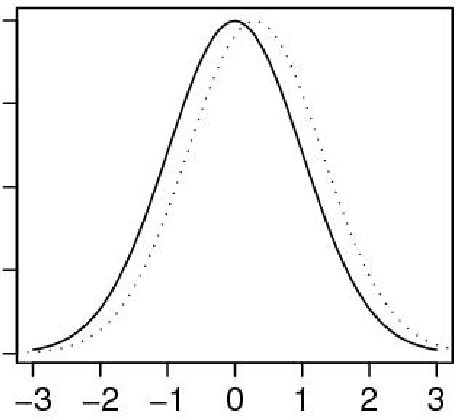
3. Using SPSS for t-tests and ANOVA

Karl B Christensen
http://publicifsv.sund.ku.dk/~kach/SPSS

Comparing two samples

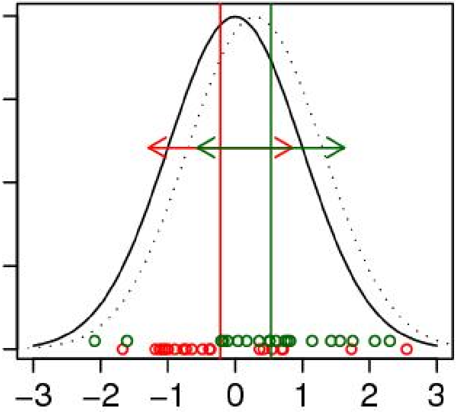
- Two groups: data $x_{11}, \ldots, x_{1n_1}$ and $x_{21}, \ldots, x_{2n_2}$
- Theoretical distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$
- Empirical mean and variance $(\bar{x}_1, s_1^2)$ and $(\bar{x}_2, s_2^2)$
- Significant difference between $\bar{x}_1$ and $\bar{x}_2$?
- Are μ_1 and μ_2 different?
- Null hypothesis $H_0 : \mu_1 = \mu_2$

Theoretical distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$



Comparing two samples

Empirical mean and variance $(\bar{x}_1, s_1^2)$ and $(\bar{x}_2, s_2^2)$



Two-sample t-test

- Standard error of mean $SEM = s/\sqrt{n}$.
- Standard error of difference of means

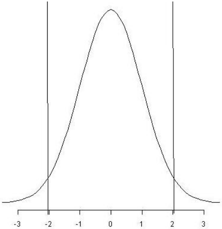
$$SEDM = \sqrt{SEM_1^2 + SEM_2^2}.$$

- T-test statistic

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\text{SEDM}}$$

measures disagreement between data and H_0

- If H_0 is true, then the distribution of t is symmetric around 0



reject if prob. of observing a more extreme value $p < 5\%$.

Equal variances?

Assume $\sigma_1^2 = \sigma_2^2$ before testing $\mu_1 = \mu_2$?

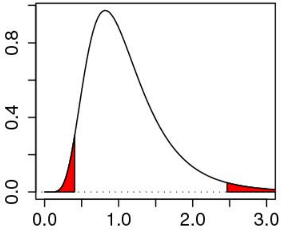
$\sigma_1^2 = \sigma_2^2$ Natural assumption under the H_0 (distributions are equal). Nice theory.

$\sigma_1^2 \neq \sigma_2^2$ Looks specifically for difference in means. Approximative theory.

Test for equal variances: Compute test statistic (Note: 2-sided test)

$$F = s_1^2 / s_2^2$$

F-distribution with (f_1, f_2) degrees of freedom, where $f_1 = n_1 - 1$ and $f_2 = n_2 - 1$



RCT data

The data set

http://publicifsv.sund.ku.dk/~kach/SPSS/RCT.sps7bdat

http://publicifsv.sund.ku.dk/~kach/SPSS/RCT.txt

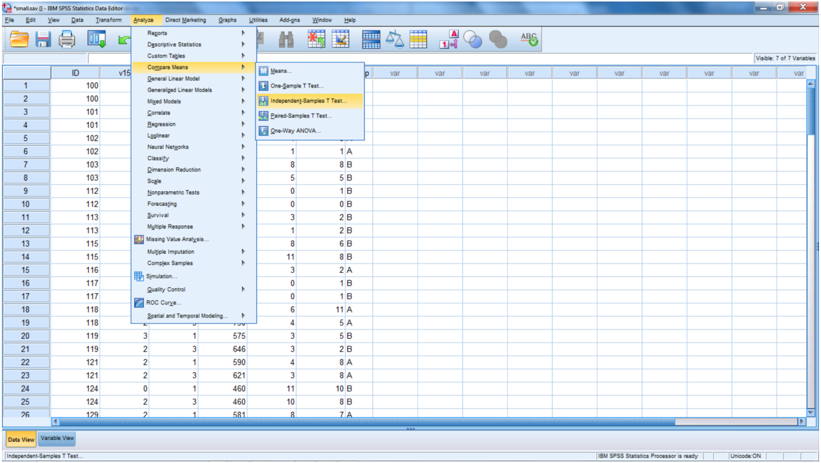
http://publicifsv.sund.ku.dk/~kach/SPSS/RCT.xlsx

