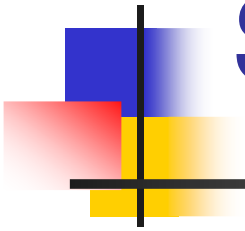
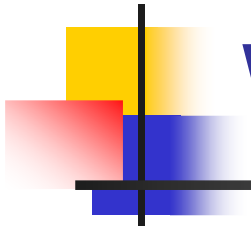


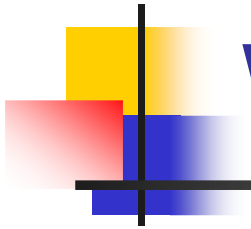
STATISTICAL ANALYSIS

Repeated Measures
and (M)ANOVA designs



Why repeated measures?

- What is the greatest source of variance in a (psycholinguistic response time) experiment?



Why repeated measures?

- What is the greatest source of variance in an experiment?
  - people

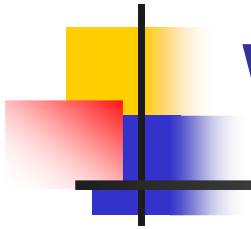
Betweem Subjects Effects
last table of Repeated measures

Tests of Between-Subjects Effects

Measure: MEASURE_1

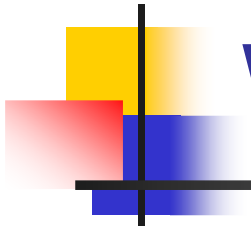
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	85863235,4	1	85863235,42	683,157	,000
Error	4399009,151	35	125685,976		



Why repeated measures?

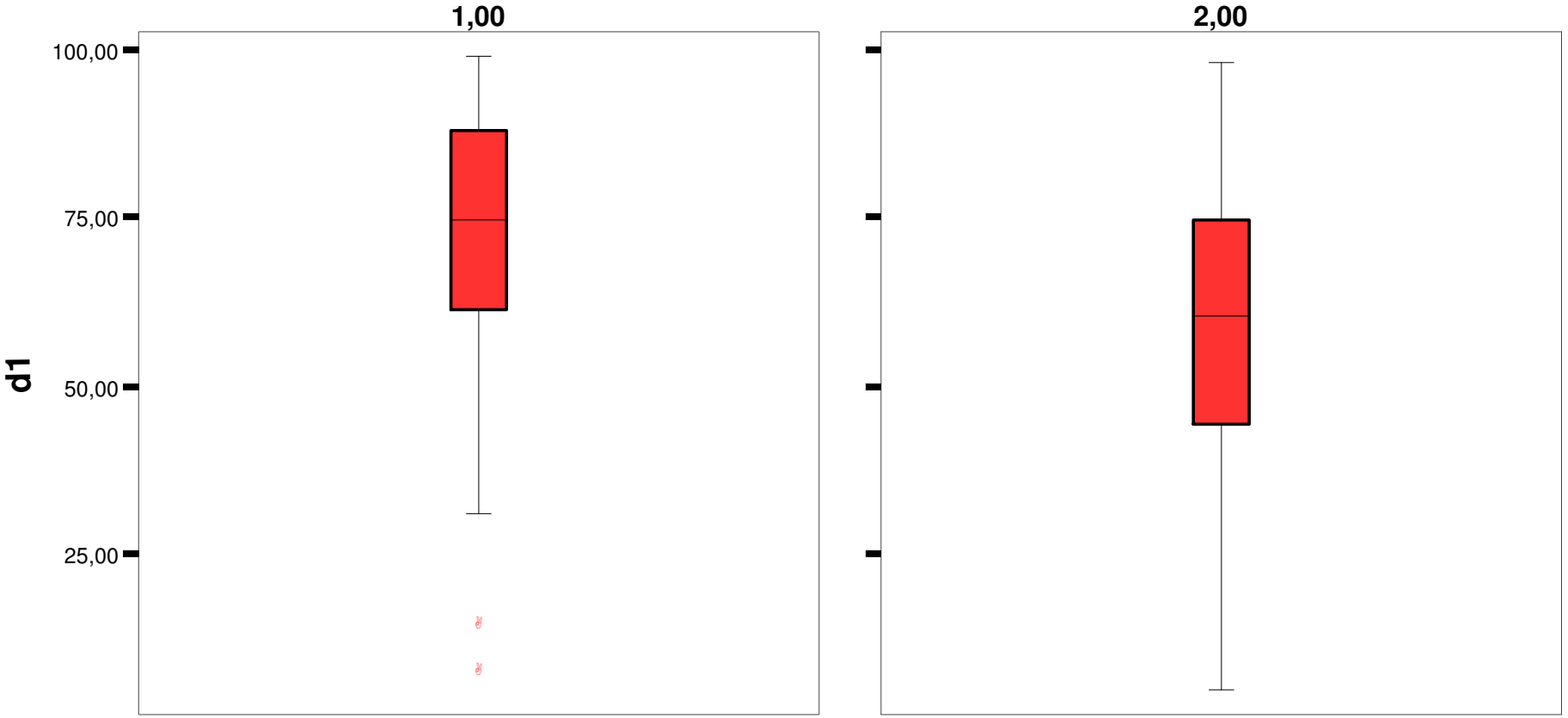
- What is the greatest source of variance in an experiment?
  - People
  - Can we get rid of this noise?
  - Repeated measures designs

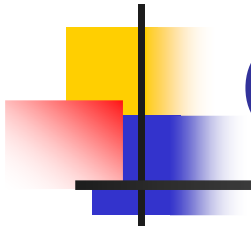


Why Repeated Measures?

- Concrete example

Are these distributions (conditions) really different?

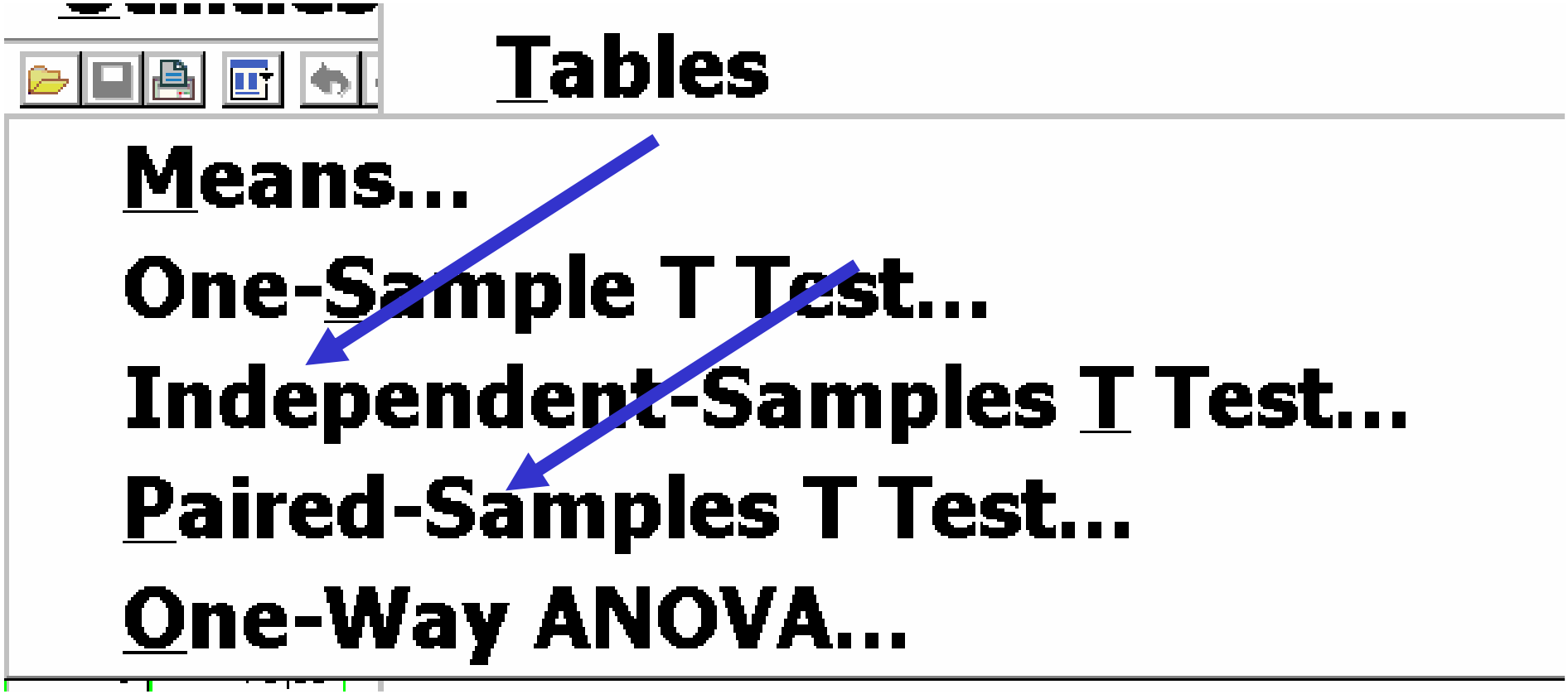




Can we use a t-test?

