Experimental Design and Biostatistics (TAMS38)
Lecture 4 – Non-parametric methods

Martin Singull

Department of Mathematics
Mathematical Statistics
Linköping University, Sweden

LINKÖPING
UNIVERSITY

Content

- Non-parametric methods
- Wilcoxon signed-rank test (one sample/pairwise observations)
- Sign test (one sample/pairwise observations)
- Tukey-Duckworth's quick test (two independent samples)
- Wilcoxon rank-sum test (two independent samples)
- Mann-Whitneys (two independent samples)
- Kruskal-Wallis (One-Way ANOVA)

Non-parametric methods

Many of the methods and analyses that we did in the basic courses are based on the assumption that data follows normal distribution

$\Rightarrow$ This is not always true!

The so-called **non-parametric methods** apply to much larger classes of distributions, e.g. all continuous distributions or any symmetric, continuous distributions.

Wilcoxons signed-rank test for one sample

Let $x_1, \ldots, x_m$ be observations from **continuous** distribution. This means that the r.v. X_i has a density function $f(x)$.

We want to test

H_0: Density function is symmetric around given value μ_0.

against

H_1: Density function is not symmetric around μ_0.

Other alternative hypothesis may occur. If H_0 is true, then $\mu_0 = E(X_i)$ if that exists, otherwise $\mu_0 =$ is a median.

Procedure

(i) Calculate differences $x_i - \mu_0$ and take away eventual zeros. There remains $y_1, \ldots, y_n$, where $y_j = x_j - \mu_0 \neq 0$.

(ii) Put the values $|y_1|, |y_2|, \ldots, |y_n|$ in increasing order (from the smallest to the biggest value) and assign the **rangs** $1, 2, \ldots, n$.

(iii) Calculate

$$T_+ = \text{sum of the ranks to the positive values } y$$
$$T_- = \text{sum of the ranks to the negative values } y$$

If H_0 is true then T_+ and T_- have the same distribution, "the signed rank distribution for sample size n".

H_0 is rejected if $T_+ \leq c$ or $T_- \leq c$ where

$$\frac{\alpha}{2} = P(T_+ \leq c) = P(T_- \leq c) = P(W_S \leq c).$$

a) $P(W_S \leq c)$ can be found in the table for $n \leq 15$.
b) If $n > 15$ to decide about the critical region we use that W_S is approx. normally distributed with

$$E(W_S) = \frac{n(n+1)}{4},$$
$$Var(W_S) = \frac{n(n+1)(2n+1)}{24}.$$

Also a one-sided test can be considered.

Confidence interval - Wilcoxons signed-rank test

Let $x_1, \ldots, x_n$ be observations of r.v. $X_1, \ldots, X_n$ with the density function that is symmetric around $\mu = E(X_i)$.

We want to find I_μ, i.e. a confidence interval for μ.

Procedure:

(i) Let $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ we the observations put in the increasing order.

(ii) Calculate for all fix i all averages

$$A_k = \frac{x_{(i)} + x_{(j)}}{2} \text{ where } i \leq j \text{ and } k = 1, ..., N.$$

In total we will have $\sum_{i=1}^{n}(n - i + 1) = \frac{n(n+1)}{2} = N$ of such averages, that we call $A_{(1)} \leq A_{(2)} \leq \ldots \leq A_{(N)}$, after putting them in the increasing order.

One can show that

$$P(A_{(k)} < \mu < A_{(N-k+1)}) = 1 - 2P(W_S \leq k - 1).$$

Interval $I_\mu = (A_{(k)},\ A_{(N-k+1)})$ can confidence level $1 - \alpha$, where k is given by the condition

$$\frac{\alpha}{2} = P(W_S \leq k - 1),$$

with W_S being Wilcoxons "signed rank statistic" for the sample size n and where $N = n(n+1)/2$.

