EXPERIMENTAL DESIGNS, STATISTICAL METHODS AND STANDARDIZATION OF AUDIOLOGICAL TESTS

An Independent Project in Audiology

Presented to the University of Mysore

In Partial Fulfilment for the M.Sc. III Semester

Examination in Speech and Hearing –December

1980

CERTIFICATE

This is to certify that the independent project in Audiology entitled " Experimental designs, Statistical Methods and Standardization of Audiological Tests " is the bona fide work, in partial fulfillment for the M.Sc III Semester Examination, carrying 50 marks, of the student with Register No. 14

Director-in-charge,
All India Institute of Speech & Hearing
Mysore-6.

CERTIFICATE

This is to certify that this independent project in Audiology has been prepared under my supervision and guidance.

GUIDE

DECLARATION

This Independent Project in Audiology is the result of my own study undertaken under the guidance of Mr. Jesudas Dayalan Samuel, Lecturer in audiology, All India Institute of Speech and Hearing, and has not been submitted earlier at any University for any other Diploma or Degree.

Mysore

December, 1980. Register No. 14

ACKNOWLEDGEMENT

The author is grateful to the invaluable guidance of Mr. Jesudas Dayalan Samuel, Lecturer in Audiology, All India Institute of Speech and Hearing, Mysore.

He is thankful to Dr. N. Rathna and Dr. S. Nikam.

He acknowledges the following who have rendered their help directly or indirectly in finishing this project.

Venkatesh, C.S.	III B.Sc. Student
Ashok Kumar, A.A.	His classmate
Ashok Kumar Sinha	His classmate.

TABLE OF CONTENTS

CHAPTER I

I. AUDIOLOGY AS A SCIENCE

Science can be defined as an organized structural knowledge, involving research and possibility of predicitions based on actual data. Natural sciences deal with physical universe and some aspects of living organism. Phenomenon is the basic element of sciences. A phenomenon is any observable event, involving variable. Variables of a phenomenon are those quantities that vary their magnitude during the production of such phenomenon. There are variables that can be quantifiedand some variables are not. Sciences that deal with quantitative variables are quantitative sciences such as physics. Sciences that deal with phenomenon including non-quantifiable variables are qualitative sciences such as psychology (Tossi, 1977).

Audiology as an applied interdisciplinary behavioral science, involves both quantitative and qualitative variables. Hence the topics for inquiry within this area are numerous, which vary in character from those that are strictly physical question to those that are principally behavioral (Peterson and Fairbanks, 1963).

As Oyer and Beasley (1973) say, when one asks questions concerning the impedance characteristics of pinna, he deals with physical problem. But when the same individual is interested in noxious acoustic elements of airport or factory, he first makes a physical measurement of the noise stimuli. Then he also makes a threshold measurement which is a psychophysical measurement. Both the measurements employ numbers in the process, and investigator describe and communicate their finding via the language of mathematics. Thus, the precision with which the description can be made of psychophysical phenomenon concerning the human functions of hearing is related to the progress and maturity of scientific study of hearing.

The study of hearing becomes scientific when it incorporates all the scientific methods of research. The scientific method is a set of rules that can be used for describing events, explaining events and predicting events (Silverman, 1977).

a) Describing events means partially answering questions beginning with `who, what, when, where, how and which' question concerning events.

Examples: 1. When is a hearing loss educationally significant?

2. Who can benefit from a particular type of hearing aid?

This descriptive approach maximizes the reliability and validity of a description and permits its reliability and validity to be estimated.

b) Explaining events means dealing with the `why' questions, that is specifying the reason or reasons for events. This approach is helpful in establishing the degree of certainity an audiologist can have in answers to `why' questions.

c) Predicting events means making possibility statements about the likelihood of their future occurrence. The degree of accuracy with which such predictions can be made will be atleast partially a function of the amount of information available about the conditions underwhich the events occur.

According to Silverman (1977) the scientific method suggests the following criterias for formulating questions, making observations to answer them and relating the answers to the existing body of knowledge.

1. Intersubjective testability – suggests that the observational procedures (methodology) used to answer questions should be described insufficient detail that they can be repeated by others.

2. Reliability-suggests that a sufficient number of observations should be made before we attempt to answer a question so that we can have a reasonable degree of confidence in the accuracy of answer. It also suggests that the study should incorporate some way of estimating the reliability of observations.

3. Definiteness and Precision – suggests that concepts referred to in questions formulated for research

purposes should be an definitely delimited as possible. It also suggests that language used to answer questions and interpreting answers should be as unambiguous as possible.

4. Coherance or Systematic structure – organization of the pieces of relevant information resulting from the observational process to make them coherent.

5. Comprehensiveness or scope of knowledge – indicates that answers to questions should be regarded as tentative and subject to change when new information becomes available.

For our purposes, we shall define research as a process of asking and answering questions that is governed by a set of rules referred to as the scientific method.

II. ROLE OF STATISTICS IN RESEARCH

Statistics is a branch of applied mathematics which specializes in procedures for describing and reasoning from observations (Williams, 1968).

A. Use of statistics in reporting observations:

Let us consider first, what is meant by a phenomenon. Phenomenon is any object or event the characteristic of which are amenable to observation. The next step is to find how characteristics of a phenomenon vary and how these phenomenon affect other phenomenon. The term Variable can be used to define the varying characteristics of a phenomenon. Variable is a phenomenon considered in terms of how its characteristics may differ according to some well defined scheme. Thus a variable or phenomenon is what is being observed, the researcher's reports of this observations are considered as data. <u>Data</u> (plural) are the reports of observation of variables. In research, the data are always in the form of some type of measurements. <u>Measurement</u> is a scheme for the assignment of numbers or symbols to specify differing characteristics of a variable. Who, or What would be one's focus for measurement, usually we are concerned with what is called a population. <u>Population</u> is any class of phenomenon arbitrarily defined upon the basis of its unique and observable characteristics.

In statistics, the word population is not limited to people; it refers to the total class of whatever is observed as a part of study. Whatever is defined as the research population is purely arbitrary; it depends upon the limits of the research problem. In many cases it is not feasible to measure an entire population, hence the researcher settles for some portion of the population called a sample. <u>Sample</u> is a collection of phenomenon so selected as to represent some well defined population. Once observation and measurement have been undertaken, what we have is a body of measurement data. The total body of data represents measures that we have taken upon every member of a sample or every member of a population.

The main role of statistics comes in reasoning from these measurement data to the overall statements that we wish to make about the data. In this sense statistics provides us with mathematical models for reasoning. Statistics provides us with a great variety of procedures each with a pattern of mathematical deduction that provides some type of statistical conclusion. We identify variables that we have observed in the real world with variable that are symbolized in the formulas. Given a set of adequate identification we then reach according to the deductive operation of the formula until we reach some desired statistical result.

The word `statistic' refers to the kind of result that a statistical formula provides. Statistically derived values which reflect an average, values which represent dispersion of measurements or values which represent other such characteristics are called descriptive statistics. Descriptive statistics are calculated values which represent certain overall characteristics of the body of data. The term sampling statistics is used when the body of knowledge comprises only a sample. Sampling statistics are calculated values which represent possible deviation of a sample characteristic from population characteristics.

B. Statistics and Research Plan:

Research plan involves problem, method and results. Statistics can have a role at each stage of such a plan.

Problem itself is defined in a statement of the problem.

It might be expressed as either a purpose or as a question and in theform of a hypothesis, a statement susceptible to testing by reasoning from observation. The statement of a given problem always has an effect upon the type of statistics to be used in carrying out th study. The problem statement also provides a preliminary definition of the population to be studied. This too will have a consequence upon the statistics to be used.

Method is the overall plan for gathering data. There are 2 methods of gathering data.

i. Descriptive method: It is a research plan undertaken to define the characteristics or relationships or both among variables based upon systematic observation of these variables. It is also known as empirical approach.

ii. Experimental method: It is a research plan undertaken to test relationships among variables based on systematic observations of all variables which are manipulated by the researcher. Here the usual intent is to test an hypothesis of cause and effect. That is manipulation upon one variable lead to consequences upon another variable. Independent variable is a phenomenon which is manipulated by the researcher and which is predicted to have an effect upon another phenomenon. Dependent variable is a phenomenon which is affected by the researcher's manipulation of another phenomenon. Statistics employed in both researches differ.

From statistical point of view we are interested in subjects because as a minimal unit of observation, the number of such subjects defines atleast the minimal number of observations made in a particular study. This `N' is required for calculation of either most descriptive statistics or sampling statistics. The most usual statistical concern with materials of a study is the type of measurement that has been involved. The choice of particular statistics depends in part upon the type of measurement used. Particular procedures carry implication concerning the use of statistics in a study.

Part of the task of designing a study involves deciding beforehand just what statistical procedures will be used and what results would require what type of reasoning from the point of view of generalization. The results of a study begin with whatever has been deduced by the use of statistical procedures. The task is then to interpret those statistical results in terms of what we identify them within the population under investigation.

Briefly, the importance of statistics in research can be stated by the following points (Guilford, 1978). They,

1. permit the most exact kind of description.
2. force us to be definite and exact in our procedures

and in our thinking.

3. enable us to summarize our results in a meaningful and convenient form.
4. enable us to draw general conclusions.
5. enable us to predict `how much’ of a thing will happen under conditions we know and have measured.
6. enable us to analyze some of the causal factors underlying complex events.

CHAPTER II

STATISTICAL METHODS AND THEIR APPLICATION TO AUDIOLOGY

Statistical methods are actually a means of reducing volumes of data to a form simple enough for the mind to grasp. They help us to make the best possible guess when encountered with uncertain facts.

What we get out of an experiment is a set of raw scores which are in any one form of scales used for measurement. Unless, these raw scores are subjected to suitable statistical manipulation, we will not be able to draw any conclusion from the observations made.

In general, statistical methods are used to

a) control the experiment and

b) make analysis of experimental results

We can divide the statistical methods used to serve the above purposes into 2 kinds:

a) Descriptive statistical techniques and

b) Inferential statistical techniques

We shall now see, how these methods are employed in audiological experimental data. The procedures for calculating each statistics are illustrated in the appendix.

a) Descriptive statistical techniques: They are used for the organization and summarization of the quantitative data. They are of 3 types:

1. measures of central tendency
2. measures of variability and
3. measures of correlation.

<u>Measures of central tendency</u>: These indices designate the \`average', \`typical' or most frequently occurring score in a set of scores. When these scores are obtained by administering say, a speech discrimination test, to a group of young children we can find 3 central tendency measures: i. the average score earned by these children ii. the score earned by the typical member of the class and iii. the score earned most often by the children in this group. Thus obtained indices represent the scores obtained by all the children. There are 3 ways of finding the representative score. They are mean, median and mode.

<u>Mean:</u> It is the sum of the scores obtained by a group divided by number of scores. For example, this measure can be used to answer the following questions:

1. Do conductive loss children perform better than sensorineural loss children on a speech discrimination test?
2. How many hours of therapy, can be required for moderately severe hearing impaired children and severely hearing impaired children to pass a auditory training task?

To answer the first question, one can average the scores obtained by both groups and compare their mean values and findout their performance. To answer the second question, average the nmber of hours taken by both the group of children, and compare their mean values. Mean is the commonest and easiest measure of central tendency: Mean is computed whenever the measurement is done in the form of interval or ratio scale.

