Parametric Test

Non Parametric Test

Back to Basics …!!!


Parametric Test
Non Parametric Test

Parametric Test

Parametric tests normally involve data expressed in absolute numbers or values rather than ranks; an example is the Student's t-test

Parametric Test

Parametric tests are restricted to data that:

1) show a normal distribution
2) are independent of one another
3) are on the same continuous scale of measurement

Parametric Test

The meaningfulness of the

results of a parametric test depends on

the validity of the assumption

Parametric Test

Parametric tests are useful as these tests are most powerful for testing the significance or trustworthiness of the computed sample statistics

Parametric Test

Non Parametric Test

Non-Parametric Test

There may be a situations where we can not meet the assumptions & conditions and thus cannot use parametric statistical procedures

In such situation we are bound to apply non-parametric statistics

Non-Parametric Test

The first meaning of non-parametric covers techniques that do not rely on data belonging to any particular distribution

Non-Parametric Test

In this the statistics is based on the ranks of observations and do not depend on any distribution of the population

Non-Parametric Test

In non-parametric statistics, the techniques do not assume that the structure of a model is fixed

Non-Parametric Test

Non-parametric statistics

- deals with small sample sizes
- are assumption free meaning these are not bound by any assumptions
- are user friendly compared with parametric statistics and economical in time

Non-Parametric Test

Non-parametric tests are used on data that:

1) show an other-than normal distribution

2) are dependent or conditional on one another

3) in general, do not have a continuous scale of measurement

Parametric Test	Non-Parametric Test

For making inferences about various population values (parameters), we generally make use of parametric and non-parametric tests

Parametric Test	Non-Parametric Test
A parametric statistical test is one that makes assumptions about the parameters of the population distribution(s) from which one's data are drawn	A non-parametric test is one that makes no such assumptions

Parametric Test	Non-Parametric Test
If the information about the population is completely known by means of its parameters then statistical test is called parametric test	If there is no information about the population but still it is required to test the hypothesis of the population, then statistical test is called non-parametric test

Parametric Test	Non-Parametric Test
A parametric statistical test specifies certain conditions such as the data should be normally distributed	Non-parametric tests are known as distribution free tests

Parametric Test	Non-Parametric Test
Null hypothesis is made on parameters of the population distribution	The null hypothesis is free from parameters

Parametric Test	Non-Parametric Test
Parametric tests are applicable only for variable	Non parametric tests are applied to both variable and attributes

Parametric Test	Non-Parametric Test
No parametric test exist for Nominal scale data	Non parametric test do exist for nominal and ordinal scale data

Parametric Test	Non-Parametric Test
Parametric test is powerful	It is not as powerful as parametric test

Parametric Test	Non-Parametric Test
e.g. the length and weight of something	e.g. did the bacteria grow or not grow

Choosing	Parametric Test	Non-Parametric Test
Correlation test	Pearson	Spearman
Independent measures, 2 groups (Comparison of 2 group)	Independent-measures t-test	Mann-Whitney test
Independent measures, >2 groups (Comparison of several group)	One-way, independent-measures ANOVA	Kruskal-Wallis test
Repeated measures, 2 conditions (Comparison of 2 group)	Matched-pair t-test	Wilcoxon test
Repeated measures, >2 conditions (Comparison of groups values on 2 variables)	One-way, repeated measures ANOVA	Friedman's test

Parametric versus Nonparametric Statistics – When to use them and which is more powerful?

