BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Statistical Methods for Bioinformatics

Course Code: 23ST08DEC

Unit-III

Statistical testing and significance

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. I. Angel Agnes Mary
Guest Faculty
Department of Statistics

UNIT – III

Statistical Testing and Significance

Hypothesis

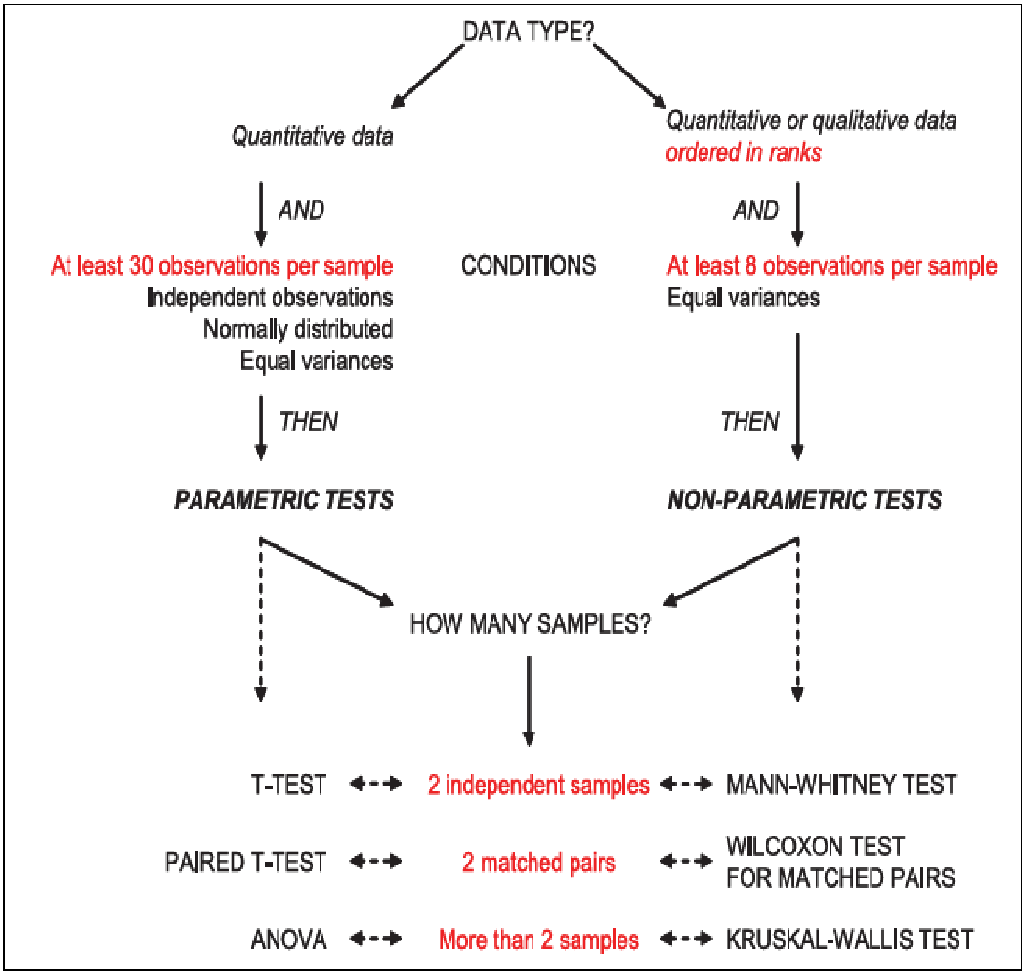
A hypothesis is a predictive statement that is tested by investigation.

Null hypothesis

The null hypothesis is a statement that you want to test. In general, the null hypothesis is that things are the same as each other, or the same as a theoretical expectation.

Alternative hypothesis

The alternative hypothesis is that things are different from each other, or different from a theoretical expectation.



Choosing the Appropriate Statistical Test

Parametric tests and Non-Parametric tests

Parametric tests allow testing hypotheses related to means. They use rigorous and often complex mathematical theory and require a probability distribution to be specified for the populations from which samples were taken (this is usually the normal distribution).

Non-parametric tests use ranks of the observations to compare medians rather than means. This removes the need for data to be normally distributed. Non-parametric tests are less powerful than parametric tests, because they don't use the actual values of the observations, but only the ranks of the observations (thus, it is often said that they lose information).

Conditions of parametric tests

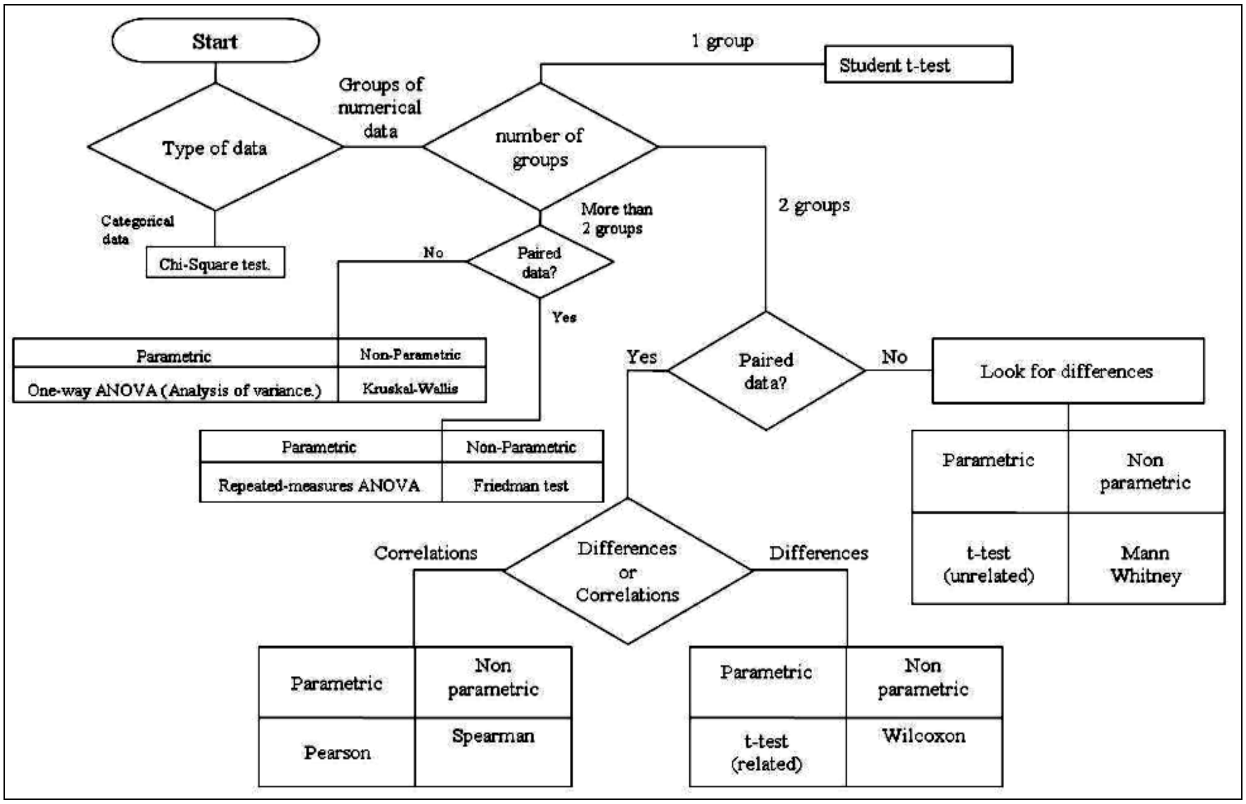
Parametric tests have three assumptions that theoretically must be met in order for the outcomes of the tests to be reliable. The assumptions of parametric tests are as follows:

