Non Parametric Tests: Hands on SPSS

N. Uttam Singh, Aniruddha Roy & A. K. Tripathi
ICAR Research Complex for NEH Region, Umiam, Meghalaya
uttamba@gmail.com, aniruddhaubkv@gmail.com, aktripathi2020@yahoo.co.in

They are called nonparametric because they make no assumptions about the parameters (such as the mean and variance) of a distribution, nor do they assume that any particular distribution is being used.

Introduction

A **parametric** statistical test is one that makes assumptions about the parameters (defining properties) of the population distribution(s) from which one's data are drawn.

A **non-parametric** test is one that makes no such assumptions. In this strict sense, "non-parametric" is essentially a null category, since virtually all statistical tests assume one thing or another about the properties of the source population(s).

Which is more powerful?

Non-parametric statistical procedures are less powerful because they use less information in their calculation. For example, a parametric correlation uses information about the mean and deviation from the mean while a non-parametric correlation will use only the ordinal position of pairs of scores.

Parametric Assumptions

- The observations must be independent
- The observations must be drawn from normally distributed populations
- These populations must have the same variances
- The means of these normal and homoscedastic populations must be linear combinations of effects due to columns and/or rows

Nonparametric Assumptions

Certain assumptions are associated with most nonparametric statistical tests, but these are fewer and weaker than those of parametric tests.

Advantages of Nonparametric Tests

- Probability statements obtained from most nonparametric statistics are exact probabilities, regardless of the shape of the population distribution from which the random sample was drawn
- If sample sizes as small as N=6 are used, there is no alternative to using a nonparametric test
- Easier to learn and apply than parametric tests
- Based on a model that specifies very general conditions.
- No specific form of the distribution from which the sample was drawn.
- Hence nonparametric tests are also known as distribution free tests.

Disadvantages of nonparametric tests

- Losing precision/wasteful of data
- Low power
- False sense of security
- Lack of software
- Testing distributions only
- Higher-ordered interactions not dealt with
- Parametric models are more efficient if data permit.
- It is difficult to compute by hand for large samples
- Tables are not widely available
- In cases where a parametric test would be appropriate, non-parametric tests have less power. In other words, a larger sample size can be required to draw conclusions with the same degree of confidence.

Few points

- The inferences drawn from tests based on the parametric tests such as t, F and Chi-square may be seriously affected when the parent population's distribution is not normal.
- The adverse effect could be more when sample size is small.
- Thus when there is doubt about the distribution of the parent population, a nonparametric method should be

used.

- In many situations, particularly in social and behavioral sciences, observations are difficult or impossible to take on numerical scales and a suitable nonparametric test is an alternative under such situations.

Measurement

The 4 levels of measurement

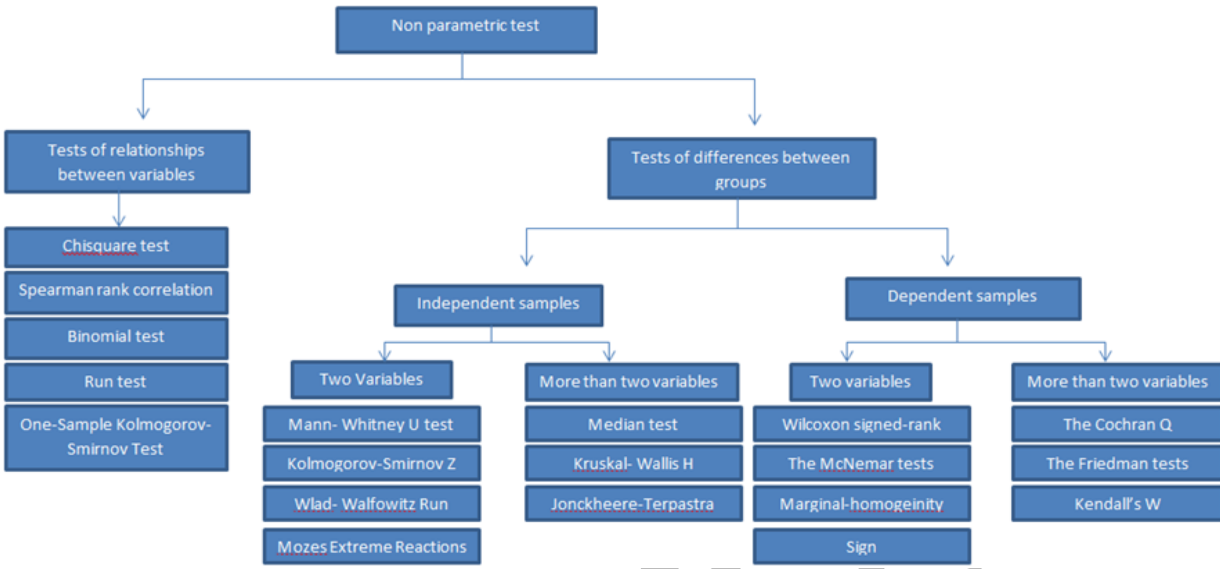
1. Nominal or Classificatory Scale
   - Gender, ethnic background, colors of a spectrum
   - In research activities a YES/NO scale is nominal. It has no order and there is no distance between YES and NO.
2. Ordinal or Ranking Scale
   - Hardness of rocks, beauty, military ranks
   - The simplest ordinal scale is a ranking.
   - There is no objective distance between any two points on your subjective scale.
3. Interval Scale
   - Celsius or Fahrenheit. It is an interval scale because it is assumed to have equidistant points between each of the scale elements.
4. Ratio Scale
   - Kelvin temperature, speed, height, mass or weight
   - **Ratio** data is interval data with a natural zero point

Parametric vs. non-parametric tests

	Parametric	**Non-parametric**
Assumed distribution	Normal	Any
Assumed variance	Homogeneous	Any
Typical data	Ratio or Interval	Ordinal or Nominal
Data set relationships	Independent	Any
Usual central measure	Mean	Median
Benefits	Can draw more conclusions	Simplicity; Less affected by outliers
Tests		
Choosing	Choosing parametric test	Choosing a non-parametric test
Correlation test	Pearson	Spearman
Independent measures, 2 groups	Independent-measures t-test	Mann-Whitney test
Independent measures, >2 groups	One-way, independent-measures ANOVA	Kruskal-Wallis test
Repeated measures, 2 conditions	Matched-pair t-test	Wilcoxon test
Repeated measures, >2 conditions	One-way, repeated measures ANOVA	Friedman's test

Nonparametric Methods

There is at least one nonparametric test equivalent to a parametric test



Tests of relationships between variables

Chi-square Test

This goodness-of-fit test compares the observed and expected frequencies in each category to test either that all categories contain the same proportion of values or that each category contains a user-specified proportion of values.

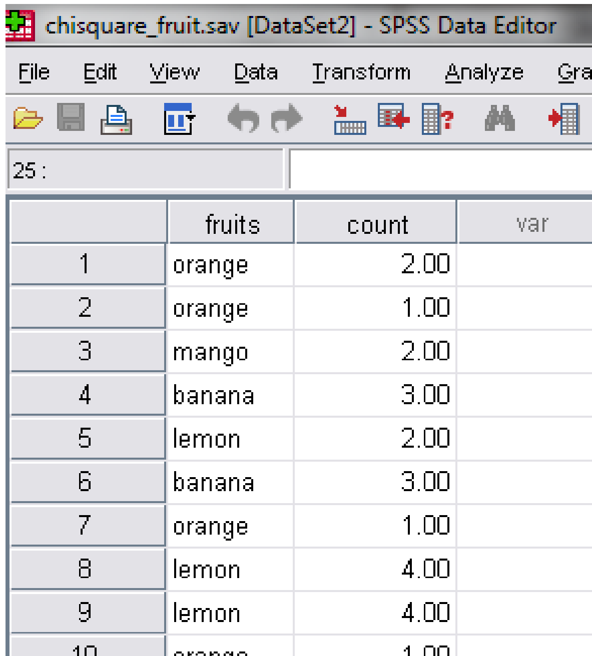
Examples

The chi-square test could be used to determine if a basket of fruit contains equal proportions of apples, bananas, oranges, and peaches.

