NONPARAMETRIC INFERENCE

MEASUREMENT SCALES

Measurement is a process whereby values (scores) are assigned to characteristics of people, places, objects or events depending upon certain attributes and according to certain rules. We might score preferences of sport or TV show, collect data on marital status or gender, count number sixes scored in a cricket match.

The goal of measurement systems is to structure the rule of assigning scores to objects in such a way that the relationship between the objects is preserved in the scores assigned to them. The different kinds of relationship preserved are called properties of the measurement system.

Magnitude:

The property of magnitude exists when an object having more of the attribute than another object is given higher score.

i.e. for all i & j, $O_i > O_j \quad \Rightarrow \quad M(O_i) > M(O_j)$, where $M(O_i)$ is the score assigned to the object O_i.

Intervals:

The property of intervals is concerned with relationship of difference between objects.

The property of interval exists, if for all i, j, k *and* l, we have

$$O_i - O_i \geq O_k - O_l \quad \Rightarrow \quad M(O_i) - M(O_j) \geq M(O_k) - M(O_l)$$

A corollary to the above definition is that, if the numbers assigned to two pairs of objects are equally different then the pair of objects must be equally different. Mathematically, for all *i, j, k and l,* we have

$$\boldsymbol{M(O_i) - M(O_j) = M(O_k) - M(O_l) \Rightarrow \quad O_i - O_i = O_k - O_l}$$

Rational Zero:

A measurement system possesses a rational zero, if an object that has none of the attribute in question is assigned the score zero by the system of rules. The object does not need to really exist in the real world, as it is somewhat difficult to visualize a "man with no height".

The property of rational zero exists, if $\boldsymbol{O_0}$ be the object with none of the attribute in question, then $\boldsymbol{M(O_0) = 0}$.

There are four types of measurement scales: *nominal, ordinal, interval and ratio*.

1. **This is the simplest and most elementary scale of measurement. The nominal scale simply names or labels or categorizes objects or event. Typical examples of nominal variables are gender, race, color, city, etc. Frequency distributions are usually used to analyze data measured on a nominal scale. The main statistic computed is the mode. Variables measured on a nominal scale are often referred to as categorical or qualitative variables. Numbers can also be used to classify or categorize, where numbers have no real 'meaning' other than differentiating between objects. For example, in case of Gender, we may use "1" to represent Male and "2" to represent Female.**

2. **The ordinal scale has all the qualities of the nominal scale plus the ability to rank objects according to some attribute. However,**

the intervals between these rankings are not necessarily equal. For example, on a five-point rating scale measuring technical capability, the difference between a rating of 2 and a rating of 3 may not represent the same between a rating of 4 and a rating of 5. There is no "true" zero point for ordinal scales, since the zero point is chosen arbitrarily. The lowest point on the rating scale in the example was arbitrarily chosen to be 1. It could just as well have been 0 or -5. The only mathematical operation allowed on ordinal data is ranking.

3. **An interval scale combines the qualities of the previous scales with equal intervals. For example, temperature, as measured in degrees Fahrenheit or Celsius, constitutes an interval scale. We can say that a temperature of 40 degrees is higher than a temperature of 30 degrees, and that an increase in heat between 20 and 30 degrees is the same as between 30 to 40 degrees. This is because each 10 degrees' interval has the same physical meaning (in terms of kinetic energy of molecules). However, the interval scales have arbitrary zero points (just because we decided to call it zero), rather than an absolute (true) zero. At 0^{o}C water freezes, but that does not mean absence of heat (i.e. absence of any molecular kinetic energy). The mathematical operations allowed are addition and subtraction, but never multiplication or division. So, we cannot say 40^{0} is twice as hot as 20^{0}.**

4. **Ratio scales has all the qualities of the interval scales. In addition, they feature an identifiable absolute zero point, thus we can make statements involving ratios of two observations such as "twice as long" or "half as fast". For example, the Kelvin temperature scale is a ratio scale as at 0 K theoretically there is no heat. Thus, a temperature of 300 Kelvin is twice as high as a**

temperature of 150 Kelvin. Our standard measures of time, distance, volume, height, weight, etc. use ratio scales. Here all mathematical operations, namely addition, subtraction, multiplication and division, are allowed.

Type of Descriptive Statistics	Measurement scale			
	Nominal	Ordinal	Interval	Ratio
Mode, Count, Frequency	Yes	Yes	Yes	Yes
Median, Minimum, Maximum, Range	No	Yes	Yes	Yes
Mean, Variance, SD	No	No	Yes	Yes

In terms of the "amount of information" content, the nominal scale is the "weakest" scale of measurement, whereas ratio scale is the "strongest" scale of measurement.

NONPARAMETRIC INFERENCE

PARAMETRIC STATISTICAL INFERENCE

Statistical procedures where decisions are based on the form of the population distribution (usually normal distribution) and relate to inferences concerning parameters of the distribution (mean, variance, etc.) are known as parametric statistical inference.

In short, if we have a basic knowledge of the underlying distribution of a variable, then we can make predictions about how, in repeated samples of equal size, this particular statistic will "behave", that is, how it is distributed.

Most methods in parametric statistics are designed primarily for use with interval (or ratio) scale data. But in many practical situations the only data available are measured either in nominal scale or in ordinal scale and in such situations methods of parametric statistics should not or cannot be used.

NONPARAMETRIC STATISTICAL INFERENCE

Statistical inferences where decisions are not concerned with the value of one or more parameters are termed *nonparametric*, whereas those inferences whose validity does not depend on a specific probability model in the population are termed *distribution-free*. Though these two terms are not synonymous, procedures of either type are known as nonparametric methods.

Moreover, for many variables of interest, we simply do not know for sure whether they follow normal distribution or not. For example,

a) **is income distributed normally in the population? - probably not,**
b) **the incidence rates of rare diseases are not normally distributed in the population,**
c) **the number of car accidents is also not normally distributed, and**
d) **neither are very many other variables.**

NONPARAMETRIC INFERENCE

A statistical procedure is nonparametric if it satisfies at least one the following criteria:

1. **The method may be used on data with a nominal scale of measurement.**
2. **The method may be used on data with an ordinal scale of measurement.**
3. **The method may be used on data with an interval or ratio scale of measurement, where the distribution functions of the random variable producing the data are not known.**

ADVANTAGES OF NONPARAMETRIC METHODS

1. **Nonparametric methods require few (or no) assumptions about the underlying populations from which the data are obtained. In particular, nonparametric procedures forgo the traditional assumptions that the underlying populations are normal. When in doubt, it is safer to use nonparametric methods.**
2. **Nonparametric techniques are often (although not always) easier to apply than their normal theory counterpart.**
3. **Nonparametric procedures are often quite easy to compute and understand.**
4. **Nonparametric methods are relatively insensitive to outlying observations.**
5. **Nonparametric procedures are applicable in many situations where normal theory procedures cannot be utilized. Many nonparametric procedures require just the rank of the observations, rather than the actual magnitude of the observations, whereas the parametric procedures require the magnitude.**
6. **Nonparametric procedures are only slightly less efficient than their normal theory competitors when the underlying populations are normal, and they can be mildly or wildly more efficient than these competitors when the underlying populations are non-normal.**

DISADVANTAGES OF NONPARAMETRIC METHODS

1. **Nonparametric methods may lack power as compared with more traditional approaches. This is a particular concern if the sample size is small or if the assumptions for the corresponding parametric method (e.g. Normality of the data) hold.**
2. **Nonparametric methods are geared toward hypothesis testing rather than estimation of effects. It is often possible to obtain nonparametric estimates and associated confidence intervals, but this is not generally straightforward.**
3. **Tied values can be problematic when these are common, and adjustments to the test statistic may be necessary.**

ORDER STATISTICS

Let $X_1, X_2, \cdots, X_n$ be random sample of size n from X, a random variable having distribution function $F(x)$. Rank the elements in order of increasing numerical magnitude, resulting in $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$, where $X_{(1)}$ is the smallest sample element $\left(X_{(1)} = min\{X_1, X_2, \cdots, X_n\}\right)$ and $X_{(n)}$ is the largest sample element$\left(X_{(n)} = max\{X_1, X_2, \cdots, X_n\}\right)$.

$X_{(i)}$is called the i-th order statistics. Often distribution of some of the order statistics is of interest, particularly the minimum and maximum sample values, $X_{(1)}$ and $X_{(n)}$ respectively.

Let $X_1, X_2, \cdots, X_n$ be continuous, independent and identically distributed (i.i.d) random variables with common CDF $F(x)$, then prove that the cumulative distribution function of above two order statistics, denoted respectively by $F_{X_{(1)}}(t)$ and $F_{X_{(n)}}(t)$ are

$$F_{X_{(1)}}(t) = 1 - [1 - F(t)]^n$$

$$F_{X_{(n)}}(t) = [F(t)]^n$$

$$
\begin{aligned}
1-F_{X_{(1)}}(t) &= 1-P\left(X_{(1)} \le t\right) \\
&= P\left(X_{(1)} > t\right) \\
&= P\left(X_1 > t, X_2 > t, \ldots, X_n > t\right) \text{ [since smallest is more than } t\text{]} \\
&= P\left(X_1 > t\right)P\left(X_2 > t\right)\ldots P\left(X_n > t\right)
\end{aligned}
$$

[Because of independence]

$$
= \left(1-F(t)\right)\left(1-F(t)\right)\ldots\left(1-F(t)\right)
$$

[Because of identical distribution]

$$
= \left(1-F(t)\right)^n
$$

So, we have

$$
\boldsymbol{F_{X_{(1)}}(t) = 1 - [1 - F(t)]^n}
$$

Moreover,

$$
\begin{aligned}
F_{X_{(n)}}(t) &= P\left(X_{(n)} \le t\right) \\
&= P\left(X_1 \le t, X_2 \le t, \ldots, X_n \le t\right) \text{ [since largest is } \le \boldsymbol{t}\text{]} \\
&= P\left(X_1 \le t\right)P\left(X_2 \le t\right)\ldots P\left(X_n \le t\right) \\
&= F(t)F(t)\ldots F(t)
\end{aligned}
$$

Thus,

$$
\boldsymbol{F_{X_{(n)}}(t) = [F(t)]^n}
$$

We can obtain the PDF of, $\boldsymbol{X_{(1)}}$and $\boldsymbol{X_{(n)}}$, by differentiating corresponding CDF's, as

$$
f_{X_{(1)}}(t) = n\left[1-F(t)\right]^{n-1} f(t) \text{, and}
$$

$$
f_{X_{(n)}}(t) = n\left[F(t)\right]^{n-1} f(t)
$$

NONPARAMETRIC INFERENCE

Many nonparametric procedures are based on ranked data. Data are ranked by ordering them from lowest to highest and assigning them, in order, the integer values from 1 to the sample size, i.e. **lowest value gets a rank of 1, second lowest gets a rank of 2, and proceeding similarly highest number gets a rank equal to the sample size**.

Ties are resolved by assigning each tied value the mean of the ranks they would have received if there were no ties, e.g., if the data sets consist of 117, 119, 119, 125, 128, then the rank data set becomes 1, 2.5, 2.5, 4, 5. (If the two 119's were not tied, they would have been assigned the ranks 2 and 3. The mean of 2 and 3 is 2.5.).

