Case Study – Non-Parametric Inference

Aim : To apply Non-Parametric tests to different type of datasets and draw appropriate conclusions out of them

Submitted by: **RAJ KUMAR**

Due Date: **10th November, 2015**

Date of Submission: **8th November, 2015**

Supervisor's Remarks

Late Submission:

Plagiarism:

Completeness:

Quality of Content:

Results and Interpretations:

Additional Remarks:

About Non-Parametric tests:

Non – Parametric inference involves estimation and testing procedures when shape of the population distribution is unknown. The theory of Non – Parametric is mainly based on Order Statistics and the Probability Integral Transform (PIT). Commonly the tests are based on counts, ranks and runs. Most of the time nonparametric testing procedures are developed for following two purposes:

1. To test a hypothesis related to some location parameter.

2. To test a hypothesis related to equality of two or more populations

As non-parametric methods make fewer assumptions, their applicability is much wider than the corresponding parametric methods. In particular, they may be applied in situations where less is known about the application in question. Also, due to the reliance on fewer assumptions, nonparametric methods are more robust.

Frank Wilcoxon Sign Test

Assumptions:

- Data are paired and come from the same population.
- Each pair is chosen randomly and independently.
- The data are measured at least on an ordinal scale, but need not be normal.

One Sample Sign Test (Binomial Test):

In general sign test is used to test for some hypothetical value of a population quantile, i.e. to test if a particular population quantile is equal to some hypothetical value. Population Median is also a quantile (Q0.5). The most application of sign test is to test "if the population median is equal to a hypothetical value".

The procedure involves calculating median and then calculating the difference between median and the sample observations. Thereafter we count either positive or negative signs in the difference and then make use of Binomial Distribution to carry out the test.

Obs	Height	Weight
1	58	115
2	59	117
3	60	120
4	61	123
5	62	126
6	63	129
7	64	132
8	65	135
9	66	139
10	67	142
11	68	146
12	69	150
13	70	154
14	71	159
15	72	164

Case 1: The following table represents observations on heights and weights of 15 females

Use sign test to test the following two hypotheses:

1. The Height of the females can be taken to be equal to 64 inches.
2. The Weight of the females can be taken to be equal to 135 lbs.

1)

H_0: Height of the females can be taken to be equal to 64 inches.

H_1: Height of the females cannot be taken to be equal to 64 inches.

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
Height	Group 1	<= 64	7	.47	.50	1.000
	Group 2	> 64	8	.53		
	Total		15	1.00		

Conclusion:

We conclude that since p-value > 0.05, we may accept H_0 at 5% l.o.s . Thus, heights of the females can be taken to be equal to 64 inches.

2) **H_0: Weight of the females can be taken to be equal to 135 lbs.**

H_1: Weight of the females cannot be taken to be equal to 135 lbs.

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
Weight	Group 1	<= 135	8	.53	.50	1.000
	Group 2	> 135	7	.47		
	Total		15	1.00		

Conclusion:

We conclude that since p-value > 0.05, we may accept H_0 at 5% l.o.s. Thus, the weight of the females can be taken to be equal to 135 lbs.

Case 2: - Win/Loss records of a certain basketball team during their 50 consecutive games are given in the following table:

Game	Outcome	Game	Outcome	Game	Outcome	Game	Outcome	Game	Outcome
1	1	11	1	21	0	31	0	41	1
2	1	12	1	22	1	32	1	42	0
3	1	13	1	23	1	33	1	43	0
4	1	14	0	24	1	34	1	44	0
5	1	15	1	25	1	35	1	45	1
6	1	16	0	26	0	36	1	46	1
7	0	17	1	27	1	37	1	47	0
8	1	18	1	28	1	38	0	48	1
9	1	19	1	29	1	39	0	49	1
10	1	20	0	30	0	40	1	50	1

Using Sign Test to test the hypothesis that win and loss are equally likely.

H_0: Win and loss ratio are equally likely.

H_1: Win and loss ratio are not equally likely.

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
Outcome	Group 1	1	36	.72	.50	.003
	Group 2	0	14	.28		
	Total		50	1.00		

Conclusion:

We conclude that since p-value < 0.05, we reject $\mathbf{H_0}$ at 5% l.o.s. Thus, we can say win and loss are not equally likely.