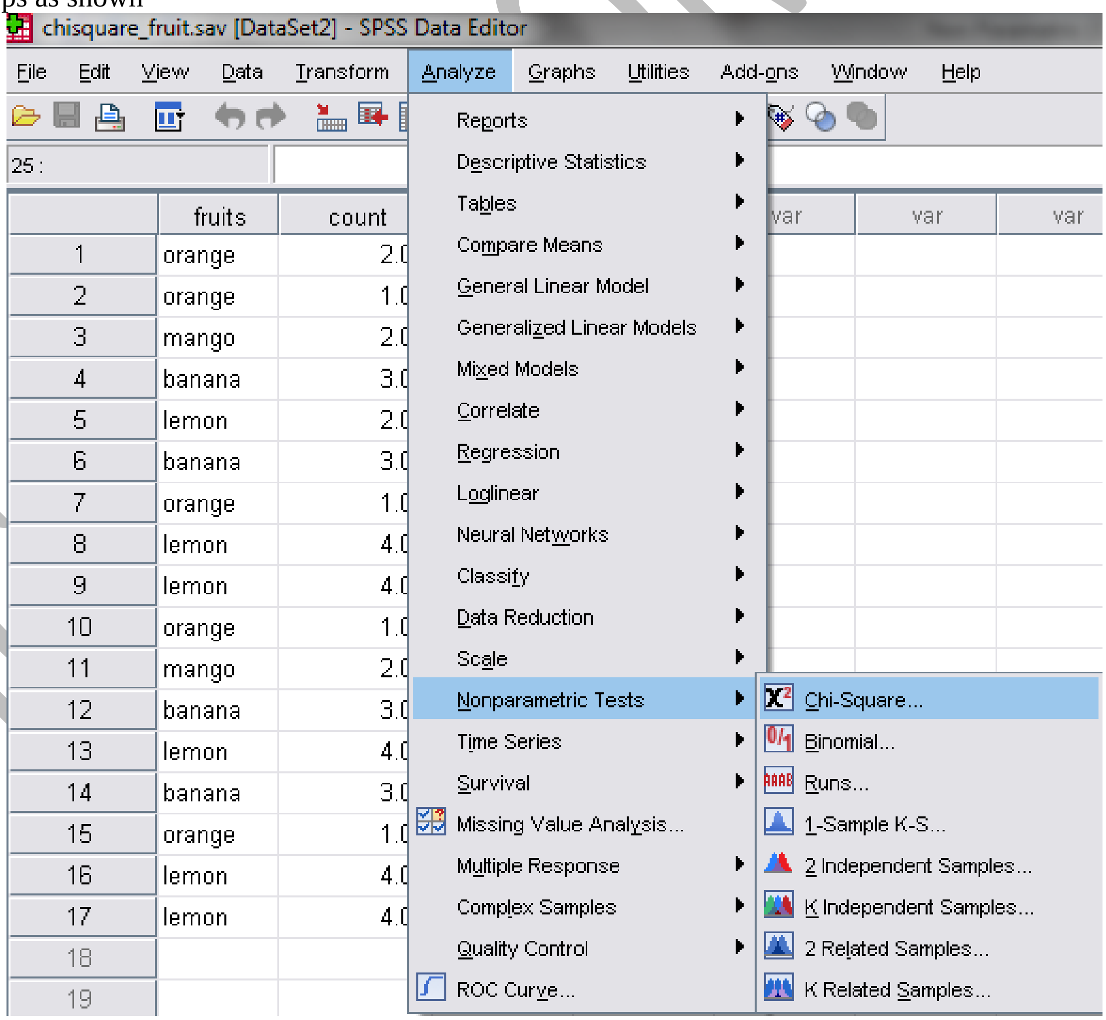
fruits	count
orange	1
orange	1
mango	2
banana	3
lemon	4
banana	3
orange	1
lemon	4
lemon	4
orange	1
mango	2
banana	3
lemon	4
banana	3
orange	1
lemon	4
lemon	4

SPSS Steps:

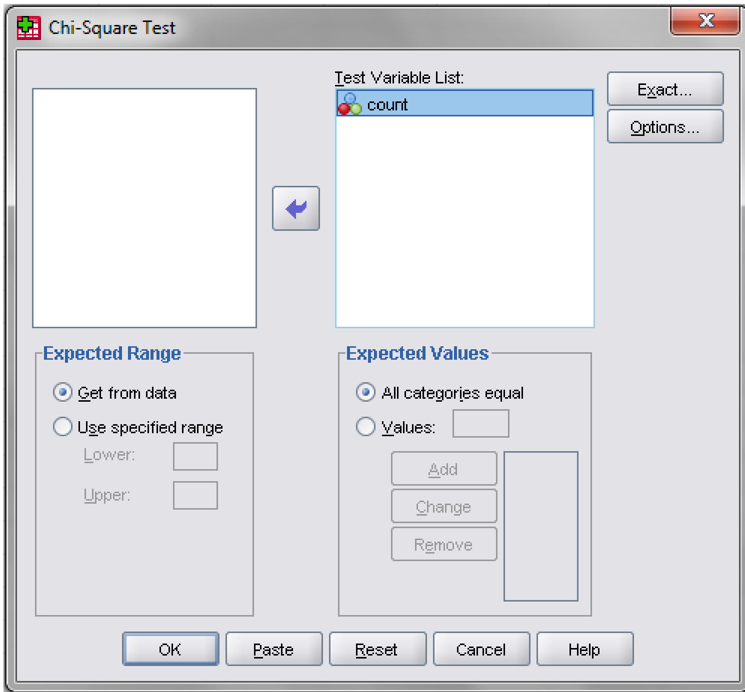
Get the data.



Follow the steps as shown



Get the count in the test variable list



Click OK and get the output as shown below

count

	Observed N	Expected N	Residual
1	4	4.2	-.2
2	4	4.2	-.2
3	4	4.2	-.2
4	5	4.2	.8
Total	17		

Test Statistics

	count
Chi-Square	.176[a]
df	3
Asymp. Sig.	.981

a. 4 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 4.3.

Interpretation:

Here p value is 0.981 which is more than 0.05. Hence it is not significant and we fail to reject the null hypothesis and conclude that there is no significant difference in the proportions of apples, bananas, oranges, and peaches.

We could also test to see if a basket of fruit contains 10% apples, 20% bananas, 50% oranges, and 20% peaches. For this we have to define the proportions by checking the button "Values" and keep on adding.

Binomial Test

The Binomial Test procedure is useful when you want to compare a single sample from a dichotomous variable to an expected proportion. If the dichotomy does not exist in the data as a variable, one can be dynamically created based upon a cut point on a scale variable (take age as example from the data). If your variable has more than two outcomes, try the Chi-Square Test procedure. If you want to compare two dichotomous variables, try the McNemar test in the Two-Related-Samples Tests procedure.

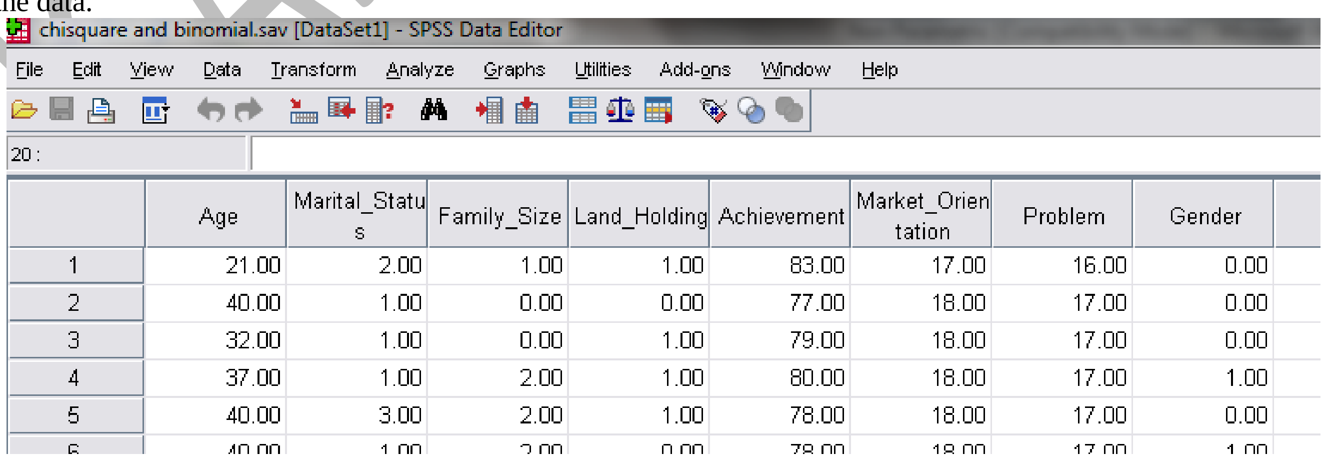
Example

Say we wish to test whether the proportion of females from the variable "gender" differs significantly from 50%, i.e., from 0.5. We will use the **exact** statement to produce the exact p-values.