NONPARAMETRIC INFERENCE

Comparing:	Dependent variable	Independent variable	Parametric test (Dependent variable is normally distributed)	Non-parametric test
The means of two INDEPENDENT groups	Continuous/ scale	Categorical/ nominal	Independent t-test	Mann-Whitney test
The means of 2 paired (matched) samples e.g. weight before and after a diet for one group of subjects	Continuous/ scale	Time variable (time 1 = before, time 2 = after)	Paired t-test	Wilcoxon signed rank test
The means of 3+ independent groups	Continuous/ scale	Categorical/ nominal	One-way ANOVA	Kruskal-Wallis test
The 3+ measurements on the same subject	Continuous/ scale	Time variable	Repeated measures ANOVA	Friedman test
Relationship between 2 continuous variables	Continuous/ scale	Continuous/ scale	Pearson's Correlation Co-efficient	Spearman's Correlation Co-efficient (also use for ordinal data)
Predicting the value of one variable from the value of a predictor variable	Continuous/ scale	Any	Simple Linear Regression	
Assessing the relationship between two categorical variables	Categorical/ nominal	Categorical/ nominal		Chi-squared test

Note: The table only shows the most common tests for simple analysis of data.

NONPARAMETRIC INFERENCE

SIGN TEST

Sign test is a nonparametric procedure for making inference about the location parameter of a continuous distribution. In sign test we use median, M as the location parameter. The median of a distribution is a value of the random variable X, such that

$$P(X \leq M) = P(X \geq M) = 0.5.$$

Since, the normal distribution symmetric, the mean of a normal distribution equals the median. Therefore, the sign test can be used to test hypotheses about the mean of a normal distribution, which is nothing but *t*-test. Note that, *t*-test is applicable for samples from a normal distribution, whereas sign test is appropriate for samples from any continuous distribution.

Procedure

Suppose that the hypotheses are

$$\boldsymbol{H_0: M = M_0}$$
$$\boldsymbol{H_1: M < M_0}$$

where $\boldsymbol{M_0}$ is some hypothesized numerical value.

The test procedure is easy to describe. Suppose that $\boldsymbol{X_1, X_2, \cdots, X_n}$ is random sample from the population with unknown median $\boldsymbol{M}$. Find the differences under null hypothesis

$$X_i - M_0, \quad i = 1,2,\cdots,n$$

Now if the null hypothesis is true, any of the above differences are equally likely to be positive or negative. Let $\boldsymbol{R^+ \& \ R^-}$ be the random variable representing the number of these differences that are positive and negative respectively and $\boldsymbol{r^+ \ \& \ r^-}$ be the corresponding

observed number of differences. Therefore, to test the null hypothesis, we are really testing whether R^+, or equivalently R^- follows binomial distribution with $p = \frac{1}{2}$. A P-value for the observed number of r^+ or r^- can be calculated from the binomial distribution. For instance, in testing above hypothesis where most of the differences likely to be negative, we will reject H_0 in favor of H_1 if the proportion of plus signs is significantly less than ½ (or in other words, whenever the observed number of plus signs is too small).

Thus, if the computed P-value

$$P = P(R^+ \leq r^+) \text{ when } p = ½$$

is less or equal to some preselected significance level α, we will reject H_0 and conclude H_1 is true.

To test the other one-sided hypothesis

$$H_0: M = M_0$$
$$H_1: M > M_0$$

we will reject H_0 in favor of H_1 if the observed number of minus signs, r^- is small or, equivalently, whenever the observed fraction of minus signs is significantly less than ½.

Thus, if the computed P-value

$$P = P(R^- \leq r^-) \text{ when } p = ½$$

is less than α, we will reject H_0 and conclude H_1 is true.

The two-sided alternative may also be tested. If the hypotheses are

$$H_0: M = M_0$$
$$H_1: M \neq M_0$$

we should reject H_0 if min *(proportion of plus signs, proportion of minus sign)* is significantly less than ½, or equivalently if min (r^+,r^-) is sufficiently small. Thus, we compute the P-value as

$$P = P(R^+ \leq r^+ \text{ when } p = ½), \text{ if } r^+ \text{ is minimum, or}$$
$$P = P(R^- \leq r^- \text{ when } p = ½), \text{ if } r^- \text{ is minimum.}$$

If the P-value is less than $\alpha/2$, we will reject H_0 and conclude H_1 is true.

Example 1

Ten samples were taken from a plating bath used in an electronics manufacturing process and the bath pH was determined. The sample pH values are 7.91, 7.85, 6.82, 8.01, 7.46, 6.95, 7.05, 7.35, 7.25, and 7.42. Manufacturing personnel believes that pH has median value of 7.0. Do the sample data indicate that this statement is correct? <Use α = 0.05>

Solution

Here the hypothesis to be tested is

$$H_0: M = 7.0$$
$$H_1: M \neq 7.0$$

Observation, *i*	1	2	3	4	5	6	7	8	9	10
pH value (x_i)	7.91	7.85	6.82	8.01	7.46	6.95	7.05	7.35	7.25	7.42
$x_i - 7.0$	+0.91	+0.85	-0.18	+1.01	+0.46	-0.05	+0.05	+0.35	+0.25	+0.42
Sign	+	+	-	+	+	-	+	+	+	+

The number of plus differences, r^+ is 8 and r^- is 2 and minimum of these two is 2. We will reject H_0 if the P-value corresponding to r^- = 2 is less than or equal to $\alpha/2$ = 0.025.

Now, P = P($R^-\leq 2$ when p = ½) = $\sum_{r=0}^{2}\binom{10}{r}(0.5)^r(0.5)^{n-r}$

$$= \sum_{r=0}^{2}\binom{10}{r}(0.5)^n = 0.0547.$$

Since, above P-value is not less than $\alpha/2$ = 0.025, we cannot reject the null hypothesis that the median pH value is 7.0. Thus, observed number of plus signs r^- = 2 was not small enough to indicate that median pH value is different from 7.0 at 5% level of significance.

Ties in the Sign Test

Since the underlying population is assumed to be continuous, probability of getting a tie, i.e. a value exactly equal to M_0, is almost zero. However, this may sometimes happen in practice because of the way the data are collected. When ties occur, they should be ignored and the sign test applied to the remaining data.

The Normal Approximation

When p = 0.5, the binomial distribution is well approximated by a normal distribution *when* $n \geq 10$. Thus, since the mean of the binomial distribution np and the variance is $np(1-p)$, the distribution of R^+ or R^- is approximately normal with mean 0.5*n* and variance 0.25*n*, whenever *n* is moderately large. Therefore, in these cases the null hypothesis can be tested using either of the following (depending upon the hypotheses to be tested):

$$Z_0 = \frac{R^+ - 0.5n}{0.5\sqrt{n}} \text{ and } Z_0 = \frac{R^- - 0.5n}{0.5\sqrt{n}} \quad .$$

The two-sided alternate would be accepted if the test statistic corresponding to the observed value, i.e. $|z_0| > z_{\alpha/2}$. Similarly for one-sided alternates, alternate would be accepted if $z_0 > z_\alpha$ [for H_1: $M > M_0$] and $z_0 < -z_\alpha$ [for H_1: $M < M_0$].

Example 2 <Example 1 using Normal approximation>

Here test statistic is $z_0 = \frac{2-0.5\times10}{0.5\sqrt{10}} = -1.897$. Since $|z_0| = 1.897 < z_{0.025} = 1.96$, we cannot reject the null hypothesis. Thus, the conclusions are identical.

SIGN TEST FOR PAIRED SAMPLES

The sign test can also be applied to paired observations drawn from continuous populations. Let $(X_{1j}, X_{2j}),\ \ j = 1, 2, \cdots, n$ be a collection of paired observations from two continuous populations, and let

$$D_j = X_{1j} - X_{2j},\ \ j = 1, 2, \cdots, n$$

be the paired differences. We wish to test the hypothesis that the two populations have a common median, that is, $M_1 = M_2$. This is equivalent to testing that the median of the differences $M_D = 0$. This can be done by applying the sign test to the n observed differences d_j.

Example 3

Two different formulations of primer can be used on aluminum panels. The drying time of these two formulations is an important consideration in the manufacturing process. Fifteen panels were selected; half of each panel is painted with primer one and the other half are painted with primer two. The drying times were observed and reported in the following table. Is there any evidence that the median drying times of the two formulations are different? <Assume α = 0.01>

Panel	Drying Time (Hrs.)		Differences	Sign
	Formulation 1	Formulation 2		
1	1.6	1.8	-0.2	-
2	1.3	1.5	-0.2	-
3	1.5	1.5		
4	1.6	1.7	-0.1	-
5	1.7	1.6	+0.1	+
6	1.9	2.0	-0.1	-
7	1.8	2.1	-0.3	-
8	1.6	1.7	-0.1	-
9	1.4	1.6	-0.2	-
10	1.8	1.9	-0.1	-
11	1.9	2.0	-0.1	-
12	1.8	1.9	-0.1	-
13	1.7	1.5	+0.2	+
14	1.5	1.7	-0.2	-
15	1.6	1.6		

Note that since in case of two panels drying times are observed to be same they are not used during analysis.

Here, α = 0.01, r^+ = 2 and n = 13. Using normal approximation to binomial, we get the test statistic as

$z_0 = \frac{2-0.5\times 13}{0.5\sqrt{13}} = -2.496$. Since $|z_0|$ = 2.496 < 2.576 = $z_{0.005}$, we cannot reject the null hypothesis that the median drying time for two formulations are same.

THE WILCOXON SIGNED-RANK TEST

The sign test for location uses only the signs of the differences between each observation and the hypothesized median M_0 (or, the difference of each pair of observations), the magnitudes of the differences are not considered. Frank Wilcoxon devised a test procedure that uses both direction (sign) and magnitude. This procedure is known as Wilcoxon Signed-Rank Test.

The Wilcoxon signed-rank test is applicable to symmetric continuous distribution. Under this assumption, mean equals to median and thus we can use this procedure to test the null hypothesis that $\mu = \mu_0$.

Procedure

Suppose we are interested in testing $H_0 : \mu = \mu_0$. Assume that $X_1, X_2, \cdots, X_n$ is a random sample from a continuous and symmetric distribution with mean (and median) μ.

- **Compute the differences $X_i - \mu_0$, $i = 1, 2, \cdots, n$.**
- **Ignore the observation(s) for which the difference(s) is (are) found to be zero (0).**
- **Arrange the absolute differences $|X_i - \mu_0|, i = 1, 2, \cdots, n$ in ascending order and assign ranks, i.e. assign rank "1" to the lowest score, rank "2" to the next lowest score and so on.**

- **Give the ranks the sign of corresponding differences.**
- **Let W^+ = sum of the positive ranks and W^- = absolute value of the sum of the negative ranks.**
- **Let W = min (W^+, W^-).**
- **Obtain corresponding critical values of W (two-sided or one-sided depending upon the case) from Table 10.6 of RMMR Table for a given value of α and let it be W_α.**

Two-sided Test

Let the alternate hypothesis be $H_1: \mu \neq \mu_0$, then if obtained value of $W \leq W_\alpha$, reject the null hypothesis $H_0: \mu = \mu_0$.