- Equivalent standard deviations?

Independent Samples vs. Paired Samples



Independent Samples

Independent Samples

		Levene's Test for Equality of Variances				
		F	Sig.	t	df	Sig
d1	Equal variances assumed	,098	,757	,771	30	
	Equal variances not assumed			,771	29,422	

Independent Samples

ɔendent Samples Test

t-test for Equality of Means					
				95% Confidence Interval of the Difference	
df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
30	,447	7,06250	9,165[illegible]9	-11,65614	25,78114
29,422	,447	7,06250	9,1655[illegible]	-11,67157	25,79657

Paired Samples (1)

Paired S

		Paired Differences		
		Mean	Std. Deviation	Std. Error Mean
Pair 1	d1 - d2	7,06250	11,45116	2,86279

Paired Samples (1)

ımples Test

95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
Lower	Upper			
,96061	13,16439	2,467	15	,026

Paired Samples (2)

Paired Sampl

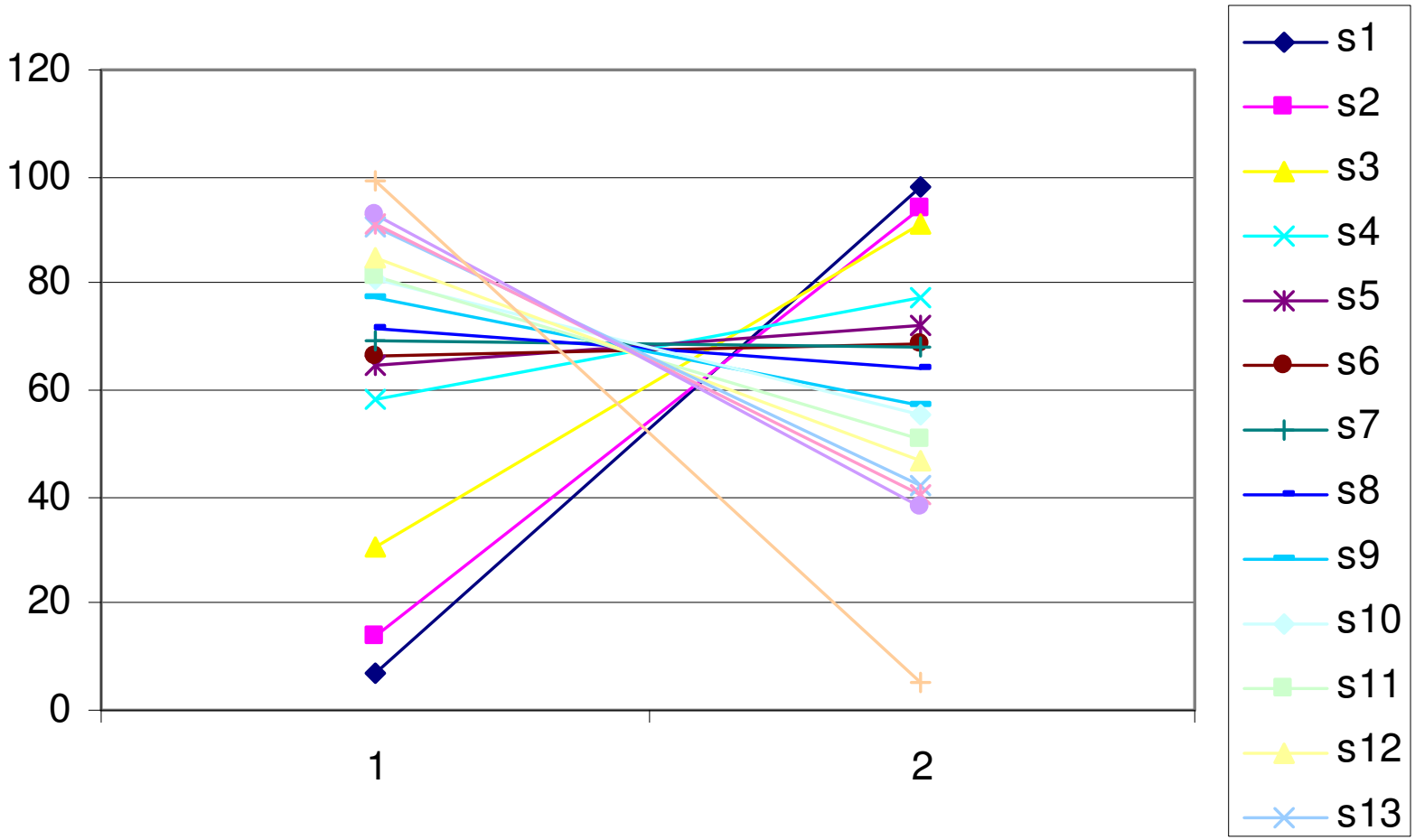
		Paired Differences			
		Mean	Std. Deviation	Std. Error Mean	95 I Low
Pair 1	d1 - d2	7,06250	50,51134	12,62784	-19,85

Paired Samples (2)

