Lecture II: Comparing Means — The Sequel

Note Title 8/30/2011

Today:
- Quiz
- Homework
- General Comments (publication)
- Analysis of Variance
  - Tests of Assumptns
  - What to do if Failure
    - 3 things

Solution: The Analysis of Variance test

also known as the ANOVA test

It compares 2+ populations

Its test statistic is F

$$F = \frac{\text{variance between groups}}{\text{variance within groups}}$$

Idea: If the groupings are meaningful, then the variance within the group becomes smaller. Thus, F increases.

We (as always) reject for large F.

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

i.e.: grouping does not matter

Assumptions of ANOVA:

Most important →

1. Measurement Normally distributed within the groups
2. Measurements have equal variance across groups
3. Measurements are independent of each other

Testing Assumptions

① Normality

ⓐ Graphical

boxplot — boxplot — symmetric & No outliers

hist — histogram — Normal-looking

qqnorm — QQ plot — §3.2.1

ⓑ Numerical

shapiro.test — Shapiro-Wilk test

ks.test — Kolgomorov-Smirnov test

Beware of Multiple Testing

- A conservative solution (Bonferroni)

Multiply your p by the number of tests you are performing.

Tests of equality of variances

1. Graphical

boxplot – boxplots ← usually good enough

hist – histograms ← tough to use

2. Numerical

var.test – F-test ← only 2 groups

bartlett.test – Bartlett test ← good enough

levene.test – Levene test ← SPSS

fligner.test – Fligner-Killeen test ← usually better

Now, what if the data/model fails a test?

Options:

1) Transform dependent variable — ch 5
2) Monte Carlo techniques — later
3) Use non-parametric tests — now

Non-Parametric tests

disadvantage: less powerful than parametric tests

advantage: more robust than parametric tests

Three non-parametric tests:

wilcox ① Wilcoxon — one group

wilcox ② Mann-Whitney — two groups

③ Kruskal-Wallis — three or more groups

kruskal.test