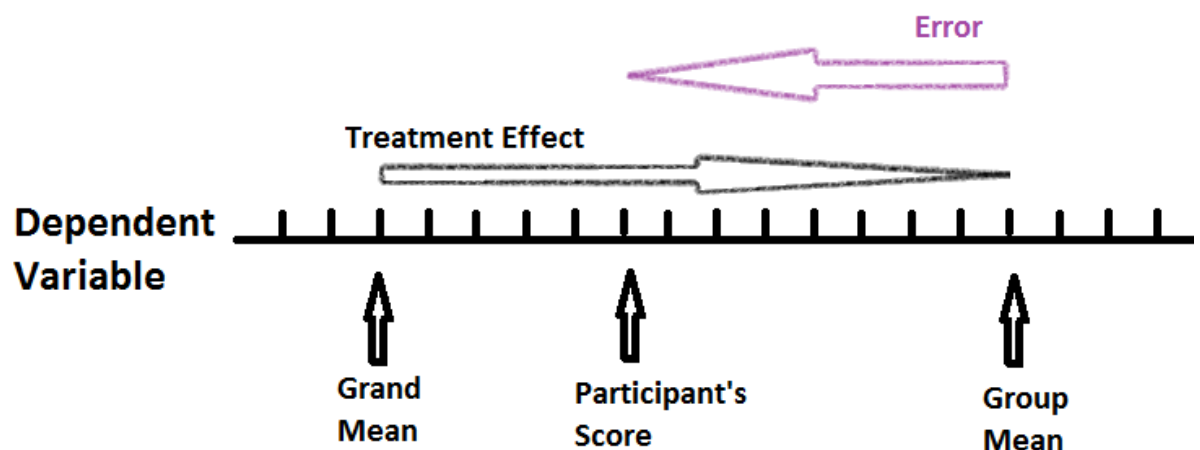
- A *factor* is an independent variable (IV).
  - It can be between-group
  - Within-subject (or repeated measures)
    - Mixed designs – a bit of both
  -
- *Main effect*
  - Effect of a factor averaged across all other factors
- *Interactions*
  - Effect of a particular combination of factors – i.e. 1 factor at a specific level of another factor.

ANOVA as Regression

- It is important to understand that regression and ANOVA are identical approaches except for the nature of the explanatory variables (IVs).
- For example, it is a small step from having three levels of a shade factor (say light, medium and heavy shade cloths) then carrying out a one-way analysis of variance, to measuring the light intensity in the three treatments and carrying out a regression with light intensity as the explanatory variable.
- Some experiments combine regression and analysis of variance by fitting a series of regression lines, one in each of several levels of a given factor (this is called analysis of covariance – ANCOVA).

- We'll come back to this later.

Why can't you just use several *t*-tests?

- If each time you do a *t*-test it carries a 5% chance of making a Type 1 Error (i.e. rejecting H_0 when it is actually true: finding an effect when none exists)
- Each time you carry out a *t*-test on the same set of data, the Type 1 Error rate increases:

Number of Tests	Type 1 Error rate
1	$1 - (.95) = 5\%$
2	$1 - (.95)^2 = 9.75\%$
3	$1 - (.95)^3 = 14.3\%$
4	$1 - (.95)^4 = 18.5\%$
5	$1 - (.95)^5 = 22.6\%$
6	$1 - (.95)^6 = 26.5\%$

- **This is the *familywise error*** – increased probability of falsely rejecting H_0
- Using ANOVA allows us make multiple comparisons with the same data, without increasing the familywise error