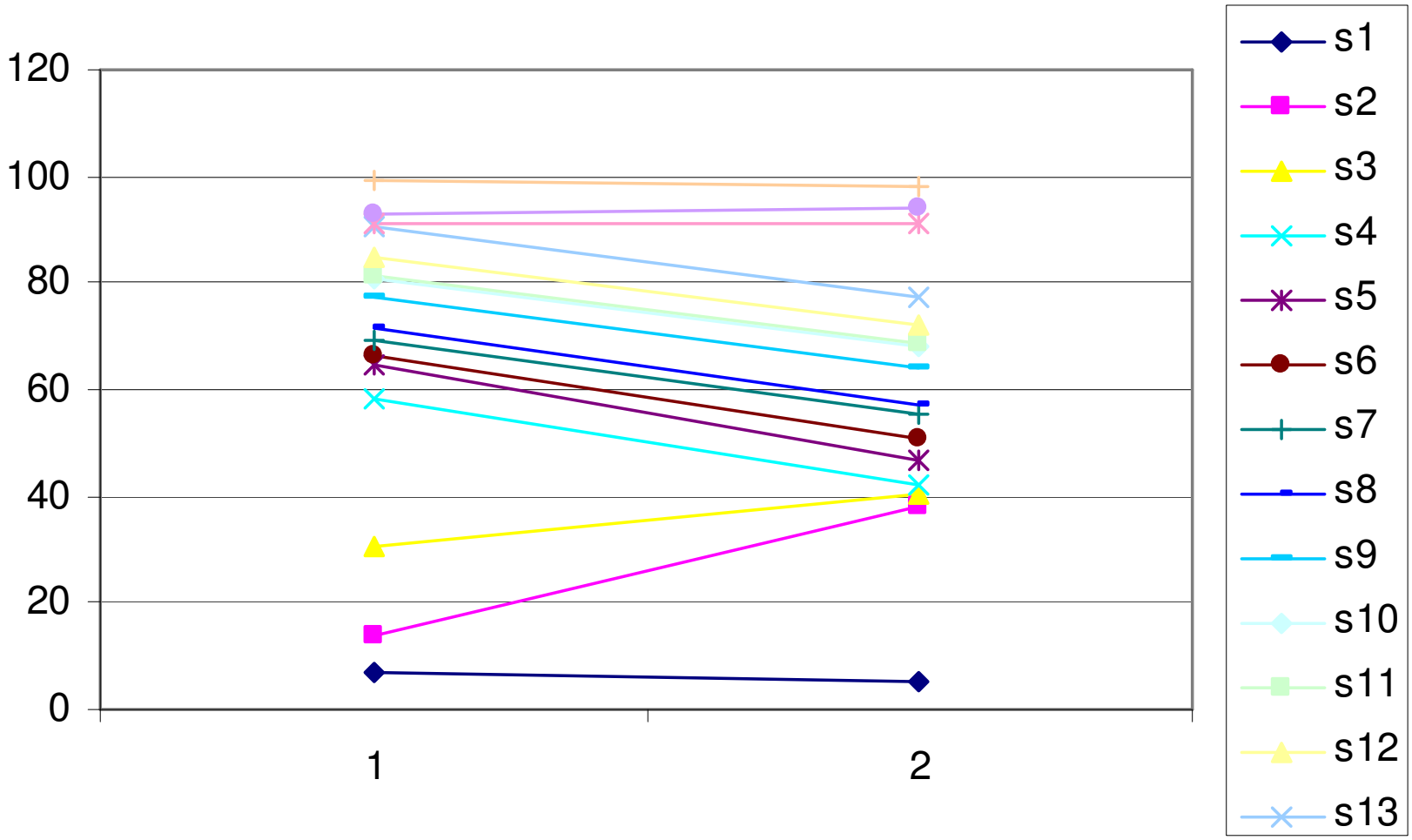
Samples Test

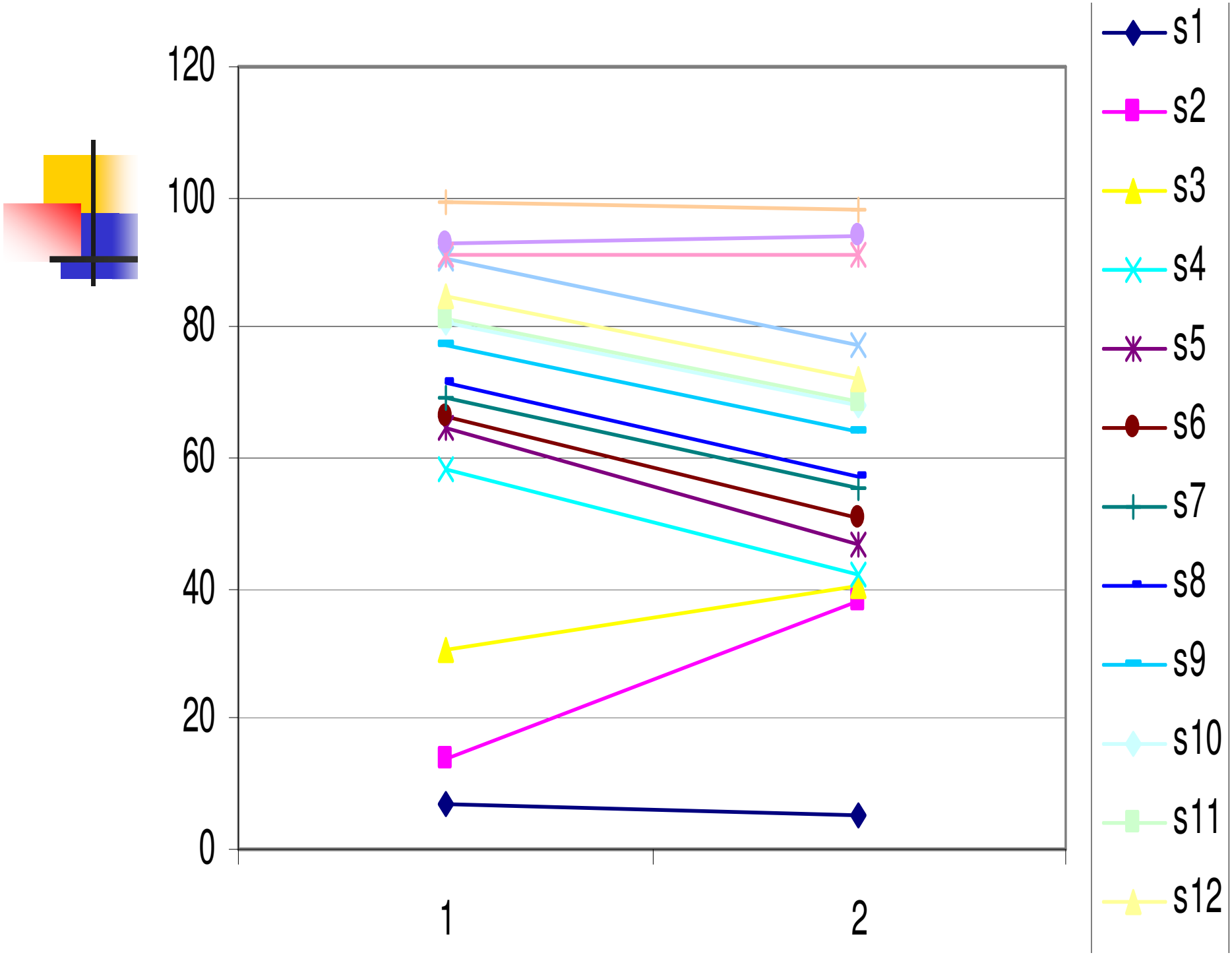
es				
95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
Lower	Upper			
-19,85310	33,97810	,559	15	,584

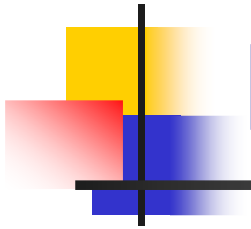
Why? Consistency



Why? Consistency



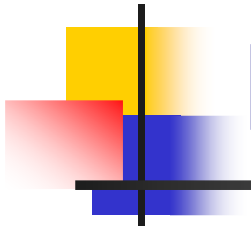

120
100
80
60
40
20
0
1
2
s1
s2
s3
s4
s5
s6
s7
s8
s9
s10
s11
s12



Repeated Measures

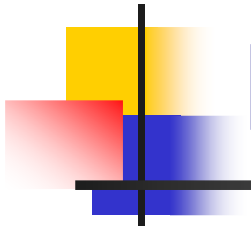
- Example with large between subject variability
- small effect size

- Here within subject (repeated measures) design is usually invaluable



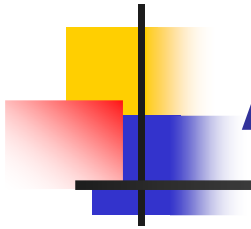
Repeated Measures

- Is a spurious effect possible using between group design when this is not the correct design?
- Yes, if there is sufficient accidental clustering of responses in one condition
  - Due to inflated degrees of freedom
  - Even thought these are not consistent



Repeated Measures

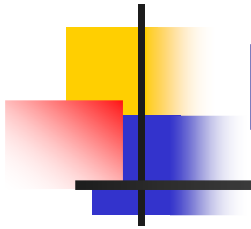
- Unless you are specifically interested in intersubject variability
- Always use repeated measures where possible



ANOVA vs. T-test

- In some experiments we have more than two levels of a factor
- So paired samples don’t work too well

- Why not?

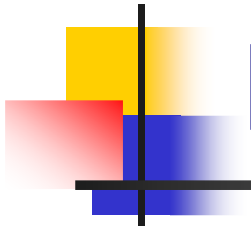


Experiment with three levels

- Sentences containing ambiguous words
- Sentence completion test
  - Context supporting more frequent (dominant) meaning
  - Context supporting less frequent (subordinate) meaning
  - Context equally consistent with each

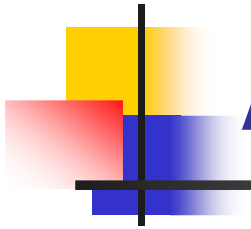
Experiment with three levels

- Het akkoord kon met weinig moeite wirden ___
- Het akkoord kon door de politici worden____
- Het akkoord kon door de pianist worden_____



Experiment with three levels

- Is there an effect of context?
- To increase completions indicating that the subordinate meaning has been selected
- T increase completions suggesting the ***dominant*** meaning has been selected?



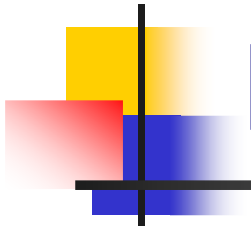
ANOVA vs. T-test

- Carry out two t-tests (dominant and subordinate context relative to neutral)?
- Or an ANOVA looking for differ4ences across all three levels?

Probability

Test

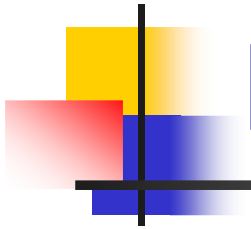
Confidence rval of the fference				
	Upper	t	df	Sig. (2-tailed)
51	13,16439	2,467	15	,026



Probability

- Probability
- The percentage of times you can expect to (extimate that you will) get this large a difference by chance if the distibutions are not different

- P = .024 meanes?

