

Statistics in experimental research

Session 3

Francesca Delogu

delogu@coli.uni-saarland.de

Overview today

- ▶ ANOVAs with more than one factor
 - ▶ Why test more than one factor?
 - ▶ Interactions
 - ▶ Planned and post hoc comparisons
- ▶ Recap of Hypothesis testing
- ▶ Which-test-to-use-when
- ▶ Software & books
- ▶ How to set up a psycholinguistic experiment

The independant variable

- ▶ Our example: coffee/ no coffee
- ▶ Can have more levels (coffee, tea, water)
 - ▶ Compare 3 groups!
- ▶ There can be more than one independent variable!
 - ▶ Coffee/ no coffee
 - ▶ Enough sleep / sleep deprivation

=> Compare 4 groups!

Why test more than 1 independent variable?

- ▶ Why not test variables separately – in two different experiments?
- ▶ Because we expect an interaction!
 - ▶ The factors coffee and sleep influence each other

Factorial design of an experiment

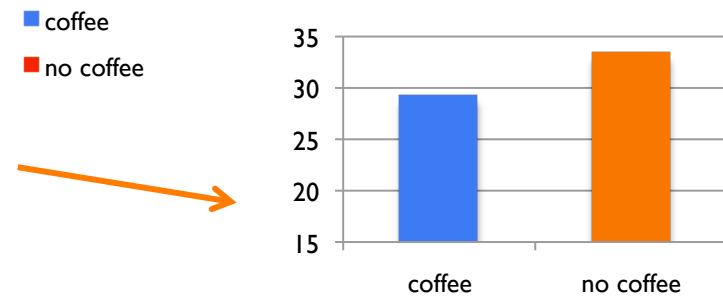
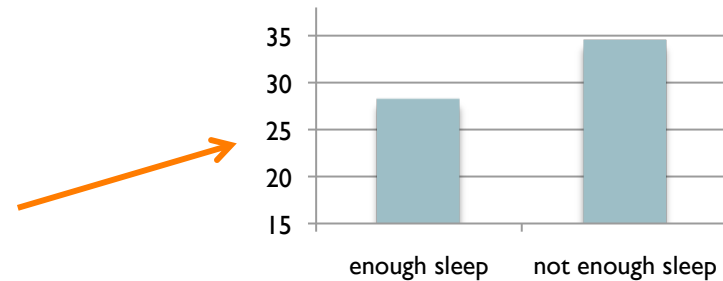
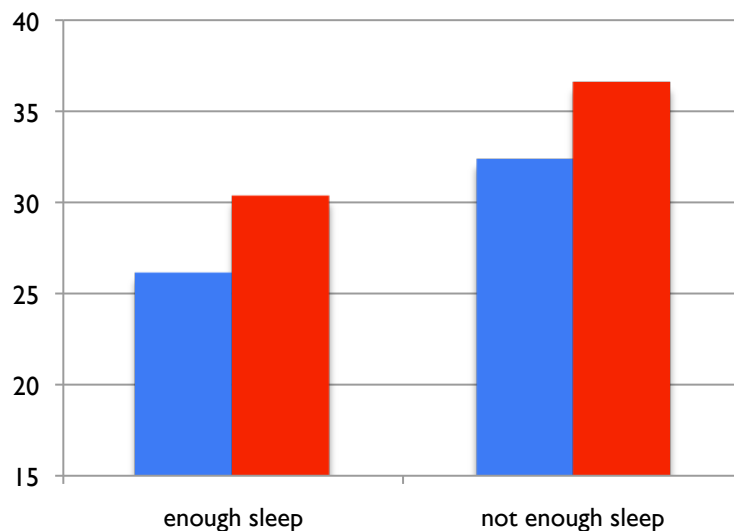
- ▶ Crossing two independent variables leads to 4 experimental conditions:
- ▶ COFFEE(coffee/nocoffee) x SLEEP(enough_sleep/not_enough_sleep)
 - ▶ Condition 1: coffee, enough_sleep
 - ▶ Condition 2: coffee, not_enough_sleep
 - ▶ Condition 3: nocoffee, enough_sleep
 - ▶ Condition 4: nocoffee, not_enough_sleep
- ▶ Experiments can have more factors
 - ▶ 2 x 3 design = 6 conditions
 - ▶ 2 x 2 x 2 design = 8 conditions

What do we want from the analysis?

- ▶ Does coffee have an effect (main effect of coffee)?
- ▶ Does sleep have an effect (main effect of sleep)?
- ▶ Do coffee and sleep influence each other (interaction of coffee and sleep)?

