BASIC MEASUREMENT AND SEALING TECHNIQUES

In every usage, measurement occurs when an established yardstick verifies the height, weight or another feature of a physical object. How well you like a song, a painting, or the personality of a friend is also a measurement. In a dictionary sense, to measure is to discover the extent, dimensions, quantity, or capacity of something, especially by comparison with a standard. We measure casually in daily life, but in research the requirements are rigorous.

Measurement in research consists of assigning numbers to empirical events in compliance with a set of rules.

This definition implies that measurement is a three-part process:

I. Selecting observable empirical events
II. Using numbers or symbols to represents aspect of the events, and
III. Applying a mapping rule to connect the observable to the symbols

Types of Scales

A scale is a device for measuring magnitude or quantity of a variable. Scales may be a series of steps, degrees, a scheme of graded amount from the highest to lowest, an indicator of relative size; scales may also designate appropriate categories such as age, sex, etc.;

There are four types of scales commonly used as levels of measurements.

a) Nominal scale

In business and social research, nominal data are probably more widely collected than any other. When you collect nominal data, you partition a set into categories that mutually exclusive and collectively exhaustive.

In this type of scale, the numbers serve only as labels or tags for identifying objects, events or characteristics. For instance, a person identity card number is a nominal scale. It only serves the function of identifying the person. We can assign numbers to football players, telephone subscribers or to products in a storeroom. These numbers or codes have no mathematical implication, and the only property conveyed by the numbers is identity. Arithmetic operations cannot be performed on these numbers, as they would have no meaning.

The only permissible mathematical operation in nominal scales is those leased upon counting such as frequencies, modes and percentages.

There are three forms of nominal scales:

(i) label nominal
(ii) category nominal scale

(iii) mixture nominal scale

I) Label nominal scale: This is the most elementary nominal scale. A label nominal scale is simply a label assigned to an object in order to identify and keep track of it. In this kind of scaling each label is unique to one object and possesses no meaning in itself.

ii) Category nominal scale: This is the most commonly used nominal scale in market research. In category nominal scale, numbers are used to represent mutually exclusive and exhaustive categories of objects. Thus, one might classify the residents of a city according to their expressed religious preferences. Classification set A given in table 8.1 is not a sound category nominal scale because it is not collectively exhaustive. Set B meets the minimum requirements, although this classification may be more useful for some research purposes than others.

Table 8

A	B
Baptist	Catholic
Catholic	Jewish
Jewish	Protestant
Lutheran	Other
Methodist	None
Presbyterian	
Protestant	

Thus each category must be assigned to one, and only one scale category, and must possess the measured common characteristics. Other examples of characteristics measured with category nominal scale include sex, tribe and so forth. For instance, in a given study men may be coded '1' and women '2' and this serves no other function apart from classification.

iii) Mixture nominal scale: This is a nominal scale which is partially a label. The numbers and labels assigned football players serve to identify the individual players, and also to place players in a category.

(b)Ordinal scales

This is a qualitative scale comprised of equal appearing intervals that rank observations from large too small. This scale indicates rank order only. It does not indicate the nature of the intervals between the ranks. For example, if several soft drinks are scaled according to sweetness, and number 1 represents the highest degree of sweetness, then

the drinks assigned number 3 would be sweeter than one assigned number 4 but less sweet than one assigned number 2.
Note that with ordinal scale the only permitted statements are of greater than or less than nature. We cannot make statements about how much less of characteristics one object posses relative to another.
Ordinal measures commonly have only three to five categories, i.e. well, better, best or:

Excellent	Very good	Average	Below average	very poor

Or

Strong agree	Agree	No opinion	Disagree	Strongly disagree

In dealing with ordinal scale, statistical description to positional measures such as median, quartile, percentile or other summary characteristics which deal with order among quantities.

c) Interval scales

Interval scale has the power of nominal and ordinal scale plus one additional strength: it incorporates the concept of equality in interval (the distance between 1 and 2 equals the distance between 2 and 3). The interval is known and equal. They can be added, subtracted and their summaries can be subjected to statistical tests. The interval scale does not have an absolute zero. The zero point of this scale is arbitrary, but it permits inferences to be made.
One common example of the interval scaling is the Fahrenheit and centigrade scales used to measure temperature. An arbitrary zero is assigned to each scale, and equal temperature differences are found by scaling equal volumes of expansion in the liquid used in the thermometer.