Two Sample Sign Test (Sign Test)

The purpose of a two sample sign test which is generally referred as the sign test is to test whether two "related" samples are coming from the same population or not? In two sample sign test we need paired or related observations on two sample. This test is sometimes called as the non-parametric counterpart of the paired t – test.

Please refer to the section Sign Test in Fundamentals of Mathematical Statistics to get the complete details. The intuition is as follows:

"If two paired samples are coming from the same population then the probability that sample observations of the first sample exceed or fall below the sample observations of the sample observations in the second sample. So if we calculate the pairwise difference between the sample observations from different samples and count the positive signs then we would expect approximately half of the signs would be positive if the samples come from the same population.

Case 3: Following data represents the marks given to the same set 22 students by two different professors in the same examination:

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Professor A	79	87	24	41	59	12	91	78	63	30	9	64	50	92	64	39	49	86	23	45	12	88
Professor B	83	91	18	39	67	34	78	89	38	45	10	45	56	89	67	35	40	82	32	38	23	92

Using Sign Test, test if the grading of both the professors can be taken to be same.

H_0: - Grading of both the professors can be taken to be same

H_1: - Grading of both the professors cannot be taken to be same

Frequencies

		N
ProfessorB - ProfessorA	Negative Differences[a]	10
	Positive Differences[b]	12
	Ties[c]	0
	Total	22

a. ProfessorB < ProfessorA

b. ProfessorB > ProfessorA

c. ProfessorB = ProfessorA

Test Statistics[a]

	ProfessorB - ProfessorA
Exact Sig. (2-tailed)	.832[b]

a. Sign Test

b. Binomial distribution used.

Conclusion:

We conclude that since p-value > 0.05, we may accept $\mathbf{H_0}$ at 5% l.o.s. Thus, the grading of both the professors can be taken to be same.

Wald-Wolfowitz Run test

Wald-Wolfowitz run test is used to examine whether two random samples come from populations having same distribution. This test can detect differences in averages or spread or any other important aspect between the two populations. This test is efficient when each sample size is moderately large (greater than or equal to 10).

A "run" of a sequence is a maximal non-empty segment of the sequence consisting of adjacent equal elements. In other words, a run is defined as a sequence of letters of one kind surrounded by a sequence of letters of other kind, and the number of element in a run is known as the length of the run.

The run test is used to test the null hypothesis that the two samples come from populations having same distribution, i.e.

H_0: Two samples come from populations having same distribution

H_1: Two samples come from populations having different distribution

Test Statistic: Let U denote the number of runs. To obtain U, list the $n_1 + n_2$ observations from two samples in order of magnitude. Denote observations from one sample by x's and other by y's. Count the number of runs.

Under H_0

$$P(U = u) = \begin{cases} \dfrac{2\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}}{\binom{n_1+n_2}{n_1}}, & if\ U = 2k \\ \dfrac{\binom{n_1-1}{k-1}\binom{n_2-1}{k}+\binom{n_2-1}{k-1}\binom{n_1-1}{k}}{\binom{n_1+n_2}{n_1}}, & if\ U = 2k+1 \end{cases}$$

Decision Rule: Reject $H0$ if $U \le U\propto$ where $U\propto$ is obtained by solving the following.

$P(U \le U\propto | \ H0) \le \propto$

<u>Intuition:</u> If two samples are coming from the same population then there would be a thorough mingling of X's and Y's and consequently number of runs in the combined ordered sample would be large. On the other hand if the samples are coming from two different populations then the number of runs in the combined ordered samples would be small.

Critical Value: Difference in location results in few runs and difference in spread also result in few number of runs. Consequently, critical region for this test is always one-sided. The critical value to decide whether or not the number of runs is few is obtained from the table. The table gives critical value U_α for n_1 (size of sample 1) and n_2 (size of sample 2) at 5% level of significance.

Large Sample Sizes: For sample sizes larger than 20 critical value U_α is given below:

$U_\alpha = \mu - 1.96\ \sigma$ at 5% level of significance

where $\mu = 1 + \dfrac{2n_1n_2}{n_1+n_2}$ and $\sigma = \sqrt{\dfrac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$

Case 4: Following is the data for prices in rupees of a certain commodity in a sample of 15 randomly selected shops from City A and those of 13 randomly selected shops from City B.