Dr. Sunil Mishra
Department of Management

Assumptions of Parametric and Non-Parametric Statistics

The general assumptions of parametric tests are

- The populations are normally distributed
- The selected population is representative of general population
- The data is in interval or ratio scale
- The observation must be independent
- These populations must have the same variance

Non-parametric tests can be applied when	Data don't follow any specific distribution
	Data measured on any scale
	No assumptions about the population are made
	The variable is continuous

Non-parametric tests can be applied when	Sample size is quite small
	Assumption like normality of the distribution of scores in the population are doubtful
	Measurement of data is available either in the form of ordinal or nominal scales

Non-parametric tests can be applied when	The data can be expressed in the form of ranks
	The nature of the population from which samples are drawn is not known to be normal
	The variables are expressed in nominal form

Advantages of Non-Parametric Statistics

Advantages of Nonparametric Statistics

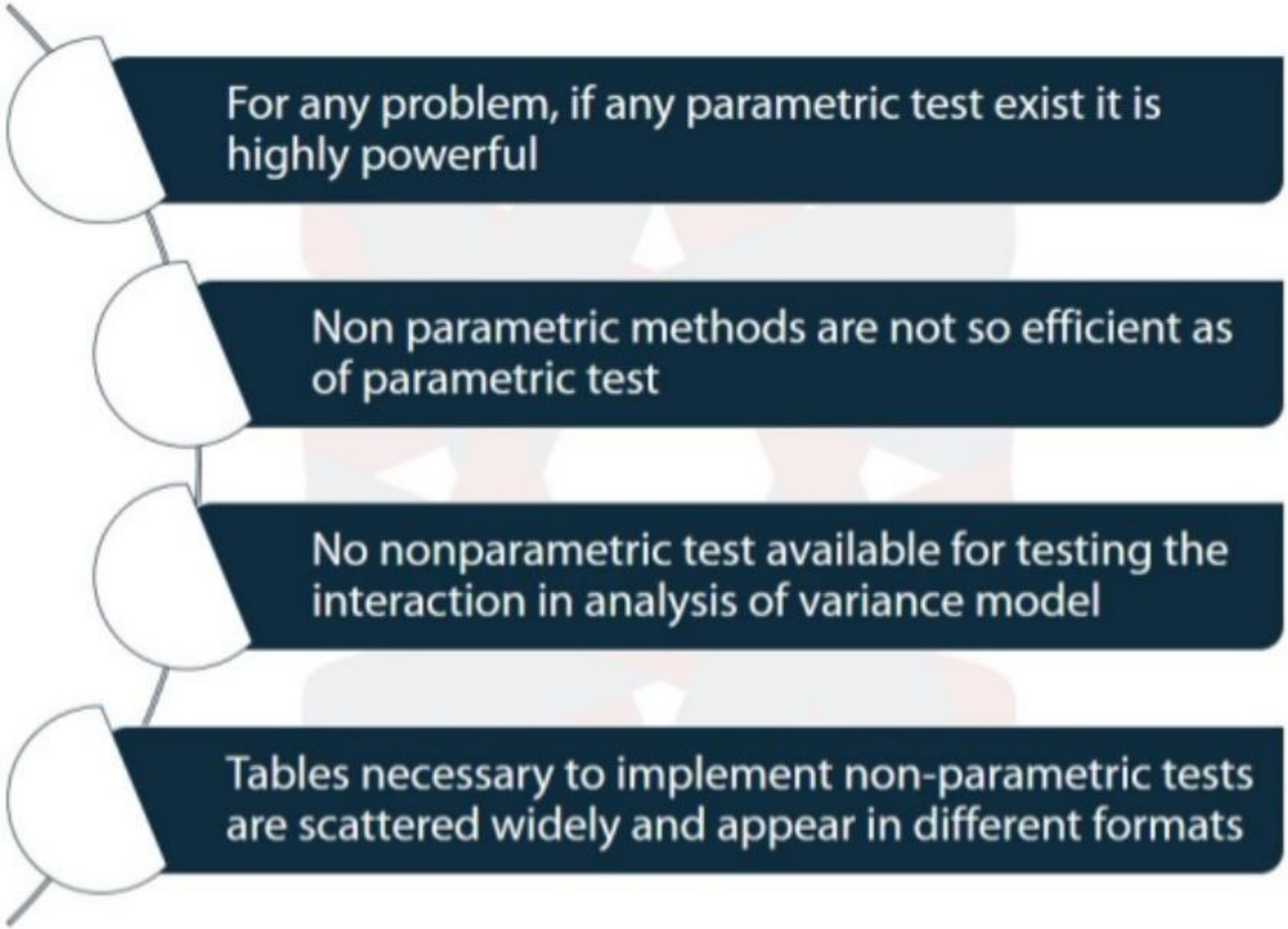
- Non parametric test are simple and easy to understand
- It will not involve complicated sampling theory
- No assumption is made regarding the parent population
- Non-parametric tests basically just need nominal or ordinal data

