Aalto University
School of Science

Experimental and statistical methods II

22.01.2014

heini.heikkila@aalto.fi

Heini Heikkilä

Brain and Mind Laboratory

Outline

- Recap: stats from the previous course
- New: repeated measures

Our tools

- Descriptive statistics
- Plots
- Group comparisons
  - t tests
  - ANOVA
- Trends in the data
  - correlation
  - regression
- NEW: repeated measures design

Example data

Blindspots vs. Spotlights experiment

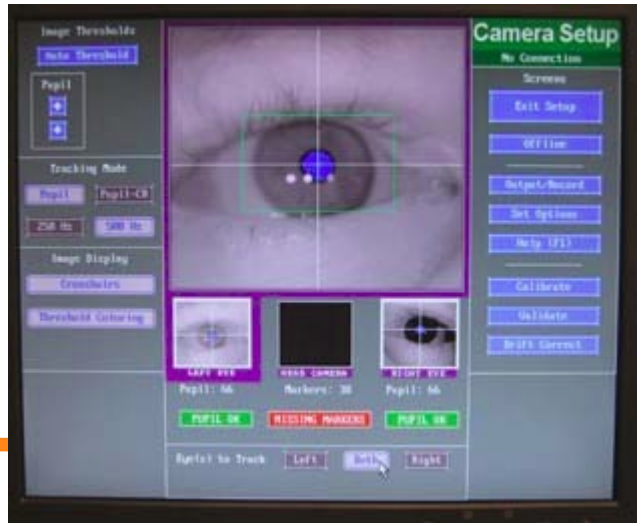
Visual search in a real-life scene

Eye-tracking
- RT to find the object

Two conditions:
- spotlight
- blindspot

N=28

http://www.nuthmann.de/antje/Site/scotoma.html



Example data

Blindspots vs. Spotlights experiment

2x3 design:
- type of image degradation (blindspots vs. spotlights)
- window size (small, medium, large)

→ two factors

Outcome variable:
- reaction time (amount of time to find a target object in the scene)

⟶ continuous variable

Preparing the data

For repeated-measures ANOVA (more soon..) in R, we need to have the data in **long** format. [Use 'reshape' if necessary.]

```
> head(data,10)
   subject degradation_type window_size       RT
1        1       blindspots       small 2647.148
2        1       blindspots      medium 2647.878
3        1       blindspots         big 2604.351
4        1       spotlights       small 4902.754
5        1       spotlights      medium 4406.960
6        1       spotlights         big 2956.953
7        2       blindspots       small 4478.689
8        2       blindspots      medium 2342.486
9        2       blindspots         big 3554.507
10       2       spotlights       small 6098.540
```

Long format

One row for each condition, multiple rows per subject

```
> head(data,10)
   subject degradation_type window_size       RT
1        1       blindspots       small 2647.148
2        1       blindspots      medium 2647.878
3        1       blindspots         big 2604.351
4        1       spotlights       small 4902.754
5        1       spotlights      medium 4406.960
6        1       spotlights         big 2956.953
7        2       blindspots       small 4478.689
8        2       blindspots      medium 2342.486
9        2       blindspots         big 3554.507
10       2       spotlights       small 6098.540
```

Wide format

One row per each subject, columns for different conditions

```
> head(data.wide)
   subject RT.small.blindspots RT.medium.blindspots RT.big.blindspots RT.small.spotlights RT.medium.spotlights RT.big.spotlights
1        1            2647.148             2647.878          2604.351            4902.754             4406.960          2956.953
7        2            4478.689             2342.486          3554.507            6098.540             3915.673          3583.427
13       3            2451.279             3286.329          2073.243            6149.233             3008.687          3430.727
19       4            3280.187             3003.340          3745.962            4585.790             5366.018          3931.450
25       5            1936.893             2628.053          3396.849            5697.515             3328.986          2790.949
31       6            3233.802             2049.298          2455.984            5529.814             2905.291          1856.470
```

