Aalto University
School of Science

t-tests, ANOVA

Experimental and Statistical Methods in Biological Sciences I

Heini Saarimäki, BECS

02/10/2014

Outline for today

- Introduction to the basic concepts

- Demos and exercises
  - Simulations: different distributions
  - Testing for normality
  - t-tests
  - ANOVA

Outline for today

Now…

1. Starters
2. Simulations of distributions
3. Tests for normality
4. t-tests
5. ANOVA

What we know after today

1. Preparing your data

When preparing your data, check that…

- Factors are factors
- Missing values are coded as NAs

2. Describing your data

- Plotting
- Descriptive statistics

What we know after today

1. Preparing your data

When preparing your data, check that…

- Factors are factors
- Missing values are coded as NAs

2. Describing your data

- Plotting
- Descriptive statistics

3. Statistical comparisons

- t-tests
- ANOVA

2. Simulations of distributions

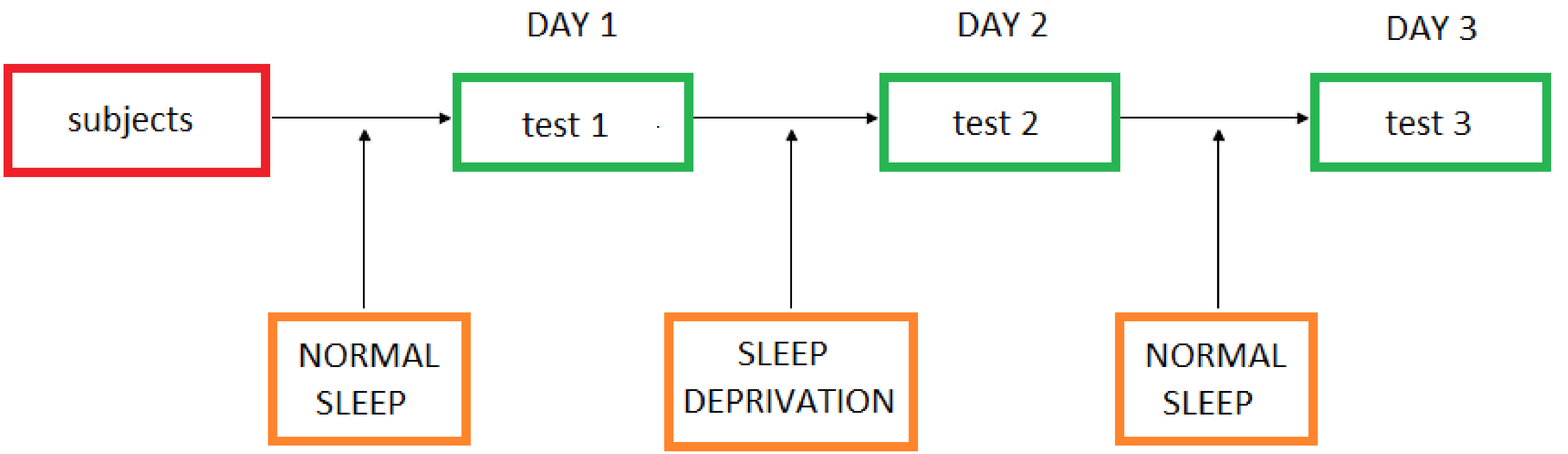
- In exercises, we will simulate the shape of three different kinds of distributions:
  - The normal distribution
  - Student's t-distribution (relevant for t-tests)
  - Fisher's F-distribution (relevant for ANOVA)

- Don't worry about the R code for simulations – just copy+paste and play around with the parameters.

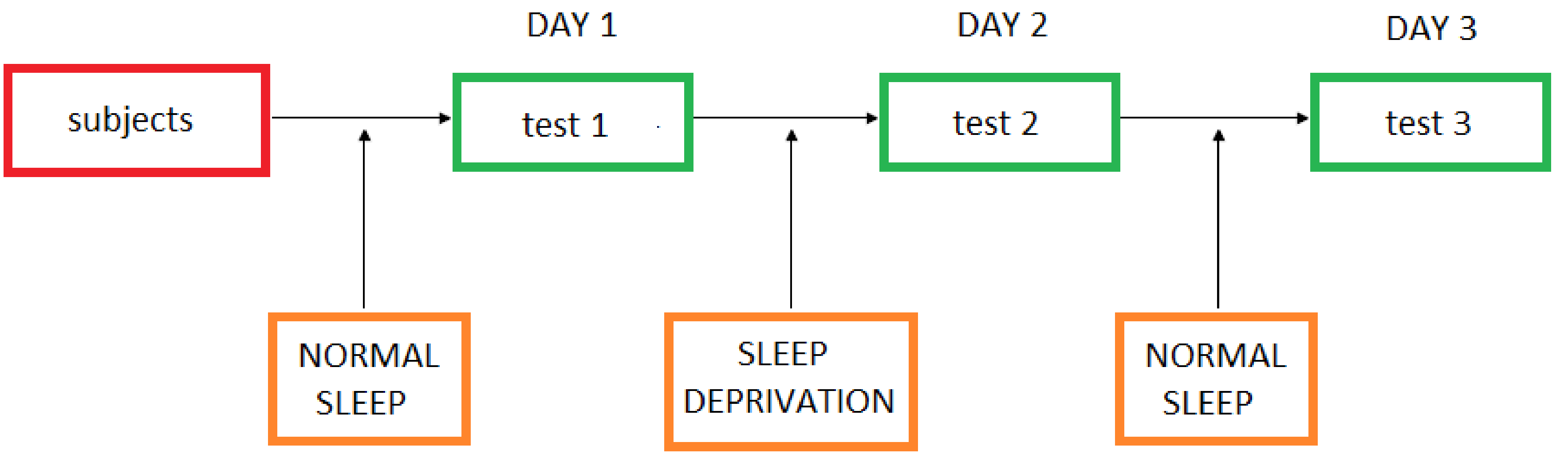
Our data today…

Variable	Type	Details
age group	categorical	1 = young 2 = old
education	categorical	1 = comp. school 2 = secondary 3 = higher
hormone treatment	categorical	user group control group
digit symbol task	continuous	at time points 1, 2, 3
benton task errors	continuous	at time points 1, 2,3