Advantages of Nonparametric Statistics

- This method is easy applicable for attribute dates
- Non-parametric statistics are more versatile tests
- Easier to calculate
- The hypothesis tested by the non-parametric test may be more appropriate

Disadvantages of Non-Parametric Statistics

Disadvantages of Nonparametric Statistics



Disadvantages of Nonparametric Statistics

- May waste information
- Require a larger sample size than corresponding parametric test in order to achieve same power
- Difficult to Compute by Hand for Large Samples
- Stat tables are not readily available

Nonparametric test

- The steps of the hypothesis test are the same, but the hypothesis is related to the median rather than the mean
  - Nonparametric tests apply to any type of distribution, even severely skewed distributions
  - We are interested in the median because the median is less affected by the tails of the distribution

Nonparametric vs Parametric

- Sign Test
- Mann-Whitney Test
- Spearman Rank Test
- Kruskal-Wallis Test
- Friedman Test

- One-sample t-test
- Two-sample t-test
- Correlation/Regression
- One-way ANOVA
- One-way blocked ANOVA

Last remarks

- Nonparametric tests are an opportunity to revise the basic ideas of statistical inference
- They are sometimes useful in biology
- They are often used in biology

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*

Measurement

What are the 4 levels of measurement?

1. Nominal or Classificatory Scale
    - Gender, ethnic background
2. Ordinal or Ranking Scale
    - Hardness of rocks, beauty, military ranks
3. Interval Scale
    - Celsius or Fahrenheit
4. Ratio Scale
    - Kelvin temperature, speed, height, mass or weight

Nonparametric Methods

- There is at least one nonparametric test equivalent to a parametric test
- These tests fall into several categories

1. Tests of differences between groups (independent samples)
2. Tests of differences between variables (dependent samples)
3. Tests of relationships between variables

Differences between dependent groups

- Compare two variables measured in the same sample
- If more than two variables are measured in same sample

Parametric	Nonparametric
t-test for dependent samples	Sign test
	Wilcoxon's matched pairs test
Repeated measures ANOVA	Friedman's two way analysis of variance
	Cochran Q

The Wilcoxon Matched-Pairs Signed Rank Test

- A non-parametric test for paired data, Wilcoxon is used in place of the *t*-test for paired samples (correlated groups) when the *t*-test assumptions are seriously violated.
- It is less powerful than *t*, but more powerful than the sign test.
- The data must be ordinal, interval, or ratio, but not nominal.

Differences between independent groups

- Two samples – compare mean value for some variable of interest

Parametric	Nonparametric
t-test for independent samples	Wald-Wolfowitz runs test
	Mann-Whitney U test
	Kolmogorov-Smirnov two sample test

Differences between independent groups

- Multiple groups

Parametric	Nonparametric
Analysis of variance (ANOVA/ MANOVA)	Kruskal-Wallis analysis of ranks
	Median test

Kruskal-Wallis

- Used in place of ANOVA for independent samples when the assumptions of ANOVA are seriously violated.
- Useful for ordinal, interval, or ratio data.

