***Case Study-Non Parametric Inference***

***Submitted by: T. Vijaya***

***Introduction***

A parametric statistical test is one that makes assumptions about the parameters (defining properties) of the population distribution(s) from which one's data are drawn. A non-parametric test is one that makes no such assumptions.Non-parametric statistical procedures are less powerful because they use less information in their calculation. For example, a parametric correlation uses information about the mean and deviation from the mean while a non-parametric correlation will use only the ordinal position of pairs of scores.

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

**Non-Parametric inference**

Non – Parametric inference involves estimation and testing procedures when shape of the population distribution is unknown. Nonparametric tests are often used in place of their parametric counterparts when certain assumptions about the underlying population are questionable. For example, when comparing two independent samples, the [Wilcoxon Mann-Whitney test](http://www.stats.gla.ac.uk/steps/glossary/nonparametric.html#wmwt) does not assume that the difference between the samples is normally distributed whereas its parametric counterpart, the [two sample t-test](http://www.stats.gla.ac.uk/steps/glossary/hypothesis_testing.html#2sampt) does.

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.

Nonparametric tests may be, and often are, more powerful in detecting population differences when certain assumptions are not satisfied. All tests involving ranked data, i.e. data that can be put in order, are nonparametric. The assumptions are:

1) Sample observations are independent.

2) Variable under study is continuous.

3) The p.d.f is continuous.

4) Lower order moments exist.

These assumptions are fewer and much weaker than those associated with parametric tests.

In [statistics](https://en.wikipedia.org/wiki/Statistics), the **p-value** is a function of the observed sample results (a [statistic](https://en.wikipedia.org/wiki/Statistic)) that is used for [testing a statistical hypothesis](https://en.wikipedia.org/wiki/Statistical_hypothesis_testing). More specifically, the p-value is defined as the probability of obtaining a result equal to or "more extreme" than what was actually observed, assuming that the hypothesis under consideration is true Before the test is performed, a threshold value is chosen, called the significance of the test, traditionally 5% or 1% and denoted as α.

*Decision rule: When the p value is less than or equal to the level of significance α, we can reject H0.*

Within this case study we will be dealing with the following cases:

**Frank Wilcoxon Sign Test: (3 Cases)**

**1.1 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.

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

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

**Use sign test to test the following two hypothesis:**

**The Height of the females can be taken to be equal to 64 inches.**

**The Weight of the females can be taken to be equal to 135 lbs.**

***Conditions of applicability***

The Binomial Test procedure is useful when you want to compare a single sample from a dichotomous variable to an expected proportion. If the dichotomy does not exist in the data as a variable, one can be dynamically created based upon a cut point on a scale variable.

The hypothesis is set as follows:

Ho:The height of the females is 64 inches.

H1: The height of the females is not 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 |  |  |

We observe that there are 2 groups. Group 1 consists of females with height <=64 and group 2 with height >64. The column labelled N tells us that there are 8 people who belong to group 2 and 7 people who belong to group 1. The Observed Prop. Column gives the observed proportions (.47 = 7 / (7 + 8)).

The next column, Test Prop., gives the value that you entered in the Test Proportion box in the Binomial Test dialog box.Since p value is 1 (>0.05) it is significant and we accept null hypothesis and conclude that the height of the females is 64 inches.

The hypothesis is set as follows:

Ho: The weight of the females is 135 lbs.

H1: The weight of the females is not 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 |  |  |

We infer that the observed proportions for group1 is (0.53=8/7+8) and for group 2 is .47. The next column, Test Prop., gives the value that you entered in the Test Proportion box in the Binomial Test dialog box. Since p value is 1.000 (>0.05) we accept thenull hypothesis and conclude that the weight of the females is 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.**

We set the following hypothesis:

Ho: win and loss are equally likely

H1: win and loss are not equally likely

| **Binomial Test** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Category | N | Observed Prop. | Test Prop. | Asymp. Sig. (2-tailed) |
| Outcome | Group 1 | <= .5 | 14 | .28 | .50 | **.003a** |
| Group 2 | > .5 | 36 | .72 |  |  |
| Total |  | 50 | 1.00 |  |  |
| a. Based on Z Approximation. | | | | | | |

Since the asymptotic significant value is 0.003 (<0.05) therefore H0 is rejected at 5% level of significance. In other words win and loss are not equally likely.

**1.2 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. 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 | 09 | 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.**

***Conditions of applicability***

1. When there are pairs of observations on 2 things being compared.
2. For any given pair each of the 2 observations is made under similar extraneous conditions.
3. Different pairs are observed under different conditions.