Median: It is that score which occurs at the midpoint of a set of scores when they are ordered from lowest to highest. It can be used to answer the following questions:

1. What would be the most `typical' dysfluency of a normal hearing individual under delayed auditory feedback of specific duration?

To answer this question, rank the disfluencies of his speech in order of frequency. The type of disfluency in the middle of the rankng would be `typical'. Unlike mean, median can be computed in both the measures having interval or ratio properties and ordinal or ranking properties.

Mode: It is the score that occurs most frequently. It can be use to answer the following questions:

1. What would be the maximum speech discrimination score that can be obtained by otosclerotics most frequently?
2. How often a noise of 90dBSPL can have deleterious effect on hearing ability?

To answer the first question, find out the maximum discrimination score obtained by each otosclerotics in a group. That is the mode. It can be used to answer questions about group tendencies. It can also answer `probability question' such as

The second one. The mode is appropriate for measures having nominal, ordinal and interval or ratio properties.

Measures of Variability: These indices define the spread, dispersion, homogeneity or variability of a set of scores. They indicate how far the obtained scores deviate from a central value. Before making any statements, such as "all levels of time compressed speech bring about the same results" one should be cautious as to what degree each level of time compressed speech bring about variations within a individual and among the individuals. In such instances we need to rely on measures of variability. The following are the commonly used measures of variability: range, semi-interquartile range and standard deviation.

Range: It is the difference between the highest and the lowest score in a group of scores. It can be used to answer question such as;

1. What would be a suitable dynamic range for hearing aid selection in a group of sensorineural cases?
2. What is the age range of clients receiving a particular therapy?

The range is appropriate for measures having ordinal, interval or ratio properties.

Semi-interquartile range (or interquartile range): Both of them are based on middle 50% of scores, that is the highest 25% and the lowest 25% are ignored, Semi-interquartile range is equal to interquartile range divided by tow. The following questions can be answered.

1. Do children's score on discrimination list A tend to be more variable than on discrimination list B?
2. Within what range of scores do the 25th and 75th percentiles fall for 6 year old deaf on a particular test of sign language?

These measures are most frequently used for measures with interval or ratio properties.

<u>Standard deviation</u>: It is the most frequently used measure of variability, which reflects the variability of all scores in a group. It is the average deviation of the scores in a group from their mean, which is called variance and the square root of which gives the standard deviation. This can be used to answer questions such as:

1. How variable are scores on a particular diagnostic test?
2. What range of scores would constitute normal limits on a particular diagnostic test? (assuming that scores on this test are normally distributed)
3. How far, on the average, did a person's thresholds for a $1000H_z$ tone deviate fro the mean threshold of this tone, when tested repeatedly?

The first question, relates the use of standard deviation as a measure of variability. The second question, illustrates its use in estimating proportion of the population whose scores fall within certain limits. The third question is used to assess the representativeness of the mean. This is appropriate for measures having interval or ratio properties.

Measures of correlation: These are the indices, used to describe the strength and direction of relationships between two groups of measures or between 2 variables. The numbers that are yielded by such indices are called as correlation coefficients.

Example: Children with long standing conductive loss have generally poorer language performance on a variety of measures than do a group of matched subjects with no conductive hearing loss.

Suppose, in this example, in order to establish an association between conductive loss and performance on language test battery one may have to rely on some sort of correlation coefficient which would give the following two kinds of information generally:

1. the strength of relationship between these variables: Correlation values usually range between plus 1.00 and minus 1.00. When the obtained indices is higher enough, say 0.90, then it is said to have good relationship.
2. the direction of relationship: It would also say, whether an increase in conductive loss brings a corresponding increase or decrease or any changes in the language abilities, that is whether it is positively correlated or negatively correlated.

The following are the commonly used types of correlation coefficients: contingency coefficient, phi-coefficient, Spearman rank order coefficient and product moment correlation.

Contigency Coefficient: It is the index of the degree of relationship between 2 attributes of events that had been assigned on the bases of these attributes to the cells of a contingency table. This contains two or more rows and two or more columns. The number of column and number of rows are determined by the number of categories into which the scale for an attribute is divided or segmented. For the attribute `sex' the scale would be divided into two categories `male and female'. For the attribute `type of hearing disorder' the scale would be divided into following categories: otitis media, otosclerosis, and meniere's disease etc.

	Male	Female
Otitis media		
Otosclerosis		
Meniere's diseases		

In the above example, the variables measured have nominal properties. Contigency coefficient can also be calculated for measures having ordinal interval or ratio properties. Suppose if one wants to relate the amount of speech discrimination to amount of hearing loss in 100 adults he can adopt a contingency table as given below. Depending on the number of individuals falling under each cell we can findout the strength and direction of relationship. Procedure is dealt with in the appendix. This coefficient can be calculated reliably when the table is larger than 2 x 2.

```
                                      Speech discrimination (in %)
                                          70        80        90
                                    |-------------|---------|----------|
Hearing              30             |-------------|---------|----------|
Loss in              50             |-------------|---------|----------|
dB                   70             |-------------|---------|----------|
```

<u>Phi-Coefficient</u>: When the contingency table is 2 x 2 type this can be employed to answer the same type of questions as the contingency coefficient. It is a useful index of correlation between dichotomous attributes such as sex.

<u>Spearman rank order correlation coefficient</u>:

When the obtained scores on any two variables have ordinal relationship with each other, this coefficient can be calculated. This coefficient is an index of the extent to which two groups of scores rank order a group of persons in the same manner. This can be used to answer the following questions:

1. Is the amount of improvement in therapy exhibited by 10 deaf children related to their IQ?

2. How well do 2 audiologists agree in their orderings of clients with regard to degree of improvement following a program of therapy?

Answering the first question requires ordering of the amount of improvement shown by each deaf child against their IQ and finding out the correlation. Answerng the second question, requires ordering the clients on the basis of each audiologist's ratings of degree of improvement.

Pearson Product moment correlation coefficient:

When measures of two variables have interval or ratio properties and the relationship between them is assumed to be linear, then to findout the relationship, the pearson product moment correlation can be applied. It is the commonly used index of correlation. It can be used to answer the following question:

1. What is the nature of the relationship between preschoolers auditory discrimination ability and their articulation proficiency?

2. What kind of relationship exists between the loudness discomfort level and hearing threshold in sensorineural loss?

Answering the first question requires assessment of the relationship between performances of different tasks by a group of persons at the same time. The second one, requires direct measurement of both loudness discomfort level and hearing threshold, and finding out the relationship.

b. Inferential statistical techniques:

When we want to generalize our results beyond our set of data or sample of subjects, to another set of data or sample or population we need to rely on a group of techniques called as inferential statistical techniques. They enable us to answer following kinds of questions about a set of data (Silvermen, 1977).

1. Reliability of differences or relationships observed in a set of data. It is by estimating the probability

that observed differences resulted from chance or random fluctuations. This probability estimations are done through the `significance tests'.

2. Generality of differences and relationships observed in a set of data. Specifically they answer the following question " How likely is it that the differences or relationships that one has observed in the sample of persons one has studied also are present in the population from which these persons were selected?". Again significance tests are applied to answer such questions. However a statistical method called `confidence intervals' would only provide an estimate of the magnitude of difference or relationships.

3. Estimating the population values of descriptive statistics and of difference between descriptive statistics. Confidence intervals would help in making such estimation.

Example: Suppose, if one finds improvement in auditory discrimination skill of young deaf children, after an auditory training program, he can find out or answer these questions with the pre-training and post-training scores on an auditory discrimination task.

a. Whether the improvement shown is really due to treatment?

b. Whether the same kind and amount of improvement can be shown by another group when subjected to training?

These questions can be answered by inferential statistical techniques and these techniques can be divided into two categories:

1. Significance tests and
2. Confidence intervals

Significance tests: They are used to answer the following question: `How likely is that, a difference or relationship observed in a set of data is the result of chance or random fluctuation?'. The differences refer to any types of measures including `mean, median, standard deviation and correlation coefficients.' The term chance or random fluctuation can be explained through the probability theory which states that measures that are not different or related can appear to be different or related. To determine whether the probability that an outcome which is due to chance or random fluctuation is adequately small, a significance test can be used to assess the viability of the null hypothesis.

Null hypothesis states that the observed differences or relationships are due to chance or random fluctuation. If the probability of a null hypothesis being true were relatively small, it would be rejected. Usual levels of probability for accepting or rejecting null hypothesis are 0.05 and 0.01 levels of confidence. Rejection of a null hypothesis only means that an observed difference or relationship is unlikely to have resulted from chance or random fluctuation. Failure to reject a null hypothesis does not necessarily mean that the difference or relationship observed is the result of chance or random fluctuation. Explanations:

1. the magnitude of difference or relationship may be a real but a smaller one, such that the significance test is not powerful enough to detect it. So when the difference is very small, we need to have larger number of subjects, to reject the null hypothesis.

2. the real difference or relationship (that is one in the population) is larger in magnitude than the observed difference or relationship. This would be due to the inadequate representativeness of the sample.

Many types of significance test have been developed for testing null hypotheses. To select a particular significance test, one has to consider the following pieces of information about the date on which the test is to be performed (Silverman, 1977).

1. the level of measurement of the measures
2. whether the measures are independent or related and
3. the number of means, medians or other measures one wish to test the difference between.

The following table gives the classification of general significance tests based on above information (Silverman, 1977).

Level of Measurement	Two sample case		More than 2 Sample	
	Related Samples	Independent Sample	Related Sample	Independent Sample
Nominal		Chi square		Chi square
Ordinal	Sign test	Mann- Whitney U test	Freidman 2 way Analysis of Variance	Kruskal-Wallis One way analysis Variance
Interval and Ratio	t- test for related measures	T test for independent measures	One way analysis of variance for independent measures	One way analysis of variance for independent measures

(The procedure for calculating each statistic is given in appendix).

The level of measurement of the measures refers to whether they possess nominal, ordinal, interval or ratio properties. There are different significant tests for each of these measures as shown in the table. Significance tests with higher level of measurement are more powerful. If the set of measures were made on the same person, they are related and apply a related measures significance test.

Example: Assessing the difference between measures made on a group of persons before and after participating in a therapy program.

When the measures are made on different persons, regarded as independent.

Example: When one wants to compare a group of persons having otosclerosis to another group of persons who do not have otosclerosis on some measure he has to use a significance test appropriate for independent measures.

When the two groups have been matched it has to be treated as related measures. When the number of means, medians, or other summary measures one wishes to assess the difference between is `two' one set of significance test is appropriate; if it is more than `two' a second set is appropriate.

<u>How to use this table</u>? If one administers a picture vocabulary test to a group of children with severe hearing loss twice, once proceding and once following their participation in a therapy program, to findout whether the obtained difference between means of two sets of scores is significant enough to reject the likelihood of chance or random fluctuation, make an assumption that the scores on vocabulary test have ordinal properties. Since the measures are taken on the same group, they are related and they are 2 sets. From the table one can find the significance test suitable for related two sets of measures having ordinal property as sign test.

Confidence Intervals: They are inferential statistics used for estimating population values of descriptive statistics and difference between such statistics. The following questions would require such an estimation:

1. By what age do 50 percent of children perform a particular task correctly?

2. What percentage of persons with a particular hearing oss derive benefit from a particular type of amplification?

3. How predictive is a person's score on a dichotic listening test before he participates in a training program of what it will be after he participates in the program.