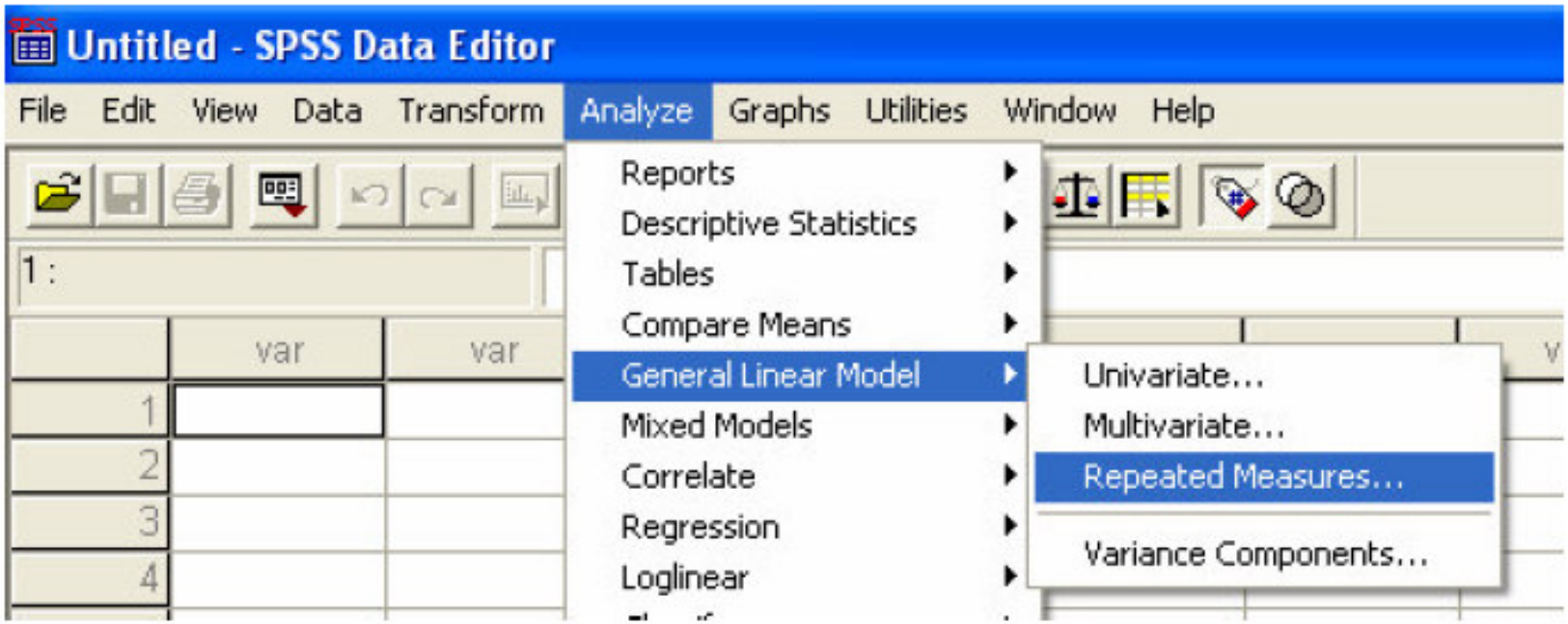


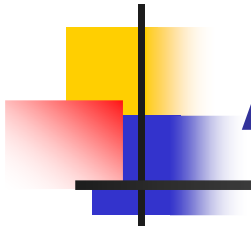
Probability

- Multiple tests raise the chance of Type 1 errors

- I.e., False positive
- Because they overestimate probability

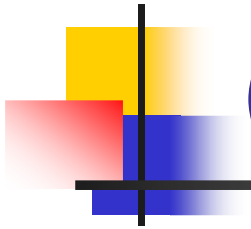
Repeated measures/ SPSS





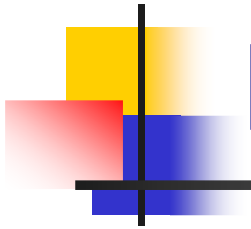
ANOVA (repeated measures)

- SPSS
- Analyze
- General Linear Model
- Repeated measures



(M)ANOVA

- H0: There is no difference between levels of a factor
- H1: There is a difference between at least one level and the others



