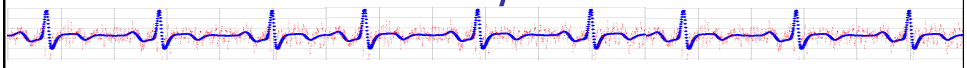
Empirical Research Methods in Information Science

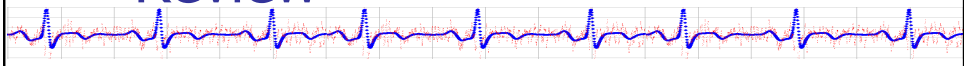
IS 4800 / CS 6350



Lecture 19

One factor, Multi-level Between-Subjects Designs

Review



- Within-subjects design
  - What is it?
  - How do you do one?
  - Criteria for choosing over between-subjects design?
  - Analysis?
  - Power analysis?

Types of Study Designs

- Qualitative
  - Ethnography
- Quantitative
  - Descriptive
  - Correlational
  - Demonstrative
  - Experimental
    - Between-subjects
      - Single factor, two-level
    - Within-subjects
      - Single factor, two-level

3

1-factor, N-level, between-subjects (N>2) Experimental Design

- Trivial generalization of two-level between-subjects design
- Randomize uniformly across the treatment levels
  - Random number generator
  - Blocked randomization still works
  - Baseline analysis generalizes to N
- Everything else is the same as 2 level

4

Accompanying Statistics

- Experimental
  - Between-subjects
    - Single factor, N-level (for N>2)
      - One-way Analysis of Variance (ANOVA)
    - Two factor, two-level (or more!)
      - Factorial Analysis of Variance
      - AKA N-way Analysis of Variance (for N IVs)
      - AKA N-factor ANOVA
  - Within-subjects (for N>2 treatments)
    - Repeated-measures ANOVA (not discussed)
      - AKA within-subjects ANOVA

5

Basic Logic of ANOVA

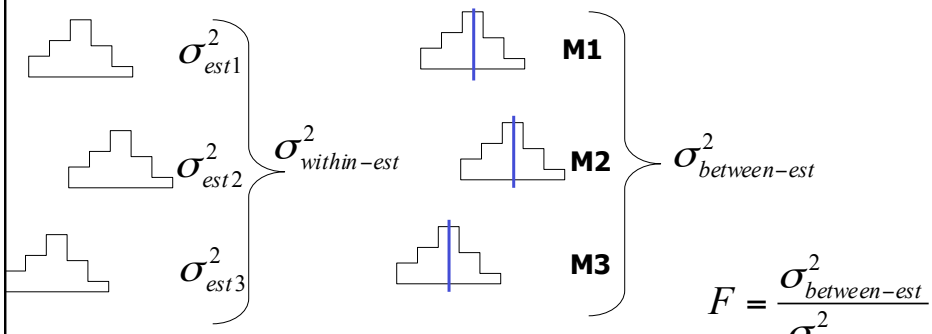
- Null hypothesis
  - Means of all groups are equal.
  - H0: $\mu_1 = \mu_2 = \mu_3 \ldots = \mu_n$
- Test: do the means differ more than expected given the null hypothesis?
- Terminology
  - Group = Condition = Cell = treatment

6

One-Way ANOVA – Assuming Null Hypothesis is True…

Within-Group Estimate Of Population Variance

Between-Group Estimate Of Population Variance



$$F = \frac{\sigma^2_{between-est}}{\sigma^2_{within-est}}$$

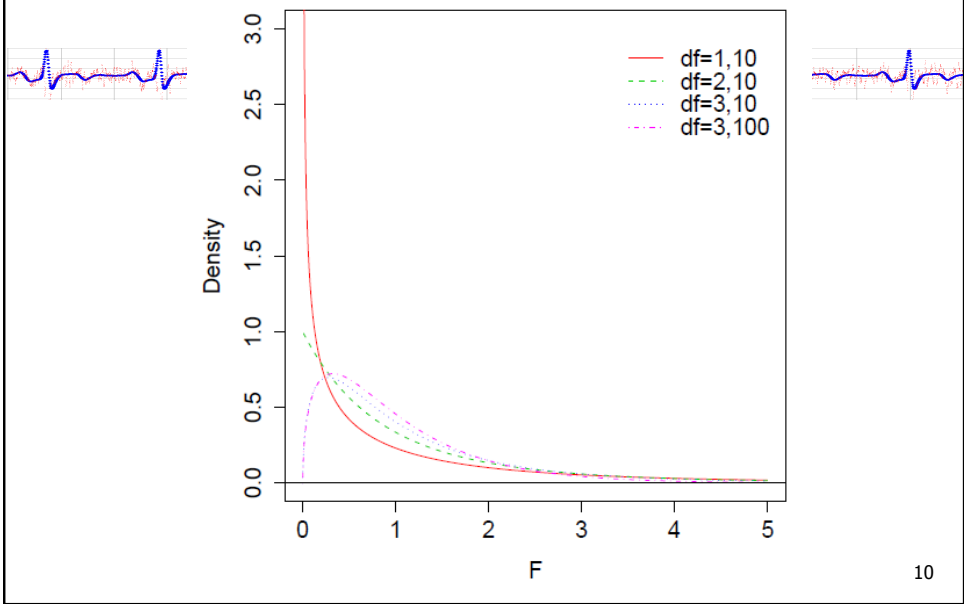
8

Degrees of freedom

- F(between-df,within-df)
- beween-df = num groups - 1
- within-df = sum df for ea group
- Each group df = N_{group}-1
  - So, within-df = total N – num groups