Interval scales permit inferences to be made about the differences between the entities to be measured (warmness); but we cannot say that any value on a specific interval scale is multiple of another. Thus a temperature of 50o F is not twice as hot as a temperature of 25o F. Also, the elapse time between 3 and 6 am equals the time between 4 and 7 am, but one cannot say 6 am is twice as late as 3am.

When a scale is interval, you use the arithmetic mean as the measure of central tendency. You **can** compute the average time of first arrival of trucks at a warehouse. The standard deviation is the measure of dispersion for arrival time. Product moment correlation, t-tests, and F-test and other parametric tests are the statistical procedures of choice.

d) Ratio Scale

This is the highest level of measurement among scales. It incorporates all the powers of the previous scales plus the provision for absolute zero or origin. Ratio scale represents the actual amounts of a variable.
Measure of physical dimensions such as weight, height, distance and are examples. In business research, we find ratio scales in many areas. These include money values, population counts, distances, return rates, productivity rates.

8.2 Sources of measurement differences

The ideal study should be designed and controlled for precise and unambiguous measurement of the variables. Since attainment of this ideal, we must recognize the sources of potential error and try to eliminate, neutralize or otherwise deal with them. Much potential error is systematic (result from a bias) while the remainder is random (occurs erratically).Seltiz C etal(1976) has pointed out several sources from which measured differences can come.

(I)The respondent as an error source

A respondent may be reluctant to express strong negative feelings or may have little knowledge about a personality i.e. the president, but be reluctant to admit ignorance. This reluctance can lead to an interview of 'guesses.

Respondent may also suffer from temporary factors like fatigue, boredom, anxiety or other distractions; these limit the ability to respond accurately and fully. Hunger, impatience, or general variations in mood may also have an impact.

(ii)Situational Factors.

Any condition that places a strain on the interview can have serious effects on interviewer-respondent rapport. If another person is present, that person can distort responses by joining in, by distracting, or merely by being present. If the respondents believe anonymity is no ensured, they may be reluctant to express certain feelings.

(iii)The measure as an error source

The interviewer can distort responses by rewarding, paraphrasing, or reordering questions. Stereotypes in appearance and action introduce bias. Inflections of voice and conscious or unconscious prompting with smiles, nods. And so forth may encourage or discourage certain replies; careless mechanical processing-checking of the wrong response or failure to record full tabulation, and faulty statistical calculation may introduce further errors.

(iv)Instrument as an error source

A defective instrument can cause distortion in two major ways. First, it can be too confusing and ambiguous. The use of complex word and syntax beyond respondent comprehension is typical. Leading questions, ambiguous meanings, mechanical defects (inadequate space for replies, response choice omissions, and poor printing), and multiple questions suggest the range of problems.

A more elusive type of instrument deficiency is poor sampling of the universe of content items. Seldom does the instrument explore all the potentially important issues.

The characteristics of sound measurement.

What are the characteristics of a good measurement tool? An intuitive answer to this question is that the tool should be an accurate counter or indicator of what we are interested in measuring. In addition, it should be easy and efficient to use. There are three (3) major criteria for evaluating a measurement tool:

- **Validity:** which refers to the extent to which a test measures what we actually wish to measure?
- **Reliability**: has to do with the accuracy and precision of a measurement procedure.
- **Practicality:** is concerned with a wide range of factors of economy, convenience, and interpretability.

Validity in research

Validity in research is achieved through the internal and external validity of the study.

Internal validity: This refers to the outcome of the study was based on function of the program; a study has internal validity if the outcome of the study is a function of the approach being tested rather than results of the causes not systematically dealt with. Internal validity is justified by the conclusions we have as researchers when we have been able to control the threats of other variables (i.e. intervening variables, or moderating variables or extraneous variables, or moderating variables or extraneous variables). The more you reduce the nuisance (other variables) affecting the study, the more you attain the internal validity.