Example 1

Assume that we have 5 observations

$$x_1 = 0.826, \quad x_2 = 0.829, \quad x_3 = 0.831, \quad x_4 = 0.836, \quad x_5 = 0.840$$

and we want to test

$$H_0: \ \mu \, (= E[X_i]) = 0.830 \quad \text{versus} \quad H_1: \ \mu \neq 0.830.$$

We have then

i	1	2	3	4	5
$y_i = x_i - 0.830$	-0.004	-0.001	0.001	0.006	0.010
$\lvert y_i \rvert$	0.004	0.001	0.001	0.006	0.010
rang	3	1.5	1.5	4	5

Calculate $A_k = \dfrac{x_{(i)} + x_{(j)}}{2}$, $k = 1, ..., 15$

$x_{(i)}/x_{(j)}$	0.826	0.829	0.831	0.836	0.840
0.826	0.826	0.8275	0.8285	0.831	0.833
0.829	-	0.829	0.830	0.8325	0.8345
0.831	-	-	0.831	0.8335	0.8355
0.836	-	-	-	0.836	0.838
0.840	-	-	-	-	0.840

and then we have

k	1	2	3	4	5	6	7	8
$A_{(k)}$	0.826	0.8275	0.8285	0.829	0.830	0.831	0.831	0.8325

k	9	10	11	12	13	14	15
$A_{(k)}$	0.833	0.8335	0.8345	0.8355	0.836	0.838	0.840

We have $n = 5$ and $N = 5 \cdot 6/2 = 15$. Table for the signed-rank test gives

$$P(W_s \leq 0) = 0.031.$$

For $k - 1 = 0$, i.e., for $k = 1$ we obtain interval with the confidence level $1 - 2 \cdot 0.031 = 0.938$, so

$$I_\mu = (A_{(1)},\ A_{(15)}) = (0.826,\ 0.840).$$

If we can assume that $x_1, \dots, x_5$ are the observations from $N(\mu, \sigma)$ we obtain 95% confidence interval

$$I_\mu = \left(\bar{x} \mp \underbrace{t_{0.975}(4)}_{=2.78} \frac{s}{\sqrt{n}}\right) = (0.825,\ 0.839)$$

In general, you get shorter intervals when employing the normal distribution theory, or the theory of another class of distributions.

Pairwise measurements

We have n' pairs of values $(x_1, y_1), (x_2, y_2), \ldots, (x_{n'}, y_{n'})$.

Here, one can think that e.g. x_i and y_i are measurements for the same object done using two different methods or measurements for the same patient before and after operation.

We want to test

H_0: In each of the pairs X_i and Y_i have the same continuous distribution.

versus

H_1: Distribution of X_i offsets from the distribution of Y_i **in the same direction** for all i.

Procedure

Step 1: Calculate differences $z_i = x_i - y_i$. Take away all zeros. There remains after renumbering values $z_1, \ldots, z_n$ that are all different than zero.

Step 2a: Signed-test.

If H_0 is true, i.e., if X_i and Y_i have the same continuous distribution, then

$$P(X_i > Y_i) = P(Y_i > X_i) = \frac{1}{2},$$

as $P(X_i = Y_i) = 0$ for any continuous distribution.

Test statistic: $\nu_+ =$ number of positive z-values.

Two-sided test: H_0 is rejected if $\nu_+ \leq a$ or $\nu_+ \geq b$, where

$$\frac{\alpha}{2} \geq P(\nu_+ \leq a \text{ if } H_0 \text{ is true })$$
$$\frac{\alpha}{2} \geq P(\nu_+ \geq b \text{ if } H_0 \text{ is true })$$

Ff H_0 is true, we have

$$\nu_+ \sim Bin\left(n, \frac{1}{2}\right) \approx N\left(n/2, \sqrt{n/4}\right)$$

for $n \geq 40$.

For the one-sided test critical region depends on the chosen alternative hypothesis.

Step 2b: Wilcoxons signed-rank test.

If H_0 is true, then the distribution of Z_i is symmetric around 0.

With help of Wilcoxons signed-rank test we can test if

$$E(Z_i) = 0$$

assuming that $E(Z_i)$ exists.