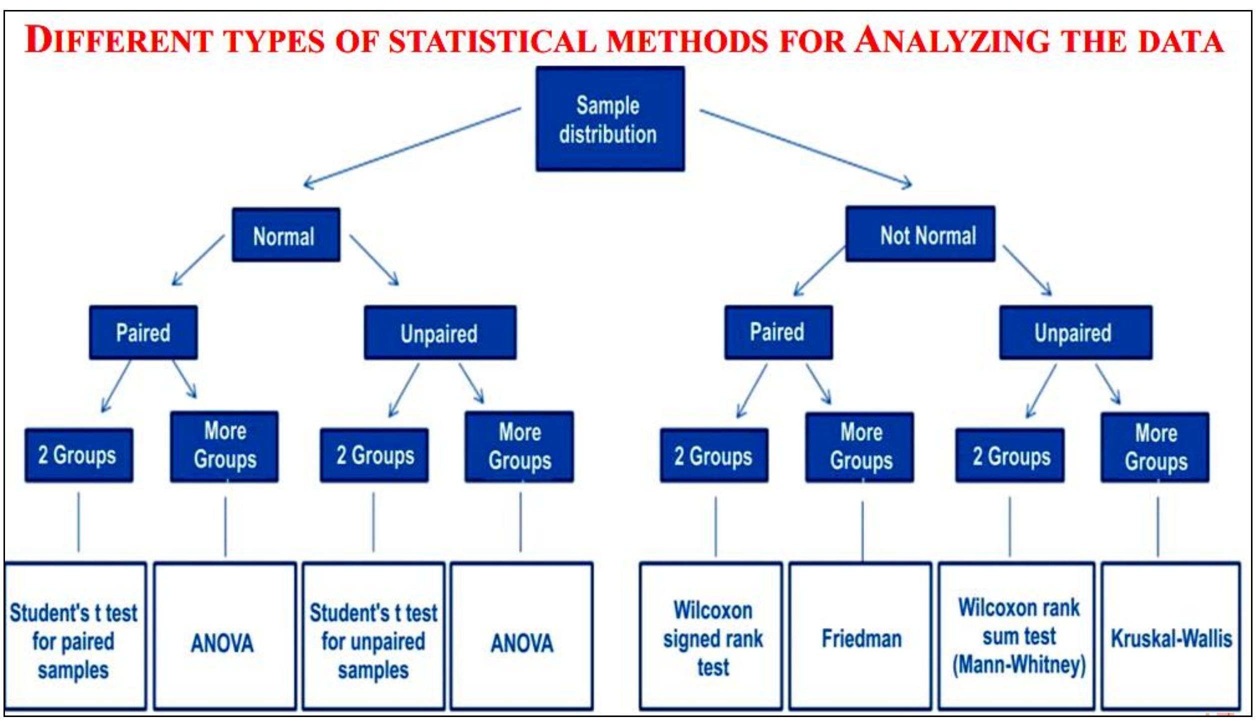
1. The samples come from **normally distributed** populations. However, parametric tests are usually fairly robust to moderate violations of this assumption providing the sample sizes of the samples to be compared are equal. Transformation of the variable to a different scale can also improve its normality.
2. The samples come from populations with **equal variances**. This assumption is more critical than that of normality but again tests are reasonably robust if sample sizes are equal. In addition, variances are often unequal because distributions are skewed.
3. The observations should be **independent** of each other, both within and between sample groups. This assumption must be considered in the design phase of the study.

Conditions of non-parametric tests

While there are fewer conditions required to run non-parametric tests, some should still be met to ensure the reliability of the tests. The assumptions of non-parametric tests are as follows:

1. The samples must have **equal variances.**
2. Distributions of the populations must be similar (but they do not have to follow the normal distribution).


Start
Type of data
Categorical data
Chi-Square test.
Groups of numerical data
number of groups
1 group
Student t-test
More than 2 groups
2 groups
Paired data?
No
Yes
Parametric
Non-Parametric
One-way ANOVA (Analysis of variance.)
Kruskal-Wallis
Parametric
Non-Parametric
Repeated-measures ANOVA
Friedman test
Paired data?
Yes
No
Look for differences
Parametric
Non parametric
t-test (unrelated)
Mann Whitney
Differences or Correlations
Correlations
Differences
Parametric
Non parametric
Pearson
Spearman
Parametric
Non parametric
t-test (related)
Wilcoxon


DIFFERENT TYPES OF STATISTICAL METHODS FOR ANALYZING THE DATA
Sample distribution
Normal
Not Normal
Paired
Unpaired
Paired
Unpaired
2 Groups
More Groups
2 Groups
More Groups
2 Groups
More Groups
2 Groups
More Groups
Student's t test for paired samples
ANOVA
Student's t test for unpaired samples
ANOVA
Wilcoxon signed rank test
Friedman
Wilcoxon rank sum test (Mann-Whitney)
Kruskal-Wallis

Student's *t*–test for one sample

The one-sample t-test is used to determine whether a sample comes from a population with a specific mean. This population mean is not always known, but is sometimes hypothesized. For example, you want to show that a new teaching method for pupils struggling to learn English grammar can improve their grammar skills to the national average. Your sample would be pupils who received the new teaching method and your population mean would be the national average score. Alternately, you believe that doctors that work in Accident and Emergency (A & E) departments work 100 hour per week despite the dangers (e.g., tiredness) of working such long hours. You sample 1000 doctors in A & E departments and see if their hours differ from 100 hours.

One-sample *t*-test assumptions

For a valid test, we need data values that are:

- Independent (values are not related to one another).
- Continuous Variable.
- Obtained via a simple random sample from the population.
- Also, the population is assumed to be normally distributed.

Problem

In the biological example of the model t-Test, 10 volunteers close their eyes, bend their knees at a 120-degree angle for a few seconds, and then rotate the knee at a 90-degree angle. Then when each person bends their knees at an angle of 120 degrees, the obtained samples are as follows:

Individual	A	B	C	D	E	F	G	H	I	J
Angle	120.6	116.4	117.2	118.1	114.1	116.9	113.3	121.1	116.9	117.0

Mean of the angle is 120. To test whether people overestimate or underestimate their knee angle.

Procedure:

- State the null and alternate hypothesis.
- Calculate the sample mean and Standard Deviation.
- Calculate the test statistic

$$t_s = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$\overline{x}$ = sample mean
μ_0 = hypothesized population mean
s = sample standard deviation
n = sample size

- Determine the Critical value from Table and Conclusion.

Calculation

Step 1:

Hypothesis

***Null hypothesis*:** The biological null hypothesis is that people don't under- or overestimate their knee angle.

***Alternative hypothesis*:** The biological null hypothesis is that people under- or overestimate their knee angle.

Step 2:

Identify the following pieces of information will need to calculate the test statistic.

- The population mean (μ). Given as 120.
- Number of observations (n) = 10.
- The sample mean (x̄) is:

$$\bar{x} = \frac{120.6 + 116.4 + 117.2 + 118.1 + 114.1 + 116.9 + 113.3 + 121.1 + 116.9 + 117.0}{10}$$

$$= \frac{1171.6}{10}$$

$$= 117.16$$

- The sample standard deviation is:

$$s = \sqrt{\frac{\sum (X - \bar{x})^2}{n - 1}}$$

X - The Value in the data distribution
$\bar{x}$ - The Sample Mean
n - Total Number of Observations

$$s = \sqrt{\frac{(11.83 + 0.58 + 0.00 + 0.88 + 9.36 + 0.07 + 14.90 + 15.52 + 0.07 + 0.03)}{(10 - 1)}}$$

$$= \sqrt{\frac{53.24}{9}}$$

s = 2.43

Step 3:

Calculate the test statistic, *t*, using this formula:

$$t_s = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$t = \frac{117.16 - 120}{2.43/\sqrt{10}}$$

$$= \frac{-2.84}{0.7684}$$

$$t = -3.69$$

Step 4:

Find the t-table value:

1. The alpha level: 5%
2. The degrees of freedom, which is the number of items in the sample (n) minus 1:

   df = (10 – 1) = 9
3. Table Value = 2.262

Conclusion:

Since calculated value is greater than the table value (3.69>2.262), we reject the null hypothesis and conclude that the people under- or overestimate their knee angle.

Student's t-test for two sample (t-test)

The Independent Samples *t* Test compares the means of two independent groups in order to determine whether there is statistical evidence that the associated population means are significantly different. The Independent Samples *t* Test is a parametric test.

This test is also known as:

- Independent *t* Test
- Independent Measures *t* Test
- Independent Two-sample *t* Test
- Student *t* Test
- Two-Sample *t* Test
- Uncorrelated Scores *t* Test
- Unpaired *t* Test
- Unrelated *t* Test

The variables used in this test are known as:

- Dependent variable, or test variable
- Independent variable, or grouping variable

Assumption

- Assumes that the dependent variable is normally distributed.
- Assumes that the variance of the two groups is the same as the dependent variable.
- Assumes that the two samples are independent of each other.
- Samples are drawn from the population at random.
- In independent sample t-test, all observations must be independent of each other.
- In independent sample t-test, dependent variables must be measured on an interval or ratio scale.

