Cognitive Foundations

Lecture 2: Experimental Methods (Cont'd)

Foundations of Language Science and Technology

Garance PARIS

12 November 2008

Counterbalancing

- Latin square: An n × n table filled with n different symbols in such a way that each symbol occurs exactly once in each row and exactly once in each column
- Divide your subjects in n groups and your items in n sets (n = number of conditions)
- Each subject sees one set of items in each condition
- Also possible: Randomization

	Cond. 1	Cond. 2
Subject Group 1	Item Set 1	Item Set 2
Subject Group 2	Item Set 2	Item Set 1

$$\begin{bmatrix}1\end{bmatrix} \quad \begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix} \quad \begin{bmatrix}1 & 2 & 3\\ 2 & 3 & 1\\ 3 & 1 & 2\end{bmatrix}$$

$$\begin{bmatrix}1 & 2 & 3 & 4\\ 2 & 1 & 4 & 3\\ 3 & 4 & 1 & 2\\ 4 & 3 & 2 & 1\end{bmatrix} \quad \begin{bmatrix}1 & 2 & 3 & 4\\ 2 & 4 & 1 & 3\\ 3 & 1 & 4 & 2\\ 4 & 3 & 2 & 1\end{bmatrix}$$

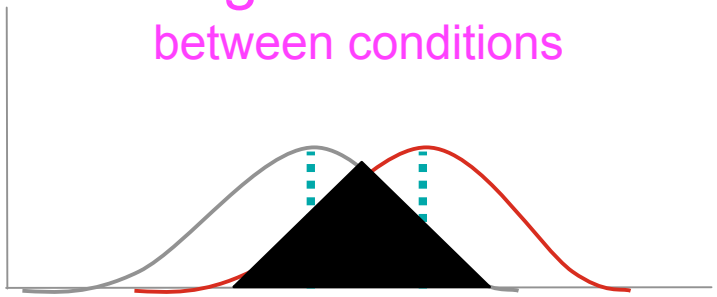
Filler Items

- Subjects should not be able to guess what the experimental manipulation is
- Filler items with similar but different structure are most often used to hide the pattern of the manipulation
- At least twice as many fillers as items is good
- Additional fillers may be necessary to demonstrate the task to the participants or for them to get used to the task

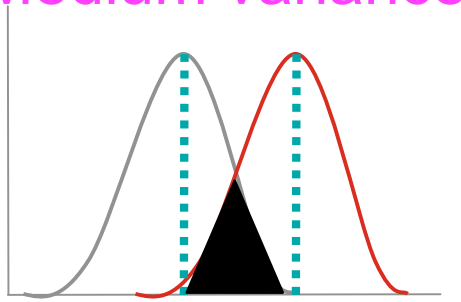
Analyzing the Data

- Most often in psycholinguistics, we try to confirm our predictions by comparing the mean reaction time, error rate, probability of an object being fixated, ... in different conditions
- But looking at the overall mean behavior is insufficient to interpret experimental results
  - For example: Even if the difference between the means is the same, the samples themselves can still differ a lot and the distribution overlap more or less

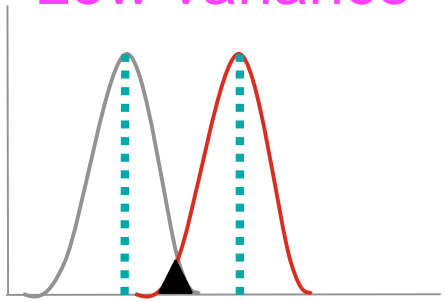
High variance between conditions



Medium variance



Low variance



Inferential Statistics

- How do we know the results we obtained from our sample hold for the population in general?
- We can make sure with inferential statistics
- To compare two means, we can use a “t-test”
  (similar but more powerful: Analysis of variance, “ANOVA”)
- Two numbers are reported that, together, tell us if the test was significant (if the data generalize) or not:
  - t: How much variance the data contains
    - t1: Analysis by subjects (averaging over items)
    - t2: Analysis by items (averaging over subjects)
  - p: The probability that we find a difference between the two samples in condition 1 and condition 2, although there is in reality none (due to sampling from a larger population)

Averaging Results

Raw data

subj.	item	cond.	D.V.
1	1	a	356
1	1	b	641
1	1	c	272
1	2	a	478
1	2	b	821
1	2	c	375
...	...	...	...
2	1	a	303
2	1	b	426
2	1	c	299
...	...	...	...

Analysis by subjects (averaging over items)

subject	condition	dependent variable
1	a	356+478+... / nb-of-items
1	b	641+821+... / nb-of-items
1	c	272+375+... / nb-of-items
2	a	303+... / nb-of-items
...	...	...

Analysis by items (averaging over subjects)

item	condition	dependent variable
1	a	356+303+... / nb-of-subjects
1	b	641+426+... / nb-of-subjects
1	c	272+299+... / nb-of-subjects
2	a	478+... / nb-of-subjects
...	...	...

Computing a Paired T-Test

- Step 1: State the null hypothesis and the alternative
  - H_0: There is no difference between conditions
  - H_1: There is a difference
- Step 2: Compute the difference for each pair, then the mean difference
- Step 4: Compute the variance, then the standard deviation, then the standard error
  - Variance: Sum of squared differences between the mean difference and the difference for each pair
  - Std. dev.: Take square root of the variance divided by n-1
  - Std. err.: Divide that number by the square root of n
- Step 5: Compute the t-value and look up the critical value in a table. If the critical value is less than the t-value you calculated, accept H_1, else reject it.