Assumptions of ANOVA:

- Measurements are random sample
- Variable is normally distributed
- **Variance is the same in all *k* populations**

How do we handle violations in these assumptions?

1. Robustness
   - If data is not normal BUT sample size is large(CLT)
   - *variances are not equal but only if sample sizes are approximately equal*
2. Data Transformation
3. Non-parametric alternative

Kruskal-Wallis Test:

- A non-parametric test similar to a single factor ANOVA

- Uses the **ranks** of the data points
  - Data points are not compared, their ranks are!
    - Using **ranks** is what frees us from having to assume normality since all distributions have similar predictions about ranks
    - All group samples are random samples
    - Distribution of the variable has the same shape in every population
  - Small samples lead to little power but when n is large, Kruskal-Wallis has the same power as ANOVA

- ***H***, sampling distribution is χ^2 with dof = k - 1

Experimental Design:

*How do we identify **which** means are different and the **magnitude** of their difference?*

- Planned comparisons:

- Unplanned comparisons:

Experimental Design:

*How do we identify **which** means are different and the magnitude of their difference?*

- Planned comparisons:
  - **A priori** comparison between means of groups that were previously identified as particularly interesting
    - **Baked into the study design**
    - **Determined BEFORE data are examined**
  - Only small number allowed so that α isn't inflated

Experimental Design:

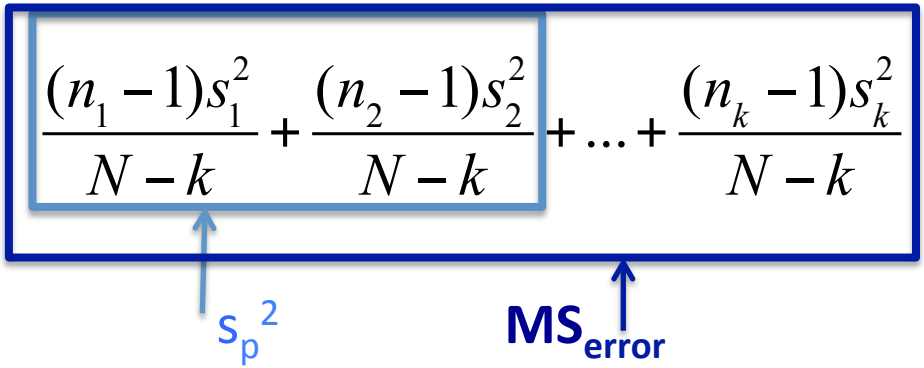
*How do we identify **which** means are different and the magnitude of their difference?*

- Planned comparisons:
  - **A priori** comparison between means of groups that were previously identified as particularly interesting
  - Only small number allowed so that α isn't inflated
  - Method:
    - Similar to two-sample t-tests
    - Use t distribution
      - Different standard error: pooled sample variance (MS_{error}) based on all *k* groups (ie. using all the information about variance rather than just a subset)
      - df of MS_{error}

$$SE = \sqrt{MS_{error}(\frac{1}{n_i} + \frac{1}{n_j})}$$

What do I mean by inflation of α?

- For a two sample t test, you are dividing up the variance of only **two** groups into the two samples.



- For a planned comparison, you are dividing up **ALL** the variance (all the total deviations of the data points) into **only two** of the **k** groups (note: you can do this because H_o assumes variance is same in all groups)
  - Big idea: this means that you have access to all the degrees of freedom provided by the data points event he ones that are in the groups we are not comparing!
  - We saw a different test that also 'absorbed' inflated error by tweaking dof (Welsh's approx t test, this reduced dof instead of expanding it)

Experimental Design:

*How do we identify **which** means are different and the magnitude of their difference?*

- Planned comparisons:
  - Example on page 472 (2nd edition)
  - **Why use instead of a two-sample t-test?**
    - Increased precision
    - Increased power
  - **Assumptions:**
    - Same as ANOVA but not as robust to violations

Experimental Design:

*How do we identify **which** means are different and the **magnitude** of their difference?*

- Planned comparisons

- Unplanned comparisons:
  - Post hoc
  - Multiple comparisons
  - Determine which means and their magnitude
  - Type **of data dredging** (interleaf) so protect against increasing α
  - Tukey-Kramer procedure tests all pairs of means

Tukey-Kramer test*:

- Already carried out a single-factor ANOVA and rejected H_0
- Compares all group means to all other group means

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

*Tukey's Honestly Significant difference (HSD) test

Tukey-Kramer test*:

- Already carried out a single-factor ANOVA and rejected H_0
- Compares all group means to all other group means

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

So why not just use a series of two-sample t-tests?