Example 2 - Sleeping pills

Results of placebo-controlled clinical trial to test the effectiveness of a sleeping drug.

	Hours of sleep			Rank
Patient	Drug (x_i)	Placebo (y_i)	Difference (z_i)	(ignoring sign)
1	6.1	5.2	0.9	3.5*
2	7.0	7.9	-0.9	3.5*
3	8.2	3.9	4.3	10
4	7.6	4.7	2.9	7
5	6.5	5.3	1.2	5
6	8.4	5.4	3.0	8
7	6.9	4.2	2.7	6
8	6.7	6.1	0.6	2
9	7.4	3.8	3.6	9
10	5.8	6.3	-0.5	1

H_0: In each of the pairs X_i and Y_i we have the same distribution, i.e., drug and placebo are as good.

versus

H_1: Systematic shift of sleeping time at the treatment.

Step 2a: Test statistic $\nu_- = 2$.

The r.v. $\nu_- \sim Bin(10, 0.5)$ if H_0 is true. H_0 is rejected on level 0.05 if $\nu_- \leq a$ or $\nu_- \geq b$, where

$$0.025 \geq P(\nu_- \leq a \text{ if } H_0 \text{ is true })$$
$$0.025 \geq P(\nu_- \geq b \text{ if } H_0 \text{ is true })$$

Table gives $a = 1$ and $b = 9$ with $\alpha = 0.0216$.

Since $1 < 2 < 9$ we can **not reject** H_0 on level 0.05.

Here, we could formulate H_1 differently and do one-sided test.

Steg 2b: Wilcoxons signed-rank test:

Test statistic $T_{-} = 4.5$.

H_0 is rejected if $T_{-} \leq c$ or $T_{+} \leq c$, where

$$0.025 \geq P(T_{-} \leq c \;\text{ if }\; H_0 \;\text{ is true })$$
$$0.025 \geq P(T_{+} \leq c \;\text{ if }\; H_0 \;\text{ is true })$$

Table for Wilcoxons signed-rank test gives $c = 8$. Since $T_{-} = 4.5 < 8$, the null hypothesis H_0 should be rejected.

As T_{-} is small, we suspect that the treatment gives a positive effects.

Two independent samples

Now, let us assume that $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ two series of measurements that are completely disconnected from each other, i.e., two independent measurement series.

We assume that $X_1, \ldots, X_{n_1}$ are independent continuous r.v. with density function f and that $Y_1, \ldots, Y_{n_2}$ are independent continuous r.v. with density function g.

We want to test

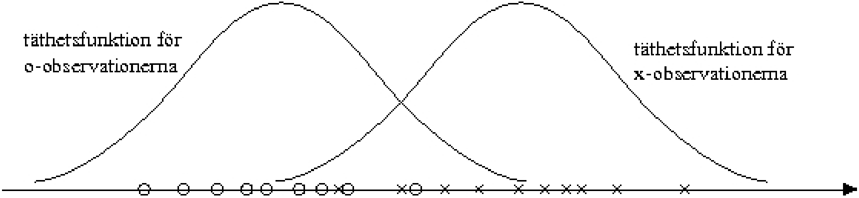
$$H_0: \ f = g \quad \text{vs.} \quad H_1: f \neq g.$$

One can use

- Tukey-Duckworth's quick test, or
- Wilcoxons rank-sum test

Example 3

Observations from the two samples has been highlighted on a number line. They are well mixed and one can well imagine that they have the same density function.



Here, we have two samples, but it no longer seems reasonable that they come from the same distribution. Density functions are probably offset relative to each other, e.g., as the figure shows.

Tukey-Duckworth's quick test.

If $4 \leq n_1 \leq n_2 \leq 30$, and $n_2 \leq \frac{4n_1}{3} + 3$ then we can test H_0 through

1. Find the smallest and the largest observation in both samples, respectively.
2. For the sample that includes the largest value from all observed values, find the number of the observations that are larger than this largest value of the other sample.
3. For the other sample, find the number of observations that are smaller than the smallest one in the first sample.
4. Let C be sum of the observations in 2. and 3. For $\alpha = 5\%$, 1% or 0.1%, we reject H_0 if $C \geq 7$, 10, or 13, respectively.