Problem

After a small number of crabs were accidentally released into a shallow pond, biologists noticed that the crabs ingested all the underwater plant population; Aquatic invertebrates such as water fleas (Daphnia sp.) are also declining. Biologists knew that goldfish were the main predators of Daphnia, and they believed that underwater plants protected Daphnia from goldfish. Daphnia lost its protection as the plants disappeared under water. Biologists designed an experiment to test their hypothesis. Place the goldfish and daphnia together in one tank with underwater vegetation and the equivalent number of goldfish and daphnia in another tank with no underwater vegetation. The number of daphnia that ate the goldfish was calculated in 30 minutes. They copied this experiment into nine additional pairs of tanks (i.e. sample size = 10, or n = 10 per group). The results of their experiment and the calculations of the test error (variance, s^2) are given in the following table.

Tanks	1 & 2	3 & 4	5 & 6	7 & 8	9 & 10	11 & 12	13 & 14	15 & 16	17 & 18	19 & 20
Plants	13	9	10	10	7	5	10	14	9	9
No Plants	14	12	15	14	17	10	15	15	18	14

Procedure

- State the null and alternate hypothesis.
- Calculate the sample mean.
- Calculate the test statistic

$$t = \frac{\mu_A - \mu_B}{\sqrt{\left[\frac{\left(\Sigma A^2 - \frac{(\Sigma A)^2}{n_A}\right) + \left(\Sigma B^2 - \frac{(\Sigma B)^2}{n_B}\right)}{n_A + n_B - 2}\right] \cdot \left[\frac{1}{n_A} + \frac{1}{n_B}\right]}}$$

$(\Sigma A)^2$: Sum of data set A, squared.
$(\Sigma B)^2$: Sum of data set B, squared.
μ_A: Mean of data set A.
μ_B: Mean of data set B.
ΣA^2: Sum of the squares of data set A.
ΣB^2: Sum of the squares of data set B.
n^A: Number of items in data set A.
n^B: Number of items in data set B.

- Determine the Critical value from Table and Conclusion.

Calculation

Step 1: Hypothesis:

Null hypothesis: There is no significant difference in the number of Daphnia in tanks with plants compared to tanks without plants

Alternative hypothesis: There is significant difference in the number of Daphnia in tanks with plants compared to tanks without plants

Step 2:

Tanks	**Plants (x_1)**	**No plants (x_2)**	**$(x_1)^2$**	**$(x_2)^2$**
1 and 2	13	14	169	196
3 and 4	9	12	81	144
5 and 6	10	15	100	225
7 and 8	10	14	100	196
9 and 10	7	17	49	289
11 and 12	5	10	25	100
13 and 14	10	15	100	225
15 and 16	14	15	196	225
17 and 18	9	18	81	324
19 and 20	9	14	81	196
Total	**96**	**144**	**982**	**2120**

Mean:

$$Plants\ (\bar{x}_1) = \frac{13+9+10+10+7+5+10+14+9+9}{10} = \frac{96}{10} = 9.6$$

$$No\ Plants\ (\bar{x}_2) = \frac{14+12+15+14+17+10+15+15+18+14}{10} = \frac{144}{10} = 14.4$$

Step 3:

$$t = \frac{\mu_A - \mu_B}{\sqrt{\left[\frac{\left(\Sigma A^2 - \frac{(\Sigma A)^2}{n_A}\right) + \left(\Sigma B^2 - \frac{(\Sigma B)^2}{n_B}\right)}{n_A + n_B - 2}\right] \cdot \left[\frac{1}{n_A} + \frac{1}{n_B}\right]}}$$

$$t = \frac{9.6 - 14.4}{\sqrt{\left(\frac{\left(982 - \frac{(96)^2}{10}\right) + \left(2120 - \frac{(144)^2}{10}\right)}{10+10-2}\right) \times \left(\frac{1}{10} + \frac{1}{10}\right)}}$$

$$t = \frac{-4.8}{\sqrt{\left(\frac{60.4+46.4}{18}\right) \times (0.2)}}, \quad t = \frac{-4.8}{\sqrt{\left(\frac{106.8}{18}\right) \times (0.2)}}, \quad t = \frac{-4.8}{1.089}, \quad t = |-4.40|$$

t = 4.40

Step 4:

Find the t-table value:

1. The alpha level: 5%
2. The degrees of freedom, (n_A-1 + n_B-1):

   df = (10 – 1)+(10 – 1) = 18
3. Table Value = 2.10.

Conclusion

Compare calculated value to table value. The calculated value is greater than the table value (4.40 > 2.10). We reject the null hypothesis. Hence, we conclude that there is significant difference in the number of Daphnia in tanks with plants compared to tanks without plants.

Paired t-test

A paired t-test (also known as a dependent or correlated t-test) is a statistical test that compares the averages/means and standard deviations of two related groups to determine if there is a significant difference between the two groups.

- A significant difference occurs when the differences between groups are unlikely to be due to sampling error or chance.
- The groups can be related by being the same group of people, the same item, or being subjected to the same conditions.

The variable used in this test is known as:

- Dependent variable, or test variable (continuous), measured at two different times or for two related conditions or units

Assumptions of a paired t-test

- The dependent variable is normally distributed
- The observations are sampled independently
- The dependent variable is measured on an incremental level, such as ratios or intervals.
- The independent variables must consist of two related groups or matched pairs.

Problem

20 students were given a diagnostic test before studying a particular module and then again after completing the module. We want to find out if, in general, our teaching leads to improvements in students' knowledge/skills (i.e. test scores).The following data was given below:

Pre-module score	18	21	16	22	19	24	17	21	23	18
	14	16	16	19	18	20	12	22	15	17
Post-module score	22	25	17	24	16	29	20	23	19	20
	15	15	18	26	18	24	18	25	19	16

Is there is significant difference in the Pre-module Score and Post-module Score.

Procedure

- State the null and alternate hypothesis.
- Calculate the Difference mean and Standard Deviation.
- Calculate the test statistic

$$T = \frac{\bar{d}}{SE(d)}$$

- Determine the Critical value from Table and Conclusion.

Calculations:

Step 1: Hypothesis:

Null hypothesis**:** There is no significant difference in the Pre-module Score and Post-module Score.

Alternative hypothesis**:** There is significant difference in the Pre-module Score and Post-module Score.

Step 2:

Let x = test score before the module, y = test score after the module

Student	Pre-module score	Post-module score	Difference
1	18	22	+4
2	21	25	+4
3	16	17	+1
4	22	24	+2
5	19	16	-3
6	24	29	+5
7	17	20	+3
8	21	23	+2
9	23	19	-4
10	18	20	+2
11	14	15	+1
12	16	15	-1
13	16	18	+2
14	19	26	+7
15	18	18	0
16	20	24	+4
17	12	18	+6
18	22	25	+3
19	15	19	+4
20	17	16	-1

Calculating the mean and standard deviation of the differences gives:

$\bar{d} = 2.05$ and $s_d = 2.837$. Therefore, $SE(d) = \frac{s_d}{\sqrt{n}} = \frac{2.837}{\sqrt{20}} = 0.634$

Step 3:

Calculate the t-statistic, which is given by $T = \frac{\bar{d}}{SE(d)}$.

So, we have:

$$t = \frac{2.05}{0.634} = 3.231 \qquad \text{on 19 df}$$

Step 4:

Table Value

Find the t-table value:

1. The alpha level: 5%
2. The degrees of freedom, (n-1):

   df = 20-1 = 19
3. Table Value = 2.093.

Conclusion

Compare calculated value to table value. The calculated value is greater than the table value (3.231 > 2.093). We reject the null hypothesis. Hence, we conclude that there is significant difference in the Pre-module Score and Post-module Score.

Non-parametric tests

Definition

Non-parametric tests are also known as **distribution-free tests.** These are statistical tests that do not require normally-distributed data.