One-sided Test

i) **If the alternate hypothesis is $H_1 : \mu < \mu_0$, then if obtained value of $W^+ \leq W_\alpha$, reject the null hypothesis $H_0 : \mu = \mu_0$. [Since under H_1, we expect W^- to be fairly large, or equivalently, W^+ to be fairly small]**

ii) **If the alternate hypothesis is $H_1 : \mu > \mu_0$, *then* if obtained value of $W^- \leq W_\alpha$, reject the null hypothesis $H_0 : \mu = \mu_0$.**

Ties in the Wilcoxon Signed-Rank Test

Because the underlying population is continuous, ties are theoretically impossible, although they will sometimes occur in practice. If several observations have the same absolute difference, they are assigned the average of the ranks that they would receive if they differed slightly from one another.

Example 4 (Wilcoxon Signed Rank Test)

Let us illustrate with data presented in Example 1. Assume that the underlying distribution is a continuous symmetric distribution.

Let the hypotheses to be tested at $\alpha = 0.05$ are

$$H_0 : \mu = 7.0$$
$$H_1 : \mu \neq 7.0$$

Computation

Observation, I	6	7	3	9	8	10	5	2	1	4
pH value (x_i)	6.95	7.05	6.82	7.25	7.35	7.42	7.46	7.85	7.91	8.01
$x_i - 7.0$	-0.05	+0.05	-0.18	+0.25	+0.35	+0.42	+0.46	+0.85	+0.91	+1.01
Signed Rank	-1.5	+1.5	-3	+4	+5	+6	+7	+8	+9	+10

The sum of the positive ranks is $W^+ = (1.5 + 4 + 5 + 6 + 7 + 8 + 9 + 10) = 50.5$ and sum of the absolute values of the negative ranks is $W^- = (1.5 + 3) = 4.5$. Therefore,

$$W = \min(50.5, 4.5) = 4.5.$$

Conclusion

Since, $W = 4.5 < W_{0.05} = 8$, we reject H_0 at 0.05 level of significance. Thus, we conclude that mean pH value differs from 7 based on the 10 sample observations.

Paired Observation

The Wilcoxon signed-rank test can be applied to paired data also. Let (X_{1i}, X_{2i}), $i = 1, 2, \cdots, n$ be a collection of paired observations from two continuous (not necessarily symmetric) distributions that differ only w.r.t. their means. This assures that distribution of the

differences $D_i = X_{1i} - X_{2i}$ is continuous and symmetric. Thus, the null hypothesis $H_0 : \mu_1 - \mu_2$ is equivalent to $H_0 : \mu_D = 0$. We initially consider the two-sided alternative $H_1 : \mu_1 \neq \mu_2$ (or, $H_1 : \mu_D \neq 0$). Note that D_i should always be $X_{1i} - X_{2i}$ and not the other way.

To use Wilcoxon signed-rank test, the differences are first ranked in ascending order of their absolute values and then the ranks are given signs of their differences. Ties are assigned average ranks. Let, W^+ and W^- be sum of the positive ranks and absolute value of sum of the negative ranks of the differences respectively, and W = min(W^+, W^-). If the observed value $W \leq W_\alpha$, the null hypothesis $H_0 : \mu_1 = \mu_2$ (or $H_0 : \mu_D = 0$) is rejected, where W_α is chosen from Table 10.6 of RMMR Table.

For one-sided tests, if the alternative is $H_1 : \mu_1 > \mu_2$ (or, $H_1 : \mu_D > 0$), reject H_0 if $W^- \leq W_\alpha$, and if $H_1 : \mu_1 < \mu_2$ (or, $H_1 : \mu_D < 0$), reject H_0 if $W^+ \leq W_\alpha$, where W_α is the one-sided critical value obtained from Table 10.6 of RMMR Table.

Note: Paired observations with "zero" differences will be ignored and such pairs will not be considered during subsequent calculations. [Number 'zero' has no sign]

Example 5

Fourteen adult males between the age of 35 and 50 participated in a study to evaluate the effect of diet and exercise on blood cholesterol levels. Total cholesterol was measured in each subject initially, and then three months after participating in an aerobic exercise program and switching to a low-fat diet. The data are given in the accompanying table. Assuming the two distributions to be continuous, use the Wilcoxon signed-rank test to ascertain that low-

fat diet and aerobic exercise results in reducing blood cholesterol levels. Use $\alpha = 0.05$.

Subject	1	2	3	4	5	6	7
BC Level Before	265	240	258	295	251	245	287
BC Level After	229	231	227	240	238	241	234

Subject	8	9	10	11	12	13	14
BC Level Before	314	260	279	283	240	238	225
BC Level After	256	247	239	246	218	219	226

Solution

Subject	Blood Cholesterol Level		Difference $[BC_B - BC_A]$	Rank	Signed-rank
	Before	After			
1	265	229	36	9	9
2	240	231	9	3	3
3	258	227	31	8	8
4	295	240	55	13	13
5	251	238	13	4.5	4.5
6	245	241	4	2	2
7	287	234	53	12	12
8	314	256	58	14	14
9	260	247	13	4.5	4.5
10	279	239	40	11	11
11	283	246	37	10	10
12	240	218	22	7	7
13	238	219	19	6	6
14	225	226	-1	1	-1

The hypothesis to be tested here is

$$\boldsymbol{H_0 : \mu_1 = \mu_2 \ (or, H_0 : \mu_D = 0 \ )}$$
$$\boldsymbol{H_1 : \mu_1 > \mu_2 \ (or, H_1 : \mu_D > 0 \ )}$$

We have, W^+ = 104 and W^- = 1. Clearly the statistic here will be W^- = 1.

Since W^- = 1< $W_{0.05}$ = 25, we reject the hypothesis and conclude that use of low-fat diet and aerobic exercise results in reduced blood cholesterol levels.

Example 6

Use Example 3 data for Wilcoxon Signed-Rank Test

Rank of 0.01 is average of (1, 2, …, 7) = 4, of 0.02 is average of (8, 9, 10, 11, 12) = 10 and of 0.03 is 13. W^+ = 14, W^- = 91 - 14 = 77, W= min (14, 77) = 14 and $W_{0.01}$ (for *n* = 13) = 10. Since W is not less than $W_{0.01}$, we cannot reject H_0.

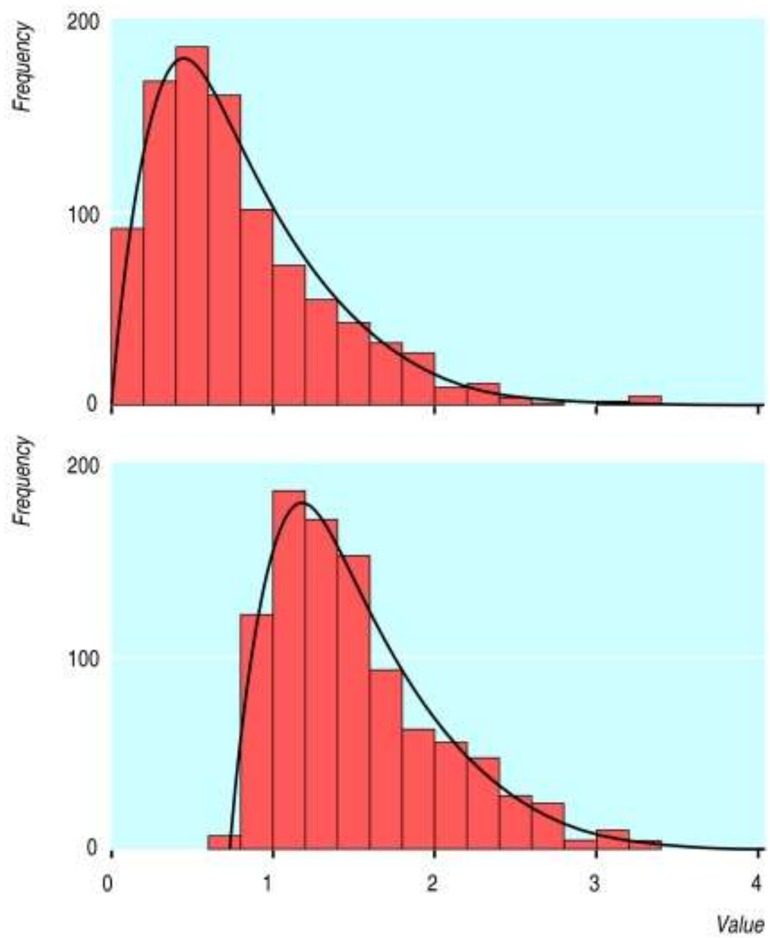
NONPARAMETRIC INFERENCE

MANN-WHITNEY U TEST

***Mann-Whitney U Test is a nonparametric procedure for making inference about the difference between two location parameters*. Let $X_1, X_2, \cdots, X_{n_1}$ (1st sample) and $Y_1, Y_2, \cdots, Y_{n_2}$ (2nd sample) be two independent random samples of size n_1 and n_2 from two continuous populations *X* and *Y* respectively, having same shape and spread. We wish to test the hypotheses**

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

This test is also known as Wilcoxon Rank-Sum Test or Wilcoxon Mann Whitney Test.



Test Procedure

Arrange all $(n_1 + n_2)$ observations in ascending order, keeping their identity, and assign ranks to them. If two or more observations are identical then assign mean rank to each of them. Let R_1 and R_2 be the sum of the ranks assigned to observations from first and second samples respectively. Clearly, $R_1 + R_2$ is the sum of first $(n_1 + n_2)$ positive integers, that is,

$$R_1 + R_2 = \frac{(n_1+n_2)(n_1+n_2+1)}{2}.$$

Now if the sample means do not differ, we will expect that the sum of the ranks for both samples, i.e. R_1 and R_2, to be nearly equal. Consequently, if the sums of the ranks differ greatly, we will conclude that there is an appreciable difference between the means of the two populations.

Mann-Whitney U statistic is defined as

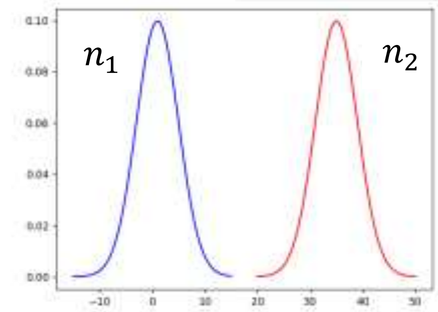
$$U_1 = R_1 - \frac{n_1(n_1+1)}{2} \text{ or } U_2 = R_2 - \frac{n_2(n_2+1)}{2} \text{ or } U = \min(U_1, U_2).$$

[**Note:** If most of the observations from the first sample is less than that of the second sample, then smaller ranks will be mostly from the first sample and consequently U_1 will be small. Moreover, in the extreme case of the distributions being non-overlapping, i.e. all the observations from the first sample are smaller than those of the second sample, ranks $1, 2, \cdots, n_1$ will be from the first sample and consequently U_1 will be zero. Similar will be the case if the opposite becomes true and in that case U_2 will be zero/close to zero/small.**]**

Graphical Presentation:

Case I: $\boldsymbol{\mu_1 < \mu_2}$

$1, 2, \cdots, n_1$ | $n_1 + 1, n_1 + 2, \cdots, n_1 + n_2$



$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2} = 0$$

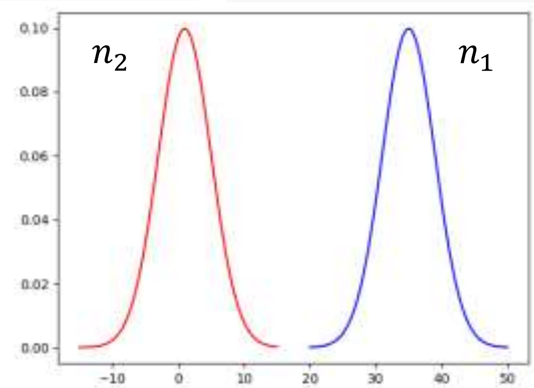
Or, in general U_1 will be smaller.