Age	Marital_Status	Family_Size	Land_Holding	Achievement	Market_Orientation	Problem	Gender
21	2	1	1	83	17	16	0
40	1	0	0	77	18	17	0
32	1	0	1	79	18	17	0
37	1	2	1	80	18	17	1
40	3	2	1	78	18	17	0
40	1	2	0	78	18	17	1
52	1	0	0	79	24	13	0
35	2	2	1	94	24	20	1
38	2	2	1	81	22	12	0
55	1	0	1	78	18	10	1
35	2	1	0	87	23	17	1
35	3	2	1	89	22	10	0
55	1	1	0	87	23	15	0
40	1	2	1	86	23	14	1
62	1	1	1	80	18	10	1
40	1	1	0	83	24	13	1
48	3	1	1	76	21	14	1
62	1	2	1	84	23	11	0
36	1	0	0	81	26	11	0
35	1	2	1	80	21	11	0
35	1	2	1	77	22	13	1
35	1	1	1	82	16	14	1
18	2	2	0	83	26	10	0

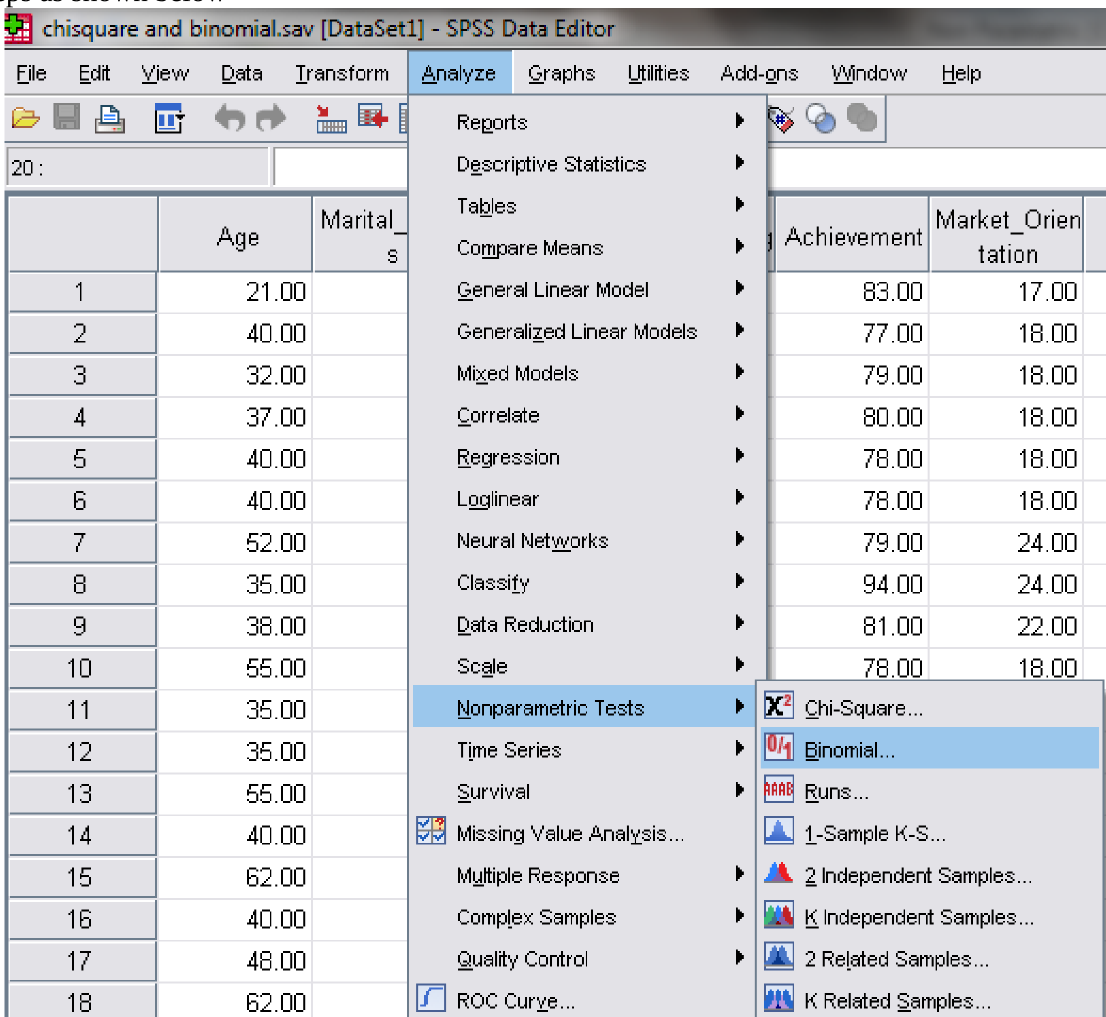
SPSS Steps:

Get the data.

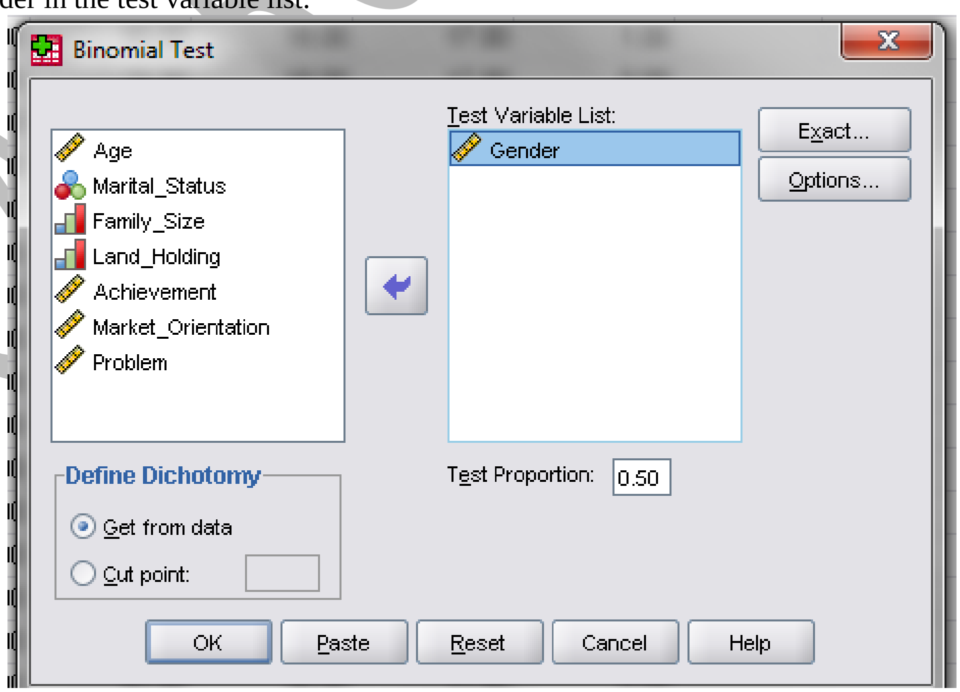


	Age	Marital_Status	Family_Size	Land_Holding	Achievement	Market_Orientation	Problem	Gender
1	21.00	2.00	1.00	1.00	83.00	17.00	16.00	0.00
2	40.00	1.00	0.00	0.00	77.00	18.00	17.00	0.00
3	32.00	1.00	0.00	1.00	79.00	18.00	17.00	0.00
4	37.00	1.00	2.00	1.00	80.00	18.00	17.00	1.00
5	40.00	3.00	2.00	1.00	78.00	18.00	17.00	0.00

Follow the steps as shown below



Get the variable gender in the test variable list.



Click OK and get the output

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
Gender	Group 1	.00	12	.52	.50	1.000
	Group 2	1.00	11	.48		
	Total		23	1.00		

Interpretation:

Since p value is 1 it is not significant and we fail to reject null hypothesis and conclude that the proportion of females from the variable "gender" does not differ significantly from 50%.

Run Test for Randomness

Run test is used for examining whether or not a set of observations constitutes a random sample from an infinite population. Test for randomness is of major importance because the assumption of randomness underlies statistical inference. In addition, tests for randomness are important for time series analysis. Departure from randomness can take many forms. The cut point is based either on a measure of central tendency (mean, median, or mode) or a custom value. A sample with too many or too few runs suggests that the sample is not random.

Example

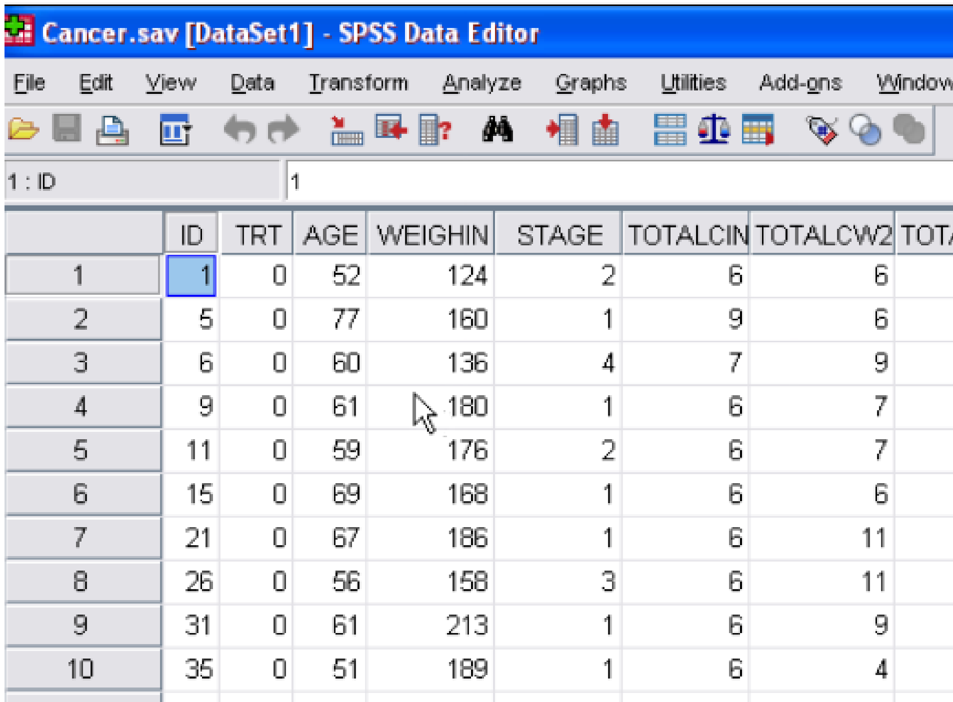
Let's see whether the variable "AGE" in the dataset below is random.