Data Dredging:

*When you use multiple tests on a data set, the **actual** probability of making **at least one** type I error, α, is larger than the significance level suggests*

- *each hypothesis test has some probability of error and these errors compound as more tests are conducted*
- *We saw this before in module 4:*
  - *two independent studies are performed to test the same null hypothesis.* What is the probability that one or both of the studies obtains a significant result and rejects the null hypothesis ***even if the null hypothesis is true***? Assume that in each study there is a 0.05 probability of rejecting the null hypothesis (Answer was 0.0975)

$P(\text{No type I errors}) = (1-\alpha)^N$, *where N = independent tests*

$P(\geq 1 \text{ type I error}) = 1 - (1-\alpha)^N$

- Why not use a series of two sample t-tests?
  - Multiple comparisons would cause the t-test to reject too many true null hypotheses
  - Tukey-Kramer adjusts for the number of tests
    - Uses larger critical value to limit Type I error
    - P(≥1 Type I error)= α
  - Tukey-Kramer also uses information about the variance within groups from all the data so it has more power than a t-test with a Bonferroni correction (data dredging interleaf): α*=α/# of tests

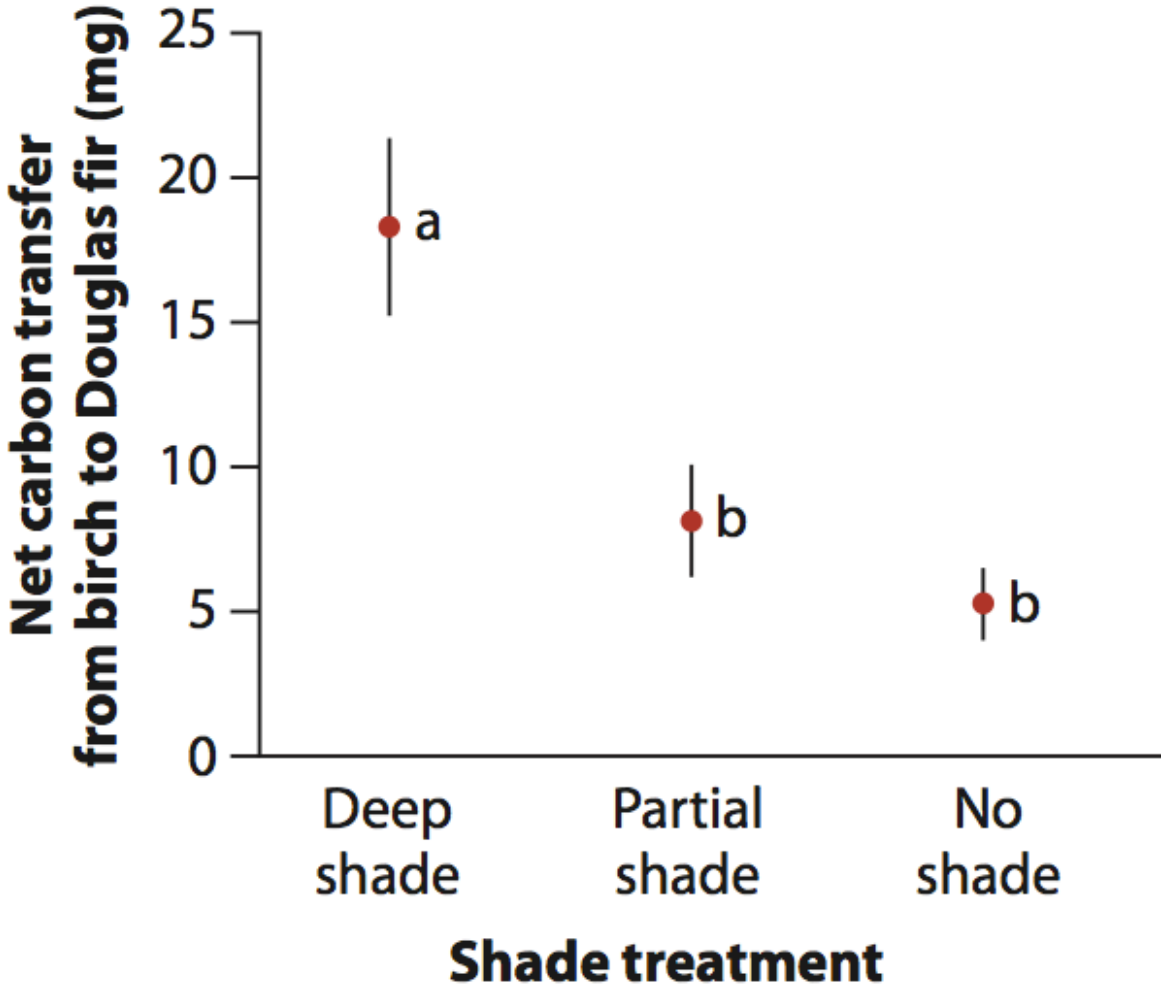
- Tukey-Kramer test:
  - Uses **q test statistic**
    - Method:
      - Order group means from smallest to largest
      - Compare each pair of group means
        - » Ex: First comparison:

