Teaching Mathematics and Statistics in Sciences, IPA HU-SRB/0901/221/088 - 2011

Mathematical and Statistical Modelling in Medicine

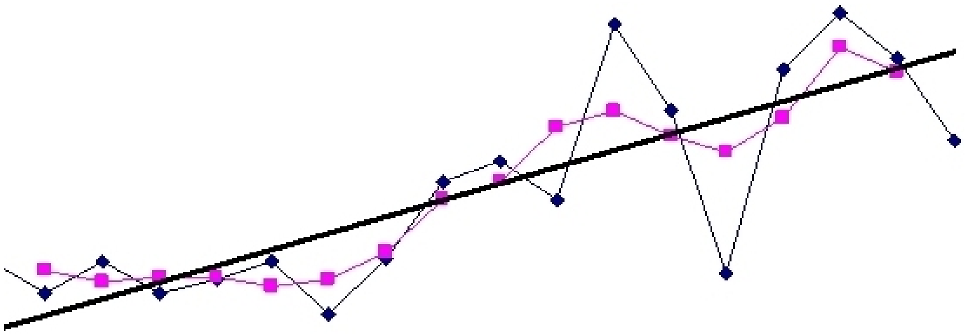
Author: **Tibor Nyári** PhD

University of Szeged
Department of Medical Physics and Informatics

www.model.u-szeged.hu
www.szote.u-szeged.hu/dmi

Nonparametric test

One sample tests
Two sample tests
Testing for three or more samples



Background

- So far we have stressed that in order to carry out hypothesis tests we need to make certain assumptions about the types of distributions from which we were sampling. For example. to do t tests we needed to assume that the populations involved were approximately normal. In the two sample t-test we needed to make the more specific assumption that the variances are equal. An important part of statistics deals with tests for which we do not need to make such specific assumptions. These tests are called nonparametric or distribution-free tests.
- These tests would ordinarily be used if a parametric test were not appropriate. This might happen. for instance. if you were working with a non normal distribution. or a distribution whose shape was not yet evident. It might also happen that you are working with some special type of data for which there was no appropriate parametric test

Ranking the data

- Nonparametric tests can't use the estimations of population parameters. They use ranks instead. Instead of the original sample data we have to use its rank. to show the ranking procedure suppose we have the following sample of measurements:
- 199. 126. 81. 68. 112. 112.
- Case 4 has the smallest value (68). it is assigned a rank of 1. Case 3 has the next smallest value. it is assigned a rank of 2. Cases 5 and 6 are equal. they are assigned a rank of 3.5. the average rank of 3 and 4. We say that case 5 and 6 are tied. The next table shows the result of ranking.

Tabulate the data

Case	Data	Rank
1	199	6
2	126	5
3	81	2
4	68	1
5	112	3.5
6	112	3.5

$$\sum_{i=1}^{n} r_i = \frac{n(n+1)}{2} = \frac{6*7}{2} = 21$$

Type of tests

- **One sample tests**
  - **Sign test**
  - Wilcoxon sign test
- **Two samples tests**
  - (Mann-Whitney test)
  - **(Wilcoxon Rank-Sum test)**
- **More than two samples**
  - Kruskall-Wallis test
  - Jonckheere-Terpstra test

Wilcoxon sign test

- Data are in pairs
  - E.g.: before-after treatment
- We have n subjects and X ($x_1.x_2...x_n$). Y ($y_1.y_2...y_n$) denotes the variable before and after treatment. respectively.
- Ignore where $x_j=y_j$.
  - $x_j=\tau+\varepsilon_i$
  - $y_j=\tau-\nu+\varepsilon_i'$
  - $d_j=x_j-y_j=\nu+\varepsilon_i-\varepsilon_i'$
- $E(d_i)=\nu$; and $E(\varepsilon_i)=E(\varepsilon_i')=0$
- H_0: $\nu=0$
- H_a:= $\nu>0$; H_a= $\nu<0$ or H_a $\nu\neq0$