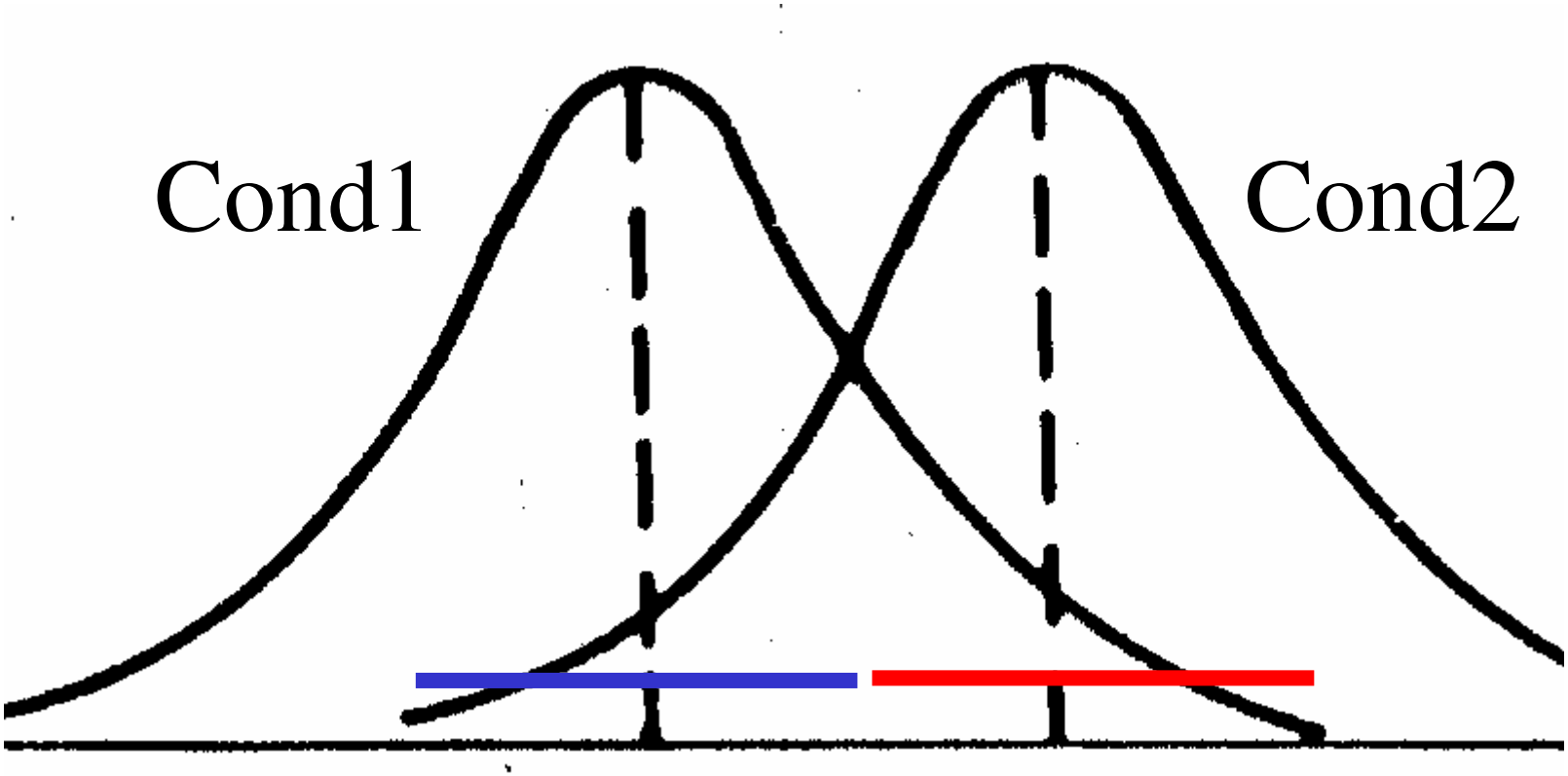
F ratio

- Estimated variance given that all observations come from a single distribution

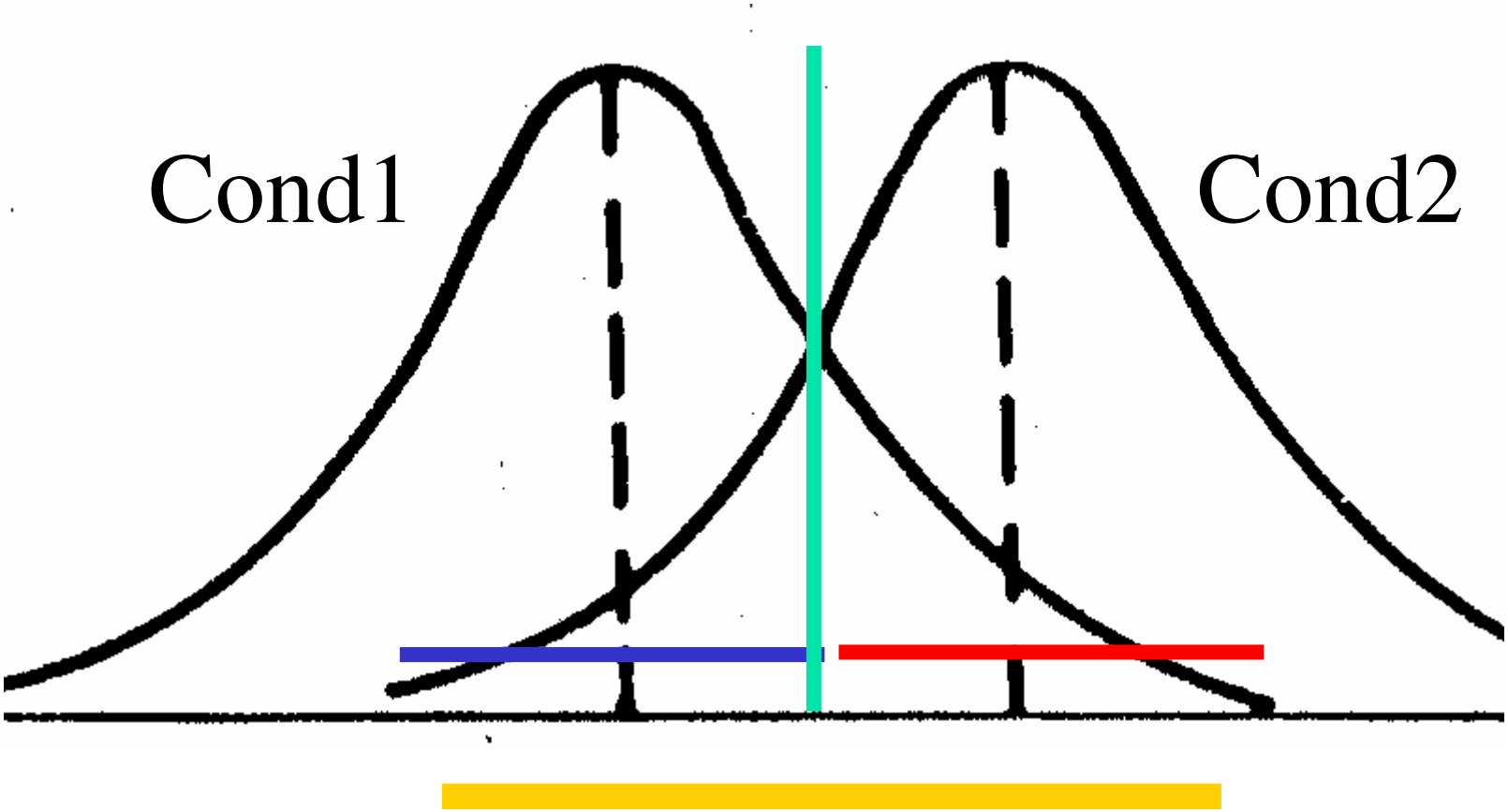
---

- Average extimated variances of each condition separately

Extimating Variance



Extimating Variance

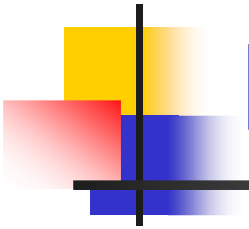


F ratio

- Estimated variance of single distribution

- Average extimated variances of condition

F ratio



- The bigger the F-ration

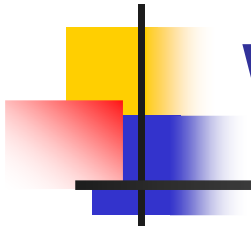
- The less likely that the conditions come from the same distribution

Experiment with three levels

Tests of Within-Subjects Effects

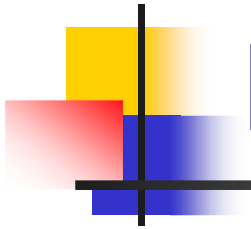
Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
context	Sphericity Assumed	2,961	2	1,480	39,786	,000
	Greenhouse-Geisser	2,961	1,834	1,615	39,786	,000
	Huynh-Feldt	2,961	1,985	1,492	39,786	,000
	Lower-bound	2,961	1,000	2,961	39,786	,000
Error(context)	Sphericity Assumed	1,712	46	,037		
	Greenhouse-Geisser	1,712	42,172	,041		
	Huynh-Feldt	1,712	45,644	,037		
	Lower-bound	1,712	23,000	,074		



Why so many tests?

- Because ANOVA and MANOVA various assumptions
  - Sphericity. For example
- And various corrections are carried out if an assumption is not valid
  - Maybe more in another presentation
  - But practically it usually does not change the significance much

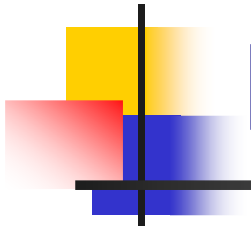


Experiment with three levels

- Okay, now you know that there is an effect (at least one of the conditions is statistically different from the others)
- But that does not answer your question

Experiment with three levels

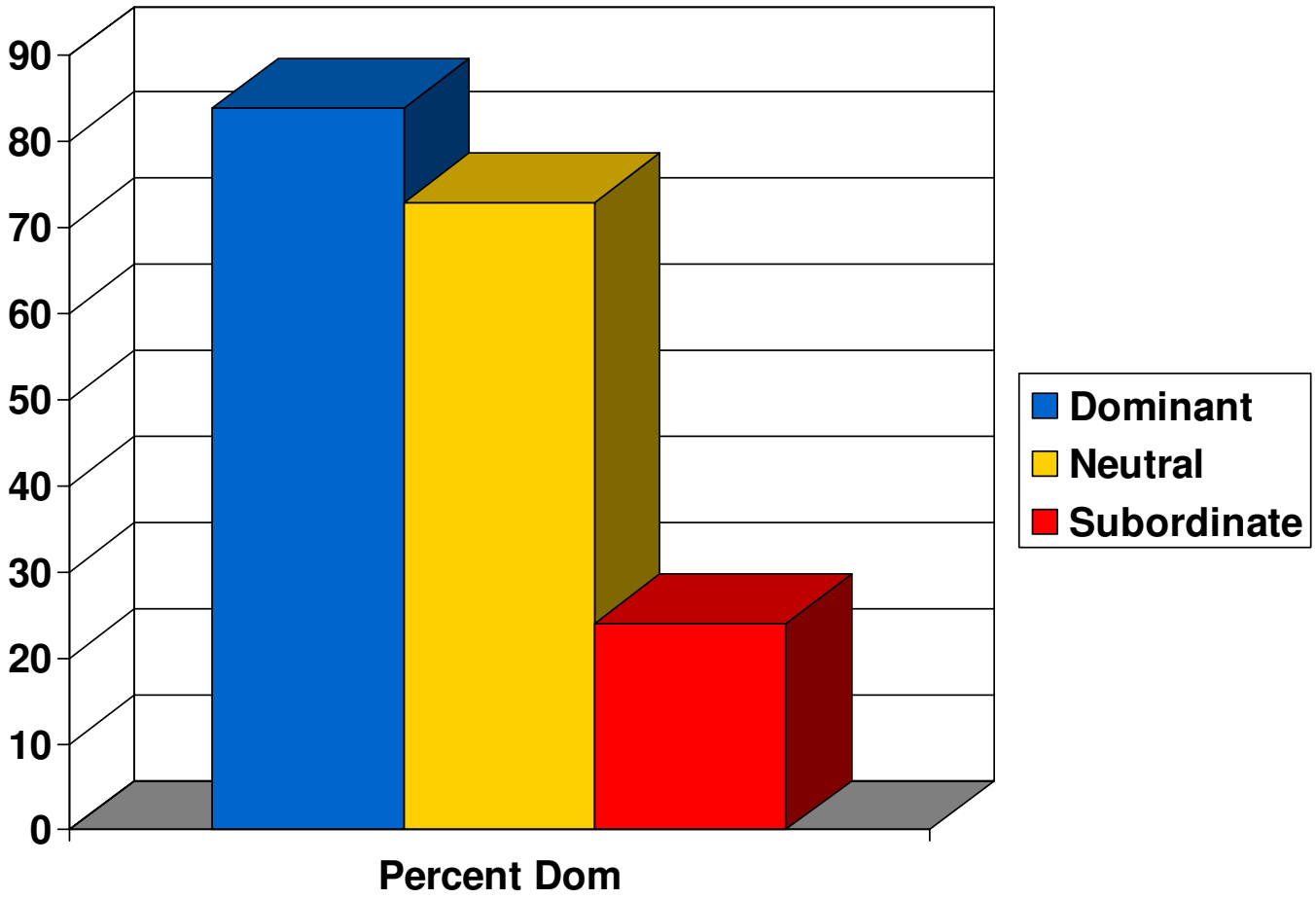
- Is there an effect of context?
- To increase completions indicating that the subordinate meaning has been selected
- To increase completions suggesting the ***dominant*** meaning has been selected?



Experiment with three levels

- The F-test shows that at least one condition is significantly different
- And you are justified in using individual ***post-hoc*** comparisons to test where it comes from

Percentage Dominant Response



Experiment with three levels

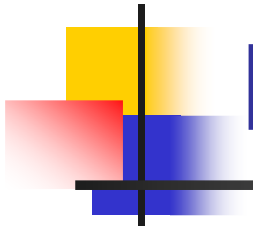
Paired Samples

		Paired Differences		
		Mean	Std. Deviation	Std. Error Mean
Pair 1	DOMCONT - NEUTCONT	,1149	,17421	,03556
Pair 2	SUBCONT - NEUTCONT	-,4908	,24763	,05055

Experiment with three levels

d Samples Test

Differences					
	95% Confidence Interval of the Difference				
Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
,03556	,0413	,1885	3,231	23	,004
,05055	-,5953	-,3862	-9,709	23	,000

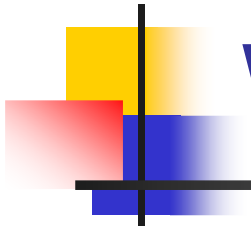


MANOVA

- Like ANOVA
- in cases where more than one factor is being manipulated

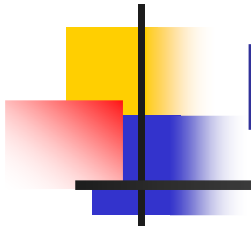
Experiment with ***nine*** levels

- Het/de
- Akkoord / overeenkomst / melodie
- kon met weinig moeite worden___
- kon door de politici worden___
- kon ddor de pianist worden___



Why this addition?

- To ***independently*** examine the effect of the context
- Really neutral, dominant supporting…?
- May steer toward one sort of completion regardless of the ambiguity

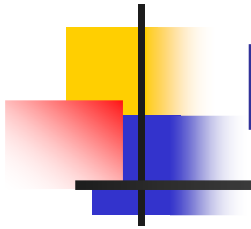


Methodological Sidestep

- How can you best judge if the completions are consistent with the dominant or subordinate meaning?