Wilcoxons test

Wilcoxons test uses the idea of putting all observations on a line and see if there are systematic differences between the samples.

One put all the observations in the increasing order. Then, we assign ranks $1, 2, ..., n_1 + n_2$ to the observations, i.e., we numerate them from 1 to $n_1 + n_2$.

As test statistic we use the rank-sum T for $x_1, ..., x_{n_1}$.

If H_0 is true, i.e., if $f = g$, so are the observations generally pretty well mixed, what means that T obtains a value relatively close to

$$\frac{n_1(n_1 + n_2 + 1)}{2},$$

i.e., the average rank.

A small or large value of T indicate that the distributions are offset relative to each other.

Limits of the critical region T_ℓ and T_r determined by conditions

$$P(T \leq T_\ell \text{ if } H_0 \text{ is true }) \leq \frac{\alpha}{2},$$

$$P(T \geq T_r \text{ if } H_0 \text{ is true}) \leq \frac{\alpha}{2}$$

For small values of n_1 och n_2 values T_ℓ and T_r can be found in the tables (observe that $2\alpha \Leftrightarrow \alpha$ in the table).

If we know the distribution of the samples except for one or a few parameters, the best parametric test is generally better in power than the Wilcoxon test.

We could also formulate H_1 differently and do one-sided test.

Normal approximation

For large values of n_1 and n_2 we use the fact that if H_0 is true, then T is approx. normally distributed. One calculates test statistic

$$U = \frac{T - n_1(n_1 + n_2 + 1)/2}{\sqrt{n_1 n_2(n_1 + n_2 + 1)/12}} \sim approx\ N(0, 1)$$

- One-sided test: H_0 is rejected if $U > c$ respectively $U < -c$. The limit of the critical region c is given by $\Phi(c) = 1 - \alpha$.
- Two-sided test: H_0 is rejected if $|U| > c'$. Here, c' is given by condition $\Phi(c') = 1 - \alpha/2$.

Each of the test has the significance level α.

Example 4

Sample 1 :	0.57	0.74	1.66	2.13	11.6		
Sample 2 :	−1.33	−1.09	−0.89	−0.66	−0.53	−0.48	−0.08
	0.07	0.28	0.37	0.49	1.23	1.47	

We want to test hypothesis H_0, given below, on level 5%.

H_0 : samples comes from the same distribution,

versus

H_1 : distribution of the sample 1 gives in general larger value than the distribution of sample 2.

The samples together and sorted in increasing order

−1.33	−1.09	−0.89	−0.66	−0.53	−0.48	−0.08	0.07	0.28
0.37	0.49	0.57*	0.74*	1.23	1.47	1.66*	2.13*	11.6*

* Observations from sample 1.

We see that the sample no. 1 gets ranks 12, 13, 16, 17, 18.

Let T denoted the rank-sum for the sample 1.

We get $T = 12 + 13 + 16 + 17 + 18 = 76$.

Using H_1 we know that H_0 should be rejected in favor of H_1 for large values of T. Table gives the limit of the critical region $T_r = 65$.

Since $76 > 65$, the hypothesis H_0 should be rejected on level 5% and we conclude that H_1 holds.

Ties

A prerequisite for the Wilcoxon test is that the samples come from continuous distributions. Theoretically, it is therefore not occur that two or more observations are equal, but in practice it happens quite often for example due to rounding effects.

These observations are called *ties*. One gives the same rank to the all observations of the same value.

Consider a piece of the put together and order samples

$$\begin{array}{lccccccccc} Obs: & \dots & 39 & 42 & 45 & 45 & 45 & 50 & 53 & \dots \\ Rang: & \dots & 10 & 11 & r & r & r & 15 & 16 & \dots \end{array}$$

Here, the rank $r = (12 + 13 + 14)/3$, i.e., it is a arithmetic average of the ranks that should be used.