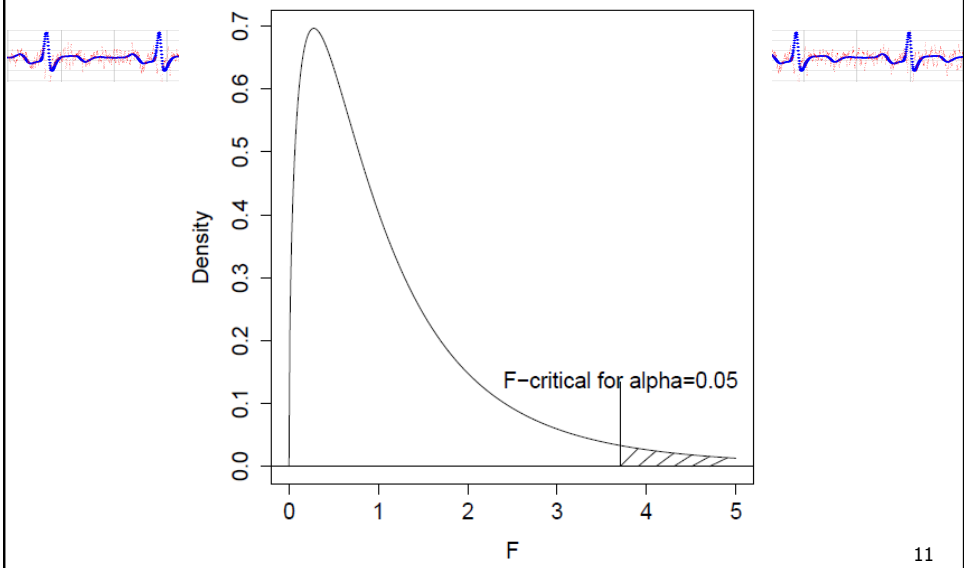
9

Sample F Distributions



10

Sample critical value for F(3,10)



11

ANOVA Assumptions

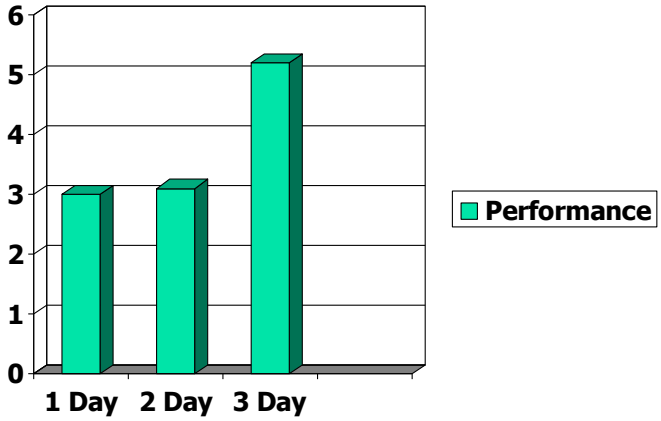
- Interval or ratio data
- Distributions for each group are normal
  - Else: Kruskall-Wallis
- Distributions for each group have equal variances
  - Check with Levene's test
  - Else Welch's F
- The errors (deviations of individual outcomes from the population group means) are assumed to be independent.

12

One-way ANOVA in R

SID	TrainingDays	Performance
1	1	4.0
2	2	3.0
3	3	6.0
4	1	3.5
5	2	4.5
6	3	6.5
7	1	2.5

Data



14

One-way ANOVA in R

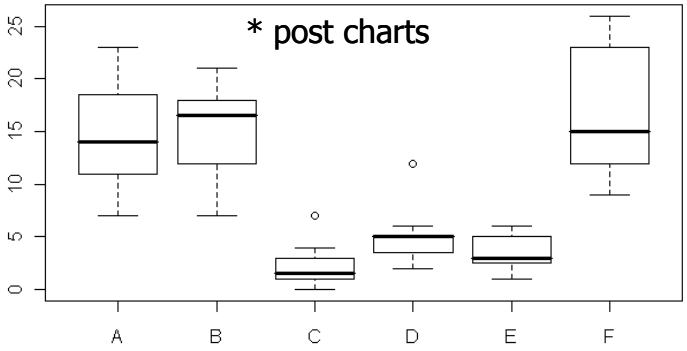
```
> one$TrainingDays <- factor(one$TrainingDays)
> res <- aov(one$Performance ~ one$TrainingDays)
> summary(res)
                 Df Sum Sq Mean Sq F value   Pr(>F)
one$TrainingDays  2 24.812  12.406  9.4417 0.001188 **
Residuals        21 27.594   1.314
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1
```

F(2,21)=9.44, p<.05

15

Visualizing results

- boxplot(DV ~ IV)



17

Post hoc analysis

- Once the ANOVA indicates there is a significant difference ("omnibus" test), you do either
  - Planned comparisons, or
  - Post hoc tests
- to determine which pairwise comparisons are significantly different
- There are many post hoc tests (B&A 452)
  - Most obvious: Least Significant Difference (LSD)
    - Same as t-tests on every pair of treatments
    - Has inflated Type I error due to multiple tests
  - Many others: Sheffe, Dunnett, Tukey, etc.