contains data from an RCT where a physical exercise intervention in cancer patients was evaluated. Consider the variables

- `ID` id number
- `VO2` Aerobic capacity (VO_2max)
- `group` Intervention/control group assignment
- `time` Time (1: baseline data 3: after intervention)

⋮

t-test in SPSS



t-test in SPSS

We want to compare aerobic capacity (VO_2max) in the two groups at follow-up

```
GET FILE='P:\small.sav'.
SELECT IF (time=3).
EXECUTE.
T-TEST GROUPS=group('A' 'B')
  /MISSING=ANALYSIS
  /VARIABLES=VO2
  /CRITERIA=CI(.95).
```

The output has three parts:

(i) `Group Statistics` showing the mean, the standard error, and the standard error of the mean. Confidence limits are also included.

(ii) `T-tests` showing two t-tests (one that assumes equal variances and one that doesn't).

(iii) `test for Equality of Variances` showing a test of equal variances.

Output (i)

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
Aerobic capacity (VO2)	A	31	2,37	,596	,107
	B	29	1,51	,432	,080

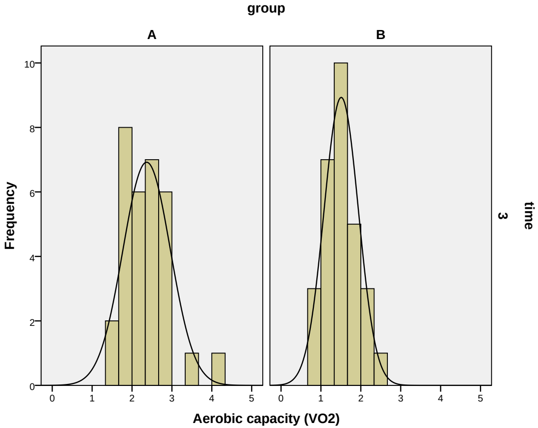
Output (ii)+(iii)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Aerobic capacity (VO2)	Equal variances assumed	1,628	,207	6,347	58	,000	,857	,135	,587	1,128
	Equal variances not assumed			6,414	54,682	,000	,857	,134	,590	1,125

Output (ii)+(iii)

The hypothesis about equal variances is not rejected (p=0.207). Difference in means is 0.857 (95% CI 0.587 to 1.128).



Exercise: t-test

1. Use graphical methods to evaluate if the distribution of VO_2max is skewed.
2. Compare the $\log(VO_2 max)$-level at follow-up in the two groups using a t-test
3. Quantify the difference. Remember confidence intervals
4. Can we interpret this difference on the original scale ?

Interpretation of difference on original scale

Example: Absolute difference in $\log(X)$:

-0.0314 (95% CI -0.1801 to 0.1173)

- t-test on X

$$\bar{x}_B - \bar{x}_A = \mu_B - \mu_A$$

- t-test on $y = \log(X)$

$$\bar{y}_B - \bar{y}_A = \log(\mu_B) - \log(\mu_A) = \log(\mu_B/\mu_A)$$

  so $\exp(\bar{y}_B - \bar{y}_A)$ is an estimate of the ratio μ_B/μ_A.

- Compute

$$\exp(-0.0314) \simeq 0.97, \exp(-0.1801) \simeq 0.84 \text{ and } \exp(0.1173) \simeq 1.12$$

  and interpret this as a relative difference in $SIGF1$ of

-3% (95% CI -16% to +12%).

Beyond the t-test

- The t-test compares two groups based on an assumption of normality, but what if data are not normally distributed or if we want to compare three or more groups?
- The t-test is robust - because means tend to be normally distributed, sometimes transformation ($x \mapsto \sqrt{x}$ or $\log(x)$) can help. Otherwise nonparametric methods.
- Compare more than three groups using analysis of variance (ANOVA).

One-way ANOVA

Comparing more than two groups

$$\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k \qquad s_1, s_2, \ldots, s_k$$

Joint test for any differences between the groups.