Relationships between variables

- Two variables of interest are categorical

Parametric	Nonparametric
Correlation coefficient	Spearman R
	Kendall Tau
	Coefficient Gamma
	Chi square
	Phi coefficient
	Fisher exact test
	Kendall coefficient of concordance

Summary Table of Statistical Tests

Level of Measurement	Sample Characteristics					Correlation
	1 Sample	2 Sample		K Sample (i.e., >2)		
		Independent	Dependent	Independent	Dependent	
Categorical or Nominal	X^2 or bi-nomial	X^2	Macnarmar's X^2	X^2	Cochran's Q	
Rank or Ordinal		Mann Whitney U	Wilcoxin Matched Pairs Signed Ranks	Kruskal Wallis H	Friendman's ANOVA	Spearman's rho
Parametric (Interval & Ratio)	z test or t test	t test between groups	t test within groups	1 way ANOVA between groups	1 way ANOVA (within or repeated measure)	Pearson's r
		Factorial (2 way) ANOVA				

(Plonskey, 2001)

Questions?

Non-parametric tests...

……include the following methods.

- The sign test
- Wilcoxon matched-pairs signed-ranks test
- The Mann-Whitney U test
- Kruskal-Wallis test
- In general, non-parametric tests are *less powerful* than parametric tests.

Run Test for Randomness

- The arrival of customers at a food bazaar of mall is given below –

  MMFFFMMMMFFFFMFMMMMMFF

  The manager wants to know whether the gender of the arriving customers is random at a significance level of 0.05.

Run Test for Randomness

- H_0 = The gender of the arriving customers is random.
- H_1 = The gender of the arriving customers is not random.

Run Test - Solution

- The frequency of occurrence of M, $n_1 = 10$
- The frequency of occurrence of F, $n_2 = 9$
- Number of Runs = 8
- Significance level = 0.05
  - Since n_1 & n_2 are less than 20, the sample is treated as small.

Run Test - Solution

- From the Table – the smaller critical value, and the larger critical value for the given combination of n_1 (10) & n_2 (9) at a significance level of 0.05 are 5 and 16, respectively.
- r_{cal} = 8 is not less than the smaller critical value (5), and not greater than the larger critical (16).

Run Test - Solution

- Accept the Null Hypothesis –
  - The gender of the arriving customers is random.

Run Test – Large Sample

- If n_1 or n_2 or both are more than 20, then the sample will be treated as large sample.
- Mean (μ) = $[2n_1n_2/(n_1+n_2)] + 1$
- Variance (σ^2) = $\dfrac{[2n_1n_2(2n_1n_2 - n_1 - n_2)]}{[(n_1+n_2)^2 (n_1 + n_2 - 1)]}$
- Testing the significance –

$$Z = (r - \mu)/ \sigma$$

Run Test for Randomness

- The arrival of customers at a food bazaar of mall is given below –

  MM FF MMMMMMM FF MMM F MMM FF MM FF MM FFF M FF MMM FF MM

  The manager wants to know whether the gender of the arriving customers is random at a significance level of 0.05.

Run Test for Randomness

- H_0 = The gender of the arriving customers is random.

- H_1 = The gender of the arriving customers is not random.

Run Test - Solution

- The frequency of occurrence of M, $n_1 = 24$
- The frequency of occurrence of F, $n_2 = 16$
- Number of Runs = 17
- Significance level = 0.05
  - Since n_1 is more than 20, the sample is treated as large.

Run Test – Large Sample

- Mean (μ) = $[2n_1n_2/(n_1+n_2)] + 1$

  $= 20.2$

- Variance (σ^2) = $\dfrac{[2n_1n_2(2n_1n_2 - n_1 - n_2)]}{[(n_1+n_2)^2 (n_1 + n_2 - 1)]}$

  $= 8.96$ **(σ = 2.993)**

- Testing the significance –

  $Z = (r - \mu)/ \sigma$

  $= -1.069$

Run Test - Solution

- Accept the Null Hypothesis –
  - The gender of the arriving customers is random.

Sign test - Example

- When patients have pancreatic cancer, often surgery is required to remove the part of the pancreas that has the cancer. When these surgeries are completed, the surgeon has the option to do a more complex surgery to preserve the spleen (splenic preservation) or to remove the spleen as part of the surgery (splenectomy)
- A study was done to compare the two surgical options in terms of health outcomes, cost and time burden on surgical staff.