When to use non-parametric tests

Non-parametric tests are tests with fewer restrictions than parametric tests. It is appropriate to use non-parametric tests in research in different cases. For example:

- When data is **nominal.** Data is nominal when it is assigned to groups; these groups are distinct and have limited similarities (e.g. responses to 'What is your ethnicity?').
- When data is **ordinal.** That is when data has a set order or scale (e.g. 'Rate your anger from 1-10'.)
- **When** there have been **outliers** identified in the data set.
- When data was collected from a **small sample.**

However, it is important to note that non-parametric tests are also used when the following criteria can be assumed:

- At least **one violation** of parametric tests assumptions. E.g., data should have similar **homoscedasticity of variance:** the amount of 'noise' (potential experimental errors) should be similar in each variable and between groups.
- Non-normal **distribution** of data. In other words, data is likely skewed.
- **Randomness:** data should be taken from a random sample from the target population.

- **Independence:** the data from each participant in each variable should not be correlated. This means that measurements from a participant should not be influenced or associated with other participants.

Non-parametric statistical tests

The table below shows examples of non-parametric tests. It includes their parametric test equivalent, the method of data analysis the test uses, and example research that is appropriate for each statistical test.

Non-parametric test	**Equivalent parametric test**	**Purpose of statistical test**	**Example**
Wilcoxon signed rank test	Paired t-test	Compares the mean value of two variables obtained from the same participants.	The difference in depression scores before and after treatment.
Mann-Whitney U test	Independent sample t-test	Compares the mean value of a variable measured from two independent groups.	The difference between depression symptom severity in a placebo and drug therapy group.
Spearman correlation	Pearson correlation	Measures the relationship (strength/direction) between two variables.	The relationship between fitness test scores and the number of hours spent exercising.
Kruskal Wallis test	One-way analysis of variance (ANOVA)	Compares the mean of two or more independent groups (uses a between-subject design and the independent variable needs to have three or more levels.)	The difference in average fitness test scores of individuals who exercise frequently, moderately, or do not exercise.
Friedman Test	One-way repeated measures ANOVA	Compares the mean of two or more dependent groups (uses a within-subject design and the independent variable needs to have three or more levels.)	The difference in average fitness test scores during the morning, afternoon, and evening.

Mann-Whitney U test

Mann-Whitney u-Test is a non-parametric test used to test whether two independent samples were selected from population having the same distribution. Another name for the Mann-Whitney U Test is Wilcoxon Rank Sum Test.

Assumptions

Mann-Whitney U test is a non-parametric test, so it does not assume any assumptions related to the distribution of scores. There are, however, some assumptions that are assumed

- The sample drawn from the population is random.
- Independence within the samples and mutual independence is assumed. That means that an observation is in one group or the other (it cannot be in both).
- Ordinal measurement scale is assumed.

Problem

In order to assess the efficacy of a new antidepressant drug, ten clinically depressed patients are randomly assigned to one of two groups. Five patients are assigned to Group 1, which is administered the antidepressant drug for a period of six months. The other five patients are assigned to Group 2, which is administered a placebo during the same six-month period. Assume that prior to introducing the experimental treatments; the experimenter confirmed that the level of depression in the two groups was equal. After six months elapse all ten subjects are rated by a psychiatrist (who is blind with respect to a subject's experimental condition) on their level of depression. The psychiatrist's depression ratings for the five subjects in each group follow (the higher the rating, the more depressed a subject):

Antidepressant Drug	11	1	0	2	0
Placebo	11	11	5	8	4

Do the data indicate that the antidepressant drug is effective?

Procedure

- State the null and alternate hypothesis.
- Perform a ranking of all the observation
- Calculate the Rank Sums
- Calculate the U Statistic for the Two Groups

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=n_1+1}^{n_2} R_i$$

Where,

U = Mann-Whitney U test

n_1 = Sample Size One

n_2 = Sample Size Two

R_i = Rank of the sample Size

- Determine the Critical value from Table

Calculation

Step 1:

Hypothesis:

***Null hypothesis*:** There is no significant difference in the antidepressant drug and Placebo.

***Alternative hypothesis*:** There is significant difference in the antidepressant drug and Placebo.

Step 2:

Antidepressant Drug	11	1	0	2	0
Rank	9	3	1.5	4	1.5
Placebo	11	11	5	8	4
Rank	9	9	6	7	5

Step 3:

Mean Rank of Drugs:

Antidepressant drug = 9 + 3 + 1.5 + 4 + 1.5 = 19 /5 = 3.8

Placebo = 9 + 9 + 6 + 7 + 5 = 36/5 = 7.2

Step 4:

The test statistic for the Mann Whitney U Test is denoted **U** and is the ***smaller*** of U_1 and U_2, defined below.

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

where R_1 = sum of the ranks for group 1 and R_2 = sum of the ranks for group 2.

$$U_1 = 5(5) + \frac{5(6)}{2} - 19 = 25 + 15 - 19 = 21$$

$$U_2 = 5(5) + \frac{5(6)}{2} - 36 = 25 + 15 - 36 = 4$$

Therefore, U = 4.

Step 5:

Table Value:

Find the t-table value:

1. The alpha level: 5%
2. The degrees of freedom, (n_1,n_2):

   df = (5,5)
3. Table Value = 2

Conclusion:

Compare calculated value to table value. The calculated value is greater than the table value (4>2). We reject the null hypothesis. Hence, we conclude that there is significant difference in the antidepressant drug and Placebo**.**

Wilcoxon signed rank test

The **Wilcoxon signed rank test** is a non-parametric test to compare dependent samples t-test data. When the word "non-parametric" is used in stats, it doesn't quite mean that you know nothing about the population. It usually means that you know the population data does not have a normal distribution. The Wilcoxon signed rank test should be used if the differences between pairs of data are non-normally distributed.

Assumptions

Two slightly different versions of the test exist:

- The **Wilcoxon signed rank test** compares your sample median against a hypothetical median.
- The **Wilcoxon matched-pairs signed rank test** computes the difference between each set of matched pairs and then follows the same procedure as the signed rank test to compare the sample against some median.

The term "Wilcoxon" is often used for either test. This usually isn't confusing, as it should be obvious if the data is matched, or not matched.

The null hypothesis for this test is that the medians of two samples are equal. It is generally used:

- As a non-parametric alternative to the one-sample t test or paired t test.
- For ordered (ranked) categorical variables without a numerical scale.

Problem

In order to assess the efficacy of electroconvulsive therapy (ECT), a psychiatrist evaluates ten clinically depressed patients before and after a series of ECT treatments. A standardized interview is used to operationalize a patient's level of depression, and on the basis of the interview each patient is assigned a score ranging from 0 to 10 with respect to his or her level of depression prior to (pretest score) and after (posttest score) the administration of ECT. The higher a patient's score, the more depressed the patient. The pretest and posttest scores of the ten patients follow:

Patient	1	2	3	4	5	6	7	8	9	10
pretest	9	2	1	4	6	4	7	8	5	1
posttest	8	2	3	2	3	0	4	5	4	0

Do the data indicate that ECT is effective?

Procedure

- State the null and alternate hypothesis.
- Perform a Differencing and ranking of all the observation
- Calculate the Rank Sums
- Calculate the z Statistic for the Two Groups

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Where,

T = Minimum Sum of the Rank Value

n = Sample Size

- Determine the Critical value from Table

Calculation

Step 1: Hypothesis:

***Null hypothesis*:** There is no significant difference in the Pretest and Post Test.
***Alternative hypothesis*:** There is significant difference in the Pretest and Post Test.

Step 2:

Patient	pretest	posttest	Difference	Sign	Rank
1	9	8	1	+	2
2	2	2	0	0	0
3	1	3	-2	-	4.5
4	4	2	2	+	4.5

5	6	3	3	+	7
6	4	0	4	+	9
7	7	4	3	+	7
8	8	5	3	+	7
9	5	4	1	+	2
10	1	0	1	+	2

Step 3:

$W^- = 4.5 = 4.5$

$W^+ = 2 + 4.5 + 7 + 9 + 7 + 7 + 2 + 2 = 40.5$

Step 4:

Calculate the z Statistic for the Two Groups

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

$$z = \frac{4.5 - 22.5}{\sqrt{71.25}} = \frac{-18}{8.44} = -2.13 = |-2.13| = 2.13$$

Step 5:

Table Value:

Find the t-table value:

1. The alpha level: 5%
2. The degrees of freedom, (n):

   df = 9
3. Table Value = 5

Conclusion:

Compare calculated value to table value. The calculated value is less than the table value (2.13<5). We accept the null hypothesis. Hence, we conclude that there is no significant difference in the Pretest and Post Test.