Logic: The *F*-Ratio

- ANOVA looks at the variance within data set to find out where it comes from
  - ➔ *Why do scores within a data set vary?*
- Compares the variance due to the IV with the variance due to individual differences.
- The difference will be significant when the between group variability is significantly more than the within-group variability.
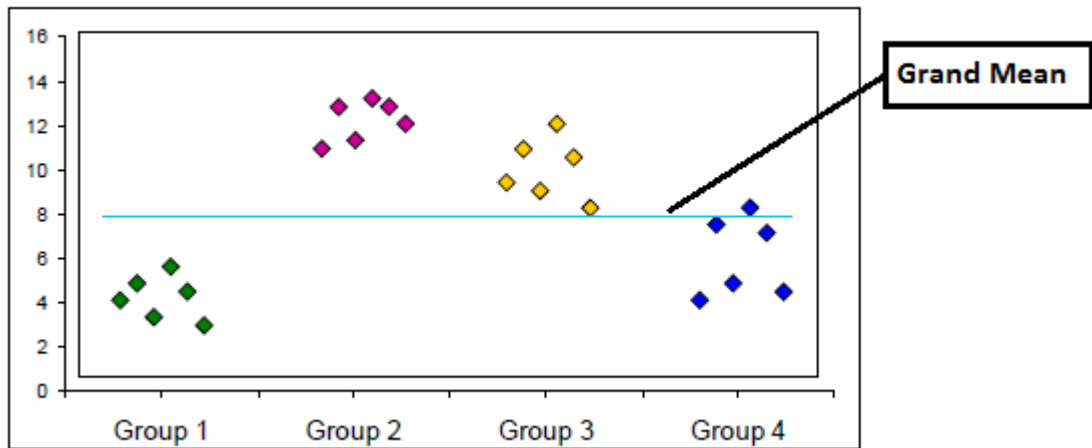
$$F = \frac{between - group\ variability}{within - group\ variability}$$
$$= \frac{treatment\ effect + error}{error}$$

If *treatment effect = 0*, then $F = \frac{error}{error} = 1$

If *treatment effect* ➔ ∞, then F ➔ ∞

(The following 3 diagrams have been adapted from Lecture notes on ANOVA1 by Dr Wendy Johnson)