Question

- A question for each technique is to determine the effect of the surgery on the platelet count in patients. Platelets are involved in clotting of patients and patients in surgery are sometimes given drugs to limit the amount of clotting during surgery. A large change in the number of platelets can be a sign that the surgery was particularly difficult.
- For each technique, the surgeons wanted to determine if there is a significant difference in the pre and post surgery platelet count.

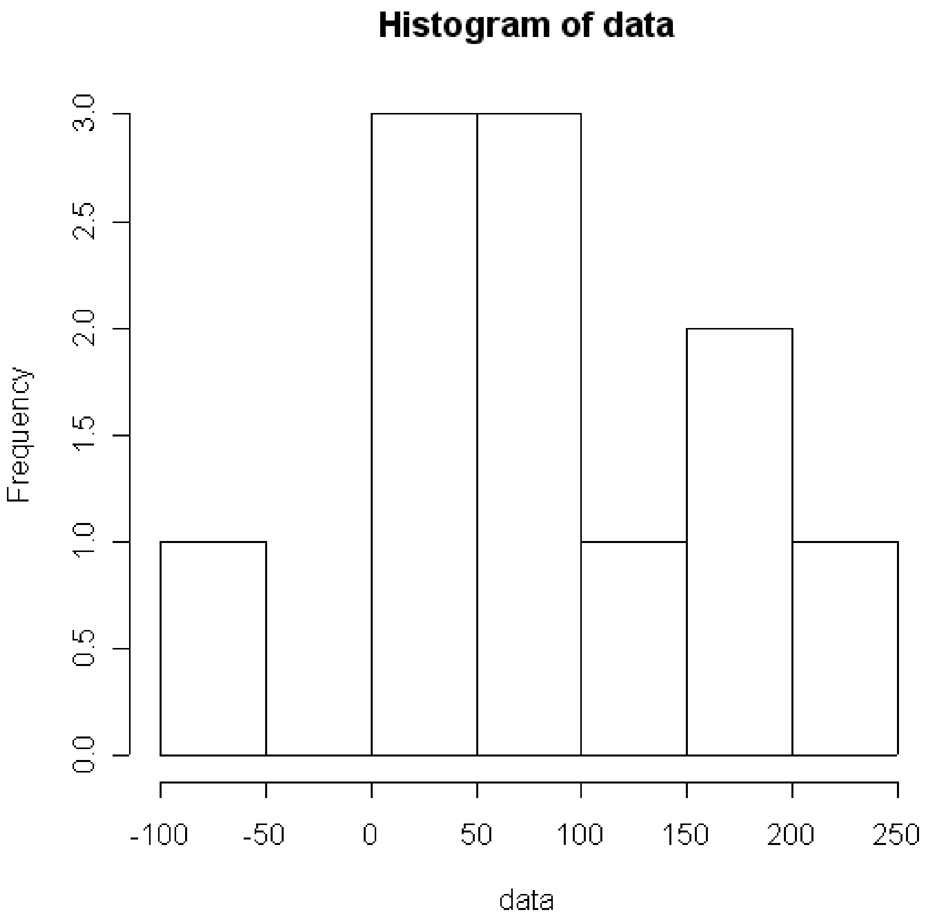
Example

- First, we will look at the splenic preservation group
- Note that we have paired observations on each of the patients
- We are interested in the difference between the two measurements
- Does it appear there is a difference?

Patient	Pre	Post	Diff
1	260	223	37
2	216	149	67
3	427	224	203
4	217	181	36
5	613	708	-95
6	245	197	48
7	371	303	68
8	236	168	68
9	421	312	109
10	677	521	156
11	363	202	161

Picture

- Since we have paired data, we could use the paired t-test.
- What can you say about the distribution of the differences?
- Does the normality assumption of the paired t-test seem appropriate?
- The difference in platelet count may be variable and contain outliers.



- The null hypothesis for our investigation is that there is no difference in the platelet count before and after the surgery.
- For the two-sample t-test, this was written as
  - H_0: mean difference (pre-post) is equal to zero (δ=0)
- In this case, we have outliers, so the mean is not a good measure of central tendency.
- What measure do you think we should use instead?
- How can we set up and test the appropriate null hypothesis?