          $H_0: \mu_1 - \mu_2 = 0$

          $H_A: \mu_1 - \mu_2 \neq 0$

      - Calculate **q** test statistic:
        - » Standard error: MS_{error} df: **k** and **N - k**
        - » **Q**-distribution (Statistical Table F)
  - Same assumptions as ANOVA but not as robust
    - P value is correct when design is balanced (approx same number of data points in each category) but it is conservative when unbalanced (makes it more difficult to reject the null hypothesis)

Results are displayed in a graph with specific symbols:



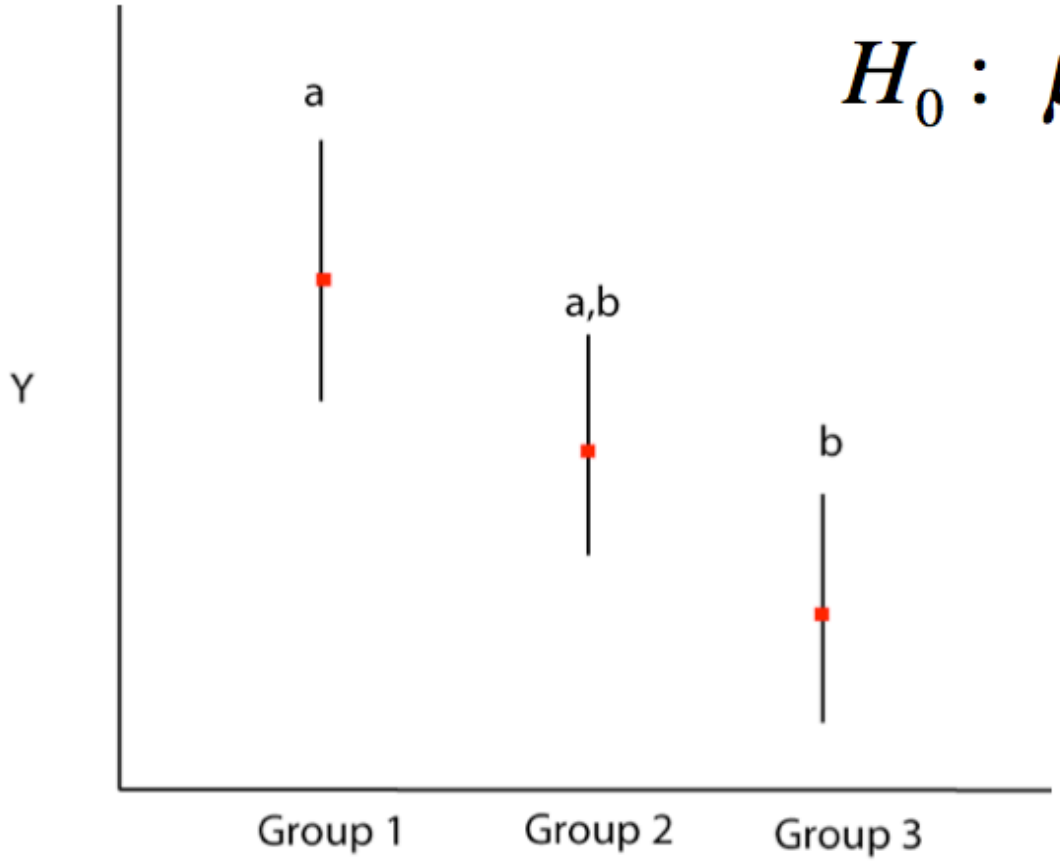
Groups whose means cannot be distinguished share the same letter

