ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology
Volume 1, Issue 3, May 2012

SIGNIFICANCE OF STATISTICS IN HEALTH SCIENCES IN RURAL AREA

Reetu Malhotra [1] , Vandana Singh [2] & Dr. Rajesh Kumar [3]

[1] Department of Applied Sciences, Chitkara University, Rajpura, Punjab –India

[2] Chitkara School of Health Sciences, Chitkara University, Rajpura, Punjab –India

[3] Rural Medical Officer, Patiala, Punjab –India

***Abstract*— The present article is aimed at significance of statistics in growing health science industries. Health and statistics are essential tools in demography; health care, medical profession and community services. Statistics plays very significant role to describe what is normal and healthy in population and to find limits of normality in variables, such as weight and pulse rate, sex and haemoglobin, age and menstruation cycle etc. Statistics in health sciences will articulates leading causes of death, sickness, whether particular diseases is rising or falling in severity and prevalence. Statistical data collected from measurements or surveillance that defines the characteristics of specific population samples. Descriptive statistics précis the utility, efficacy and expenses of medical goods and facilities. Progressively, health care organizations employ statistical analysis to measure their performance outcomes. This type of study is important in finding the correlation between two variables, comparison in the action of different drugs, to find an association between two attributes, in field of Anatomy & Physiology, Pharmacology, Medicines, and Public health & Community medicines respectively. By collecting the waiting times of five different patients chosen at random from Govt dispensary Mallewal, district Patiala, Govt dispensary Dhablan, district Patiala , Govt dispensary Todarpur, district Patiala Govt dispensary Bahadurgarh, district Patiala(treated as out-patient clinics A, B, C, and D)to determine if there are differences between the dispensaries using Kruskal-Wallis Application (statistics) one can know how statistics is helpful in rural area.**

***Keywords*— Health sciences, Statistics , demography, Anatomy & Physiology, Pharmacology, Medicines, and Public health & Community medicines**

I. INTRODUCTION

At the outset, it might be useful to point out to different meanings attached to the term 'statistics'. Statistics is the term use to indicate facts and figures of any kind: business statistics, health statistics, and vital statistics. It is a field of study concerned with techniques or methods of collection of data, classification, summarizing, interpretation, drawing inferences, test of hypothesis, etc. when only a part of data is used. The statistics plays an important role in health sciences. Everything in medicine be it research, diagnosis or treatment, depends on counting or measurement. A glance at any medical journal shows the extensive use of statistical methods in the collection, evaluation and presentation of biological data. There is a reflection of a need for precise quantitative assessment of different type of phenomena which are encountered in this field. Research workers, students, healthcare professionals in this field will need to study the application of statistics so they can design and evaluate their own work. A combination of biology and statistics referred as biometry or biometrics. Metric is derived from the word measure. Knowledge of the logical basis of the statistical approach and rationale of the commonly used statistical techniques will be an essential perquisite.

In the article, we have presented the application statistical methods used in different fields such as Anatomy & Physiology, Pharmacology, Medicines, and Public health & Community medicines etc.

II. AREAS OF APPLICATIONS AND USES

The history of the development of modern and epidemiology are replete with examples of how pioneers made startling discoveries using simple statistical methods of designing experiments analysing the findings. One of the chief functions of statistics in the field of public health is to present numerical information that is relevant, reliable, comprehensive, analogous and up to date. Statistical methods are beginning to be integrated into medical informatics, public health informatics, bioinformatics and computational biology

- **Public health**, including epidemiology, health services research, nutrition, environmental health and healthcare policy & management.
- **Design and investigation** of clinical trials in medicine.
- **In Anatomy and Physiology**: Statistics helps
  - To define what is normal or healthy in a population and to find limits of normality in variables such as weight and pulse rate.

- To find the difference between means and proportions of normal at two places or in different periods.
- To find correlation between two variables such as height and weight -whether weight increases or decreases proportionately with height and if so by how much, has to be found.

• **Population genetics and statistical genetics** in order to link variant in genotype with a variation in phenotype. This has been used in farming to improve crops and animals .In biomedical research, this work can assist in finding candidates for gene alleles that can cause or influence susceptibility to disease in human heredities.

• **In Pharmacology***:*

- To compare the action of two different drugs or two successive dosages of the same drug.
- To find the relative potency of a new drug with respect to a standard drug.
- To find the action of drug- a drug is given to animals or humans to see whether the changes produces are due to the drug or by chance.

• Exploration of genomics data, for example from microarray or proteomics investigates the concern diseases and its stages.

• **In Medicine***:*

- To find an association between two attributes such as cancer and smoking or filariasis and social class- an appropriate test is applied for this purpose.
- To compare the efficacy of a particular drug , operation or line of treatment
- To identify signs and symptoms of a disease or syndrome. Cough in typhoid is found in almost found by chance and fever is found in almost every case.