Possible outcomes

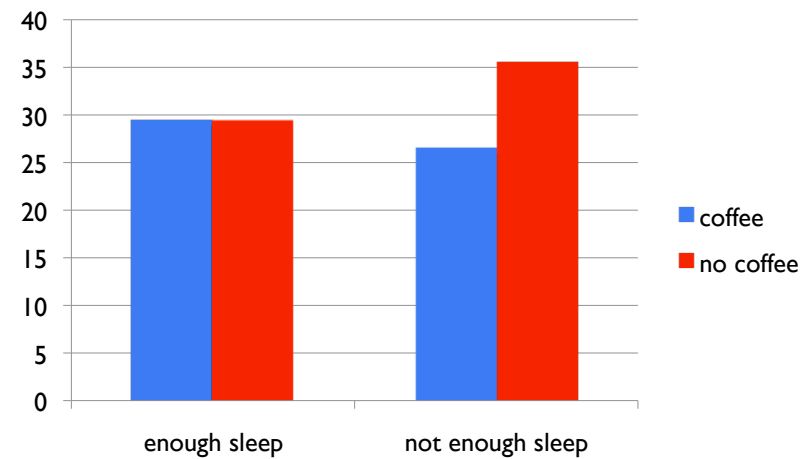
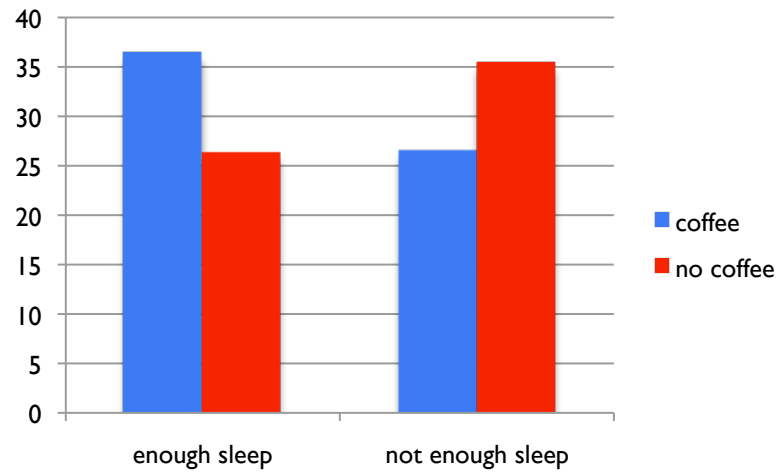
- ▶ Two main effects



- ▶ People are faster with coffee
- ▶ People are faster with enough sleep

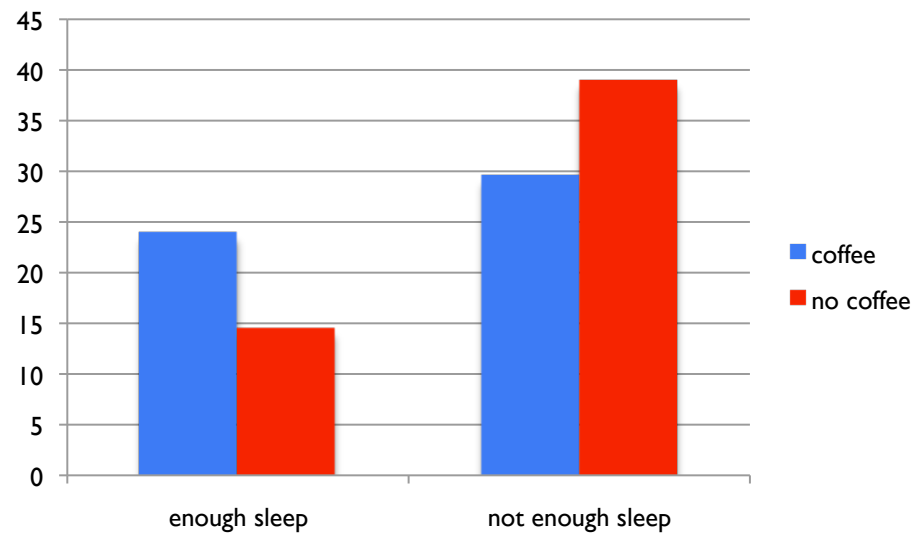
Possible outcomes II

► Interaction



Possible outcomes III

► Main effect of sleep+ interaction



How do we interpret an interaction?

- ▶ The information we get from the ANOVA is that there is an interaction, not what kind of interaction
- ▶ At least we know that the two factors influence each other (are not independent)
- ▶ We don't know which differences between individual conditions are significant
- ▶ Pairwise comparisons!