Descriptive statistics

```
> summary(data)
    subject       degradation_type window_size      RT          
 1      :  6   blindspots:84      big   :56    Min.   : 937.2  
 2      :  6   spotlights:84      medium:56    1st Qu.:2405.3  
 3      :  6                      small :56    Median :3023.0  
 4      :  6                                   Mean   :3237.0  
 5      :  6                                   3rd Qu.:3903.5  
 6      :  6                                   Max.   :6602.3  
 (Other):132                                                   
```

factors

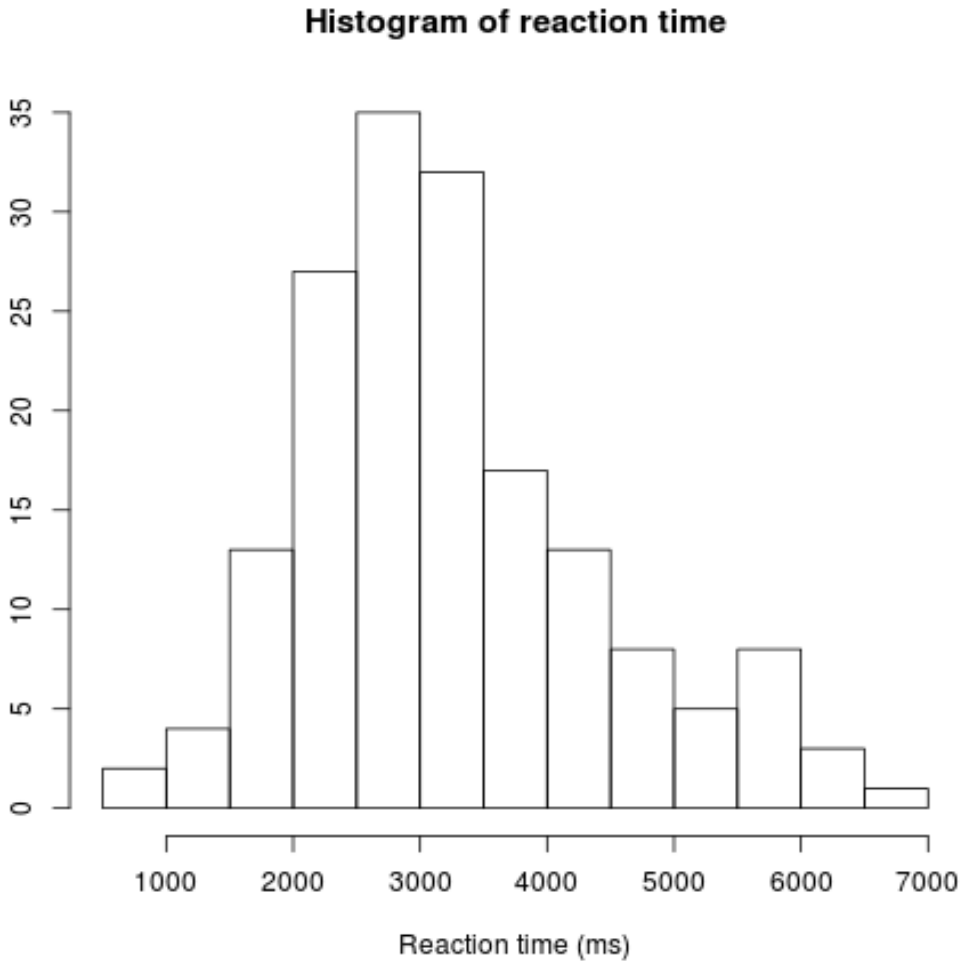
continuous variable

```
> describe(data)
                  var   n    mean      sd  median trimmed     mad     min     max   range skew kurtosis    se
subject*            1 168   14.50    8.10   14.50   14.50   10.38    1.00   28.0   27.00  0.0    -1.22  0.63
degradation_type*   2 168    1.50    0.50    1.50    1.50    0.74    1.00    2.0    1.00  0.0    -2.01  0.04
window_size*        3 168    2.00    0.82    2.00    2.00    1.48    1.00    3.0    2.00  0.0    -1.52  0.06
RT                  4 168 3237.02 1175.11 3023.02 3139.81 1048.46  937.15 6602.3 5665.15  0.7     0.06 90.66
```