- If all groups are the same then any variance between groups is due to chance – the mean of each group is the same as the grand mean, and you can predict any one score from the grand mean (of all your samples together)
  - H_0 can't be rejected; no evidence for differences between the means.
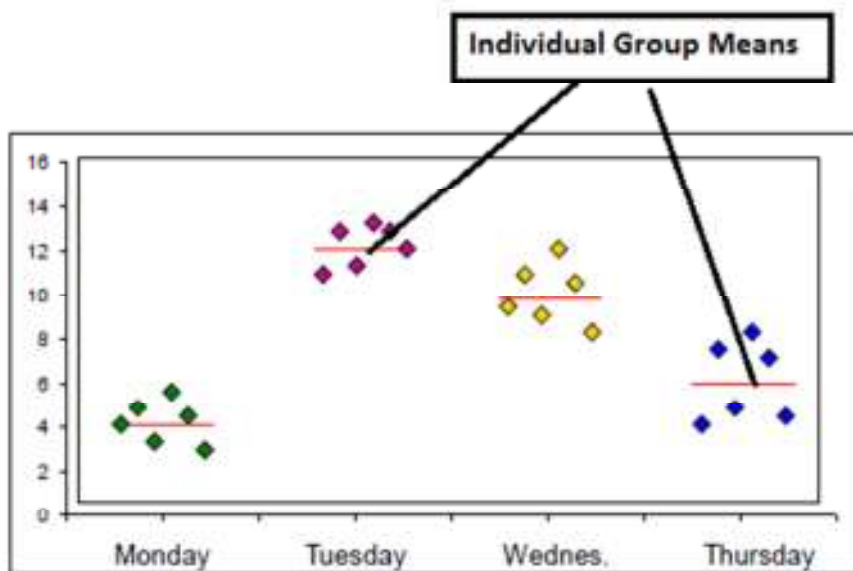
- If groups are different then the variance between groups is not due to chance alone, but a treatment effect (between groups variation) plus error (within group variation).
- You can predict one score from the mean of its individual group
  - H_0 is rejected in favour of H_1 – there is a difference between the means of the groups
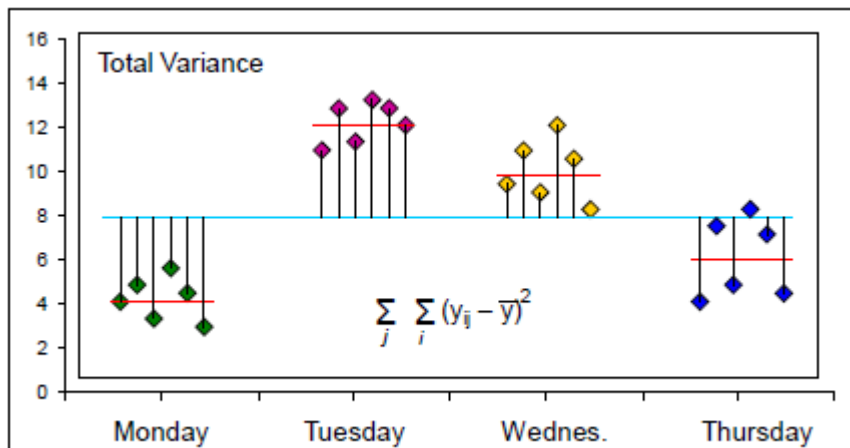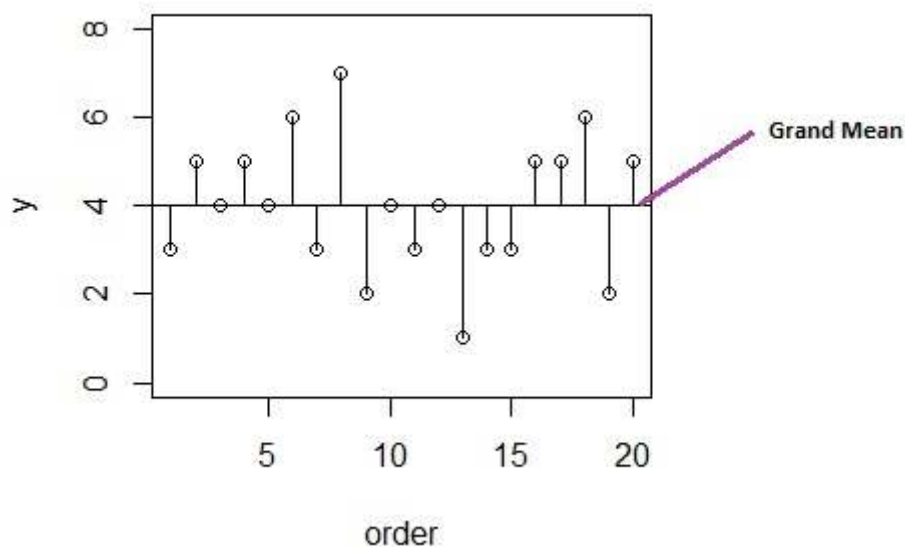
- To decide which is the better predictor, we divide all the variance into *within group variance* (a measure of how much each score differs from its group mean) and *between group variance* (how much each score differs from the grand mean)

Steps for one-way ANOVA

1. **Calculate SS_T**

   Total sum of squares

- Based on the best guess for each participant being the grand mean
- Total sum of squared differences of each score from the grand mean.

$$SS_T = \sum (x_i - \bar{x}_{grand})^2$$

$$SS_T = s^2_{grand}(n-1)$$

$$df_T = (N - 1)$$

(i.e. one less than the total sample size)

2. Calculate SS$_M$

Model sum of squares (AKA Treatment sum of squares, SS$_A$).