Our data today…

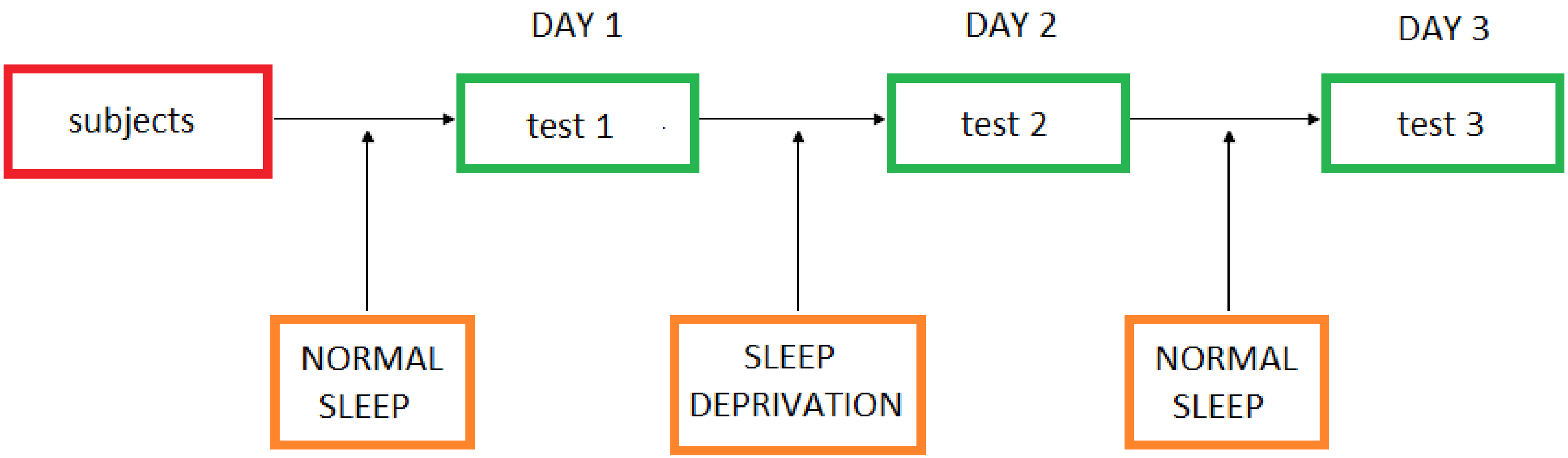


Our data today…



Categories: age group, education, hormone treatment

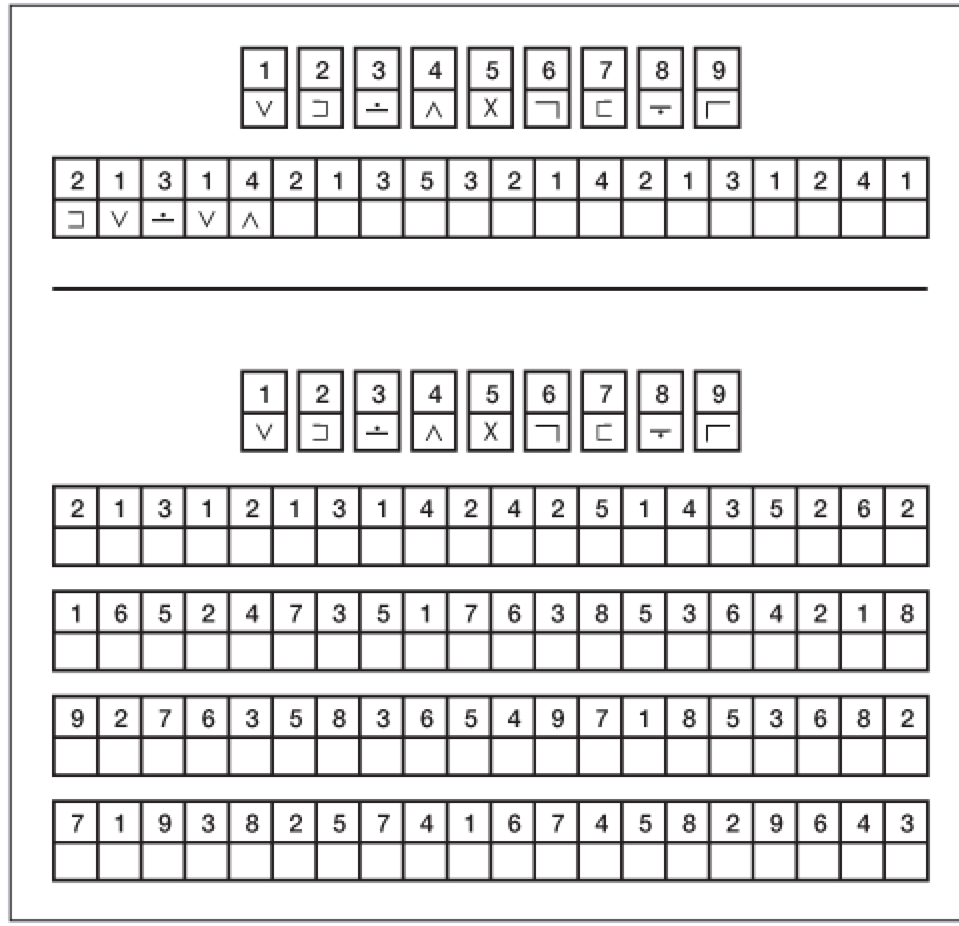
Our data today…



Longitudinal tests: Digit symbol task, Benton task errors

Our data today…

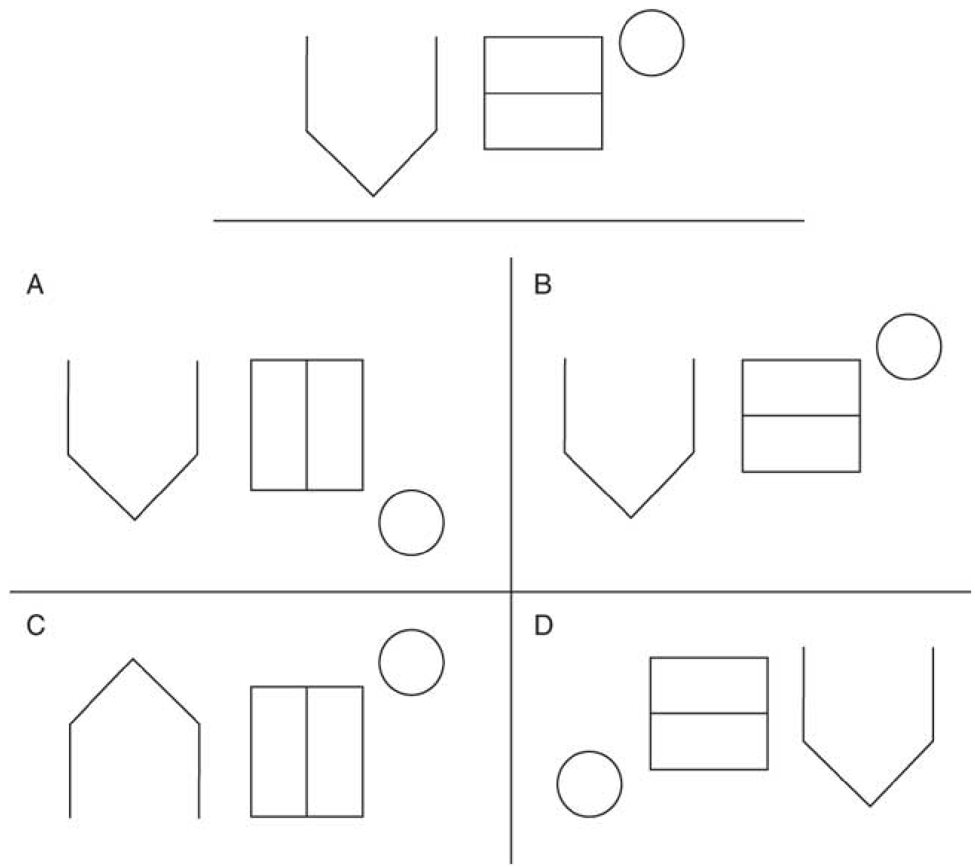
Digit symbol task



Our data today…

Benton Visual

Retention Task



3. Testing for normality

Parametric tests have assumptions…

- t-tests
  - Sample size > 20
  - Normality
  - Continuous variables
- ANOVA
  - Dependent variable is continuous
  - One or more discrete (categorical) variables defining group membership
  - Sample size > 15 per group
  - Normality
  - Equality of variances

3. Testing for normality

Parametric tests have assumptions…

- t-tests
  - Sample size > 20
  - **Normality**
  - Continuous variables
- ANOVA
  - Dependent variable is continuous
  - One or more discrete (categorical) variables defining group membership
  - Sample size > 15 per group
  - **Normality**
  - Equality of variances

3. Testing for normality

- Descriptive statistics:
  - Mean, st dev, median, skewness, …
- Plotting:
  - Histograms with normal curve added
  - Quantile-quantile plots
- Tests:
  - Kolmogorov-Smirnov
  - Shapiro-Wilk