• **Ecology, ecological predictions**

• **Biological sequence analysis**

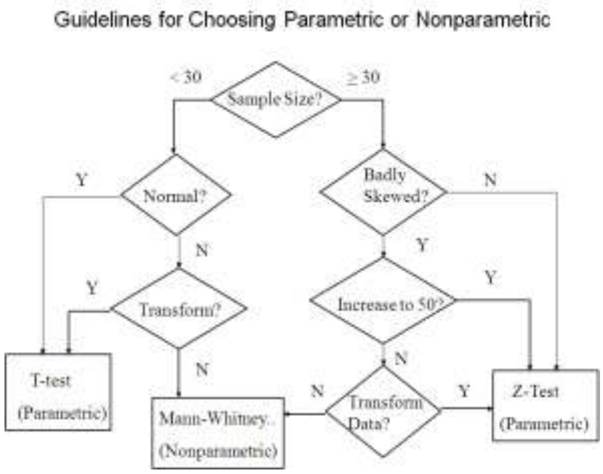
• **Systems biology** for gene network inference or pathways analysis.

• **In Community Medicine and public health**

- To test usefulness of sera and vaccines in the field – percentage of attacks or deaths among the vaccinated subjects is compared with that among the unvaccinated ones to find whether the difference observed is statistically significant.
- In epidemiological studies- the role of causative factors is statistically tested. Deficiency of iodine as an important cause of goitre in a community is confirmed only after comparing the incidence of goitre cases before and after giving iodized salt.
- Constantly newer and newer statistical ratios are being established to help measure the **mutable patterns** of various disease and health disorders. Addition of HIV health indicators, counting radiation doses, forensic criminal investigation ratios are amid the novel ones.
- **Electrocardiogram (ECG):** graphical representation of impulses of heart (cardiac system) on graph paper using statistics help in the diagnosis of various heart abnormalities.

III. IMPLEMENTATION OF STATISTICS

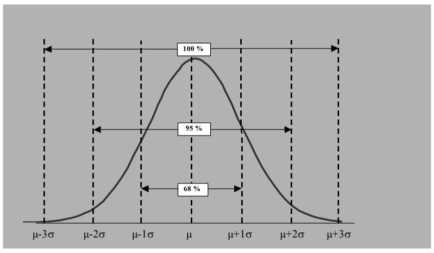
Statistics is an important element of Heath Sciences, Tests and their results in health sciences are very complex. With the help of statistics results can be easily interpreted in useful manner .Tests for differences in averages are either parametric or nonparametric. **Parametric tests** are considered more powerful in measuring differences in means and can be used to measure differences in the amount of dispersion. **Nonparametric tests** can be used to compare non-normal distributions .



• **Parametric tests**

- z test for comparing the averages of two large samples (n>30)
- t test for comparing the averages of two small samples.
- t test for matched pairs.
- Minimum requirements for these tests are the average, standard deviation and sample size. Individual tests are preferred in order to confirm normality and some statistical software packages require the individual data for input.

Parametric (Normal) Distribution



- **Nonparametric tests**
  - Man-Whitney U-test for two unmatched samples.
  - Kruskal-Wallis for more than two unmatched samples
  - Wilcoxon test for matched pairs.
  - Minimum requirements for these tests are the individual tests (for ranking).

IV. DESCRIPTION ABOUT VARIOUS TESTS

Exact test for goodness-of-fit: test fit of observed frequencies to expected frequencies used for small sample sizes (less than 1000) count the number of males and females in a small sample, test fit to expected 1:1 ratio

G-test for goodness-of-fit: test fit of observed frequencies to expected frequencies used for large sample sizes (greater than 1000) count the number of red, pink and white flowers in a genetic cross, test fit to expected 1:2:1 ratio

Chi-square test for goodness-of-fit: test fit of observed frequencies to expected frequencies used for large sample sizes (greater than 1000) count the number of red, pink and white flowers in a genetic cross, test fit to expected 1:2:1 ratio

Chi-square test for goodness-of-fit: test fit of observed frequencies to expected frequencies used for large sample sizes (greater than 1000) count the number of red, pink and white flowers in a genetic cross, test fit to expected 1:2:1 ratio

Randomization test for goodness-of-fit: test fit of observed frequencies to expected frequencies used for small sample sizes (less than 1000) with a large number of categories count the number of offspring in a thyroid genetic cross, test fit to expected 27:9:9:9:3:3:3:1 ratio

G-test of independence: test hypothesis that proportions are the same in different groups large sample sizes (greater than 1000) count the number of apoptotic vs. non-apoptotic cells in liver tissue of organic chemists, molecular biologists, and regular people, test the hypothesis that the proportions are the same