Table: Cancer dataset

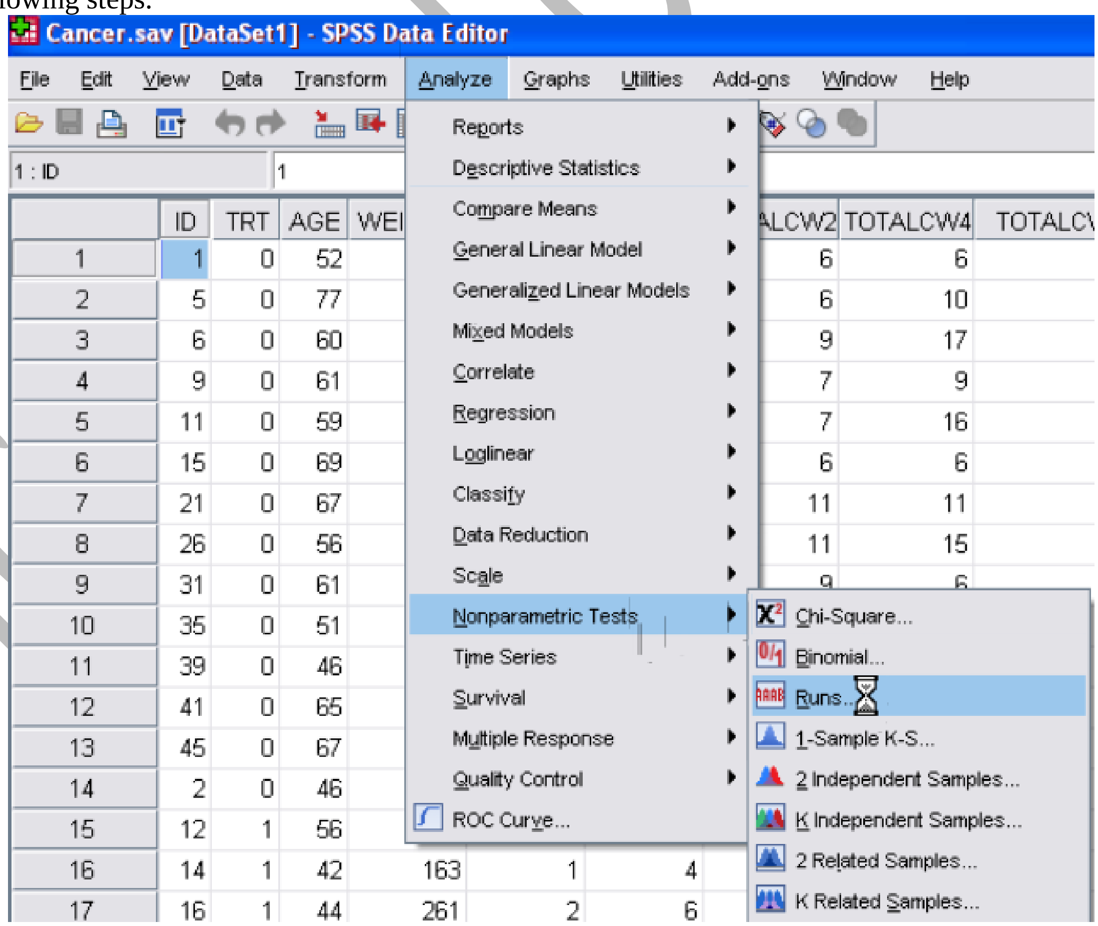
ID	TRT	AGE	WEIGHIN	STAGE	TOTALCIN	TOTALCW2	TOTALCW4	TOTALCW6
1	0	52	124	2	6	6	6	7
5	0	77	160	1	9	6	10	9
6	0	60	136.5	4	7	9	17	19
9	0	61	179.6	1	6	7	9	3
11	0	59	175.8	2	6	7	16	13
15	0	69	167.6	1	6	6	6	11
21	0	67	186	1	6	11	11	10
26	0	56	158	3	6	11	15	15
31	0	61	212.8	1	6	9	6	8
35	0	51	189	1	6	4	8	7
39	0	46	149	4	7	8	11	11
41	0	65	157	1	6	6	9	6
45	0	67	186	1	8	8	9	10
2	0	46	163.8	2	7	16	9	10
12	1	56	227.2	4	6	10	11	9
14	1	42	162.6	1	4	6	8	7
16	1	44	261.4	2	6	11	11	14
22	1	27	225.4	1	6	7	6	6
24	1	68	226	4	12	11	12	9
34	1	77	164	2	5	7	13	12
37	1	86	140	1	6	7	7	7
42	1	73	181.5	0	8	11	16	
44	1	67	187	1	5	7	7	7
50	1	60	164	2	6	8	16	
58	1	54	172.8	4	7	8	10	8

SPSS Steps:

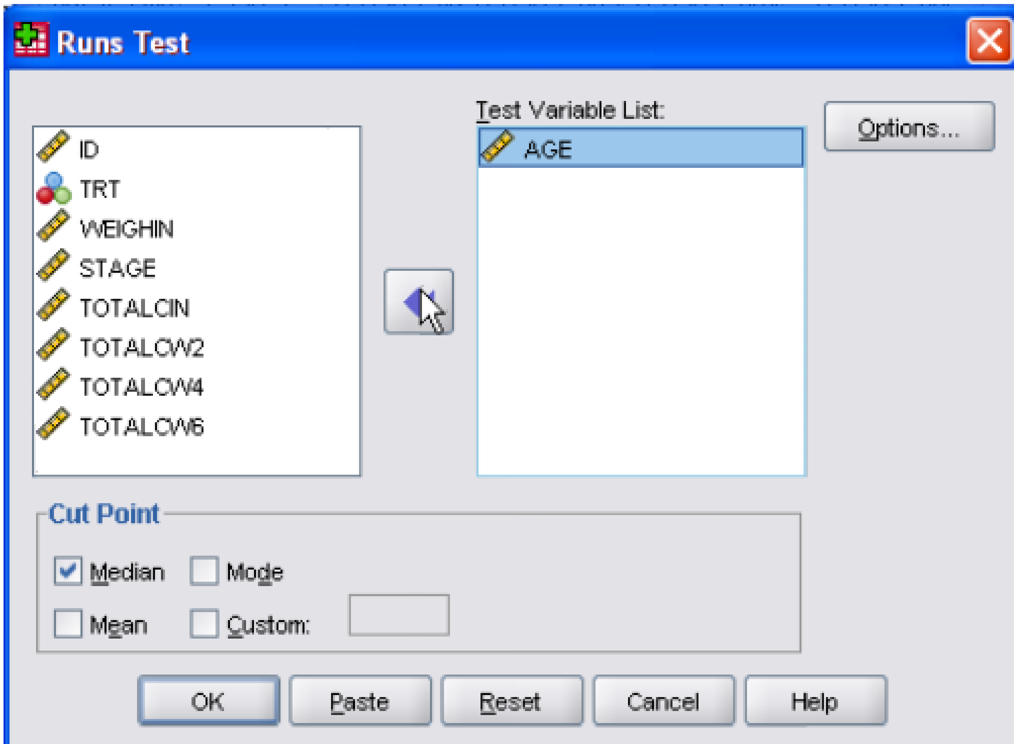
Load the data.



Follow the following steps.



Select “AGE” in the test variables list.



This variable “AGE” must be divided into two spate groups. Therefore we must indicate a cut point. Now lets take Median as the cut point. Any value blow the median point will belong to one group and any value greater than or equal to median will belong to the other group. Now click OK to get output.

Runs Test

	AGE
Test Value[a]	60
Cases < Test Value	11
Cases >= Test Value	14
Total Cases	25
Number of Runs	11
Z	-.755
Asymp. Sig. (2-tailed)	.450

a. Median

Interpretation:

Now p value is 0.450. So it is not significant and we cannot say that AGE is not random.