- This is how much variance comes from the fact that data points belong to different groups
- Based on the best guess for each participant being the mean of the group to which they belong
- The difference between each participant's predicted score (i.e. the mean of the group to which it belongs) and the grand mean
    a. Calculate difference between each group mean and the grand mean
    b. Square each of these differences
    c. Multiply each result by the number of participants within that group
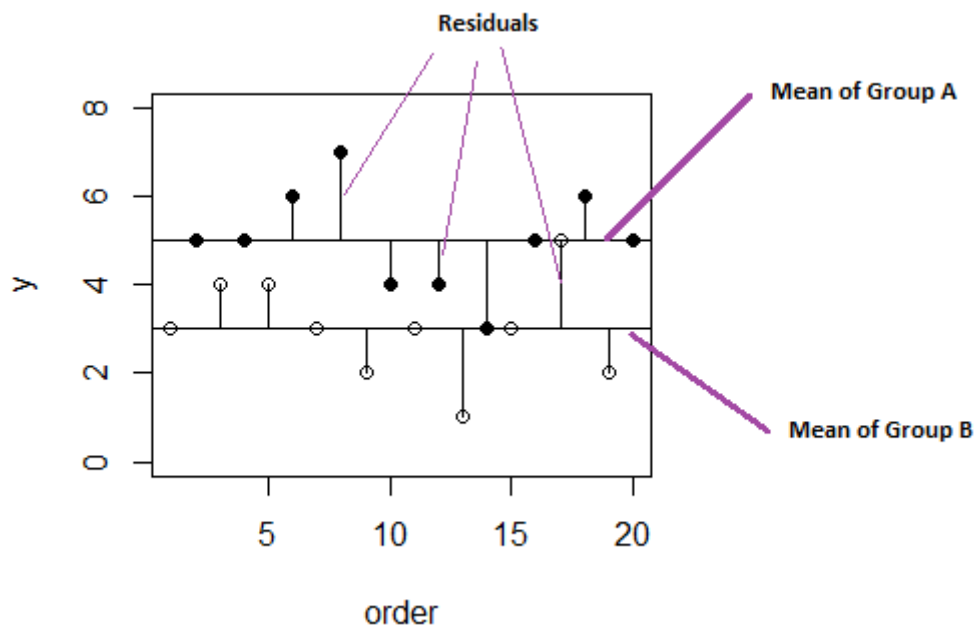   d. Add the values for each group together

$$SS_M = \sum n_k \left(\bar{x}_k - \bar{x}_{grand} \right)^2$$

$$df_M = k - 1$$

Where k is the number of groups

3. Calculate SS$_R$

Residual sum of squares (AKA error SS, SS$_E$)

- How much variation in the scores cannot be explained by the model?

- o Could be caused by individual differences – basically anything other than the experimental manipulation
- Can be calculated by:

$$SS_R = SS_T - SS_M$$

- Or – look at the differences between each score and the mean of the group it belongs to.
  - o Distance between each data point and its group mean is squared and summed.

$$SS_R = \sum (x_{ik} - \bar{x}_k)^2$$

- Or, using variance (s^2):

$$SS_R = \sum s_k^2 (n_k - 1)$$

i.e. The sum of the variance of each group multiplied by one less than the number of participants in each group.

$$df_R = df_T - df_M$$
$$= (N - 1) - (k - 1)$$
$$= N - k$$

i.e. Total sample size (N) minus the number of groups (k)

4. **Calculate MS_M and MS_R**
   - SS_M and SS_R deal with totals (sums), and are therefore influenced by how many in each sample
     - o Need to divide by df before calculating the F ratio

$$MS_M = \frac{SS_M}{df_M}$$

$$MS_R = \frac{SS_R}{df_R}$$

5. **Calculate F**

$$F(df_M, df_R) = \frac{MS_M}{MS_R}$$

1-way ANOVA table

Source	SS	df	MS	F	Critical F
Factor	SS_M	k-1	$MS_{M=}$ SS_M/k-1	$\frac{MS_M}{MS_R}$	qf(0.95,df_M,df_R)
Error	SS_R	k(n-1)	$MS_{R=}$ SS_R/k(n-1)		
Total	SS_T	kn-1			

6. **For higher-way designs**
   - **Factorial** ANOVA looks at effects of 2 or more IVs on a DV, as well as the effect of the interaction between them.