There are three widely accepted classification of internal validity:

There are three widely accepted classification of internal validity:

(i) content validity

(ii) criterion- related validity
(iii) construct validity

(i) Content validity

The content validity of a measuring instrument is the extent to which it provides adequate coverage of the topic under study. If the instrument contains a representative sample of the universe of the subject matter of interest, then the content validity is good. To evaluate the content validity of an instrument, one must first agree on what element constitutes adequate coverage of the problem.

(ii) Criterion-Related Validity

Criterion-related validity reflects the success of measures used for prediction or estimation. You may want to predict an outcome or estimate the existence of a current behavior or condition.

These are *predictive and concurrent validity*, respectively. They differ only in a time perspective. An opinion questionnaire that correctively forecasts the outcome of a union election has predictive validity. An observational method that correctively categorizes families by current income class has concurrent validity. While these examples appear to have simple and ambiguous validity criteria, there are difficult in estimating validity. Consider the problem of estimating family income. There clearly is a knowable true income for the family. However, we may find it difficult to secure this figure. Thus, while the criterion is conceptually clear, it may be unavailable.

(iii) Construct Validity

One may also wish to measure or infer the presence of abstract characteristics for which no empirical validation seems possible. Attitude scales and aptitude and personality tests generally concern concepts that fall in this category. Although this situation is much more difficult, some assurance is still needed that the measurement has an acceptable degree of validity.

In attempting to evaluate construct validity, we consider both the theory and the measuring instrument being used. If we were interested in measuring the effect of ceremony on organizational culture, the way in which ceremony was operational defined would have to respond empirically grounded theory. Once assured that the construct was meaningful in a theoretical sense, we would next investigate the adequacy of the instrument. If a known measure of ceremony in organizational culture was available, we might correlate the results obtained using this measure with those derived from our new instrument. Such an approach would provide us with preliminary indications of *convergent* validity.

Reliability

Reliability means many things to many people, but in most contexts the notion of consistency emerges. A measure is reliable to the degree that it supplies consistent results. **Reliability** is a contributor to validity can be simply illustrated with the use of a bathroom scale. If the scale measures your weight correctly (using a concurrent criterion such as a scale known to be accurate), then it is both reliable and valid. If it consistently overweighs you by six pounds, then the scale is reliable but not valid. If the scale measures erratically from time to time, then it is not reliable and therefore cannot be valid.

Reliability is concerned with estimates of the degree to which a measurement is free of random or unstable error. It is not as valuable as validity determination, but it is much easier to assess. Reliable instruments can be used with confidence that transient and situational factors are not interfering. Reliable instruments are robust; they work well at different times under different conditions.

Factors affecting the interval validity of a study

Among the many threats to internal validity, we consider the following:

(a)History

During the time that a research is taking place, some events may occur that confuse the relationship being studied. These are events may either increase or decrease the expected outcomes of the project. These events which are not part of the project and they are not planned for. They may just happen in the research and have tremendous effects on the results of the study.

(b) Testing

The process of taking a test can affect the scores of a second test. The mere experience of taking the first test can have a learning effect that influences the results of the second test. Subjects who are given a pretest are likely to remember some of the questions or some of the errors they made when they are taking the posttest. They are also likely to do somewhat better on the posttest than they did on the pretest.

(c) Instrumentation

This threat to internal validity results from changes between observations, in measuring instruments or in observer. Using different questions at each measurement is an obvious source of potential trouble, but using different observers or interviewers also threaten

validity. Observer experience, boredom, fatigue and anticipation of results can all distort the results of separate observations. For example, an experienced interviewer may obtain more complete information from a correspondent than an inexperienced interviewer. The additional information may be due to the fact that the interviewer has become more skilled in asking questions or observing events and not due to the effect of the program or observing the effects of the treatment.

(d) Maturation:

Changes may also occur within the subjects that are a function of the passage of time and not specific to any particular event. These are of special concern when the study covers a long time, but they may also be factors in tests that are as short as an hour or two. A subject can become hungry, bored or tired in a short time, and these conditions can response results.

(e) Selection

An important threat to internal validity is the differential selection of subjects for experimental and control groups. Validity considerations require that groups be equivalent in every respect. If subjects are randomly assigned to experimental and control groups, this selection problem can be largely overcome. Additionally, matching the members of the groups on key factors can enhance the equivalence of the groups. Validity considerations require that the groups be largely overcome. Additionally, matching the members of the groups on key factors can enhance the equivalence of the groups.