The only assumptions are:

1) Deviations can be expressed in terms of positive and negative signs.

2) Variables have continuous distribution.

3) The deviations are independent.

| **Frequencies** | | |
| --- | --- | --- |
|  |  | N |
| Professor B - Professor A | Negative Differencesa | 10 |
| Positive Differencesb | 12 |
| Tiesc | 0 |
| Total | 22 |
| a. Professor B < Professor A | | |
| b. Professor B > Professor A | | |
| c. Professor B = Professor A | | |

This table depicts the differences in marks of professor A and professor B.

| **Test Statisticsb** | |
| --- | --- |
|  | Professor B - Professor A |
| Exact Sig. (2-tailed) | .832a |
| a. Binomial distribution used. | |
| b. Sign Test | |

We set the following hypothesis:

H0: grading of both the professors are same

H1: grading of both the professors are not same

Since the significant value is 0.832 (>0.05) therefore H0 is accepted at 5% level of significance. In other words grading of both the professors is same.

**Wald – Wolfowitz Run Test: (1 Case)**

**Definition.** 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.

Test Statistic: Let  be the combined ordered sample then define U = # of runs in the combined ordered sample

**Intuition** If two samples are coming from the same population then there would be a thorough mingling of X’s and Y’s and consequently # of runs in the combined ordered sample would be large. On the other hand if the samples are coming from two different populations then # of runs in the combined ordered samples would expected to be small. But “HOW SMALL??” so that we can accept the alternative

**Decision Rule:**

Reject  if  where is obtained by solving the following.



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

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

We set the following hypothesis:

H0: Prices in City A and City B can be taken to be following same probability distribution

H1: Prices in City A and City B cannot be taken to be following same probability distribution

| **Frequencies** | | |
| --- | --- | --- |
|  | F | N |
| Com | 1.00 | 15 |
| 2.00 | 13 |
| Total | 28 |

It gives number of observations belonging to city A and city B in the combined sample.

| **Test Statisticsb,c** | | | | |
| --- | --- | --- | --- | --- |
|  |  | Number of Runs | Z | Exact Sig. (1-tailed) |
| Com | Minimum Possible | 14a | -.166 | **.436** |
| Maximum Possible | 14a | -.166 | **.436** |
| a. There are 1 inter-group ties involving 2 cases. | | | | |
| b. Wald-Wolfowitz Test | | | | |
| c. Grouping Variable: F | | | | |

When ties are broken in all possible ways, the minimum number of runs is14, and the maximum is 14. The smallest possible exact p value is thus equal to the largest possible exact p value which is 0.436. It implies that you may accept the null hypothesis that prices in city A and city B are the same.

**Mann – Whitney – Wilcoxon U – Test: (1 Case)**

U-test is a non-parametric test to test a hypothesis regarding equality of two population distributions.

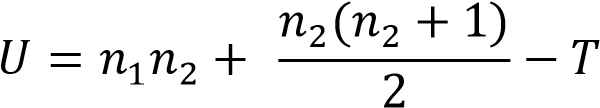
Let  be a RS of size n from  and  be a RS of size n from. We are interested in testing a hypothesis regarding equality of two population distributions on the basis of the available samples.

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

Null Hypothesis: 

Alternative Hypothesis: (Two Tailed)

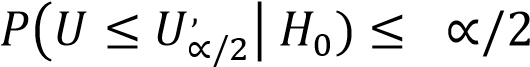
**Test Statistic:** Let  be the combined ordered sample. Let  = sum of the ranks of  in the combined ordered sample then define:



**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.

Obtaining exact distribution of U is quite troublesome. However, it has been obtained for small values of  and.

**Decision Rule:** For a pre-specified level of significance we reject  where and are obtained by solving the following.

And

**Mann-Whitney TestCase 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 |

Use U-test to test if two different teaching methods for reading ability can be taken as equally effective.

**Conditions of applicability**

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

The two-sample Kolmogorov-Smirnov test: To test the null hypothesis that two samples have the same distribution Wald-Wolfowitz Run: Used to examine whether two random samples come from populations having same distribution

We set the following hypothesis:

H0: two different teaching methods for reading ability can be taken as equally effective

H1: two different teaching methods for reading ability cannot be taken as equally effective

| **Ranks** | | | | |
| --- | --- | --- | --- | --- |
|  | Flags | N | Mean Rank | Sum of Ranks |
| Combined marks | 1.00 | 15 | 14.47 | 217.00 |
| 2.00 | 15 | 16.53 | 248.00 |
| Total | 30 |  |  |

We infer that the above table provides information regarding the output of the actual Mann-Whitney U test. It shows mean rank and sum of ranks for the two groups tested.