Example 5 - Blood cells

The following data shows the number of red blood cells (unit: million per cubic millimeter) for eight men and ten women

M (x_i) :	5.02	4.58	5.57	4.52	4.84	5.36	4.27	5.15		
W (y_j) :	4.15	3.89	4.56	4.40	4.38	4.20	4.31	4.73	4.26	3.89

We want, on level 5%, test

H_0: The number of red blood cells have the same distribution for women och men

vs.

H_1: There is a systematic difference between the distributions for women and men.

We will use the Wilcoxon rank sum test. We start by organizing measurement in the increasing order within each sample.

M :	4.27	4.52	4.58	4.84	5.02	5.15	5.36	5.57		
W :	3.89	3.89	4.15	4.20	4.26	4.31	4.38	4.40	4.56	4.73

Rank-sum for men, i.e., for $n_1 = 8$ observations:

$T =$

We have $n_1 = 8$ and $n_2 = 10$ and a two-sided test on level 5%.

Hence, we should use table with $2\alpha = 0.05$. Table gives that H_0 should be rejected if $T \leq 53$ or $T \geq 99$.

Conclusions:

Mann-Whitneys test

This test is equivalent to the Wilcoxon rank sum test, but the test statistic has been modified.

Mann-Whitneys test uses instead the test statistic U_{n_1} that is calculated with use of the differences $d_{ij} = x_i - y_j$, where $i = 1, \ldots, n_1$ och $j = 1, \ldots, n_2$. One has

$$U_{n_1} = [\text{ number of } d_{ij} \text{ such that } d_{ij} > 0\ ] + $$

$$+ [\text{ number of } d_{ij} \text{ such that } d_{ij} = 0\ ] \cdot \frac{1}{2}$$

Theoretically there is no d_{ij} such that $d_{ij} = 0$, but this definition is in line with our approach to manage ties, i.e., two or more equal observations.

We will now study relation between T and U_{n_1}. Let

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n_1)} \quad \text{and} \quad y_{(1)} \le y_{(2)} \le \ldots \le y_{(n_2)}$$

be our observations that are ordered increasingly.
For the case without ties

$$\begin{aligned} T &= [\text{ number of } y\text{-observations that are} < x_{(n_1)}] + n_1 \\ &\quad + [\text{ number of } y\text{-observations that are} < x_{(n_1-1)}] + n_1 - 1 \\ &\quad + \ldots + [\text{ number of } y\text{-observations that are} < x_{(1)}] + 1 \\ &= U_{n_1} + n_1 + (n_1 - 1) + \ldots + 1 = U_{n_1} + n_1 \cdot \frac{n_1 + 1}{2}. \end{aligned}$$

Two-sided test: H_0 is rejected if $T \le T_l$ or $T \ge T_r$ which corresponds to the

$$U_{n_1} \le T_l - \frac{n_1(n_1+1)}{2} \text{ or } U_{n_1} \ge T_r - \frac{n_1(n_1+1)}{2},$$

where T_l and T_r can be found in Wilcoxon-table.

We assume that the two cdfs F and G can be written as

$$F(x) = H(x - \theta_1) \quad och \quad G(x) = H(x - \theta_2).$$

Here, θ_1 and θ_2 are called **position parameters** and $\theta_1 - \theta_2$ describes how much the probability mass of x-observations is shifted with respect to probability mass of y-observations.

Furthermore, $H_0 : \; F = G$ att $\theta_1 = \theta_2$.

A common situation is that

$$X_i = \theta_1 + \varepsilon_i \quad and \quad Y_j = \theta_2 + \varepsilon'_j$$

where ε_i and ε'_j are independent and identically distributed (but not necessary normally distributed).