Wilcoxon Sign Test

- Calculate absolute values of z_i.
- Sort them.
- Calculate δ_i.
- The test statistics T^+

$$\delta_i = \begin{cases} 1, & \text{if } d_i > 0 \\ 0, & \text{if } d_i < 0 \end{cases}$$

$$T^+ = \sum_{i=1}^{n'} \delta_i R_i$$

Decision rule

- Use standard normal distribution table

$$E(T^{+}) = \frac{n(n+1)}{4}; D^{2}(T^{+}) = \frac{n(n+1)(2n+1)}{24}$$

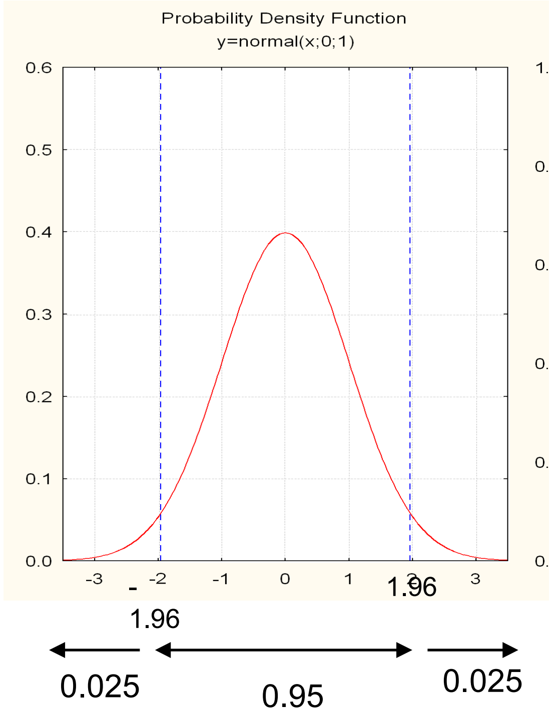
$$z = T^{*} = \frac{T^{+} - E(T^{+})}{D(T^{+})}$$

Decision

- If the calculaterd |z| score is greater than 1.96, then Nullhypothesis is rejected , and the alternative hypothesis is accepted, namely the diffence is significant
- If the calculaterd |z| score is less than 1.96, then Nullhypothesis is accepted, namely the diffence is NOT significant.

Standard normal probabilities

z	Φ(x): proportion of area to the left of Z
-4	0.0003
-3	0.0013
-2.58	0.0049
-2.33	0.0099
-2	0.0228
-1.96	0.0250
-1.65	0.0495
-1	0.1587
0	0.5
1	0.8413
1.65	0.9505
1.96	0.975
2	0.9772
2.33	0.9901
2.58	0.9951
3	0.9987
4	0.99997



 10

Example

- There is a treatment using a new drug at 9 patients.
- Data are summarised in the next table.
- X is the baseline hormone level
- Y is the after treatment hormone level
- Is there any changes at hormone levels after treatment?

The data

i	x_i	y_i	d_i	$\lvert d_i \rvert$	R_i	δ_i	$\delta_i R_i$
1	1.83	0.878	-0.952	0.952	8	0	0
2	0.5	0.647	0.147	0.147	3	1	3
3	1.62	0.598	-1.022	1.022	9	0	0
4	2.48	2.05	-0.43	0.43	4	0	0
5	1.68	1.06	-0.62	0.62	7	0	0
6	1.88	1.29	-0.59	0.59	6	0	0
7	1.55	1.06	-0.49	0.49	5	0	0
8	3.06	3.14	0.08	0.08	2	1	2
9	1.3	1.29	-0.01	0.01	1	0	0

- H_0: $\nu=0$
- H_a $\nu \neq 0$
- Test statistics

$$T^+ = \sum_{i=1}^{n'} \delta_i R_i = 5$$

- $T_{\alpha/2.\ n=9}=39$
- The intervall:
- $T^+ \leq 6$ or $T^+ \geq 39$
- So we reject H_0

$$T^+ \leq \frac{9*10}{2} - T_{\alpha/2,n=9} or T^+ \geq 39$$

STATA results

```
       sign |      obs   sum ranks    expected
   ---------+---------------------------------
   positive |        7          40        22.5
   negative |        2           5        22.5
       zero |        0           0           0
   ---------+---------------------------------
        all |        9          45          45

   unadjusted variance           71.25
   adjustment for ties            0.00
   adjustment for zeros           0.00
                            ----------
   adjusted variance             71.25

   H0: xi = yi
                z =   2.073
       Prob > |z| =   0.0382
```

t-Test: Paired Two Sample for Means

	before	*after*
Mean	1.766666667	1.334777778
Variance	0.512075	0.643738944
Observations	9	9
Pearson Correlation	0.847876519	
df	8	
t Stat	3.035375416	
P(T<=t) one-tail	0.008088314	
t Critical one-tail	1.859548033	
P(T<=t) two-tail	0.016176627	
t Critical two-tail	2.306004133	