Methodological Sidestep

- Blind rating
- Non-blind overestimated the likelihood of dominant completions
- Particularly in the dominant condition

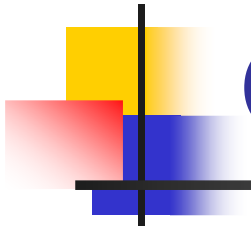


Nine Levels!

- This is the wrong way to look at the data
- Here you have two factors with three levels which combine to give you 9 conditions

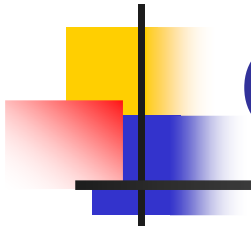
Experiment with 3 x 3 design

- Ambiguity: 3 levels
  - Ambiguous
  - Dominant control
  - Subordinate Control
- Sentence context
  - Neutral
  - Dominant supporting
  - Subordinate supporting



Orthogonality

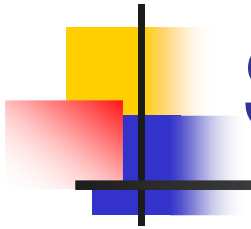
- Carry out tests that are ***independent*** of each other
- And thus do not lead to overestimation of probability



Orthogonality

- The answers to the following are independent, i.e. do not influence each other

- Main effect 1: $a + b =?= c + d$
- Main effect 2: $a + c =?= b + d$
- Interaction: $a - b =?= c - d$



SPSS

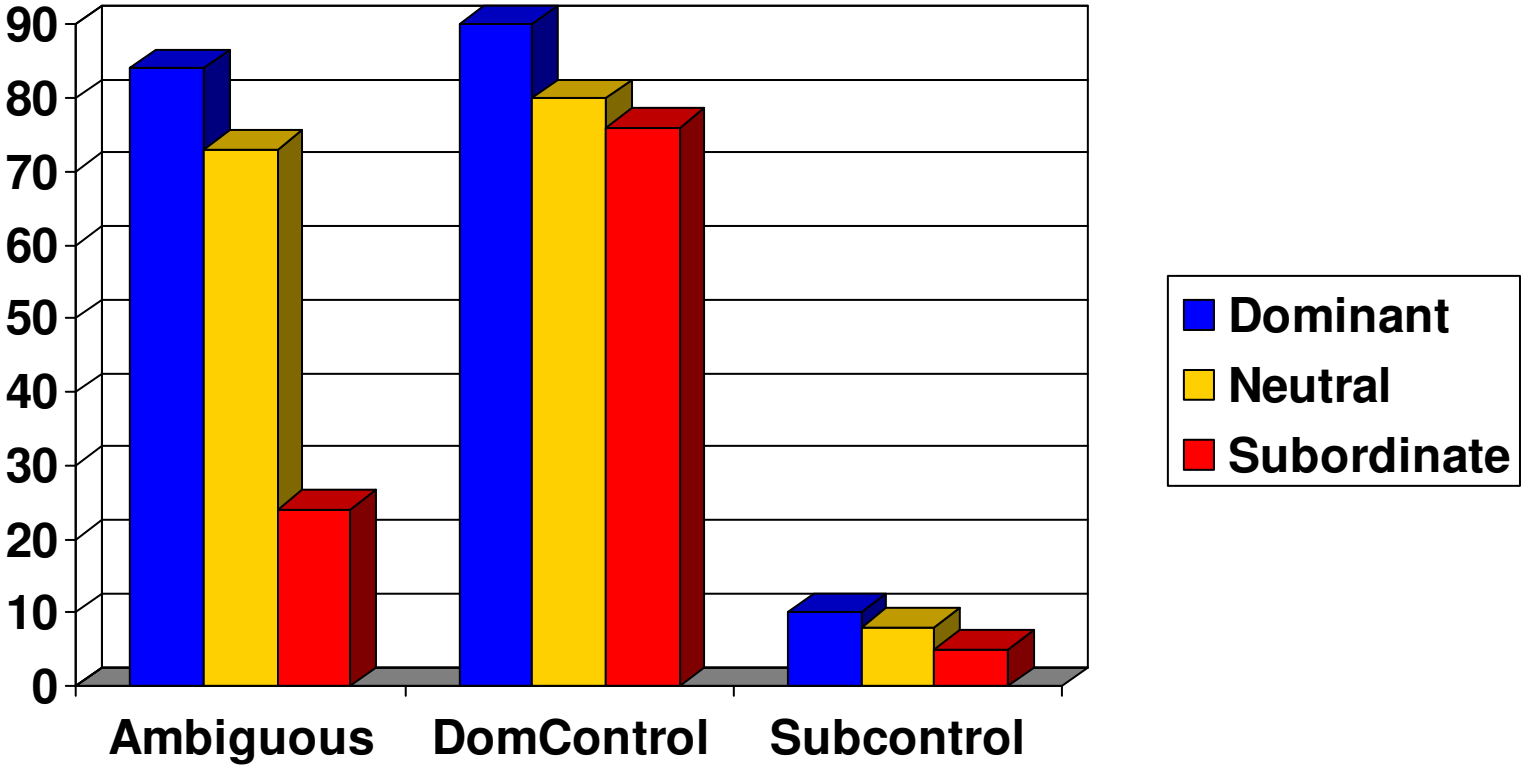
- Repeated measures
- Factor Ambiguity 3 levels add
- Factor Context 3 levels add

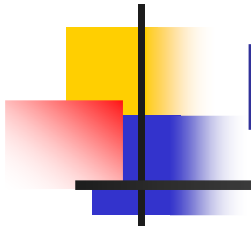
- Fill in design matrix

Matrix

- Fill in matrix with variables
  - Ambigdom (1,1)
  - Ambigneut (1,2)
  - Ambigsub (1,3)
  - Dcontdom (2,1)
  - Dcontneut (2,2)
  - Dcontsub (2,3)
  - Scontdom (3,1)
  - Scontneut (3,2)
  - Scontsub (3,3)

Percentage Dominant Response

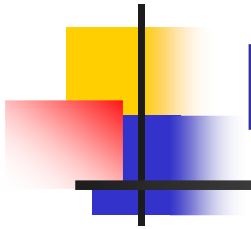




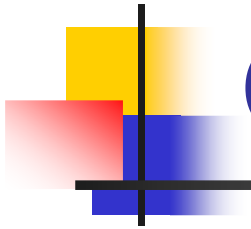
Main Effects and Interactions

- When you have orthogonal factors you can investigate
- main effects of each factor
- Interactions between factors

Statistical Reasons for MANOVA



- Fragmented univariate ANOVAs lead to type 1 errors
  - seeing effects that aren't really there.
- Univariate ANOVAs throw away info - correlation among dependent variables.



Clearer Example of Interaction

- We know that both word frequency and irregularity of spelling contribution to how quickly words are recognized
- But are these effects independent of each other?
  - Repeated measures design with naming task

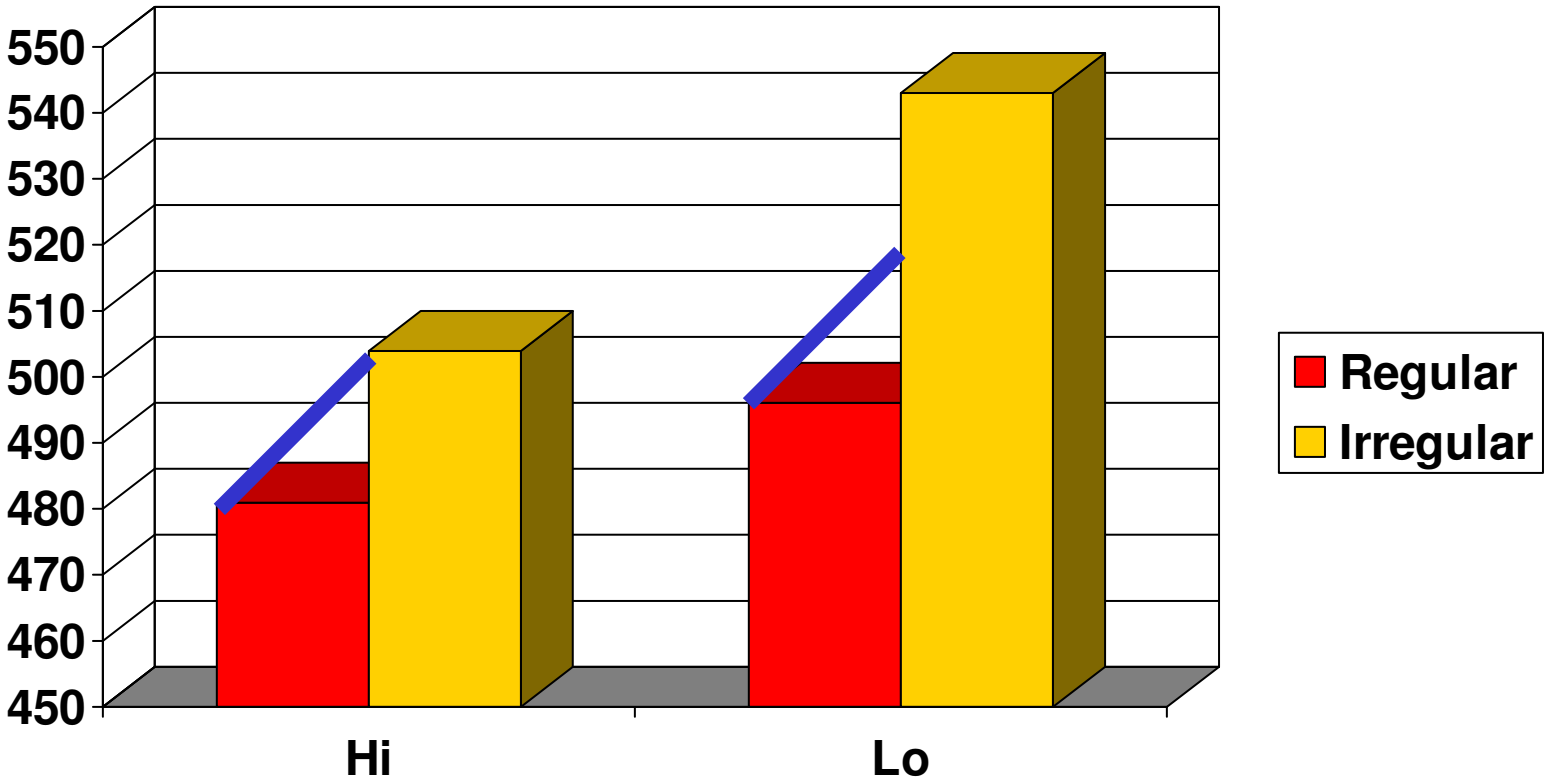
SPSS Output : Chance to check design!