Plots

- examining your data

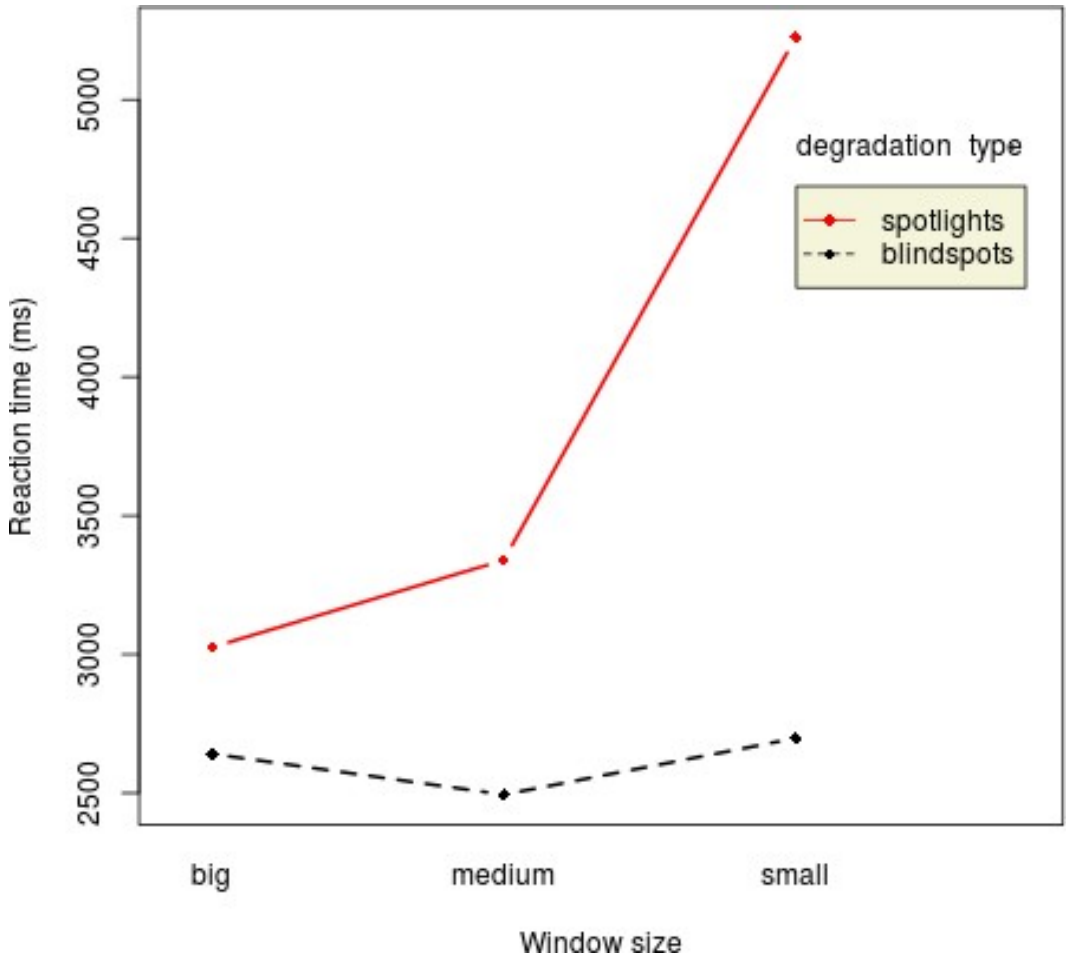
- high-quality graphs for reports and publications