3. Testing for normality

Descriptive statistics: `describe`

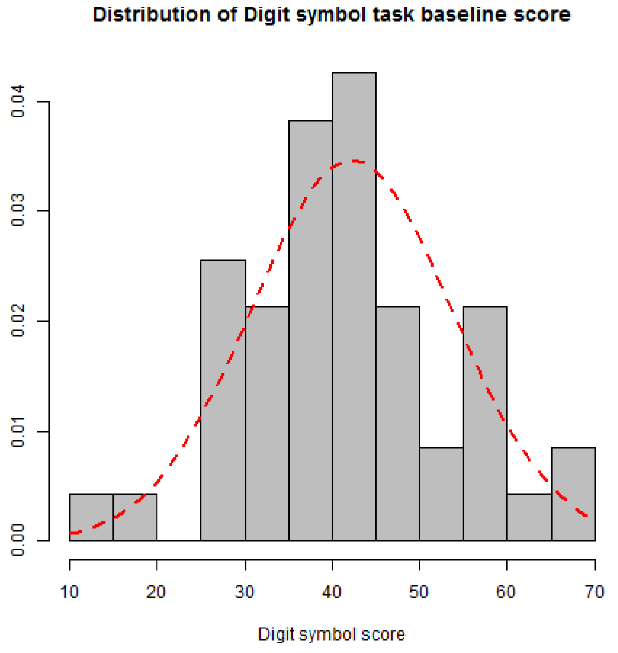
- Mean, st dev, median, skewness, …

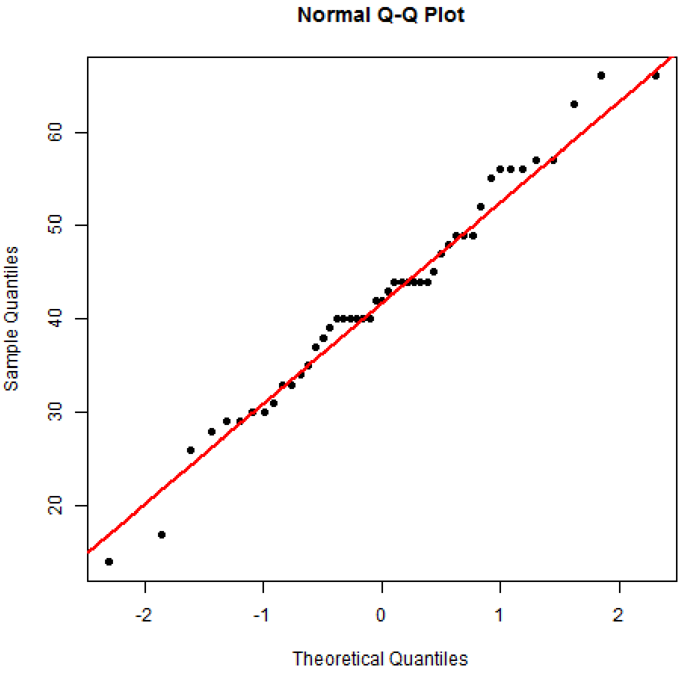
```
> describe(deprivation[,5:10])
               var  n  mean    sd median trimmed   mad min max range  skew kurtosis   se
digitsymbol_1    1 47 42.23 11.51     42   42.26 10.38  14  66    52 -0.05    -0.13 1.68
digitsymbol_2    2 47 42.91 12.40     44   42.49 10.38  17  70    53  0.28    -0.15 1.81
digitsymbol_3    3 47 48.49 13.81     48   48.08 11.86  15  81    66  0.23     0.28 2.01
bentonerror_1    4 47  6.04  2.42      6    5.97  2.97   2  11     9  0.20    -0.91 0.35
bentonerror_2    5 47  5.45  2.87      6    5.26  2.97   1  13    12  0.55    -0.06 0.42
bentonerror_3    6 47  3.77  2.42      4    3.64  2.97   0  11    11  0.63     0.61 0.35
```

3. Testing for normality

Plotting:

- Histograms with normal curve added: `hist`
- Quantile-quantile plots: `qqnorm, qqline`





3. Testing for normality

Tests:

- Kolmogorov-Smirnov: `ks.test`
- Shapiro-Wilk: `shapiro.test`

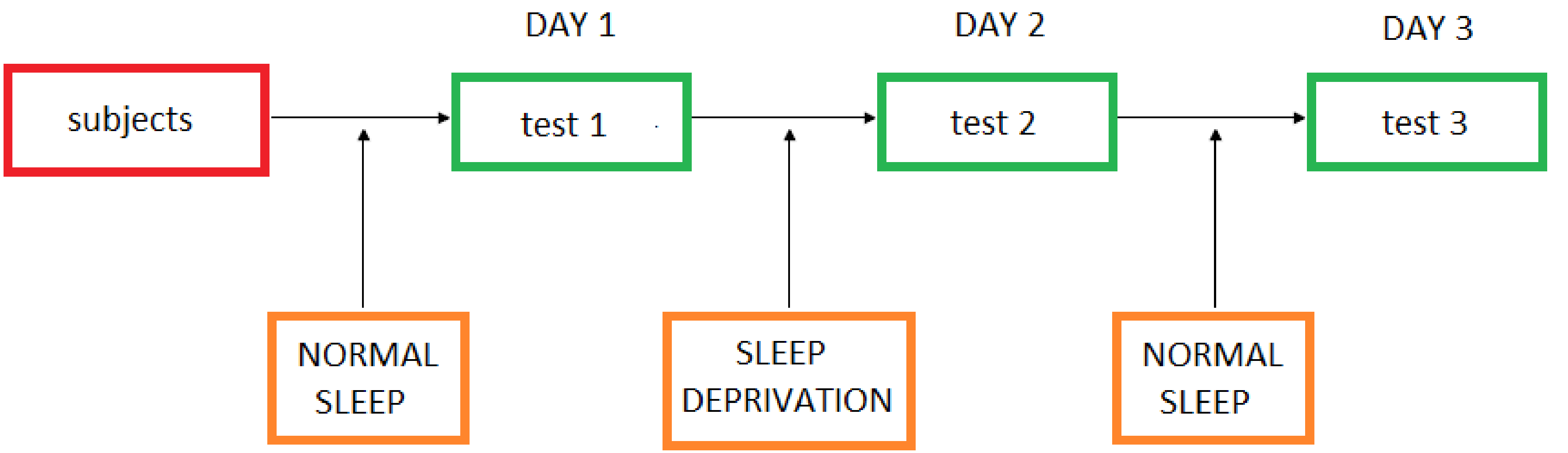
```
> shapiro.test(digitsymbol_1)

        Shapiro-Wilk normality test

data:  digitsymbol_1
W = 0.9809, p-value = 0.6287


> ks.test(digitsymbol_1, "pnorm", mean=mean(digitsymbol_1), sd=sd(digitsymbol_1))

        One-sample Kolmogorov-Smirnov test

data:  digitsymbol_1
D = 0.0986, p-value = 0.7506
alternative hypothesis: two-sided
```

4. t-tests

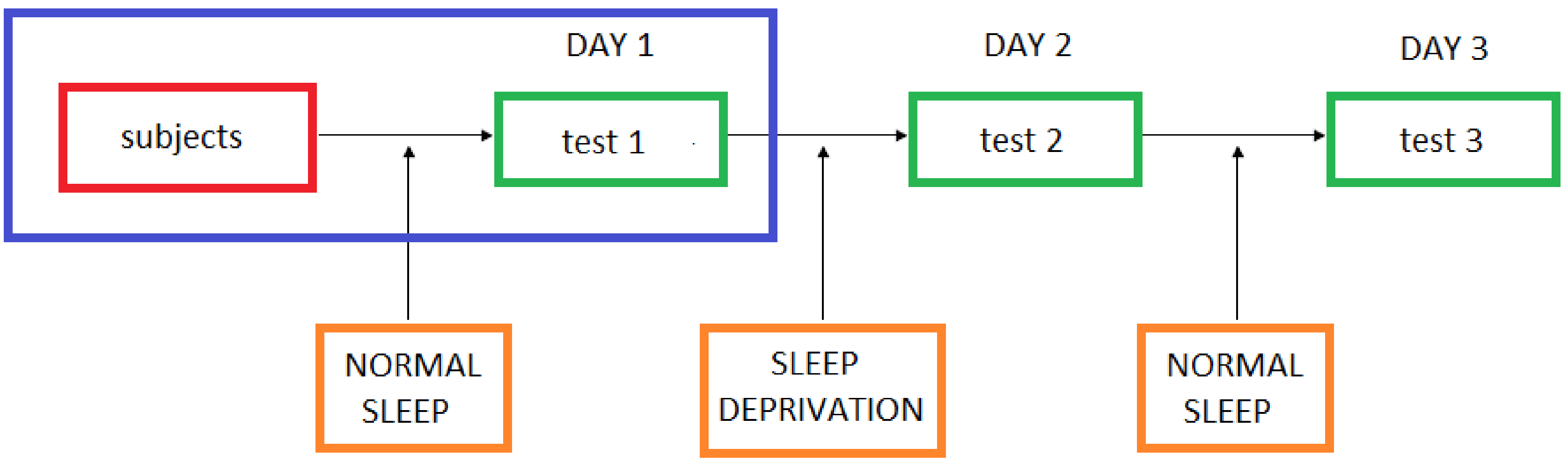
- One-sample t-test
- Independent samples t-test
- Repeated measures t-test