Case II: $\boldsymbol{\mu_1 > \mu_2}$

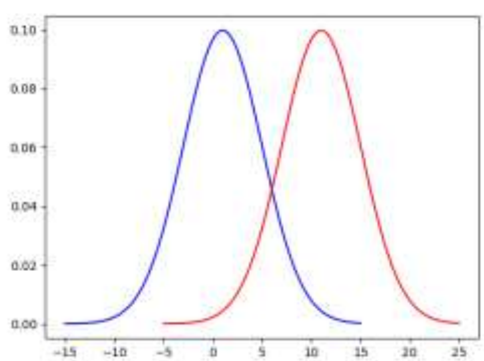
$1, 2, \cdots, n_2$ | $n_2 + 1, n_2 + 2, \cdots, n_2 + n_1$

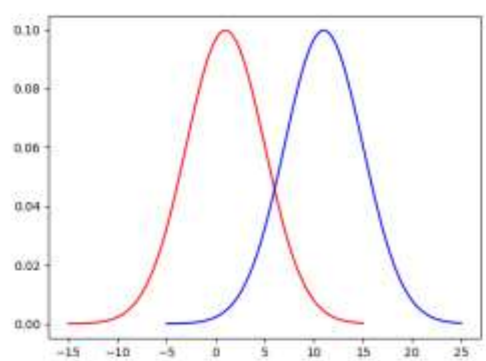
$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2} = 0$$

Or, in general U_2 will be smaller.



General Case: $\boldsymbol{\mu_1 \neq \mu_2}$





$$U = Minimum\,[U_1, U_2]$$

For two-sided alternative, the null hypothesis is rejected if the value of U, as calculated above, is less than or equals to the value of U_α for a given level of significance α. U_α can be obtained from Table 10.4 of RMMR Table.

In case of one-sided alternatives, rejection criteria for null hypothesis will be as follows:

Alternate Hypothesis	Reject Null Hypothesis, if
H_1: $\mu_1 < \mu_2$	$U_1 \leq U_\alpha$
H_1: $\mu_1 > \mu_2$	$U_2 \leq U_\alpha$

For one-sided alternatives, critical values of U_1 or U_2 are obtained from Table 10.4 (First page) of RMMR Table.

Example 7

The following are the weight gains (in kg) of two random samples young people's fed two different diets but otherwise kept under identical conditions:

Diet 1: 9.7, 10.8, 7.0, 8.9, 5.4, 6.3, 8.5, 8.7, 6.9, 9.1, 6.7, 8.6, 9.4, 9.6
Diet 2: 7.4, 4.6, 4.9, 6.1, 6.8, 5.4, 6.5, 4.6, 5.4, 6.7, 10.7, 6.8, 6.6, 8.3, 6.0, 6.4

Use U test to test the null hypothesis the populations sampled is identical against the alternate hypothesis that on the average the first diet produces a greater weight gain. <Use α = 0.01>

Solution

The hypothesis to be tested is

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

Ranking the data jointly in ascending order gives

Diet #	Weight Gain	Rank	Diet #	Weight Gain	Rank
2	4.6	1.5	2	6.8	15.5
2	4.6	1.5	1	6.9	17
2	4.9	3	1	7.0	18
2	5.4	5	2	7.4	19
2	5.4	5	2	8.3	20
1	5.4	5	1	8.5	21
2	6.0	7	1	8.6	22
2	6.1	8	1	8.7	23
1	6.3	9	1	8.9	24
2	6.4	10	1	9.1	25
2	6.5	11	1	9.4	26
2	6.6	12	1	9.6	27
2	6.7	13.5	1	9.7	28
1	6.7	13.5	2	10.7	29
2	6.8	15.5	1	10.8	30

The sum of the ranks for diet 2 is R_2 = 176.5 and that for diet 1 is R_1= 288.5.
<check: R_1 + R_2 = 465 = sum(1, 2, …, 30)>

$U_2 = R_2 - n_2(n_2 + 1)/2$ = 176.5 – 16(16+1)/2 = 176.5 – 136 = 40.5. Critical value of U (one-sided) for α = 0.01 corresponding to n_1 =14 & n_2 =16 is 56. Since U_2 = 40.5 < U_α= 56, we reject the null hypothesis and conclude that first diet produces greater weight gain.

Comparison with corresponding parametric test

Asymptotic Relative Efficiency (ARE) provides a measure for relative performance of two types of tests, when the sample size is large.

ARE of one test relative to another is the limiting ratio of sample sizes necessary to obtain identical power for both the tests. Suppose ARE of a nonparametric test relative to a parametric test is reported to be 0.98. Loosely interpreted, these means that the nonparametric test based on, say 100 observations, is as efficient as the corresponding parametric test based on 98 observations.

Test of Hypothesis	Nonparametric Test	Analogous Parametric Test	ARE
Location Parameters			
One / Paired Sample(s)	**Sign Test**	***t*-test**	**0.637**
One / Paired Sample(s)	**Wilcoxon signed-rank Test**	***t*-test**	**0.955**
Two Independent Samples	**Mann-Whitney U Test**	***t*-test**	**0.955**
Association Analysis			
Two Related Samples	**Spearman / Kendal Test**	**Pearson Product – moment correlation**	**0.912**

RUN TEST

In case of any statistical inference, it is necessary that the inference based on a random sample. Run Test is a nonparametric method of testing the randomness of observed data on the basis of the order in which they were obtained. This technique is based on theory of runs, where each run is a sequence data having the same characteristic (i.e. identical letters or, symbols, numerals, etc.), that is preceded and followed by a data with a different characteristic or no data at all. The number of data in a run is called the length of the run.

Too few runs and too many runs are very rare in truly random sequences and can therefore serve as a statistical criterion for the rejection of randomness.

Example 8

The gender of babies born in a hospital in one month was recorded in order of birth, where *F* represents a female and *M* represents a male. Determine the number of runs and the length of each run.

F F F M M F F M F M M M F F F M M M M

$$\overbrace{FFF\quad MM\quad FF\quad M\quad F\quad MMM\quad FFF\quad MMMM}^{\text{There are 8 runs}}$$

	FFF	MM	FF	M	F	MMM	FFF	$MMMM$
Length of runs:	3	2	2	1	1	3	3	4

Example 9

Students in the English professor's afternoon class also record whether he is late (L) or on time (T) for class each day. The results of one student, for 26 days, are shown below. At α = 0.05, can you conclude that this sequence is not random?

T	T	L	T	L	L	L	L	L	T	T	L	T
T	T	L	L	L	T	L	T	T	T	T	T	L

H_0: **Sequence of arrivals are random**
H_1: **Sequence of arrivals are not random (Claim)**

n_1 = Number of L's = 12 and n_2=number of T's = 14
Number of runs = 12

From the table given in the next page, depending upon the values of n_1 and n_2, we find

lower critical value = 8, and
upper critical value =20.

Since, the number of run (12) is between the critical values 8 and 20, we fail to reject the null hypothesis.

So, at 5% level of significance, there is not enough evidence to support the claim that the sequence of arrivals is not random.

Critical values for number of runs $(\alpha = 0.05)$

This table gives the critical values at $\alpha = 0.05$ for a two-tailed test. Reject the null hypothesis if the number of runs is less than or equal to the smaller value or greater than or equal to the larger value.

Value of n_1	Value of n_2: 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	1 6	1 6	1 6	1 6	1 6	1 6	1 6	1 6	1 6	1 6	2 6	2 6	2 6	2 6	2 6	2 6	2 6	2 6	2 6
3	1 6	1 8	1 8	1 8	2 8	2 8	2 8	2 8	2 8	2 8	2 8	2 8	2 8	3 8	3 8	3 8	3 8	3 8	3 8
4	1 6	1 8	1 9	2 9	2 9	2 10	3 10	3 10	3 10	3 10	3 10	3 10	3 10	3 10	4 10	4 10	4 10	4 10	4 10
5	1 6	1 8	2 9	2 10	3 10	3 11	3 11	3 12	3 12	4 12	4 12	4 12	4 12	4 12	4 12	4 12	5 12	5 12	5 12
6	1 6	2 8	2 9	3 10	3 11	3 12	3 12	4 13	4 13	4 13	4 13	5 14	5 14	5 14	5 14	5 14	5 14	6 14	6 14
7	1 6	2 8	2 10	3 11	3 12	3 13	4 13	4 14	5 14	5 14	5 14	5 15	5 15	6 15	6 16	6 16	6 16	6 16	6 16
8	1 6	2 8	3 10	3 11	3 12	4 13	4 14	5 14	5 15	5 15	6 16	6 16	6 16	6 16	6 17	7 17	7 17	7 17	7 17
9	1 6	2 8	3 10	3 12	4 13	4 14	5 14	5 15	5 16	6 16	6 16	6 17	7 17	7 18	7 18	7 18	8 18	8 18	8 18
10	1 6	2 8	3 10	3 12	4 13	5 14	5 15	5 16	6 16	6 17	7 17	7 18	7 18	7 18	8 19	8 19	8 19	8 20	9 20
11	1 6	2 8	3 10	4 12	4 13	5 14	5 15	6 16	6 17	7 17	7 18	7 19	8 19	8 19	8 20	9 20	9 20	9 21	9 21
12	2 6	2 8	3 10	4 12	4 13	5 14	6 16	6 16	7 17	7 18	7 19	8 19	8 20	8 20	9 21	9 21	9 21	10 22	10 22
13	2 6	2 8	3 10	4 12	5 14	5 15	6 16	6 17	7 18	7 19	8 19	8 20	9 20	9 21	9 21	10 22	10 22	10 23	10 23
14	2 6	2 8	3 10	4 12	5 14	5 15	6 16	7 17	7 18	8 19	8 20	9 20	9 21	9 22	10 22	10 23	10 23	11 23	11 24
15	2 6	3 8	3 10	4 12	5 14	6 15	6 16	7 18	7 18	8 19	8 20	9 21	9 22	10 22	10 23	11 23	11 24	11 24	12 25
16	2 6	3 8	4 10	4 12	5 14	6 16	6 17	7 18	8 19	8 20	9 21	9 21	10 22	10 23	11 23	11 24	11 25	12 25	12 25
17	2 6	3 8	4 10	4 12	5 14	6 16	7 17	7 18	8 19	9 20	9 21	10 22	10 23	11 23	11 24	11 25	12 25	12 26	13 26
18	2 6	3 8	4 10	5 12	5 14	6 16	7 17	8 18	8 19	9 20	9 21	10 22	10 23	11 24	11 25	12 25	12 26	13 26	13 27
19	2 6	3 8	4 10	5 12	6 14	6 16	7 17	8 18	8 20	9 21	10 22	10 23	11 23	11 24	12 25	12 26	13 26	13 27	13 27
20	2 6	3 8	4 10	5 12	6 14	6 16	7 17	8 18	9 20	9 21	10 22	10 23	11 24	12 25	12 25	13 26	13 27	13 27	14 28

Using Run Test to test randomness of numerical values:

- **Let there is a sample of n observations, which are recorded in the order they are collected.**
- **Calculate the median.**
- **Give $(+)$ sign to all the observations in the sample greater than the median value and $(-)$ sign to those smaller than the median value. If there is an odd number of an observation, ignore the median observation. This ensures that number of + signs and number of – signs are equal.**
- **Count the number + and – signs, n_1 and n_2, respectively.**
- **Count the number of runs R in the sample.**
- **Determine the critical values of R from the table given previously.**
- **If the value of R falls in the critical region (i.e. R is either ≤ lower critical value or ≥ upper critical value), then reject the null hypothesis that observations in the sample occurred in a random manner.**

Example 10

Consider the following data set and test for randomness.