(i) If expected values $E(\varepsilon_i) = E(\varepsilon'_j)$ exist then

$$E(X_i) - E(Y_j) = \theta_1 + E(\varepsilon_i) - (\theta_2 + E(\varepsilon'_j)) = \theta_1 - \theta_2$$

(ii) If expected values $E(X_i)$ and $E(Y_j)$ do not exist then

$$\theta_1 - \theta_2 = \widetilde{m}_1 - \widetilde{m}_2,$$

where $\widetilde{m}_1$ and $\widetilde{m}_2$ stands for medians.

The median $\widetilde{m}$ is given by $F(\widetilde{m}) = \frac{1}{2}$, where F is the cdf.)

Confidence interval for $\theta_1 - \theta_2$

To find the confidence interval for $\theta_1 - \theta_2$, we will use the differences $d_{ij} = x_i - y_j$, i.e.,

1. Let $d_{(1)} \leq d_{(2)} \leq \ldots \leq d_{(n_1 n_2)}$ be $n_1 n_2$ differences d_{ij} ordered increasingly.
2. Let $c = T_l - \dfrac{n_1(n_1+1)}{2}$ be a **left limit of the critical region** for Mann-Whitneys test statistic U_{n_1} for two-sided test on level $\leq \alpha$. (T_l comes from Wilcoxon-table.)
3. Interval

$$I_{\theta_1-\theta_2} = \left(\, d_{(c+1)},\ d_{(n_1 n_2 - c)}\right)$$

is a confidence interval for $\theta_1 - \theta_2$ with confidence level at least $1 - \alpha$.

For a derivation, see Lehmann, E. L. (1975). *Nonparametrics - Statistical Methods Based on Ranks*.

Point estimators

We estimate $\theta_1 - \theta_2$ with

$$\hat{\theta}_1 - \hat{\theta}_2 = \text{ median for } d_{(1)}, \ldots, d_{(n_1 n_2)},$$

i.e.,

$$\hat{\theta}_1 - \hat{\theta}_2 = \begin{cases} d_{(k+1)} & \text{if } n_1 n_2 = 2k+1 \\ \dfrac{d_{(k)} + d_{(k+1)}}{2} & \text{if } n_1 n_2 = 2k \end{cases}$$

Example 6 - Augmenters and reducers (Lehmann, 1975)

Petrie has developed an interesting classification of persons into augmenters (who perceive an exaggerated impression of sensory stimuli), reducers (whose perception tends to diminish such stimuli), and and intermediate category of moderates.

As a check on his classification, Petrie tested the reactions of a number of subjects, once with audio analgesia (an acoustic treatment which tends to increase the tolerence to pain) and once without. According to his theory, the reduction in response due to analgesia should be more marked for the augmenters than for reducers.

Augmenters:	17.9	13.3	10.6	7.6	5.7	5.6	5.4
	3.3	3.1	0.9				
Reducers:	7.7	5.0	1.7	0.0	-3.0	-3.1	-10.5

Here, we have $n_1 = 10$, $n_2 = 7$ and

$$U_{n_1} = \text{ number of pairs } (x_i, y_j) \text{ where } x_i > y_j.$$

Hypothesis $H_0 : F = G \Leftrightarrow \Delta = \theta_1 - \theta_2 = 0$ should be rejected on level $\leq 2\alpha = 0.10$ if

$$U_{n_1} \leq 72 - 10 \cdot 11/2 = 17 = c$$

or if

$$U_{n_1} \geq 108 - 10 \cdot 11/2 = 53.$$

```
MTB > WDifferences 'Aug' 'Red' c3.
MTB > sort diff sortdiff
MTB > print sortdiff
```

Data Display

```
sortdiff
-6.8  -4.6  -4.4  -4.1  -2.3  -2.1  -2.0  -1.9  -1.7  -0.8
-0.1   0.4   0.6   0.7   0.9   1.4   1.6   2.6   2.9   3.1
 3.3   3.7   3.9   3.9   4.0   4.0   5.4   5.6   5.6   5.6
 5.7   5.9   6.1   6.2   6.3   6.4   7.6   8.3   8.4   8.5
 8.6   8.7   8.7   8.8   8.9  10.2  10.6  10.6  10.7  11.4
11.6  12.9  13.3  13.6  13.6  13.7  13.8  15.9  16.1  16.2
16.2  16.3  16.4  17.9  18.1  20.9  21.0  21.1  23.8  28.4
```

We have point estimate $\hat{\theta}_1 - \hat{\theta}_2 = [\text{ median for } d_{(i)} \,] = 6.35$. and the interval

$$I_{\theta_1-\theta_2} = \left(d_{(c+1)},\ d_{(70-c)}\right) = \left(d_{(18)},\ d_{(53)}\right) = (2.6,\ 13.3)$$

with confidence level 90%. Compare with Minitab.