- Assumptions – check for these before starting:
  - Normality
  - Sample size > 20
  - Continuous variables

Our data today…



4. t-tests: one-sample t-test



Example:

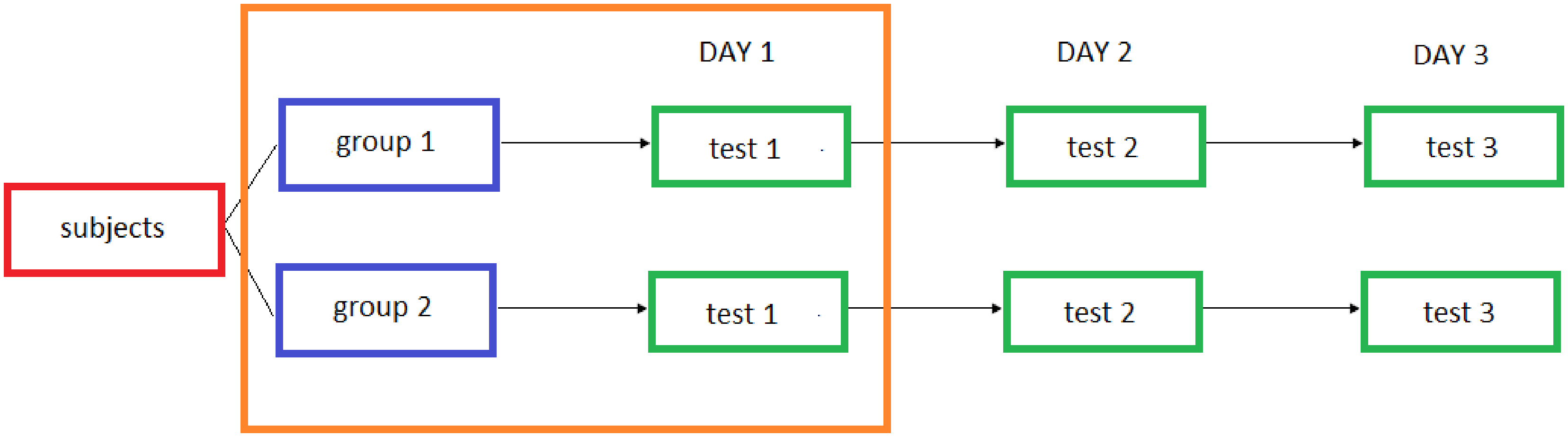
Does our sample mean in baseline Digit symbol task differ from the known population mean (=40)?

4. t-tests: one-sample t-test

```
> t.test(digitsymbol_1, mu=40)

        One Sample t-test

data:  digitsymbol_1
t = 1.3304, df = 46, p-value = 0.1899
alternative hypothesis: true mean is not equal to 40
95 percent confidence interval:
 38.85393 45.61415
sample estimates:
mean of x
 42.23404
```

No differences between our sample mean and the population mean ($t(46) = 1.33$, $p > .05$).

4. t-tests: independent samples t-test



Example:

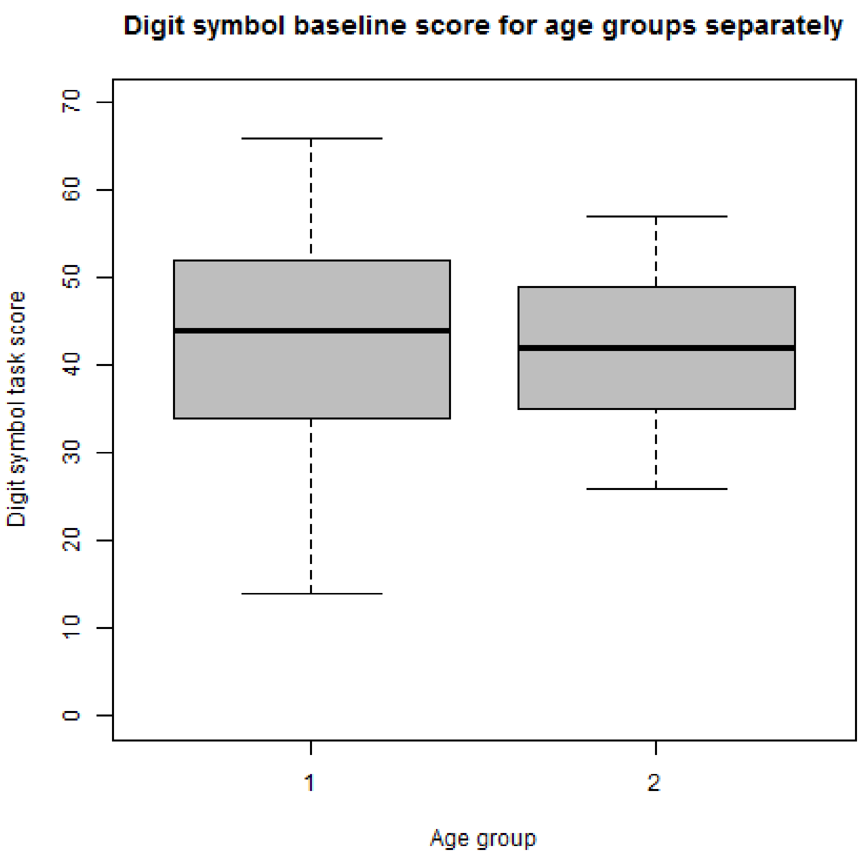
Do the baseline Digit task scores differ between younger and older participants?

4. t-tests: independent samples t-test

```
> t.test(digitsymbol_1 ~ age)

        Welch Two Sample t-test

data:  digitsymbol_1 by age
t = 0.2655, df = 35.118, p-value = 0.7922
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.161346  8.015891
sample estimates:
mean in group 1 mean in group 2 
       42.72727        41.80000 
```

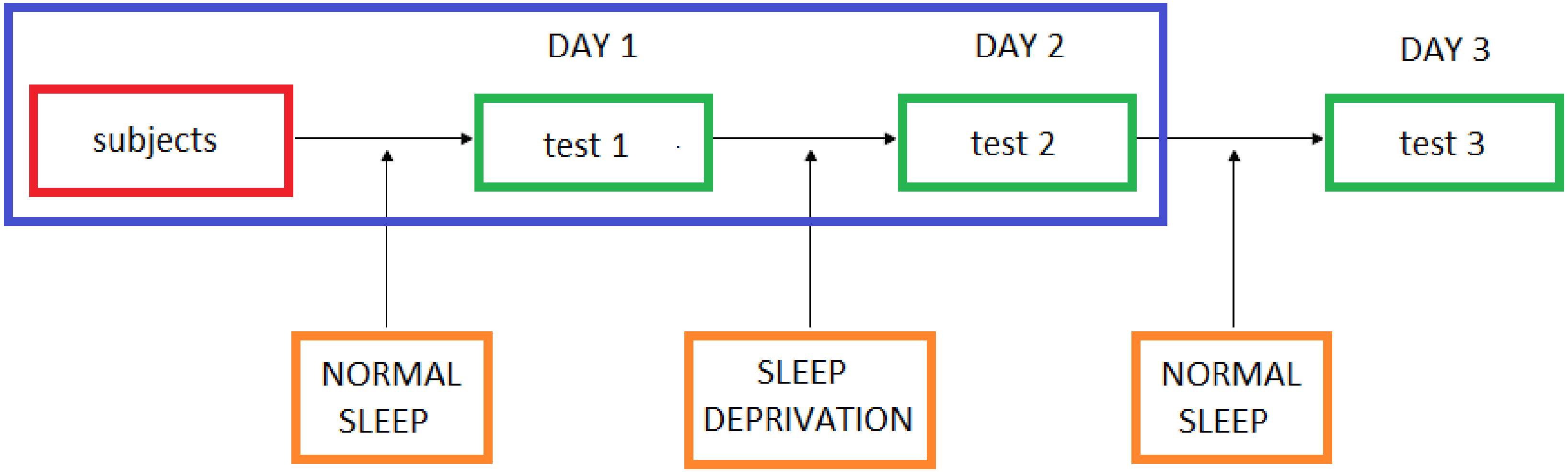
No differences between age groups in Digit symbol task scores ($t(35) = 0.27$, $p > .05$).