City A (prices in rupees)	7.41	7.77	7.44	7.4	7.38	7.93	7.58	8.28	7.23	7.52	7.82	7.71	7.84	7.63	7.68
City B (prices in rupees)	7.08	7.49	7.42	7.04	6.92	7.22	7.68	7.24	7.74	7.81	7.28	7.43	7.47		

Use Run Test to determine if the prices in City A and City B can be taken to be following same probability distribution.

H_0: Prices in City A and City B come from the same probability distribution

H_1: Prices in City A and City B do not come from the same probability distribution

Frequencies

	City	N
Prices	0	15
	1	13
	Total	28

Test Statistics[a,b]

		Number of Runs	Z	Exact Sig. (1-tailed)
Prices	Minimum Possible	14[c]	-.166	.436
	Maximum Possible	14[c]	-.166	.436

a. Wald-Walfowitz Test

b. Grouping Variable: City

c. There are 1 inter-group ties involving 2 cases.

Conclusion:

Since, test statistic for both the minimum and maximum possible runs which are the same for this data. And also p-value > 0.05, we may accept H_0 at 5% l.o.s. Hence we conclude that the prices in City A and City B can be taken to be following same probability distribution.

Mann-Whitney-Wilcoxon U test

In statistics, the Mann–Whitney *U* test (also called the Mann–Whitney–Wilcoxon (MWW), Wilcoxon rank- sum test, or Wilcoxon–Mann–Whitney test) is a nonparametric test of the null hypothesis that two populations are the same against an alternative hypothesis especially that a particular population tends to have larger values than the other. The Mann-Whitney test for testing independent samples is useful for determining if there exist significant differences between two independent samples. The Mann- Whitney test is the nonparametric version of the two-independent samples test.

It has greater efficiency than the t-test on non-normal distributions, such as a mixture of normal distributions, and it is nearly as efficient as the t-test on normal distributions. It is often used when the assumptions of the t-test have been violated. Thus it is useful if:

- The dependent variable is ordinal scaled instead of interval or ratio.
- The assumption of normality has been violated in a t-test (especially if the sample size is small.)
- The assumption of homogeneity of population variances has been violated in a t-test.

Let $X_1, \ldots, X_{n1}$ be a random sample of size n1 from $X \sim F_X$ and $Y_1, \ldots, Y_{n2}$ be a random sample of size n2 from $Y \sim F_Y$. We are interested in testing a hypothesis regarding equality of two population distributions on the basis of the available samples.
Let us set up the null and alternative hypotheses as follows:

Null Hypothesis H0: $F_X = F_Y$

Alternative Hypothesis H1: $F_X \neq F_Y$ (Two Tailed)

Test Statistic: Let $Z_1, \ldots, Z_{n1+n2}$ be the combined ordered sample. Let T = sum of the ranks of Y's in the combined ordered sample then define:

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

Intuition: If two samples are coming from different populations then one of the samples is expected to fall below or above the other sample, i.e., the sum of the ranks of sample units in one sample will be either very small or very large.

Decision Rule: For a pre-specified level of significance we reject $H0$ if $U \geq U_{\propto/2}$ or $U \leq U_{\propto/2}$, where $U_{\propto/2}$ and $U'_{\propto/2}$, are obtained by solving the following.

$$P\ (U \geq U_{\propto/2} |\ H0) \leq \propto/2 \qquad \text{and} \qquad P\ (U \leq U'_{\propto/2}\ ,|\ H0) \leq \propto/2$$

Case 5: An experiment on reading ability of students was conducted, where at the beginning of the year a class was randomly divided into two groups. One group was taught to read using a uniform method, where all the students progressed from one stage to the next at the same time, following the instructor's direction. The second group was taught to read using an individual method, where each student progressed at his own rate according to a programmed work book under the supervision of the instructor. At the end of the year each student was given a reading ability test and following were their scores.

First Group	227	176	252	149	16	55	234	194	247	92	184	147	88	161	171
Second Group	202	14	165	171	292	271	151	235	147	99	63	284	53	228	271

Ranks

	Group	N	Mean Rank	Sum of Ranks
Marks	0	15	14.47	217.00
	1	15	16.53	248.00
	Total	30		

Test Statistics[a]

	Marks
Mann-Whitney U	97.000
Wilcoxon W	217.000
Z	-.643
Asymp. Sig. (2-tailed)	.520
Exact Sig. [2*(1-tailed Sig.)]	.539[b]

a. Grouping Variable: Group

b. Not corrected for ties.

