NEERAJ®

M.P.C.-6
Statistics in Psychology

Chapter Wise Reference Book
Including Many Solved Sample Papers

Based on

I.G.N.O.U.

& Various Central, State & Other Open Universities

By: *Renu Gupta*

NEERAJ PUBLICATIONS
(Publishers of Educational Books)

Mob.: 8510009872, 8510009878 E-mail: info@neerajbooks.com
Website: www.neerajbooks.com

MRP ₹ 320/-

Content

STATISTICS IN PSYCHOLOGY

S.No.	*Chapterwise Reference Book*	*Page*

INTRODUCTION TO STATISTICS

CORRELATION AND REGRESSION

NORMAL DISTRIBUTION

NON-PARAMETRIC STATISTICS

■ ■

Sample Preview of the Solved Sample Question Papers

Published by:

NEERAJ PUBLICATIONS

www.neerajbooks.com

QUESTION PAPER

June – 2024

(Solved)

STATISTICS IN PSYCHOLOGY

M.P.C.-6

Time: 2 Hours] ***[Maximum Marks: 50***

Note: *All sections are compulsory. Use of simple calculator is permitted.*

SECTION–A

Note: Answer the following questions:

Q. 1. Discuss the assumptions of parametric and non-parametric statistics.

Ans. Ref.: See Chapter-1, Page No. 2, 'Assumptions of Parametric', 'Non-Parametric Statistics'.

Q. 2. Explain the concept of normal curve with the help of a diagram. Describe the characteristics of normal probability curve.

Ans. Ref: See Chapter-9, Page No. 65, 'Concept of Normal Curve', Page No. 66, 'Characteristics or Properties of Normal Probability Curve (NPC)'.

Q. 3. Compute ANOVA for the following data:

Group A	**6**	**7**	**3**	**4**	**2**	**4**	**6**	**7**	**8**	**1**
Group B	**4**	**7**	**3**	**4**	**7**	**3**	**2**	**7**	**4**	**2**
Group C	**3**	**6**	**2**	**6**	**8**	**6**	**3**	**9**	**6**	**3**

Critical Value : 3.35 at 0.05 level

5.49 at 0.01 level

Ans. The data provided is for three groups (A, B, and C). We need to compute ANOVA, which involves comparing the means of these three groups to determine if there are any statistically significant differences between them.

Step 1: State the Hypotheses.

Null Hypothesis (H_0): There is no significant difference between the means of the groups.

Alternative Hypothesis (H_a): There is a significant difference between the means of the groups.

Step 2: Organize the Data.

The given data is:

- Group A: 6, 7, 3, 4, 2, 4, 6, 7, 8, 1
- Group B: 4, 7, 3, 4, 7, 3, 2, 7, 4, 2
- Group C: 3, 6, 2, 6, 8, 6, 3, 9, 6, 3

Step 3: Calculate the Means for Each Group.

We calculate the mean for each group as follows:

Mean of Group A=

$$\frac{6+7+3+4+2+4+6+7+8+1}{10} = \frac{48}{10} = 4.8$$

Mean of Group B =

$$\frac{4+7+3+4+7+3+2+7+4+2}{10} = \frac{43}{10} = 4.3$$

Mean of Grup C =

$$\frac{3+6+2+6+8+6+3+9+6+3}{10} = \frac{52}{10} = 5.2$$

Step 4: Compute the Overall Mean.

The overall mean is calculated by taking the mean of all the observations across the groups:

$$\text{Overall Mean} = \frac{48+43+52}{30} = \frac{143}{30} = 4.77$$

Step 5: Calculate the Sum of Squares.

We now need to calculate the Sum of Squares Between Groups (SSB) and Sum of Squares Within Groups (SSW).

1. Sum of Squares Between Groups (SSB)

The formula for SSB is:

$$SSB = n \sum(\bar{X}_t - \bar{X})^2$$

Where *(n)* is the number of observations in each group, ($\bar{X}_t$) is the mean of each group, and ($\bar{X}$) is the overall mean.

$$SSB = 10[(4.8 - 4.77)^2 + (4.3 - 4.77)^2 + (5.2 - 4.77)^2]$$

$$SSB = 10[0.0009 + 0.2209 + 0.1849] = 10 \times 0.4067 = 4.067$$

2. Sum of Squares Within Groups (SSW)

The formula for SSW is:

$$SSW = \sum(X_{ij} - \bar{X}_t)^2$$

Where (X_{ij}) is each individual observation, and \(\bar{X}_{i}\) is the group mean.