Estimate the population median to answer the first question and population percentage to second and population correlation coefficient to answer the final question. In the same way, the differences between descriptive statistics can also be answered.

These confidence intervals are designated by two values (a lower limit and an upper limit) between which we can be given a percent certain (usually 95 or 99%) that a population value falls. Confidence intervals can be computed from measures having ordinal, interval or ratio properties.

CHAPTER III

EXPERIMENTAL DESIGNS IN AUDIOLOGY

I. Meaning and Nature of Experimental Design:

Experiment is a means of collecting evidence to show the effect of one variable upon another. In the ideal case, the experimenter manipulates the independent variable, holds all other variable constant and then observes the changes in the dependent variable. Here any changes in dependent variable is considered to be the result of manipulation of independent variable (Plutchik, 1968).

The different ways in which data can be collected in an experiment is the subject matter of experimental design. The major purposes of an experimental design are two fold:

1. they provide the ways of arranging the condition of an experiment in order to answer the questions we are concerned with.

2. they have the purpose of eliminating or minimizing sources of error or bias so that unequivocal causal connection can be established.

An independent variable used in an experiment may be either a `treatment' variable or a `classification' variable. A treatment variable involves modification in the experimental subjects, that is subjects are treated in some way by the experimenter. Experimental subjects may be classified on a characteristic which was present, prior to the experiment, and does not result from the manipulation of experimenter, such a variable is a classification variable. Examples are sex, age, IQ, socioeconomic status etc.

An independent variable is also called as a `factor'. When an experiment involves a single treatment or classification variable (independent variable) with 2 or more levels, it is called as single factor experiment. Experiments which investigate simultaneously the effects of two independent variables are called as two-factor or two-way classification experiments.

The external variables affecting the experimental results can be divided into 1. Subject variables: examples are age, sex, intelligence, motivation etc and 2. Situational variables: those which are associated with conditions underwhich the experiment is conducted. Usually all the experiments have some way of controlling these variables.

Controlling the subject's variables: Here we want to ensure that the groups of subjects tested under each experimental condition are as similar as possible on all the dimensions along which people can differ. This is done by carefully controlling the way in which subjects are allocated to the experimental conditions. Three methods are adopted to ensure such control.

a. Repeated measures design: Use the same subjects in each group. Each subjects perform under each conditions of the experiment, so that the effects of subject variables will balance out. By analyzing the difference between treatment scores belonging to each subject we obtain a very sensitive measure of the effects of independent variable.

Example: Studying the effect of delayed auditory feedback on spondee threshold. Use the same group of subjects under each level of delayed auditory feedback.

Independent variable

No delay	0.2sec delay	0.4sec delay
S1	S1	S1
S2	S2	S2
S3	S3	S3
.	.	.
.	.	.

Here a new irrelevant variable is the order in which various delays are received by each subject. If subject 1 receives the delay of 0.4sec first and 0.2sec second time, there can be practice effect. So we must ensure that the order in which the delays are received must be counter balanced across subjects, that is 1/3rd of the subjects follow one order, and the remaining two 1/3rds follow different order. If the order effects are asymmetrical; example. One delay produces more fatigue than the other then we can not expect the order effects to be neutralized by counterbalancing or any similar procedure. Thus repeated measures design should only be used when we judge the order effects to be symmetrical or insignificant.

b. Matched subjects design: means having pairs of individuals who are very similar on the variables that influence the behavior under study.

Example: Comparision of the effects of two different types of instruction on the speed with which pure tone thresholds are elicited.

Here, since we cannot test a subject two times, the repeated measures design is inappropriate. It is better to use a matched subjects design in which each pair of subjects is matched with respect to age, sex, hearing ability. Then we can allocate one member of each pair to the first condition and the other member to the second condition. The allocation is made on an ordered random basis.

	Instruction 1	Instruction 2
Pair 1	S1a	S1b
Pair 2	S2a	S2b
Pair 3	S3a	S3b
.	.	.
.	.	.

The difference between each score would reflect the influence of type of instruction.

c. Independent groups design: Here, a group of experimental subjects are divided into `K' independent groups using a random method such that each subject has an equal chance of being allocated tone group or the other. A different treatment is applied to each group. One group may be a control group, that is a group to which no treatment is applied. A meaningful interpretation of the experiment may require a

Comparison of results obtained under treatment with results obtained in the absence of treatment. Comparisons may be made between treatments and a control, between treatments of both.

Controlling situational variables: such as background noise, apparatus changes, experimenter's behavior and so on. These factors can easily confound the effects of the independent variable if they changed systematically from one condition to another. Effective way of avoiding this is to hold the variables in question constant throughout the experiment. Thus background noise can be kept at a minimum or eliminated by conducting the experiment in a sound proof room. Hold the experimenter's behavior by tape recording the instruction. Thus systematic bias and random effects are eliminated.

Some variables simple cannot be held constant, for example it may be necessary to test subjects on different days of the week or to use the several experimenters during the course of the study. The ways to deal with such variables is to balance their effects across conditions of the experiment, that is by applying the same pattern of changes to each condition.

When both methods are not possible, test the subjects in a random order rather than dealing with each condition in turn. The general purpose of randomization is to protect the validity of the experiment by controlling the biasing influence of extraneous variables.

In more specialized designs, as the experimental conditions increases the complexity of the method of allocation of subjects to each condition also increases, in order to have a good control over all the subjective and situational variables.

" The most frequent errors occurring today involve the more complex analyses, particularly those designs involving several experimental dimension and repeated measurements of the experimental subjects " (Feldt, 1959).

Some of the criteria for good experimental designs are as follows: (Winer, 1962)

1. the analyses resulting from the design should provide unambiguous information on the primary objectives of the experiment. In particular the design should lead to unbiased estimates.
2. the model and it's underlying assumptions must be appropriate for the experimental material.
3. the design should provide maximum information with respect to the major objectives of the experiment per minimum amount of experimental effort.
4. the design should provide some information with respect to all the objectives of the experiment and
5. the design must be feasible within the working conditions that exist for the experimenter.

II. Description and Review of Experimental Designs in Audiology

" Statistical procedure and experimental design are only two different aspects of the same whole and that whole

Comprises all the logical requirements of the complete process of adding to naturalknowledge by experimentation" (Fisher, 1951).

In this section, some of the experimental designs, which have been made use of in the recently published audiological studies have been reviewed.

A) Single factor experimental designs:

Basic designs such as repeated measures design, matched subjects design have already dealt with examples in the previous section. Single factor experimental designs other than those described previously are described below.

1. Group design: It permits the average (mean, median or mode) performance of the subjects in a group under the experimental condition or conditions to be determined. Two general categories of group designs can be identified:

i. those intended to determine whether a \`difference' exists – these designs permit answering of questions such as

> Does hearing aid A tend to result in better speech discrimination ability for persons with conductive loss than hearing aid B?

ii. those intended to determine whether a \`relationship' exists – these permit the answering of questions such as

> Are scores on audiometric test A negatively related to scores on audiometric test B?

Moncur and Dirks (1967) making use of group design studied the monaural and binarual speech intelligibility under 4 types of reverberation times combined with competing and noncompeting message condition in a group of 48 adults with normal hearing. Reverberation conditions were randomized. Analysis of results made in terms of mean, standard deviation and t test.

Perozzi and Kunze (1971) with a group of 30 normal children investigated the relationship between 2 measures of speech sound discrimination skill and specific as well as general language ability. Pearson product moment correlation coefficient was found to signify the relationship.

2. Experimental and Control Group design: Feldman and Reger (1967) studied the relation between selected measures of audition (pure tone threshold, speech reception threshold and speech discrimination) and known measures of central function (visual, tactile and auditory reaction time) in 5 experimental groups divided according to the age and a control group. Tests were administered randomly. Mean, standard deviation and 3 reaction time for each group were found.

Prins (1963) using a experimental group of 26 children with defective articulation matched with a control group of 19 children with respect to age, intelligence studied the relationship between specific articulatory deviations and speech sound discrimination skills. Correlation coefficient was applied to find the relation.

Cozed's (1971) study of speech reading skill and communication difficulty ofchildren and young adults with unilateral hearing loss involved a randomly selected 9 males and 9 females with unilateral hearing loss matched with a control group with respect to age and sex. To prove the hypothesis " if unilateral group did have more frequent listening difficulty than the normal hearing subjects that they would have relied on and in turn developed better speech reading skills," both groups have been tested with 3 lip reading tests A, B, and C in a counterbalanced order and Mann-Whitney U test of significance was employed to verify the hypothesis.

Kronvall and Diehl (1954) using a matched experimental control group design studied the relationship between auditory discrimination and organic articulatory defects in 2 groups of elementary grade children (30). The matching variables included age, sex, grade and intelligence.

Stafford (1962) compared the problem solving ability of 29 congenital deaf children with 29 normal hearing children matched on the basis of IQ, sex and age and studied whether these differences varied with the difficulty of the problem and the age of the child. Both the groups were divided based on mental age into 3 such as low, median and high mental age. A t test was employed to analyze the number of problems solved and a chi square test, to analyze the number of trials required to solve each problem.

Pederson (1974) tested the hypothesis that normal cochlea is essential tonormal temporal integration using a before-after experimental control group design. He studied the perception of brief tones, prior to and after salicylate treatment in 14 experimental subjects and 7 control subjects. Mean and standard deviation of the temporal integration for each frequency with each group was found.

Vasudha (1972) using a series of before-after experimental control group design investigated the improvement in intellectual levels among 29 hearing impaired and 29 normal children consequent to speech and language therapy.

3. Independent Subjects design: Here the selected group of subjects are independent of each other in their characteristic. Data for each group of subjects is collected independently.

Siegenthaler (1949) after having selected 3 groups of subjects (5 each), one with flat hearing loss, second group with high frequency hearing loss and final 5 as control subjects studied the relation between hearing loss and intelligibility of selected CVC words. 3 types of word variables (voicing of consonants, pressure pattern of consonants and influence of one sound upon another) were studied. Variables such as intensity level, syllable, phonetic elements, pitch and voice quality of the announcer were held constant.

Shapiro (1979) using a moderately severe sensorineural hearing loss group of 10 subjects and severe-profound hearing loss group of 10 subjects evaluated the relationship between loudness discomfort level for wide band noise at frequencies 500, 1000, 2000 and 4000H_z and hearing threshold. Analysis of variance was done to find the variance between 2 groups.

Harford and Jerger (1959) studied the effect of loudness recruitment on delayed speech feedback presented at various supra threshold levels. Experimental groups involved independent subjects of 5 groups with 10 subjects each. Hearing impaired subjects with recruitment, hearing impaired subjects without recruitment (bilateral otosclerosis), normal subjects, hearing impaired subjects with recruitment but without speech discrimination loss and a final group of 10 normal subjects matched to the otosclerotics in terms of age. The 6 presentations level were apparately randomized for each subject. Median error score at each level for each group was compared.

4. Independent Groups design: As previously stated a group of experimental subjects are subdivided into smaller groups through randomization and each group receives different treatment.

Fairbanks and Guttman (1957) assigned independent groups of subjects to 5 experimental condition which represented a series of compression ranging from 0-100% and studied their effect on comprehension of connected speech. Three factor analysis of variance was employed to study the interaction between time compression, listener aptitude and message effectiveness.

