Transformations. Non-parametric tests

Jon Michael Gran
Department of Biostatistics, UiO

MF9130 Introductory course in statistics
Monday 06.06.2011
Overview

Transformations. Non-parametric methods
(Aalen chapter 8.8, Kirkwood and Sterne chapter 13 and 30.2)

- Logarithmic transformation
- Choice of transformation
- Non-parametric methods based on ranks (Wilcoxon signed rank test and Wilcoxon rank sum test)
Transformations

Motivation

• Non-normality means that the standard methods cannot be used
• This includes linear regression, which is often a main tool of analysis → major problem
• For data that are skewed to the right, one can often use log-transformations
• This does not only give normally distributed data; it may also give equal variances in different groups
Logarithmic transformation

- So what do we mean by doing a log transformation?
- Take the logarithm of all your data values $x_i$ for all $i$’s, and do your analysis on the new dataset of $u_i$’s, where $u_i = \ln(x_i)$
Example: Ln transformation

- Skew distribution. Example of observations: 0.40, 0.96, 11.0
Example: Ln transformation

- Skew distribution. Example of observations: 0.40, 0.96, 11.0
- Ln transformed distribution: $\ln(0.4) = -0.91$, $\ln(0.96) = -0.04$, $\ln(11) = 2.40$
Example: Ln transformation

Skew distribution. Example of observations: 0.40, 0.96, 11.0

Ln transformed distribution: \( \ln(0.4) = -0.91 \), \( \ln(0.96) = -0.04 \), \( \ln(11) = 2.40 \)

Do analysis on the ln transformed data

In SPSS: transform \( \rightarrow \) compute
Choice of transformations

- The logarithmic transformation is by far the most frequently applied
- Appropriate for removing positive skewness (data skew to the right)
- There are other types of transformation for data that are stronger or weaker skewed, or data skew to the left
Other transformations

- **Skewed to the right:**
  - Lognormal: Logarithmic ($u = \ln x$)
  - More skewed than lognormal: Reciprocal ($u = 1/x$)
  - Less skewed than lognormal: Square root ($u = \sqrt{x}$)

- **Skew to the left:**
  - Moderately skewed: Square ($u = x^2$)
  - More skewed: Cubic ($u = x^3$)

- **Non-linear relationship**: Transform only one of the two variables
Non-parametric statistics

Motivation

- So what if transforming our data doesn’t help and it is still not normally distributed?
- Use non-parametric test
Non-parametric tests

• In tests we have done so far, the null hypothesis has always been a stochastic model with a few parameters. Models based on for example:
  ▶ Normal distribution
  ▶ T-distribution
  ▶ Binomial distribution

• In nonparametric tests, the null hypothesis is not a parametric distribution, rather a much larger class of possible distributions
Parametric methods we have seen

- **Estimation**
  - Confidence interval for $\mu$
  - Confidence interval for $\mu_1 - \mu_2$

- **Testing**
  - One sample T-test
  - Two sample T-test

- The methods are based on the assumption of normally distributed data (or normally distributed mean)
Typical assumptions for parametric methods

1. **Independence**: All observations are independent. Achieved by taking random samples of individuals; for paired t-test independence is achieved by using the difference between measurements.

2. **Normally distributed data** (Check: histograms, tests for normal distribution, Q-Q plots)

3. **Equal variance or standard deviations in the groups**
How to test for normality

- **Visual methods**, like histograms and Q-Q plots, are very useful.
- Also have several **statistical tests** for normality:
  - Kolmogorov-Smirnov test
  - Shapiro-Wilk
- However, with enough data, visual methods are more useful!
Tests for normality

- Example of SPSS output from a **Kolmogorov-Smirnov test**:

<table>
<thead>
<tr>
<th>GENDER</th>
<th>Kolmogorov-Smirnov&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Statistic</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEFSFMTM 1.00</td>
<td>.105</td>
<td>53</td>
<td>.200*</td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td>.067</td>
<td>52</td>
<td>.200*</td>
<td></td>
</tr>
</tbody>
</table>

* This is a lower bound of the true significance.