4. t-tests: independent samples t-test




```
> boxplot(digitsymbol_1 ~ age,

    xlab="Age group",
    ylab="Digit symbol task score",
    main="Digit symbol baseline
    score for age groups
    separately,
    col="grey",
    ylim=c(0,70))
```

4. t-tests: repeated measures t-test



Example:

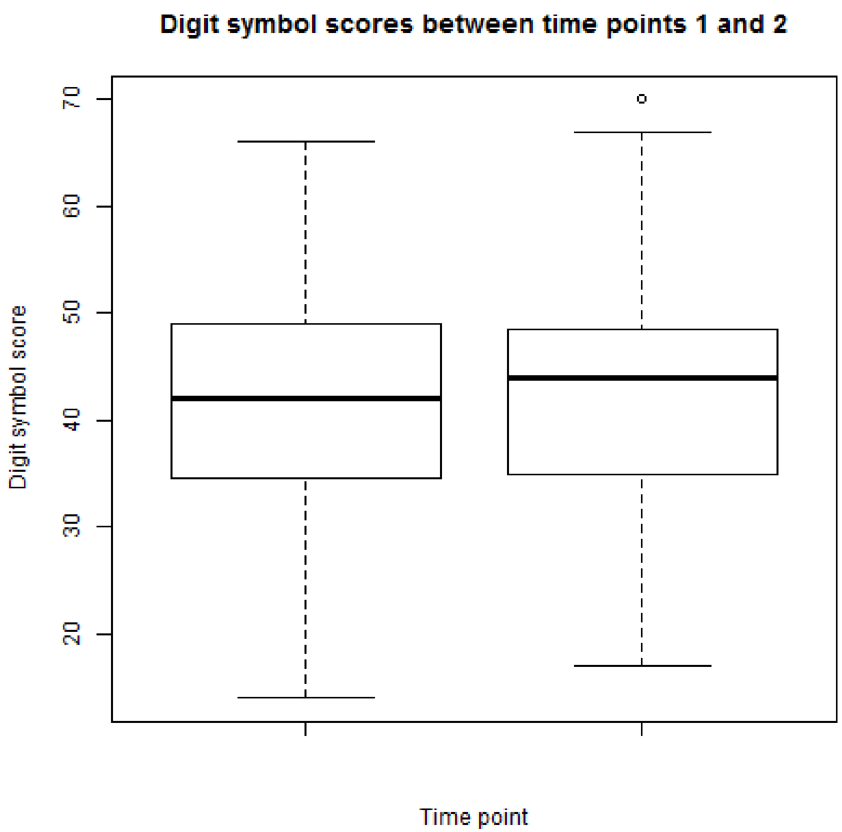
Does the Digit score task performance differ between baseline and when tested after sleep deprivation?

4. t-tests: repeated measures t-test

```
> t.test(digitsymbol_1, digitsymbol_2, paired=T)

        Paired t-test

data:  digitsymbol_1 and digitsymbol_2
t = -0.786, df = 46, p-value = 0.4359
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.424453  1.062751
sample estimates:
mean of the differences
             -0.6808511
```

No differences in Digit symbol task scores between baseline and when tested after sleep deprivation ($t(36) = -0.79$, $p > .05$).

4. t-tests: repeated measures t-test

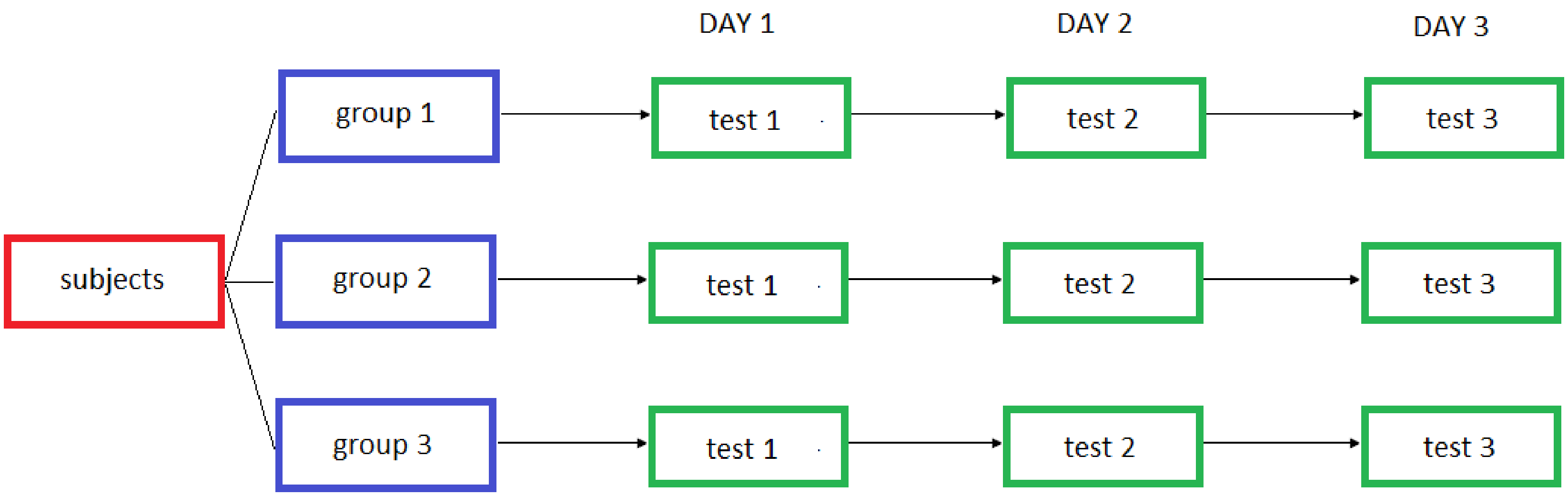



```
> boxplot(digitsymbol_1,
   digitsymbol_2,
   xlab="Time point",
   ylab="Digit symbol score",
   main="Digit symbol scores
   between time points 1 and 2")
```

But…

What if we have more than three groups?

What if we want to take into account multiple categorizations at the same time?