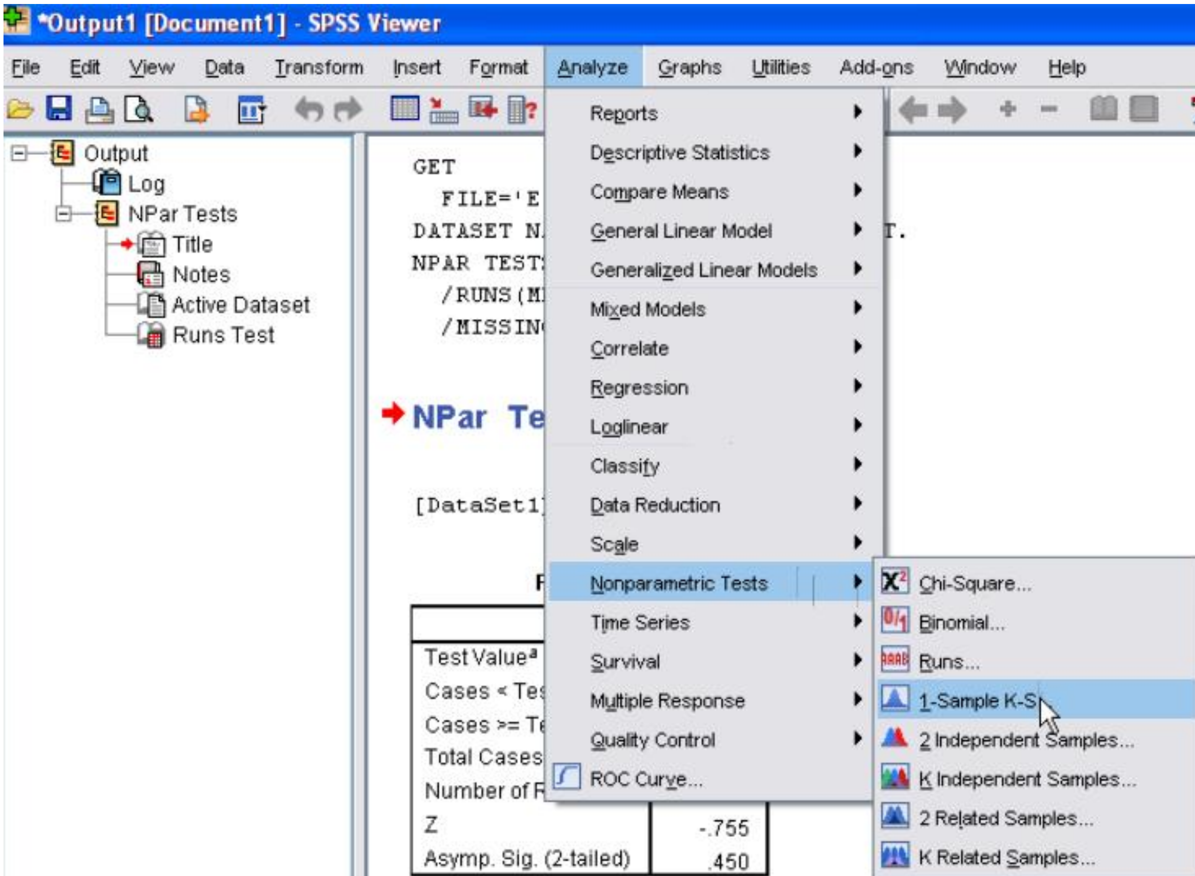
One-Sample Kolmogorov-Smirnov Test

The One-Sample Kolmogorov-Smirnov procedure is used to test the null hypothesis that a sample comes from a particular distribution. Four theoretical distribution functions are available-- normal, uniform, Poisson, and exponential. If we want to compare the distributions of two variables, use the two-sample Kolmogorov-Smirnov test in the Two-Independent-Samples Tests procedure.

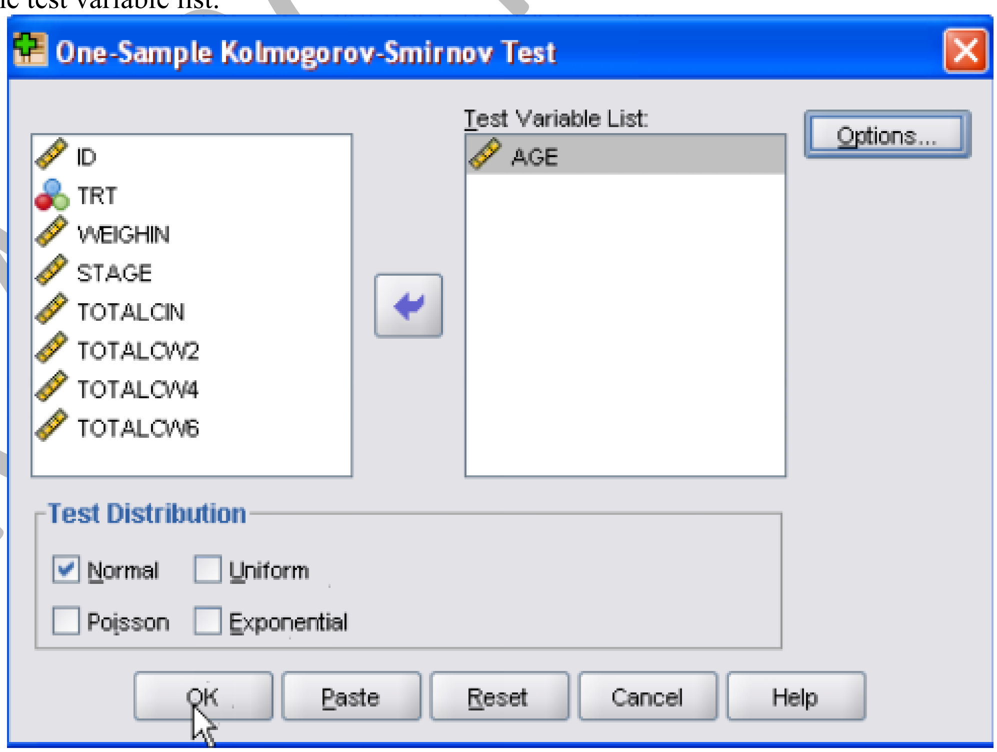
Example: Let us test the variable “AGE” in the cancer dataset used for Run test above is normal distribution or uniform distribution.

SPSS Steps

Get the data as done before. Then…



Select "AGE" in the test variable list.



Check the distribution for which you want to test. Click OK and get the output.

One-Sample Kolmogorov-Smirnov Test

		AGE
N		25
Normal Parameters[a]	Mean	59.64
	Std. Deviation	12.932
Most Extreme Differences	Absolute	.080
	Positive	.075
	Negative	-.080
Kolmogorov-Smirnov Z		.401
Asymp. Sig. (2-tailed)		.997

a. Test distribution is Normal.

Interpretation:

The p value is 0.997 which is not significant and therefore we cannot say that "AGE" does not have an approximate normal distribution. If the p value were less than 0.05 we would say it is significant and AGE does not follow an approximate normal distribution.

Two-Independent-Samples Tests

The nonparametric tests for two independent samples are useful for determining whether or not the values of a particular variable differ between two groups. This is especially true when the assumptions of the t test are not met.

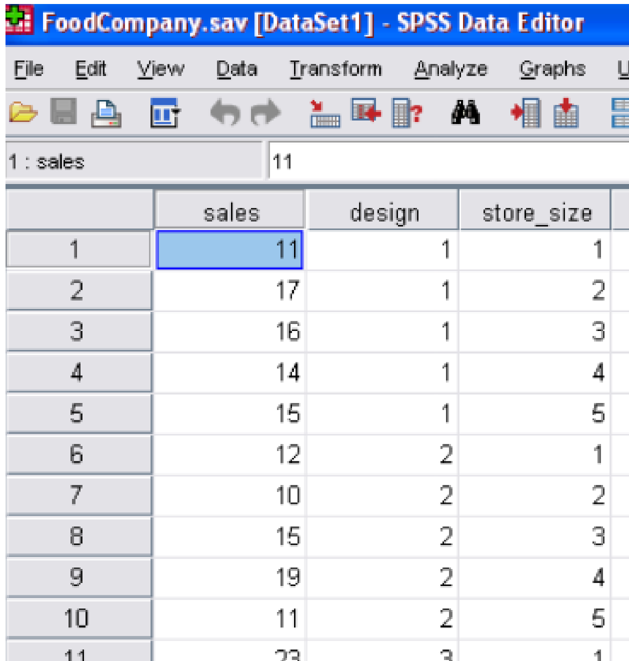
- **Mann-Whitney U test:** To test for differences between two groups
- **The two-sample Kolmogorov-Smirnov test:** To test the null hypothesis that two samples have the same distribution
- **Wlad-Walfowitz Run:** Used to examine whether two random samples come from populations having same distribution
- **Mozes Extreme Reactions:** Exact Test

Example: We want to find out whether the sales are different between two designs.

