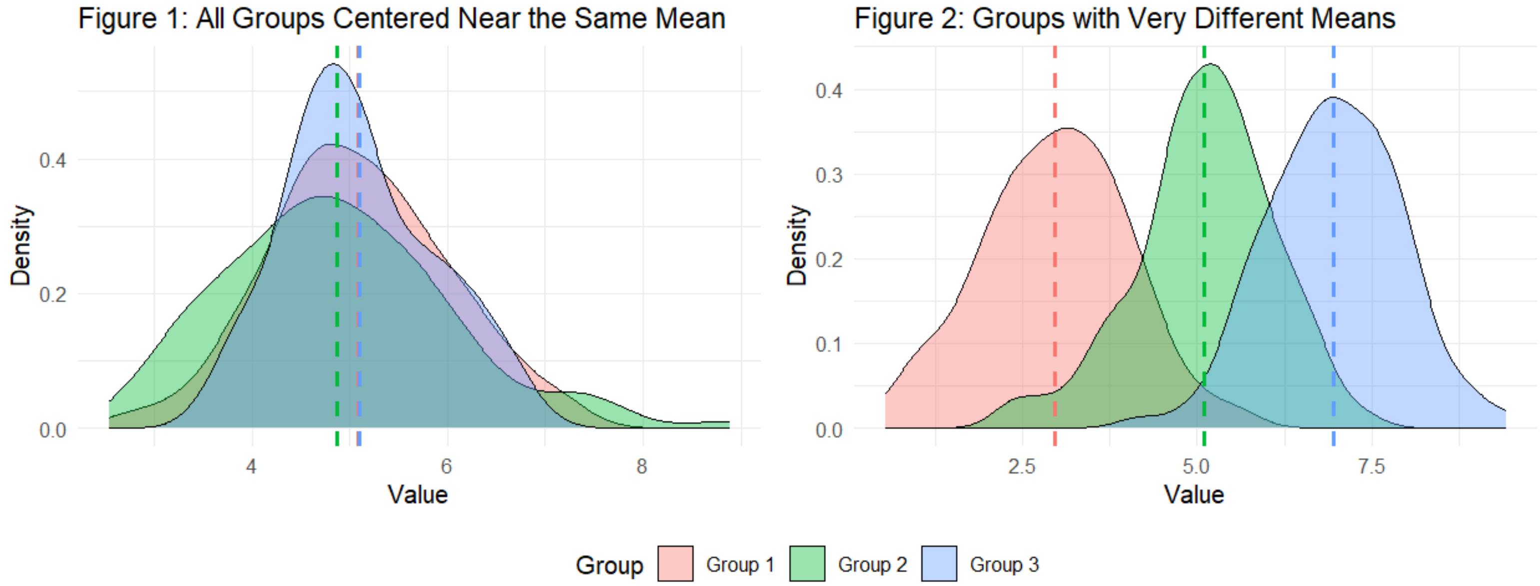
Applied Statistics for Data Scientists with R