<sup>a</sup> Lilliefors Significance Correction

(measurements of lung function, separated by gender)

- According to this test, if the p-value is less than 0.05 the data cannot be considered as Normally distributed
Tests for normality

- Example of SPSS output from a **Kolmogorov-Smirnov test**: (measurements of lung function, separated by gender)

<table>
<thead>
<tr>
<th>GENDER</th>
<th>Kolmogorov-Smirnova</th>
<th>Statistic</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEFSITM 1.00</td>
<td>.105</td>
<td>53</td>
<td>.200*</td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td>.067</td>
<td>52</td>
<td>.200*</td>
<td></td>
</tr>
</tbody>
</table>

* This is a lower bound of the true significance.

a. Lilliefors Significance Correction

- According to this test, if the p-value is less than 0.05 the data cannot be considered as Normally distributed

- Note though that **these tests are a bit problematic**. They’re not very effective at discovering departures from normality. The power is low, so even if you don’t get a significant result it’s a bit of a leap to assume normality
Q-Q plots

- Graphical way of comparing two distributions, plotting their quantiles against each other
- Q for quantile
- If the distributions are similar, they Q-Q plot should be close to a straight line
Histogram and Q-Q plot for Gender = 1

**Histogram**

For GENDER= 1.00

Frequency

PEFSITM

**Normal Q-Q Plot of PEFSITM**

For GENDER= 1.00

Expected Normal

Observed Value
Histogram and Q-Q plot for Gender = 2

**Histogram**

For GENDER 2.00

**Normal Q-Q Plot of PEFSITM**

For GENDER 2.00

Std Dev = 86.31
Mean = 59.7
N = 52.00

PEFSITM

Observed Value
Non-parametric statistics

• The null hypothesis is for example that the median of the distribution is zero
• A test statistic can be formulated, so that
  ▶ it has a known distribution under this hypothesis
  ▶ it has more extreme values under alternative hypotheses
Non-parametric methods

• Estimation
  ▶ Confidence interval for the median

• Testing
  ▶ Paired data: Sign test and Wilcoxon signed rank test
  ▶ Two independent samples: Mann-Whitney test/Wilcoxon rank-sum test

• Make (almost) no assumptions regarding distribution
Example: Confidence interval for the median

- Beta-endorphin concentrations (pmol/l) in 11 individuals who collapsed during a half-marathon (in increasing order): 66.0 71.2 83.0 83.6 101.0 107.6 122.0 143.0 160.0 177.0 414.0
- We find that the median is 107.6
- What is the 95% confidence interval?
Confidence interval for the median in SPSS

- Use ratio statistics, which is meant for the ratio between two variables. Make a variable that has value 1 for all data (call it unit)
- **Analyze → Descriptive statistics → Ratios**
  Numerator: beta Denominator: unit
- Click **Statistics** and under **Central tendency** check **Median** and **Confidence Intervals**
SPSS output

- 95% confidence interval for the median is (71.2, 177.0)
Non-parametric tests

- Most tests are **based on rank-sums**, and not the observed values.
- The sums of the rank are assumed approximately **normally distributed**, so we can use a normal approximation for the tests.
- If two or more values are equal, the tests use a mean rank.
The sign test

• Assume the null hypothesis is that the median is zero
• Given a sample from the distribution, there should be roughly the same number of positive and negative values
• More precisely, number of positive values should follow a binomial distribution with probability 0.5
• When the sample is large, the binomial distribution can be approximated with a normal distribution
• Sign test assumes independent observations, no assumptions about the distribution
• SPSS: Analyze → nonparametric tests → two related samples. Choose sign under test type
Example: Energy intake kJ