e.g. 2-way unrelated design

$$SS_{Total} = \mu + SS_\alpha + SS_\beta + SS_{\alpha\beta} + SSe_{ijl}$$

$$SS_{Total} = grand\ mean + SS_{IVa} + SS_{IVb} + SS_{IVab\ interaction} + individual\ error$$

Calculate MS

		Number of levels in A		
Source	SS	df	MS	F
Between	$\sum_j n_j(\bar{y}_j - \bar{y})^2$	$j - 1$		
A (rows)	$\sum n_a\,(\bar{y}_a - \bar{y})^2$	$a - 1$	SS_A / df_A	MS_A / MS_R
B (columns)	$\sum n_b\,(\bar{y}_b - \bar{y})^2$	$b - 1$	SS_B / df_B	MS_B / MS_R
A x B	$SS_{Bet} - SS_A - SS_B$	$df_{Bet} - df_A - df_B$	SS_{AxB} / df_{AxB}	MS_{AxB} / MS_R
Within	$\sum_j \sum_i (y_{ij} - \bar{y}_j)^2$	$df_T - df_{Bet}$		
Total	$\sum_j \sum_i (y_{ij} - \bar{y})^2$	$N - 1$	**Number of levels in B**	

ANOVA – Assumptions – when can you use an ANOVA?

When parametric assumptions are met:
1. DV is *interval* or *ratio*
2. Data is usually in *row* x *column* format – each row represents a case or participant and each column a group / category
3. Data comes from a *random sample*
4. Observations are *independent* of each other
   - Shouldn't be able to predict an observation from either other observations in its group, or from corresponding cases in other groups.
5. *Homogeneity of Variance*
   - Variability is equal across all treatment conditions/groups
     - Most common violation: floor/ceiling effects; As mean → floor/ceiling, variability → 0
     - E.g. Likert scales, Counts, Time/length/volume measurements (>0 by definition)
     - Apply Welch correction for violations – default in 1-way ANOVA in R.
   - Test using Levene's test – check for R

- o Levene test measures *heteroscedasticity* – violations of homogeneity of variance
- o If significant, then there is some reason to be worried
- o p-value will only be approximately correct
- o Brown-Forsythe/Welch tests – "robust" tests
- ANOVA is robust against this if cell sizes are equal

6. DV is *normally distributed*
- Look at frequency distribution charts
  - o Check for skew (long tails) – floor/ceiling effects, Multimodal distributions
  - o Fields: P-P plots of data versus expected
  - o 1-sample Kolmogorov-Smirnoff test
  - o Compares empirical cumulative distributions to hypothetical distribution with same mean and standard deviation
- BUT ANOVA is robust if this is violated, if sample size is large, and there is Homogeneity of Variance

7. *Sphericity* – all the covariances across pairs of levels of the IV are similar
- Mauchly's test for sphericity: if significant, the usual F-tables don't apply
- Use corrections in the corrected *F* tables
  - o Greenhouse-Geisser or Huynh-Feldt probabilities.

Types of ANOVA

1. **One-way ANOVA**
One independent Variable (IV), explanatory variable or *factor*, with 3 or more *levels* (otherwise you'd use a *t*-test – which would give the same result as a 2-level 1-way ANOVA)

e.g. Comparing the effects of 3 different teaching methods (A, B & C – 3 levels of the IV 'teaching method') on exam results

Independent Variable – Teaching Method		
Method A	**Method B**	**Method C**
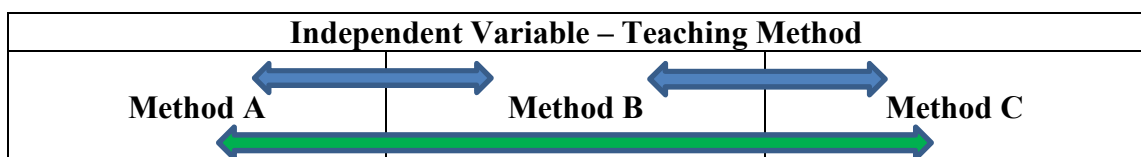
2. **Factorial**

    score= grand mean + main effects + interactions + noise

- A factor is an independent variable
- 2 or more factors each with 2 or more levels each
- All combinations of each IV occur
  - o Between-group / Within-subject (or repeated measures)