Sign test

- The simplest nonparametric test is the sign test
- The null and alternative hypothesis for the sign test
  - H_0 : median of differences (pre-post) = 0
  - H_A : median of differences (pre-post) not = 0
- Under the null hypothesis, we would expect the same number of positive and negative signs. Therefore, P(positive sign)=0.5 under the null hypothesis
- If most or all of the differences are positive, there would be some evidence against the null hypothesis. How much?

Sign test

- We have now included the sign column
- If there was truly no effect of the therapy, we would assume that there would be an equal number of + and - signs
- What can you see about the signs of the differences? Is there a significant difference between the two groups? How can we calculate the p-value?

Patient	Pre	Post	Diff	Sign
1	260	223	37	+
2	216	149	67	+
3	427	224	203	+
4	217	181	36	+
5	613	708	-95	-
6	245	197	48	+
7	371	303	68	+
8	236	168	68	+
9	421	312	109	+
10	677	521	156	+
11	218	202	16	+

Hypothesis test

1) Paired data, alpha level=0.05
2) Hypotheses
   - H_0: median of differences = 0
   - H_A: median of differences != 0
3) Test statistic is 10+ signs
4) Reject null hypothesis
5) Conclusion: There is a significant difference between the pre- and post-surgery platelet values for patients who had the splenic preservation surgery

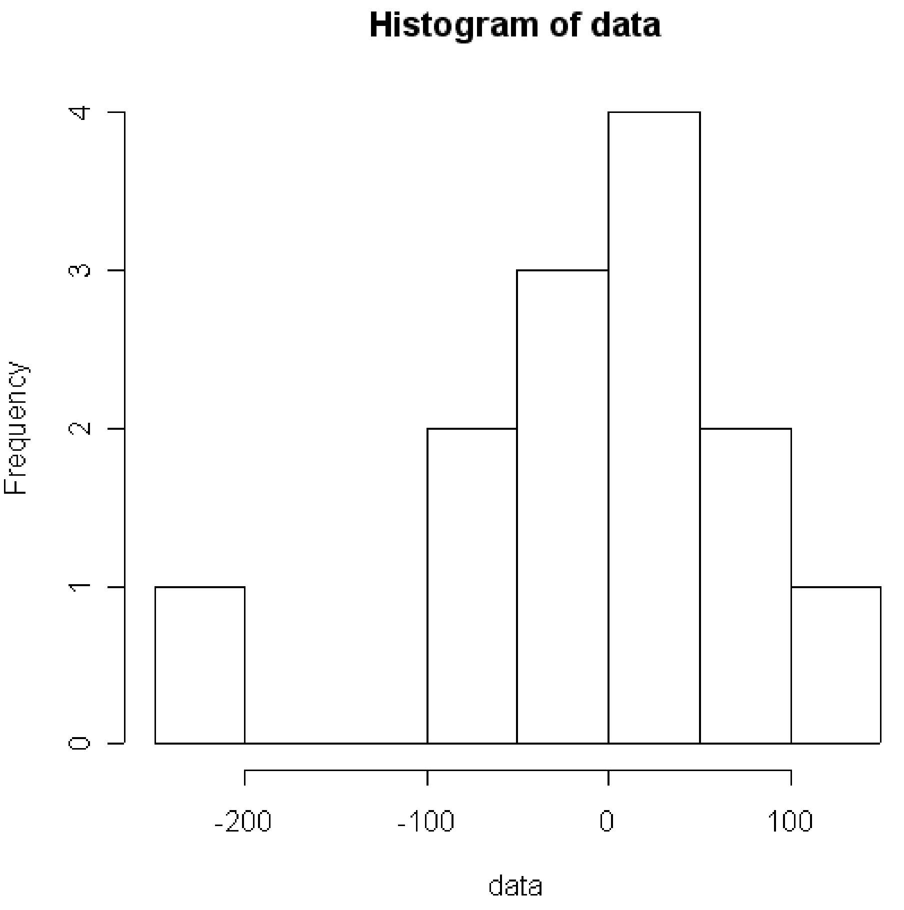
Example

- Now, we can look at the splenectomy group
- Again, we have paired observations on each of the patients, and we are interested in the difference between the two measurements
- Does it appear there is a difference?