Why not just pairwise t-tests?

- Mass significance (type I error)
- Loss of overview

The fewer tests, the better.

Notation in ANOVA model

x_{ij} observation no. j in group no. i, e.g. x_{35} the 5th observation in group 3. Model

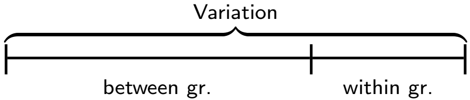
$$X_{ij} = \mu_i + \epsilon_{ij} \qquad \epsilon_{ij} \sim N(0, \sigma^2)$$

The null hypothesis (no differences between groups)

$$\mu_1 = \mu_2 = ... = \mu_k$$

Variation within groups and variation between groups

- Main idea behind analysis of variance (ANOVA): If the variation between group means is large compared to the variation within groups, it is a sign that the null hypothesis is wrong.
- The model (grouping) *explains* part of the variation



Let $\bar{x}_i$ denote the mean for group i and let $\bar{x}_.$ denote the total (grand) mean

Variation **W**ithin groups:

$$\text{SSD}_W = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2$$

Variation **B**etween groups:

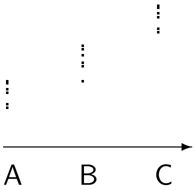
$$\text{SSD}_B = \sum_i \sum_j (\bar{x}_i - \bar{x}_.)^2$$

Can be mathematically proven that

$$\text{SSD}_B + \text{SSD}_W = \text{SSD}_{\text{total}} = \sum_i \sum_j (x_{ij} - \bar{x}_.)^2$$

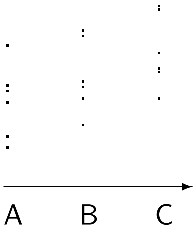
Var. between groups large compared to var. within groups

Small variation within groups



high 'between' variation
small 'within' variation
F is large
H_0 is rejected

Large variation within groups



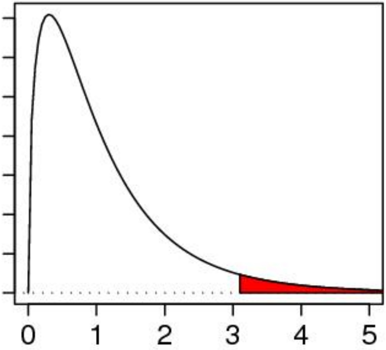
high 'within' variation
small 'between' variation
F is small
H_0 is *not* rejected

F-test for identical group means

Reject the hypothesis if the variation between groups is large compared to the variation within groups.

$$F = [\mathrm{SSD}_B/(k-1)]/[\mathrm{SSD}_W/(N-k)]$$

If null hypothesis is true we know distribution of F



Reject hypothesis that group means are identical if F too large.

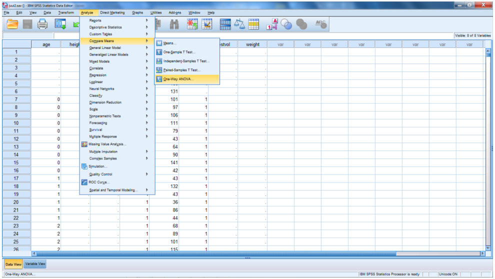
One-way ANOVA in SPSS

Data set `juul2.sav`. Compare boys in different Tanner stage with respect to their log SIGF1

1. Generate a new data set
2. Select (sexnr=1, age<20)
3. model: What is described by what? (`sigf1` by `tanner`)
4. SPSS knows that tanner is a grouping

ANOVA

Point-and-click



syntax

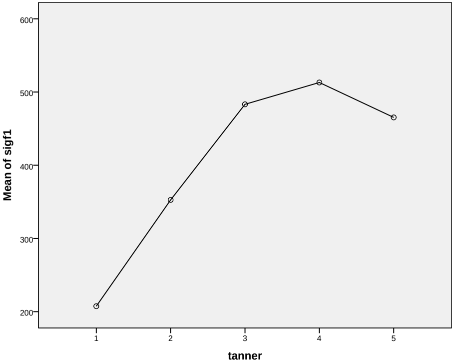
```
GET FILE='P:\juul2.sav'.

ONEWAY sigf1 BY tanner
  /PLOT MEANS
  /MISSING ANALYSIS.
```