5. <u>Repeated Measures Design</u>: Here, same group of subjects are being used under all experimental conditions, as previously mentioned.

Orchik et. al (1979) using repeated measures design studied the interaction between time compressed speech discrimination score and reading readiness skills in a group of 34 children of age range 5 – 7 yrs.

Simhadri (1977) studied the interaction effect between the familiarity with language of the competing signal and various signal to noise ratio, using a experimental control group repeated measures design. Subjects have been randomly assigned to different signal to noise ratio condition. A two way analysis of variance was made use of.

6. <u>Individual Subject design</u>: It permits the performance of individual subjects under the experimental condition or conditions to be reliably determined. Reliable conclusions can be reached concerning the effects of an experimental condition on the behavior of an individual subject. It is also known as single subject design.

It permits answering questions such as `Has a client improved following a period of therapy?' This would be a before-after design where a single measure is made twice just prior to treatment and following treatment.

Symbolically,

X O X where X is the measure and O is the treatment.

According to Silverman (1977) research that utilizes this design is difficult to interpret since it contains no way of assessing reliability or controlling of souorces of error. Some ways of increasing it's level of reliability are as follows:

1. making several measurements just prior to and after treatment

X X X X X O X X X X X

Example: Obtaining pre-aided discrimination scores and comparing it with post-aided scores.

2. finding out two measures simultaneously prior to and after treatment

X1 X1 X1 X1 X1 X1 X1 X1

O

X2 X2 X2 X2 X2 X2 X2 X2

Example: Studying the articulatory skills and vocabulary level of a deaf child just prior to a specific therapy and after it has been concluded.

3. another feasible approach is to replicate the treatment or experimental condition, symbolically

X X X O X X X O X X X O X X X

Example: Studying the effect of masking noise on time compressed speech discrimination (30%). X would be the speech discrimination score under time compression. O would be the additional masking noise.

4. the above replicated design can be further modified, by randomly administering the treatments

5. measure two behaviors at given points in time – one that treatment should modify and one that the treatment would not be expected to modify. If the treatment modified the one and not the other this would be interpreted as evidence of it's effectiveness.

Before-after designs are otherwise known as pretest-post test designs and one of their variant is called time series design where before-after measures are made at fixed set of time intervals.

B. Two-Factor and More than two-factor Designs:

1. Factorial Designs: They are used to study simultaneously the effects of two or more independent variables. These designs permit the evaluation of interaction effects between variables. Designs in which the treatments are combinations of levels of two or more factors are called as factorial designs. If all possible treatment combinations are studied then the design is called as Complete factorial design On the other hand, if reduced number of systematically selected treatment combinations are used they are called as Incomplete factorial designs.

If a study involves simultaneous study of two levels of one independent variable and two levels of another independent variable then it is called as 2 x 2 factorial design. In factorial designs, the number of subjects under each cell is kept constant. As the level of variables increases, design becomes complex and interpretation becomes difficult.

McCoy et. al (1977) using a 2 x 2 factorial design studied the effect of age and sex on a dichotic listening test. The age variable had two levels-young adults, older adults. Using an appropriate analysis of variance procedure studied the main effects of age, sex and their interaction effects. Schematically, it can be represented as follows.

Mean Age (yrs)

		23	66	
Sex	Male			N-16
	Female			

Mcleod and Greenberg (1979), using a 2 X 3 factorial design with repeated measures studied the relationship between loudness discomfort level and acoustic reflex threshold in 15 sensorineural cases and 15 normal subjects for noise stimuli at 1000, 2000 and 4000Hz.

Stimuli (frequency)in Hz

Group	1000	2000	4000
Normal			
SN loss			

Smith (1969) using a 2 X 2 factorial design studied the effect of hypnosis and suggestion upon auditory threshold in 64 college students. The design can be represented as given below:

Suggestion

		Hypnotic	Waking
Hypnotic state	Present	16	16
	Absent	16	16

Sticht (1969) studied the intelligibility of time compressed words as a function of age and hearing loss by using a 2 X 2 factorial design with repeated measures on both the factors. A three way analysis of variance was applied to study the interaction between age, hearing loss and percentage of time compression.

		Age in Years	
		More than 60	Less than 60
Hearing Loss	Normal	14	14
	SN loss	14	14

Nikam, Beasley and Rintleman (1976) studied the perception of four time compressed CNC monosyllabic lists of B form of Northwestern University Auditory test No. 6 in selected two groups of non-native speaker/listeners of English, each group comprising of 72 Subjects. The design is a multivariate analysis of variance with repeated measures design. Here, each subject was assigned randomly to a fixed time compressed condition and received each of the 4 lists at sensation levels of 8, 16, 24, 32 and 40dB, each list presented at a different sensation level, except that the list used for 3dBSL was also used at 40dBSL. List presentation order and sensation levels were randomized for each subject. Design is represented schematically in the next page.

Sensation level

% of Time Compression	8	16	24	32	40
	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
0%					
30%					
40%					
50%					
60%					
70%					

(1,2,3 & 4 are word list number).

Beasley and Beasley (1973) studied the auditory reassembly abilities of black and white first and third grade children using a four factor analysis of variance with repeated measures design. There are 10 subjects per cell for a total of 160 subjects, 80 at each grade level (40 black and 40 white) and 40 subjects for each of the 4 interphonemic interval levels (20 black and 20 white). The 160 subjects were divided into 16 separate groups, each of the 4 race by grade level groups was tested for one of the 4 conditions of interphonemic interval, so that 10 black and 10 white children at each grade level heard one interphonemic interval condition. This design can be represented as given below:

		Race			
		Black		White	
Grade	Ist	G1	G2	G5	G6
		G3	G4	G7	G8
	IInd	G9	G10	G13	G14
		G11	G12	G15	G16

N-10for each group

		Interphonemic interval (in msecs)			
		100	200	300	400
Grade	Ist	G1 G5	G2 G6	G3 G7	G4 G8
	IInd	G9 G13	G10 G14	G11 G15	G12 G16

2. Latin Square Design: It is a variation of two-factor design. Here, when we have 2 independent variables, say for example time compression and masking noise, each factor having 3 levels. Subjects are assigned to each treatment condition in such a way that each group of subject reveives each type of treatment only once. For example A, B and C are the 3 groups. Groups are assigned randomly to each treatment condition and also maintained that each group appear only once in a column and row. Schematically.

		Noise in dB		
		30	60	90
Time Compression	0 %			
	30%			
	60%			

Instead of assigning different groups of subjects randomly, we can also assign, available levels of treatment in a random fashion (counterbalancing) in a way that each each group of subject receive only one sequence of treatments and while the other group of subject receives another sequence of treatment.

Example: Hirsh et.al (1950) during the construction of W1 CID auditory test material, presented 6 word lists to 6 listeners at 6 hearing threshold levels using the following latin square. The order of presentation for each list was same. But the order in which different levels appeared was varied for each listener and each list. The presentation levels were plus 4,2,0 and minus 2,4,6dB above a relative threshold level. For convenience, call them as A,B,C,D,E and F respectively. Schematically this design would be as follow as:

		Word lists					
		1	2	3	4	5	6
	1	A	B	C	D	E	F
	2	B	C	D	E	F	A
	3	C	D	E	F	A	B
Subjects	4	D	E	F	A	B	C
	5	E	F	A	B	C	D
	6	F	A	B	C	D	E

Latin square designs in general are used for 3 purposes

1. to control for possible bias resulting from variables which may be thought to correlate with the dependent variables which are not of primary concern in the investigation.

2. they are used to simplify 3 way and higher order factorial experiments by using a reduced and specifically selected number of groups of subjects and treatment combinations.

3. they are used in a variety of ways in experiments with repeated measurements to counterbalance the order of treatment effects

C. Miscellaneous Designs: Designs which are infrequently used fall under this category.

1. ABBA Design: It is useful in situations where it is desirable to use a single subject and to compare two or more experimental conditions. It is used when the order of presentations of treatment conditions are negligible, a method adopted to balance out the fatigue effects and practice effects across the experimental conditions.

Example: Study of auditory reaction time for puretones at 500 and 4000Hz with a group of 2 adult males.

Say, under each condition subjects are given 10 trials. The subjects would be exposed to 500Hz (condition A) first, then the 4000Hz (condition B), followed by the 4000Hz condition (B), and the 500Hz condition (A). So symbolically the sequence of conditions is represented as ABBA. In analyzing the results, the two A conditions are combined and the 2 B conditions are combined. If the frequencies are increased to include 500, 2000 and 4000Hz then the sequence of testing would be as follows:

A B C D E E D C B D

2. Randomized block design: It is a design in which a basis exists for arranging subjects into subgroups or blocks. The blocks have some degree of homogeneity with respect to a variable which may be correlated with the dependent variable under study. Consider an experiment involving four different methods of teaching speech reading to the deaf children. Subjects may be classified into 4 groups according to performance on an intelligence test, which is the blocking variable. Subjects within the blocks may be allocated at random to the 4 methods. Such design is called as randomized block design. Analysis ofvariance is the method of evaluation of randomized block designs.

When our intention is to study the effect of one independent variable upon two or more dependent variable simultaneously, then special designs called multivariate designs are available. Since, very rarely such conditions are encountered during audiological experiments, they have not been adopted in the past literature.

III. Choosing a desingn and appropriatestatistical analysis:

Selection of an appropriate design depends on the following factors of a study.

1. number of independent variables under study
2. levels of treatment condition within each independent variable
3. population or sample to be studied – single subject group of subjects
4. availability of appropriate statistical analysis.

When the study involves two or more factor with a group of subjects, then any one form of factorial design is adequate. Appropriate tests of significance and analysis of variance to study the interaction effect are available with each type of factorial designs.

When all treatment combinations cannot be administered to ‘K’ group of subjects, and if we want to control the levels of experimental (treatment) condition in a random fashion to counterbalance all possible biases, sources of error then one can resort to a Latin square design (choosen to study the order or sequential effect of treatment conditions). Statistical analysis usually involve a type of analysis of variance.

When it is a single subject of a single group of subject studied under 2 or more experimental conditions, each of them being their own control, and the order of experimental conditions found to have a very negligible effect on can adopt ABBA design. Double blind designs are sometimes used to eliminate the experimenter and subject’s sources of bias.

When an experiment involves only a single factor, then the choice of a design depends on how reliably a selected design can provide information adequate to the study.

The reliability of single factor designs can be studied by dividing them into 2 basic designs, individual subject and group design and critically evaluating their advantages

And disadvantages (Silverman, 1977).

Individual Subject Design	Group Design
1. It makes possible to detect individual differences. Provide data concerning the 'typical' behavior of an individual subject under an experimental condition.	1. Provide data concerning the behavior of the 'typical' member of a group under experimental condition.
2. Here, it is not necessary to assume that subjects respond similarly to an experimental condition.	2. It is necessary to assume that subjects in a group respond similarly to an experimental condition.
3. Necessary for subjects to be run more than once under each experimental condition.	3. Not necessary for subjects to be run more than once under each experimental condition.
4. Results can be generalized to typical behavior of individual studied under experimental condition.	4. It can generalize to 'typical' behavior of mean or median group member under experimental condition.
5. May not be relatively easy to control for order and sequence effects.	5. Usually relatively easy to control for order and sequence effects.
6. Statistical procedures for assessing the reliability of research finding are not well developed.	6. They have a well developed body of statistical procedure
7. Minimum number of subjects necessary is 1	7. It is 10.
8.Generalize to population from which subjects are selected on logical basis	8. Generalize to population from which subjects are selected on statistical basis.