The table because above is very useful it indicates which group can be considered as having the higher marks, overall; namely, the group with the highest mean rank. In this case, group 2 had the higher marks.

| **Test Statisticsb** | |
| --- | --- |
|  | Combined marks |
| Mann-Whitney U | 97.000 |
| Wilcoxon W | 217.000 |
| Z | -.643 |
| Asymp. Sig. (2-tailed) | .520 |
| Exact Sig. [2\*(1-tailed Sig.)] | .539a |
| a. Not corrected for ties. | |
| b. Grouping Variable: Flags | |

This table shows us the actual significance value of the test. Specifically, the **Test Statistics** table provides the test statistic, U statistic, as well as the asymptotic significance (2-tailed) p-value.

Depending on the size of your groups, SPSS Statistics will produce both exact and asymptotic statistical significance levels. Now two p values are displayed, asymptotic which is appropriate for large sample and exact which is independent of sample size. Therefore we will take the exact p value (as it is a small sample of size 30) i. e. 0.539(>0.05) hence we may accept null hypothesis and conclude that there is no significant difference in marks between group 1 and group 2 or the 2 teaching methods are equally effective.

**Run Test for Randomness**: (1 Case)

**Assumptions:**

The sample data are arranged according to some scheme (such as time series).

The data falls into two separate categories (such as above and below a specific value).

The runs test is based on the order in which the data occur; not on the frequency of the data.

Let 𝑋1…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:

Null Hypothesis: 𝐻0:Data is random

Alternative Hypothesis: 𝐻1:Data is not random

**Rationale:**

Suppose we have a sample of size 9, where each unit is either H (High) or L (Low). Further suppose there are 4 H and 5 L. As occurrence H or L is probabilistic. There can be are 9! / (4!\*5!) = 126 different arrangements of H and L.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **# of Runs** | **Cases** | **Probability** |
| 9 | (𝐿𝐻𝐿𝐻𝐿𝐻𝐿𝐻𝐿) |  |
| 8  … | (𝐿𝐿𝐻𝐿𝐻𝐿𝐻𝐿𝐻), (𝐿𝐻𝐿𝐿𝐻𝐿𝐻𝐿  (𝐿𝐻𝐿𝐻𝐿𝐿𝐻𝐿𝐻), (𝐿𝐻𝐿𝐻𝐿𝐻𝐿𝐿  (𝐻𝐿𝐿𝐻𝐿𝐻𝐿𝐻𝐿), (𝐻𝐿𝐻𝐿𝐻𝐿𝐻𝐿  (𝐻𝐿𝐻𝐿𝐿𝐻𝐿𝐻𝐿), (𝐻𝐿𝐻𝐿𝐻𝐿𝐿𝐻 |  |
| 2 | (𝐻𝐻𝐻𝐻𝐻𝐿𝐿𝐿𝐿),  (𝐿𝐿𝐿𝐿𝐻𝐻𝐻𝐻𝐻) |  |

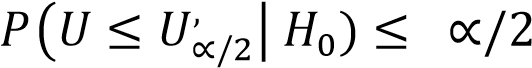
If sample size is sufficiently large (>20) then we can use a normal approximation by using, obviously we don’t want to do this forever. There are tables for small values, and there is a normal approximation for large data.

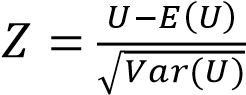
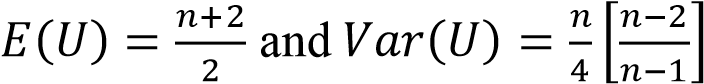
We can observe that for a data to be random (meaning by non-existence of any pattern) will be supported if # of runs is neither too large nor too small.

**Test Statistic:** Let 𝑀 be the sample median then for each observation Define an indicator variable 𝛿𝑖= 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 ≥  where are obtained by solving the following.

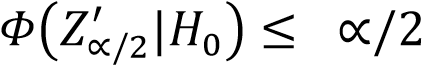
(𝑈 ≥ 𝑈∝/2| 𝐻0) ≤ ∝/2 and

 ~ N (0, 1) under𝐻0where  under 𝐻0.

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 ≥  where 𝑍∝/2 and are obtained by solving the following.

1 − (𝑍∝/2|𝐻0) ≤ ∝/2

And

Where(.) is the CDF of a S.N.V.