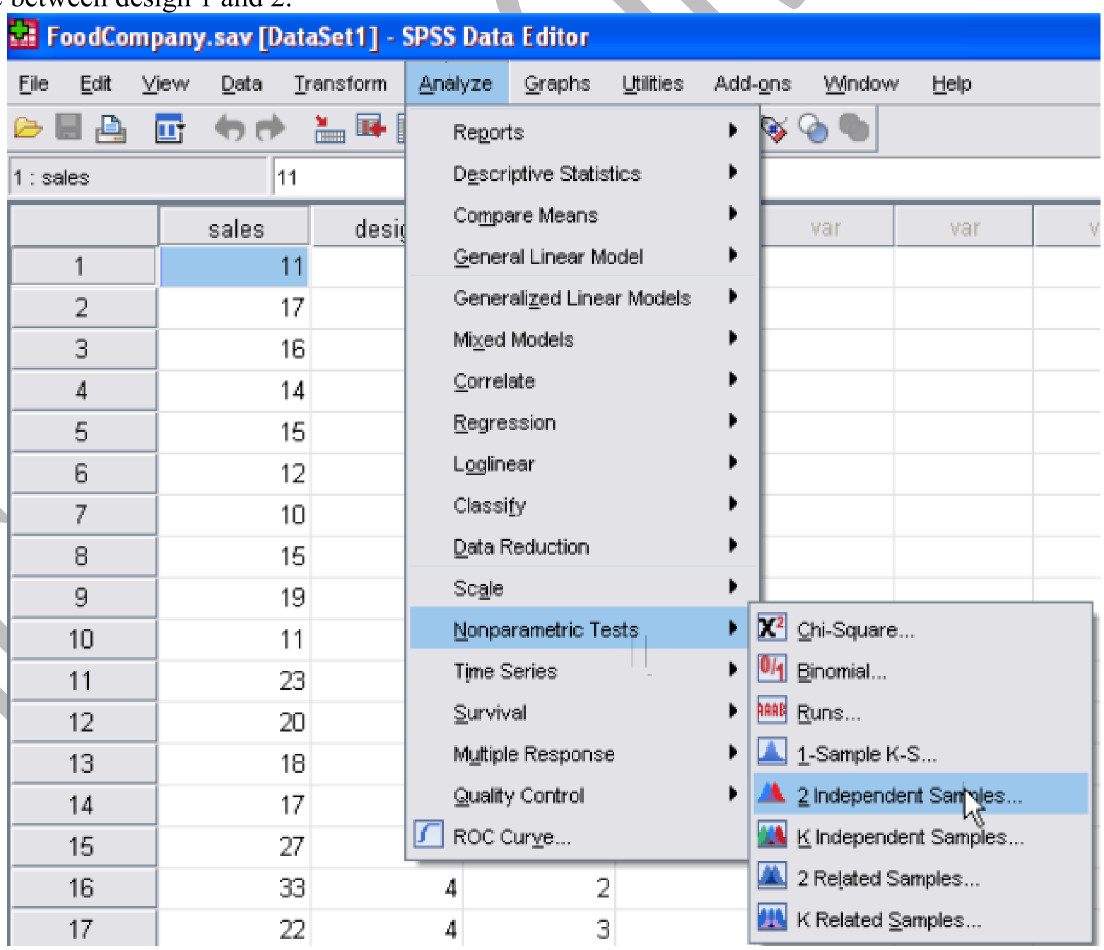
sales	design	store_size
11	1	1
17	1	2
16	1	3
14	1	4
15	1	5
12	2	1
10	2	2
15	2	3
19	2	4
11	2	5
23	3	1
20	3	2
18	3	3
17	3	4
27	4	1
33	4	2
22	4	3
26	4	4
28	4	5

SPSS Steps:

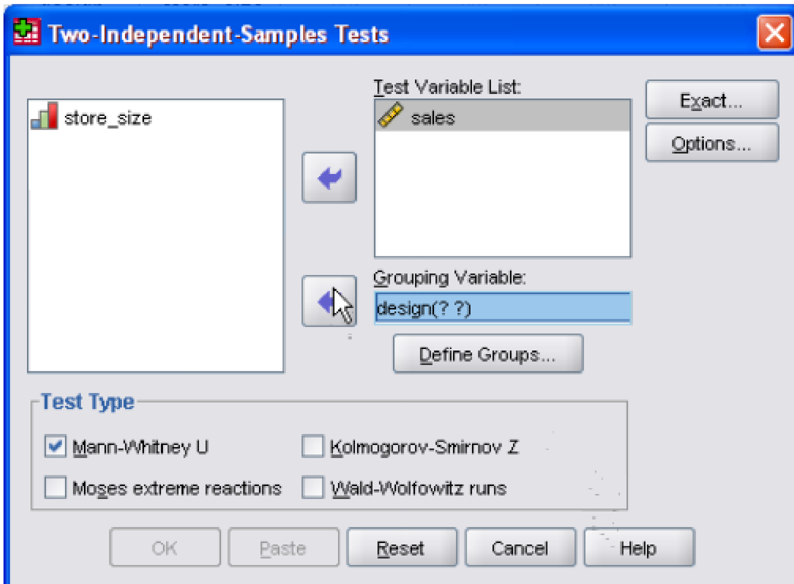
Open the dataset



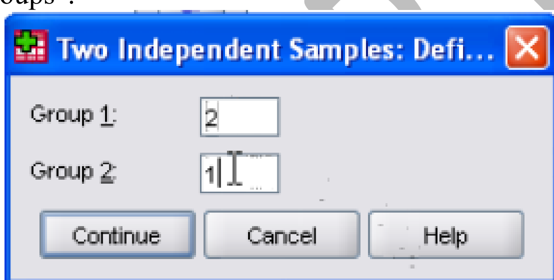
Let's compare between design 1 and 2.



Enter variable sales in test variable list and design in grouping variable.



Since we are performing two independent sample tests we have to designate which two groups in our factor design we want to compare. So click "Define groups".



Here we type group 2 and 1. Order is not important, only we have to enter two distinct groups. Then click continue and OK to get output.

Ranks

	de...	N	Mean Rank	Sum of Ranks
sales	1	5	6.20	31.00
	2	5	4.80	24.00
	Total	10		

Test Statistics[b]

	sales
Mann-Whitney U	9.000
Wilcoxon W	24.000
Z	-.736
Asymp. Sig. (2-tailed)	.462
Exact Sig. [2*(1-tailed Sig.)]	.548[a]

a. Not corrected for ties.

b. Grouping Variable: design

Interpretation:

Now two p values are displayed, asymptotic which is appropriate for large sample and exact which is independent of sample size. Therefore we will take the exact p value i. e. 0.548 which is not significant and we conclude that there is no significant difference in sales between the design group 1 and group 2.

Multiple Independent Samples Tests

The nonparametric tests for multiple independent samples are useful for determining whether or not the values of a particular variable differ between two or more groups. This is especially true when the assumptions of ANOVA are not met.

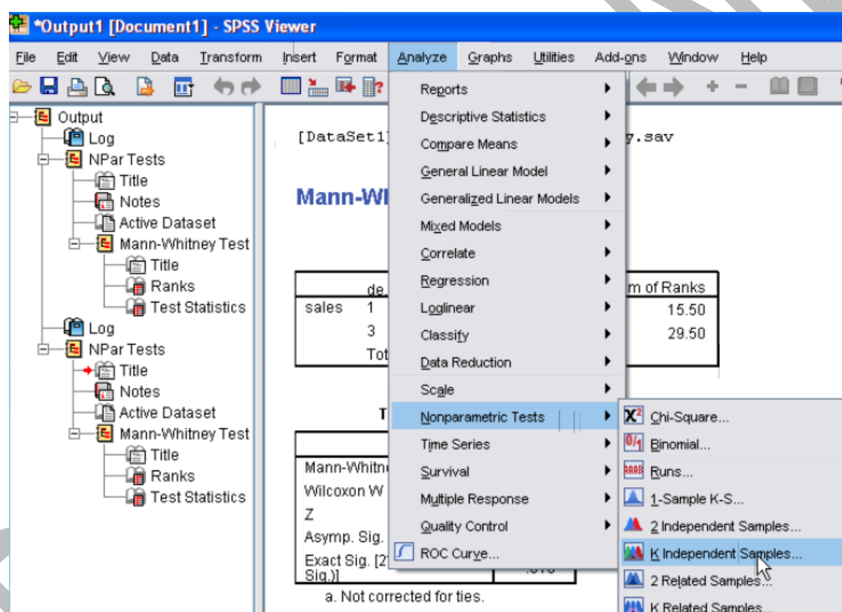
- **Median test:** This method tests the null hypothesis that two or more independent samples have the same median. It assumes nothing about the distribution of the test variable, making it a good choice when you suspect that the distribution varies by group
- **Kruskal-Wallis H:** This test is a one-way analysis of variance by ranks. It tests the null hypothesis that multiple independent samples come from the same population.
- **Jonckheere-terpstra test:** Exact test

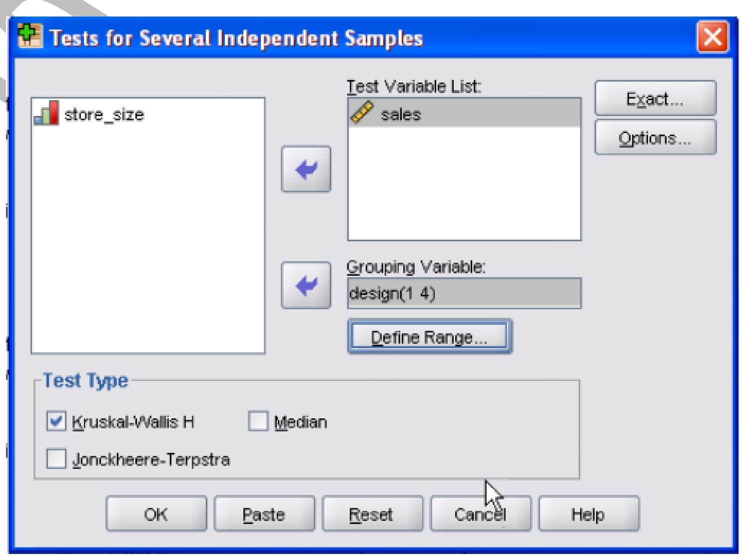
Example:

We want to find out whether the sales are different between the designs (Comparing more than two samples simultaneously)

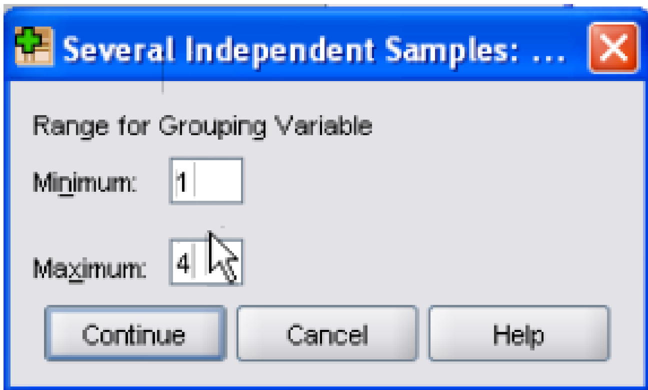
SPSS Steps:

Get the data in SPSS window as done before. Then…





Define range



Click continue then OK to get output.