Class 17: ANOVA and Post-hoc analysis

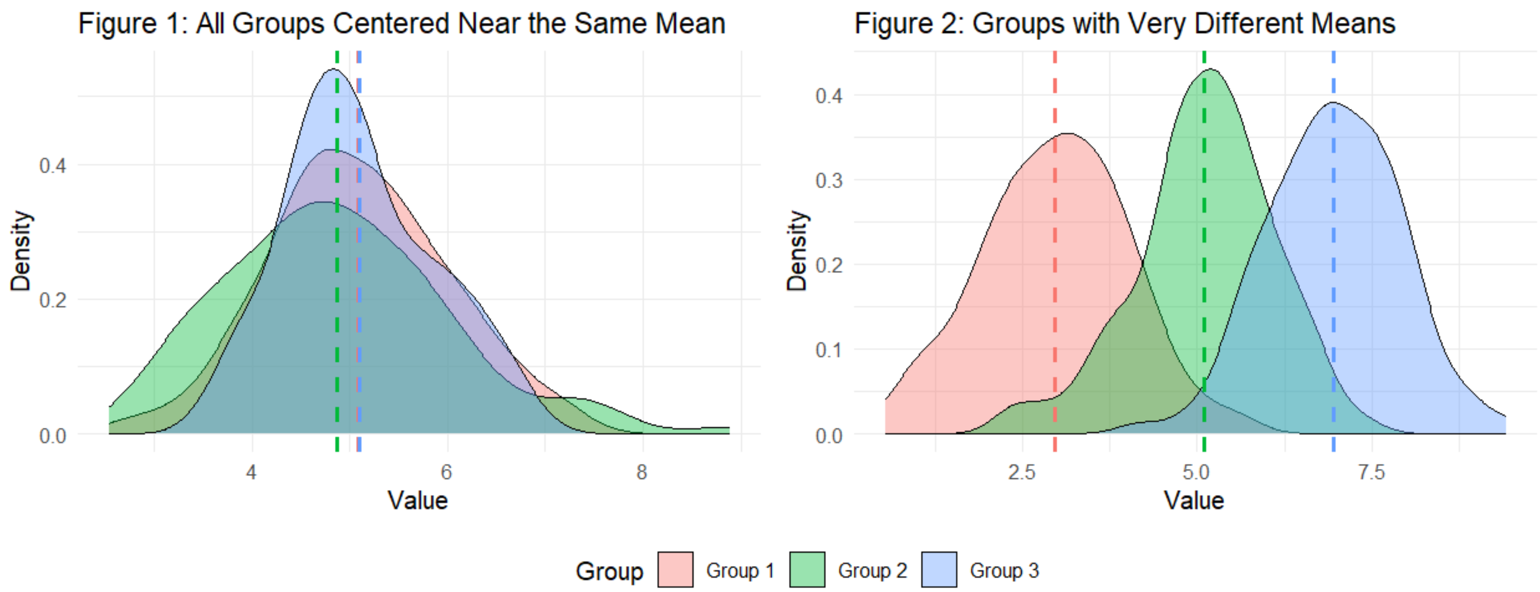
ANOVA: What is this?

- ANOVA is acronym for **An**alysis **o**f **Va**riance
- It is used to **compare** the **means** of two or more groups by analyzing variance
- ANOVA decomposes the total variability in the data into:
  - **Between-group variance** (Variation due to differences in group means)
  - **Within-group variance** (Variation within each group, assumed to be due to random error)



ANOVA: How does it work?

- If the group means were equal, the between-group variability would be small, and most of the variability would be due to differences within the groups (i.e., the error).
- By comparing the variability between groups to the variability within groups, ANOVA determines whether the observed differences between group means are greater than what we'd expect due to random variation.