18

LSD aka unadjusted t-tests

```
> pairwise.t.test(DV, IVfactor,
          p.adjust="none", pool.sd = T)
Pairwise comparisons using t tests with pooled SD data:
DV and IVfactor

       Compact       Other         Pickup
Other  0.50197       -             -
Pickup 0.32786       0.72507       -
Sports 5.9e-05       0.00019       0.00064

P value adjustment method: none
```

Note: p.adjust can also be "holm", "hochberg", "hommel", "bonferroni", "BH", "BY"

19

Post-hoc tests in R Tukey HSD ("Honest Sig Diffs")

```
> res <- aov(one$Performance ~ one$TrainingDays)
> TukeyHSD(res)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = one$Performance ~ one$TrainingDays)

$`one$TrainingDays`
      diff        lwr      upr     p adj
2-1 0.0625 -1.3821563 1.507156 0.9934676
3-1 2.1875  0.7428437 3.632156 0.0027729
3-2 2.1250  0.6803437 3.569656 0.0035777
```

20

Publication format

The overall ANOVA was significant, $F(2,21)=9.44$, $p<.05$, indicating significant differences among the three study treatments.

Tukey HSD post-hoc tests (at .05 significance) indicated significant differences between 3-day training and the other conditions, but not between 1-day and 2-day training.

21

Group Exercise

- For each problem, write
  1. What kind of study design is it?
  2. Research & Null hypotheses in English
  3. Research & Null hypotheses in terms of Pop means
  4. Test criteria
  5. Test results
     - Formal report format
     - English

22

Effect Size in R

- For between-group comparisons
  - Package "compute.es"
  - "Means to Effect Size" – translates group means do effect size estimate (d)

mes(m1,m2,sd1,sd2,n1,n2)

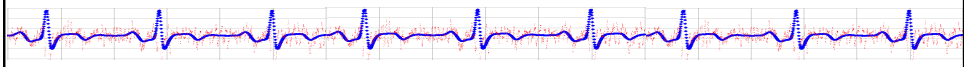
24

Power Analysis

- 'N' computed for your criteria for a between-subjects design is for *each cell* of your experimental design
- A one-factor x four-level design has *four cells*
- B&A: Need at least 5 Ss per cell
- See Aron Table 10-16 (pg 410)
  - Example: medium effect size, 4 group requires 33x4 = 132 Ss!

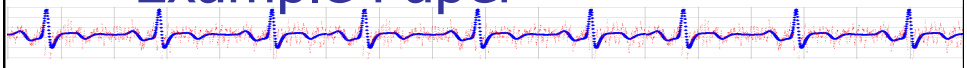
25

Power Analysis in R



- power.anova.test(groups, n, between.var, within.var, sig.level, power)
  - between.var, within.var - variances
  - Leave one of these parameters NULL to compute it from the others.
  - n is *per group!*

Example Paper



- Thank you – I did not see that: In-car, speech-based information systems for older adults.
- Critique?

Design Example

- You want to evaluate which of 2 games leads to greatest engagement. You randomly assign participants to play RockBand or GuitarHero and keep track of how long they play. When they are done you let them play the other game for as long as they want and keep track of the time.
- Kind of study?
- Primary outcome
  - Measure?
  - Statistic?

Design Example

- You want to evaluate which of 2 games leads to greatest satisfaction. You randomly assign participants to play RockBand or GuitarHero, ask them to rate satisfaction on a scale from 1 to 10, then send them home.
- Kind of study?
- Primary outcome
  - Measure?
  - Statistic?

Design Example

- You want to evaluate which of 3 games leads to greatest satisfaction. You randomly give participants BioShock, StarCraftII or TombRaider, ask them to go home and play for a week, then fill out a 12-item composite measure of satisfaction (then they are done).
- Kind of study?
- Primary outcome
  - Measure?
  - Statistic?

Design Example

- You want to evaluate which of 2 games leads to greatest satisfaction. You bring subjects to the lab, ask them to play PacMan or DonkeyKong, then fill out a 12-item composite measure of satisfaction, before sending them home.
- Kind of study?
- Primary outcome
  - Measure?
  - Statistic?

Experiment

32

Homework

- Full-factorial/Multi-factor ANOVA
  - B&A Ch 10, 326-331, B&A Ch 14, 453-457
  - Example paper (Rickenberg & Reeves)
- Next class – 4/3 - T2!

33