Spearman correlation

Spearman rank correlation is a non-parametric test that is used to measure the degree of association between two variables. The Spearman rank correlation test does not carry any assumptions about the distribution of the data and is the appropriate correlation analysis when the variables are measured on a scale that is at least ordinal.

The following formula is used to calculate the Spearman rank correlation:

$$r = 1 - \frac{6\sum D^2}{n^3 - n}$$

r = Spearman rank correlation
d = the difference between the ranks of corresponding variables
n= number of observations

Types of research questions a Spearman Correlation can examine:

Is there a statistically significant relationship between participant's level of education (high school, bachelors, or graduate degree) and their starting salary?

Assumptions

The assumptions of the Spearman correlation are that data must be at least ordinal and the scores on one variable must be monotonically related to the other variable.

Problem

A pediatrician speculates that the length of time an infant is breast fed may be related to how often a child becomes ill. In order to answer the question, the pediatrician obtains the following two scores for five three-year-old children: The number of months the child was breast fed (which represents the X variable) and the number of times the child was brought to the pediatrician's office during the current year (which represents the Y variable). The scores for the five children follow:

Child	1	2	3	4	5
breast fed	20	0	1	12	3
child becomes ill	7	0	2	5	3

Do the data indicate that the length of time a child is breast fed is related to the number of times a child is brought to the pediatrician?

Procedure

- State the null and alternate hypothesis.

- Putting into rank of two groups.
- Calculate the Rank Difference
- Calculate the test Statistic for the Two Groups

$$r = 1 - \frac{6\sum D^2}{n^3 - n}$$

- Determine the Critical value from Table

Calculation:

Step 1: Hypothesis:

Null hypothesis**:** There is no significant relationship between the lengths of time a child is breast fed is related to the number of times a child is brought to the pediatrician.

Alternative hypothesis**:** There is significant relationship between the lengths of time a child is breast fed is related to the number of times a child is brought to the pediatrician.

Step 2:

No. of times breast fed	Rank	No. of times child ill	Rank	D	D^2
20	5	7	5	0	0
0	1	0	1	0	0
1	2	2	2	0	0
12	4	5	4	0	0
3	3	3	3	0	0

Step 3:

$$r = 1 - \frac{6\sum D^2}{n^3 - n}, \quad r = 1 - \frac{6 \times 0}{5^3 - 5}, \quad r = 1 - \frac{0}{125 - 5}, \quad r = 1 - 0, \quad r = 1$$

Step 4:

Table Value: Find the t-table value:

1. The alpha level: 5%
2. The degrees of freedom, (n): df = 5
3. Table Value = 0.900

Conclusion

Since, r = 1, the variables are highly positively correlated. Compare calculated value to table value. The calculated value is greater than the table value (1>0.900). We reject the null hypothesis. Hence, we conclude that there is significant relationship between the lengths of time a child is breast fed is related to the number of times a child is brought to the pediatrician.

Kruskal Wallis test

The Kruskal-Wallis H test is a rank-based nonparametric test that can be used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable. It is considered the nonparametric alternative to the one-way ANOVA, and an extension of the Mann-Whitney U test to allow the comparison of more than two independent groups.

Assumptions

There are certain assumptions in the Kruskal-Wallis test.

- It is assumed that the observations in the data set are independent of each other.
- It is assumed that the distribution of the population should not be necessarily normal and the variances should not be necessarily equal.
- It is assumed that the observations must be drawn from the population by the process of random sampling.

Problem

In order to assess the efficacy of a new antidepressant drug, 15 clinically depressed patients are randomly assigned to one of three groups. Five patients are assigned to Group 1, which is administered the antidepressant drug for a period of six months. Five patients are assigned to Group 2, which is administered a placebo during the same six-month period. Five patients are assigned to Group 3, which does not receive any treatment during the six-month period. Assume that prior to introducing the experimental treatments; the experimenter confirmed that the level of depression in the three groups was equal. After six months elapse, all 15 subjects are rated by a psychiatrist (who is blind with respect to a subject's experimental condition) on their level of depression. The psychiatrist's depression ratings for the five subjects in each group follow.

Antidepressant drug	8	10	9	10	9
Placebo	7	8	5	8	5
No treatment	4	8	7	5	7

Do the data indicate that the antidepressant drug is effective?

Procedure

- State the null and alternate hypothesis.
- Putting into rank of all the groups.
- Calculate the Rank sum all group values
- Calculate the test Statistic

$$H = \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1)$$

Where,

$\sum \frac{R_j^2}{n_j}$ = sum of squares of all groups

n = Sample Size

- Determine the Critical value from Table

Calculation:

Step 1:

Hypothesis:

Null hypothesis**:** All the three groups are exposed to same levels of noise.

Alternative hypothesis**:** there is a significant difference between at least two of the three groups exposed to different levels of drugs.

Step 2: Calculate rank of all the groups.

Antidepressant drug	8	10	9	10	9	Total	$(\text{Total})^2$
Rank	***9.5***	***14.5***	***12.5***	***14.5***	***12.5***	***63.5***	***4032.25***
Placebo	7	8	5	8	5		
Rank	***6***	***9.5***	***3***	***9.5***	***3***	***31***	***961***
No treatment	4	8	7	5	7		
Rank	***1***	***9.5***	***6***	***3***	***6***	***25.5***	***650.25***

Step 3: Compute test statistic

$$\sum \frac{R_j^2}{n_j} = \frac{4032.25}{5} + \frac{961}{5} + \frac{650.25}{5} = 1128.7$$

n = 15

$$H = \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1) = \frac{12}{15 \times 16} \times 1128.7 - 3(16)$$

= 56.435 – 48

H = 8.435

Step 4: Table Value: Find the t-table value:

1. The alpha level: 5%
2. The degrees of freedom, (Number of Groups Minus 1):
   df = 3 – 1 = 2
3. Table Value = 5.991

Conclusion

Compare calculated value to table value. The calculated value is greater than the table value (8.435>5.991). We reject the null hypothesis. Hence, we conclude that there is a significant difference between at least two of the three groups exposed to different levels of drugs.

Friedman Test

Friedman Test is a non-parametric test alternative to the one way ANOVA with repeated measures. It tries to determine if subjects changed significantly across occasions/conditions. For example: - Problem-solving ability of a set of people is the same or different in Morning, Afternoon, Evening. It is used to test for differences between groups when the dependent variable is ordinal. This test is particularly useful when the sample size is very small.

Elements of Friedman Test

- One group that is measured on three or more blocks of measures overtime/experimental conditions.
- One dependent variable which can be Ordinal, Interval or Ratio.

Assumptions of Friedman Test

- The group is a random sample from the population
- No interaction between blocks (rows) and treatment levels (columns)
- The one group that is measured on three or more different occasions
- Data should be at least an ordinal or continuous
- The samples are do not need to be normally distributed

Problem

In order to assess the efficacy of a drug which a pharmaceutical company claims is effective in treating hyperactivity, six hyperactive children are evaluated during the following three time periods: a) One week prior to taking the drug; b) After a child has taken the drug for six consecutive months; and c) Six months after the drug is discontinued. The children are observed by judges who employ a standardized procedure for evaluating hyperactivity. During each time period a child is assigned a score between 0 and 10, in which the higher the score, the higher the level of hyperactivity. During the evaluation process, the judges are blind with respect to whether or not a child is taking medication at the time he or she is evaluated. The hyperactivity scores of the six children during the three time periods follow:

Child	One week prior to taking the drug	After a child has taken the drug for six consecutive months	Six months after the drug is discontinued
1	9	7	4
2	10	8	7
3	7	5	3
4	10	8	7
5	7	5	2
6	8	6	6

Do the data indicate that the drug is effective?

Procedure

- State the null and alternate hypothesis.
- Assign Ranks for the drugs corresponding to each person and find the sum.
- Calculate the test Statistic

$$F_R = \frac{12}{nk(k+1)} \sum R_i^2 - 3n(k+1)$$

Where,

n = total number of subjects/participants.
k = total number of blocks to be measured.
R_i = sum of ranks of all subjects for a block i

- Determine the Critical value from Table

Calculation:

Step 1: Hypothesis:

Null hypothesis**:** All three drugs have the same effective.
Alternative hypothesis**:** At least two of them differ from each other.