(f) Experiment Mortality

This occurs when the composition of the study groups changes during the test. Attrition is especially likely in the experimental groups, and with each dropout, the group changes. Because members of the control group are not affected by the testing situation, they are less likely to withdraw. In a compensation incentive study, some employees might not like the change in compensation method and withdraw from the test group; this action could distort the comparison with the control group that has continued working under the established system, perhaps without knowing a test is under way.

Factors affecting the external validity of the study

Internal validity factors cause confusion about whether the experiment treatment (x) or extraneous factors are the source of observation differences. In contrast, external validity is concerned with the interactions of the experimental treatment with other factors and the resulting impact on abilities to generalize to times, settings or persons.

(a) Reactive effects of testing:
If pre-testing has been used and which sensitizes the experimental subjects to the particular treatment, then the effect of the treatment may be partially the result of the sensitization of the pre-test.

(b) Interaction effects of selection bias
If the samples draws from the study are not representative of the larger population, then it would be difficult to generalize findings from the samples to the population, and this may arise when the samples are not drawn randomly from the population. Consider a study in which you a cross-section of a population to participate in an experiment, but a substantial number refuse. If you do the experiment only with those who agree to participate, can the results be generalized to the population?

(c) Other reactive factors
- The experimental settings themselves may have a biasing effect on a subject's response to the treatment. An artificial setting can obviously give results that are not representatives of large populations. Suppose workers who are given an incentive pay are moved to a different work area to separate them from the control group. These new conditions alone could create a strong reaction condition.
- If subjects know they are participating in an experiment, there may be a tendency to role-play in a way that distorts the effect of the experimental treatment.

Common effects related to the research process
There are other situations in which the internal and external validity of the study may both be threatened simultaneously. This is brought about by what we call research effects, which have nothing to do with the treatment.

1. **Hawthorn Effect**

This refers to a situation where subject awareness of being in an experimental group motivates them to perform better. Therefore, the most influential factor on the subjects is not the independent variable but their awareness of being in a special group.

2. **The placebo Effect**

This is common to medical studies. Research observes that a drug administered to any group of parties has two effects.

(i) Chemical effect
(ii) Psychological effect

To counteract this effect, researchers use a placebo and this is an inactive substance which has the same color and tests as the active drug and the other Alf (control group) are given the placebo inactive drug.
If there is a significance difference between those two groups, the drug may be said to have a significance effect.

Review Questions

1. What are the essential differences among nominal, ordinal, interval and ratio scales?
2. You have data from a corporation on the annual salary of each of its 200 employees.
    a) Illustrate how the data can be presented as ratio, interval, ordinal, and nominal data.
    b) Describe the successive loss of information on the presentation changes from ratio to nominal.

3. Briefly explain validity in research

1.

10.0 Analysis and Presentation of Data

Once the data begin to flow in, attention turns to data analysis. The steps followed in data collection influence the choice of data analysis techniques. The main preliminary steps that are common to many studies are:

- ✓ Editing
- ✓ Coding and
- ✓ Tabulation

(i) Editing

Editing involves checking the raw data to eliminate errors or points of confusion in data. The main purpose of editing is to set quality standards on the raw data, so that the analysis can take place with minimum of confusion. In other words, editing detects errors and omissions, corrects them when possible and certifies that minimum data quality standards are achieved. The editor's purpose is to guarantee that data are:

- Accurate
- Consistent with other information
- Uniformly entered
- Complete and

- Arranged to simply coding and tabulation.

In the following questions asked of military officers, one respondent checked two categories, indicating that he was in the reserves and currently serving on active duty.

_____
Please indicate your current military status:
☐ Active duty officer
☐ National Guard officer
☐ Reserve officer
☐ Separated officer
☐ Retired officer

The editor's responsibility is to decide which of the responses is consistent with the intent of the question or other information in the survey and is most accurate for this individual respondent. There are two stages in editing;

(i) the field edit
(ii) the central office edit

(i) The Field Edit

Field edit is the preliminary edit whose main purpose is to detect the obvious inaccuracies and omissions in the data. This also helps the researcher to control the fieldworkers and to clear misunderstanding of the procedures and of specific questions.