Chi-square test of independence: test hypothesis that proportions are the same in different groups large sample sizes (greater than 1000) count the number of apoptotic vs. non-apoptotic cells in liver tissue of organic chemists, molecular biologists, and regular people, test the hypothesis that the proportions are the same

Fisher's exact test: test hypothesis that proportions are the same in different groups used for small sample sizes (less than 1000)count the number of left-handed vs. right-handed grad students in Biology and Animal Science, test the hypothesis that the proportions are the same

Randomization test of independence: test hypothesis that proportions are the same in different groupsused for small sample sizes (less than 1000) and large numbers of categories count the number of cells in each stage of the cell cycle in two different tissues, test the hypothesis that the proportions are the same

Mantel-Haenzel test: test hypothesis that proportions are the same in repeated pairings of two groups count the number of left-handed vs. right-handed grad students in Biology and Animal Science at several universities, test the hypothesis that the proportions are the same; alternate hypothesis is a consistent direction of difference

Arithmetic mean: description of central tendency of data median description of central tendency of data more useful than mean for very skewed data median height of trees in forest, if most trees are short seedlings and the mean would be skewed by the few very tall trees

Range: description of dispersion of data used more in everyday life than in scientific statistics variance description of dispersion of data forms the basis of many statistical tests; in squared units, so not very understandable -

standard deviation: description of dispersion of data in same units as original data, so more understandable than variance

standard error of the mean :description of accuracy of an estimate of a mean confidence interval description of accuracy of an estimate of a mean

One-way ANOVA, model I: model test the hypothesis that the mean values of the continuous variable are the same in different groups model I the nominal variable is meaningful, differences among groups are interesting compare mean heavy metal content in mussels from Nova Scotia, Maine, Massachusetts, Connecticut, New York and New Jersey, to see whether there is variation in the level of pollution

One-way ANOVA, model II: estimate the proportion of variance in the continuous variable "explained" by the nominal variable model II: the nominal variable is somewhat arbitrary, partitioning variance is more interesting than determining which groups are different compare mean heavy metal content in mussels from five different families raised under common conditions, to see if there is heritable variation in heavy metal uptake

Sequential Dunn-Sidak method: after a significant one-way model I ANOVA, test the homogeneity of means of planned, non-orthogonal comparisons of groups compare mean heavy metal content in mussels from Nova Scotia+Maine vs. Massachusetts+Connecticut, also Nova Scotia vs. Massachusetts+Connecticut+New York

Gabriel's comparison intervals: after a significant one-way model I ANOVA, test for significant differences between all pairs of groups compare mean heavy metal content in mussels from Nova Scotia vs. Maine, Nova Scotia vs. Massachusetts, Maine vs. Massachusetts, etc.

Tukey-Kramer method: after a significant one-way model I **ANOVA,** test for significant differences between all pairs of groups compare mean heavy metal content in mussels from Nova Scotia vs. Maine, Nova Scotia vs. Massachusetts, Maine vs. Massachusetts, etc.

Bartlett's test: test the hypothesis that the variance of a continuous variable is the same in different groups usually used to see whether data fit one of the assumptions of an ANOVA

Nested ANOVA: test hypothesis that the mean values of the continuous variable are the same in different groups, when each group is divided into subgroups subgroups must be arbitrary (model II) compare mean heavy metal content in mussels from Nova Scotia, Maine, Massachusetts, Connecticut, New York and New Jersey; several mussels from each location, with several metal measurements from each mussel

Two-way ANOVA: test the hypothesis that different groups, classified two ways, have the same means of the continuous variable compare cholesterol levels in blood of male vegetarians, female vegetarians, male carnivores, and female carnivores

Paired t-test: test the hypothesis that the means of the continuous variable are the same in paired data compare the cholesterol level in blood of people before vs. after switching to a vegetarian diet

Linear regression: see whether variation in an independent variable causes some of the variation in a dependent variable; estimate the value of one unmeasured variable corresponding to a measured variable -measure chirping speed in crickets at different temperatures, test whether variation in temperature causes variation in chirping speed; or use the estimated relationship to estimate temperature from chirping speed when no thermometer is available

Correlation: see whether two variables covary measure salt intake and fat intake in different people's diets, to see if people who eat a lot of fat also eat a lot of salt

Multiple regression: fit an equation relating several X variables to a single Y variable measure air temperature, humidity, body mass, leg length, see how they relate to chirping speed in crickets

Polynomial regression: test the hypothesis that an equation with X2, X3, etc. fits the Y variable significantly better than a linear regression - -

Analysis of covariance: test the hypothesis that different groups have the same regression lines first step is to test the homogeneity of slopes; if they are not significantly different, the homogeneity of the Y-intercepts is tested measure chirping speed vs. temperature in four species of crickets, see if there is significant variation among the species in the slope or y-intercept of the relationships

Sign test: test randomness of direction of difference in paired data often used as a non-parametric alternative to a paired t-test compare the cholesterol level in blood of people before vs. after switching to a vegetarian diet, only record whether it is higher or lower after the switch

Kruskal–Wallis test: test the hypothesis that rankings are the same in different groups often used as a non-parametric alternative to one-way ANOVA 40 ears of corn (8 from each of 5 varieties) are judged for tastiness, and the mean rank is compared among varieties

Spearman rank correlation: see whether the ranks of two variables covary often used as a non-parametric alternative to regression or correlation 40 ears of corn (8 from each of 5 varieties) are judged for tastiness and prettiness, see whether prettier corn is also tastier.