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>PREMENST</th>
<th>POSTMENS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5260.0</td>
<td>3910.0</td>
</tr>
<tr>
<td>2</td>
<td>5470.0</td>
<td>4220.0</td>
</tr>
<tr>
<td>3</td>
<td>5640.0</td>
<td>3885.0</td>
</tr>
<tr>
<td>4</td>
<td>6180.0</td>
<td>5160.0</td>
</tr>
<tr>
<td>5</td>
<td>6390.0</td>
<td>5645.0</td>
</tr>
<tr>
<td>6</td>
<td>6515.0</td>
<td>4680.0</td>
</tr>
<tr>
<td>7</td>
<td>6805.0</td>
<td>5265.0</td>
</tr>
<tr>
<td>8</td>
<td>7515.0</td>
<td>5975.0</td>
</tr>
<tr>
<td>9</td>
<td>7515.0</td>
<td>6790.0</td>
</tr>
<tr>
<td>10</td>
<td>8230.0</td>
<td>6900.0</td>
</tr>
<tr>
<td>11</td>
<td>8770.0</td>
<td>7335.0</td>
</tr>
</tbody>
</table>

- Want to test if energy intake is different before and after menstruation.
- $H_0$: Median difference $= 0$
  $H_1$: Median difference $\neq 0$
- All differences are positive
Using the sign test

- The number of positive differences should follow a binomial distribution with $P = 0.5$ if $H_0$: median difference between premenst and postmenst is 0

- What is the probability of observing 11 positive differences?

- Let $P(x)$ count the number of positive signs

- $P(x)$ is Binomial($n = 11, P = 0.5$).

- The p-value of the test is the probability of observing 11 positives or something more extreme if $H_0$ is true, hence $P(x \geq 11) = P(x = 11) = \frac{11!}{0!(11-0)!} 0.5^{11} = 0.0005$.

- However, because of two-sided test, the p-value is $0.0005 \cdot 2 = 0.001$.

- Clear rejection of $H_0$ on significance-level 0.05.
Using the sign test

- The number of positive differences should follow a binomial distribution with $P = 0.5$ if $H_0$: median difference between premenst and postmenst is 0
- What is the probability of observing 11 positive differences?
Using the sign test

- The number of positive differences should follow a binomial distribution with $P = 0.5$ if $H_0$: median difference between premenst and postmenst is 0
- What is the probability of observing 11 positive differences?
- Let $P(x)$ count the number of positive signs
- $P(x)$ is $Bin(n = 11, P = 0.5)$

$p-value$ of the test is the probability of observing 11 positives or something more extreme if $H_0$ is true, hence $P(x \geq 11) = P(x = 11) = \binom{11}{0} \cdot 0.5^0 \cdot (1 - 0.5)^{11} = 0.0005$

However, because of two-sided test, the $p$-value is $0.0005 \cdot 2 = 0.001$

Clear rejection of $H_0$ on significance-level 0.05
Using the sign test

• The number of positive differences should follow a binomial distribution with \( P = 0.5 \) if \( H_0 \): median difference between premenst and postmenst is 0
• What is the probability of observing 11 positive differences?
• Let \( P(x) \) count the number of positive signs
• \( P(x) \) is \( Bin(n = 11, P = 0.5) \)
• The p-value of the test is the probability of observing 11 positives or something more extreme if \( H_0 \) is true, hence
\[
P(x \geq 11) = P(x = 11) = \frac{11!}{0!(11-0)!} \cdot 0.5^0 (1 - 0.5)^{11} = 0.0005
\]
Using the sign test

- The number of positive differences should follow a binomial distribution with \( P = 0.5 \) if \( H_0 \): median difference between premenst and postmenst is 0
- What is the probability of observing 11 positive differences?
- Let \( P(x) \) count the number of positive signs
- \( P(x) \) is \( Bin(n = 11, P = 0.5) \)
- The p-value of the test is the probability of observing 11 positives or something more extreme if \( H_0 \) is true, hence \( P(x \geq 11) = P(x = 11) = \frac{11!}{0!(11-0)!}0.5^0(1 - 0.5)^{11} = 0.0005 \)
- However, because of two-sided test, the p-value is \( 0.0005 \cdot 2 = 0.001 \)
- Clear rejection of \( H_0 \) on significance-level 0.05
In SPSS