**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 Valuea | 41.50 |
| Cases < Test Value | 15 |
| Cases >= Test Value | 15 |
| Total Cases | 30 |
| Number of Runs | 17 |
| Z | .186 |
| Asymp. Sig. (2-tailed) | **.853** |
| a. Median | |

We set the following hypothesis:

𝐻0:Data is random

𝐻1:Data is not random

Now p value is 0.853 (>0.05) so we may accept null hypothesis and conclude that the data is random.

**Kolmogorov – Smirnov (KS) Test: (2 Cases)**

**5.1 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:

1. Normal
2. Exponential
3. Poisson
4. 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 |

**We set the following hypothesis**

H01:the 1st sample comes from Normal Distribution

H11:the 1st sample does not come from Normal Distribution

| **One-Sample Kolmogorov-Smirnov Test** | | |
| --- | --- | --- |
|  |  | S1 |
| N | | 20 |
| Normal Parametersa,,b | Mean | 1.661817346100E0 |
| Std. Deviation | 1.0284977352978E0 |
| Most Extreme Differences | Absolute | .141 |
| Positive | .141 |
| Negative | -.108 |
| Kolmogorov-Smirnov Z | | .632 |
| Asymp. Sig. (2-tailed) | | **.819** |
| a. Test distribution is Normal. | | |
| b. Calculated from data. | | |

The p value is 0.819 which is not significant and therefore we say that the random sample has an approximate normal distribution. (If the p value were less than 0.05 we would say it is significant and random sample does not follow an approximate normal distribution.)

**We set the following hypothesis**

H02: the 2nd sample comes from Exponential Distribution

H12: the 2nd sample does not come from Exponential Distribution

| **One-Sample Kolmogorov-Smirnov Test** | | |
| --- | --- | --- |
|  |  | S2 |
| N | | 20 |
| Exponential parameter.a,,b | Mean | .197328281200 |
| Most Extreme Differences | Absolute | .136 |
| Positive | .136 |
| Negative | -.085 |
| Kolmogorov-Smirnov Z | | .607 |
| Asymp. Sig. (2-tailed) | | **.855** |
| a. Test Distribution is Exponential. | | |
| b. Calculated from data. | | |

The p value is 0.855 which is not significant and therefore we say that the random sample has an approximate exponential distribution.

**We set the following hypothesis**

H03: the 3rd sample comes from Poisson distribution

H13: the 3rd sample does not come from Poisson distribution**.**

| **One-Sample Kolmogorov-Smirnov Test** | | |
| --- | --- | --- |
|  |  | S3 |
| N | | 20 |
| Poisson Parametera,,b | Mean | 3.20 |
| 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. | | |

The p value is 1.00 which is not significant and therefore we can say that the random sample has an approximate Poisson distribution.

**We set the following hypothesis**

H04: the 4th sample comes from Uniform Distribution

H14: the 4th sample does not come from Uniform Distribution

| **One-Sample Kolmogorov-Smirnov Test** | | |
| --- | --- | --- |
|  |  | S4 |
| N | | 20 |
| Uniform Parametersa,,b | Minimum | 1.0123791950E1 |
| Maximum | 1.9687383850E1 |
| Most Extreme Differences | Absolute | .147 |
| Positive | .147 |
| Negative | -.066 |
| Kolmogorov-Smirnov Z | | .660 |
| Asymp. Sig. (2-tailed) | | **.777** |
| a. Test distribution is Uniform. | | |
| b. Calculated from data. | | |

The p value is 0.777 which is not significant and therefore we say that the random sample has an approximate uniform distribution

**5.2 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 |

| **Frequencies** | | |
| --- | --- | --- |
|  | Flag | N |
| Com | 1.00 | 20 |
| 2.00 | 20 |
| Total | 40 |

The sample 1 is assigned a flag of 1 and sample 2 is assigned a flag of 2.

| **Test Statisticsa** | | |
| --- | --- | --- |
|  |  | Com |
| Most Extreme Differences | Absolute | .400 |
| Positive | .400 |
| Negative | .000 |
| Kolmogorov-Smirnov Z | | 1.265 |
| Asymp. Sig. (2-tailed) | | **.082** |
| a. Grouping Variable: Flag | | |

The asymptotic significant p value is 0.082. This demonstrates that, there is a statistically significant difference between the two samples therefore H0 is accepted at 5% level of significance. In other words both the samples are coming from the same population.

**Chi – Square Tests**

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 Chi-square tests.

**6.1 Karl Pearson’s Goodness of Fit: (1 Case**)

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 𝑋1…𝑋𝑛 be a random sample of size n from 𝑋~𝐹𝑋. We are interested in testing a hypothesis if  where  is some hypothetical distribution.