For Group A:

$SSW_A = (6 - 4.8)^2 + (7 - 4.8)^2 + ... + (1-4.8)^2 = 22.96$

For Group B:

$SSW_B = (4 - 4.3)^2 + (7 - 4.3)^2 + ... + (2 -4.3)^2 = 23.1$

For Group C:

$SSW_C = (3 - 5.2)^2 + (6 - 5.2)^2 + ...+ (3 - 5.2)^2 = 33.6$

$SSW = SSW_A + SSW_B + SSW_C = 22.96 + 23.1 + 33.6 = 79.66$

Step 6: Calculate the Degrees of Freedom.

The degrees of freedom for between groups and within groups are calculated as follows:

- *df* between (*df*B) = number of groups – 1 = 3 – 1 = 2
- *df* within (*df*W) = total number of observations - number of groups = 30 – 3 = 27

Step 7: Compute the Mean Squares.

The mean square for between groups and within groups is calculated by dividing the sum of squares by the corresponding degrees of freedom:

- MSB = SSB / dfB = 4.067 / 2 = 2.0335
- MSW = SSW / dfW = 79.66 / 27 = 2.95

Step 8: Compute the F-Ratio.

The F-ratio is the ratio of the mean square between groups to the mean square within groups:

$$F = \frac{MSB}{MSW} = \frac{2.0335}{2.95} = 0.689$$

Step 9: Compare with the Critical Value.

The critical value at the 0.05 significance level with *df*B = 2 and *df*W = 27 is given as 3.35. Since our calculated F-value (0.689) is less than the critical value (3.35), we fail to reject the null hypothesis.

Q. 4. Compute Chi-square for the following data:

	Responses			
Categories	**Never**	**Rarely**	**Sometimes**	**Always**
Group A	**10**	**12**	**8**	**10**
Group B	**6**	**4**	**7**	**3**
Group C	**11**	**9**	**4**	**6**
Group D	**7**	**3**	**4**	**6**

Ans. To compute the Chi-square for the following data:

Responses	Never	Rarely	Sometimes	Always
Group A	10	12	8	10
Group B	6	4	7	3
Group C	11	9	4	6
Group D	7	3	4	6

The steps for calculating Chi-square (χ^2) are as follows:

Step 1: State the hypotheses.

- **Null hypothesis (H_0):** The distribution of responses is independent of the groups.
- **Alternative hypothesis (H_1):** The distribution of responses is dependent on the groups.

Step 2: Compute the expected frequencies.

The expected frequency for each cell in a contingency table is calculated as:

$$E = \frac{(\text{Row Total}) \times (\text{Column Total})}{\text{Grand Total}}$$

Let's calculate the totals first:

- Row totals:
- Group A: (10 + 12 + 8 + 10 = 40)
- Group B: (6 + 4 + 7 + 3 = 20)
- Group C: (11 + 9 + 4 + 6 = 30)
- Group D: (7 + 3 + 4 + 6 = 20)
- Column totals:
- Never: (10 + 6 + 11 + 7 = 34)
- Rarely: (12 + 4 + 9 + 3 = 28)
- Sometimes: (8 + 7 + 4 + 4 = 23)
- Always: (10 + 3 + 6 + 6 = 25)
- Grand total: (40 + 20 + 30 + 20 = 110)

Step 3: Calculate expected frequencies for each cell.

Using the formula mentioned earlier, we calculate the expected frequencies for each cell. For example:

- Expected frequency for Group A, Never:

$$E = \frac{(40 \times 34)}{110} = 12.36$$

- Expected frequency for Group A, Rarely:

$$E = \frac{(40 \times 28)}{110} = 10.18$$

(We would do similar calculations for all other cells.)

Step 4: Compute the Chi-square statistic.

The formula for the Chi-square statistic is:

$$\chi^2 = \sum \frac{(0 - E)^2}{E}$$

Sample Preview of The Chapter

Published by:

NEERAJ PUBLICATIONS

www.neerajbooks.com

STATISTICS IN PSYCHOLOGY

Block-1: Introduction to Statistics

Parametric and Non-Parametric Statistics

1

INTRODUCTION

We discuss various aspects of parametric and non-parametric statistics here. In a parametric statistical test certain conditions like normally distributed data, etc. is specified. In non-parametric statistics no such condition is required. They are known as distribution free tests.