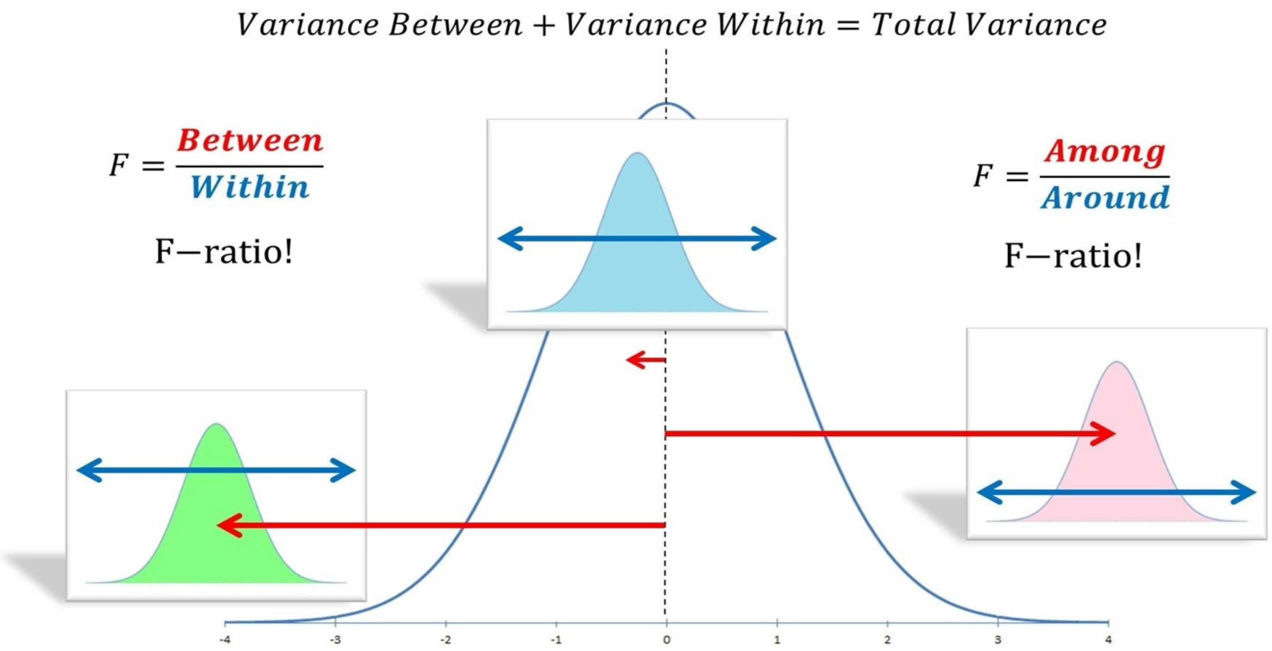
ANOVA: Hypothesis

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

H_1: At least one group mean differs from the others..

ANOVA: F-Statistic

- The test statistic in ANOVA is F-statistic.
- It is the ratio of **mean square between** (MSB) to the **mean square within** (MSW): $F = \frac{\mathrm{MS}_b}{\mathrm{MS}_w}$

