Non-Parametric Tests

Dr. Chanchal Kundu

Professor

Department of Mathematical Sciences

Rajiv Gandhi Institute of Petroleum Technology

Jais 229304, U.P., India

Why Use Nonparametric Tests?

Parametric Tests

Definition: Statistical technique to test a hypothesis based on some restrictive assumptions about the population.

- Parametric hypothesis tests require the estimation of one or more unknown parameters (e.g., population mean or variance).
- Often, unrealistic assumptions are made about the normality of the underlying population.
- Large sample sizes are often required to invoke the Central Limit Theorem.
- Parametric tests require quantitative measurement of the sample data in the form of an interval or ratio scale

Why Use Nonparametric Tests?

Nonparametric Tests

Definition: Non-parametric tests are nor dependent upon the restrictive normality assumption of the population.

- *Nonparametric* or distribution-free tests
  - usually focus on the sign or rank of the data rather than the exact numerical value.
  - do not specify the shape of the parent population.
  - can often be used in smaller samples.
  - can be used for ordinal data.

Advantages and Disadvantages of Nonparametric Tests

Advantages:

- *Non-parametric tests can be used to analyze nominal as well as ordinal level of data.*
- *When sample size is small, non-parametric tests are easy to compute.*
- *Non-parametric tests are not based on the restrictive normality assumption of the population or any other specific shape of the population.*

Disadvantages:

- *When all the assumptions of parametric tests are met, non-parametric tests should not be applied.*
- *When compare to parametric tests, availability and applicability of non-parametric tests are limited.*
- *When sample size is large, non-parametric tests are not easy to compute.*

Chi-Square Goodness-of-fit Test:

- To decide whether a particular probability distribution e.g., binomial, poisson, normal etc. is the appropriate distribution or, whether an observed sample fits an assumed population distribution.

The hypothesis are

H_0 : The given set of observations follows the assumed distribution.

H_1 : The given set of observations does not follow the assumed distribution.

The test statistic is $\chi^2_{cal} = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e}$ where

f_0 : observed frequency and f_e : expected frequency.

Example:

The post office is interested in modeling the mangled-letter problem. It has been suggested that any letter sent to a certain area has a 0.15 chance of being mangled. Because the post office is so big, it can be assumed that two letters' chances of being mangled are independent. A sample of 310 people was selected and two test letters were mailed to each of them. The number of people receiving zero, one or two mangled letters was 260, 40, and 10, respectively. At 0.1 level of sig., is it reasonable to conclude that the number of mangled letters received by people follows a binomial distribution with $p=0.15$?

Example: The sales manager of a cement company feels that the quarterly demand for cement in tons follows uniform distribution. The observed frequencies of demand values are summarized in the following table:

Demand (in tons)	30	31	32	33	34	35	36	37	38	39
Observe Freq.	2	3	1	4	2	2	4	2	4	1

Check whether the given data follow uniform distribution using Chi-square goodness-of-fit test at a sig. level of 0.05.

Thank You