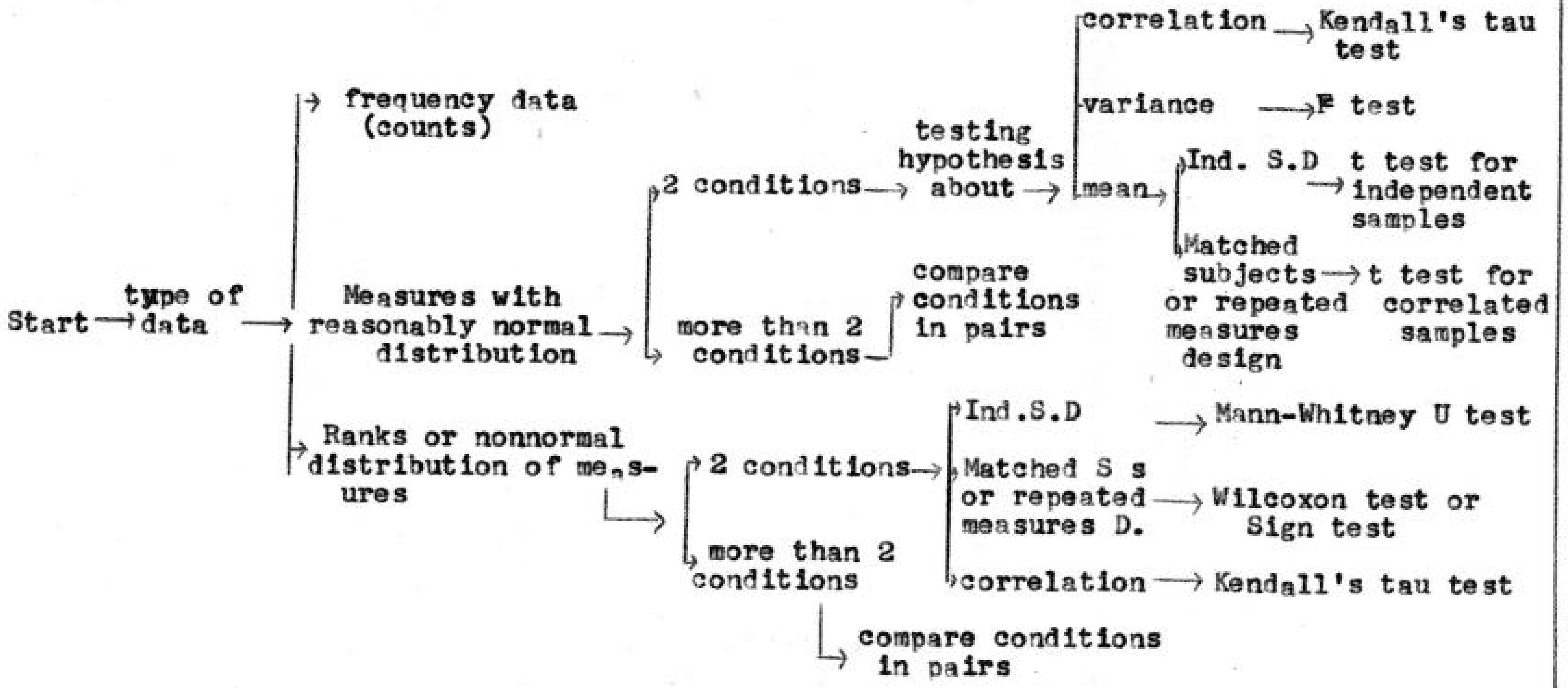
From the above critical evaluation it must be clear that usually a number of factors must be considered while deciding as to whether to use an individual subject design or group design to answer a question. When a research question is clear cut and involves any one of the following features, then the selection is consideration of any one design.

1. Questions that state unequivocally whether they refer to individuals or typical members of typical members of groups
2. Questions for which all treatments could not be administered to individual subjects and
3. Questions for which an insufficient N would be available for a group design.

For the questions where the choice is not clear-cut one must weigh the advantages and disadvantages of each design and select the one that seems most likely to yield data of adequate validity, reliability and generality to answer the questions. The flow diagram given below represents the selection of appropriate statistical procedure for different types of data and basic designs (Robson, 1973).

“If the design of an experiment is faulty, any method of interpretation which makes it out to be decisive must be faulty too. It is true that there are a great many experimental procedures which are well designed in that they may lead to decisive conclusions, but on other occasion may fail to do so: in such cases if the decisive conclusions are infect drawn, when they are unjustified we may say that the fault is wholly in interpretation, not in the design. Both have to be understood. “(Fisher, 1951).

FLOW DIAGRAM



Ind.S.D = Independent subjects design

CHAPTER IV

TEST STANDARDIZATION

I. Meaning and Need for standardization of Audiological tests:

A clinician or researcher can rarely follow a subject for 24 hours a day to record those behaviors which are of interest to him. It is far more feasible to take a sample of what occurs so he can estimate the true condition of his subject in relation to this behavior. This sampling of behavior whether in the form of questions, observations or tasks comprises what is usually termed a TEST. However, the raw data resulting from this sampling are meaningless without further interpretation. In most instances, this interpretation is achieved by comparing what the subject has done to the results achieved on the same test by a group of comparable individuals. It is these results or norms that form the interpretative basis of a standardized test. Such norms result from a process of standardization, that is the collection of responses or scores from a sample of people who are representative of the people on whom one may wish to make observations.

Some of the standardized test in the field of Audiology are:

1. Synthetic Sentences Identification Test (Jerger, Speaks and Tillman, 1968)
2. Staggered Spondaic Word Test (Katz, 1968)
3. Doerfler-Stewert Test (Epstein and Hopkins, (1956)
4. PAL PB-50 Word list (Egan, 1944)
5. CID W-22 Word lists and CID W-1 & CID W-2 Spondaic word tests (Hirsh et.al, (1952)

The remaining tests though not standardized fully, have been used in the clinical[practice just for identification purposes.

The main aim of research is to provide systematic observation of events. This systematic observations, can be accomplished, when we make use of standardized test during the process of data collection. The goals of standardization are:

1. to be objective and formal in our test procedure
2. to control the unwanted variables
3. to exclude the subjective factors
4. to obtain valid and reiable results and
5. to interpret the results more meaningfully and objectively using the norms established.

The clinician and researcher must be certain that the standardized test is the most appropriate instrument for the purposes of routine examinations and research, that it is appropriate for the subjects studied and that it is the best instrument of many available (Ventry, 1980).

The task of doing and evaluating research becomes more complicated when nonstandardized tests are employed. An unreliable or invalid nonstandardized test makes suspect any data reported by the investigator.

Test scores, like all numerals assigned to attributes of events are subject both to random and to systematic measurement error. These errors can be overcome by using standardized test with standard administration procedures and interpretations.

II. Steps in Standardization:

" Clinician or researcher must be aware of the nature of standardizations of any test that he is considering for use. Only by careful study of this factor can be avoid erroneous interpretations and perhaps harmful decision. Tests should be chosen which are appropriate for the opulation to be studied not only in terms of test's validity and reliability, but also in terms of the relevance of the standardization group. Those who would seek to collect norms for a test whether for local or general use must be fully aware of procedures for standardization," (Weiner and Hoock, 1973).

A well constructed test is said to be one which involves a test which measures what it says it does (validity) which is long enough to yield stable scores (reliability) and which has a balanced selection of items relating to the behavior under study (item selection and item analysis) (Weiner and Hoock, 1973).

Adequate standardization requires a number of procedures each of which must be carefully considered (Weiner & Hoock, 1973).

1. The major variables that affect scores on the behavior being tested must be determined. Such variable sare of concern in interpreting an individual's score and are therefore relevant to the choice of the reference groups which must then be selected for the standardization sample.

Example: When one constructs a test of vocabulary for deaf children, he should remember that knowledge of vocabulary varies with a number of factors such as age, intelligence level, sex, urban or rural residence, degree of hearing less and so on. These and other variables can be dealt with in several ways in establishing norms for a vocabulary test. The most commonly used method is to establish a single standardization group with cells (subgroups) formed by the interaction of the variables. Thus if there are 5 age levels, 5 intelligence levels, 2 sex groups, 2 urban-rural groups, and 3 degrees of hearing loss, 300 cells or subgroups would be needed to develop reference or norm tables for member of all groups, which might be studied. There are circumstances, in which the use of one table of norms might result in meaningless comparisons. If the phonemic systems of southern block and northern white children differ widely it may be meaningless to compare the articulation intelligibility test results of a child from either group with norms based on a mixed group.

2. A sample size must be determined that will provide an adequate range of scores for each variable of concern. A small sample may distort the norms obtained by restricting the range of scores (extremes may not appear) and by reducing reliability. The results obtained from 2 small samples may

differ widely by chance. As a rule, 100 to 200 subjects should be included in each cell if adequate range and reliability are to be obtained. In the vocabulary example given above, 30,000 children would have to be included in the sample in order to meet the requirements met. Where the individual testing is involved, as in speech and hearing examination (evaluation), this kind of sampling becomes impossible to realize. Obviously either fewer cells must be used in the standardization of such individual tests or fewer children should be included in each cell. The result is often a compromise, less than ideal but practically useful.

3. The method of choosing the sample must involve a randomization process so that no selection bias showsup in the results. The subjects to be included in a standardization sample must be selected as much as possible by a process which involves randomization, that is a process by which each subject in the population has an equal chance of being chosen. Selecting subjects simply because they are readily available may introduce a bias that will destroy th value of any norms gathered. Subjects obtained in a clinical setting my be representative of a clinical population but certainly not of the general population.

Using the following 3 methods one can derive random samples.

a. Fisher's random number table (1963)

b. Systematic sampling: it is the process whereby all the members of the desired population or subpopulation is alphabetically arranged, and selection of every `n'th person from that group is made.

c. Stratified random sampling: it requires dividing process of a population into various stratas, each strata representing the population, in any of the desired variable such as socio-economic status, male:female and so on. Then members are drawn at random from the various strata. If the members are drawn such that the proportions in the various strata in the sample are the same as the proportions in those strata in the population, the sample is a `proportional stratified sample'.

The most appropriate method is to locate a large pool of relevant subjects for example children in the public schools of one are a. The subjects to be included in each cell of the standardization group are selected from the pool on a random basis. As many subjects are collected as time and money allow, but at the least every variable to be treated must be covered by the minimum number decided upon. Availability of subjects will always introduce some bias.

4. The constructed test must be administered to all subjects within few weeks or months in order not to distort any of the

variables which may be subject to change over time. Follow the same procedure of administration with each subjects.

5. The data need adequate statistical treatment. First it must be determined whether any of the variables choosen are not unique (for example, sex groups may not differ). If so, the groups representing these variables could be combined. The final step in standardization, thestatistical treatment of the raw data obtained is necessary to provide clear-cut reference points for interpretation of individual scores. The test score obtained by any subject can be compared with the summary statistics of the standardization group. Plot a frequency distribution with frequencies on the x-axis and scores on y-axis. The most effective statistics depends on the particular situation and the most frequently used measures being percentiles and some form of standard scores.