Conclusion:

Since p-value > 0.05, we may accept H_0 at 5% l.o.s. Thus, we conclude that the two different teaching methods for reading ability can be taken to be equally effective.

Run Test for Randomness

Assumptions:

1. The sample data are arranged according to some scheme (such as time series).
2. The data falls into two separate categories (such as above and below a specific value).
3. The runs test is based on the order in which the data occur; not on the frequency of the data.

Let $X1, \ldots, Xn$ be a Sample of size n. We are interested in testing a hypothesis whether the sample is random or not.

Let us set up the null and alternative hypotheses as follows:

H_0: Data is random .

H_1: Data is not random.

Test Statistic: Let M be the sample median then for each observation Define an indicator variable $\delta i = I(Xi > M)$, a realization of δi's can be 1001110010100011, define

U = # of run in the realization of δi's

Decision Rule: For a pre-specified level of significance we reject the hypothesis of randomness of data if $U \geq U_{\propto/2}$ or $U \leq U_{\propto/2}$, where $U_{\propto/2}$ and $U'_{\propto/2}$, are obtained by solving the following.

$$P(U \geq U_{\propto/2} | H0) \leq \propto/2 \quad \text{and} \quad P(U \leq U'_{\propto/2} , | H0) \leq \propto/2$$

If sample size is sufficiently large (>20) then we can use a normal approximation by using,

$Z = (U-(U))/\sqrt{Var(U)} \sim$ N(0,1) Under H_0 where $E(U) = \frac{n+2}{2}$ and $Var(U) = \frac{n}{4}[\frac{n-2}{n-1}]$ under H_0.

Now we can apply the normal test as follows:

Decision Rule: For a pre-specified level of significance we reject the hypothesis of randomness of data if $Z \geq Z_{\propto/2}$ or $Z \leq Z'_{\propto/2}$, where $Z_{\propto/2}$ and $Z'_{\propto/2}$, are obtained by solving the following.

$$1 - \Phi (Z_{\propto/2}|H_0) \leq \propto/2$$

and $$\Phi(Z'_{\propto/2} \; |H_0) \leq \propto/2$$

where $\Phi(.)$ is the CDF of a SNV.

Case 6: Test the randomness of following sample of size 30 using Run Test:

15, 77, 01, 65, 69, 69, 58, 40, 81, 16, 16, 20, 00, 84, 22, 28, 26, 46, 66, 36, 86, 66, 17, 43, 49, 85, 40, 51, 40,10

Runs Test

	sample
Test Value[a]	42
Cases < Test Value	15
Cases >= Test Value	15
Total Cases	30
Number of Runs	17
Z	.186
Asymp. Sig. (2-tailed)	.853

a. Median

Conclusion:

Since p-value > 0.05, we may accept H_0 at 5% l.o.s. Thus, we conclude that the sample (set of observations considered) is random.

KOLMOGOROV-SMIRNOV (KS) TEST :

The Kolmogorov–Smirnov test (K–S test) is a nonparametric test of the equality of continuous, one dimensional 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).

One Sample test:

One Sample KS Test is used to test if a sample taken from a specified population. The test statistic is calculated as a measure of Distance between the theoretical (to be tested) and empirical (observed) distribution functions. SPSS provides functionality to test if the sample is from one of the following distributions:

- Normal
- Exponential
- Poisson
- Uniform

Case 7: For the following four samples test if they are drawn from Normal, Exponential, Poisson and Uniform distributions respectively.

Observation	Sample 1	Sample 2	Sample 3	Sample 4
1	1.089781309	0.046136443	4	10.25068427
2	1.962787672	0.296905535	2	10.12379195
3	1.724451834	0.013852846	1	17.81733259
4	1.63955842	0.149763684	1	18.87337658
5	0.144050286	0.216846562	2	15.32347378
6	0.232942589	0.549152735	3	11.97729205
7	1.68271611	0.075868307	4	16.79090476
8	3.633887711	0.147932045	3	14.40535435
9	1.81341443	0.29035859	3	14.10547096
10	1.683039558	0.027180583	4	14.99055234
11	1.659612162	0.163903305	3	12.68408943
12	0.8396626	0.8104371	1	10.62609998
13	3.427254188	0.078686029	8	15.59840961
14	1.127955432	0.153359897	4	17.59452935
15	1.552543896	0.141724322	5	10.60249139
16	0.214796062	0.066255849	2	17.17324608
17	0.475882672	0.085298693	4	11.59441059
18	3.013061127	0.507875983	5	14.11860911
19	2.73502768	0.104899753	0	19.68738385
20	2.583921184	0.020127363	5	17.20303417