We can obtain the same confidence interval using Minitab.

```
MTB > Mann-Whitney 90.0 'Aug' 'Red';
SUBC>   Alternative 0.
```

Mann-Whitney Test and CI: Aug, Red

```
      N  Median
Aug  10    5.65
Red   7    0.00

Point estimate for ETA1-ETA2 is 6.35
91.2 Percent CI for ETA1-ETA2 is (2.60,13.30)
W = 114.0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.0218
```

Assuming normal distribution we have:

```
MTB > TwoSample 'Aug' 'Red';
SUBC>   Pooled;
SUBC>   Confidence 90.
```

Two-Sample T-Test and CI: Aug, Red

```
Two-sample T for Arg vs Red

      N   Mean  StDev  SE Mean
Aug  10   7.34   5.20      1.6
Red   7  -0.31   5.99      2.3

Difference = mu (Aug) - mu (Red)
Estimate for difference:  7.65
90% CI for difference:  (2.88, 12.43)
T-Test of difference = 0 (vs not =): T-Value = 2.81
                                P-Value = 0.013  DF = 15
Both use Pooled StDev = 5.5277
```

Example 7 - Pooling and subdiets (Lehmann, 1975)

In a comparative study of certain diet groups, it is desired to test whether it is necessary to distinguish between the diets within the same group.

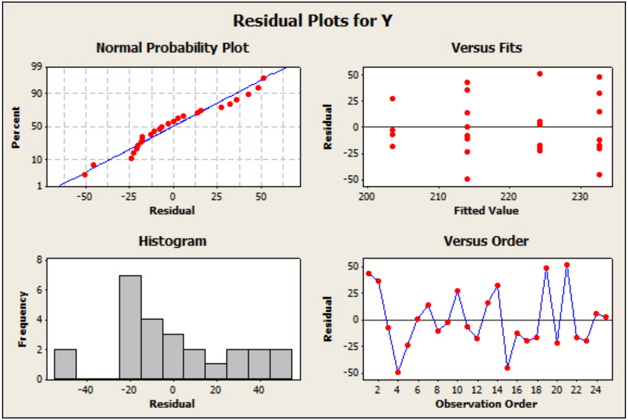
The following table gives the growth of 25 rats after 12 weeks on four subdiets of diet A together with their ranks.

Diet	Growth Figures	Ranks
A_1	257, 250, 206, 164, 190, 214, 228, 203	22,21,10,1,4,13,17,8
A_2	201, 231, 197, 185	6,19,5,2
A_3	248, 265, 187, 220, 212, 215, 281	20,23,3,15,12,14,25
A_4	202, 276, 207, 204, 230, 227	7,24,11,9,18,16

One-way ANOVA: Y versus A

Source	DF	SS	MS	F	P
A	3	2596	865	0,99	0,415
Error	21	18272	870		
Total	24	20868			

S = 29,50



Kruskal-Wallis test for One-Way ANOVA

We want to compare a treatments and for treatment i we have n_i independent observations

	Observations
Treatment 1	$y_{11},\ y_{12}, \ldots, y_{1n_1}$
$\vdots$	$\vdots$
Treatment a	$y_{a1},\ y_{a2}, \ldots, y_{an_a}$

and we want to test hypothesis

H_0 : a samples comes from the same continuous distribution
i.e., treatments are equally good.

versus

H_1 : There is difference between distributions.