Planned vs post hoc

- ▶ **Planned comparisons:**

- ▶ Your hypothesis predicts a particular data pattern, e.g.
Coffee makes students faster, but only if they are tired before

- ⇒ Difference between the conditions:

- ▶ coffee, not_enough_sleep
 - ▶ nocoffee, not_enough_sleep

- ⇒ No Difference between the conditions:

- ▶ coffee, enough_sleep
 - ▶ nocoffee, enough_sleep

- ▶ **Perform two t tests with Bonferroni correction**

Planned vs post hoc

- ▶ Post hoc tests:
 - ▶ Your hypothesis didn't state particular differences (possibly because you did not expect an interaction)
 - ▶ Test all possible pairs!
 - ▶ Have to use a more conservative correction here
 - ▶ Tukey's Test

Different types of ANOVA

- ▶ ANOVA with 1 factor between subjects
 - ▶ one-way ANOVA
- ▶ ANOVA with more than 1 factor between subjects
 - ▶ e.g., 2x2 between subjects ANOVA
- ...
- ▶ ANOVA with 1 or more factors within subjects
 - ▶ Repeated measures ANOVA
 - ▶ e.g., 2x2 within subjects ANOVA
- ...
- ▶ Mixed ANOVA
 - ▶ e.g., 2x2 ANOVA with one factor within and one factor between

Example of mixed design

- ▶ Coffee → within
 - ▶ all students tested with and without coffee
- ▶ Sleep → between
 - ▶ Half students tested after sleep deprivation and half tested after enough sleep

Summary ANOVA

- ▶ Dependant variable: continuous
- ▶ One or more independent variables with 2 or more levels each
- ▶ Gives significance-values for
 - ▶ Main effect (effect of one factor)
 - ▶ Interaction (influence of factors on each other)
- ▶ Usually requires additional testing
 - ▶ Planned comparisons
 - ▶ Post hoc tests

Hypothesis testing

- ▶ Identify the hypothesis
 - ▶ Be as specific as you can be!
- ▶ Define your dependant variable and your independant variable (s)
- ▶ Classify your variables
 - ▶ Continuous or categorical?
- ▶ Do you test the same entity (person) in all conditions?
 - ▶ Use the paired or repeated measure variant
- ▶ Choose an appropriate test
 - ▶ T test, chi square test, ANOVA, something else