Step 2:

Assign Ranks for the drugs corresponding to each person and find the sum.

Child	One week drug	R_1	drug for six months	R_2	Six months after the drug is discontinued	R_3
1	9	***3***	7	***2***	4	***1***
2	10	***3***	8	***2***	7	***1***
3	7	***3***	5	***2***	3	***1***
4	10	***3***	8	***2***	7	***1***
5	7	***3***	5	***2***	2	***1***
6	8	***3***	6	***1.5***	6	***1.5***
Total		***18***		***11.5***		***6.5***
Mean		***3***		***1.92***		***1.08***

Step 3:

Calculate the test Statistic

$$F_R = \frac{12}{nk(k+1)} \sum R_i^2 - 3n(k+1)$$

$$= \frac{12}{(6)(3)(3+1)}[(18)^2 + (11.5)^2 + (6.5)^2] - (3)(6)(3+1)$$

$F_R = 11.08$

Step 4:

Table Value: Find the t-table value:

1. The alpha level: 5%
2. The degrees of freedom, F(k = 3, n = 6, α = 0.05)
   df = 7
3. Table Value = 5.991

Conclusion

Compare calculated value to table value. The calculated value is greater than the table value (11.08>7). We reject the null hypothesis. Hence, we conclude that there is a significant difference between at least two of them differ from each other.

Pearson's Correlation

Correlation coefficients are used to measure how strong a relationship is between two variables. There are several types of correlation coefficient, but the most popular is Pearson's.

Pearson's correlation (also called Pearson's *R*) is a correlation coefficient commonly used in linear regression.

Pearson's correlation coefficient returns a value between -1 and 1. The interpretation of the correlation coefficient is as under:

- If the correlation coefficient is -1, it indicates a strong negative relationship. It implies a perfect negative relationship between the variables.
- If the correlation coefficient is 0, it indicates no relationship.
- If the correlation coefficient is 1, it indicates a strong positive relationship. It implies a perfect positive relationship between the variables.

A higher absolute value of the correlation coefficient indicates a stronger relationship between variables. Thus, a correlation coefficient of 0.78 indicates a stronger **positive correlation** as compared to a value of say 0.36. Similarly, a correlation coefficient of -0.87 indicates a stronger **negative correlation** as compared to a correlation coefficient of say - 0.40.

Problem

A psychologist conducts a study employing a sample of five children to determine whether there is a statistical relationship between the number of ounces of sugar a ten-year-old child eats per week (which will represent the X variable) and the number of cavities in a child's mouth (which will represent the Y variable). The two scores (ounces of sugar consumed per week and number of cavities) obtained for each of the five children follow:

Child	1	2	3	4	5
Sugar consumption	20	0	1	12	3
Number of cavities	7	0	2	5	3

Is there a significant correlation between sugar consumption and the number of cavities?

Procedure:

- Find the mean of the two series x and y.
- Square the deviations and get the total, of the respective squares of deviations of x and y and denote by ΣX^2 , ΣY^2 respectively.
- Multiply the deviations of x and y and get the total and Divide by n. This is covariance.
- Substitute the values in the formula.

$$r = \frac{COV(x.y)}{\sigma_x.\sigma_y} = \frac{\sum(x-\bar{x})\ (y-y)/n}{\sqrt{\frac{\sum(x-\bar{x})}{n}}\cdot\sqrt{\frac{\sum(y-\bar{y})}{n}}}$$

The above formula is simplified as follows

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2\ .\Sigma Y^2}} \qquad X = x - \bar{x}, Y = y - \bar{y}$$

Calculation

Step – 1

Find the mean of the two series x and y.

$\bar{x} = \frac{\sum x_i}{n} = \frac{36}{5}$ $\bar{y} = \frac{\sum y_i}{n} = \frac{17}{5}$

$\bar{x} = 7.2$ $\bar{y} = 3.4$

Child	X	Y	X^2	Y^2	XY
1	20	7	400	49	140
2	0	0	0	0	0
3	1	2	1	4	2
4	12	5	144	25	60
5	3	3	9	9	9
Total	**36**	**17**	**554**	**87**	**211**

Step – 3

Substitute the values in the formula

$$r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}}$$, $r = \frac{211}{\sqrt{554*87}}$, $r = \frac{211}{\sqrt{48198}}$, $r = \frac{211}{219.54}$, $r = 0.961$

Conclusion

Since, r = + 0.961, the variables are highly positively correlated. Hence we conclude that, there a significant correlation between sugar consumption and the number of cavities.

Resampling

Resampling techniques are a set of methods to either repeat sampling from a given sample or population, or a way to estimate the precision of a statistic. Although the method sounds daunting, the math involved is relatively simple and only requires a high school level understanding of algebra.

Informally, resample can mean something a little simpler: repeat any sampling method. For example, if you're conducting a Sequential Probability Ratio Test and don't come to a conclusion, then you resample and rerun the test. For most intents and purposes though, if you read about resampling (as opposed to "resample"), then the author is most likely talking about a specific resampling technique.

Resampling Techniques

1. Bootstrapping and Normal resampling (sampling from a normal distribution).
2. Permutation Resampling (also called Rearrangements or Rerandomization),
3. Cross Validation.

1. Bootstrapping and Normal Resampling

Bootstrapping is a type of resampling where large numbers of smaller samples of the same size are repeatedly drawn, with replacement, from a single original sample. **Normal resampling** is very similar to bootstrapping as it is a special case of the normal shift model—one of the assumptions for bootstrapping (Westfall et al., 1993). Both bootstrapping and normal resampling both assume that samples are drawn from an actual population (either a real one or a theoretical one). Another similarity is that both techniques use sampling *with* replacement.

Ideally, you would want to draw large, non-repeated, samples from a population in order to create a sampling distribution for a statistic. However, limited resources may prevent you from getting the ideal statistic. Resampling means that you can draw small samples over and over again from the same population. As well as saving time and money, the samples can be quite good approximations for population parameters.

2. Permutation Resampling

Unlike bootstrapping, permutation resampling doesn't need any "population"; resampling is dependent only on the assignment of units to treatment groups. The fact that you're dealing with actual samples, instead of populations, is one reason why it's sometimes referred to as the Gold standard bootstrapping technique (Strawderman and Mehr, 1990). Another important difference is that permutation resampling is a without replacement sampling technique.

3. Cross Validation

Cross-validation is a way to validate a predictive model. Subsets of the data are removed to be used as a validating set; the remaining data is used to form a training set, which is used to predict the validation set.

Post Hoc Test in ANOVA

ANOVA can be used to determine if three or more means are different, it provides no information concerning where the difference lies. For example, if Ho: $mean_1 = mean_2 = mean_3$ is rejected, then there are three alternate hypotheses that can be tested: $mean_1 \neq mean_2 \neq mean_3$, $mean_1 \neq mean_2 = mean_3$, or $mean_1 = mean_2 \neq mean_3$. Methods have been constructed to test these possibilities, and they are termed multiple comparison post-tests. There are several tests are as followed.

- Duncan's multiple range test [DMRT]
- Tukey's multiple Comparison Test

MULTIPLE RANGE TEST [MRT]:

In the case of significance F, the null hypothesis rejected then the problem is known which of the treatment means are significantly different. Many test procedures are available for this purpose. The most commonly used test is,

- Lest significance difference [is known as critical difference]
- Duncan's multiple range test [DMRT]

Critical difference (C.D):

The critical difference is a form of t-test is formula is given by

C.D = t.S.E(d)

Where SE = Standard Error

$$S.E(d) = \sqrt{EMS\left(\frac{1}{r_i} - \frac{1}{r_j}\right)}$$

EMS= Error mean Square

In the case of same replication the standard is

$$S.E = \sqrt{\left(\frac{2EMS}{r}\right)}$$

In this formula t is the critical (table) value of t for a specified level of significance and error degrees of freedom r_i and r_j for the number of replications for the i^{th} and j^{th} treatment respectively, the formula for t-test is

$$t = \frac{Y_i - Y_j}{S\sqrt{\frac{1}{r_i} - \frac{1}{r_j}}}$$

The two treatment means are declared significantly different at specified level of significance. If the difference exceeds the calculated CD value, otherwise they are not significant CD value.