81.02 80.08 80.05 79.70 79.13 77.09 80.09 79.40 80.56 80.97 80.17 81.35 79.64 80.82 81.26 80.75 80.74 81.59 80.14 80.75 81.01 79.09 78.73 78.45 79.56 79.80

Solution

The median of the above data set is 80.12.

81.02	80.08	80.05	79.70	79.13	77.09	80.09	79.40	80.56	80.97	80.17	81.35	79.64
+	-	-	-	-	-	-	-	+	+	+	+	-

80.82	81.26	80.75	80.74	81.59	80.14	80.75	81.01	79.09	78.73	78.45	79.56	79.80
+	+	+	+	+	+	+	+	-	-	-	-	-

Clearly, in above data set n_1 = n_2 = 13 and *R* = 6. From above table, we find the critical values at α = 0.05 as 8 (lower) and 20 (upper). Since *R* = 6 is less than 8, we reject the null hypothesis and conclude that the fluctuations are not random.

KOLMOGOROV-SMIRNOV TEST

Kolmogorov–Smirnov test (K–S test) is a nonparametric test for the equality of continuous probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample K–S test).

The Kolmogorov–Smirnov statistic quantifies a distance between the empirical cumulative distribution function (of the sample) and the cumulative distribution function of the assumed distribution, or between the empirical cumulative distribution functions of the two samples (two sample case). In this test the null hypothesis is that the samples are drawn from the same distribution (in the two-sample case) or that the sample is drawn from the assumed distribution (in the one-sample case). In each case, the distributions considered under the null hypothesis are continuous distributions only.

The two-sample K-S test is one of the most useful and general nonparametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples.

The Kolmogorov–Smirnov test can be modified to serve as a goodness of fit test. In the special case of testing for normality of the distribution, samples are standardized and compared with a standard normal distribution.

The empirical cumulative distribution function (ECDF) F_n for n i.i.d. observations X_i is defined as

$$F_n(X_i) = \frac{1}{n}\sum_{j=1}^{n} I\left(X_j \le X_i\right),$$

where the indicator function, $I\left(X_j \le X_i\right)$ is defined as

$$I\left(X_j \le X_i\right) = \begin{cases} 1, & X_j \le X_i \\ 0, & \text{otherwise} \end{cases}.$$

So, ECDF corresponding to a given observation is defined as the proportion of observations in the sample less or equal to the given observation.

The K-S logic is as follows: if the maximum departure between the assumed CDF and ECDF is small, then the assumed CDF will likely be correct. But if this discrepancy is "large" then the assumed CDF is unlikely to be proper.

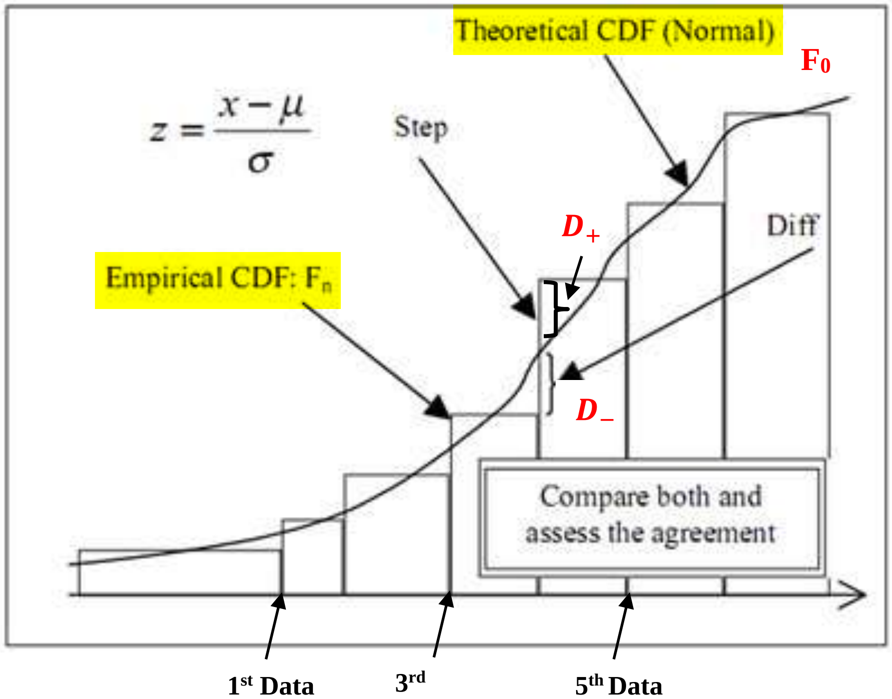
One Sample Kolmogorov-Smirnov Test

The hypothesis to be tested here is

$$H_0 : F_n = F_0 \qquad H_1 : F_n \neq F_0$$

In the distance tests, under null hypothesis, the theoretical (assumed) CDF, denoted by F_0, closely follows the empirical cumulative distribution function (ECDF), denoted by F_n. This is conceptually illustrated in Figure below. The data here are given as an ordered sample $\{X_1 \leq X_2 \leq \cdots \leq X_n\}$. We compare the theoretical and empirical results. If they are close to each other (probabilistically), then the data supports the assumed distribution. If they do not, the distribution assumption is rejected.

Assuming the null hypothesis to be true, estimate the parameter(s) of the assumed distribution. Then, obtain both the theoretical (assumed CDF) distribution (F_0) as well as the empirical (F_n) at each data point. Since K-S is a distance test, we need to find the maximum absolute distance $|F_n - F_0|$ between the empirical and theoretical distributions.



Let $F_0(X_i)$be the assumed cumulative distribution function evaluated at X_i and $F_n(X_i)$be the corresponding empirical cumulative distribution function. Define

$$D_+ = F_n(X_i) - F_0(X_i) \quad \text{and}$$
$$D_- = F_0(X_i) - F_n(X_{i-1})$$

for every data point X_i [clearly, $F_n(< X_1) = 0$]. The Kolmogorov-Smirnov statistic is:

$$D(n) = \max[\max(D_+), \max(D_-)]$$

Clearly, larger the value of $D(n)$ greater will be the chance of the null hypothesis to be rejected. Thus, if the calculated value of $D(n)$ exceeds the critical value, obtained from Table 10.1 of RMMR Table, we reject the null hypothesis.

Example 11

Following data set gives six tensile strength values obtained from samples drawn from same population.

338.7	308.5	317.7	313.1	322.7	294.2

Can it be assumed that the data belong to a normal population?

Solution

We first arrange above data in ascending order.

Calculated values of mean and standard deviation is $\mu = 315.82$**,** $\sigma = 14.85$**.**

As an example, we calculate the values for the smallest data point (294.2)

- $F_0\left(294.2\right) = \Phi\left(\dfrac{294.2 - 315.82}{14.85}\right) = \Phi\left(-1.456\right) = 0.0727$**, which is the theoretical probability at 294.2.**
- $F_n(294.2) = 1/6 = 0.1667;$ **, which is the empirical probability at 294.2**
- Clearly, $F_n(< 294.2) = 0$**.**
- $D_+\left(294.2\right) = F_n\left(294.2\right) - F_0\left(294.2\right) = 0.1667 - 0.0727 = 0.0940$
- $D_-\left(294.2\right) = F_0\left(294.2\right) - F_n\left(< 294.2\right) = 0.0727 - 0 = 0.0727$

Proceeding similarly for all the observations, we get

Row	Data	Z	F_0	F_n	F_{n-1}	$D+=F_n-F_0$	$D-=F_0-F_{n-1}$
1	294.2	-1.456	0.0727	0.1667	0.0000	0.0940	0.0727
2	308.5	-0.493	0.3110	0.3333	0.1667	0.0223	**0.1443**
3	313.1	-0.183	0.4274	0.5000	0.3333	0.0726	0.0941
4	317.7	0.127	0.5505	0.6667	0.5000	0.1162	0.0505
5	322.7	0.463	0.6783	0.8333	0.6667	**0.1550**	0.0116
6	338.7	1.541	0.9383	1.0000	0.8333	0.0617	0.1050
					Max:	**0.1550**	**0.1443**

$D(n)$ **= Maximum** (D_+, D_-) **= 0.1550**

Critical value for $D(n)$**, as obtained from Table 10.1 of RMMR Table, is 0.519.**

Since calculated value of *D* is much less than corresponding critical value, the null hypothesis cannot be rejected. So, the data set follows normal distribution.

Two Sample Kolmogorov-Smirnov Test

This test is used to test the null hypothesis of two identical populations. Here the test statistic is a function of the difference between empirical cumulative distribution functions of the two samples.

Let $X_{11}, X_{12}, \ldots, X_{1n}$ and $X_{21}, X_{22}, \ldots, X_{2m}$ be two independent random samples of size *n and m* from two continuous populations with distribution functions $F_1(x)$ and $F_2(x)$ respectively. The hypothesis is

$$H_0 : F_1(x) = F_2(x) \qquad \text{for all } x$$
$$H_1 : F_1(x) \neq F_2(x) \qquad \text{for at least one } x$$

Procedure

- **Arrange all (n + m) observations in ascending order, keeping their identity.**

- **For every X, calculate the empirical distribution functions for both the samples, that is**

$$S_1 = \frac{\text{number of observations in the first sample } \leq X}{n}$$

$$S_2 = \frac{\text{number of observations in the second sample } \leq X}{m}$$

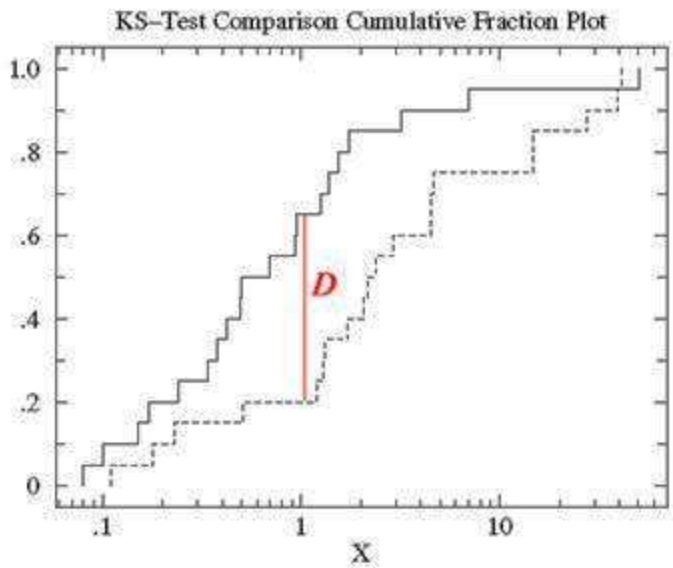
- **If the two samples are drawn from identical populations, there should be reasonable agreement between S_1 and S_2 for all value of X. The absolute values of the differences are a measure of the disagreement. If the maximum absolute difference is small, then all differences are small.**

- **The two-sample Kolmogorov-Smirnov statistic is defined as**

$$D(n, m) = max|S_1 - S_2|$$

- **If above value of D exceeds or equals to $D_{Critical} = 1.36\sqrt{\frac{1}{n} + \frac{1}{m}}$, we reject the null hypothesis at 5% level of significance.**

The graph below shows how a K-S two sample test distribution looks like.