ANOVA

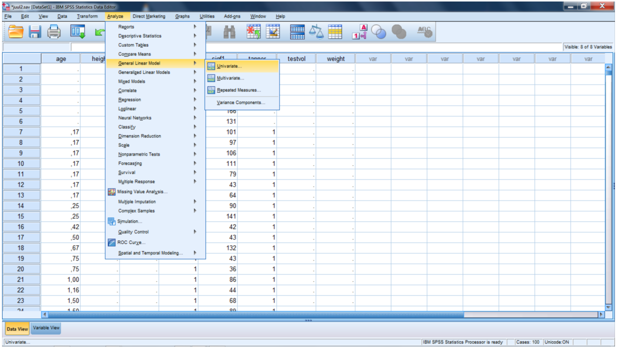
ANOVA

sigf1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12696216,59	4	3174054,147	228,353	,000
Within Groups	10939116,28	787	13899,767		
Total	23635332,86	791			



ANOVA using UNIANOVA



ANOVA exercise

Look at ANOVA in the Juul data. Try

```
UNIANOVA sigf1 BY tanner
  /CONTRAST(tanner)=Simple
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN=tanner.
```

This yields contrasts. Try to interpret these. If you have time do the analysis on the log-transformed SIGF1.

Mann-Whitney test, Wilcoxon test, Kruskal-Wallis test

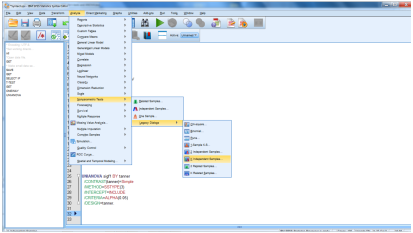
Nonparametric statistics: 't-test' (Mann-Whitney test, Wilcoxon test) or 'ANOVA' (Kruskal-Wallis test) on *ranks*

$$\underset{\text{Data}}{\boxed{\begin{array}{c} 3, 9, 17, 50 \\ 4, 7, 8.7, 200 \end{array}}} \longrightarrow \underset{\text{Rank}}{\boxed{\begin{array}{c} 1, 5, 6, 7 \\ 2, 3, 4, 8 \end{array}}}$$

Distribution is (in principle) known under null hypothesis. Does not depend on data following a normal distribution. Other "scores" than ranks can also be used

Non parametric tests in SPSS

Use



or

```
NPAR TESTS
  /K-W=sigf1 BY tanner(1 5)
  /STATISTICS QUARTILES
  /MISSING ANALYSIS.
```

Non parametric tests in SPSS

NPar Tests

Descriptive Statistics

	N	Percentiles		
		25th	50th (Median)	75th
sigf1	1018	202,00	313,50	463,25
tanner	1099	1,00	2,00	5,00

Kruskal-Wallis Test

Ranks

	tanner	N	Mean Rank
sigf1	1	311	186,77
	2	70	398,32
	3	45	557,81
	4	58	609,28
	5	308	544,22
	Total	792	

Test Statistics[a,b]

	sigf1
Chi-Square	462,346
df	4
Asymp. Sig.	,000

a. Kruskal Wallis Test

b. Grouping Variable: tanner

More non parametric tests. Sign test

Paired data where patients rate two drugs A and B

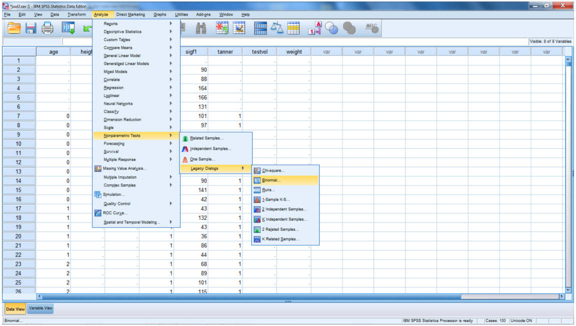
$$M_i = \begin{cases} 1, & \text{if } A_i > B_i \\ 0, & \text{if } B_i < A_i \end{cases}$$

for $i = 1, \ldots, 20$. Under the null hypothesis

$$H_0 : P(A_i > B_i) = \frac{1}{2}$$

the test statistic $M = \sum_{i=1}^{20} M_i$ is binomially distributed

More non parametric tests. Sign test



More non parametric tests

- Friedmans test.
- Jonckheere-Terpstra test

Exercise: RCT data

1. Compare baseline aerobic capacity across the two groups using histograms, boxplots and descriptive statistics.
2. Test the null hypothesis that baseline aerobic capacity does not differ across the two groups. Use t-test, t-test on log-transformed VO_2max and non parametric statistics.
3. Discuss how change scores

   $$(VO_2\text{max after intervention}) - (\text{baseline } VO_2\text{max})$$

   could be evaluated.