Patient	Pre	Post	Diff
1	492	375	117
2	297	382	-85
3	272	325	-53
4	367	585	-218
5	206	181	25
6	284	237	47
7	338	273	65
8	212	243	-31
9	161	147	14
10	384	326	58
11	224	214	10
12	251	292	-41
13	224	263	-39

Picture

- Again, the distribution of the differences does not appear normal
- We could use the sign test, but there is another more powerful test called the Wilcoxon signed rank test.



Computing Wilcoxon

- From the paired scores, compute a column of D(ifference) scores, as with the sign test and *t* for correlated groups.
- Rank the absolute values (sizes) of the differences.
- Re-apply the algebraic signs of the differences to the rank values.
- Find the sum of the positive ranks, and the sum of the negative ranks.
- The smaller sum is the Wilcoxon statistic T.
- Compare the obtained value of T to table.

Wilcoxon Signed Rank Test

- The sign test looks only at the sign of the differences, but the Wilcoxon signed rank uses the sign and rank of the differences.
- The null and alternative hypotheses are the same as for the sign test
  - H_0: median diff = 0
  - H_A: median diff not = 0

Patient	Pre	Post	Diff	Rank
1	492	375	117	12
2	297	382	-85	-11
3	272	325	-53	-8
4	367	585	-218	-13
5	206	181	25	3
6	284	237	47	7
7	338	273	65	10
8	212	243	-31	-4
9	161	147	14	2
10	384	326	58	9
11	224	214	10	1
12	251	292	-41	-6
13	224	263	-39	-5

- **The test statistic of this test is the sum of the positive ranks.**
- Under the null hypothesis, half of the ranks should be positive and half of the ranks should be negative. **Evidence against the null would be having the sum of the positive ranks either being very high or very low.**

alternative hypothesis: true mu is not equal to 0

Hypothesis test

- Paired data, wilcoxon test, alpha=0.05
- Hypotheses
  - Null: median difference = 0
  - Alternative: median difference not = 0
- Test statistic: Sum of positive ranks = 44
- Fail to reject null hypothesis
- Conclusion: There is no evidence of a difference between the pre and post platelet counts for patients who had a splenectomy during their surgery.

Comments

- When we have paired data and the assumptions of a paired t-test are not met, we have two ways to complete the hypothesis test
- The Wilcoxon test is always preferred over the sign test because it uses more of the data (since it uses the ranks). The Wilcoxon test has much more power to detect a significant difference.
- There is not a large loss of power in using a Wilcoxon test compared to a t-test when the normality assumption holds. The Wilcoxon is much more powerful when the normality assumption does not hold.
- Therefore, the Wilcoxon test is more appropriate if there is any reason to doubt the normality assumption.

Mann-Whitney U Test
Wilcoxon rank-sum test

- Nonparametric alternative to two-sample t-test (independent samples)
- Actual measurements not used – ranks of the measurements used
- Data can be ranked from highest to lowest or lowest to highest values
- Calculate Mann-Whitney U statistic

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

The Mann-Whitney *U* test

- Like Wilcoxon, requires ordinal, interval, or ratio data, but not nominal.
- *U* is obtained by ranking the combined scores of both groups. Then, find the sum of the ranks for each group separately.
- Apply the *U* formula.