- **For the special case of $n = m = N$, a computed value of $N \times D(N, N)$ that exceeds or is equal to the critical value given in Table 10.2 of RMMR Table, lead to rejection of the null hypothesis of two identical populations.**

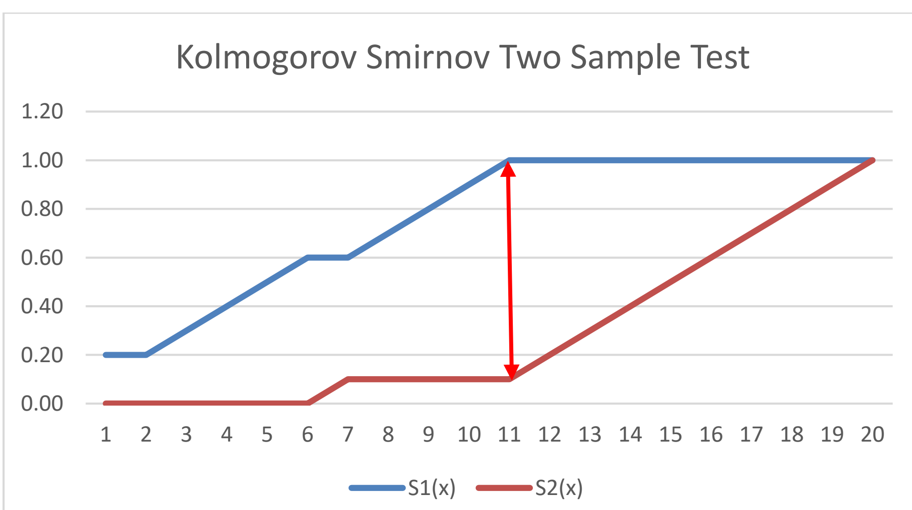
Example 12

Let two samples, each of size 10, were collected from two populations. Can it be assumed that both samples are from same population?

Sample 1	**0.6**	**1.2**	**1.6**	**1.7**	**1.7**	**2.1**	**2.8**	**2.9**	**3.0**	**3.2**
Sample 2	**2.1**	**4.4**	**7.4**	**10.5**	**13.7**	**16.9**	**20.4**	**24.0**	**28.8**	**36.0**

Solution

x	Sample #	$S_1(x)$	$S_2(x)$	Absolute Difference
0.6	1	1/10	0	0.1
1.2	1	2/10	0	0.2
1.6	1	3/10	0	0.3
1.7	1	4/10	0	0.4
1.7	1	5/10	0	0.5
2.1	1	6/10	0	0.6
2.1	2	6/10	1/10	0.5
2.8	1	7/10	1/10	0.6
2.9	1	8/10	1/10	0.7
3.0	1	9/10	1/10	0.8
3.2	1	10/10	1/10	0.9 ← Largest
4.4	2	10/10	2/10	0.8
7.4	2	10/10	3/10	0.7
10.5	2	10/10	4/10	0.6
13.7	2	10/10	5/10	0.5
16.9	2	10/10	6/10	0.4
20.4	2	10/10	7/10	0.3
24.0	2	10/10	8/10	0.2
28.8	2	10/10	9/10	0.1
36.0	2	10/10	10/10	0



So, maximum D = 0.9 and N×D = 9. Now from Table 10.2 of RMMR table we get critical vale as 7 (corresponding to n = 10 and α = 0.05). Since the computed value of 9 exceeds the critical value of 7, we reject the null hypothesis and conclude that two samples come from different samples.

SPEARMAN'S RANK CORRELATION TEST

This test measures the correlation in a random variable that is used in situations where the data consists of pair of observations. Each pair of observations represents two measurements taken on the same object or individual.

Procedure

Suppose that the data consists of n pair of observations from a bivariate population, $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$. Let $R(X_i)$ be the rank of X_i as compared with other X value, for $i = 1, 2, \cdots, n$. So that $R(X_i) = 1$ if X_i is the smallest observed value of X. Similarly, let $R(Y_i)$ = 1, 2, …, or n, depending upon the relative magnitude of Y_i as compared with $Y_1, Y_2, \cdots, Y_n$, for each i.

In case of ties, assign to each tied observation the average of the ranks that would have been assigned if there had been no ties.

The measure of correlation as given by Spearman is usually denoted by ρ, and if there are no ties, is defined as

$$\rho = \frac{\sum_{i=1}^{n}\left[R(X_i) - \frac{n+1}{2}\right]\left[R(Y_i) - \frac{n+1}{2}\right]}{n(n^2-1)/12}.$$

An equivalent, but computationally easier form is given by

$$\boldsymbol{\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}}, \quad \text{where } \boldsymbol{d_i = R(X_i) - R(Y_i)} \qquad (\mathbf{1})$$

If there are many ties, then $\boldsymbol{\rho}$ is defined as

$$\rho = \frac{\sum_{i=1}^{n} R(X_i)R(Y_i) - n\left(\frac{n+1}{2}\right)^2}{\sqrt{\sum_{i=1}^{n}\{R(X_i)\}^2 - n\left(\frac{n+1}{2}\right)^2}\sqrt{\sum_{i=1}^{n}\{R(Y_i)\}^2 - n\left(\frac{n+1}{2}\right)^2}}, \qquad \mathbf{(2)}$$

$$\left[\text{Since, we know that } r = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}\sqrt{\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2}}\right]$$

which is simply Pearson's $\boldsymbol{r}$ computed on the ranks and average ranks. If a moderate number of ties (10-20% of the observations tied in each *X* and *Y*) are present in the data, equation (1) is recommended for computational simplicity.

The Spearman's rank correlation coefficient is used as a test statistic to test

H_0 : No correlation between X and Y

H_1 : There exist correlation between X and Y

and the null hypothesis is rejected if $|\rho|$ exceeds the critical value given in Table 10.7 (if n ≤ 10) or Table 7.1 with d. f. ν = n-2 (if n > 10) of RMMR Table.

Example 13

Twelve sets of identical twins were given psychological tests to measure their aggressiveness. The emphasis is on testing the degree of similarity between twins within the same set. The data were measures of aggressiveness and are given below.

Twin Set, i	1	2	3	4	5	6	7	8	9	10	11	12
First-born, X_i	86	71	77	68	91	72	77	91	70	71	88	87
Second-born, Y_i	88	77	76	64	96	72	65	90	65	80	81	72

Solution

Here the hypotheses to be tested are

H_0 **: The measures of aggressiveness of two identical twins are independent, i.e. un-correlated.**
H_1 **: There exists correlation between the two measures of aggressiveness.**

i	X_i	Y_i	$R(X_i)$	$R(Y_i)$	$[R(X_i) - R(Y_i)]^2$
1	86	88	8	10	4
2	71	77	3.5	7	12.25
3	77	76	6.5	6	0.25
4	68	64	1	1	0
5	91	96	11.5	12	0.25
6	72	72	5	4.5	0.25
7	77	65	6.5	2.5	16
8	91	90	11.5	11	0.25
9	70	65	2	2.5	0.25
10	71	80	3.5	8	20.25
11	88	81	10	9	1
12	87	72	9	4.5	20.25
				$\sum d_i^2$	75.00

Therefore, $\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6\times 75}{12(12^2-1)} = 0.7378.$ As a point of interest, the value of ρ obtained using Equation 2 is 0.7354. Critical value of ρ, as obtained from Table 7.1 of RMMR Table, is 0.708 for d. f. = 12-2 = 10 and α = 0.01.

The computed value of ρ is more than the critical value of 0.708, so we reject the null hypothesis and conclude that the measures are correlated or in other words twins within the same set are similar in terms of the aggressiveness measures.

KENDALL'S RANK CORRELATION TEST

Here also data consists of a bivariate random sample of size n, $(X_i,\ Y_i)$ for $i = 1, 2, \cdots, n$.

Two pair of observations, e.g. (1.3, 2.2) and (1.6, 2.8), are called *concordant* if both members of one observation are larger than their counterparts in the other observation. Again a pair of observations, such as (1.3, 2.2) and (1.6, 1.4), are called *discordant* if the two members in one observation differ in opposite direction from the respective members in the other observation. Pairs with ties between respective members are neither concordant nor discordant.

Let N_C and N_D represent the number of concordant and discordant pairs of observations respectively. Because the *n* observations may be paired in $\binom{n}{2} = \frac{n(n-1)}{2}$ different ways, number of concordant pairs N_C plus number of discordant pairs N_D plus number of pairs with ties should add up to $\frac{n(n-1)}{2}$.

The measure of correlation proposed by Kendall is

$$\boxed{\tau = \frac{N_C - N_D}{\frac{n(n-1)}{2}}.}$$

If all pairs are concordant, i.e. if there is a direct association between the variables, then τ equals 1.0, whereas if all pairs are discordant, i.e. if there exists an inverse association, then τ equals -1.0. If $N_C > N_D$, then we have a direct association between the *X* and *Y* rankings and τ is positive. On the other hand, if $N_C < N_D$, then we have an inverse association between the *X* and *Y* rankings and τ is negative. The computation of τ is simplified if the observations (X_i, Y_i) are arranged in a column according to increasing values of *X* [we say that the *X's* are in *natural order*]. Then by comparing each *Y*-value only with those below it, the corresponding number of concordant or discordant pairs can be determined easily.

NONPARAMETRIC INFERENCE

X	Y	X (Natural Order)	Y	NC	ND
60	118	10	124	2	7
55	117	20	126	0	8
25	120	25	120	4	3
50	121	30	123	0	5
40	119	30	125	0	5
40	122	40	119	1	2
30	123	40	122	0	3
10	124	50	121	0	2
30	125	55	117	1	0
20	126	60	118	0	0

Kendall's τ statistic can also be used to test the following hypotheses

H_0 **: *No association between X and Y [i.e. $\tau = 0$]***
H_1 **: *Association exists between X and Y [i.e. $\tau \neq 0$]***

Here only the numerator part of τ is used as statistic and is defined by

$$T = N_C - N_D.$$

If the *absolute value of T* is found to be more than critical value given in following table, the null hypothesis is rejected.

Critical Values for Kendall's Test Statistic, T

N	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5% Level	6	8	11	13	16	18	21	25	28	32	35	39	44	48	51	55	60
1% Level	6	10	13	17	20	24	27	31	36	42	45	51	56	62	67	73	78

Example 14

For illustration, we will use the data given in Example 10. Arranging the data values $(X_i, \ Y_i)$ in terms of the increasing values of X gives the following:

	(X_i, Y_i)	*Concordant Pairs below* $(X_i, \ Y_i)$	*Discordant Pairs below* $(X_i, \ Y_i)$
	(68, 64)	11	0
	(70, 65)	9	0
Ties	(71, 77)	4	4
	(71, 80)	4	4
	(72, 72)	5	1
Ties	(77, 65)	5	0
	(77, 76)	4	1
	(86, 88)	2	2
	(87, 72)	3	0
	(88, 81)	2	0
Ties	(91, 90)	0	0
	(91, 96)	0	0
TOTAL		$N_C = 49$	$N_D = 12$

Kendall's τ is given by

$$\tau = \frac{N_C - N_D}{\frac{n(n-1)}{2}} = \frac{49-12}{\frac{12\times 11}{2}} = 0.5606 \, .$$

Thus there is positive rank correlation between the aggression scores. Now to test the hypotheses

$$H_0 : \textit{No association between the aggression measures}$$
$$H_1 : \textit{Association exists between the aggression measures,}$$

we compute

$$T = N_C - N_D = 49 - 12 = 37.$$

Since, absolute value of T exceeds the critical value of 28 [at n = 12 & α = 0.05], we reject the null hypothesis and conclude that there exist a positive association between the aggression measures in the twins.