Plots

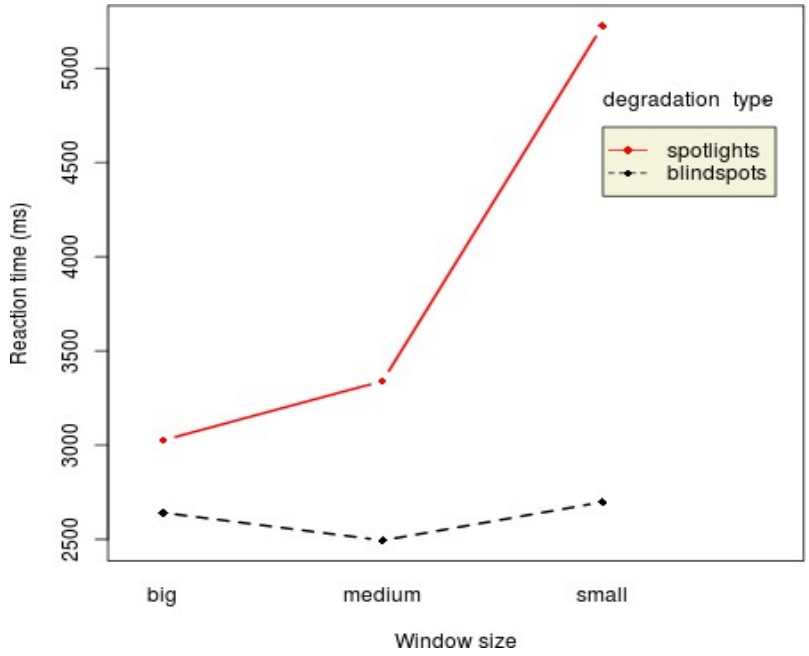
- examining your data

- high-quality graphs for reports and publications



Tables

- examining your data
- high-quality tables for reports and publications



		Window size		
		big	medium	small
Degradation type	blindspots	2641.2	2494.4	2697.5
	spotlights	3025.5	3340.0	5223.5

Statistical tests

Group differences:

- Factor with two levels (i.e., comparing 2 groups): t tests
- Factor with ≥ three levels: ANOVA

Trends in the data:

- Two ontinuous variables: correlations
- Multiple continuous variables, factors: regression (linear models)

T tests and ANOVA: assumptions

T-tests

- Sample size > 20
- Normality
- Continuous variables

ANOVA

- Dependent variables is continuous
- One discrete variable defining group membership
- Sample size > 15 per group
- Normality
- Equality of variances

→ If these are violated, use nonparametric tests or transformations!

Repeated measures design

e.g.

- longitudinal data
- experiments with multiple conditions
    - in example data: spotlights and blindspots

Why repeated measures design?

subject	storeA	storeB	storeC	storeD
lettuce	1.17	1.78	1.29	1.29
potatoes	1.77	1.98	1.99	1.99
milk	1.49	1.69	1.79	1.59
eggs	0.65	0.99	0.69	1.09
bread	1.58	1.7	1.89	1.89
cereal	3.13	3.15	2.99	3.09
ground.beef	2.09	1.88	2.09	2.49
tomato.soup	0.62	0.65	0.65	0.69
laundry.detergent	5.89	5.99	5.99	6.99
aspirin	4.46	4.84	4.99	5.15

Why repeated measures design?

subject	storeA	storeB	storeC	storeD
lettuce	1.17	1.78	1.29	1.29
potatoes	1.77	1.98	1.99	1.99
milk	1.49	1.69	1.79	1.59
eggs	**0.65**	0.99	0.69	1.09
bread	1.58	1.7	1.89	1.89
cereal	3.13	3.15	2.99	3.09
ground.beef	2.09	1.88	2.09	2.49
tomato.soup	0.62	0.65	0.65	0.69
laundry.detergent	**5.89**	5.99	5.99	6.99
aspirin	4.46	4.84	4.99	5.15

Repeated measures t test

Are there differences in reaction times between degradation types?

```
> head(data,10)
   subject degradation_type window_size       RT
1        1       blindspots       small 2647.148
2        1       blindspots      medium 2647.878
3        1       blindspots         big 2604.351
4        1       spotlights       small 4902.754
5        1       spotlights      medium 4406.960
6        1       spotlights         big 2956.953
7        2       blindspots       small 4478.689
8        2       blindspots      medium 2342.486
9        2       blindspots         big 3554.507
10       2       spotlights       small 6098.540
```

Repeated measures t test

```
> t.test(RT~degradation_type, paired=T)

 Paired t-test

data:  RT by degradation_type
t = -9.3553, df = 83, p-value = 1.271e-14
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1518.1290  -985.7879
sample estimates:
mean of the differences
              -1251.958
```

Oops! → degrees of freedom vs N=28??