5. ANOVA

1. ANOVA assumptions

2. Different types of ANOVA
   - One-way between subject
   - Two-way between subject
   - One-way within subjects
   - (Two-way within subjects)

3. Post-hoc comparisons

5. ANOVA: assumptions

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

5. ANOVA: building your model

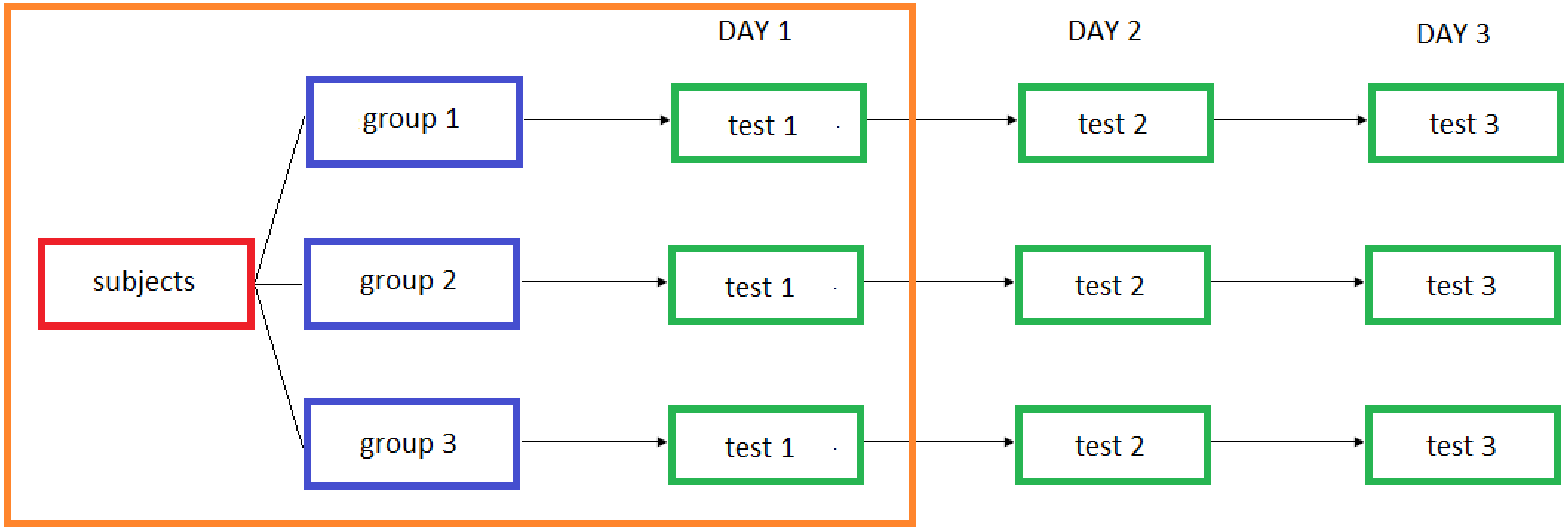
Note the notation in R:

```
y ~ x                                # one main effect
y ~ x1 + x2                          # two main effects
y ~ x1 * x2 (= y ~ x1 + x2 + x1*x2) # two main effects, interaction
                                       effect
```

where y = dependent variable

x1, x2 = independent variables

5. ANOVA: one-way between subjects



Example:

Do baseline Digit symbol task scores differ between education groups?

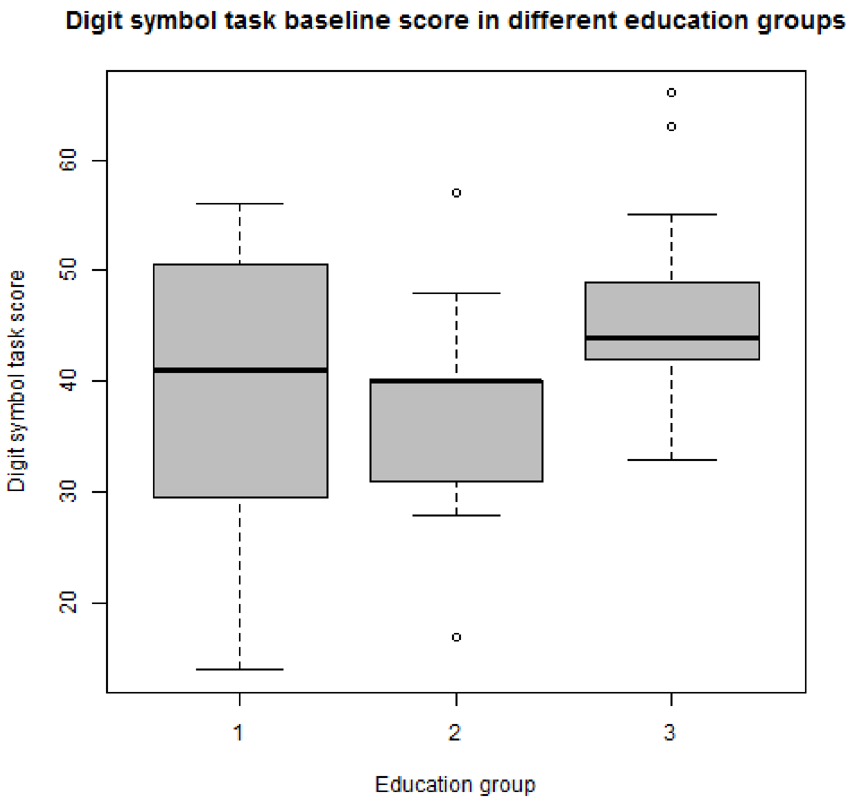
5. ANOVA: one-way between subjects

```
> A1 <- aov(digitsymbol_1 ~ education)
> summary(A1)
            Df Sum Sq Mean Sq F value Pr(>F)
education    2    544   271.9   2.155  0.128
Residuals   44   5553   126.2
> model.tables(A1, "means")
Tables of means
Grand mean

42.23404

 education
        1     2    3
    40.19 38.85 46.5
rep 16.00 13.00 18.0
```

```
aov(x ~ y)
summary(model)
model.tables
```

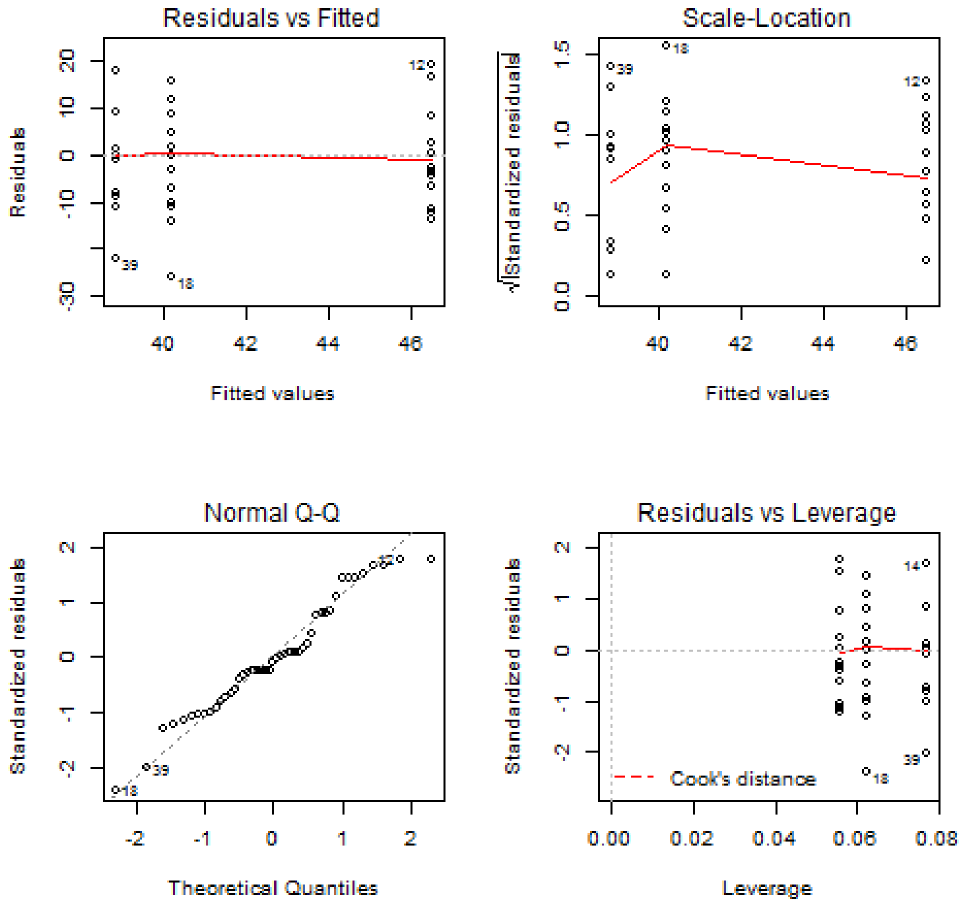
No differences between education groups in Digit symbol task scores ($F(2,44) = 2.16$, $p > .05$).