Example of Mann-Whitney U test

- Two tailed null hypothesis that there is no difference between the heights of male and female students
- H_o: Male and female students are of the same height
- H_A: Male and female students are not of the same height

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U = (7)(5) + \frac{(7)(8)}{2} - 30$$

$$U = 35 + 28 - 30$$

$$\mathbf{U = 33}$$

$$U' = n_1 n_2 - U$$

$$U' = (7)(5) - 33$$

$$U' = 2$$

$$\mathbf{U_{0.05(2),7,5} = U_{0.05(2),5,7} = 30}$$

As 33 > 30, H_o is rejected

Heights of males (cm)	Heights of females (cm)	Ranks of male heights	Ranks of female heights
193	175	1	7
188	173	2	8
185	168	3	10
183	165	4	11
180	163	5	12
178		6	
170		9	
$n_1 = 7$	$n_2 = 5$	$R_1 = 30$	$R_2 = 48$

Zar, 1996

REVISION

Level of measurement	One - Sample Test	Two-sample case		K-sample case	
		Related Samples	Independent samples	Related samples	Independent samples
Nominal	Binomial	McNemar for significance of changes	Fisher exact probability Chi-square	Cochran Q (Dichotomous)	Chi-square
Ordinal	Kolmogorov Smirnov Runs	Sign Wilcoxon matched-pair signed-ranks	Mann-Whitney U Kolmogorov-Smirnov Wald-Wolfowitz runs Moses of extreme reactions	Friedman two-way analysis of variance Kendall's W	Kruskal-Wallis one-way analysis of variance
Interval		Walsh	Randomization		

Parametric tests-nonparametric equivalent

- Paired t-test – Wilcoxon signed rank
- Two sample t-test – Wilcoxon rank sum
- ANOVA – Kruskal-Wallis test
  - When you have two or more independent samples and the assumptions of ANOVA are not met, you can use the Kruskal-Wallis test. This is a rank based test.
  - The command in R is kruskal.test
  - As a homework problem, try to complete the ANOVA analyses from last class using the Kruskal-Wallis test

Sign test

- tests whether the numbers of differences (+ve or –ve) between two samples are approximately the same. Each pair of scores (before and after) are compared.
- When “after” > “before” (+ sign), if smaller (- sign). When both are the same, it is a tie.
- Sign-test did not use all the information available (the size of difference), but it requires less assumptions about the sample and can avoid the influence of the outliers.

Wilcoxon matched-pairs signed-ranks test

- **Wilcoxon matched-pairs signed-ranks test** – Similar to sign test, but take into consideration the ranking of the magnitude of the difference among the pairs of values. (Sign test only considers the direction of difference but not the magnitude of differences.)
- The test requires that the differences (**of the true values**) be a sample from a symmetric distribution (but not require normality). It's better to run stem-and-leaf plot of the differences.

Two-sample case (independent samples)

- **Mann-Whitney U** – similar to Wilcoxon matched-paired signed-ranks test except that the samples are independent and not paired. It’s the most commonly used alternative to the independent-samples *t* test.
- *Null hypothesis*: the population means are the same for the two groups.
- The actual computation of the Mann-Whitney test is simple. You rank the combined data values for the two groups. Then you find the average rank in each group.
- Requirement: the population variances for the two groups must be the same, but the shape of the distribution does not matter.

Kolmogorov-Smirnov Z

- to test if two distributions are different. It is used when there are only a few values available on the ordinal scale. K-S test is more powerful than M-W U test if the two distributions differ in terms of dispersion instead of central tendency.

- **Wald-Wolfowitz Run** – Based on the number of runs within each group when the cases are placed in rank order.
- **Moses test of extreme reactions** – Tests whether the range (excluding the lowest 5% and the highest 5%) of an ordinal variables is the same in the two groups.

K-sample case (Independent samples)

- **Kruskal-Wallis One-way ANOVA** – It’s more powerful than Chi-square test when ordinal scale can be assumed. It is computed exactly like the Mann-Whitney test, except that there are more groups. The data must be independent samples from populations with the same shape (but not necessarily normal).

K related samples

- **Friedman two-way ANOVA** – test whether the k related samples could probably have come from the same population with respect to mean rank.

- **Cochran Q** – determines whether it is likely that the k related samples could have come from the same population with respect to proportion or frequency of “successes” in the various samples. In other words, it only compares dichotomous variables.