Solution: reshape to wide or use ANOVA

Repeated measures ANOVA

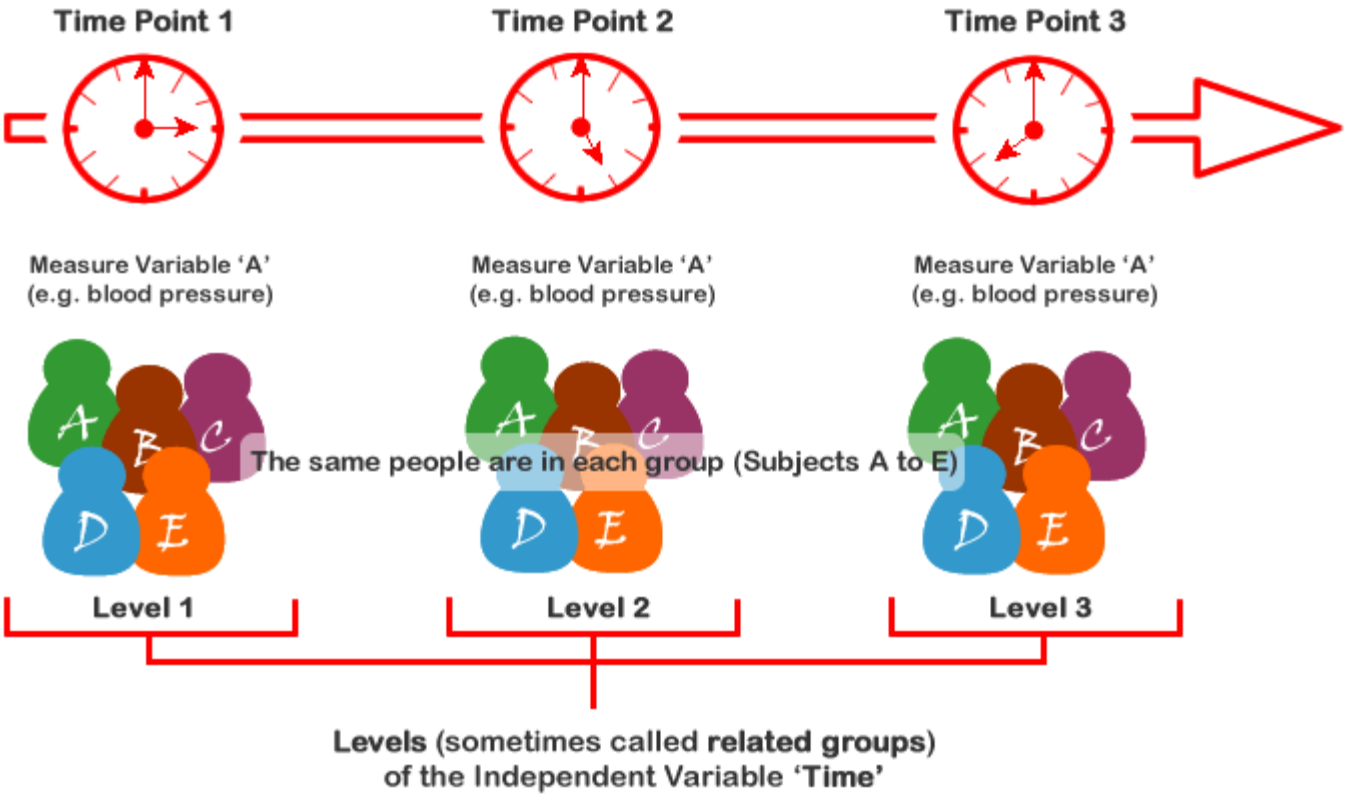
Assumptions:

- balanced design
- random assignment to groups
- individual differences are error