5. ANOVA: one-way between subjects




```
> boxplot(digitsymbol_1 ~ education,
   xlab="Education group",
   ylab="Digit symbol task score",
   main="Digit symbol task baseline
   score in different education
   groups",
   col="grey")
```

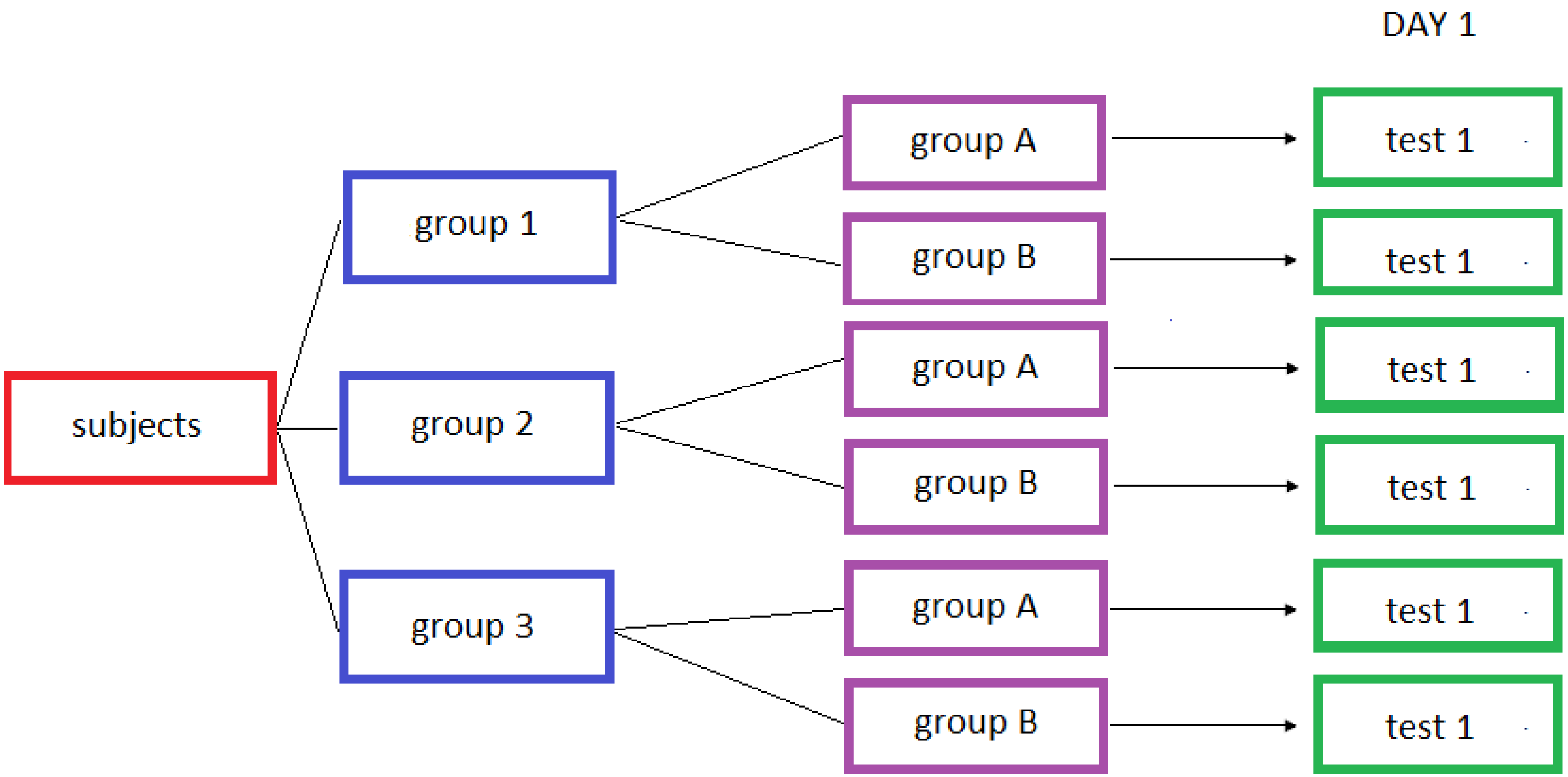
5. ANOVA: one-way between subjects



Diagnostic plots:

```
> layout(matrix(c(1,2,3,4),2,2))
> plot(A1)
```

5. ANOVA: two-way between subjects



Example:

Do education and age effect the baseline Digit symbol task scores?

5. ANOVA: two-way between subjects

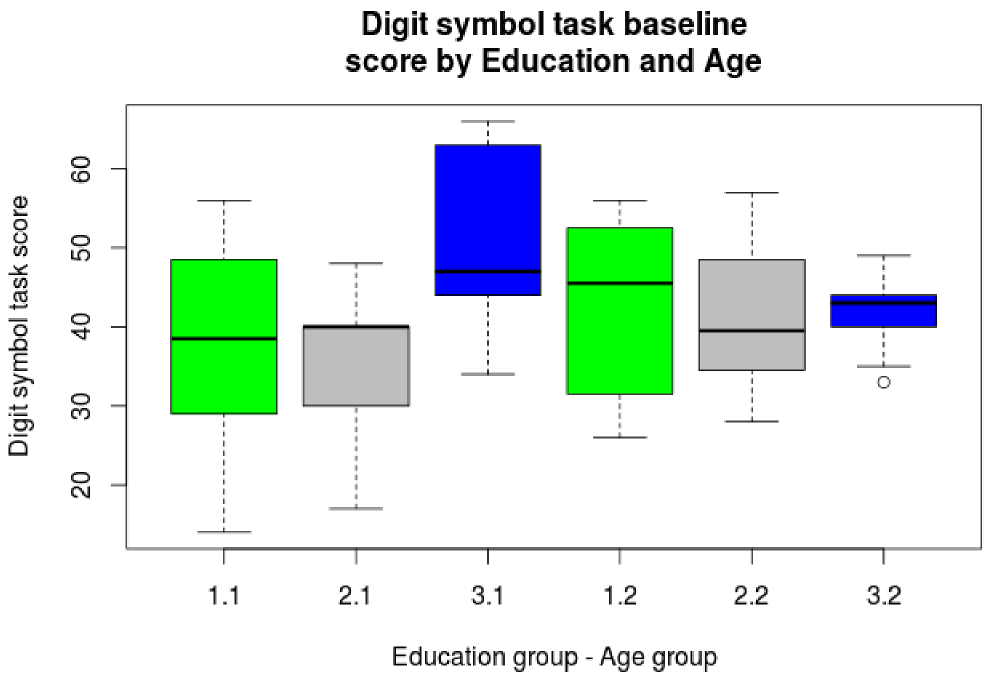
```
> A2 <- aov(digitsymbol_1 ~education * age)
> summary(A2)
              Df Sum Sq Mean Sq F value Pr(>F)
education      2    544   271.9   2.276  0.115
age            1      3     2.9   0.024  0.877
education:age  2    652   326.2   2.731  0.077 .
Residuals     41   4897   119.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

42.23404

 education
        1     2    3
    40.19 38.85 46.5
rep 16.00 13.00 18.0

 age
       1  2
    42.5 42
rep 22.0 25

 education:age
           age
education 1     2
        1   37.75 42.62
        rep  8.00  8.00
        2   35.00 41.25
        rep  5.00  8.00
        3   51.44 41.56
        rep  9.00  9.00
```

```
aov(y ~ x1 * x2)
summary(model)
model.tables
```