H_0: Sample 1 is drawn from normal population

H_1: Sample 1 is not drawn from normal population

One-Sample Kolmogorov-Smirnov Test1

		sample1	sample2	sample3	sample4
N		20	20	20	20
Normal Parameters[a,b]	Mean	1.66	.20	3.20	14.58
	Std. Deviation	1.028	.206	1.852	3.027
Most Extreme Differences	Absolute	.141	.264	.133	.118
	Positive	.141	.264	.133	.105
	Negative	-.108	-.187	-.117	-.118
Test Statistic		.141	.264	.133	.118
Asymp. Sig. (2-tailed)		.200[c,d]	.001[c]	.200[c,d]	.200[c,d]

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

d. This is a lower bound of the true significance.

Conclusion:

Since p-value > 0.05, we may accept H_0 at 5% l.o.s. Hence we conclude that the sample (set of observations considered) are drawn from Normal Distribution i.e. $Sample\ 1 \sim (1.66{,}1.05779)$

H_0: Sample 2 is drawn from exponential population

H_1: Sample 2 is not drawn from exponential population

One-Sample Kolmogorov-Smirnov Test 2

		sample1	sample2	sample3	sample4
N		20	20	20	20
Uniform Parameters[a,b]	Minimum	0	0	0	10
	Maximum	4	1	8	20
Most Extreme Differences	Absolute	.229	.512	.325	.147
	Positive	.229	.512	.325	.147
	Negative	-.054	-.050	-.075	-.066
Kolmogorov-Smirnov Z		1.023	2.288	1.453	.660
Asymp. Sig. (2-tailed)		.246	.000	.029	.777

a. Test distribution is Uniform.

b. Calculated from data.

Conclusion:

Since p-value > 0.05, we may accept H_0 at 5% l.o.s. Hence we conclude that the sample (set of observations considered) are drawn from Exponential Distribution i.e.$Sample\ 2 \sim E(5.0677)$

The Mean of Exponential Distribution is inverse of the parameter = 0.197

H_0: Sample 3 is drawn from Poisson population;

H_1: Sample 3 is not drawn from Poisson population

One-Sample Kolmogorov-Smirnov Test 3

		sample1	sample2	sample3	sample4
N		20[c]	20[d]	20	20[e]
Poisson Parameter[a,b]	Mean	1.66	.20	3.20	14.58
Most Extreme Differences	Absolute			.055	
	Positive			.055	
	Negative			-.053	
Kolmogorov-Smirnov Z				.248	
Asymp. Sig. (2-tailed)				1.000	

a. Test distribution is Poisson.

b. Calculated from data.

c. Poisson variables are non-negative integers. The value 0 occurs in the data. One-Sample Kolmogorov-Smirnov Test cannot be performed.

d. Poisson variables are non-negative integers. The value 0 occurs in the data. One-Sample Kolmogorov-Smirnov Test cannot be performed.

e. Poisson variables are non-negative integers. The value 10 occurs in the data. One-Sample Kolmogorov-Smirnov Test cannot be performed.

Conclusion:

Since p-value > 0.05, we may accept H_0 at 5% l.o.s. Hence we conclude that the sample (set of observations considered) are drawn from Poisson Distribution i.e. $Sample\ 3 \sim Po(3.20)$

H_0: Sample 4 is drawn from uniform population;

H_1: Sample 4 is not drawn from uniform population

One-Sample Kolmogorov-Smirnov Test 4

		sample1	sample2	sample3	sample4
N		20	20	20[c]	20
Exponential parameter.[a,b]	Mean	1.66	.20	3.37	14.58
Most Extreme Differences	Absolute	.257	.136	.237	.501
	Positive	.112	.136	.227	.259
	Negative	-.257	-.085	-.237	-.501
Kolmogorov-Smirnov Z		1.150	.607	1.034	2.239
Asymp. Sig. (2-tailed)		.142	.855	.235	.000

a. Test Distribution is Exponential.

b. Calculated from data.

c. There is 1 value outside the specified distribution range. This value is skipped.

Conclusion:

Since p-value > 0.05, we may accept H_0 at 5% l.o.s. Hence we conclude the sample (set of observations considered) are drawn from Uniform Distribution i.e. $Sample\ 4 \sim$ (10.124,19.687).