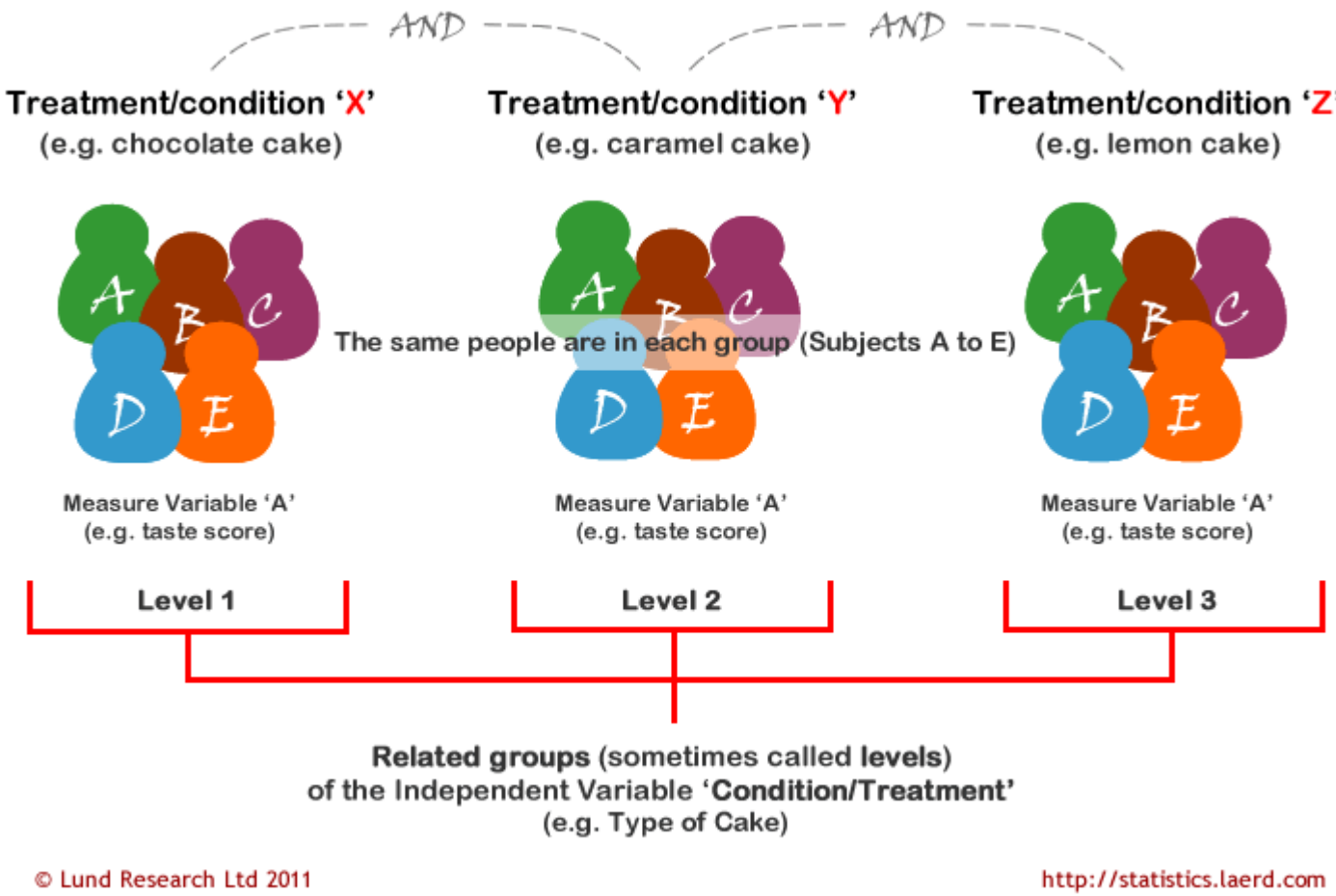
Hypothesis:

- Are there any differences between related population means?
- Null hypothesis: means are equal.