A. How to convert a raw score into a derived score?

To establish, norms the obtained scores should be plotted as a frequency distribution, having bell shaped curve. For this purpose, raw scores are transformed or converted into some standard scores, which can be directly plotted as a frequency curve.

The transformation is the process of any systematic alteration in a set of scores whereby certain characteristics of the set are changed and other characteristics remain unchanged, thus changing the shape of frequency distribution of the variables, too often to a normal form. This involve a change in mean, standard deviation as well as a change in skewness and kurtosis. The conversion of a set of frequencies f1,f2,f3, …… fk to proportions by dividing each frequency by N, or to percentage by dividing by N and multiplying by 100 is a simple transformation. The ordering of the transformed values is the same as the ordering of the original frequencies.

A test is said to be standardized, when transformed scores are available, based on a reference group of acceptable size. The transformed scores themselves are called norms.

Methods of Transformation:

1. Sigma score: It is the deviation from the mean divided by the standard deviation.

$$Z = \frac{X - X1}{s}$$

Where X is the mean
X1 is the raw score
s is standard deviation

Symbolically, it can be represented by either s or Z. After finding out sigma score we can apply a formula for conversion of a given raw score.

$$X' = \frac{s'}{s}(X - M)\quad M'$$

Where X- a score in the original distribution; X'- a standard score in the new distribution; M & M' – means of the raw score and standard score distribution; s & s'- standard deviation of raw score and standard scores. When the distributions of raw scores of 2 or more variables are of same shape, then comparision between standard scores of variables is possible.

2. Percentile point and percentile ranks: If `K' percent of members of a sample have scores less than a particular value, that value is the Kth percentile point (Pp). On an examination, if 85% of individual scores less than 60, then 60 is the 85th percentile point (P85). A percentile rank (PR) is a value on the transformed scale corresponding to the percentile point. If 60 is a score below which 85% of individuals fall, then 85 is the corresponding percentile rank. For small N, the computation of PR, Pp is not a very meaningful procedure.

When the data is ungrouped the following formula would aid in the calculation of PR.

$$PR = \frac{R - 0.5}{N} x100$$

Where R- rank of individual counting from the bottom; N- total number of cases.

When the data is grouped, the steps for calculating PR are as follows:

1. findout the exact lower limit of the interval containing the score X whose percentile rank is required
2. findout the difference between X and the lower limit of the interval containing it
3. divide this by the class interval and multiply by the % within the interval, and
4. add this to the percentile rank corresponding to the bottom of the interval.

3. T scores: They are used to transform a variable to the normal form. They are normally distributed with a mean of 50 and standard deviation of 10. Percentile ranks corresponding to certain points on the score scale are calculated. A table of areas under the normal curve is used to find the points on the base line of the unit normal curve is used to find the points on the base line of the unit normal curve corresponding to these percentile ranks. These points correspond to the percentile points on the original score scale. The normal standard scores are multiplied by a constant to obtain any desired standard deviation of the transformed values. A transformed value corresponding to any score value on the original scale may be obtained by interpolation.

4. Stanine scale: It is an approximate normal transformation. The transformed values are assigned the integers 1 to 9. The mean of stanine scale is 5 and standard deviation is 1.96. The percentages of cases in stanine score categories from 1 to 9 are 4, 7, 12, 17, 20, 17, 12, 7, and 4. Thus 4% have a stanine score 1, 7% have a score 2, 12% a score 3 and so on.

A standardized test always accompanies some statement about the reliability and validity of the test constructed.

B. Reliability Estimation: A test score is called reliable when we have reasons for believing the score to be stable and trustworthy. Score should really reflect one’s ability to handle tasks like those represented by the test. Suppose, if a child gets a discrimination score of 80% in test A, and the child's discrimination score on a comparable test B, should not deviate much from 80%. The correlation of the test with itself is called the reliability coefficient of the test. The following methods are used to calculate reliability coefficient.

i. <u>Test-retest reliability</u>: Here the same test is administered twice to a sample of subjects selected randomly from the standardization sample and the scores are correlated. The time interval between 2 administrations should be optimum. It is a close estimate of the stability of the test scores.

ii. <u>Parallel form method</u>: Parallel or equivalent forms of a test may be administered to the same group of subjects, and the paird observations are correlated. Criteria for parallelism are required. Test content, type of item, instructions for administering and the like should be similar for the different forms. Also the parallel forms should have approximately equal means, standard deviation and intercorrelations.

iii. <u>Split-half method</u>: Here, the test is first divided into 2 equivalent halves and the correlation found for these half tests. From the reliability of the half-test, the self-correlation of the whole test is then estimated by the Spearman-Brown formula. This method is employed when it is not possible to construct parallel forms of the test nor advisable to repeat the test itself. It is the best method of measuring test

Reliability, because all data for computing reliability are obtained upon one occasion.

iv. Method of rational equivalence: Two forms of a test are defined as `equivalent' when corresponding items a,A,b,B etc are interchangeable and when the inter-item correlations are the same for both forms. A formula is available for calculating such reliability coefficient from test item statistics.

Among the above 4 methods, the test-retest reliability coefficient is the most commonly calculated value in the area of standardizing audiological tests.

C. Validity Estimation: Any test is valid when the performances which it measures correspond to the same performance as otherwise independently measured or objectively defined. Some of the following methods can be adopted for estimating validity.

i. Validity estimation by means of judgements: The validation of content through competent judgements is most satisfactory when the sampling of items is wide and when adequate standardization groups are utilized. This is also known as face validity. A test is said to have face validity when it appears to measure whatever the test constructor had in mind, namely what he though he was measuring. Face validity is necessary when we must decide what items are suitable for children and which are acceptable to adults.

ii. Experimental estimation of validity: The validity of a test is determined experimentally by finding the correlation between the test and some independent criterion. A criterion may be an objective measure of performance. For example, we can validate speech discrimination tests for children against their articulation proficiency. A high correlation between a test and a criterion is evidence of validity provided i. the criterion was set up independently and ii. both the test and the criterion are reliable. The index of reliability is sometimes taken as a measure of validity.

iii. <u>Factorial validity</u>: It is a complex procedure, wherein the intercorrelations of a large number of tests are examined and if possible accounted for in terms of much smaller number of more general `factors' through a statistical procedure called factorial analysis. It is sometimes found that 3 or 4 factors will account for the intercorrelations obtained among 10 or more tests. The validity of a given test is defined by it's factor loadings – and these are given by the correlation of the test with each other.

Commonly used methods of validity estimations are content validity and predictive validity.

<u>Relation between validity and reliability</u>:

1. the two concepts validity and reliability refer to different aspects of what is called test efficiency
2. A reliable test is theoretically valid, but may be practically invalid as judged by the correlations with various independent criteria
3. a highly valid test cannot be unreliable since it's correlation with a criterion is limited by it's own index of reliability.

III. A critique of specific tests: After having considered the above steps in standardizing a test, one may be interested in knowing how well some specific tests can be criticized on the above grounds. Such a critique might serve purposes like illustrating some of the procedures described and also give some indication of the pit falls test constructors face in establishing fully adequate norms.

i. Vocabulary Norm for deaf children (Silverman-Dresner, 1972)

The purpose of this study was to establish reading vocabulary norms for deaf children, an initial vocabulary pool of 14852 words were selected from various soures and reduced to 7300 words. These words were fed into a computer which produced 73 sets of 100 randomly selected words each. The sets were then converted into 73 vocabulary tests and administered to 13,207 deaf children between the ages of 8 and 17 years, from 89 residential schools for the deaf, by respective class room teacher. Results were analyzed to establish norms for each age group, sex and 6 word classes.

As per the first step in our standardization procedure, the variables, that would affect the obtained score such as age, sex and word class were taken into account, however test constructor did not consider the variables such as socio-economic distribution, racial differences and degree o deafness. The distribution of deafness among males and females in population, whether it was equally represented in the selected sample was also not mentioned. The second step is to check the size of the sample, since the variables for norm estimation involved only 2 sex group, 5 age groups and 6 word classes, the selected sample in the present study is appropriate. Sample selection as the third step, was not stated to involve randomized selection of deaf schools nd randomized selection of subjects within that. Whether the proportionality of sample with respect to age and sex, selected from each school was not delineated. Though the administration procedure, stated to have uniformity, since various class room teacher handled the task of administration, possibilities of procedural variation arises. The span of time taken to finish the testings, was not reported. Frequency distributions have been plotted for number of correct responses, for age, sex groups and word classes but no attempt was made to transform the scores.

ii. Goldman-Fristoe-Woodcock test of Auditory Discrimination (1970)

This test was intended to assess a subject's ability to distinguish between various speech sounds, presented with and without background interference. The major variable considered in forming the standardization group was age, the range being from 3 yrs to 84 yrs. Some attention was given to geographical region with the samples being drawn from unspecified location. While no mention is made of the hearing

Sensitivity of the preschool children, it is noted that school age children or adults with known moderate or severe hearing losses were excluded from the sample. The hearing of most adults over 60 was screaned. The standardization sampl was composed of 745 subjects but the number of subjects at the various age levels is not clear. In many instances, the numbers were clearly small, for example there were only 6 subjects at the level of 3 years 8 months to 3 years 11 months. The exact method of selecting subjects is not described. The time period within which the data were gathered is also not indicated. Only age and sex are considered to have received serious attention in this standardization. While there were fewer cases at each age level than seems desirable, the matter of sex group comparison seems to have been reasonably handled. Dialect, intelligence level and socioeconomic level were all ignored. Subjects from only 3 regions were included in the sample. Finally the method of selection of subjects remains undescribed, apparently any available subjects were taken. The possibility for biased selection is obvious.

APPENDIX

In order to have an understanding of all statistical methods and their procedure of calculation, a single set of data has been worked out. All the possible and the frequently calculated statistical measures have been derived. Questions are stated and answers are drawn using appropriate statistic. The calculation of familiar descriptive statistical measures have been found out without going into their detailed procedures.

Table: the following are the obtained discrimination scores by 3 age groups of children on 3 conditions of time compressed speech (0%, 3% and 60%) when PBK-50 was presented at 32 dBSL.

		Percentage of time compression		
S.No.	Age(yrs)	0%	30%	60%
1	4	76	72	52
2	4	74	70	54
3	4	72	74	50
4	4	70	68	52
5	4	74	60	54
6	4	76	64	50
7	4	72	68	48
8	4	70	66	50
9	4	76	70	52
10	4	70	68	54
11	6	88	80	64
12	6	86	82	62
13	6	84	78	66
14	6	82	80	60
15	6	86	76	58
16	6	88	78	62
17	6	84	80	64
18	6	86	76	66
19	6	82	78	64
20	6	84	70	60

S.No.	Age(yrs)	Percentage of time compression 0%	 30%	 60%
21	8	94	84	74
22	8	92	86	70
23	8	90	88	72
24	8	96	82	74
25	8	98	80	76
26	8	92	84	72
27	8	88	80	76
28	8	90	78	68
29	8	98	86	72
30	8	96	82	70
	Total:	2514	2287	1866
	Mean:	83.8	76.3	62.2

Note: 0% time compression will be referred to as condition A
30% time compression will be referred to as condition B
60% time compression will be referred to as condition C

Expressions of symbols used: Variable X

SX - sum of the scores
SX^2 - sum of the squares of scores
$(SX)^2$ - square of sum scores
X - mean score
X^2 - squared mean score

Variable Y

SY - sum of the scores
SY2 - sum of the squares of scores
$(SY)^2$ - square of sum scores
Y - mean score
Y^2 - squared mean score
N - total number of children

Calculation of Descriptive statistics:

1. Mean ($\overline{?}$): Question: What are the speech discrimination scores obtained by an average child, under 3 conditions of time compression?