Two Sample KS Test

Two Sample KS Test is used to test if two samples are taken from same population. The test statistic is calculated as a measure of Distance between the empirical (observed) distribution functions of the samples.

Case 8: For the following two samples test if they can be taken to be coming from same population

Observation	Sample 1	Sample 2
1	0.075204597	1.319177696
2	0.282203071	0.255423126
3	0.473605304	0.250284353
4	0.171775727	0.941835437
5	0.084642496	3.078396099
6	0.601160542	0.270368067
7	0.212552515	0.413272132
8	0.294969478	0.05425652
9	0.026919861	1.340734424
10	0.054462148	0.127618122
11	0.076084169	0.060699583
12	0.021943532	0.208278913
13	0.486042232	0.104869289
14	0.083376869	1.126610877
15	0.62800881	1.179774988
16	1.317637268	2.015836491
17	0.431532897	0.43267859
18	0.151809043	0.686019322
19	0.645182388	1.210587738
20	0.018898663	0.230682213

H_0: The samples come from the same population

H_1: The samples do not come from the same population

Frequencies

	group	N
sample	1	20
	2	20
	Total	40

Test Statistics[a]

		sample
Most Extreme Differences	Absolute	.400
	Positive	.400
	Negative	.000
Kolmogorov-Smirnov Z		1.265
Asymp. Sig. (2-tailed)		.082

a. Grouping Variable: group

Conclusion:

Since p-value > 0.05, we may accept H_0 at 5% l.o.s. Thus, we conclude that the samples (2 sets of observations considered) can be taken to be coming from same population.

Chi –Square test

Chi – Square test is perhaps the most widely used statistical test. Most of the common Chi – Square tests viz. goodness of fit, independence of attributes etc. do not assume any assumption on the shape of the distribution and hence are nonparametric in nature.

For this case study we will be dealing with the following four ChiSquare tests.

Karl Pearson's Goodness of fit

It is very powerful non-parametric test for testing the significance of discrepancy between theory and experiment. It enables to find out if the deviation of the experiment form theory is just by chance or is it really to inadequacy of the theory to fit the observed data.

Let $X_1, \ldots, X_n$ be a random sample of size n from $X \sim F_X$. We are interested in testing a hypothesis if $F_X = F_X^0$ where F_X^0 is some hypothetical distribution.

Let us set up the null and alternative hypotheses as follows:

$H_0: F_X = F_X^0$

$H_1: F_X \neq F_X^0$

Test Statistic: Classify the sample data into k different groups/ classes and observe the frequencies (f_i ; i= 1, … , k) in the different classes. Now obtain the expected frequencies (e_i ; i = 1, … , k) for these groups using the distribution as F_X^0. Define

$$\chi^2 = \sum_{i=1}^{k} \left[\frac{(f_i - e_i)^2}{e_i}\right]$$

$$\sim \chi^2(n-1)$$

Where $\sum_{i=1}^{k} f_i = \sum_{i=1}^{k} e_i$

Decision Rule: Reject $_0$ if $\chi^2 \leq \chi^2_\alpha(n-1)$ where $\chi^2_\alpha(n-1)$ is obtained by solving the following.

$$P(\chi^2 \geq \chi^2_\alpha(n-1) | H_0) \leq \alpha$$

Case 9: A sample survey of 800 families each with 4 children was conducted and following distribution was observed.

# of Male Children	# of Female Children	Type	# of Families
0	4	1	32
1	3	2	178
2	2	3	290
3	1	4	236
4	0	5	64

Is the observed distribution consistent with the hypothesis that male and female births are equally probable?

H_0: Male and female births are equally probable.

H_1: Male and female births are not equally probable.

Test Statistics

	male
Chi-Square	20.211[a]
df	4
Asymp. Sig.	.000

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 48.0.

male

	Observed N	Expected N	Residual
0	32	48.0	-16.0
1	178	200.0	-22.0
2	290	304.0	-14.0
3	236	200.0	36.0
4	64	48.0	16.0
Total	800		

Conclusion:

Since p-value < 0.05, we reject $\boldsymbol{Ho}$ at 5% level of significance and hence conclude that the observed distribution is inconsistent with the hypothesis that male and female births are equally probable.

Independence of Attributes

The Chi-Square test for independence of attributes is based on exactly the same ideas as the goodness of fit test. Indeed they are the same tests.