Mann-Whitney Test

- (Non-parametric independent two-group comparisons)
- Definition: A non-parametric test (distribution-free) used to compare two independent groups of sampled data.
- Assumptions: Unlike the parametric t-test. this non-parametric makes no assumptions about the distribution of the data (e.g.. normality).
- Characteristics: This test is an alternative to the independent group t-test. when the assumption of normality or equality of variance is not met. This. like many non-parametric tests. uses the ranks of the data rather than their raw values to calculate the statistic. Since this test does not make a distribution assumption. it is not as powerful as the t-test.
- Test: The hypotheses for the comparison of two independent groups are:
- Ho: The two samples come from identical populations
- Ha: The two samples come from different populations

 16

Mann-Whitney (M-W) procedure

- To compute the test. the observations from both samples are first combined and ranked from smallest to largest value. The statistic for testing the null hypothesis that the two distributions are equal is the sum of the ranks for each of the two groups. If the groups have the same distribution. their sample distributions of ranks should be similar. If one of the groups has more than its share of small or large ranks. there is reason to suspect that the two underlying distributions are different.
- If the total sample size is less than 30. tables can be used where an interval for R_{min}-R_{max} is given. If one of our test statistic is in the interval. we do not reject the null hypothesis. For large sample size a normal approximation is possible to get the p-value

 17

M-W test

- Notice that the hypothesis makes no assumptions about the distribution of the populations. These hypotheses are also sometimes written as testing the equality of the central tendency of the populations.

- The test statistic for the Mann-Whitney test is U. This value is compared to a table of critical values for U based on the sample size of each group. If U exceeds the critical value for U at some significance level (usually 0.05) it means that there is evidence to reject the null hypothesis in favor of the alternative hypothesis.

- Note: Actually. there are two versions of the U statistic calculated. where $U' = n_1 n_2 - U$ where n1 and n2 are the sample sizes of the two groups. The largest of U or U' is compared to the critical value for the purpose of the test.

- **Note: For sample sizes greater than 8. a z-value can be used to approximate the significance level for the test. In this case. the calculated z is compared to the standard normal significance levels**.

- Note: The U test is usually perform as a two-tailed test. however some text will have tabled one-tailed significance levels for this purpose. If the sample size if large. the z-test can be used for a one-sided test.

Example (M-W)

- Professor Testum wondered if students tended to make better scores on his test depending if the test were taken in the morning or afternoon. From a group of 19 similarly talented students. he randomly selected some to take a test in the morning and some to take it in the afternoon. The scores by groups were:

The Data

Morning	Afternoon
89.8	87.3
90.2	87.6
98.1	87.3
91.2	91.8
88.9	86.4
90.3	86.4
99.2	93.1
94.0	89.2
88.7	90.1
83.9	

Calculate ranks

Morning	Afternoon	Morning Ranks	Afternoon Ranks
89.8	87.3	10	4.5
90.2	87.6	12	6
98.1	87.3	18	4.5
91.2	91.8	14	15
88.9	86.4	8	2.5
90.3	86.4	13	2.5
99.2	93.1	19	16
94	89.2	17	9
88.7	90.1	7	11
83.9		1	

 21

Sum of ranks

- $\Sigma_{\text{Morning ranks}} = 119$
- $\Sigma_{\text{Afternoon ranks}} = 71$
- M-W critical value is 75-125
- 119є[75-125]
- So we accept null hypothesis.

STATA Results of Mann-Whitney test

```
Two-sample Mann-Whitney rank-sum test

   group |      obs    rank sum    expected
---------+---------------------------------
       1 |       10         119         100
       2 |        9          71          90
---------+---------------------------------
combined |       19         190         190

unadjusted variance      150.00
adjustment for ties       -0.26
                     ----------
adjusted variance        149.74

Ho: data(group==1) = data(group==2)
             z =   1.553
    Prob > |z| =   0.1205
```

t-Test: Two-Sample Assuming Equal Variances

	Morning	*Afternoon*
Mean	91,43	88,8
Variance	20,83566667	5,85
Observations	10	9
Pooled Variance	13,78358824	
Hypothesized Mean Difference	0	
df	17	
t Stat	1,541768106	
P(T<=t) one-tail	0,070769125	
t Critical one-tail	1,739606716	
P(T<=t) two-tail	0,14153825	
t Critical two-tail	2,109815559	