Ranks

	de...	N	Mean Rank
sales	1	5	6.30
	2	5	5.20
	3	4	12.12
	4	5	16.80
	Total	19	

Test Statistics[a,b]

	sales
Chi-Square	13.707
df	3
Asymp. Sig.	.003

a. Kruskal Wallis Test

b. Grouping Variable: design

Interpretation:

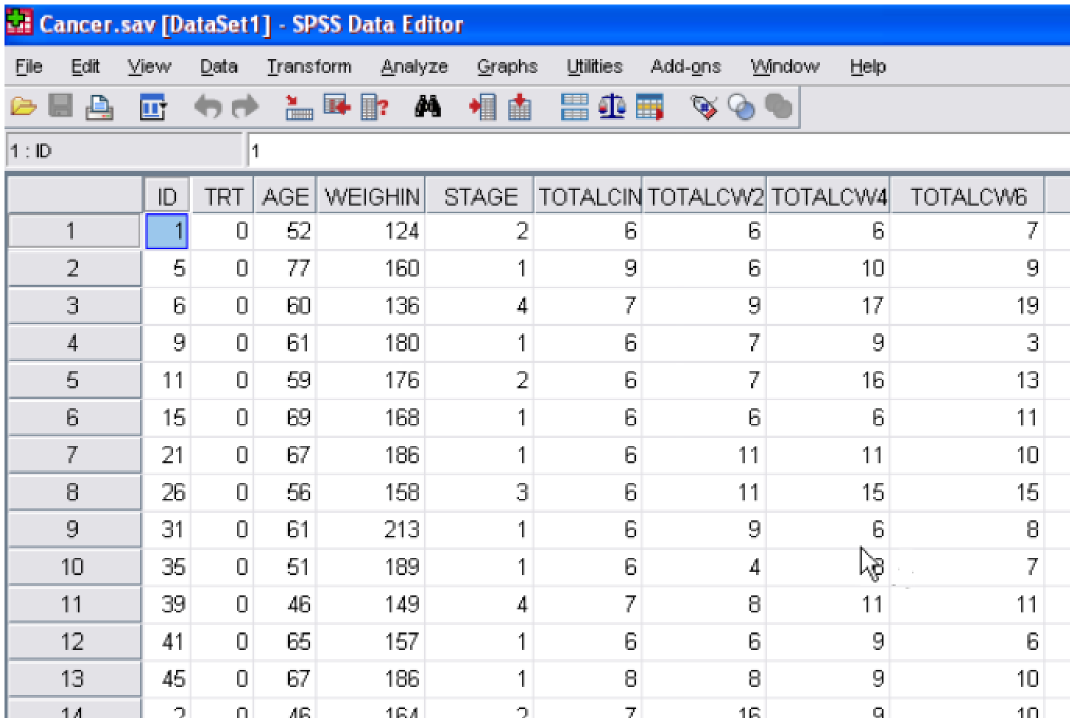
P value is 0.003 which is significant. Therefore we conclude that there is significant difference between the groups (meaning- at least two groups are different)

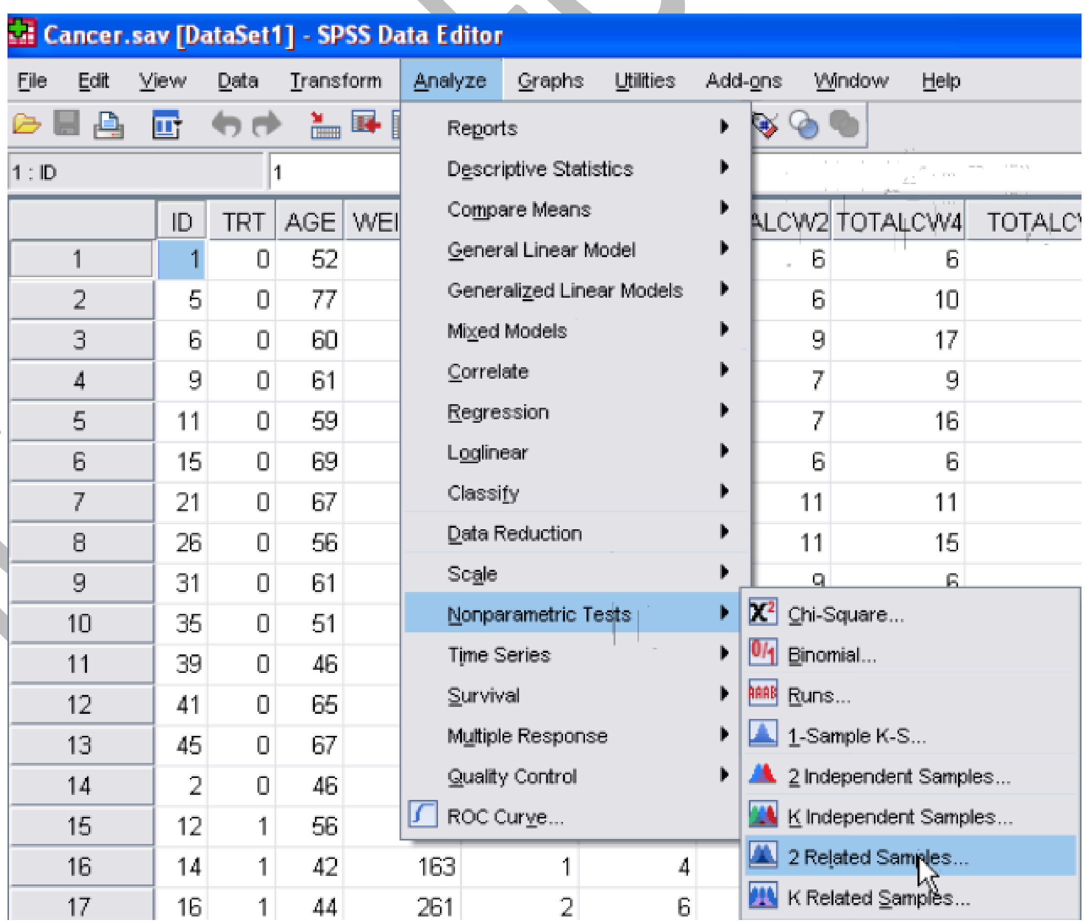
Tests for Two Related Samples

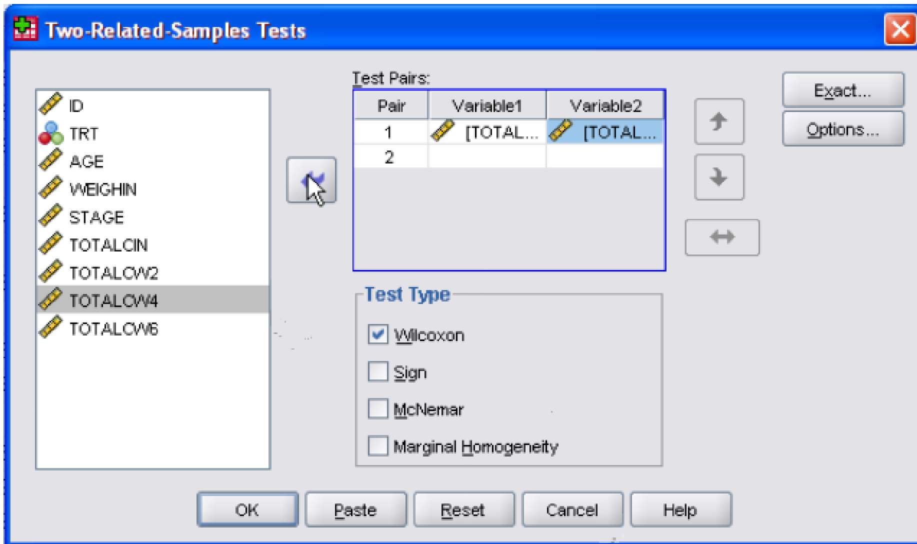
The nonparametric tests for two related samples allow you to test for differences between paired scores when you cannot (or would rather not) make the assumptions required by the paired-samples t test. Procedures are available for testing nominal, ordinal, or scale variables.

