International Journal Of Advanced Research In Engineering Technology & Sciences

ISSN: 2394-2819

Email: editor@ijarets.org

July- 2015 Volume 2, Issue-7

www.ijarets.org

# SOME CASE STUDIES FOR NON-PARAMETRIC TESTS FOR ORDINAL DATA

**Sahana Prasad**
Associate Professor
Department of Statistics
Christ University, Bangalore, India

**ABSTRACT:**
Ordinal scale is used to indicate the place of a category in an ordered series and they state the position of the characteristic with respect to others. Thus, in an ordinal scale, the values are hierarchical and state the relative position of a characteristic and unlike nominal scales, ordinal scales allow comparisons of the degree to which two subjects possess the characteristic. Some valid statistical tools which can be used for ordinal data are Median, Quartile deviation, Spearman's rank correlation, Mann-Whitney U test etc. Nonparametric tests are also called "distribution-free tests" as they are not based assumptions of normality . Parametric tests involve specific probability distributions and estimation of the key parameters of that distribution from the sample data. Generally , nonparametric tests are less powerful than their parametric counterparts . This paper explores some case studies using ordinal data.
Keywords: Ordinal data, Non parametric tests, One sample test, Chi-square, Mann- Whitney test

**INTRODUCTION:**
In an ordinal scale, the values are hierarchical and state the relative position of a characteristic and unlike nominal scales, ordinal scales allow comparisons of the degree to which two subjects possess the characteristic. Generally, the number 1 is assigned to that which has larger quantity of what is being studied just like a student with the highest marks gets the first rank. The others are ranked according to their marks. It has to be remembered though that the difference between two levels of an ordinal scale is not the same as the difference between two other levels. For e.g.: If 1 denotes highly satisfied, 2 denotes satisfied and 3 denotes moderately satisfied, 2-1 is not the same as 3-2. The following are the various scenarios under ordinal data.

**Case I:** One sample- to test for randomness of sample
**Tool:**
Run test for both small and large samples
**Case II**: Two independent samples
**Tool:**
1. Median test- two independent samples are tested for their difference in central tendencies
2. Mann- Whitney U test

**Case III:** two related samples
**Tool:**
1.Sign test
2.Wilcoxon matched pairs signed- rank test
**Case IV:** k independent samples
**Tool:**
1. Median test
2. Kruskal- Wallis one way ANOVA by ranks

**Case V**: k related samples
**Tool:**
Friedman two way ANOVA by ranks

In the social and behavioural sciences, characteristics are measured on an ordinal level. We often ask if people "Strongly disagree" till "Slightly Disagree". We then assign a value of '1' if they strongly disagree with a statement, up to a '5' if they strongly agree with a statement. This type of measurement is ordinal as "Strongly Agree" reflects more agreement than "Slightly Agree". This type of measurement is not an interval or a ratio level of measurement, because we cannot state with certainty that the interval between "Strongly Disagree" and "Slightly Disagree" is equivalent to the interval between "Slightly Disagree" and "Neutral". Also, we say that there is an absolute zero point for level of agreement. But it has been found that for the many studies, it is not wrong to treat ordinal data (such as variables which have been measured using Strongly Disagree to Strongly Agree response alternatives) as interval level data, and conduct statistical tests that are appropriate for interval level data.

**SOME CASE STUDIES:**

**Case Study 1:**

In an industrial production line, items are inspected periodically for defectives. The following is a sequence of defective items, D, and non-defective items, N, produced by the production line: D D N N N D N N D D N N N N N D D D N N D N N N N D N D. Use the large sample theory for the runs-test, with a significance level of 0.05, to determine whether the defectives are occurring at random.

**SOLUTION:**

The suitable tool is "One sample runs test for large sample" as the null hypothesis refers to the randomness of a single group of events.

Null Hypothesis H0: The defectives are occurring at random

Alternative Hypothesis H1: The defectives are not occurring at random

Test Statistic: $z = \dfrac{r-\mu_r}{\sigma_r} \sim N(0,1)$

Level of significance: $\alpha = 0.05$

Calculations:

$$\underset{1}{\underline{D\,D}}\ \underset{2}{\underline{N\,N\,N}}\ \underset{3}{\underline{D}}\ \underset{4}{\underline{N\,N}}\ \underset{5}{\underline{D\,D}}\ \underset{6}{\underline{N\,N\,N\,N\,N}}\ \underset{7}{\underline{D\,D\,D}}\ \underset{8}{\underline{N\,N}}\ \underset{9}{\underline{D}}\ \underset{10}{\underline{N\,N\,N\,N}}\ \underset{11}{\underline{D}}\ \underset{12}{\underline{N}}\ \underset{13}{\underline{D}}$$

r=13, $n_1$=11, $n_2$=17

Mean, $\mu_r = \dfrac{2n_1n_2}{n_1+n_2} + 1 = \dfrac{2(11)(17)}{11+17} + 1 = 14.36$