   Answer: An average child has obtained a mean discrimination score of

   83.8% under condition A

   76.3% under condition B

   62.2% under condition C

2. Median: Question: What percentage of discrimination score did a typical child get under condition of A, B and C?

   Answer: Median is the score midway between N/2 score and N/2 plus 1 score when they are ordered from low to high scores.

   A typical child gets – 85% median score under condition A

   78% median score under condition B

   63% median score under condition C

The median value approximates the mean values under all 3 conditions.

3. Mode: Question: What are all percentage discrimination scores typical of each condition?

   Answer: Mode is the score which most often occurred under each condition. Since this is not an appropriate descriptive statistic, it has not been worked out.

4. Range: Question: What range of discrimination scores were obtained under conditions of no time compression, 30% time compression and 60% time compression?

   Answer: Range is the difference between lowest score and highest score

   Range of discrimination score is:

   28% under condition A

   28% under condition B

   28% under condition C

This shows that scores were equally fluctuating under all conditions.

5. Standard Deviation (S):

Question: How far children's discrimination score differed from each other's score under each condition of time compression?

Answer: S

Children's discrimination score differed equally under conditions A and C and the difference was greater under condition B

6. Spearman Rank order Correlation coefficient (r_s):

Question: How similar were the ordering of 4 yr old children with regard to their discrimination score under condition B and C?

Procedure: 1. Assign ranks to children with regard to their discrimination score under both condition

2. find the difference between each child's two ranks (d1)

3. square this (d1) values and find the sum

4. calculate coefficient using the formula

$$r_S = 1 - \frac{6Sd12}{N^3-N} \qquad \text{Here, } = 1 - \frac{6 \times 161}{10^0-10}$$

$$= 1 - \frac{966}{990}$$

$$= 0.05$$

<u>Interpretation</u>: Rank order coefficient usually ranges from -1.00 to +1.00. If the value is close to 0.00, it is considered to be weak association.

The obtained value here (0.03) shows a very weak similarity between discrimination scores of condition B and C for the 4 yr old children, that is no rank correlation is seen.

7. <u>Pearson Product Moment correlation</u> (r):

<u>Question</u>: Is there any relationship between discrimination scores of children under conditions A and B?

<u>Procedure</u>: 1. Compute SX (condition-A) SX^2 and SXY

SY (condition-B) SY^2

2. for the present data,

SX	2514	SX2	212988
SY	2287	SY2	182716
		SXY	193304

3. calculate the value r using the formula

r =

4. Here,

<u>Interpretation</u>: The r also vary from -1.00 to +1.00. These 2 values indicate the degree and direction of relationship. The obtained r (0.376) indicate a weak relationship between the scores of children under condition A and B.

8. Contigency Coefficient(c):

Question: Were the children who obtained high discrimination scores under condition A the same ones who obtained the high discrimination score under condition B, and vice versa?

Procedure: 1. Assign each of the 30 children to one of the 4 cells of a 2 x 2 table given below

Condition A

		Below median	Above Median
ConditionB	Below Median	4 a	13 b
	Below Median	11	2 d

2. compute the value of chi square (X^2) for the frequencies in the table

$$?^{?} \quad ? \quad \frac{? \; ? \; ?? \; ? \; ? \; ? \; ? \; ?? \; \frac{?}{?} ?^{?}}{?? \; ? \; ? \; ??? \; ? \; ? \; ??? \; ? \; ? \; ????? \; ? \; ?}$$

$$= \frac{? \; ? \; ?? \; ? \; ? \; ? \; ? \; ? \; ? \; ??}{??? \; ?? \; ? \; ?? \; ? \; ??} \qquad = \qquad = \; 8.12$$

3. contingency coefficient $c = ?\overline{\frac{?^{?}}{? \; ?^{?}}}$

$$= ?\overline{\frac{?G?^{?}}{?? \; ?G?^{?}}} \qquad =0.68$$

Interpretation: The possible values of C range from 0.00 to 1.00. The obtained value 0.68 shows a close relationship between the discrimination scores of 2 conditions and reveals the general tendency that the children those who obtained high scores under condition A obtained high score under condition B and vice versa.

9. Phi-Coefficient: (? ?

Question: as in contingency coefficient.

Procedure: steps are same as in c but only the formula differs

$$? \quad ? \quad \frac{?? \, ?? \, ??? \, ?}{? \, \sqrt{??? \; ? ???? \; ? ???? \; ? ??? ? \; ? ?}}$$

Substituting the values from the table = $\frac{???}{???G}$ = 0.61

Interpretation: Values ranges from -1.00 to +1.00, weak association between variables when value is close to 0.00. Since the obtained values are more close to 1.00, we can draw same interpretation as in c.

Calculation of Inferential statistics: Nominal Data

1. Chi square test (X^2):

Question: Is there any tendency for the children who obtained discrimination score above the median under condition A, also to obtain discrimination score above median under condition B?

Procedure: 1. choose probability level of 95% or alpha level of 0.05

2. calculate chi square as in contingency coefficient

3. the previously obtained value is 8.12

Interpretation: Since we are dealing both positive with and negative relationships the test for rejection of null hypothesis would be a two-tailed test. The degrees of freedom would be 1.

By refering to the X^2 interpretation table (p.359, Silverman, 1977) it was found that the value of chi square required for rejection at 0.05 level when df is 1 is 3.84. Since the value obtained here is larger than this, the null hypothesis will be rejected, Showing a definite tendency for the children who obtained discrimination score above the median under condition A also to obtain discrimination score above median under condition B.

2. Sign test:

Question: Did the 8 yr group children tend to obtain high discrimination score under condition B than under condition C?

Procedure: 1. choosing an alpha level or probability level of 0.05

2. assign to each of 10 children at or a- depending on the discrimination score under condition B or under condition C is higher

3. assign + when discrimination score (B) is higher than discrimination score (C), assign – when it is reverse, and assign \`O' when both are equal.

4. find the number of children for whom the two discrimination scores are not the same (N) and the number of times the sign occurs (+ or -) which occurs least often (x) are determined. Here, the values are 10(N), 10(x).

Interpretation: The test for rejecting the null hypothesis would be one tailed because we are finding the difference in one direction. Based on binomial distribution when N is 10, the probability of 10 pluses(x) occurring by chance is 1.0 (p.320, Silverman, 1977). Since this value is larger than 0.05, the null hypothesis would not be rejected concluding that no tendency for the 8 yr old children to obtain high discrimination score under condition B than under condition C.

We can make use of the same sign test oto answer the following question also. Did the 4yr, 6yr and 8yr old children tend to obtain high discrimination score under condition B than under condition C?

3. Mann-Whitney U test:

Question: Did the typical 6 yr old children have a high discrimination score under condition B than did the typical 4yr and 8yr old children?

Procedure:
1. Here the null hypothesis rejection is based on 0.05 level and one tailed test.
2. assign ranks to all 30 children for the discrimination scores under condition B
3. find the sum of ranks for 6 yr olds (R1) and the combined sum for 4th and 8yr old children (R2).
4. for the present set of data R1 = 146.5 & n1 = 10

   R2 = 318.8 & n2 = 20
5. compute the value using the formula

$$U = a)\ n1n2+\frac{?\,?\,??\,?\,?\ \,?\,?}{?} - R1 = 108.5$$

$$b)\ n1n2+\frac{?\,?\,??\,?\,?\,?\,?}{?} - R2 = 91.2$$

6. use the lowest value of U (91.2) to compute the value of Z, the test statistic.

$$Z = \frac{?\,?\,\frac{?\,?\,?\,?}{?}}{\frac{?\,\overline{??\,?\,?\,?\,???\,?\,?\,?\,??\,??}}{??}} = \frac{??\,G\,?\,?\,?\,?}{?\,?\,?}$$

$$= 0.359.$$

Interpretation: Based on the normal distribution (p.321, Silverman, 1977), the probability of this outcome occurring by chance is 0.359, since this value is larger than 0.05, the null hypothesis will not be rejected, retaining that the typical 6 yr old children did not have a high discrimination score under condition B than did the typical child of 4yr or 8yr old.

4. Friedman Two-way analysis of variance (X_r^2):

Question: Did the 4yrold children tend to get high discrimination score under any one condition than the other conditions?

Procedure: 1. choose the alpha level of 0.05

2. assign ranks to the discrimination score under each of the 3 conditions separately

3. find total ranks for each condition (Rj) here,
for condition A - 55
for condition B – 55
for condition C – 55

4. compute the value of Xr^2 using the following formula

N- number of subjects – 10

k- number of variables- 3

$$Xr^2 = \frac{??}{? ? ??? ??} ?? \overset{\frac{?}{?}}{?} (Rj)^2 - 3N (k+1)$$

$$= \frac{??}{???}(55^2+55^2+55^2) - 120$$

$$= 787.5$$

Interpretation: To determine whether this value of X_r^2 is larger enough for the null hypothesis to be rejected, the value of chi square required for rejection is located in the table (p.319, Silverman, 1977) for df of 2. The value for rejection at 0.05 level is 5.99. Since the obtained value is larger than this, the null hypothesis would be rejected, the alternate hypothesis that the 4yr old children tend to get high discrimination scores under any one or some condition than the other.

5. Kruskal-Wallis one-way analysis of variance (H):

Question: Is there a tendency for the amount of discrimination score under condition (or variable) B obtained by children to vary as a function of age level?

Procedure: 1. alpha level choosen is 0.05 level

2. give ranks to all 30 children on the basis of their discrimination score under condition B

3. find Rj for each age level, which is the sum of the ranks assigned to that age level

here, Rj- for 4yr is 232

for 6yr is 146.5

for 8yr is 66.5

4. then compute H using the formula

$$H = \frac{??}{? ?? ? ??} \; \bullet^{?} \; ?? \; ? \; \frac{? ?^{?}}{??} \; - \; 3(N+1)$$

Here N- total number of subjects (30)
nj- subjects under each age (10)
k- number of column (3)

$$H = \frac{??}{???} \left(\frac{???^{?}}{??} \; ? \; \frac{???? ?^{?}}{??} \; ? \; \frac{??? ?^{?}}{??} \right) - 93 = 9.66$$

<u>Interpretation</u>: It is distributed as chi square, provided the number of groups is more than 3. By looking into the same table for (p.319, Silverman, 1977) df of 2, the value of chi square required for rejection at 0.05 level is 5.99. Since the obtained value is larger than this the null hypothesis would be rejected, showing a definite tendency for the amount of discrimination score under condition B obtained by children to vary as a function of age level.

<u>Inferential Statistical tests</u> (interval and ratio data)

1. <u>t test for related measures</u> (t):

<u>Question</u>: Did the children on the average tend to get high discrimination score under condition B than discrimination score under condition C?

<u>Procedure</u>: 1. 0.05 alpha level is selected. Use one tailed test for rejecting null hypothesis

2. subtract discrimination score under condition B fro condition C for each child(D), find SD, D, SD^2, D^2.