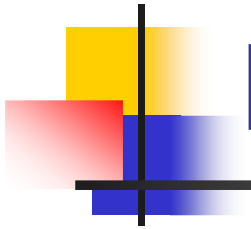
Within-Subjects Factors

Measure: MEASURE_1

reg	freq	Dependent Variable
1	1	reghi
	2	reglo
2	1	irreghi
	2	irreglo

Results


550
540
530
520
510
500
490
480
470
460
450
Hi
Lo
Regular
Irregular



Main Effects and Interaction?

- Low frequency RTs > High Frequency?
- Irregular RTs > Regular ?
- Mayve combination is bigger than either alone?

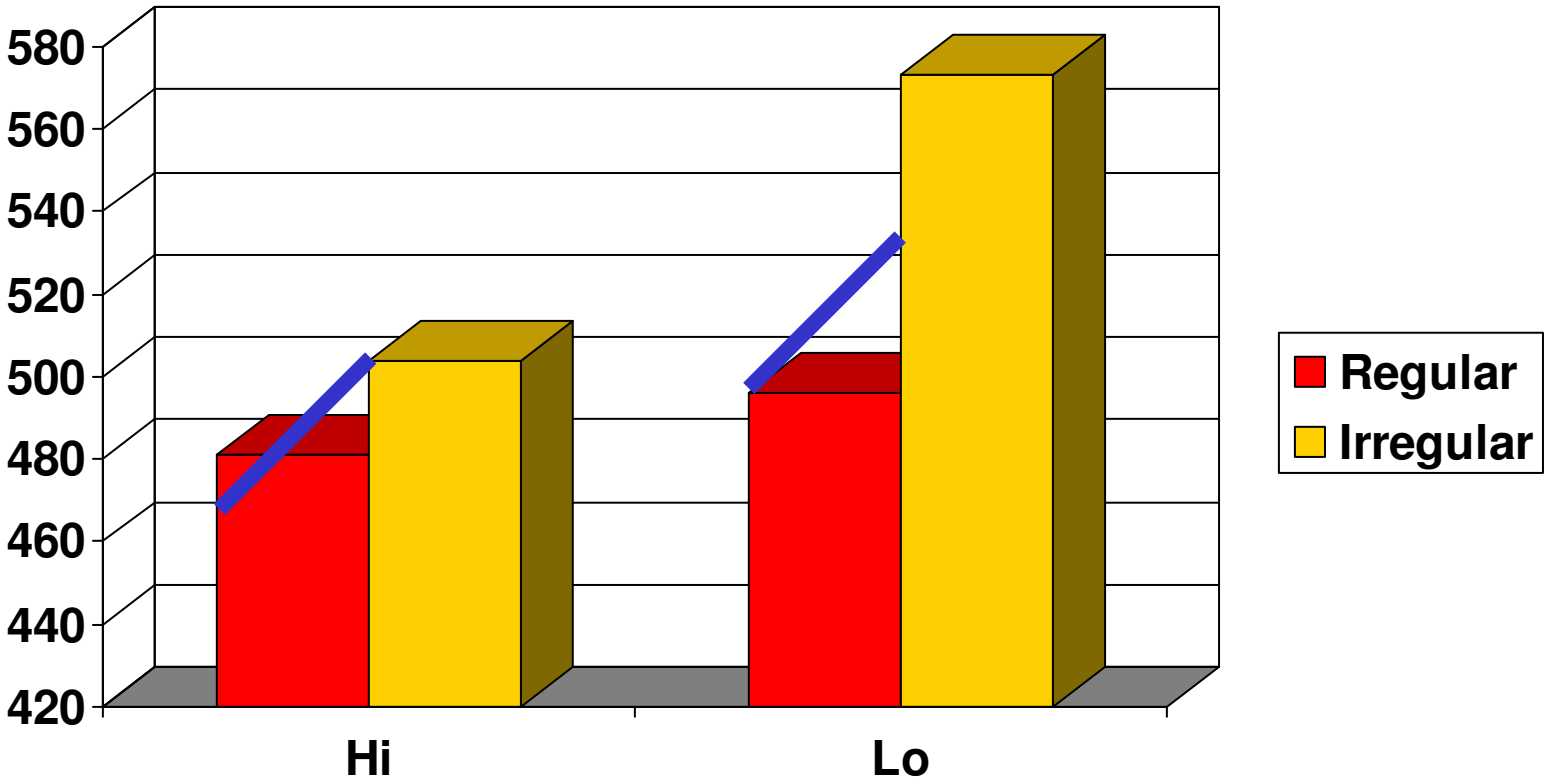
Main Effects and Interaction?

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	reg	freq	Type III Sum of Squares	df	Mean Square	F	Sig.
reg	Linear		19321,000	1	19321,000	34,966	,000
Error(reg)	Linear		8288,500	15	552,567		
freq		Linear	11881,000	1	11881,000	14,079	,002
Error(freq)		Linear	12658,500	15	843,900		
reg * freq	Linear	Linear	2475,063	1	2475,063	2,329	,148
Error(reg*freq)	Linear	Linear	15940,438	15	1062,696		

Results: More like real results



Main Effects and Interaction?