Case 10: Out of 8000 graduates in a town 800 are females, out of 1600 graduate employees 120 are females. Use Chi – Square test to test if any sex discrimination is made in the employment.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	13.889[a]	1	.000		
Continuity Correction[b]	13.544	1	.000		
Likelihood Ratio	14.785	1	.000		
Fisher's Exact Test				.000	.000
N of Valid Cases	8000				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 160.00.

b. Computed only for a 2x2 table

Conclusion:

Since p-value < 0.05, we reject $\boldsymbol{Ho}$ at 5% level of significance and hence conclude that there is sex discrimination in the employment scenario.

McNemar's test

McNemar's test is a normal approximation used on nominal data. It is applied to 2 × 2 contingency tables with a dichotomous trait, with "matched pairs" of subjects, to determine whether the row and column marginal frequencies are equal ("marginal homogeneity"). It is named after Quinn McNemar, who introduced it in 1947.

The test is applied to a 2 × 2 contingency table, which tabulates the outcomes of two tests on a sample of n subjects, as follows

	Test 2 positive	Test 2 negative	Row total
Test 1 positive	a	b	a + b
Test 1 negative	c	d	c + d
Column total	a + c	b + d	N

The null hypothesis of marginal homogeneity states that the two marginal probabilities for each outcome are the same, i.e. $p_a + p_b = p_a + p_c$ and $p_c + p_d = p_b + p_d$. Thus the null and alternative hypotheses are given as follows:

H0: $p_b = p_c$;

H1: $p_b \neq p_c$

Here p_a, etc., denote the theoretical probability of occurrences in cells with the corresponding label. The McNemar test statistic is given as follows:

$$\chi^2 = \frac{(b-c)^2}{b+c}$$

The statistic with Yates's correction for continuity is given by:

$$\chi^2 = \frac{(|b-c|-0.5)^2}{b+c}$$

Under the null hypothesis, with a sufficiently large number of discordants (cells b and c), χ^2 has a chi-squared distribution with 1 degree of freedom. If either *b* or *c* is small ($b + c < 25$) then χ^2 is not well-approximated by the chi-squared distribution.

If the χ^2 result is significant, this provides sufficient evidence to reject the null hypothesis, in favour of the alternative hypothesis that $p_b \neq p_c$, which would mean that the marginal proportions are significantly different from each other.

<u>Case 11</u>: A researcher attempts to determine if a drug has an effect on a particular disease. Counts of individuals are given in the table, with the diagnosis (disease: *present* or *absent*) before treatment given in the rows, and the diagnosis after treatment in the columns. The test requires the same subjects to be included in the before-and-after measurements (matched pairs).

Effect of Treatment	After: present	After: absent	Row total
Before: present	101	121	222
Before: absent	59	33	92
Column total	160	154	314

Using McNemar's Test, test the hypothesis of "marginal homogeneity", i.e. there was no effect of the treatment.

before * after Crosstabulation

			after		Total
			absent	present	
before	absent	Count	33	59	92
		Expected Count	45.1	46.9	92.0
	present	Count	121	101	222
		Expected Count	108.9	113.1	222.0
Total		Count	154	160	314
		Expected Count	154.0	160.0	314.0

Chi-Square Tests

	Value	Exact Sig. (2-sided)
McNemar Test		.000[a]
N of Valid Cases	314	

a. Binomial distribution used.

Conclusion:

The Mc-Nemar's test statistic has significance (P-value) equal to 0.000. Thus, we conclude that at 5% l.o.s. there is a significant effect of drug on the particular disease. We would conclude that the drug has a significant effect on the particular disease

Cochran–Mantel–Haenszel test

The Cochran–Mantel–Haenszel test (which is sometimes called the Mantel–Haenszel test) is used for repeated tests of independence. There are three nominal variables; we want to know whether two of the variables are independent of each other, and the third variable identifies the repeats. The most common situation is that you have multiple 2×2 tables of independence, so that's what we will talk about here. There are versions of the Cochran–Mantel–Haenszel test for any number of rows and columns in the individual tests of independence, but we will cover only the case when repeated individual tests for independence have 2×2 only contingency tables.