- o Or both – Mixed designs
- Random versus fixed factors

e.g. Comparing the effects of different teaching methods (between group variable) and gender (between-group variable) on exam results

Gender	Teaching Method		
	Method A	**Method B**	**Method C**
Male	Male Method A	Male Method B	Male Method C
Female	Female Method A	Female Method B	Female Method C

- *Main effects (Green dashed line in table below):*
  - o The effect of a factor averaged across all others
    - Effect of teaching method, averaged over gender
    - Effect of gender, averaged over teaching method
  - o Note that it may be misleading to discuss main effects if there is a significant interaction.
- *Interactions / Simple Main Effects (Blue solid line in table below):*
  - o Effect of a particular combination of factors
    - effect of gender at different levels of teaching method
      - e.g. is there a difference between male and female under method A?
    - effect of teaching method at different levels of gender
      - e.g. is there a difference between teaching method A, teaching method B and teaching level C within the males only?

Gender	Teaching Method		
	Method A	**Method B**	**Method C**
Male	Male Method A	Male Method B	Male Method C
Female	Female Method A	Female Method B	Female Method C

3. Repeated Measures

Each individual is exposed to multiple treatments / multiple levels of an IV

Longitudinal studies usually have repeated measures design – testing the same person over time.

e.g. Each student is tested before, during and after going through each teaching method to see if there's a difference in exam score.

4. Mixed Design

A combination of within / between design

e.g. Each student is assigned a teaching method (A, B, C), Gender is recorded (Male, Female) - both between-groups variables – and test score before and after administration of teaching method (Before, After) – within-participants variable.

Relationships between factors can be:

- CROSSED:
    - Each level of each factor is encountered at each level of each other factor (see example above)
- NESTED:
    - 2 factors are nested if each level of one occurs only with one level of the other:

e.g. Include year of study

Gender	Teaching Method					
	Method A		**Method B**		**Method C**	
Male	Male Method A		Male Method B		Male Method C	
	Year1	Year2	Year 3	Year 4	Year 5	Year 6
Female	Female Method A		Female Method B		Female Method C	
	Year1	Year2	Year 3	Year 4	Year 5	Year 6

Steps in an ANOVA
1. Check assumptions
2. Visualise data - graphs
3. Find out if there is *any* difference between any of the groups - ANOVA
   - *Omnibus Test;* H_0: $\mu_1 = \mu_2 = \mu_3$; H_1: there is a difference between at least 2 of the groups – not all means are equal.
4. Find out between which groups these differences lie:
   - *Planned comparisons* or *Post Hoc tests.*
       - e.g. H_0: $\mu_2 = \mu_3$; H_1: $\mu_2 \neq \mu_3$, or directional: H_0: $\mu_2 = \mu_3$; H_1: $\mu_2 > \mu_3$
       - use multiple *t*-tests with *Bonferonni* comparison – adjust for experimentwise error rate
       - or Fisher's Least Significant Difference (LSD)

o Dunnett 2-tailed
o **Tukey-Kramer (HSD) - preferred**
o Scheffe's

Interpreting and Reporting an ANOVA output

In this example :

- Main effect of diet: $F(2,36) = 83.52; p < .01$
- Main effect of supplement: $F(3,36) = 17.82; p < .01$
- No interaction between diet and supplement; $F(6,36) = 0.33; p = .917$

NB – report the exact p-value for a non-significant outcome, and p < .05 or p < .01 for a significant one.

Planned Comparisons and Post Hoc Tests

- Having found a difference, we need to find where the difference lies