 24

Wilcoxon Rank-Sum Test

- (Non-parametric independent two-group comparisons)
- Definition: A non-parametric test (distribution-free) used to compare two independent groups of sampled data.
- Test: The hypotheses for the comparison of two independent groups are:

- H_0: The two samples come from identical populations
- H_a: The two samples come from different populations

 25

Wilcoxon Rank Sum test

- We have M=m+n observations in two groups:
  - X ($x_1.x_2...x_m$). Y ($y_1.y_2...y_n$) denotes the variables.
- We suppose:
  - $x_j=\varepsilon_i$ i=1,2,..m
  - $y_j=\Delta+\varepsilon_{m+j}$, j=1,2,..., n
  - x_j,y_j are the observed frequencies.
- H_0: $\Delta = 0$
- H_a:= $\Delta > 0$

Wilcoxon Rank-Sum Test

- Sort in ascending order the total M observations
  - (Merge the two groups).
- If R_j denotes the ranks of y_j then calculate the sum of R_js.

$$W = \sum_{j=1}^{n} R_j$$

- Test statistics (z) is approximately N(0,1) distributed for large M:

$$z = W^* = \frac{W - E(W)}{D(W)} = \frac{W - n(m + n + 1)/2}{(mn(n + m + 1)/12)^{1/2}}$$

27

Example

- We have the following measurements of serum triglyceride level in two groups:
- Control (X; m=6) :
  - 1.29 1.60 2.27 1.31 1.81 2.21
- Treated (Y; n=3):
  - 0.96 1.14 1.59
- Conbine them and assign the ranks:

Example

- Conbine them and assign the ranks:
  - X: 1.29 1.31 1.60 1.81 2.21 2.27
  - Y: 0.96 1.14 1.59
  - R: 1 2 3 4 5 6 7 8 9
- W= 1+2+5=8
- Critical interval for W is [7-23] at $\alpha=0.05$. Thus, we accept H_0.

STATA Results of Wilcoxon ranksum test

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

   group |      obs    rank sum    expected
---------+---------------------------------
 control |        6          37          30
 treated |        3           8          15
---------+---------------------------------
combined |        9          45          45

unadjusted variance          15.00
adjustment for ties           0.00
                        ----------
adjusted variance            15.00

Ho: data(group_==0) = data(group==1)
             z =   1.807
    Prob > |z| =   0.0707
```

EXAMPLE

- After a randomised trial comparing aspririn with placebo for hadache, 8 patients on aspirin and 10 on placebo rated their improvement on a 10 cm kine. A measure of 0 indicating no improvement and one of 10 indicating very much better.

Data

Group	Improvement
Aspirin	7.5
Aspirin	8.3
Aspirin	9.1
Aspirin	6.2
Aspirin	5.4
Aspirin	8.3
Aspirin	6.5
Aspirin	8.4
Placebo	3.1
Placebo	5.6
Placebo	4.5
Placebo	6.2
Placebo	5.1
Placebo	5.3
Placebo	5.5
Placebo	4.1
Placebo	4.3
Placebo	4.2

32

Stata results

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

mw_group |      obs    rank sum    expected
---------+---------------------------------
 Aspirin |        8       112.5          76
 Placebo |       10        58.5          95
---------+---------------------------------
combined |       18         171         171

unadjusted variance      126.67
adjustment for ties       -0.26
                     ----------
adjusted variance        126.41

Ho: improvem(mw_group==Aspirin) = improvem(mw_group==Placebo)
             z =   3.246
    Prob > |z| =   0.0012
```


33

Kruskall-Wallis test

- We have more than two groups.
- (Non-parametric independent two-group comparisons)
- Definition: A non-parametric test (distribution-free) used to compare more than two independent groups of sampled data.
- Test: The hypotheses for the comparison of independent groups are:

- H_0: The samples of all groups come from identical populations
- H_a: The samples of all groups come from different populations

 34

Kruskall-Wallis test

1	2	…	i	…	k
X_{11}	X_{12}	…	X_{1i}	…	X_{1k}
X_{21}	X_{22}	…	X_{2i}	…	X_{2k}
	$Xn_{2}2$				
			$Xn_{i}i$		
					$Xn_{k}k$
$Xn_{1}1$					

 35

Kruskall-Wallis test

- $x_{jj}=\mu+\tau_i+\varepsilon_{ij}$, j=1,2,..., n_i, i=1,2,..., k and $N=\Sigma n_i$. (i=1,2, ...k)
  - where μ is the unknown expected value
  - τ_i is the effect of ith treatment.
- H_0: $\tau_1 = \tau_2 = ...= \tau_k$
- H_A: $\tau_o \neq \tau_p$, there is at least one group differs from others.