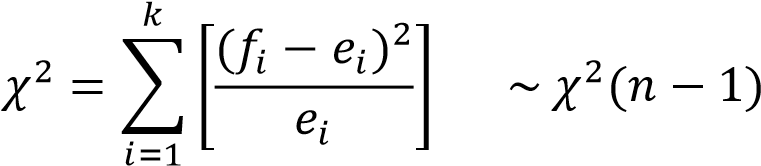
Let us set up the null and alternative hypotheses as follows:

Null Hypothesis: 

Alternative Hypothesis: 𝐻1:𝐹𝑋≠𝐹𝑋0

**Test Statistic:** Classify the sample data into k different groups/ classes and observe the frequencies (𝑓𝑖; i =

1… k) in the different classes. Now obtain the expected frequencies (𝑒𝑖; i = 1… k) for these groups using the distribution as. Define



Where.

**Decision Rule:** Reject  where  is obtained by solving the following.

(𝜒2 ≥ 𝜒∝2 (𝑛 − 1)| 𝐻0) ≤ ∝

**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** | **# of Families** |
| **0** | **4** | **32** |
| **1** | **3** | **178** |
| **2** | **2** | **290** |
| **3** | **1** | **236** |
| **4** | **0** | **64** |

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

We set the following hypothesis:

H0:male and female births are equally probable.

H1:male and female births are not equally probable.

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

**Sampling Chi-Square Test**

| **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 |  |  |

| **Test Statistics** | |
| --- | --- |
|  | Male |
| Chi-Square | 20.211a |
| df | 4 |
| Asymp. Sig. | **.000** |
| a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 48.0. | |

Here p value is 0.000 (< 0.05). Hence it is significant and we reject the null hypothesis at 5 % level of significance and conclude that there is significant difference in the proportions of males and females. Hence it implies that male and female births are not equally probable.

**6.2Independence of Attributes: (1 Case)**

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.

We set the following hypothesis-

H0: gender and employment are independent i.e. Sex discrimination is not made in the employment

H1: gender and employment are not independent i.e. Sex discrimination is made in the employment

**Crosstabs**

| **Sex \* EmploymentCross tabulation** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  |  |  | Employment | | Total |
|  |  |  | Employed | Unemployed |
| Sex | F | Count | 120 | 680 | 800 |
| Expected Count | 160.0 | 640.0 | 800.0 |
| M | Count | 1480 | 5720 | 7200 |
| Expected Count | 1440.0 | 5760.0 | 7200.0 |
| Total | | Count | 1600 | 6400 | 8000 |
| Expected Count | 1600.0 | 6400.0 | 8000.0 |

This table gives the chi-square table of observed and expected frequencies for each possible combination of the two variables. In this example, 120 personsare females and employed. The expected frequency for this cell under H0 is 160(from the Expected Count row of the F row and employed column.)

| **Chi-Square Tests** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
| Pearson Chi-Square | **13.889a** | 1 | **.000** |  |  |
| Continuity Correctionb | 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%) have expected count less than 5. The minimum expected count is 160.00. | | | | | |
| b. Computed only for a 2x2 table | | | | | |

The value of the chi-squared statistic is 13.889.The p value equals .000.As p value<0.05 we may reject null hypothesis and conclude that sex discrimination is made. The footnote at the bottom of the Chi-Square output tells us that 0% of the cells have expected frequencies less than 5. Thus, none of the assumptions of chi-square has been violated and the results are meaningful

**6.3 McNemar’s Test: (1 Case)**

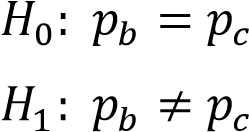
**McNemar's test** is a statistical test used on nominal data. It is applied to 2 × 2 contingency tables with a dichotomous trait, with “matched pairs” of subjects, to determine whether the row and column marginal frequencies are equal ("marginal homogeneity"). It is named after Quinn McNemar’s, who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

The McNemar’s test is used to test the equality of binary response rates from two populations in which the data consist of paired, dependent responses, one from each population. It is typically used in a repeated measurements situation in which each subject’s response is elicited twice, once before and once after a specified event (treatment) occurs. The test then determines if the initial response rate (before the event) equals the final response rate (after the event).