Repeated measures: longitudinal design

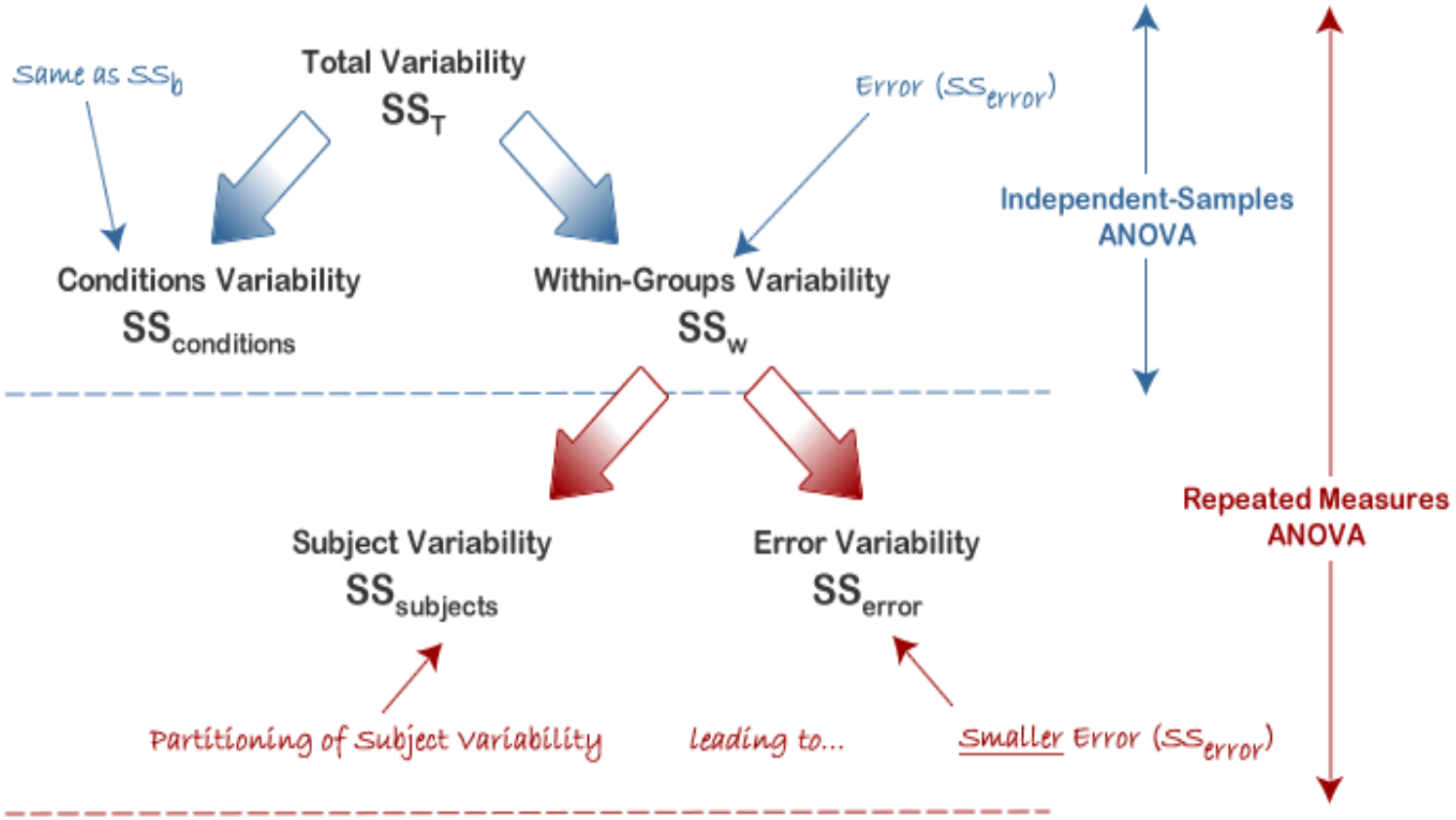


Repeated measures: multiple conditions



The already familiar between-subjects ANOVA

Our newer friend within-subjects ANOVA



Why repeated measures design?

subject	storeA	storeB	storeC	storeD
lettuce	1.17	1.78	1.29	1.29
potatoes	1.77	1.98	1.99	1.99
milk	1.49	1.69	1.79	1.59
eggs	**0.65**	0.99	0.69	1.09
bread	1.58	1.7	1.89	1.89
cereal	3.13	3.15	2.99	3.09
ground.beef	2.09	1.88	2.09	2.49
tomato.soup	0.62	0.65	0.65	0.69
laundry.detergent	**5.89**	5.99	5.99	6.99
aspirin	4.46	4.84	4.99	5.15