Duncan's multiple range test (DMRT):

In a set of t-treatments if the comparison of all possible pairs of treatment mean is required. We can use Duncan's multiple range test. The DMRT can be used irrespective of whether F is significant or not.

Procedure:

Step: 1 Arrange the treatments in descending order that is to range.

Step: 2 Calculate the S.E of mean as

$$S.E(\bar{Y}) = \sqrt{\left(\frac{SQ^2}{r}\right)} = \sqrt{\left(\frac{EMS}{r}\right)}$$

Step: 3 From statistical table write the significant student zed range as (rp), p=1,2,………t treatment and error degrees of freedom.

Step: 4 Calculate the shortest significance range as R_p where $R_{p}= r_p \,.\, S.E(\bar{Y})$

Step: 5 From the largest mean subtract the R_p for largest P. Declare as significantly different from the largest mean. For the remaining treatment whose values are larger than the difference (largest mean-largest R_p). Compare the difference with appropriate R_p value.

Step: 6 Continue this process till all the treatment above.

Step: 7 Present the results by using either the line notation (or) the alphabet notation to indicate which treatment pair which are significantly different from each other.

Tukey's range test:

Tukey's range test is also known as Tukey's test, Tukey's HSD (Honest significance difference) test. It can be used on raw data or in cons unction with an ANOVA (post-hoc analysis) to find means that are significantly different from each other. Tukey's test compares the means of every treatment to the means of every other treatment.

The test statistic: Tukey's test is based on a formula very similar to that of the t-test. In fact, Tukey's test is essentially a t-test, except that is corrects for experiment wise error rate. Formula to,

$$q_s = \frac{Y_A - Y_B}{S.E}$$

Where Y_A is a larger of the two means being compared. Y_B is the smaller of the two means being compared. S.E is the standard error. This q_s value can then be compared to a q value from the studentized range distribution. If the q_s value is larger than the q critical value obtained from the distribution. The two means are said to significantly different. The studentized range distribution:

$$q = \frac{\dfrac{(\bar{y}_{\max} - \bar{y}_{\min})}{S}}{\sqrt{\dfrac{2}{n}}}$$

Starting with the ANOVA

Hypothesis

Null Hypothesis: All group means are equal.

Alternative Hypothesis: Not all group means are equal.

Example One-Way ANOVA to Use with Post Hoc Tests

We want to determine whether the mean differences between the strengths of these four materials are statistically significant. We obtain the following ANOVA results. To follow along with this example:

Material	*B*	*C*	*D*	*A*	*A*	*C*	*A*	*B*
Strength	37.9	36	38	40	36.9	39.4	33.4	26.2
Material	*B*	*B*	*D*	*B*	*D*	*D*	*C*	*A*
Strength	24.9	30.3	40.8	32.6	45.9	40.4	36.3	42.3
Material	*A*	*C*	*A*	*D*	*D*	*C*	*B*	*C*
Strength	39.1	29.5	34.7	39.9	41.4	34.9	37.5	39.8

One-way ANOVA: Strength versus Material

```
Method

Null hypothesis         All means are equal
Alternative hypothesis  At least one mean is different
Significance level      α = 0.05

Equal variances were assumed for the analysis.


Factor Information

Factor    Levels  Values
Material       4  A, B, C, D


Analysis of Variance

Source    DF  Adj SS  Adj MS  F-Value  P-Value
Material   3   281.7   93.90     6.02    0.004
Error     20   312.1   15.60
Total     23   593.8

Means

Material  N   Mean  StDev       95% CI
A         6  37.73   3.36  (34.37, 41.10)
B         6  31.57   5.50  (28.20, 34.93)
C         6  35.98   3.73  (32.62, 39.35)
D         6  41.07   2.64  (37.70, 44.43)
```

The p-value of 0.004 indicates that we can reject the null hypothesis and conclude that the four means are not all equal. The Means table at the bottom displays the group means. However, we don't know which pairs of groups are significantly different. To compare group means, we need to perform post hoc tests, also known as multiple comparisons. Before we get to these group comparisons, you need to learn about the experiment-wise error rate.

Experiment-wise Error Rate

Post hoc tests perform two vital tasks. Yes, they tell you which group means are significantly different from other group means. Crucially, they also control the experiment-wise or family wise, error rate.

For every hypothesis test, there is a type I error rate, which your significance level (alpha) defines. In other words, there's a chance that you'll reject a null hypothesis that is actually true, it's a false positive. When you perform only one test, the type I error rate equals your significance level, which is often 5%. However, as you conduct more and more tests, your chance of a false positive increases. If you perform enough tests, you're virtually guaranteed to get a false positive! The error rate for a family of tests is always higher than an individual test.

Imagine you're rolling a pair of dice and rolling two ones (known as snake eyes) represents a Type I error. The probability of snake eyes for a single roll is ~2.8% rather than 5%, but you get the idea. If you roll the dice just once, your chances of rolling snake eyes aren't too bad. However, the more times you roll the dice, the more likely you'll get two ones. With 25 rolls, snake eyes become more likely than not (50.8%). With enough rolls, it becomes inevitable.

Family Error Rates in ANOVA

In the ANOVA context, you want to compare the group means. The more groups you have, the more comparison tests you need to perform. For our example ANOVA with four groups (A B C D), we'll need to make the following six comparisons.

- A – B
- A – C
- A – D
- B – C
- B – D
- C – D

Our experiment includes this family of six comparisons. Each comparison represents a roll of the dice for obtaining a false positive. What's the error rate for six comparisons? Unfortunately, as you'll see next, the experiment-wise error rate snowballs based on the number of groups in your experiment.

The Experiment-wise Error Rate Quickly Becomes Problematic!

The table below shows how increasing the number of groups in your study causes the number of comparisons to rise, which in turn raises the family-wise error rate. Notice how quickly the quantity of comparisons increases by adding just a few groups! Correspondingly, the experiment-wise error rate rapidly becomes problematic.

The table starts with two groups, and the single comparison between them has an experiment-wise error rate that equals the significance level (0.05). Unfortunately, the family-wise error rate rapidly increases from there!

All Pairwise Comparisons Alpha = 0.05		
Groups	Comparisons	Experimentwise Error Rate
2	1	0.05
3	3	0.142625
4	6	0.264908109
5	10	0.401263061
6	15	0.53670877
7	21	0.659438374
8	28	0.762173115
9	36	0.842220785
10	45	0.900559743
11	55	0.940461445
12	66	0.966134464
13	78	0.981700416
14	91	0.990606054
15	105	0.995418807
		From StatisticsByJim.com

The formula for the maximum number of comparisons you can make for N groups is:

$$(N*(N-1))/2.$$

The total number of comparisons is the family of comparisons for your experiment when you compare all possible pairs of groups (i.e., all pair wise comparisons). Additionally, the formula for calculating the error rate for the entire set of comparisons is $1 – (1 – \alpha)^C$. Alpha is your significance level for a single comparison, and C equals the number of comparisons.

The experiment-wise error rate represents the probability of a type I error (false positive) over the total family of comparisons. Our ANOVA example has four groups, which produces six comparisons and a family-wise error rate of 0.26. If you increase the groups to five, the error rate jumps to 40%! When you have 15 groups, you are virtually guaranteed to have a false positive (99.5%)!

Post Hoc Tests Control the Experiment-wise Error Rate

The table succinctly illustrates the problem that post hoc tests resolve. Typically, when performing statistical analysis, you expect a false positive rate of 5%, or whatever value you set for the significance level. As the table shows, when you increase the number of groups from 2 to 3, the error rate nearly triples from 0.05 to 0.143. And, it quickly worsens from there!

These error rates are too high! Upon seeing a significant difference between groups, you would have severe doubts about whether it was a false positive rather than a real difference.

If you use 2-sample t-tests to systematically compare all group means in your study, you'll encounter this problem. You'd set the significance level for each test (e.g., 0.05), and then the number of comparisons will determine the experiment-wise error rate, as shown in the table.

Fortunately, post hoc tests use a different approach. For these tests, you set the experiment-wise error rate you want for the entire set of comparisons. Then, the post hoc test calculates the significance level for all individual comparisons that produces the family wise error rate you specify.