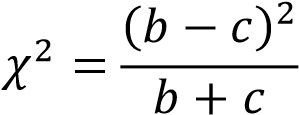
The test is applied to a 2 × 2 contingency table, which tabulates the outcomes of two tests on a sample of *n* subjects, as follows:

|  |  |  |  |
| --- | --- | --- | --- |
|  | 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 |

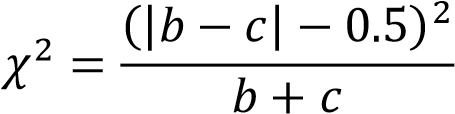
The null hypothesis of marginal homogeneity states that the two marginal probabilities for each outcome are the same, i.e. and. Thus the null and alternative hypotheses are given as follows:



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



The statistic with[Yates's correction for continuity](http://en.wikipedia.org/wiki/Yates%27s_correction_for_continuity)is given by:



Under the null hypothesis, with a sufficiently large number of discordances (cells and),  has a chi-squared distribution with 1 degree of freedom. If either  or is small  then is not well-approximated by the chi-squared distribution and we then use the Binomial distribution.

If the 𝜒2 result is significant, this provides sufficient evidence to reject the null hypothesis, in favour of the alternative hypothesis that𝑝𝑏 ≠ 𝑝𝑐, which would mean that the marginal proportions are significantly different from each other.

**Case 11:** 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).

|  |  |  |  |
| --- | --- | --- | --- |
|  | **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.

**When to use (McNemar’s Test assumptions)**

The 2 groups of our dependent variable must be mutually exclusive.

2) The cases (participants) are a random sample from the population of interests.

If the samples are not independent, but instead are before-and-after observations on the same individuals, we should use [McNemar's test](http://www.biostathandbook.com/fishers.html#mcnemars).

**We set the following hypothesis:**

H0: there was no effect of the treatment

H1: there was some effect of the treatment

| **Before \* After Cross tabulation** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  |  |  | 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 |

We can see that there were **121** participants who were suffering from the disease but after the treatment they became disease free. These participants can be considered as the treatments successes. However, by consulting the top right cell we can see that 59 non-diseased participants actually contracted the disease following the treatment.

In all the table denotes the different directions of travel that the participants took.

| **Chi-Square Tests** | | |
| --- | --- | --- |
|  | Value | Exact Sig. (2-sided) |
| McNemar Test |  | .000a |
| N of Valid Cases | 314 |  |
| a. Binomial distribution used. | | |

Since the significant value is 0.000 (<0.05) therefore we have a statistically significant resultand the proportion of non-diseased before and after the treatment is statistically significantly different.H0 is rejected at 5% level of significance. In other words marginal homogeneity exists or there was no effect of the treatment.

**Note: We may not have the exact p value but an asymptotic p value instead. This is because SPSS statistics calculates the p value differently depending on the no. of discordant pairs in our cross tabulation table.**

**6.4 Cochran – Mantel – Haenszel Test: (1 Case)**

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 analyse 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.

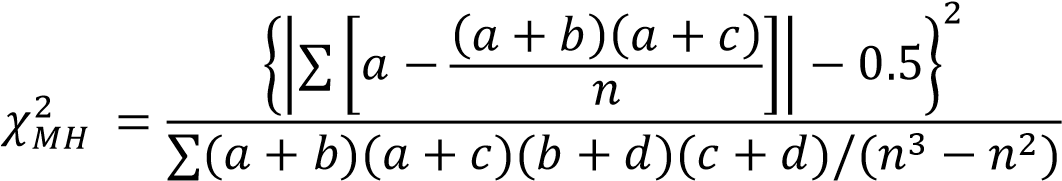
Technically, the null hypothesis of the Cochran–Mantel–Haenszel test is that the odds ratios within each repetition are equal to 1. The odds ratio is equal to 1 when the proportions are the same, and the odds ratio is different from 1 when the proportions are different from each other. As proportions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

**How does the test work – An Illustration:**

If the four numbers in a 2×2 test of independence are labelled like this:

|  |  |
| --- | --- |
| a | b |
| c | d |

And, the equation for the Cochran–Mantel–Haenszel test statistic can be written like this:



The numerator contains the absolute value of the difference between the observed value in one cell (*a*) and the expected value under the null hypothesis,, so the numerator is the squared sum of deviations between the observed and expected values. It doesn't matter how we arrange the 2×2 tables, any of the four values can be used as *a*. The 0.5 is subtracted as a continuity correction. The denominator contains an estimate of the variance of the squared differences.

The test statistic, , gets bigger as the differences between the observed and expected values get larger, or as the variance gets smaller (primarily due to the sample size getting bigger). It is chi-square distributed with one degree of freedom.

