3. Using SPSS for t-tests and ANOVA

Karl B Christensen
http://biostat.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 $\mu_1$ and $\mu_2$ different?
- Null hypothesis $H_0 : \mu_1 = \mu_2$
Comparing two samples

Theoretical distributions $N(\mu_1, \sigma^2_1)$ 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
  \[ t = \frac{\bar{x}_2 - \bar{x}_1}{SEDM} \]
  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 $p < 5\%$. 

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

3. Using SPSS for $t$-tests and ANOVA
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)

$$F = \frac{s_1^2}{s_2^2}$$

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

http://biostat.ku.dk/~kach/SPSS/RCT.sps7bdat
http://biostat.ku.dk/~kach/SPSS/RCT.txt
http://biostat.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 ($VO_2$max)

**group** Intervention/control group assignment

**time** Time (1: baseline data 3: after intervention)
3. Using SPSS for t-tests and ANOVA
We want to compare aerobic capacity (VO$_{2\text{max}}$) 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.
### Group Statistics

<table>
<thead>
<tr>
<th>group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerobic capacity (VO2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>31</td>
<td>2.37</td>
<td>0.596</td>
<td>0.107</td>
</tr>
<tr>
<td>B</td>
<td>29</td>
<td>1.51</td>
<td>0.432</td>
<td>0.080</td>
</tr>
</tbody>
</table>
### Independent Samples Test

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Sig.</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
</tr>
<tr>
<td>Aerobic capacity (VO2)</td>
<td></td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>1,628</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>6,414</td>
</tr>
</tbody>
</table>
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 $VO_2\text{max}$ is skewed.
2. Compare the log($VO_2\text{max}$)-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?
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 $y = \log(X)$
  
  $$\bar{y}_B - \bar{y}_A = \log(\mu_B) - \log(\mu_A) = \log(\frac{\mu_B}{\mu_A})$$

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

- Compute
  
  $$\exp(-0.0314) \simeq 0.97, \exp(-0.1801) \simeq 0.84 \ and \ \exp(0.1173) \simeq 1.12$$

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).
Comparing more than two groups

\[ \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k \quad s_1, s_2, \ldots, s_k \]

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

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

The null hypothesis (no differences between groups)

$$\mu_1 = \mu_2 = \ldots = \mu_k$$
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

\[ \text{Variation} = \text{between gr.} + \text{within gr.} \]
Sums of squares

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

Variation Within groups:

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

Variation Between groups:

$$SSD_B = \sum_i \sum_j (\bar{x}_i - \bar{x})^2$$

Can be mathematically proven that

$$SSD_B + SSD_W = SSD_{total} = \sum_i \sum_j (x_{ij} - \bar{x})^2$$
Var. between groups large compared to var. within groups

Small variation within groups

Large variation within groups

A  B  C

high 'between' variation
small 'within' variation

F is large

H₀ is rejected

A  B  C

high 'within' variation
small 'between' variation

F is small

H₀ 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.

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

If null hypothesis is true we know distribution of \( F \)

Reject hypothesis that group means are identical if \( F \) too large.
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
Point-and-click

Syntax

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

ONESWAY sigf1 BY tanner
   /PLOT MEANS
   /MISSING ANALYSIS.
### ANOVA

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>12696216.59</td>
<td>4</td>
<td>3174054.147</td>
<td>228.353</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>10939116.28</td>
<td>787</td>
<td>13899.767</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23635332.86</td>
<td>791</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Mean of sigf1

<table>
<thead>
<tr>
<th>Tanner</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>300</td>
</tr>
<tr>
<td>3</td>
<td>400</td>
</tr>
<tr>
<td>4</td>
<td>500</td>
</tr>
<tr>
<td>5</td>
<td>600</td>
</tr>
</tbody>
</table>
3. Using SPSS for t-tests and ANOVA
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.
Nonparametric statistics: 't-test' (Mann-Whitney test, Wilcoxon test) or 'ANOVA' (Kruskal-Wallis test) on ranks

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

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

or

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

Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>25th</th>
<th>50th (Median)</th>
<th>75th</th>
</tr>
</thead>
<tbody>
<tr>
<td>sigf1</td>
<td>1018</td>
<td>202</td>
<td>313.50</td>
<td>463.25</td>
</tr>
<tr>
<td>tanner</td>
<td>1099</td>
<td>1.00</td>
<td>2.00</td>
<td>5.00</td>
</tr>
</tbody>
</table>

Kruskal-Wallis Test

Ranks

<table>
<thead>
<tr>
<th>tanner</th>
<th>N</th>
<th>Mean Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>sigf1</td>
<td>1</td>
<td>311</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>308</td>
</tr>
<tr>
<td>Total</td>
<td>792</td>
<td></td>
</tr>
</tbody>
</table>

Test Statisticsa,b

<table>
<thead>
<tr>
<th></th>
<th>sigf1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>462.346</td>
</tr>
<tr>
<td>df</td>
<td>4</td>
</tr>
<tr>
<td>Asymp. Sig.</td>
<td>.000</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>age</th>
<th>height</th>
<th>sigc</th>
<th>tanner</th>
<th>textbook</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90</td>
<td>0</td>
<td>1</td>
<td>101</td>
<td>97</td>
</tr>
<tr>
<td>2</td>
<td>88</td>
<td>1</td>
<td>1</td>
<td>131</td>
<td>166</td>
</tr>
<tr>
<td>3</td>
<td>164</td>
<td>1</td>
<td>1</td>
<td>131</td>
<td>166</td>
</tr>
<tr>
<td>4</td>
<td>101</td>
<td>1</td>
<td>1</td>
<td>131</td>
<td>166</td>
</tr>
<tr>
<td>5</td>
<td>97</td>
<td>1</td>
<td>1</td>
<td>131</td>
<td>166</td>
</tr>
</tbody>
</table>

Using SPSS for t-tests and ANOVA
More non parametric tests

- Friedmans test.
- Jonckheere-Terpstra test
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$_2$max and non-parametric statistics.

3. Discuss how change scores
   \[(\text{VO}_2\text{max after intervention}) - (\text{baseline VO}_2\text{max})\]
   could be evaluated.