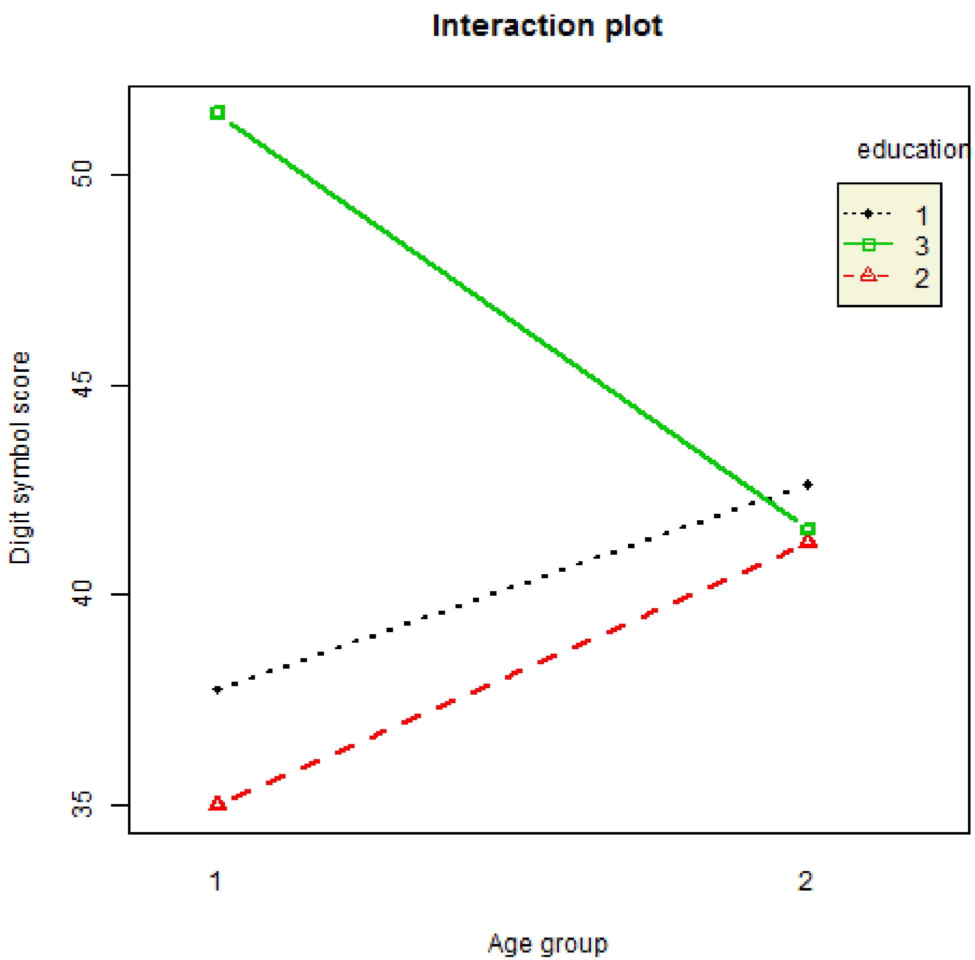
No differences between education groups in Digit symbol task scores (F(2,44) = 2.16, p > .05).

5. ANOVA: two-way between subjects



- boxplot(digitsymbol_1 ~ education * age, xlab="Education group - Age group", ylab="Digit symbol task score", main="Digit symbol task baseline score by Education and Age", col=c("green", "grey“, “blue”))

5. ANOVA: two-way between subjects




```
interaction.plot(age,
  education, digitsymbol_1,
  type="b", col=c(1:3),
  leg.bty="o",leg.bg="beige",
  lwd=2, pch=c(18,24,22),
  xlab="Age group",
  ylab="Digit symbol score",
  main="Interaction plot")
```

5. ANOVA: two-way between subjects

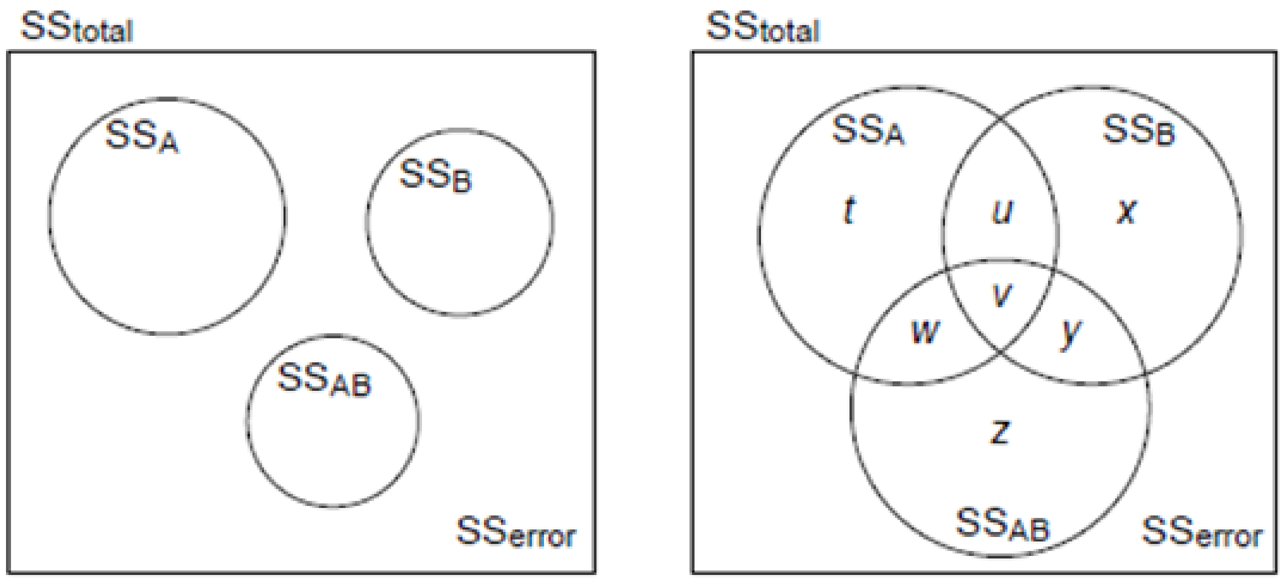
- Note: Type of sum of squares!!
  - the order in which independent variables are fed
  - sequantial, hierarchical, or orthogonal
- R (`aov`) uses Type I by default
- e.g. SPSS uses Type III by default

5. ANOVA: two-way between subjects

Type of sum of squares:

Type III: orthogonal

Type I: nonorthogonal



5. ANOVA: two-way between subjects

Type of sum of squares: how to compare different types

- take type I SS with `aov` first
- check other types with `Anova()`

5. ANOVA: post-hoc comparisons

I get significant ANOVA, but what causes the difference between groups?

E.g., after the simple one-way ANOVA:

```
Anova1 <- aov(digitsymbol_1 ~ education)

pairwise.t.test(digitsymbol_1, education, p.adj=”none”)

pairwise.t.test(digitsymbol_1, education, p.adj=”bonf”)

pairwise.t.test(digitsymbol_1, education, p.adj=”holm”)

TukeyHSD(Anova1)
```

5. ANOVA: post-hoc comparisons

BUT…

- Better than post-hoc tests are the planned comparisons, i.e. contrasts!
- We will learn how to do this in R when using regression next week...

Other things regarding exercises…

Data can be presented either in a wide or long format: function `reshape` modifies the format

wide format:

ID	age	edu	ht	ds1	ds2	ds3	bt1	bt2	bt3
1	1	1	user	35	36	39	10	11	7
2	2	2	control	45	42	47	5	5	4
3	1	3	control	40	39	49	6	8	3

Other things regarding exercises…

Data can be presented either in a wide or long format: function `reshape` modifies the format

wide format:

ID	age	edu	ht	ds1	ds2	ds3	bt1	bt2	bt3
1	1	1	user	35	36	39	10	11	7
2	2	2	control	45	42	47	5	5	4
3	1	3	control	40	39	49	6	8	3

Other things regarding exercises…

long format:

(remember 'naming' dataset?)

ID	age	edu	ht	time	ds_score	bt_score
1	1	1	user	1	35	10
1	1	1	user	2	36	11
1	1	1	user	3	39	7
2	2	2	control	1	45	5
2	2	2	control	2	42	5
2	2	2	control	3	47	4
3	1	3	control	1	40	6
3	1	3	control	2	39	8
3	1	3	control	3	49	3

For exercises and homework

Remember to report your results properly:

- Not just R code
- But written report and interpretation of the results
- Should include the standard reporting of values (see lecture slides)