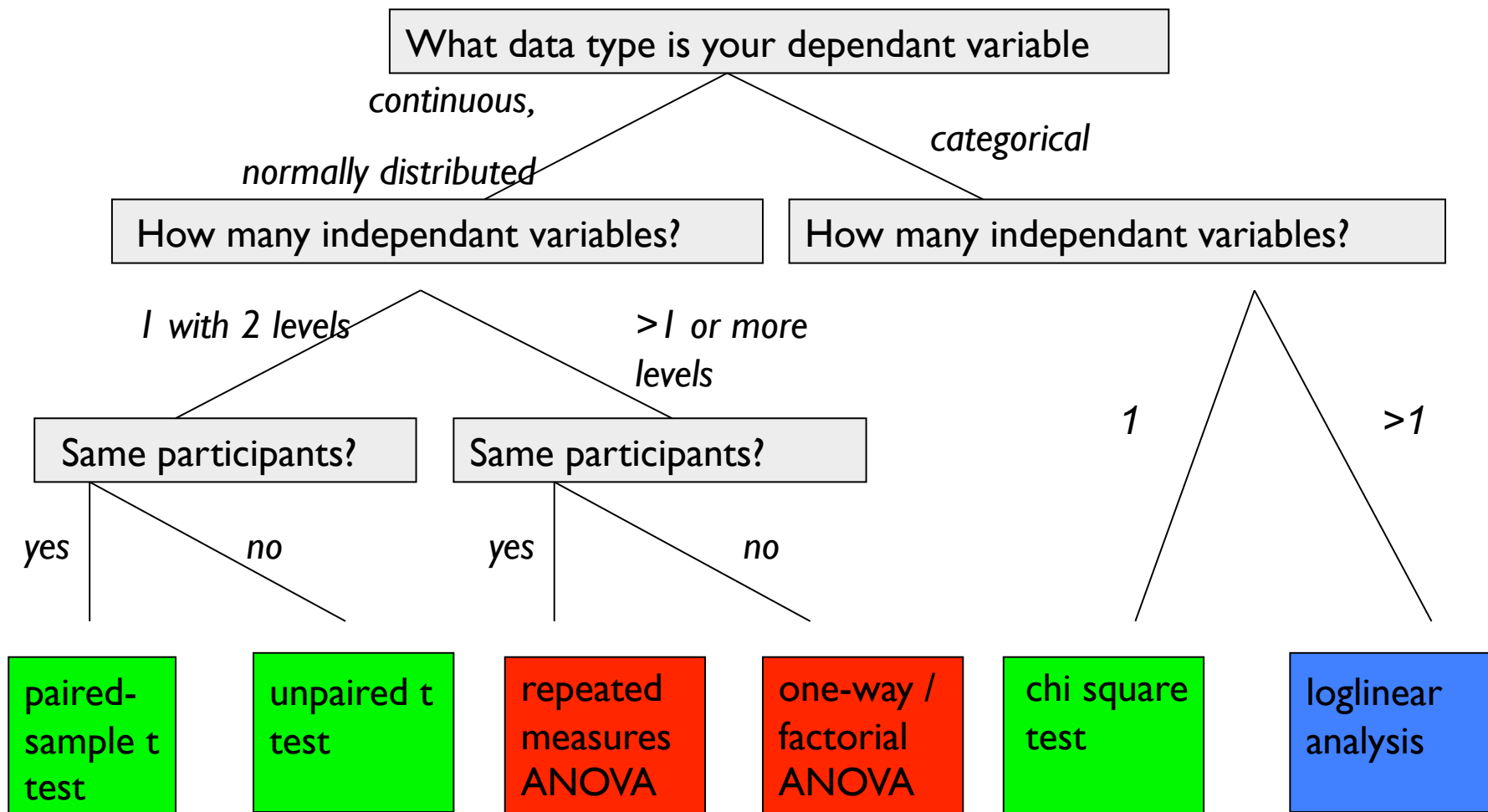
Hypothesis testing II

- ▶ Calculate the test statistic
 - ▶ Or have a programm do this for you ;)
- ▶ Compare the test statistic to the critical value depending on your α
- ▶ If the test statistic is above the critical value
 - ▶ Your result is significant, i.e.
 - ▶ The probability of observing your data if the Null-Hypothesis were true is below α ($p < 0.05$)
- ▶ If the test statistic is below the critical value
 - ▶ Your result is not significant
 - ▶ You can't reject the Null-Hypothesis
 - ▶ DOES NOT MEAN THAT THERE IS NO DIFFERENCE!

Important concepts

- ▶ Dependant vs independant variables
- ▶ Data types: continuous vs categorical
- ▶ Level of significance:
 - ▶ α is the predefined boundary
 - ▶ p-value is the actual probability of our observation if the Null-Hypothesis is true
- ▶ Population and sample

Which test to use



Questions?

- ▶ What is α ?
- ▶ What does it mean for a difference to be statistically significant?
- ▶ When do we use the t test?
- ▶ What is a continuous/categorical variable?
- ▶ What is a dependant/independant variable?
- ▶ Why do we use the variance of the sample in calculating t?
- ▶ Where do we get our hypotheses from?
- ▶ What is the Null Hypothesis (H_0)?
- ▶ What is an interaction?
- ▶ What kinds of errors can we make in hypothesis testing?

Software for statistical analysis

▶ Excel:

- Chi square test, T tests
- Descriptives: mean, variance, graphs
- + probably available to everybody 😊

▶ SPSS:

- ANOVAs, loglinear, non-parametric tests
- Everything that Excel can do
- licences are expensive 😞

available in the psycholinguistics department, but not on all machines

Software for statistical analysis

► R

- Everything that Excel and SPSS can do
- Can do other models (mixed effects models etc)
- Without graphical user interface
 - ⇒ You have to know what you are doing!
- + can be downloaded for free 😊

Helpful readings

Statistics in SPSS:

Field, Andy (2009). *Discovering statistics using SPSS*. London, England: SAGE.