V. ILLUSTRATED EXAMPLE & NUMERICAL RESULTS

Kruskal-Wallis Application: A healthcare worker & Dr. Rajesh coauthor collects the waiting times of five different patients chosen at random from Govt dispensaries(treated as a out-patient clinics) **Govt dispensary Mallewal, district Patiala, Govt dispensary Dhablan, district Patiala , Govt dispensary Todarpur, district Patiala Govt dispensary Bahadurgarh, district Patiala** (A, B, C, and D) to determine if there are differences between the clinics**.**

A	B	C	D
15	19	23	71
28	47	4	45
17	50	1	56
9	31	9	82
8	12	14	40

Kruskal-Wallis Results:

Kruskal-Wallis Test on Time Clinics

	N	Median	Ave Rank	Z
A	**5**	**15.000**	**7.301**	**-1.40**
B	**5**	**31.000**	**12.40**	**0.83**
C	**5**	**9.000**	**5.101**	**-2.36**
D	**5**	**56.000**	**17.201**	**2.92**
Overall	**20**		**10.5**	

H = 12.56 DF = 3 P = 0.006

H = 12.57 DF = 3 P = 0.006 (adjusted for ties)

In this way, using Statistics we can compare differences between the clinics.

VI. CONCLUSIONS

The approach is to equip medicos and the other users to the extent that they may be able to appreciate the utility and usefulness of statistics in medical and biosciences. The Kruskal-Wallis test helps in testing the hypothesis that several populations (dispensaries) have the same continuous distribution versus the alternative that measurements tend to be higher in one or more of the populations(dispensaries).

The article directly highlights the application of statistics in diagnosis, prognosis, prescription and management of diseases in individuals and community.

VII. ACKNOWLEDGMENT

The authors are thankful to **Rural Medical Officers** who help us in data collection. The authors are also thankful to readers /reviewers. The suggestions of reviewers for the improvement of this work are thankfully acknowledged.

Reetu Malhotra, is working as a Assistant Professor in Applied Sciences department & persuing her Ph.D degree from Chitkara University, Rajpura punjab, India.She had completed her M.Sc(Mathematics) from Guru Nanak Dev University, Amritsar with distinction and M.Phill(Mathematics) from Alagappa University, Karaikudi, UGC NET qualified, and having 8 years of experience in teaching.

Vandana Singh is working as a Assistant Professor in Chitkara School of Health Sciences in Chitkara University, Rajpura punjab, India having 6 years of experience in teaching.

Dr. Rajesh Kumar is Rural medical officer posted at Govt. Dispensary Mallewal. He has done his M.B.B.S from Govt. Medical College Patiala.

.

REFERENCES

[1] B.K.MAHAJAN,, the text book *methods in biostatistics* for medical students and research scholars sixth edition .
[2]Managing performance measurement data in *health care* - page 51
[3] Nelson, p. r. (1988), "application of the analysis ofmeans," sas users group international: proceedings of the thirteenth annual conference, 225-230
[4] American *journal* of public *health*: JPH: Volume 99, Issues 1-3; Volume 99, Issues 1-3
[5] *Journal* of the American *Statistical* Association: Volume 101, Pages 863-1751
[6] Ramig, P. R. (1983), "Applications of the Analysis ofMeans," Journal of Quality Technology , 15, 19-25.
[7] Al-Assaf, A. F. and Schmele, J. A. (1993), the textbook of Total Quality in Healthcare, Delray Beach, Florida: St. Lucie Press.
SAS Institute Inc.
SAS Institute Inc. (1990b), SAS/STAT User's Guide, Version
[8] Alwan, L. C. and Roberts, H. V.lsNelson, P. R. (1982), "Exact Critical Values for Use withthe Analysis of Means," Communications in Statistics, A11, 699-709
[9] .Ramig, P. R. (1983), "Applications of the Analysis ofMeans," Journal of Quality Technology , 15, 19-25.
[10] Benneyan, J. (1995), "Applications of Statistical Process Control (SPC) to Improve Health Care," conference presentation, Healthcare Information and Mangement SystemsSociety.
[11] Govt dispensary Mallewal, district Patiala, Govt dispensary dhavlan, district Patiala ,Govt dispensary todarpur, district Patiala Govt dispensary bahadurgarh district Patiala.