Kruskall-Wallis test

- Combine and sort all x_{ij} values in ascending order. r_{ij} denotes the rank of x_{ij}.
- We know:

$$\sum_{i=1}^{k} R_i = \frac{N(N+1)}{2}$$

$$R_i = \sum_{j=1}^{n_i} r_{ji}$$

$$R_{.i} = \frac{R_i}{n_i}$$

$$R_{..} = \frac{\sum_{i=1}^{k} R_i}{N} = \frac{N+1}{2}$$

37

Test Statistics

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i (R_{.i} - R_{..})^2 = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

- H statistics is approximately chi-square distributed with k-1 degrees of freedom

 38

Example

- We have results of three treatments

A	B	C
6.4	2.5	1.3
6.8	3.7	4.1
7.2	4.9	4.9
8.3	5.4	5.2
8.4	5.9	5.5
9.1	8.1	8.2
9.4	8.2	
9.7		

 39

Assign ranks

A	B	C
11	2	1
12	3	4
13	5.5	5.5
17	8	7
18	10	9
19	14	15.5
20	15.5	
21		
131	58	42

	A	B	C
average of ranks	16.4	8.3	7.0

$$H = \frac{12}{21(21+1)}\left(\frac{131^2}{8} + \frac{58^2}{7} + \frac{42^2}{6}\right) - 3(21+1) = 9.84$$

STATA Result for Kruskal-Wallis test

- Test: Equality of populations (Kruskal-Wallis Test)

Groups	_Obs	_RankSum
1	8	131.00
2	7	58.00
3	6	42.00

- chi-squared = 9.836 with 2 d.f.
- probability = 0.0073

Spearman's rank correlation coefficient

- The rank correlation coefficient is the Pearson correlation coefficient based on the ranks of the data if there are no ties (adjustments are made if some of the data are tied). If the original data for each variable have no ties. the data for each variable are first ranked. and then the Pearson correlation coefficient between the ranks for the two variables is computed. Like Pearson correlation coefficient. the rank correlation ranges between -1 and +1. where -1 and +1 indicate a perfect linear relationship between the ranks of the two variables. The interpretation is therefore the same except that the relationship between ranks. and not values. is examined

42

Ranks of the 1.sample	Ranks of the 2.sample	Difference
r_1	q_1	$d_1=r_1-q_1$
r_2	q_2	$d_2=r_2-q_2$
...	...	...
r_n	q_n	$d_n=r_n-q_n$

 43

Test statistics

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^{\,2}}{n^3 - n}$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

44

The t-test

- H_0: correlation coefficient in population = 0, in notation: $\rho = 0$
- H_a: $\rho \neq 0$
- This test can be carried out by expressing the t statistic in terms of r. It can be proven that the statistic has t-distribution with n-2 degrees of freedom

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}} = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$

- **Decision using statistical table**: If t_{table} denotes the value of the table corresponding to n-2 degrees of freedom and probability,
  - if $|t| > t_{table}$, we reject H_0 and state that the population correlation coefficient, ρ is different from 0.
- **Decision using p-value**: if $p < \alpha$ (=0.05) we reject H_0 and state that the population correlation coefficient, ρ is different from 0

Example for Spaerman rank correlation

- The effectiveness of a treatment was measured on a scale between 0-12.
- The scores were determined by both the patients and doctors.
- Is there any relationship between the patients’ and doctors’ scores?