Statistics in R:

Baayen, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

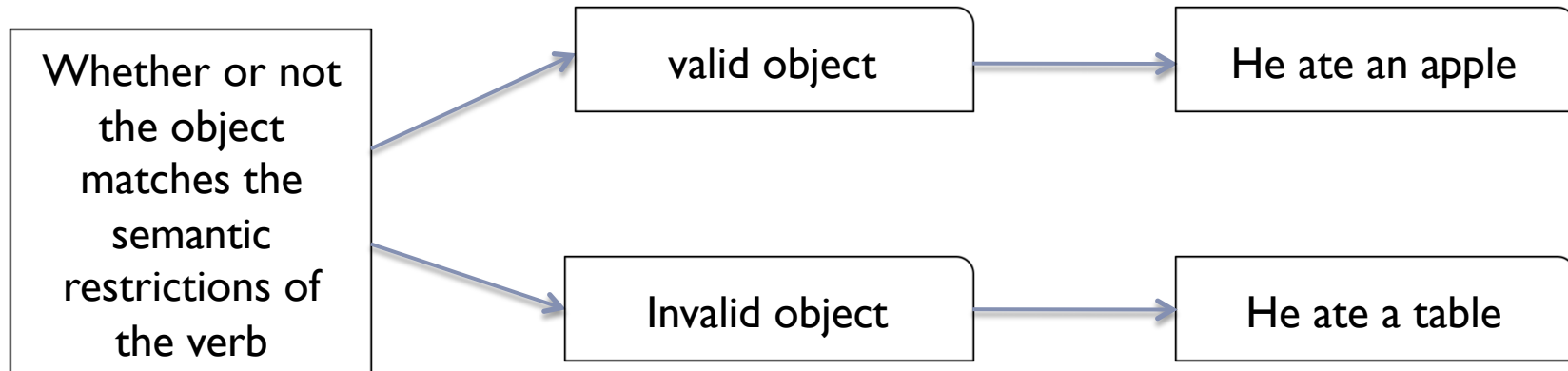
Setting up a psycholinguistic experiment

- ▶ Suppose you want to test whether people take longer to read a word if it does not match the semantic restrictions of the verb
 - ▶ e.g, He eats an apple. vs He eats a table.
- ▶ What is your IV? and your DV?
 - ▶ How many levels?
- ▶ What is the H_0 ? And the H_1 ?
- ▶ Where do we sample from?

IV

Conditions

Example



DV → reading times

Statistical hypotheses

- ▶ H_0 : people are neither faster nor slower to read a word if it does not match the semantic restrictions of the verb
 - ▶ people are neither faster nor slower to read “table” in “he ate a table” than “apple” in “he ate an apple”
- ▶ H_1 : people are faster (or slower) to read a word if it does not match the semantic restrictions of the verb
 - ▶ people are faster (or slower) to read “table” in “he ate a table” than “apple” in “he ate an apple”

Where do we sample from

- ▶ All English speakers
 - ▶ Participants
 - ▶ a random sample of English speakers
- ▶ All English nouns
 - ▶ Items
 - ▶ a constructed sample of English sentences containing a selective verb and a noun
- ▶ How do we create the items?

Items: how to reduce variation

- ▶ Test the same item in all conditions:

valid	invalid
1 Peter eats an apple	Peter drives an apple
2 Paul plants a tree	Paul smokes a tree
3 Suzy reads a book	Suzy drinks a book

...

- ▶ Control for additional variation: the verb
 - ▶ Because reading times on a word may be influenced by reading times on the preceding word if the latter is longer or less frequent....
.... we need to control for frequency and length of the verb
- ▶ Verb restrictions might be of different strength:
 - ▶ Use the same verb in the other condition, too!

Constructing items: counterbalancing

► Original version:

valid

- 1 Peter eats an apple
- 2 Paul plants a tree
- 3 Suzy reads a book

...

invalid

- Peter drives an apple
Paul smokes a tree
Suzy drinks a book

► Counterbalancing version:

valid

- 1 Peter drives a car
- 2 Paul smokes a cigar
- 3 Suzy drinks a beer

...

invalid

- Peter eats a car
Paul plants a cigar
Suzy reads a beer

Constructing lists

- ▶ Usually, we show each participant each item in only one condition
 - ▶ They might react differently when reading the same word again
 - ▶ 2 conditions and a counterbalancing version for the item
 - ⇒ 4 experimental lists
 - Each should be tested an equal number of times
- ▶ Make sure that every condition appears equally often
- ▶ Randomize the list!

Validity

- ▶ If your participants can guess the goal of your experiment, they might behave differently!

⇒ Don't tell them the purpose

⇒ Try to distract them from the purpose

- e.g. put in more sentences, that don't have anything to do with the experiment (filler items)

- ▶ What's next?
- ▶ Run the experiment!

Summary

- ▶ You sample from two populations:
 - ▶ Participants
 - ▶ Items
- ▶ Try to eliminate as much variation in your materials as you can!
 - ▶ Controll for factors
 - ▶ Counterbalance your materials
- ▶ Try to prevent your participant from behaving strategically
 - ▶ Introduce filler items to distract from the real purpose