Kruskal-Wallis - Procedure

1. Merge all the samples into a large sample of $N = n_1 + n_2 + \ldots + n_a$ observations. Arrange them in order and assign them ranks $1, 2, \ldots, N$. Let

$$r_{ij} = \text{ rank for } y_{ij},$$

$$s_i = \sum_{j=1}^{n_i} r_{ij} = \text{ rank-sum for sample } i,$$

$$S_a = \sum_{i=1}^{a} \frac{s_i^2}{n_i}.$$

2. Calculate test statistics

$$T = \frac{12 S_a}{N(N+1)} - 3(N+1)$$

or

$$T_{ties} = \frac{(N-1)(S_a - C)}{S_r - C},$$

where $S_r = \sum_{i,j} r_{ij}^2$ och $C = \frac{1}{4}N(N+1)^2$, if we have **ties**.

3. Deviation from the null hypothesis is manifested by large values of the test statistics. Thus we reject H_0 if $T \geq c$.
   (i) For the small values of $n_1, \ldots, n_a$ value c is given by the table where $\alpha = P(T \geq c$ if H_0 true).
   (ii) For "large" values of $n_1, \ldots, n_a$ we are using that

$$T \text{ is appr } \chi^2(a-1),$$

   if H_0 is true, then c can be found in $\chi^2(a-1)$-table.

See Lehmann, E. L. (1975) *Nonparametrics - Statistical Methods Based on Ranks*, for more details.

Explanations of the test statistic.

Test statistics can be written as

$$T = \frac{12}{N(N+1)} \sum_{i=1}^{a} n_i \left(\frac{s_i}{n_i} - \frac{N+1}{2} \right)^2.$$

This involves comparing the average rank $\frac{s_i}{n_i}$ for sample i with the average rank $\frac{N+1}{2}$ for the combined sample, compare

$$SS_{TREAT} = \sum_{i=1}^{a} n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2.$$

If all the samples come from the same distribution all $\frac{s_i}{n_i}$ are close to $\frac{N+1}{2}$. Deviation from the null hypothesis leads large values of T.

Example 7, cont.

Here, $s_1 = \sum_{j=1}^{8} r_{1j} = 1 + 4 + \cdots + 18 + 22 = 96$ and

$$s_2 = 32, \; s_3 = 112, \; s_4 = 85$$

$$S_4 = \frac{96^2}{8} + \frac{32^2}{4} + \frac{112^2}{7} + \frac{85^2}{6} = 4404.2$$

$$T = \frac{12 S_4}{25 \cdot 26} - 3 \cdot 26 = 3.31$$

The r.v. T is *appr* $\chi^2(3)$ if H_0 is true.

H_0 is rejecte if $T \geq c$. Table gives $c = 7.82$.

$3.31 < 7.82$. No detectable difference between diets. Here, the significance level is rather approximate because the sample sizes n_i are not equal.

```
MTB > stack c1-c4 c5;
SUBC> subscripts c6.
MTB > Kruskal-Wallis 'Y' 'A'.
```

Kruskal-Wallis Test: Y versus A

```
Kruskal-Wallis Test on Y

A         N  Median  Ave Rank      Z
A1        8   210,0      12,0  -0,47
A2        4   199,0       8,0  -1,48
A3        7   220,0      16,0   1,27
A4        6   217,0      14,2   0,45
Overall  25              13,0

H = 3,31  DF = 3  P = 0,347

* NOTE * One or more small samples
```

Pairwise comparisons for One-Way ANOVA

If you want to make pairwise comparisons between the position parameters of the various samples, you can use the Wilcoxon-Mann-Whitney test (rank-sum test) and the method for the confidence interval construction associated with the test.

Mann-Whitney Test and CI: A1; A2

```
    N  Median
A1  8  210,00
A2  4  199,00


Point estimate for ETA1-ETA2 is 11,00
96,6 Percent CI for ETA1-ETA2 is (-32,98;55,99)
W = 57,0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0,4447
```

Linköping University - Research that makes a difference

LINKÖPING UNIVERSITY