Some statisticians recommend that you test the homogeneity of the odds ratios (follow <https://en.wikipedia.org/wiki/Odds_ratio>and <http://vassarstats.net/odds2x2.html>to read about odds ratio) in the different repeats, and if different repeats show significantly different odds ratios, you shouldn't do the Cochran– Mantel–Haenszel test. In our arthritis-legwarmers example, they would say that if legwarmers have a significantly different effect on pain in the different seasons, you should analyse each experiment separately, rather than all together as the Cochran–Mantel–Haenszel test does. The most common way to test the homogeneity of odds ratios is with the Breslow–Day test.

Other statisticians will tell you that it's perfectly okay to use the Cochran–Mantel–Haenszel test when the odds ratios are significantly heterogeneous. The different recommendations depend on what your goal is. If your main goal is hypothesis testing—you want to know whether legwarmers reduce pain, in our example—then the Cochran–Mantel–Haenszel test is perfectly appropriate. A significant result will tell you that yes, the proportion of people feeling ankle pain does depend on whether or not they're wearing legwarmers. If your main goal is estimation—you want to estimate how well legwarmers work and come up with a number like "people with ankle arthritis are 50% less likely to feel pain if they wear fluorescent pink polyester knit legwarmers"—then it would be inappropriate to combine the data using the Cochran–Mantel–Haenszel test. If legwarmers reduce pain by 70% in the winter, 50% in the spring, and 30% in the summer, it would be misleading to say that they reduce pain by 50%; instead, it would be better to say that they reduce pain, but the amount of pain reduction depends on the time of year.

***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 |
|  | non94 | 40 | 77 |
| Yaquina | 94 | 61 | 257 |
|  | non94 | 57 | 301 |
| Alsea | 94 | 73 | 65 |
|  | non94 | 71 | 79 |
| Umpqua | 94 | 71 | 48 |
|  | non94 | 55 | 48 |

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

We set up the following hypothesis

H0: there was no difference in the proportion of Lap 94 alleles between the marine and estuarine habitats, after controlling for area

H1: there was difference in the proportion of Lap 94 alleles between the marine and estuarine habitats, after controlling for area

There are 3 nominal variables allele (94, non94), habitat (marine, estuarine) and area (Tillamook Yaquina Alsea Umpqua).This table shows the number of 94 and non-94 alleles at each location. There is a smaller proportion of 94 alleles in the estuarine location of each estuary when compared with the marine location; we wanted to know whether this difference is significant.

|  |  |  |  |
| --- | --- | --- | --- |
| **Location** | **Allele** | **Marine** | **Estuarine** |
| Tillamook | 94 | 56 | 69 |
|  | non-94 | 40 | 77 |
|  | percent 94 | 58.3% | 47.3% |
| Yaquina | 94 | 61 | 257 |
|  | non-94 | 57 | 301 |
|  | percent 94 | 51.7% | 46.1% |
| Alsea | 94 | 73 | 65 |
|  | non-94 | 71 | 79 |
|  | percent 94 | 50.7% | 45.1% |
| Umpqua | 94 | 71 | 48 |
|  | non-94 | 55 | 48 |
|  | percent 94 | 56.3% | 50.0% |

**Crosstabs**

| **Alleles \* Habitat \* Location Crosstabulation** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| Location | | | | Habitat | | Total |
| 1.00 | 2.00 |
| 1.00 | Alleles | 1.00 | Count | 56 | 69 | 125 |
| Expected Count | 49.6 | 75.4 | 125.0 |
| 2.00 | Count | 40 | 77 | 117 |
| Expected Count | 46.4 | 70.6 | 117.0 |
| Total | | Count | 96 | 146 | 242 |
| Expected Count | 96.0 | 146.0 | 242.0 |
| 2.00 | Alleles | 1.00 | Count | 61 | 257 | 318 |
| Expected Count | 55.5 | 262.5 | 318.0 |
| 2.00 | Count | 57 | 301 | 358 |
| Expected Count | 62.5 | 295.5 | 358.0 |
| Total | | Count | 118 | 558 | 676 |
| Expected Count | 118.0 | 558.0 | 676.0 |
| 3.00 | Alleles | 1.00 | Count | 73 | 65 | 138 |
| Expected Count | 69.0 | 69.0 | 138.0 |
| 2.00 | Count | 71 | 79 | 150 |
| Expected Count | 75.0 | 75.0 | 150.0 |
| Total | | Count | 144 | 144 | 288 |
| Expected Count | 144.0 | 144.0 | 288.0 |
| 4.00 | Alleles | 1.00 | Count | 71 | 48 | 119 |
| Expected Count | 67.5 | 51.5 | 119.0 |
| 2.00 | Count | 55 | 48 | 103 |
| Expected Count | 58.5 | 44.5 | 103.0 |
| Total | | Count | 126 | 96 | 222 |
| Expected Count | 126.0 | 96.0 | 222.0 |