- Find \( p\)-value = 0.000 using t-test, so \( p\)-value from sign-test is a bit larger
- This will often be the case
Wilcoxon signed rank test

- Takes into account the strength of the differences, not just if they are positive or negative
- Here, the **null hypothesis is a symmetric distribution with zero median**. Do as follows:
  - Rank all values by their absolute values, from smallest to largest. Let $T^+$ be the sum of ranks of the positive values, and $T^-$ corresponding for negative values.
  - Let $T$ be the minimum of $T^+$ and $T^-$.
  - Under the null hypothesis, $T$ has a known distribution.
- For large samples, the distribution can be approximated with a normal distribution.
- Assumes independent observations and symmetrical distributions.
- SPSS: **Analyze → nonparametric tests → two related samples**. Choose **Wilcoxon** under **test type**.

## Energy intake example

<table>
<thead>
<tr>
<th>Difference</th>
<th>Rank (+)</th>
<th>Rank(-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>premenst-postmenst</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1350,00</td>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>1250,00</td>
<td>4</td>
<td>negative differences!</td>
</tr>
<tr>
<td>1755,00</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>1020,00</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>745,00</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1835,00</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>1540,00</td>
<td>8.5</td>
<td></td>
</tr>
<tr>
<td>1540,00</td>
<td>8.5</td>
<td></td>
</tr>
<tr>
<td>725,00</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1330,00</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>1435,00</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Rank sum:</td>
<td>T+:66</td>
<td>T-:0</td>
</tr>
</tbody>
</table>

- Min(66,0)=0
- Test $H_0$: Median difference =0 vs. $H_1$: Median difference≠0 with $\alpha=0.05$
- Find $T=11$ for $n=11$ and $\alpha/2=0.025$
- Reject $H_0$
- Also, see that p-value is smaller than 0.01
In SPSS

### Ranks

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>postmenst - premenst</td>
<td>11&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6.000</td>
<td>66.000</td>
</tr>
<tr>
<td>Positive Ranks</td>
<td>0&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Ties</td>
<td>0&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- postmenst < premenst
- postmenst > premenst
- postmenst = premenst

### Test Statistics

<table>
<thead>
<tr>
<th></th>
<th>postmenst - premenst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>-2.9364</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>0.003</td>
</tr>
<tr>
<td>Exact Sig. (2-tailed)</td>
<td>0.001</td>
</tr>
<tr>
<td>Exact Sig. (1-tailed)</td>
<td>0.000</td>
</tr>
<tr>
<td>Point Probability</td>
<td>0.000</td>
</tr>
</tbody>
</table>

- Based on positive ranks.
- Wilcoxon Signed Ranks Test

**Sum of ranks**

**P-value based on normal approximation**

**P-value based on Wilcoxon distribution**
Sign test and Wilcoxon signed rank test

- **Main difference**: The sign test assume nothing about the distribution, while the Wilcoxon signed rank test assumes symmetry.
- In other words: It is a little bit harder to get significant results using the Sign test, because of lesser assumptions
- Often used on **paired data**
Wilcoxon rank sum test (or the Mann-Whitney U test)

- **Not for paired data**, but rather \( n_1 \) values from group 1 and \( n_2 \) values from group 2
- Assumptions: Independence within and between groups, same distributions in both groups
- We want to test whether the values in the groups are samples from different distributions:
  - Rank all values as if they were from the same group
  - Let \( T \) be the sum of the ranks of the values from group 1
  - Under the assumption that the values come from the same distribution, the distribution of \( T \) is known
  - The expectation and variance under the null hypothesis are simple functions of \( n_1 \) and \( n_2 \)
• For large samples, we can use a normal approximation for the distribution of $T$
• The Mann-Whitney U test gives exactly the same results, but uses slightly different test statistic
• In SPSS:
  ▶ Analyze $\rightarrow$ nonparametric tests $\rightarrow$ two independent samples tests
  ▶ Test type: Mann-Whitney U
Example 1

- We have observed values
  - Group 1: 1.3, 2.1, 1.5, 4.3, 3.2
  - Group 2: 3.4, 4.9, 6.3, 7.1
- Are the groups different?
- If we assume that the values in the groups are normally distributed, we can solve this using the T-test
- Otherwise we can try the normal approximation to the Wilcoxon rank sum test
- Test $H_0$: Groups have equal medians vs. $H_1$: Groups have different medians on 5% significance level
Example 1 (cont.)