In this chapter various descriptive statistical measures like centres tendency, variability, normal probability curve, etc. will be discussed. Concepts and assumptions of parametric tests will be explain. Assumptions and applications of analysis of variance and co-variance for testing the significance of difference between means of three and more samples will also be discussed. The Distribution free (non-parametric) tests will be discussed including – Chi-Square Test, Median Test, Man-Whitney U Test, Sign Test and Wil Xoyon-Matched Pairs and Signed Ranks Test.

CHAPTER AT A GLANCE

DEFINITION OF PARAMETRIC AND NON-PARAMETRIC STATISTICS

Non-parametric statistics cover techniques which do not rely on data belonging to any particular distribution. They have:

1. Distribution Free Methods: They are not from normally distributed population and consist of statistical models, inference and statistical tests.

2. Non-parametric Statistics: Rank of observations are used.

3. No Assumption of a Structure of a Model: Techniques do not assume that the structure of a model is fixed. The individual variables are typically assumed to be from parametric distributions and assumptions of types of connections among variables are also made. These techniques also have:

(a) Non-parametric Regression

(b) Non-parametric Hierarchical Bayesian Models.

The structure of relationship is treated non-parametrically in non-parametric regression. Based on Dirichlet Process which permits the number of latent variables to grow as necessary to fit the data in regard to Bayesian Models.

4. Classical and Standard Tests. The assumptions are not applied.

Parametric Tests

Like in student's *t*-test, parametric tests normally involve data expressed in absolute numbers or values and not ranks. These tests operate under certain conditions, which are assumed. If assumptions are valid then, test will be meaningful. Apart from assumption of normally distribution, it is assumed that scene being analysed comes from at least one interval scale.

Population is the entire group that researcher intends to study and understand *t*, *z*, *f* are parametric statistical tests.

T-tests: The text is used to find if the scores of two groups differ on a variable.

t-test: It is used to *t*-test for differences in mean scores. *t*-tests and *f*-test are quite robus even when some assumptions are not met. Parametric tests have assumptions on population and type of scale.

1. Nominal data or classified data for nominal scale is used.
2. For interval data, ordinal data is used.
3. A constant scale without natural zero, the interval scale deals with ordered data with interval.

4. Ratio scale employed for ordered constant scale with a natural zero e.g., for height, weight, age, lengths, etc.

In case of small number of items in sample, we use non-parametric statistical tests, as there is no normal distribution. How to decide it? Take small number when $N \leq 30$ and for more the parametric test may be used.

Non-parametric statistical test can be used for nominal data with ordinal data.

Assumptions are of independent, continuity of variable under study for non-parametric test and are fewer and weaken than in case of parametric test; they are used for small sample sizes and non bounded by any assumptions, are user friendly and less time is required to use them. However, they must be used only with data based upon ratio and interval measurements. They are less precise and have less power than the parametric tests.

ASSUMPTIONS OF PARAMETRIC AND NON-PARAMETRIC STATISTICS

Assumptions of Parametric Statistics: Sometimes assumptions and conditions cannot be meet so parametric procedures cannot be used. Hence, only option is apply on non-parametric statistics.

If we have samples of small size, distribution is not normally, distributed and/or sample is in nominal or ordinal scale forms, then non-parametric test is good. Hence use it if–

- Sample size is quite small; say 5 or 6 then only non-parametric test can be used.
- Normality of distribution is doubtful or is not normal.
- Data is in either in form or ordinal or nominal scale.
- Data can be in + or – or good-bad etc. form
- Data can be expressed in form of ranks.
- Variables are expressed in nominal form.

Assumptions of Non-parametric Statistics

Non-Parametric Statistics is based on the ranks of observations and do not depend upon any distribution of the population non-parametric statistics is applied in the situation where we cannot meet the assumptions and conditions. The sample with small number of items are treated with non-parametric statistics. It is based on the model that specifies only very general conditions. In case of counted or ranked data we made use of non-parametric statistics. Non-parametric statistics are more user friendly than parametric statistics.