Example of Using Tukey's Method with One-Way ANOVA

For our ANOVA example, we have four groups that require six comparisons to cover all combinations of groups. We'll use a post hoc test and specify that the family of six comparisons should collectively produce a family wise error rate of 0.05. The post hoc test we'll use is Tukey's method. There are a variety of post hoc tests you can choose from, but Tukey's method is the most common for comparing all possible group pairings.

There are two ways to present post hoc test results—adjusted p-values and simultaneous confidence intervals. I'll show them both below.

Adjusted P-values

The table below displays the six different comparisons in our study, the difference between group means, and the adjusted p-value for each comparison.

Tukey Simultaneous Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
B - A	-6.17	2.28	(-12.55, 0.22)	-2.70	0.061
C - A	-1.75	2.28	(-8.14, 4.64)	-0.77	0.868
D - A	3.33	2.28	(-3.05, 9.72)	1.46	0.478
C - B	4.42	2.28	(-1.97, 10.80)	1.94	0.245
D - B	9.50	2.28	(3.11, 15.89)	4.17	0.002
D - C	5.08	2.28	(-1.30, 11.47)	2.23	0.150

Individual confidence level = 98.89%

The adjusted p-value identifies the group comparisons that are significantly different while limiting the family error rate to your significance level. Simply compare the adjusted p-values to your significance level. When adjusted p-values are less than the significance level, the difference between those group means is statistically significant. Importantly, this process controls the family-wise error rate to your significance level. We can be confident that this entire set of comparisons collectively has an error rate of 0.05.

In the output above, only the D – B difference is statistically significant while using a family error rate of 0.05. The mean difference between these two groups is 9.5.

t Distribution: Critical Values of t

Degrees of freedom		*Significance level*					
	Two-tailed test:	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
21		1.721	2.080	2.518	2.831	3.527	3.819
22		1.717	2.074	2.508	2.819	3.505	3.792
23		1.714	2.069	2.500	2.807	3.485	3.768
24		1.711	2.064	2.492	2.797	3.467	3.745
25		1.708	2.060	2.485	2.787	3.450	3.725
26		1.706	2.056	2.479	2.779	3.435	3.707
27		1.703	2.052	2.473	2.771	3.421	3.690
28		1.701	2.048	2.467	2.763	3.408	3.674
29		1.699	2.045	2.462	2.756	3.396	3.659
30		1.697	2.042	2.457	2.750	3.385	3.646
32		1.694	2.037	2.449	2.738	3.365	3.622
34		1.691	2.032	2.441	2.728	3.348	3.601
36		1.688	2.028	2.434	2.719	3.333	3.582
38		1.686	2.024	2.429	2.712	3.319	3.566
40		1.684	2.021	2.423	2.704	3.307	3.551
42		1.682	2.018	2.418	2.698	3.296	3.538
44		1.680	2.015	2.414	2.692	3.286	3.526
46		1.679	2.013	2.410	2.687	3.277	3.515
48		1.677	2.011	2.407	2.682	3.269	3.505
50		1.676	2.009	2.403	2.678	3.261	3.496
60		1.671	2.000	2.390	2.660	3.232	3.460
70		1.667	1.994	2.381	2.648	3.211	3.435
80		1.664	1.990	2.374	2.639	3.195	3.416
90		1.662	1.987	2.368	2.632	3.183	3.402
100		1.660	1.984	2.364	2.626	3.174	3.390
120		1.658	1.980	2.358	2.617	3.160	3.373
150		1.655	1.976	2.351	2.609	3.145	3.357
200		1.653	1.972	2.345	2.601	3.131	3.340
300		1.650	1.968	2.339	2.592	3.118	3.323

Critical Values of the Mann-Whitney U
(Two-Tailed Testing)

n_2	α	n_1																	
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	.05	--	0	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
	.01	--	0	0	0	0	0	0	0	0	1	1	1	2	2	2	2	3	3
4	.05	--	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
	.01	--	--	0	0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5	.05	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
	.01	--	--	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6	.05	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
	.01	--	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7	.05	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
	.01	--	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8	.05	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
	.01	--	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9	.05	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
	.01	0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10	.05	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
	.01	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11	.05	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
	.01	0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	48
12	.05	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
	.01	1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13	.05	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
	.01	1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60
14	.05	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
	.01	1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15	.05	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
	.01	2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16	.05	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
	.01	2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17	.05	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
	.01	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
18	.05	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
	.01	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19	.05	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
	.01	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
20	.05	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
	.01	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105

Critical Values of the Wilcoxon Signed Ranks Test

n	Two-Tailed Test		One-Tailed Test	
	α = .05	α = .01	α = .05	α = .01
5	--	--	0	--
6	0	--	2	--
7	2	--	3	0
8	3	0	5	1
9	5	1	8	3
10	8	3	10	5
11	10	5	13	7
12	13	7	17	9
13	17	9	21	12
14	21	12	25	15
15	25	15	30	19
16	29	19	35	23
17	34	23	41	27
18	40	27	47	32
19	46	32	53	37
20	52	37	60	43
21	58	42	67	49
22	65	48	75	55
23	73	54	83	62
24	81	61	91	69
25	89	68	100	76
26	98	75	110	84
27	107	83	119	92
28	116	91	130	101
29	126	100	140	110
30	137	109	151	120

Critical Values of Spearman's Rank Correlation Coefficient R_s

	Nominal α					
n	0.10	0.05	0.025	0.01	0.005	0.001
4	1.000	1.000	-	-	-	-
5	0.800	0.900	1.000	1.000	-	-
6	0.657	0.829	0.886	0.943	1.000	-
7	0.571	0.714	0.786	0.893	0.929	1.000
8	0.524	0.643	0.738	0.833	0.881	0.952
9	0.483	0.600	0.700	0.783	0.833	0.917
10	0.455	0.564	0.648	0.745	0.794	0.879
11	0.427	0.536	0.618	0.709	0.755	0.845
12	0.406	0.503	0.587	0.678	0.727	0.818
13	0.385	0.484	0.560	0.648	0.703	0.791
14	0.367	0.464	0.538	0.626	0.679	0.771
15	0.354	0.446	0.521	0.604	0.654	0.750
16	0.341	0.429	0.503	0.582	0.635	0.729
17	0.328	0.414	0.488	0.566	0.618	0.711
18	0.317	0.401	0.472	0.550	0.600	0.692
19	0.309	0.391	0.460	0.535	0.584	0.675
20	0.299	0.380	0.447	0.522	0.570	0.662
21	0.292	0.370	0.436	0.509	0.556	0.647
22	0.284	0.361	0.425	0.497	0.544	0.633
23	0.278	0.353	0.416	0.486	0.532	0.621
24	0.271	0.344	0.407	0.476	0.521	0.609
25	0.265	0.337	0.398	0.466	0.511	0.597
26	0.259	0.331	0.390	0.457	0.501	0.586
27	0.255	0.324	0.383	0.449	0.492	0.576
28	0.250	0.318	0.375	0.441	0.483	0.567
29	0.245	0.312	0.368	0.433	0.475	0.558

CHI-SQUARE TABLE

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Critical values for Friedman's two-way analysis of Variance by ranks

	$k = 3$		$k = 4$		$k = 5$		$k = 6$	
b	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
2	-	-	6.000	-	7.600	8.000	9.143	9.714
3	6.000	-	7.400	9.000	8.533	10.13	9.857	11.76
4	6.500	8.000	7.800	9.600	8.800	11.20	10.29	12.71
5	6.400	8.400	7.800	9.960	8.960	11.68	10.49	13.23
6	7.000	9.000	7.600	10.20	9.067	11.87	10.57	13.62
7	7.143	8.857	7.800	10.54	9.143	12.11		
8	6.250	9.000	7.650	10.50	9.200	12.30		
9	6.222	9.556	7.667	10.73	9.244	12.44		
10	6.200	9.600	7.680	10.68				
11	6.545	9.455	7.691	10.75				
12	6.500	9.500	7.700	10.80				
13	6.615	9.385	7.800	10.85				
14	6.143	9.143	7.714	10.89				
15	6.400	8.933	7.720	10.92				
16	6.500	9.375	7.800	10.95				
17	6.118	9.294	7.800	11.05				
18	6.333	9.000	7.733	10.93				
19	6.421	9.579	7.863	11.02				
20	6.300	9.300	7.800	11.10				
21	6.095	9.238	7.800	11.06				
22	6.091	9.091	7.800	11.07				
23	6.348	9.391						
24	6.250	9.250						