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

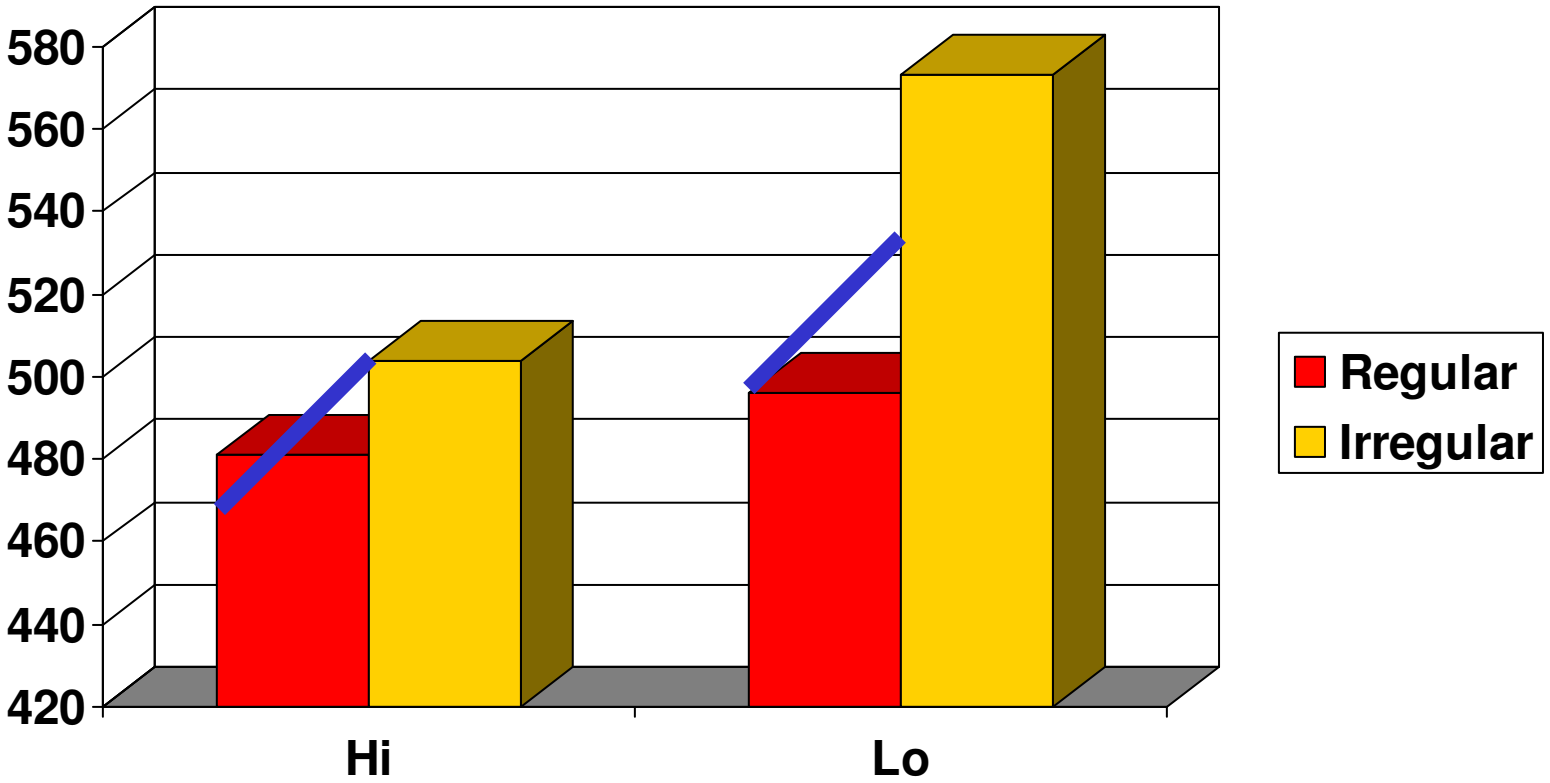
Source	reg	freq	Type III Sum of Squares	df	Mean Square	F	Sig.
reg	Linear		39601,000	1	39601,000	63,463	,000
Error(reg)	Linear		9360,000	15	624,000		
freq		Linear	28561,000	1	28561,000	31,178	,000
Error(freq)		Linear	13741,000	15	916,067		
reg * freq	Linear	Linear	12045,063	1	12045,063	10,825	,005
Error(reg*freq)	Linear	Linear	16690,938	15	1112,729		

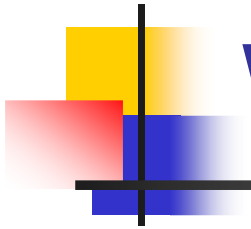
What are ‚ainm effects and interactions?

- The difference between the differences

- Main effect 1: $a + b =?= c + d$
- Main effect 2: $a + c =?= b + d$
- Interaction: $a - b =?= c - d$

Results: More like real results





What is an interaction?

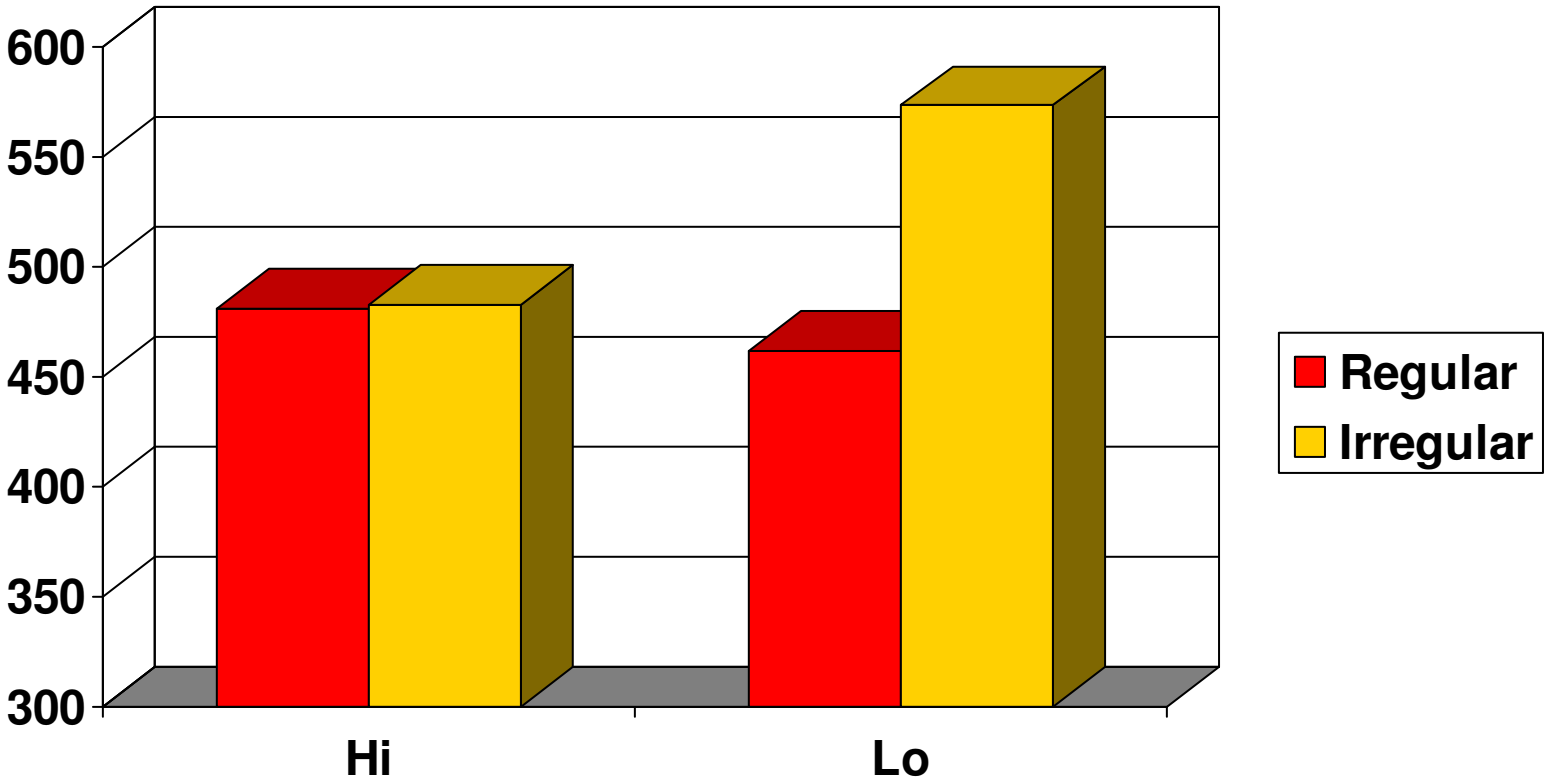
- The difference between the differences

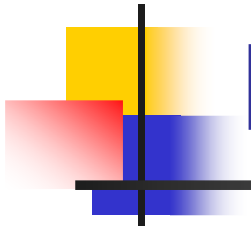
- Main effect 1: $(a + b) - (c + d) =?= 0$
- Main effect 2: $(a + c) - (b + d) =?= 0$
- Interaction: $(a - b) - (c - d) =?= 0$

Main Effects and Interaction?

- If therre is an interaction between regularity and frequency
- What can we conclude about the effects of regularity and frequency?
  - Can’t be sure that they are not due to the interaction

Results: More like real results





Main Effects and Interaction?

- How can we best interpret the interaction between regularity and frequency and the main effects
- Post-hoc analyses

Interaction/main effect posthocs

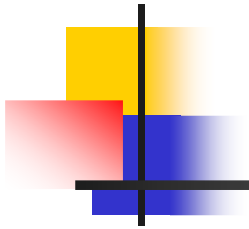
Paired Samples Test

		Paired Differences			
		Mean	Std. Deviation	Std. Error Mean	95% C Interv Diff Lower
Pair 1	reghi - reglo	-14,81250	9,33251	2,33313	-19,78545
Pair 2	reghi - irreghi	-22,31250	27,66639	6,91660	-37,05488

Interaction/main effect posthocs

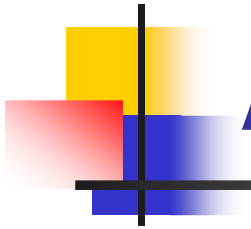
aired Samples Test

ifferences					
	95% Confidence Interval of the Difference				
d. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
33313	-19,78545	-9,83955	-6,349	15	,000
91660	-37,05488	-7,57012	-3,226	15	,006



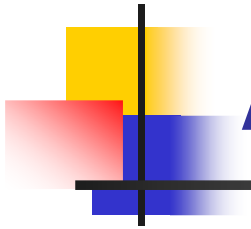
Interaction/main effect posthocs

- Frequency and regularity both have effects (< posthocs)
- But the combination of the two leads to greater difficulty than just the sum of the two effects (< existence of interaction)



Assumptions of MANOVA

- Independence of observations (as in univariate ANOVA)
- Multivariate normality - all dependent variables and linear combinations of them are distributed normally
- Equality of covariance matrices (cf homogeneity of variance in univariate)



Assumptions of MANOVA

- Second and third assumptions are more stringent than corresponding univariate assumptions in univariate ANOVA.