- **Wilcoxon signed-ranks:** A nonparametric alternative to the paired-samples t test. The only assumptions made by the Wilcoxon test are that the test variable is continuous and that the distribution of the difference scores is reasonably symmetric.
- **McNemar** method tests the null hypothesis that binary responses are unchanged. As with the Wilcoxon test, the data may be from a single sample measured twice or from two matched samples. The McNemar test is particularly appropriate with nominal or ordinal test variables for binary data. Unlike the Wilcoxon test, the McNemar test is designed for use with nominal or ordinal test variables.
- **Marginal-homogeinity:** If the varialbles are mortinomial i.e if they have more than two levels.
- **Sign test:** Wilkoxon and Sign are used for contineous data and of the two wilkoxon is more powerful

Example: Use the cancer data deployed in Run Test to test whether the condition of the cancer patient at the end of 2nd week and 4th week are significantly different. (here higher the reading, better is the condition)







Output:

Ranks

		N	Mean Rank	Sum of Ranks
TOTALCW4 - TOTALCW2	Negative Ranks	3[a]	9.17	27.50
	Positive Ranks	16[b]	10.16	162.50
	Ties	6[c]		
	Total	25		

a. TOTALCW4 < TOTALCW2

b. TOTALCW4 > TOTALCW2

c. TOTALCW4 = TOTALCW2

Test Statistics[b]

	TOTALCW4 - TOTALCW2
Z	-2.723[a]
Asymp. Sig. (2-tailed)	.006

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Interpretation:

P value is 0.006 which is significant. This indicates that the condition of cancer patient at the end of 2nd week and 4th week are different.

Tests for Multiple Related Samples

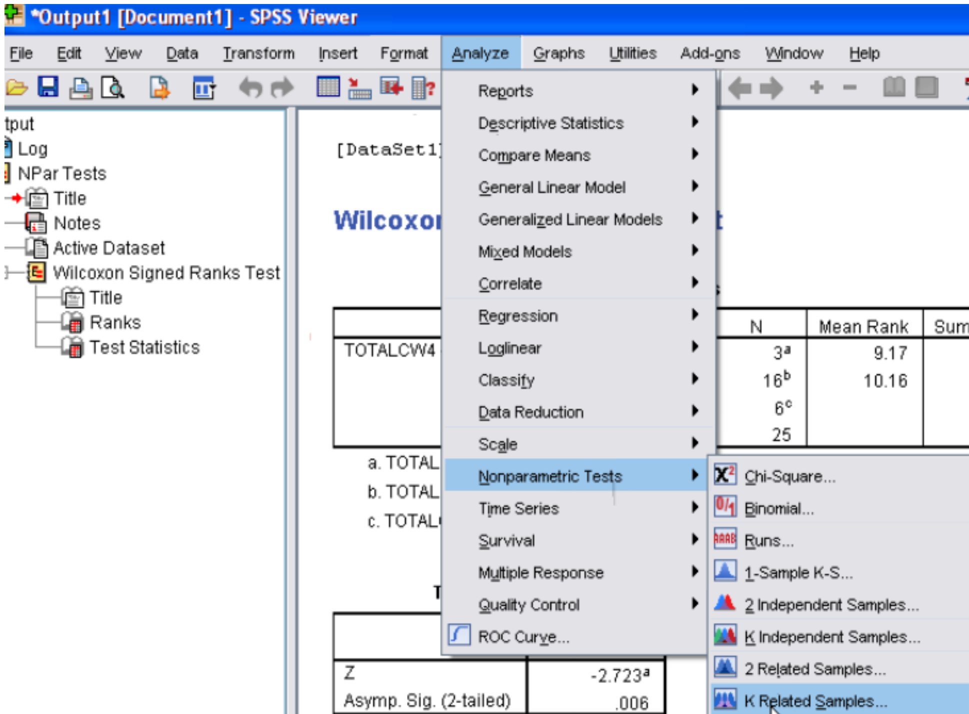
The nonparametric tests for multiple related samples are useful alternatives to a repeated measures analysis of variance. They are especially appropriate for small samples and can be used with nominal or ordinal test variables.

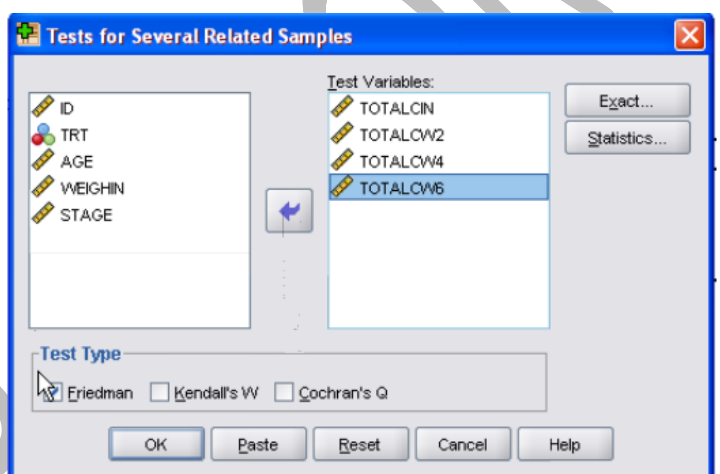
Friedman test is a nonparametric alternative to the repeated measures ANOVA. It tests the null hypothesis that multiple ordinal responses come from the same population. As with the Wilcoxon test for two related samples, the data may come from repeated measures of a single sample or from the same measure from multiple matched samples. The only assumptions made by the Friedman test are that the test variables are at least ordinal and that their distributions are reasonably similar.

Cochran's Q: It tests the null hypothesis that multiple related proportions are the same. Think of the Cochran Q test as an extension of the McNemar test used to assess change over two times or two matched samples. Unlike the Friedman test, the Cochran test is designed for use with binary variables.

Kendall's W: is a normalization of Friedman test and can be interpreted as a measure of agreement

SPSS steps:





Output

Ranks

	Mean Rank
TOTALCIN	1.48
TOTALCW2	2.43
TOTALCW4	3.22
TOTALCW6	2.87

Test Statistics[a]

N	23.000
Chi-Square	27.381
df	3.000
Asymp. Sig.	.000

a. Friedman Test

Interpretation:

P value is less than 0.05. Hence there is significant difference between the four groups (meaning- at least two groups are different)

Exact Tests and Monte Carlo Method

These new methods, the exact and Monte Carlo methods, provide a powerful means for obtaining accurate results when your data set is small, your tables are sparse or unbalanced, the data are not normally distributed, or the data fail to meet any of the underlying assumptions necessary for reliable results using the standard asymptotic method.

The Exact Method

By default, IBM® SPSS® Statistics calculates significance levels for the statistics in the Crosstabs and Nonparametric Tests procedures using the **asymptotic method**. This means that *p* values are estimated based on the assumption that the data, given a sufficiently large sample size, conform to a particular distribution.

However, when the data set is small, sparse, contains many ties, is unbalanced, or is poorly distributed, the asymptotic method may fail to produce reliable results. In these situations, it is preferable to calculate a significance level based on the exact distribution of the test statistic. This enables you to obtain an accurate *p* value without relying on assumptions that may not be met by your data.

The Monte Carlo Method

Although exact results are always reliable, some data sets are too large for the exact *p* value to be calculated, yet don't meet the assumptions necessary for the asymptotic method. In this situation, the Monte Carlo method provides an unbiased estimate of the exact *p* value, without the requirements of the asymptotic method.
The Monte Carlo method is a repeated sampling method. For any observed table, there are many tables, each with the same dimensions and column and row margins as the observed table. The Monte Carlo method repeatedly samples a specified number of these possible tables in order to obtain an unbiased estimate of the true *p* value.

The Monte Carlo method is less computationally intensive than the exact method, so results can often be obtained more quickly. However, if you have chosen the Monte Carlo method, but exact results can be calculated quickly for your data, they will be provided.

When to Use Exact Tests

Calculating exact results can be computationally intensive, time-consuming, and can sometimes exceed the memory limits of your machine. In general, exact tests can be performed quickly with sample sizes of less than 30. Table 1.1 provides a guideline for the conditions under which exact results can be obtained quickly.

Table 1.1 Sample sizes (N) at which the exact p values for nonparametric tests are computed quickly

One-sample inference	
Chi-square goodness-of-fit test	$N \le 30$
Binomial test and confidence interval	$N \le 100,000$
Runs test	$N \le 20$
One-sample Kolmogorov-Smirnov test	$N \le 30$
Two-related-sample inference	
Sign test	$N \le 50$
Wilcoxon signed-rank test	$N \le 50$
McNemar test	$N \le 100,000$
Marginal homogeneity test	$N \le 50$
Two-independent-sample inference	
Mann-Whitney test	$N \le 30$
Kolmogorov-Smirnov test	$N \le 30$
Wald-Wolfowitz runs test	$N \le 30$
K-related-sample inference	
Friedman's test	$N \le 30$
Kendall's W	$N \le 30$
Cochran's Q test	$N \le 30$
K-independent-sample inference	
Median test	$N \le 50$
Kruskal-Wallis test	$N \le 15, K \le 4$
Jonckheere-Terpstra test	$N \le 20, K \le 4$
Two-sample median test	$N \le 100,000$

Test Questions

References

NONPARAMETRIC TESTS
Eldho Varghese and Cini Varghese
Indian Agricultural Statistics Research Institute, New Delhi - 110 012
eldho@iasri.res.in, cini_v@iasri.res.in

IBM SPSS Exact Tests
Cyrus R. Mehta and Nitin R. Patel

IBM SPSS Statistics Base 20