CONFIDENCE INTERVAL FOR A QUANTILE

Let the data consists of observations on $X_1, X_2, \cdots, X_n$, which are independent and identically distributed random variables. Let, $X_{(1)} \le X_{(2)} \le \cdots \le X_{(r)} \le \cdots \le X_{(s)} \le \cdots \le X_{(n)}$ represent the ordered sample, where $1 \le r < s \le n$. We wish to find a confidence interval for the (unknown) p^* th quantile, where p^* is some number between 0 and 1. The measurement scale of the $X_i's$ should be at least ordinal. Note that, *p-th quantile is that value of the ordered set of data such that p proportion of data will be below that*.

Method

To find the approximate 100(1-α) % confidence interval, we use the Binomial distribution table to find the values of r and s as follows:

1. **Determine the largest value of K' such that $P(K \le K' | n, p^*) \le \alpha/2$, and take $r = K' + 1$. Let us denote the exact cumulative probability corresponding to above value of r by α_1.**
2. **Determine the largest value of K' such that $P(K \le K' | n, p^*) \le 1 - \frac{\alpha}{2}$, and take $s = K' + 1$. Let us denote the exact cumulative probability corresponding to s by α_2.**
3. **So, the confidence interval will be $[X_{(r)}, X_{(s)}]$.**
4. **The exact confidence coefficient will be $100(\alpha_2 - \alpha_1)\%$.**

For the special case of $p^* = 0.5$, we get the confidence interval for the median.

Example 15

Fifteen radio tubes are selected at random from a large batch of radio tubes and are tested. The number of hours until failure is recorded for each tube. We wish to find a confidence interval for the median with confidence coefficient close to 95%.

The results of the testing, arranged in increasing order, are as follows.

$X_{(1)} = 46.9$ $\quad X_{(2)} = 47.2$ $\quad X_{(3)} = 49.1$ $\quad X_{(4)} = 56.5$

$X_{(5)} = 56.8$ $\quad X_{(6)} = 59.2$ $\quad X_{(7)} = 59.9$ $\quad X_{(8)} = 63.2$

$X_{(9)} = 63.3$ $\quad X_{(10)} = 63.4$ $\quad X_{(11)} = 63.7$ $\quad X_{(12)} = 64.1$

$X_{(13)} = 67.1$ $\quad X_{(14)} = 67.7$ $\quad X_{(15)} = 73.3$

The median of the above observations is 63.2

Binomial Distribution table for n = 15 and p = 0.5

K'	Probability	Cum. Prob.
0	0.00003	0.00003
1	0.00046	0.00049
2	0.00320	0.00369
3	0.01389	0.01758
4	0.04165	0.05923
5	0.09165	0.15088
6	0.15274	0.30362
7	0.19638	0.50000
8	0.19638	0.69638
9	0.15274	0.84912
10	0.09165	0.94077
11	0.04165	0.98242
12	0.01389	0.99631
13	0.00320	0.99951
14	0.00046	0.99997
15	0.00003	1.00000

To find the lower and upper limits of the confidence interval, we find the values of *r* and *s* using the above reproduction of the Binomial Table. Since

$P(k \leq 3 \mid 15, 0.5) = 0.01758$.
So we take, $r = 3+1 = 4$.

Though this probability is much less than 0.025, but the next cumulative probability (i.e. 0.05923) is much more than 0.025.

Similarly since, $P(k \leq 10 \mid 15, 0.5) = 0.94077$, so we take $s = 10 + 1 = 11$.

So, lower limit = $X_{(4)} = 56.5$ and upper limit = $X_{(11)} = 63.7$.

Now, the exact cumulative probabilities for the lower and upper bounds are 0.05923 and 0.98242 respectively. So, the confidence coefficient is 100[0.98242 – 0.05923] = 92.319.

Thus, we can say "interval from 56.5 hours to 63.7 hours, inclusive, is a 92.3% confidence interval for the median".

KRUSKAL-WALLIS TEST

The Kruskal-Wallis test is a nonparametric alternative to F-test in case of single factor ANOVA, and it requires only that ε_{ij} have the same continuous distribution for all factor levels $(i = 1, 2, \cdots, n)$.

In this case the data consists of $N = \sum_{i=1}^{a} n_i$ observations, where n_i is the number of observations in *i*-th level of the factor.

Procedure:

Combine all N observations and arrange them in ascending order. Let, R_{ij} be the rank of j-th observation in i-th factor level (Note: these ranks will be numbers between 1 to N). Let

$$R_{i\bullet} = \sum_{j=1}^{n_i} R_{ij} \; and \; \bar{R}_{i\bullet} = \frac{R_{i\bullet}}{n_i}.$$

Clearly,

$$\bar{R}_{\circ\circ} = \frac{1}{N}\sum_{i=1}^{a}\sum_{j=1}^{n_i} R_{ij} = \frac{1}{N} \times \frac{N(N+1)}{2} = \frac{N+1}{2}$$

If the null hypothesis $H_0: \mu_1 = \mu_2 = \cdots = \mu_a$ is true, the N observations come from the same distribution and all possible assignments of the N ranks to the a samples are equally likely, we would expect the ranks $1, 2, \cdots, N$ to be mixed evenly among the a samples. On the contrary, if the null hypothesis is false, some samples will contain observations having predominantly small ranks, while other samples will consist of observations having predominantly large ranks.

When the null hypothesis is true i.e. all observations are from the same distribution, expected value of level-wise average rank, i.e. $E(\bar{R}_{i\cdot})$ will be equal to the overall average $\bar{R}_{\bullet\bullet}$, i.e.

$$E(\bar{R}_{i\cdot}) = \frac{N+1}{2}.$$

The Kruskal-Wallis test statistic measures the degree to which the actual observed average ranks $\bar{R}_{i\cdot}$ differ from their expected value of $(N+1)/2$. If the difference is large then H_0 is rejected.

The test statistic used is

$$H = \frac{1}{S^2}\sum_{i=1}^{a} n_i\left[\bar{R}_{i\bullet} - E\left(\bar{R}_{i\bullet}\right)\right]^2 = \frac{1}{S^2}\left[\sum_{i=1}^{a}\frac{R_{i\bullet}^2}{n_i} - \frac{N(N+1)^2}{4}\right] \qquad (1)$$

where

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left[R_{ij} - \bar{R}_{..}\right]^2 = \frac{1}{N-1}\left[\sum_{i=1}^{a}\sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4}\right]$$

Please note that S^2 is just the variation of the ranks. If there are no ties, then S^2 simplifies to $N(N+1)/2$ and the test statistic simplifies to

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{a}\frac{R_{i\bullet}^2}{n_i} - 3(N+1) \qquad (2)$$

If the number of ties is moderate, say 10 to 20%, there will be very little difference between equations (1) and (2), so the simpler form (2) may be preferred.

If n_i are reasonably large, say $n_i \geq 5$, then H is distributed approximately as χ^2_{a-1} under H_0. Since large value of H will lead to the rejection of H_0, we will reject H_0, if the observed value h exceeds corresponding critical value, i.e.

$$h \geq \chi^2_{\alpha,a-1}$$

where $\chi^2_{\alpha,a-1}$ is upper α percentile point of chi-square distribution with $a-1$ degrees of freedom.

Example 16

An experiment was conducted to assess whether the cotton content in a synthetic fiber has any effect on fiber tensile strength. The sample data and ranks for this experiment are given below.

Percentage of Cotton									
15	Rank	20	Rank	25	Rank	30	Rank	35	Rank
7	2.0	12	9.5	14	11.0	19	20.5	7	2.0
7	2.0	17	14.0	18	16.5	25	25.0	10	5.0
15	12.5	12	9.5	18	16.5	22	23.0	11	7.0
11	7.0	18	16.5	19	20.5	19	20.5	15	12.5
9	4.0	18	16.5	19	20.5	23	24.0	11	7.0
TOTAL	27.5		66.0		85.0		113.0		33.5

Since there are quite a large number of ties, we will use equation (1).

$$s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{a}\sum_{j=1}^{n_i} r_{ij}^2 - \frac{N(N+1)^2}{4}\right]$$

$$= \frac{1}{24}\left[5510 - \frac{25(26)^2}{4}\right] = 53.54$$

So, the test statistics becomes

$$h = \frac{1}{s^2}\left[\sum_{i=1}^{a}\frac{r_{i\bullet}^2}{n_i} - \frac{N(N+1)^2}{4}\right]$$

$$= \frac{1}{53.54}\left[5245.7 - \frac{25(26)^2}{4}\right] = 19.06 .$$

Since $h > \chi^2_{0.01,4} = 13.28$, we would reject the null hypothesis and conclude that cotton content affects fiber tensile strength.

FRIEDMAN TEST

The Friedman test is a non-parametric statistical test developed by Milton Friedman. Similar to the parametric repeated measures ANOVA, it is used to detect differences in treatments across multiple test attempts to given subjects, treated as blocks. The procedure involves ranking observations under each block (row/subjects), then considering the values of ranks by columns (i.e. treatments).

Analysis Procedure:

1. **Given data $\{y_{ij}\}_{r\times c}$, that is, a matrix with r rows (the *blocks*), c columns (the *treatments*) and a single observation at the intersection of each block and treatment, determine the ranks within each block. If there are tied values, assign to each tied value the average of the ranks that would have been assigned without ties. Replace the data with a new matrix $\{R_{ij}\}_{r\times c}$ where the entry R_{ij} is the rank of y_{ij} within the block *i*.**

2. **Find the values (treatment-wise average rank):**

$$* \ \bar{R}_{\cdot j} = \frac{1}{r}\sum_{i=1}^{r} R_{ij}$$

It may be noted that under the null hypothesis of no treatment effect, expected value of treatment-wise average rank will be equal to overall average rank, i.e.

$$E\left(\bar{R}_{\bullet j}\right) = \bar{R}_{\bullet\bullet}$$

* Since, $R_{\circ\circ} = \sum_{i=1}^{r}\sum_{j=1}^{c} R_{ij} = \sum_{i=1}^{r}\frac{c(c+1)}{2} = \frac{rc(c+1)}{2}$, $\quad \bar{R}_{\circ\circ} = \frac{c+1}{2}$

* $$SS_T = r\sum_{j=1}^{c}\left(\bar{R}_{\bullet j} - \bar{R}_{\bullet\bullet}\right)^2 = \frac{1}{r}\sum_{j=1}^{c} R_{\bullet j}^2 - \frac{rc(c+1)^2}{4}$$

* $$S^2 = \frac{1}{r(c-1)}\left[\sum_{i=1}^{r}\sum_{j=1}^{c}\left(R_{ij} - \bar{R}_{\bullet\bullet}\right)^2\right] = \frac{1}{r(c-1)}\left[\sum_{i=1}^{r}\sum_{j=1}^{c} R_{ij}^2 - \frac{rc(c+1)^2}{4}\right]$$
$$= \frac{c(c+1)}{12}$$

[S^2 is simplified to the final expression under the assumption that there are no ties within a block]

3. The test statistic is given by

$$F_r = \frac{SS_T}{S^2} = \begin{cases} \dfrac{r\sum_{j=1}^{c}\left(\bar{R}_{.j} - \bar{R}_{..}\right)^2}{S^2}, \\ \dfrac{12}{rc(c+1)}\sum_{j=1}^{c} R_{.j}^2 - 3r(c+1) \end{cases} \qquad (1)$$

- ✓ **Second expression of (1) can be used if there are no ties within a block.**
- ✓ **Whereas in the event of ties within a block as in the case of first expression of (1) , we need to use $s^2 = \frac{1}{r(c-1)}\left[\sum_{i=1}^{r}\sum_{j=1}^{c} R_{ij}^2 - \frac{rc(c+1)^2}{4}\right]$.**

4. As the number of blocks gets large (i.e., greater than 4, though not a very strict requirement), one can approximate the test

statistic F_r by using the chi-square distribution with c - 1 degrees of freedom. Thus, for any selected level of significance α, the null hypothesis will be rejected if the computed value of F_r is greater than the corresponding upper-tail critical value for the chi-square distribution with c–1 degrees of freedom. Otherwise, null hypothesis of equal medians cannot be rejected.