Variance, $\sigma_r{}^2 = \dfrac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$ =6.113

$$|z| = \left|\frac{r-\mu_r}{\sigma_r}\right| = \left|\frac{13-14.36}{\sqrt{6.113}}\right| = |-0.55| = 0.55$$

$z_{\alpha/2} = z_{0.025} = 1.96$

Critical region: $|z| \ge z_{\alpha/2}$

Since the obtained value of z (0.55) lies outside the critical region, we accept H0 and conclude that the defectives are occurring at random.

**Case Study 2:**

Below are the age distributions of two samples drawn from the same population. Test if they really are random samples drawn from the same population.

| Age | Sample 1 | Sample 2 | Combined | Cf |
|---|---|---|---|---|
| 15-19 | 5 | 10 | 15 | 15 |
| 20-24 | 10 | 13 | 23 | 38 |
| 25-29 | 20 | 25 | 45 | 83 |
| 30-34 | 13 | 10 | 23 | 106 |
| 35-39 | 2 | 2 | 4 | 110 |
| Total | n1=50 | n2=60 | n1+n2=110 | |

**SOLUTION:**

The suitable tool is "$\chi^2$ test with correction for continuity" as None of cell frequencies is less than 5 and the total sample size is appreciably large.

$$Mdn = l + \frac{\frac{n}{2} - F}{f_m} c = 24.5 + \frac{55-38}{45} \times 0.5 = 26.39$$

In sample l, the number of observations that are ≤26.39 is $15 + \frac{26.39-24.5}{5} \times 0.20 = 22.55 \leq 23$

The same in sample 2 is $23 + \frac{1.89}{5} \times 0.25 = 32.45 \geq 32$

With this information we construct a 2×2 contingency table with median split.

| | Sample 1 | Sample 2 | Total |
|---|---|---|---|
| Above median | 27(A) | 28(B) | 55(A+B) |
| Below median | 23(C) | 32(D) | 55(C+D) |
| Total | 50(A+C) | 60(B+D) | 110 |

$H_0$: Proportion of individuals in cells A and B (or C and D) is the same, against the alternative hypothesis $H_1$: they are different.

Since none of cell frequencies is less than 5 and the total sample size is appreciably large, the appropriate test for testing the significance of difference in cell frequencies is " Chi-square with correction for continuity". It will be a two-tailed test and the level of significance selected is 0.05

Computations:

$$\chi^2 = \frac{n(|AD - BC| - n/2)^2}{(A+B)(B+D)(A+C)(C+D)}$$

$$= \frac{110(|264-644|-55)^2}{55 \times 50 \times 55 \times 60} = 0.33$$

Critical region $\chi^2 \geq \chi^2_{0.05,1}$

Tabulated value $\chi^2_{0.05,1}$ =3.84

Since obtained value $\chi^2$ is far less than the tabulated value 3.84, we accept $H_0$ and conclude that the two samples are drawn from the same population.

**Case study 3:**

Given below are the mileages (in thousands of miles) of two samples of automobile tyres of different brands, say I and II, before they wear out.

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tyre I | 34 | 32 | 37 | 35 | 42 | 43 | 47 | 58 | 59 | 62 | 69 | 71 | 78 | 84 |
| Tyre II | 39 | 48 | 54 | 65 | 70 | 76 | 87 | 90 | 111 | 118 | 126 | 127 | | |

Test whether the tyre II gives more median mileage then tyre I. Use $\alpha$ = 0.05

**SOLUTION:**

The suitable Tool is " Mann Whitney test" as we are comparing two samples which are independently drawn

Null hypothesis H0: The median mileages of the two brands of tyres are the same.

Alternative hypothesis H1: The median mileage of tyre I is less than that of tyre II.