ADVANTAGES OF NON-PARAMETRIC STATISTICS

In case sample size is small, it is only useful test and in case of fewer assumptions too it is more appropriate for research investigation. They are easier to learn and apply. Also their interpretation is often more direct.

DISADVANTAGES OF NON-PARAMETRIC STATISTICAL TESTS

If all assumptions of parametric are met then the non-parametric test is wasteful which (wastefulness) can be expressed in percentage of its power efficacy.

Next they may have widely scrattered different formats.

PARAMETRIC STATISTICAL TESTS FOR DIFFERENT SAMPLES

Sampling error in a sample should not vary from sample to sample but should usually deviate from the population mean. The means of these samples can be then put in a frequency distribution, known as sampling distribution of means.

Central Limit Theorem which describes the characteristics of sample means. For it select large number of equal sized samples, $N \geq 3$ in size for an infinite population. The means of samples are normally distributed. Take the average value of sample means.

= Mean of population

Standard deviation SE_M or σ_m of mean

= Standard error of mean

This is the clue to how such sample mean may be expected to deviate from the population mean.

Then standard error of mean in a large sample.

$$SE_M \text{ or } \sigma_m = \frac{\sigma}{\sqrt{N}}$$

Where, σ = Standard deviation of the population

N = Size of sample.

For small samples, sampling distribution is not normal. They are called *t*-distribution.

t-distribution lies below normal.

PARMETRIC STATISTICAL MEASURES FOR CALCULATING DIFFERENCE BETWEEN MEANS

To determine true difference between population parameters of two samples, like means, standard deviation, correlations, etc. techniques will be needed. For example, if mean of males in M.A. Psychology sample be 55 and for females 59. Then difference 4 is too large to be due to sampling error or chance.

To test the significance of this difference 4, first find standard error of difference of two means because it is reasonable to express the difference between two means will be subject to sampling errors. Then we can determine whether difference exists between the population means by the difference between sample means and its standard error.

Significance of Difference between the Means of Two Independent Large and Small Samples

Means are called uncorrelated or independent if computed for samples drawn at random from totally different and unrelated groups.

Large Samples: As frequency distribution of large sample means taken from same population are normally distributed around M_{PP}, the population mean, it can reasonably be expected that the frequency distribution of difference of means of samples drawn from two different populations will also tend to be normal with mean = 0 and standard deviation, known as the standard error of difference of means.

Let σ_{dm} = standard error, calculated from standard errors of sample means σ_{m_1} σ_{m_2} of two sample means then

$$\sigma_{dm} = \sqrt{\sigma_m + \sigma_m}$$

where σ_{m_1} = S.E. of mean of first sample

σ_{m_2} = S.E. of mean of second sample

for N_1, N_2 the number of cases in first, second cases respectively.

Significance of the Differences between the Means of Two Dependent Samples

Correlation or dependent means are obtained from the scores of the same test put to the same sample on two ocassions or same test administered to (put to) equivalent samples in which members of group have been matched for item to item by one or more attributes.

$$T = \frac{M_1 + M_2}{\sigma M_1^2}$$

Where

M_1 = mean

Significance of the difference between the Means of Three or More Samples.

F test is used for testing the significance between the means of three or more samples. It involves the use of analysis of variance on analysis of co-variance. This analysis is based on certain assumptions.

1. The population distribution should be normal.
2. All the groups should be randomly chosen from the sub-population having the same criteria that is the sub-group under investigation should have the same variability. For this F_{max} test is applied

F_{max} = Largest variance/Smallest variance

Variances are said to be homogenous or in significant if F_{max} < appropriate F critical value.

PARAMETRIC STATISTICS MEASURES RELATED TO PEARSON'S '*r*'

Pearson's co-efficeint of correlation is complicated mathematical basis because it has a nature of sampling distribution. The sampling distribution of *r* is not normal for all but when population *r* is near zero and size of ($N \geq 30$) sample is large.

For $r \cong 80$ or more and N < 30, the sample distribution of *r* is skewed. And the same is held for low $r \cong 0.20$ or less.