For example, let's say we have found several hundred pink knit polyester legwarmers that have been hidden in a warehouse since they went out of style in 1984. We decide to see whether they reduce the pain of ankle osteoarthritis by keeping the ankles warm. In the winter, you recruit 36 volunteers with ankle arthritis, randomly assign 20 to wear the legwarmers under their clothes at all times while the other 16 don't wear the legwarmers, then after a month you ask them whether their ankles are pain-free or not. With just the one set of people, you'd have two nominal variables (legwarmers vs. control, pain-free vs. pain), each with two values, so we would analyze the data using the usual Chi-Square test for independence of attributes.

However, let's say we repeat the experiment in the spring, with 50 new volunteers. Then in the summer we repeat the experiment again, with 28 new volunteers. We could just add all the data together and do the usual Chi-Square test for independence of attributes on the 114 total people, but it would be better to keep each of the three experiments separate. Maybe the first time we did the experiment there was an overall higher level of ankle pain than the second time, because of the different time of year or the different set of volunteers. We want to see whether there's an overall effect of legwarmers on ankle pain, but we want to control for possibility of different levels of ankle pain at the different times of year.

Null and Alternative Hypotheses: The null hypothesis is that the two nominal variables that are tested within each repetition are independent of each other; having one value of one variable does not mean that it's more likely that we will have one value of the second variable. For the legwarmers experiment, the null hypothesis would be that the proportion of people feeling pain was the same for legwarmer-wearers and non-legwarmer wearers, after controlling for the time of year. The alternative hypothesis is that the proportion of people feeling pain was different for legwarmer and non-legwarmer wearers.

Technically, the null hypothesis of the Cochran–Mantel–Haenszel test is that the odds ratios within each repetition are equal to 1. The odds ratio is equal to 1 when the proportions are the same, and the odds ratio is different from 1 when the proportions are different from each other. I think proportions are easier to grasp than odds ratios, so I'll put everything in terms of proportions.

Case 12: McDonald and Siebenaller (1989) surveyed allele frequencies at the Lap locus in the mussel Mytilus trossulus on the Oregon coast. At four estuaries, samples were taken from inside the estuary and from a marine habitat outside the estuary. There were three common alleles and a couple of rare alleles;

based on previous results, the biologically interesting question was whether the Lap ("94") allele was less common inside estuaries, so all the other alleles were pooled into a "non-Lap"("non-94") class.

There are three nominal variables: allele (94 or non-94), habitat (marine or estuarine), and area (Tillamook, Yaquina, Alsea, or Umpqua). The following table shows the number of 94 and non-94 alleles at each location.

Location	Allele	Marine	Estuarine
Tillamook	94	56	69
	non-94	40	77
Yaquina	94	61	257
	non-94	57	301
Alsea	94	73	65
	non-94	71	79
Umpqua	94	71	48
	non-94	55	48

Using Cochran–Mantel–Haenszel test, test the null hypothesis that at each area, there is no difference in the proportion of Lap alleles between the marine and estuarine habitats, after controlling for area.

Tests of Homogeneity of the Odds Ratio

	Chi-Squared	df	Asymp. Sig. (2-sided)
Breslow-Day	.529	3	.912
Tarone's	.529	3	.912

Tests of Conditional Independence

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	5.338	1	.021
Mantel-Haenszel	5.050	1	.025

Conclusion:

Using both Breslow-Day as well as Tarone's test, we get significance (P-value) equal to 0.912. Thus, we may conclude that at 5% l.o.s. odds ratio are homogeneous for each location/area.

Using both Cochran's as well as Mantel-Haenszel test, we get significance (P-value) equal to 0.021 and 0.025 respectively. Thus, we may conclude that at 5% l.o.s. the areas are not independent of each other.

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			1.317
ln(Estimate)			.276
Std. Error of ln(Estimate)			.119
Asymp. Sig. (2-sided)			.021
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	1.042
		Upper Bound	1.665
	ln(Common Odds Ratio)	Lower Bound	.042
		Upper Bound	.510

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

The estimate of the Mantel-Haenszel common odds ratio as 1.317 and its natural logarithm as 0.276. It gives the standard error of natural logarithm of estimate as 0.119. It also provides the asymptotic 95% confidence interval for the common odds ratio and natural logarithm of common odds ratio as (1.042 , 1.665) and (0.042 , 0.510) respectively.

Sign- off Note

When we use non-parametric test we make a trade off. We lose sharpness but we gain the ability to use less information and to calculate faster. Thus, non-parametric tests are capable of being used in all problem situations.

For sufficiently large samples , power of non-parametric tests approaches to power of parametric tests.