### Mann-Whitney Test

#### Ranks

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values 1,00</td>
<td>5</td>
<td>3.20</td>
<td>16.00</td>
</tr>
<tr>
<td>2,00</td>
<td>4</td>
<td>7.25</td>
<td>29.00</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>7.25</td>
<td>29.00</td>
</tr>
</tbody>
</table>

#### Test Statistics

<table>
<thead>
<tr>
<th>Value</th>
<th>Mann-Whitney U</th>
<th>Wilcoxon W</th>
<th>Z</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>Exact Sig. [2*(1-tailed Sig.)]</th>
<th>Exact Sig. (2-tailed)</th>
<th>Exact Sig. (1-tailed)</th>
<th>Point Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,000</td>
<td>16,000</td>
<td>-2.205</td>
<td>0.027</td>
<td>0.032a</td>
<td>0.032</td>
<td>0.016</td>
<td>0.008</td>
</tr>
</tbody>
</table>

- b. Grouping Variable: Group

*Based on normal approximation* based on *wilcoxon distribution*
### Example 1 (cont.)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>3.4</td>
</tr>
<tr>
<td>6</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>4.9</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>6.3</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>7.1</td>
</tr>
</tbody>
</table>

**Rank sum:**
- Group 1: 16
- Group 2: 29

**Wilcoxon T:** 16

\[
E(T) = \frac{n_1(n_1+n_2+1)}{2}
\]
\[
= 25
\]

\[
SE(T) = \sqrt{n_1n_2(n_1+n_2+1)/12}
\]
\[
= 4.08
\]

\[
Z = \frac{(T-E(T))/SE(T)}
\]
\[
= (16-25)/4.08
\]
\[
= -2.20
\]

**p-value:** 0.032

Reject \( H_0 \)
Example 2

<table>
<thead>
<tr>
<th>ID</th>
<th>GROUP</th>
<th>ENERGY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>6.13</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>7.05</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>10.15</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>10.88</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>8.79</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>9.19</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>11.85</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>12.79</td>
</tr>
</tbody>
</table>

Example:
Energy expenditure in two groups, lean and obese. Want to test if they come from different distributions.

Number of cases read: 22
Number of cases listed: 22
Example 2 (cont.)

- Wilcoxon Rank Sum Test in SPSS:

<table>
<thead>
<tr>
<th>group</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>energy</td>
<td>13</td>
<td>7.92</td>
<td>103.00</td>
</tr>
<tr>
<td>1.00</td>
<td>9</td>
<td>16.67</td>
<td>150.00</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sum of ranks

<table>
<thead>
<tr>
<th>Test Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U</td>
</tr>
<tr>
<td>Wilcoxon W</td>
</tr>
<tr>
<td>Z</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
</tr>
<tr>
<td>Exact Sig. [2*(1-tailed Sig.)]</td>
</tr>
</tbody>
</table>

- We would have gotten exactly the same result by using the T distribution!
Some useful general comments

• Always start by checking whether it is appropriate to use t-tests (normally distributed data)
• If in doubt, a useful tip is to compare p-values from t-tests with p-values from non-parametric tests
• If similar results, stay with the parametric tests
Exercises

- Data files can be found on www.med.uio.no/imb/stat/kursfiler
- They are called Children, oppg-9-2 and oppg-9-4
Summary

Key words

• Skew data
• Logarithmic transformations (and other alternatives)
• Testing for normality
• Non-parametric tests:
  ▶ Sign test and Wilcoxon (paired data)
  ▶ Mann-Whitney/Wilcoxon rank-sum test (two independent samples)