Note: Remember, blocks could be row-wise or column-wise. For the sake of explanation, we have assumed blocks to be row-wise above. In general, idea is to find the

- **block-wise ranks,**
- **treatment-wise ranks total,**
  [i.e. if blocks are row-wise (or, column-wise), treatment-wise ranks total will be column-wise (or, row-wise) total]
- **grand total of ranks,**
- **number of subjects (r), and**
- **number of treatments (c).**

Example 17

Given below is the acne severity scores compared at baseline and after each 4 weeks following a course of Erythromycin 500 mg taken twice each day.

Acne Severity Scores (1-12 scale) following Erythromycin.

Subjects	Week 0	Week 4	Week 8	Week 12
1	12	8	6	5
2	11	5	5	6
3	6	4	4	5
4	4.5	1	2	1
5	5.5	5	5	5
6	7	6	6	5.5
7	3.5	3	3	3
8	8	8	7	5
9	9	8.5	7	4
10	10	8	6	4
11	12	8	6	5
12	12	7	5	4

Is there any variation in acne severity scores over the 12-week period?

Solution>

Subject-wise ranks data for Acne Severity Score

Subjects	Week 0	Week 4	Week 8	Week 12
1	4	3	2	1
2	4	1.5	1.5	3
3	4	1.5	1.5	3
4	4	1.5	3	1.5
5	4	2	2	2
6	4	2.5	2.5	1
7	4	2	2	2
8	3.5	3.5	2	1
9	4	3	2	1
10	4	3	2	1
11	4	3	2	1
12	4	3	2	1
Total	**47.5**	**29.5**	**24.5**	**18.5**

In this example, clearly number of subjects (*r*) = 12 and number of treatments (*c*) = 4.

$$SS_T = \frac{1}{12}\left(47.5^2 + 29.5^2 + 24.5^2 + 18.5^2\right) - \frac{12\times 4\left(4+1\right)^2}{4}$$
$$= 339.083 - 300 = 39.083$$

Since there are ties within blocks, $\sum_{i=1}^{12}\sum_{j=1}^{4} R_{ij}^2 = 353.5$

$$S^2 = \frac{1}{12\left(4-1\right)}\left[353.5 - \frac{12\times 4\left(4+1\right)^2}{4}\right]$$
$$= \frac{1}{36}\left[353.5 - 300\right] = 1.486$$

$$F_r = \frac{39.083}{1.486} = 26.301$$

Now, $\chi^2(0.05, 3) = 7.81$, so we reject the null hypothesis and arrive at the conclusion that acne severity score is different over the 12-week period.

ANDERSON DARLING TEST FOR NORMALITY

One method of testing the hypothesis that n observations have been drawn from a population with a specified distribution function is to compare the specified cumulative distribution function, $F(x)$, with its sample analogue, the empirical cumulative distribution function, $F_n(x)$. The Anderson-Darling test (Anderson and Darling, 1952) belong to the class of quadratic CDF based statistic, which measures the square of the difference between $F(x)$ and $F_n(x)$, where

$$F_n(x) = \begin{cases} 0, & x < x_1 \\ \frac{i}{n}, & x_i \le x < x_{i+1}, \forall i = 1, 2, \cdots, n-1 \\ 1, & x \ge x_n \end{cases}$$

Obviously, x_i's has to be arranged in ascending order.

Hypothesis to be tested is $H_0 : F_n(x) = F(x)$
$H_1 : F_n(x) \neq F(x)$

Or in other wards

H_0: **the data follow a specified distribution.**
H_1: **the data do not follow the specified distribution.**

Criterion suggested by AD is

$$A^2 = n\int_{-\infty}^{\infty}\left[F_n(x)-F(x)\right]^2 w\big(F(x)\big)dF(x)$$

where $w\big(F(x)\big)$ is a suitable non-negative weight function.

For a given x and hypothetical distribution $F(x)$, the random variable $nF_n(x)$ – that can be thought of as number of success - has a binomial distribution with probability of success $F(x)$ under null hypothesis. The expected value and variance of $nF_n(x)$ is $nF(x)$ and $nF(x)[1-F(x)]$ respectively. Anderson-Darling Test statistic places more weight on the observations in the tails of the distribution. In that case the choice of $w\big(F(x)\big)$ is

$$w\big(F(x)\big) = \frac{1}{F(x)[1-F(x)]}$$

so that the weight is large when $F(x)$ is close to 0 and 1, i.e. near the tails, and small when $F(x) = 1/2$, i.e. near the medians. The Anderson-Darling distance places more weight on observations in the tails of the distribution, thereby making it more sensitive to deviations in the tails of the selected distribution.

Now, for a specified x, $\frac{n[F_n(x)-F(x)]}{\sqrt{nF(x)[1-F(x)]}} = \sqrt{n}\frac{F_n(x)-F(x)}{\sqrt{F(x)[1-F(x)]}}$

has zero mean and unity variance.

The Anderson Darling statistic is

$$\frac{1}{n}A^2 = \int_{-\infty}^{\infty} \frac{\left[F_n(x) - F(x)\right]^2}{F(x)\left[1 - F(x)\right]} dF(x)$$

or,

$$\frac{1}{n}A^2 = \int_{-\infty}^{x_1} \frac{F^2(x)}{F(x)[1-F(x)]} dF(x) + \int_{x_1}^{x_2} \frac{[F_n(x) - F(x)]^2}{F(x)[1-F(x)]} dF(x) + \cdots + \int_{x_n}^{\infty} \frac{[1-F(x)]^2}{F(x)[1-F(x)]} dF(x)$$

Above expression upon integration and collection of like terms gives

$$A^2 = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[ln\left(u_j\right) + ln\left(1 - u_{n+1-j}\right)\right],$$

where $u_j = F\left(x_j\right)$ and x_j are the ordered data.

The null hypothesis is rejected if A^2 is large.

Test Procedure – to test for Normality of observation

1) Arrange the given values in ascending order: $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$

2) Calculate the standardized values $z_{(j)} = \frac{x_{(j)} - \bar{x}}{s}$, $j = 1, 2, \cdots, n$.

3) Determine $u_j = \Phi\left[z_{(j)}\right]$ using the table for standard normal distribution.

4) Calculate the value of A^2.

5) Compute the modified statistic (for small samples):

$$A_*^2 = A^2\left[1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right].$$

6) Reject the null hypothesis of normality, if A_*^2 exceeds 0.752 or 1.035 at 0.05 and 0.01 levels of significance respectively.

Note: Above correction is suggested by D'Agostino and Stephens (1986) (Ref. *Goodness-of-fit Techniques*, Marcel Dekker, NY**)**

Example 18

Test the data given below for normality using Anderson Darling procedure.

338.7 308.5 317.7 313.1 322.7 294.2

Mean = 315.82 and Std. deviation = 14.9.

$$A^2 = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[\ln u_j + \ln\left(1-u_{n+1-j}\right)\right], \text{where } u_j = \Phi\left(z_j\right)$$

$$A_*^2 = A^2\left[1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right]$$

j	x_j	Z_j	u_j	$ln\ u_j$	Z_{n+1-j}	u_{n+1-j}	$ln(1-u_{n+1-j})$	Col 5 + Col 8	$(2j-1)$Col 9
Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7	Col 8	Col 9	Col 10
1	294.2	-1.45101	0.07339	-2.61198	1.53557	0.93768	-2.77544	-5.38742	-5.38742
2	308.5	-0.49128	0.31162	-1.16598	0.46174	0.67787	-1.13279	-2.29878	-6.89633
3	313.1	-0.18255	0.42758	-0.84962	0.12617	0.55020	-0.79896	-1.64858	-8.24292
4	317.7	0.12617	0.55020	-0.59747	-0.18255	0.42758	-0.55787	-1.15534	-8.08739
5	322.7	0.46174	0.67787	-0.38880	-0.49128	0.31162	-0.37341	-0.76221	-6.85990
6	338.7	1.53557	0.93768	-0.06435	-1.45101	0.07339	-0.07622	-0.14057	-1.54627
								Total	-37.02024

So, A^2 = 0.17004 and A_*^2= 0.20192.

Since, A_*^2< 0.752, we cannot reject the null hypothesis and conclude that the data can be assumed to follow a normal distribution.

Exercises

1) In semiconductor manufacturing, wet chemical etching is often used to remove silicon from the backs of wafers prior to metallization. The etch rate is an important characteristic in this process and known to follow a continuous distribution. Two different etching solutions have been compared using two random samples of 10 wafers for each solution. The observed etch rates are as follows (in ml per minute):

Sol 1	9.9	9.4	9.3	9.6	10.2	10.6	10.3	10.0	10.3	10.1
Sol 2	10.2	10.6	10.7	10.4	10.5	10.0	10.2	10.7	10.4	10.3

Use Wilcoxon signed-rank test to investigate the claim that the mean etch rate is same for both solutions. Use α = 0.05.

2) Test scores obtained by two group of students in Mathematics examined by same Professor are as given below:

Group 1	74	78	68	72	76	69	71	74
Group 2	75	80	87	81	72	73	80	76

Using the Mann-Whitney Test and a significance level of $\alpha = 0.05$, determine if the locations of the two distributions are equal (i.e., if the medians are equal).

3) An experiment was performed to investigate the effect of three different conditioning methods on the breaking strength of cement briquettes. The data are shown in the following table. Use Kruskal-Wallis procedure to test whether the conditioning method affects breaking strength? Assume α=0.05.

Conditioning Method	Breaking Strength				
1	553	550	568	541	537
2	553	599	579	545	540
3	492	530	528	510	571

4) The table below shows the hours of relief provided by two analgesic drugs in 12 patients suffering from arthritis. Is there any evidence that one drug provides longer relief than the other?

Patient #	Drug 1	Drug 2	Patient #	Drug 1	Drug 2
1	2.0	3.5	7	14.9	16.7
2	2.6	2.9	8	6.6	6.0
3	3.6	5.7	9	2.3	3.8
4	2.6	2.4	10	2.0	4.0
5	7.3	9.9	11	6.8	9.1
6	3.4	3.3	12	8.5	20.9