LOCAL SEARCH METHODS
APPLICATIONS AND ENGINEERING

Lecture 13

Empirical Methods
Applications

Marco Chiarandini

Outline

1. Analysis of Variance
2. Multi-comparisons
3. Examples of analysis
4. Sequential Testing: the Racing Algorithm
5. Other Methods

Outline

1. Analysis of Variance
2. Multi-comparisons
3. Examples of analysis
4. Sequential Testing: the Racing Algorithm
5. Other Methods

Two Factors Analysis of Variance

Model:

$$X_{hi} = \mu + \alpha_i + \theta_h + \epsilon_{hi}$$

Statistical Tests

- ANOVA through F-ratio and Fisher test (`aov`)
- Friedman test (`friedman.test`)

Two Factors Repeated Measures Analysis of Variance

Model:

$$X_{hit} = \mu + \alpha_i + \theta_h + \epsilon_{hit}$$

or alternatively:

$$X_{hit} = \mu + \alpha_i + \theta_h + \alpha\theta_{hi} + \epsilon_{hit}$$

Statistical Tests

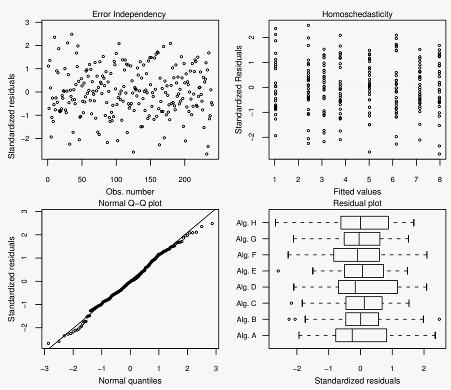
- ANOVA through F-ratio and Fisher test (`aov`)
- Friedman test (not yet available in R!)

If the interactions between factors are of interest interaction plots are used to visualize them

ANOVA Assumptions

- Parametric analysis: ANOVA through F-ratio and Fisher test
  - independent
  - normally distributed
  - homoschedastic
- Nonparametric analysis: through Rank based tests
  - independent
  - homoschedastic

Check Assumptions



Outline

1. Analysis of Variance
2. Multi-comparisons
3. Examples of analysis
4. Sequential Testing: the Racing Algorithm
5. Other Methods

Multi-comparisons – Post-hoc Analysis

Post-hoc analysis: Once the effect of factors has been recognized a finer grained analysis is performed to distinguish where important differences are.

$$H_{0(ij)}: \mu_i = \mu_j \qquad H_{1(ij)}: \mu_i \neq \mu_j$$

$$c = \binom{n}{2} \quad \text{comparisons}$$

Repeating two-samples tests for all pairs the error of Type I is not α but higher:

$$\alpha_{EX} = 1 - \alpha^c$$

Eg, for $\alpha = 0.05$ and $c = 3$ $\alpha_{EX} = 0.14$!

Statistical Tests for Multi-comparisons

- Parametric analysis: Pairwise t-test
- Nonparametric analysis:
  - One Factor AOV: Pairwise Kruskall Wallis or Mann Whitney
  - Two Factors AOV: Pairwise Friedman Test or Wilcoxon Test
  - Two Factors Repeated Measures AOV: Pairwise Friedman Test

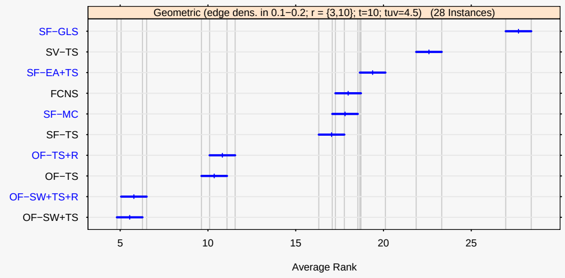
Adjustment method

- Tukey Honest Significance Method (for parametric analysis)
- Bonferroni $\alpha = \alpha_{EX}/c$ (conservative)
- Holm (step-wise)

Outline

1. Analysis of Variance
2. Multi-comparisons
3. Examples of analysis
4. Sequential Testing: the Racing Algorithm
5. Other Methods

An Example with Confidence Intervals



Outline

1. Analysis of Variance
2. Multi-comparisons
3. Examples of analysis
4. Sequential Testing: the Racing Algorithm
5. Other Methods

Unreplicated Designs

procedure *Race*
  repeat only one candidate left or no more unseen instances
    Randomly select an unseen instance and test all candidates on it
    Perform *all-pairwise comparison* statistical tests
    Drop all candidates that are significantly inferior to the best algorithm
  end
end *Race*

Replicated Designs

procedure *Race*
  repeat only one candidate left or maximal number of runs exceeded
    Run all candidates once on all instances
    Perform *all-pairwise comparison* statistical tests
    Drop all candidates that are significantly inferior to the best algorithm
  end
end *Race*

Outline

1. Analysis of Variance
2. Multi-comparisons
3. Examples of analysis
4. Sequential Testing: the Racing Algorithm
5. Other Methods

Distribution Fitting

Parameters of theoretical distributions to fit the data through the Maximum Likelihood Method can be found by in R

`fitdistr(x, ‘‘exp’’)`

Kolmogorov-Smirnov Tests and Goodness of Fit

The test compares empirical cumulative distribution functions.

It computes the maximal difference between the two curves and assess how likely is this value by permutation methods or approximation to the χ^2 distribution.

The test can be done in R with `ks.test`.

The test can be used as a two-samples or single-sample test.
In single-sample the second distribution is a specified continuous distribution, eg, `‘‘pweibull’’`. The parameters of the distribution must be pre-specified and not estimated from the data.

Multiple Regression, Non Linear Regression and Smoothing

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_p x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i$$

$$(y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i)$$

In R: `lm`

R^2 tells the fraction of the variance explained by the model

---

$$y = f(x, \beta)\epsilon$$

In R: `nls`

Smooth curves in scatter plots (no idea on the functional form of the curve)

In R: `poly`, `smooth.spline`, `loess.smooth`, `supsmu`, `ksmooth`

Then add to a scatter plot with `lines`