Results are displayed in a graph with specific symbols:

$H_0: \mu_1 = \mu_2$ Cannot reject

$H_0: \mu_1 = \mu_3$ Reject

$H_0: \mu_2 = \mu_3$ Cannot reject



All of the statements below about assumptions of analysis of variance (ANOVA) are true except?

A) The assumptions of ANOVA are the same as the ones used in a two-sample t-test as long as it holds for all groups.

B) The variable is normally distributed in all groups.

C) explanatory variable and response variable have a linear relationship across all groups.

D) The variance is the same in all groups.

Fixed Effects:

- Also called Model 1 ANOVA
  - What we have been using so far

- Different categories of explanatory variable are predetermined and repeatable
  - **Results cannot be generalizable**
  - Example: specific drug treatment, specific diets, specific season

Random Effects:

- Also called Model 2 ANOVA

- Different categories of explanatory variable are *randomly sampled from a larger population of groups*
  - **Results are generalizable**
    - conclusions reached about difference among groups can be generalized to the whole population
  - Example:
    - family in a study about resemblance of IQ
      - Chose a random family in a population of families
      - Family: group
      - Replicates: different children within each family
    - The population and not the particular families involved is the target of study

- Random effects
  - Used to estimate *variance components*
    - MS_{error} and MS_{groups}
    - Determine the contributions of genes and the environment to heritability:

$$V_p = V_G + V_E$$

Random Effects:

- **Random** variation within groups
  - Assumes every group has the same true variance, σ^2
  - MS_{error}
- **Variance Components:**
  - σ^2 and σ_A^2
  - Account for all the variance in the response variable
- **Random** variation among groups
  - Each group assumed to have its own mean
  - All groups means are normally distributed with grand mean: μ_A, σ_A^2

Fixed Effects:

- **Random** variation within groups
  - MS_{error}

Estimate σ_A^2 from ANOVA:

$$s_A^2 = \frac{MS_{groups} - MS_{error}}{n}$$

Repeatability:

- Applied to Random effects ANOVA
- Overall similarity of repeat measurements taken on the same group
- 0 < Repeatability < 1
- ~0 means that almost all variance in response variables results from error
- ~1 means that repeated measurements on the same group give the same answer each time

$$\text{Re}\,peatability = \frac{s_A^2}{s_A^2 + MS_{error}}$$

How do assumptions for random-effects ANOVA and fixed-effects ANOVA **differ**?

---

a) They have the same assumptions

b) The fixed-effects ANOVA assumes the groups are randomly sampled and the group means have a normal distribution in the population whereas the random-effect ANOVA does not.

c) The random-effects ANOVA assumes the groups are randomly sampled and the group means have a normal distribution in the population whereas the fixed-effect ANOVA does not.

d) The fixed-effects ANOVA assumes the variance is the same in all *k* populations and the variable is normally distributed in each of the *k* populations whereas the random-effect ANOVA does not.