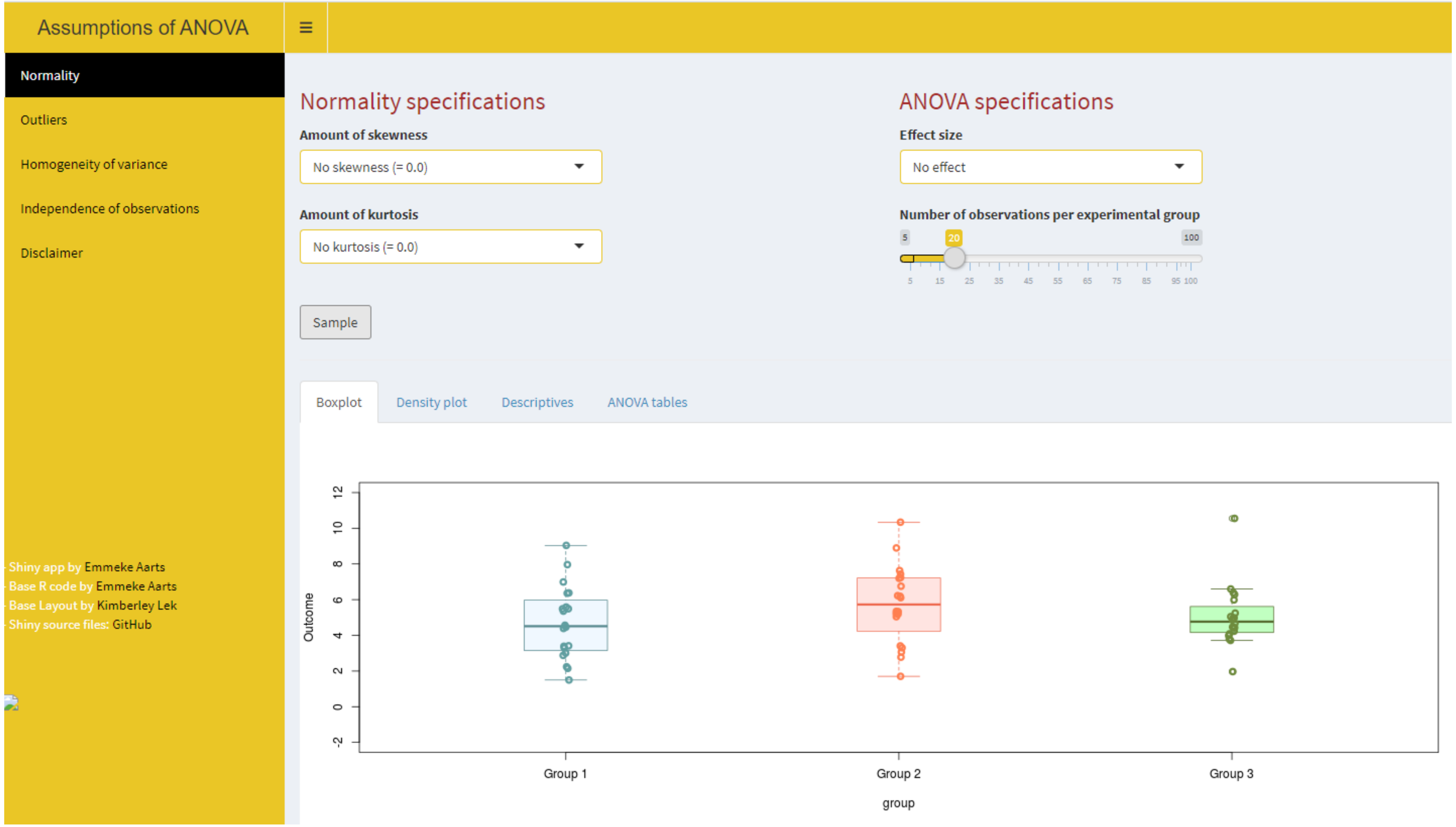


ANOVA Table Construction

Source	Sum of Squares (SS)	df	Mean Square (MS)	F
Between (b)	$SS_b = n\sum_{i=1}^{k}(\overline{X}_i - \overline{X})^2$	$k-1$	$\frac{SS_b}{k-1}$	$\frac{MS_b}{MS_w}$
Within (w)	$SS_w = \sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \overline{X}_i)^2$	$k(n-1)$	$\frac{SS_w}{k(n-1)}$	–
Total (t)	$SS_t = SS_b + SS_w$	$kn-1$	–	–

- Group mean: $\overline{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij},$
- Grand mean: $\overline{X} = \frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}.$
- Mean squares: $MS_b = \frac{SS_b}{k-1}, \quad MS_w = \frac{SS_w}{k(n-1)}$
- F-statistic: $F = \frac{MS_b}{MS_w}$

A Shiny App for Demonstration



https://utrecht-university.shinyapps.io/ANOVA-assumptions/

Post-hoc Analysis

- ANOVA tells us there is a difference, but not where

- A post hoc analysis is designed to:
  - Pinpoint specific group differences
  - Control for the inflated Type I error rate

Solution to Inflated Type-I Error

Multiple Comparison Problem

Example: Let’s say, we aim to compare the means in 3 groups (A, B, C)

Hence, we have to perform **3** separate t-tests:

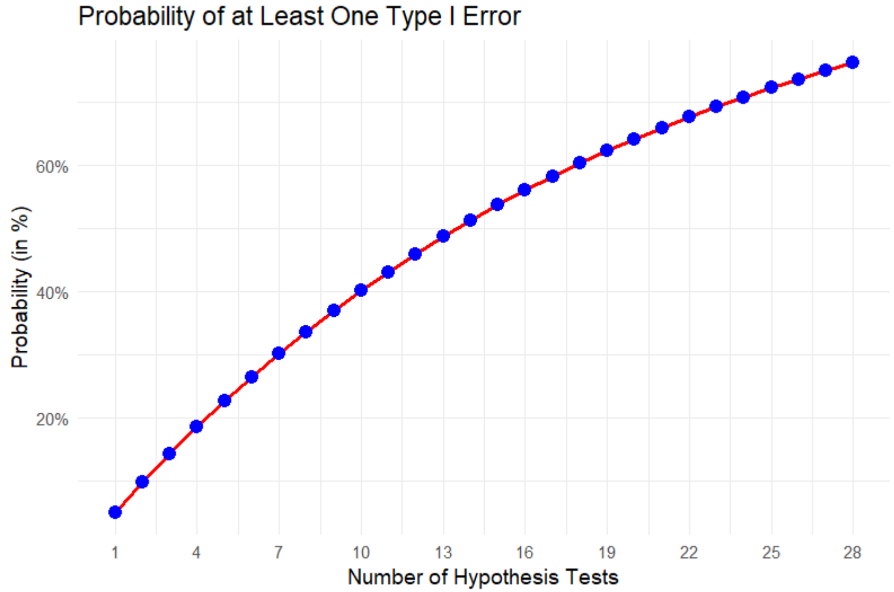
A vs. B   A vs. C   B vs. C

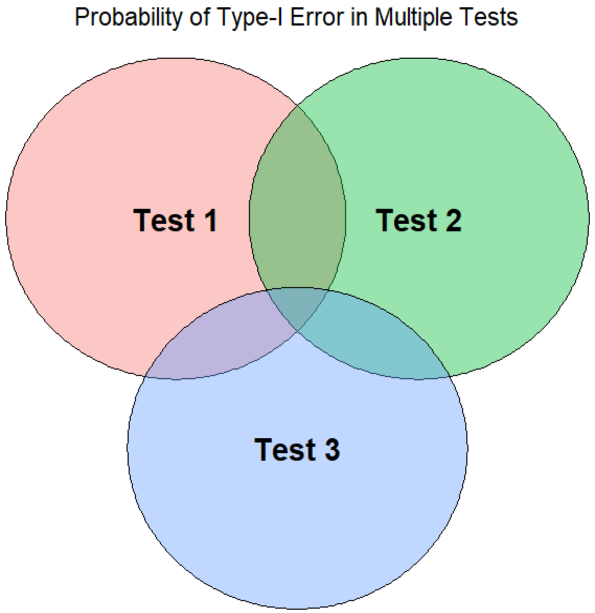
Each test has a 5% (α) chance of making a Type I error.

- Probability of no type-I Error in 1 test: $1-\alpha = 1-0.05 = 0.95$
  So, there's a **95% chance** that we do not make a Type-I error in a single test.
- If we perform 2 **independent** tests, and each test has a 95% chance of being correct, then the probability that **2 tests** with no type-I error is: $0.95 \times 0.95 \; = \; 0.9025$
- If we perform 3 **independent** tests, and each test has a 95% chance of being correct, then the probability that **3 tests** with no type-I error is: $0.95 \times 0.95 \times \cdots \times 0.95 \; = \; 0.95^3 = 0.8573$

Post-hoc Analysis

- So, probability of making at least 1 type I error if we perform 1 independent test $= 1 - 0.95 = 0.05$
- Probability of making at least 1 type I error if we perform 2 independent tests $= 1 - 0.95^2 = 0.097$
- Probability of making at least 1 type I error if we perform 3 independent tests $= 1 - 0.95^3 = 0.143$
- So, probability of **making at least one** Type I error if we perform n independent tests: $1 - (1 - \alpha)^n$





Post-hoc Analysis

- Each additional test increases the risk of a false positive.
- With just 5 groups, requiring 10 tests, the risk of making at least one Type I error jumps from 5% to 40%.
- Instead of running multiple t-tests, ANOVA performs a single test that evaluates all group means simultaneously while controlling for overall Type I error.
- If ANOVA finds a significant result, we can use post-hoc tests (e.g., Bonferroni, Tukey) that adjust for multiple comparisons to determine which means differ.

Post-hoc Analysis

- Common Post Hoc Tests

  - Tukey's HSD:
    Typically best when group sizes are equal or nearly equal, and homogeneity of variance holds

  - Bonferroni Correction:
    Divides the desired significance level by the number of comparisons to keep the overall Type I error rate in check.