Test statistic: $z = \dfrac{U - \mu_u}{\sigma_u} \sim N(0,1)$

Level of significance, $\alpha = 0.05$

Calculations:

| Tyre I mileages | Ranks | Tyre II mileages | Ranks |
|---|---|---|---|
| 34 | 2 | 39 | 5 |
| 32 | 1 | 48 | 9 |
| 37 | 4 | 54 | 10 |
| 35 | 3 | 65 | 14 |
| 42 | 6 | 70 | 16 |
| 43 | 7 | 76 | 18 |
| 47 | 8 | 87 | 21 |
| 58 | 11 | 90 | 22 |
| 59 | 12 | 111 | 23 |
| 62 | 13 | 118 | 24 |
| 69 | 15 | 126 | 25 |
| 71 | 17 | 127 | 26 |
| 78 | 19 | | |
| 84 | 20 | | |
| | R1=138 | | |

$R_1=138$, $n_1=14$, $n_2=12$

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 14(12) + \frac{14(14+1)}{2} - 138 = 135$$

$$\mu_u = \frac{n_1 n_2}{2} = 84, \quad \sigma_u = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = 19.44$$

$$z = \frac{U - \mu_u}{\sigma_u} = \frac{135 - 84}{19.44} = 2.623$$

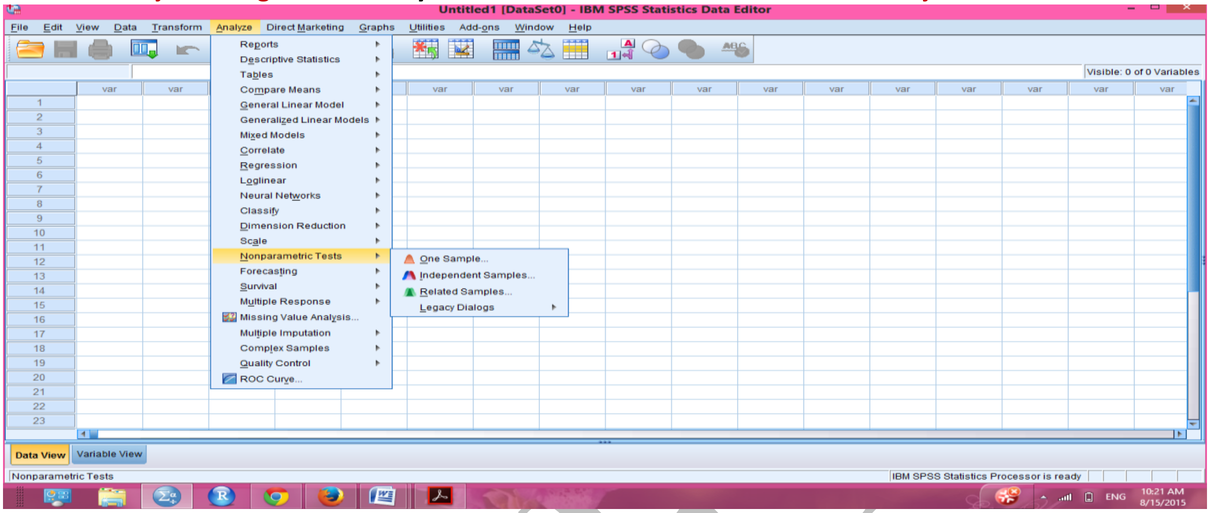
$z_\alpha = 2.58$

Critical region $z \geq z_\alpha$

Since the obtained value of z (2.623) lies in the critical region, we reject $H_0$.

Conclusion: The median mileage of Tyre I is less than that of tyre II.

**NON PARAMETRIC TESTS USING SPSS:**




## CONCLUSION:

Nonparametric statistical tests when the data are nominal or ordinal , instead of interval or ratio and when data are not normally distributed, or have heterogeneous variance in spite of having interval or ratio scale. Non parametric statistics are particularly useful with variables like. ranks or frequencies and such situations occur in qualitative and descriptive studies and not so frequently in experimental studies.

## ACKNOWLEDGMENT:

I acknowledge the support of Centre for Research, Christ University for preparation of monograph " Applications of Statistical Tools in Social Sciences Research" from which this paper is derived.

## REFERENCES:

1. B Y Brase, C. H. (n.d.). Understanding basic statistics .
2. Healey, J. (n.d.). Statistics- A tool for Social Research . Newport University.
3. Linneman, T. J. (n.d.). Social Statistics- Basics and Beyond .
4. P.K.Majumdar. (n.d.). Statistics- A tool for Social Sciences .
5. Utts, J. H. (n.d.). Mind on Statistics.
6. http://onlinestatbook.com/2/introduction/ levels_of_measurement.html