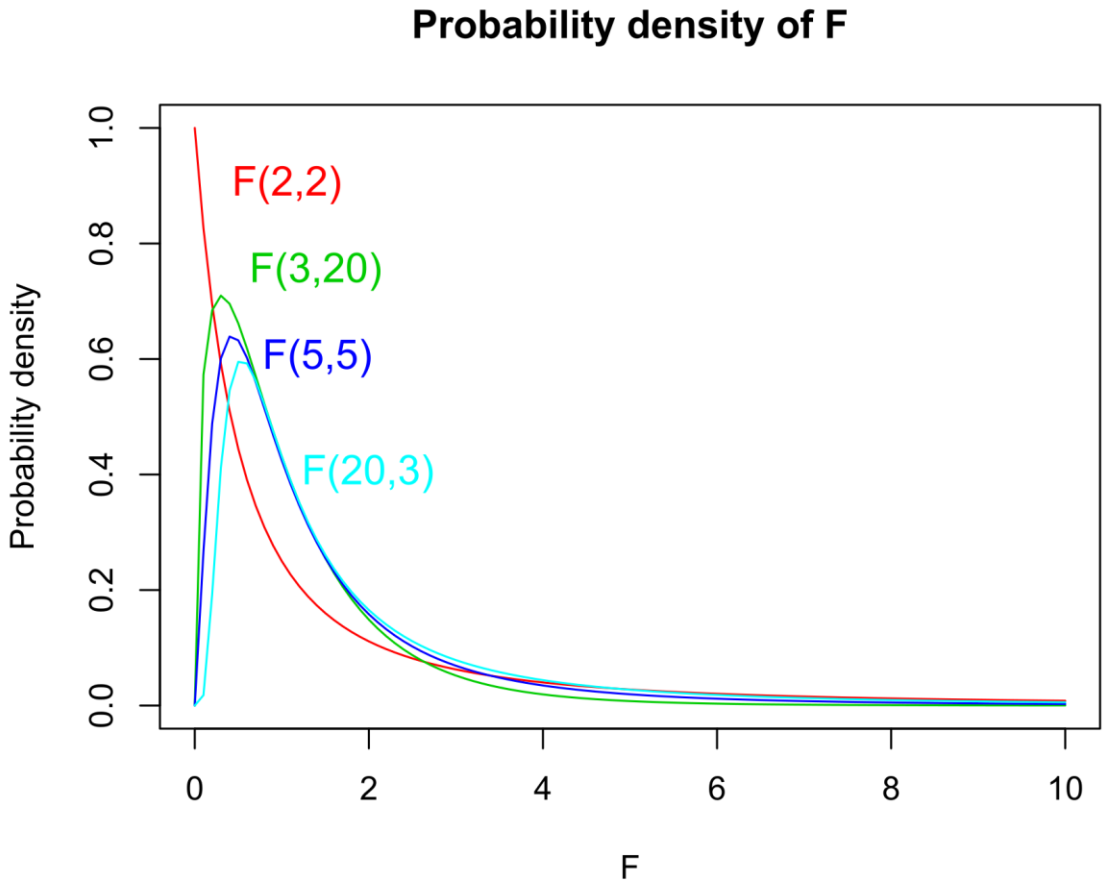
# Chapter 8
# *F-test*
# *ANOVA*

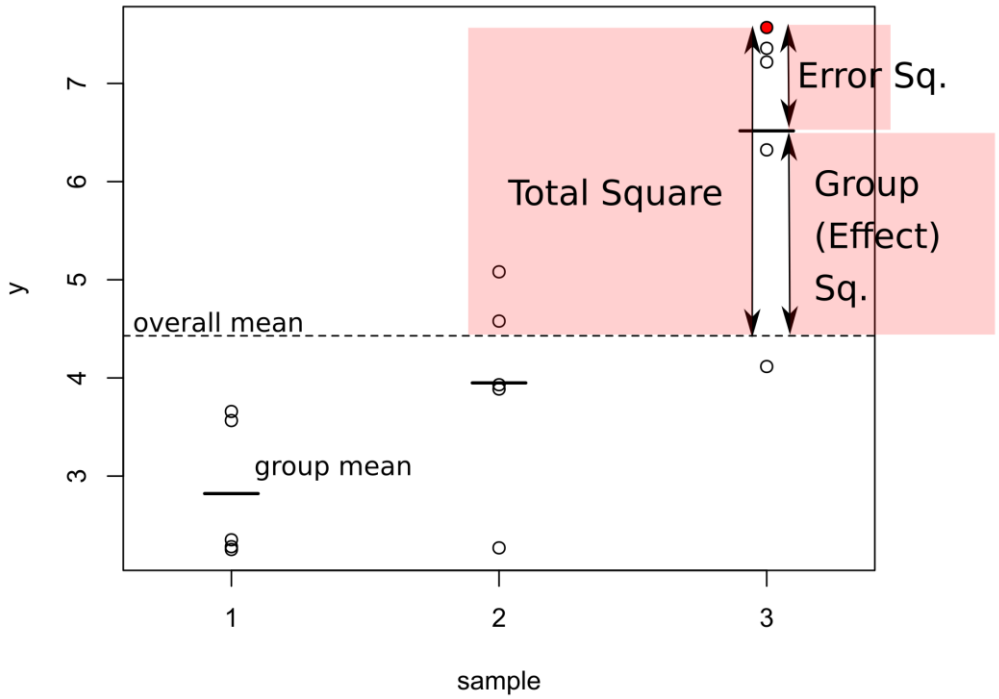
# F-ratio

- Ratio of two variances
- $F = \frac{s_1^2}{s_2^2}$
- Can be used to test H0 that variances of two variables are equal
- Two DFs
  - DF(numerator) = n (sample 1) – 1
  - DF(denominator) = n(sample 2) – 1



# Analysis of variance (ANOVA)

- Application of the F-test principle
- Allows comparisons of multiple mean values
- H0: means of all groups are identical
- Decomposes the total Sum of Squares (= numerator in the variance formula) into
  - Systematic component (Sum of squares effect) – can be scaled by corresponding $DF_{effect}$ (=number of groups – 1) to get Mean Square ($MS_{effect}$)
  - Residual variability (Error Sum of Sq.) – can be scaled by $DF_{error}$ (= number of obs. – 1 – $DF_{effect}$ ) to obtain the $MS_{error}$
- $F = MS_{eff}/MS_{error}$ to be compared with the F distribution with corresponding DFs
- $R^2$ (proportion of explained variability) = $SS_{effect}/ SS_{total}$



# ANOVA assumptions

- Homogeneity of variances (variances in all groups are equal)
- Normal distribution of residuals (= of values within individual groups)
- ANOVA may be unbalanced (= unequal sample size within groups)
- Formal tests exist to check the assumptions
  - Difficult to interpret
  - Graphical inspection of residuals using plot(anova.object) is a better option

# Post-hoc comparisons

- Rejecting H0 in ANOVA means that all group means are not identical, i.e. at least one is different
- We wonder, which mean is significantly different from which
- Post-hoc pairwise comparisons are used for that
  - Tukey HSD test and others
  - Various control levels of type-I error in multiple comparisons
  - Results are best displayed on graphs (groups with different letters are significantly different)