This table gives us the expected frequencies for the 2 types of alleles in different locations.

| **Chi-Square Tests** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| Location | | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
| 1.00 | Pearson Chi-Square | 2.844a | 1 | **.092** |  |  |
| Continuity Correctionb | 2.418 | 1 | .120 |  |  |
| Likelihood Ratio | 2.853 | 1 | .091 |  |  |
| Fisher's Exact Test |  |  |  | .115 | .060 |
| Linear-by-Linear Association | 2.832 | 1 | .092 |  |  |
| N of Valid Cases | 242 |  |  |  |  |
| 2.00 | Pearson Chi-Square | 1.243c | 1 | .265 |  |  |
| Continuity Correctionb | 1.027 | 1 | .311 |  |  |
| Likelihood Ratio | 1.241 | 1 | .265 |  |  |
| Fisher's Exact Test |  |  |  | .267 | .155 |
| Linear-by-Linear Association | 1.241 | 1 | .265 |  |  |
| N of Valid Cases | 676 |  |  |  |  |
| 3.00 | Pearson Chi-Square | .890d | 1 | .345 |  |  |
| Continuity Correctionb | .682 | 1 | .409 |  |  |
| Likelihood Ratio | .891 | 1 | .345 |  |  |
| Fisher's Exact Test |  |  |  | .409 | .205 |
| Linear-by-Linear Association | .887 | 1 | .346 |  |  |
| N of Valid Cases | 288 |  |  |  |  |
| 4.00 | Pearson Chi-Square | .883e | 1 | .347 |  |  |
| Continuity Correctionb | .646 | 1 | .421 |  |  |
| Likelihood Ratio | .883 | 1 | .347 |  |  |
| Fisher's Exact Test |  |  |  | .415 | .211 |
| Linear-by-Linear Association | .879 | 1 | .348 |  |  |
| N of Valid Cases | 222 |  |  |  |  |
| a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 46.41. | | | | | | |
| b. Computed only for a 2x2 table | | | | | | |
| c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 55.51. | | | | | | |
| d. 0 cells (.0%) have expected count less than 5. The minimum expected count is 69.00. | | | | | | |
| e. 0 cells (.0%) have expected count less than 5. The minimum expected count is 44.54. | | | | | | |

Pearson s chi square under (Asymp. Sig. (2-sided)) should be 0.05 or lower to be significant. From above table none of the Pearson chi square statistics are significant.

| **Tests of Homogeneity of the Odds Ratio** | | | |
| --- | --- | --- | --- |
|  | Chi-Squared | df | Asymp. Sig. (2-sided) |
| Breslow-Day | .529 | 3 | .912 |
| Tarone's | .529 | 3 | .912 |

Ho- The odds ratios are equal across the different repeats.

H1-The odds ratios are not equal across the different repeats.

The Breslow-Day test for the data shows no significant evidence for heterogeneity of odds ratios (that is we accept null hypothesis and conclude that odds ratios are homogeneous) (χ2=0.529, 3 D.F., P=0.912).

| **Tests of Conditional Independence** | | | |
| --- | --- | --- | --- |
|  | Chi-Squared | df | Asymp. Sig. (2-sided) |
| Cochran's | 5.338 | 1 | .021 |
| Mantel-Haenszel | **5.050** | 1 | **.025** |
| Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0. | | | |

From the above table χ2=5.050, 1 D.F., P=0.025. We can reject the null hypothesis that the proportion of Lap94 alleles is the same in the marine and estuarine locations.

| **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. | | | |

1) The closer our lower bound or upper bound is to 1 , the less confident we should be in the common odds ratio estimate because it means that there is a chance that an actual common odds ratio is close to 1,which means there isn’t a difference.

**Note:** In addition to testing the null hypothesis, the Cochran-Mantel-Haenszel test also produces an estimate of the common odds ratio, a way of summarizing how big the effect is when pooled across the different repeats of the experiment. This require assuming that the odds ratio is the same in the different repeats. We can test this assumption using the Breslow-Day test.

If some repeats have a big difference in proportion in one direction, and other repeats have a big difference in proportions but in the opposite direction, the Cochran-Mantel-Haenszel test may give a non-significant result.

**References**

1) Discovering Statistics Using SPSS by Andy Field

2) Abhishek Sir s notes

3) Fundamentals of Mathematical Statistics by S.C. Gupta and V.K.Kapoor