One-way within-subjects ANOVA

- When testing for the effect of one factor.
- Are there differences in reaction times between degradation types?
  - Expect the subjects to vary in the effect of degradation type.
- Logic: in ANOVA, add an error term that reflects that we have 'degradation types nested within subjects'

```
> A1 ← aov(RT ~ degradation_type + Error(subject/degradation_type))
```

Compare to the one-way ANOVA:

```
> A0 ← aov(RT ~ degradation_type)
```

One-way within-subjects ANOVA

```
> A1 <- aov(RT~degradation_type + Error(subject/degradation_type))
> summary(A1)

Error: subject
          Df   Sum Sq Mean Sq F value Pr(>F)
Residuals 27 39964303 1480159

Error: subject:degradation_type
                 Df   Sum Sq  Mean Sq F value Pr(>F)
degradation_type  1 65830798 65830798   330.9 <2e-16 ***
Residuals        27  5371007   198926
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Error: Within
           Df    Sum Sq Mean Sq F value Pr(>F)
Residuals 112 119442844 1066454
```

degrees of freedom looks better now!

One-way within-subjects ANOVA

```
> model.tables(A1, 'means')
Tables of means
Grand mean

3237.023

 degradation_type
degradation_type
blindspots spotlights
      2611       3863
```

Two-way within-subjects ANOVA

- When testing for the effect of two or more factors.
- Are there differences in reaction times if we vary **degradation type**?
  Are there differences in reaction times if we vary **window size**?
  Are there differences in reaction times caused by **an interaction effect of degradation type and window size**?
  - Expect the subjects to vary in the degradation type effect, in the window size effect, and in the interaction effect.

```
> A2 ← aov(RT ~ degradation_type*window_size +
            Error(subject/degradation_type*window_size)
```

Two-way within-subjects ANOVA

```
> A2 <- aov(RT~degradation_type*window_size +
Error(subject/(degradation_type*window_size)))
> summary(A2)

Error: subject
          Df   Sum Sq Mean Sq F value Pr(>F)
Residuals 27 39964303 1480159
```

main effect of degradation type

```
Error: subject:degradation_type
                 Df   Sum Sq  Mean Sq F value Pr(>F)
degradation_type  1 65830798 65830798   330.9 <2e-16 ***
Residuals        27  5371007   198926
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

main effect of window size

```
Error: subject:window_size
            Df   Sum Sq  Mean Sq F value   Pr(>F)
window_size  2 44164320 22082160   65.45 3.69e-15 ***
Residuals   54 18219221   337393
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

interaction effect

```
Error: subject:degradation_type:window_size
                           Df   Sum Sq  Mean Sq F value   Pr(>F)
degradation_type:window_size  2 35575390 17787695   44.71 3.52e-12 ***
Residuals                    54 21483912   397850
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Two-way within-subjects ANOVA

```
> model.tables(A2, 'means')
Tables of means
Grand mean

3237.023

 degradation_type
degradation_type
blindspots spotlights
      2611       3863

 window_size
window_size
   big medium  small
  2833   2917   3961

 degradation_type:window_size
                window_size
degradation_type big  medium small
      blindspots 2641 2494   2698
      spotlights 3026 3340   5223
```

Where are the differences?

If ANOVA gives you significant results, you might want to look where the differences are.

- post-hoc comparisons
- or planned comparisons (contrasts)

A simple way: run all possible pairwise comparisons with t tests (remember to correct for multiple comparisons!)

http://xkcd.com/882/

	big	medium	small
blindspots	2641.2	2494.4	2697.5
spotlights	3025.5	3340.0	5223.5

Need to use a nonparametric test?

- repeated measures t test:
  Wilcoxon Signed-Rank Test
  > ?wilcox.test

- repeated measures ANOVA:
  Friedman Rank Sum Test
  > ?friedman.test

More?

Repeated measures ANOVA:

Laerd Statistics Guide

William B. King's tutorial

Help with R/SPSS/Matlab:

- Online resources
- Ask us :)
-

Next demo sessions deal with common pitfalls in analysis and reporting based on our observations on the written reports – don't hesitate to email us with your concerns or questions!