3. for the present data. SD-444; ? -14.8; SD^2-7144 and D^2-219

4. compute t using the formula

$$t = \; ? \; \frac{? \overline{? ? ?}}{\overline{? \frac{??^{?}}{?} ? \quad ?^{?}}} \quad = \quad \frac{??G ? ?G?}{?G?} \quad = \quad 16.22$$

Interpretation: By checking the t value required for rejecting the null hypothesis (p.322, Silverman, 1977) at 0.05 level, for a one tailed test, having df of (N-1) 29 it's value is found to be 1.69. Since computed value is larger than this value, the null hypothesis is givenoff accepting that the children on the average tend to get high discrimination score under condition B than under condition C.

2. t test for independent measures:

Question: Did the 6 yr old children tend to get higher discrimination score under condition C on the average than the 4 and 8yr old children?

Procedure: 1. Select 0.05 level and one tailed test of testing hypothesis

2. compute the following for 6yr olds

$$\text{SX1; X1; SX?}^{?}\text{ ? S1}^{2} = \frac{???^{?}}{??} - \text{X1}^{2}$$

3. in the same way the combined values for 4 and 8 yr olds

$\text{SX2; X2; SX2}^{2}\text{; S2}^{2}$

4. for the present data the values are SX1-626; X1-62.6; SX1^{2}-39252; S1^{2}-6.44; SX2-1240; X2-62; SX2^{2}-57664; S2^{2}- -1039.2

5. use the following formula

$$t = \frac{\bar{?}??\bar{?}?}{?\frac{\overline{????^{?}?????^{?}}}{???????}\quad ?\frac{?}{??}?\frac{?}{??}?} = \frac{?G}{?\overline{???G}} = 0.057$$

Interpretation: By checking t value requi-red for rejecting the hypothesis (p.322, Silverman, 1977) at 0.05 level, for a ne tailed test having a df of (N-2) 28, it's value is found to be 1.701. since the computed value is smaller than this the null hypothesis would not be rejected, leading to the conclusion that 6 yr old children do not obtain high discrimination score under condition C on the average than the 4 and 8yr old children.

3. <u>t test for significance of Pearson product moment correlation</u>

<u>Question</u>: Is there any significant relationship between discrimination scores of children under conditions A and B?

<u>Procedure</u>: 1. A two-tailed test at 0.05 level is applied to check the hypothesis

2. as the r values has already been calculated, the calculation of t would be using the following formula (r-0.376)

$$t = \frac{?}{?\overline{??\ ?^{?}}} \text{ x } ?\overline{?\ \ ?\ \ ?} \quad = \quad \frac{?G??}{?G??} \text{ x } 5.29 = 2.15$$

<u>Interpretation</u>: By checking the t value required for rejecting the hypothesis (p.322, Silverman, 1977) for two-tailed test, having df of (N-2) 28, it's value is 1.701. Since the computed value is larger than this value, the null hypothesis is rejected, emphasizing a stronger relationship between discrimination score under condition A and B.

<u>Estimating Population Values</u>:

1. <u>Mean</u>: <u>Question</u>: Within what range the mean discrimination score of the population from which the children were selected must likely to fall under condition A?

<u>Procedure</u>: 1. Usually a 95% confidence interval is calculated.

2. compute the upper (?) and lower (__) limits of the confidence interval using the formula

$$? = ? + \frac{??}{?\overline{?\ ?\ ?}}\ (1.96)$$

$$\_\_ = {}^{-} - \frac{??}{?\overline{???}}\ (1.96)$$

for the present data, the values X and SD are 83.8 and 8.78 respectively.

$$\overline{} = 83.8 + \frac{?G?}{?G?} \text{ x } 1.96 = 86.99$$

$$\overline{} = 83.8 - \frac{?G?}{?G?} \text{ x } 1.96 = 80.60$$

<u>Interpretation</u>: The mean discrimination score of the population from which the children were selected is approximately 95% certain to fall between 80.6 and 86.99 under condition A.

2. <u>Median</u>: <u>Question</u>: Within what range is the median discrimination score of the population from which children were selected most likely to fall?

<u>Procedure</u>: 1. Calculate 95% confidence interval using the formula given below

2. upper limit $\overline{?} = \text{mdn} + \frac{?G???}{?\overline{???}} (1.96)$

Lower limit $\underline{?} = \text{mdn} - \frac{?G???}{?\overline{???}} (1.96)$

3. here the median is 85 and SD is 8.78

4. So $\overline{?} = 85 + \frac{?G?}{?G?} \text{ x } 1.96 = 88.2$

$\underline{?} = 85 - \frac{?G?}{?G?} \text{ x } 1.96 = 81.8$

<u>Interpretation</u>: The median discrimination score of the population from which the children were selected is approximately 95% certain to fall between 81.8 and 88.2 under condition A.

BIBLIOGRAPHY

Beasley, D.S and Beasley, D.C. " Auditory reassembly abilities of black and white first and third grade children," Journal of Speech and Hearing Research, 16, 1973, 213.

Cozad, R.L., " Speech reading skill and communication difficulty of children and young adults with unilateral hearing loss," Journal of auditory Research, 17, 1977, 25.

Egan, J.P., " Articulation testing methods. II," OSRD Report, 3802, 1944 cited in J.Katz (ed.) Handbook of Clinical Audiology, 2nd edn. Baltimore: Williams & Wilkins Company, 1978.

Epstein, A and Hopkinson, N.T., " An investigation into the clinical validity of the Doerfler-Stewart test in the determination of the non-organic hearing loss," Paper presented at the convention of the ASHA, Nov. 1956, cited in J.Eata (ed.) Handbook of Clinical Audiology, 2nd edn. Baltimore: Wiliams & Wilkins Company, 1978.

Fairbanks, G1, Guttman, N and Miron, M.S., " Effects of time compression upon the completion of connected speech," Journal of Speech and Hearing Disorders, 22, 1957, 10.

Feldman, R.M and Reger, S.N., " Relations among hearing reaction time and age," Journal of Speech and Hearing Research, 10, 1967, 479.

Feldt, L.D., " The Latin square design in speech and hearing research," Journal of Speech and Hearing Research, 2, 1959, 216.

Fisher, R.A., The design of experiments, 7th edn. Edinburgh: Oliver and Boyd Ltd., 1951.

Garrett, H.E., Statistics in psychology and education, New York: David McKay Company, Inc., 1979.

Guilford, J.P., Fundamental statistics in psychology and education, New York: McGraw-Hill Book Company, 1978.

Harford, E.R and Jerger, J.F., " Effect of loudness recruitment on delayed speech feedback," Journal of Speech and Hearing Research. 2, 1959, 361.

Hirsh, I.J.et.al., " Development of materials for speech audiometry," Journal of Speech and Hearing Disorders. 17, 1950, 321.

Jerger, J., Speaks, C and Trammell, J.A. " A new approach to speech audiometry," Journal of Speech and Hearing Disorders, 33, 1968, 318.

Katz, J., " the SSW test- an interim report," Journal of Speech and Hearing Disorders, 33, 1968, 132.

Kerlinger, F.N., Foundation of Behavioral research, New York: Holt, Rinehart & Winston, Inc., 1964.

Eronvall, E.L and Diehl, B.F., " The relationship of auditory discrimination to articulatory defects of children with known organic impairment," Journal of Speech and Hearing Disorders. 19, 1954, 335.

Lindquist, E.F., Design and analysis of experiments in psychology and education, Boston: Houghton Eiffin company, 1953.

McCoy, C.et.al., " Effects of age and sex on dichotic listening: The SSW test," Journal of Auditory Research, 17, 1977, 263.

McLeod, A.L and Greenberg, H.J, " Relationship between loudness discomfort level and acoustic reflex threshold for normal and sensorineural loss group," Journal of Speech and Hearing Research, 22, 1979, 873.

Miller, S., Experimental design and statistics, Essential psychology series, London: Methuen & Co,Ltd., 1975.

Monour,J.P and Dirks,D. " Binaural and Monaural speech intelligibility in reverberation," Journal of Speech and Hearing Research, 10, 1957, 186.

Nikam, S., Beasley,D.S and Rintlemann, W.F.," Perception of time-compressed CNC monosyllables by non-native speaker/listeners of English," Journal of American Auditory Society, 2, 1976, 45.

Orchik,D.J.et.al., " Time compressed speech discrimination, reading readiness and the effect of clinical method," Audiology,18, 1979, 80.

Oyer,H.J and Beasley, D.S., " Speech and hearing science as related to clinical audiology," Journal of All India Institute of Speech and Hearing, 4, 1973, 67.

Pederson, C.B, " Brief tone audiometry in persons treated with salicylate," Audiology, 13, 1974, 311.

Perozzi,J.A and Kunze,L.H. " Relationship between speech sound discrimination skills and language abilities," Journal of Speech and Hearing Research, 14, 1971, 382.

Peterson,G.E and Fairbanks, G. " Speech and Hearing science," ASHA, 5, 1963, b39.

Plutchik,R., Foundations of experimental research, New York: Harper & Row publishers, 1965.

Prins,D., " Relations among specific articultory deviations and responses to a clinical measure of sound discrimination ability," Journal of Speech and Hearing Disorders, 28, 1963, 382.

Robson, C. Experiment, Design and statistics in psychology, New York: Penguin Books Lts. 1973.

Shapiro,I. " Evaluation of relationship between hearing threshold and loudness discomfort level in sensorineural loss," Journal of Speech and Hearing Disorders, 44, 1979, 31.

Siegenthaler, B.M. " a study of relationship between measured hearing loss and intelligibility of selected words," Journal of Speech and Hearing Disorders, 14, 1949, 111.

Siegel,S., Nonparametric statistics for the behavioral sceiences, Tokyo: McGraw-Hill, Inc. 1956.

Silverman-Dresner, T., Guilfoyle, G.R., Vocabulary norms for deaf children, Washington, D.C: Alexander Graham Bell Association for the Deaf, Inc., 1972.

Silverman, F.H. Research design in Speech pathology and audiology, New Jersy: Prentice-Hall, Inc., 1977.

Simhadri, R. " Effect of familiarity on the performance of subjects in a competing message task," Unpublished Master's Dissertation, University of Mysore, 1977.

Smith, W.H., " The effects of hypnosis and suggestion upon auditory threshold," Journal of Speech and Hearing Research, 10, 1969, 161.

Stafford,K. " Problem solving ability of deaf and hearing children," Journal of Speech and Hearing Research, 5, 1962, 169.

Sticht, T.G and Gray, B.B. " The intelligibility of time compressed words as a function of age and hearing loss," Journal of Speech and Hearing Research, 12, 1969, 443.

Tossi, O. " Acoustic Phonetics" Handout for Ph.D course, Michigan State University, 1977.

Vasudha, A.R. " IQ variation consequent to speech and language therapy among hard of hearing children," Unpublished Master's Dissertation, University of Mysore, 1972.

Ventry, I.M. " Effects of conductive hearing loss: Fact or Fiction," Journal of Speech and Hearing Disorders, 45, 1980, 143.

Weiner, P.S and Hoock, W.C. " The standardization of tests:criteria and criticisms," Journal of Speech and Hearing Research, 16, 1973, 616.

Williams, F. Reasoning with statistics: simplified examples in common research, New York: Holt, Rinehart & Wington, Inc., 1968.

Winer, B.J. Statistical principles in experimental design, New York: McGraw-Hill Book Company, 1962.