 46

Data

patient	doctor
2	1.5
10	9.1
7.1	8.1
2.3	1.5
3	3.1
4.1	5.2
10	1
10.5	9.6
11.0	7.6
12	9

47

The results

patients	doctors	Rank (patients')	Rank (doctor)	difference	d_i^2
2	1.5	1	2.5	-1.5	2.25
10	9.1	6.5	9	-2.5	6.25
7.1	8.1	5	7	-2	4
2.3	1.5	2	2.5	-0.5	0.25
3	3.1	3	4	-1	1
4.1	5.2	4	5	-1	1
10	1	6.5	1	5.5	30.25
10.5	9.6	8	10	-2	4
11.0	7.6	9	6	3	9
12	9	10	8	2	4

 48

Results

- H_0: correlation coefficient in population = 0, in notation: ρ =0
- H_a: ρ ≠ 0

$$r_s = 1 - \frac{6 * \sum_{i=1}^{n} d_i}{n^3 - n} = 1 - \frac{6 * 62}{1000 - 10} = 0.62$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.6242\sqrt{10-2}}{\sqrt{1-0.6242^2}} = 2.26$$

STATA results

- Number of obs = 10
- Spearman's rho = 0.6220
-
- Test of Ho: patient and doctor independent
- Pr > |t| = 0.0549

Jonckheere-Terpstra Test (JP)

- The Jonckheere-Terpstra test. which is a nonparametric test for ordered differences among classes.
- It tests the null hypothesis that the distribution of the response variable does not differ among classes.
- It is designed to detect alternatives of ordered class differences. which can be expressed as (or). with at least one of the inequalities being strict. where denotes the effect of class *i*.
- For such ordered alternatives. the Jonckheere-Terpstra test can be preferable to tests of more general class difference alternatives. such as the Kruskal - Wallis test.
- The Jonckheere-Terpstra test is appropriate for a contingency table in which an ordinal column variable represents the response. The row variable. which can be nominal or ordinal. represents the classification variable. The levels of the row variable should be ordered according to the ordering you want the test to detect

Jonckheere-Terpstra statistics

- The Jonckheere-Terpstra test statistic is computed by first forming $R(R$-1)/2 Mann-Whitney counts $M_{i.i'.}$ where $i < i'$. for pairs of rows in the contingency table .

 52

Null and alternative hypothesis

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_k$$

$$H_A : \tau_1 \leq \tau_2 \leq \ldots \leq \tau_k$$

 53

Test statistics

$$T_{uv} = \sum_{i=1}^{n_u} \sum_{i'=1}^{n_v} \delta(X_{iu}, X_{i'v})$$

$$\delta = \begin{cases} 1, & \text{if} \quad a < b \\ \frac{1}{2}, & \text{if} \quad a = b \\ 0, & \text{otherwise} \end{cases}$$

$$J = \sum_{u<v} T_{uv} = \sum_{u=1}^{k-1} \sum_{v=1}^{k} T_{uv}$$

54

Example:Do five different chemotherapy methods differ significantly in treatment response?

- A small pilot study was performed with five chemotherapy regimens: Cytoxan (CTX) alone, Cyclohexyl-chloreoethyl nitrosourea (CCNU) alone, Methotrexate (MTX) alone, CTX and MTX together, and CTX, CCNU, and MTX together. Tumor regression was measured on a three-point scale: no response, partial response, and complete response. The results are displayed in the following Table.

Example

No. of Patients

Chemo	No Response	Partial Response	Complete Response
CTX	2	0	0
CCNU	1	1	0
MTX	3	0	0
CTX+CCNU	2	2	0
CTX+CCNU+MTX	1	1	4

 56

Ranks

	No. of Patients		
Chemo	No Response	Partial Response	Complete Response
CTX	12	3,5	3,5
CCNU	8,5	8,5	3,5
MTX	14	3,5	3,5
CTX+CCNU	12	12	3,5
CTX+CCNU+MTX	8,5	8,5	15

 57

Test statistics

$$T_{12} = \sum_{i=1}^{n_1} \sum_{i'=1}^{n_2} \delta\,(X_{i1}, X_{i'2}) =$$

$$\delta\,(2{,}0) + \delta\,(2{,}1) + \delta\,(2{,}0) + \delta\,(2{,}2) + \delta\,(2{,}1) +$$

$$\delta\,(1{,}0) + \delta\,(1{,}1) + \delta\,(1{,}0) + \delta\,(1{,}2) + \delta\,(1{,}1) + \ldots$$

$$+ \delta\,(1{,}0) + \delta\,(1{,}1) + \delta\,(1{,}0) + \delta\,(1{,}2) + \delta\,(1{,}1) = 20$$

$$T_{13} = 20 \qquad T_{23} = 16$$

$$J = 56 \qquad J_{critical} = 54$$

 58