Hence for very low or very high *r*, we convert it to Fighter's co-efficient Z using conversion table and find the standard error of Z, which is normal regardless of size of N and size of population *r*. The formula of S.E. of Z of size N is:

$$\sigma_2 = \text{S.E.}_2 = \frac{1}{\sqrt{N-3}}$$

To determine the S.E. diffz of difference between Pearson's coefficient of correlation of two samples, we use

$$SE_{diffz} = \sigma_{z_1 - z_2}$$

$$= \sqrt{\frac{1}{N_1 - 3} + \frac{1}{\sqrt{N_2 - 3}}}$$

N_1, N_2 are sizes of two samples.

To test the significance of difference between the two Z's is CR given by:

$$CR = \frac{Z_1 - Z_2}{SE_{diffz}}$$

Non-Parametric Tests Used for Inference

The following three tests are most frequently used in case of unrelated or independent samples:

1. Chi-square Test
2. Median Test
3. Mann-Whitney 'U' Test

The Chi-Square $(X)^2$ Test

This is applied to discrete data only, i.e., to the data that are counted rather than measured. It is to test independence, i.e., to estimate the likelihood that some factor and not the chance accounts for observed

relationship. X^2 does not measure the degree of relationship and merely evaluates the probability that observed relationship occurs from chance (or not).

The formula is:

$$X^2 = \sum\left[\frac{fo - fe)^2}{fe}\right]$$

Where, *fo* = frequency of occurrence of observed or experimental determined facts

fe = expected frequency of occurrence

The Median Test

To check whether the two independent samples differ in central tendency, we use the Median Test. The test gives information regarding the possibility of same median of two independent samples drawn from populations. If samples are on ordinal scale, even then it is useful. Combined medians of both samples is obtained then both sets of scores at the combined median are dichotomized and data put into 2 × 2 matrix, in rows, putting median, first below and then above medians and in the two columns, their corresponding samples, 1 and 2 respectively.

The Mann-Whitney U-Test

It is more useful than median test. It can be a better alternative to the Parametric *t*-test when parametric conditions are not met and measurements are expressed in ordinal scale values.

SOME NON-PARAMETRIC TESTS FOR RELATED SAMPLES

Several tests are used in drawing statistical inferences in case of related samples. We study sign test, and Wilcoxon Matched-Pairs Signed-Ranks Test here.

The Sign Test

It is the simplest test of significance is the category of non-parametric tests. Here '+' and '–' signs are used and is particularly useful when quantitative measurement are impossible or inconvenient, but ranks can be obtained of the two members of each pair.

If researcher wants to establish that two conditions of a single sample are different, then it is quite useful. Here only assumption is that the variable under investigation has a continuous distribution and no assumption of form of distribution is made.

The Wilcoxon Matched Pairs Signed Ranks Test

It is more powerful than the sign test. It test not only direction but also magnitude of difference within pairs of matched groups. It also deals with only dependent groups. Here, null hypothesis would assume that the magnitude and direction of pair difference would be about the same.

UNIT END QUESTIONS

Q. 1. Define parametric statistics.

Ans. Parametric statistics test operate under certain conditions, i.e., data is expressed in absolute numbers or values, normally distributed and assumes that the scene being analysed results from measurement in atleast an interval scale.

Q. 2. Discuss non-parametric statistics.

Ans. Non-parametric covers techniques that do not result from any particular distribution and hence they have distribution free methods. Mostly, we can have no condition imposed and ranks may be used.

Q. 3. Write various assumptions of parametric statistics.

Ans. Assumptions of parametric statistics are:

- The population from which the sample has been drawn are normally distributed.
- The variables involved must have been measured or ratio scale. Sample size are large.

Q. 4. What are the advantages of non-parametric statistics?

Ans. They are easier to learn and apply and their interpretation is often more direct. Sample size is small and have a fewer assumptions.

Q. 5. Differentiate between parametric and non-parametric statistics?

Ans. Parametric statistics deals with a few assumptions like they are drawn from a normal distribution with large sample sizes. In non-parametric statistics, sample sizes are small and data is not drawn from a normal distribution.

Parametric tests are useful as they are quite robust for testing significance or trustworthiness of the computed sample trustworthiness.

Q. 6. List the assumptions on which the use of Parametric Tests is base.

Ans. Assumptions of parametric test are based on nature of the population of distribution and on the way the type of scale is used to quantify the data measures. Parametric test based on normal distribution also assumes that the scene being analysed results from measurements in at least one interval scale.

Q. 7. Describe the characteristics of Central Limit Theorem.

Ans. Central Limit Theorem describes the characteristics of sample mean. According to this