The best arrangement is to have the field edit conducted soon after the data collection form has been administered. The following items are checked in the field edit:

a) *Completeness:* Checking the data form to ensure that all the questions have been answered. The respondent may refuse to answer some questions or may just not notice them.
b) *Legibility:* A questionnaire may be difficult for others to comprehend but the interviewer can easily clarify it, if asked in good time.
c) *Clarity:* A response may be difficult for others to comprehend but the interviewer can easily clarify it, if asked in good time.
d) *Consistency:* The responses provided may also lack consistency. These can be corrected by the fieldworker if detected early.

In large projects, field editing review is a responsibility of the field supervisor. It, too, should be done soon after the data have been gathered. A second important control

function of the field supervisor is to *validate the field results.* This normally means s/he will re-interview some percentage of the respondent, at least on some questions.

(ii) Central Editing

This comes after edit, At this point; data should get a thorough editing. Sometimes it is obvious that an entry is incorrect, is entered in the wrong place or states times in months when it was requested in weeks. A more difficult problem concerns faking. Arm chair interviewing is difficult to spot, but the editor is in the position to do so. One approach is to check responses to open-ended questions. These are the most difficult to fake. Distinctive response patterns in other questions will often emerge if faking is occurring. To uncover this, the editor must analyze the instruments used by each interviewer.

1. **Coding**

   It involves assigning numbers or other symbols to answers so the responses can be grouped into a limited number of classes or categories. The classifying of data into limited categories sacrifices some data detail but is necessary for efficient analysis. Instead of requesting the work male or female in response to a question that asks for the identification of one's gender, we could use the codes 'M' or 'F'. Normally, this variable would be coded 1 for male and 2 for the female or 0 and 1.

   Coding helps the researcher to reduce several thousand replies to a few categories containing the critical information needed for analysis.

   The first step involves the attempt to determine the appropriate categories into which the responses are to be placed. Since multiple choice and dichotomous questionnaire have specified alternative responses, coding the responses of such questions is easy. It simply involves assigning a different numerical code to each different response category.

   Open questions present different kinds of problems for the editors. The editor has to categorize the responses first and then each question is reviewed to identify the category into which it is to be placed. There is a problem in that there can be a very wide range of responses, some of which are not anticipated at all. To ensure consistency in coding, the task of coding should be apportioned by questions and not by questionnaires. That is, one person may handle question one to six, in all the questionnaires instead of dividing the coding exercise by questionnaires. The next step involves assigning the code numbers to the established categories.

For example, a question may demand that a respondent lists the factors s/he considers when buying a pair of shoes. The respondent is free to indicate anything s/he thinks of. The responses may range from color, size, comfort, price, materials, quality, durability, style, uniqueness and manufacturer among others. The response may have to be coded into just three or four categories and each response has to be placed within a specific category and coded as such.

The 'don't know' (DK) response present special problems for data preparation. When the DK response group is small, it is not troublesome. But there are times when it is of major concern, and it may even be the most frequent response received. Does this mean the question that elicited this response is useless? It all depends. But the best way to deal with undesired DK answer is to design better questions at the beginning.

2. Tabulation

This consists of counting the number of responses that fit in each category. The tabulation may take the form of simple tabulation.

- Simple tabulation involves counting a single variable. This may be done for each of the variables of the study. Each variable is independent of the others.
- In gross tabulation two or more variables are handled simultaneously. This may be done by hand or machine.

Where hands tabulation is used. A tally sheet is used. For example, if the question read:
How many cigarettes do you smoke in a day?
The tally for a sample of size 35 would look like this:

Classes	Code	Tally	Frequency
0	1	////	4
1-5	2	//// //	7
6-10	3	//// //// ///	13
11-15	4	//// //	7
15 and over	5	////	4

Cross tabulation can be created when we combine the number of cigarettes smoked in a day with the age of the respondent

This can be done to establish the relationships between the number of cigarettes smoked and age. The table below shows the cross tabulation for the variables.

Age of respondents						
No of cigarettes smoked	15-20	21-25	26-30	31-35	36+	
1. 2. 3. 4. 5.	//'	/		/		4
Totals	6	6	5	13	5	35

The cross tabulations indicate for example, that all the respondents smoking more than 5 cigarettes a day are in the 36 and over years of age category.

This kind of tabulation is only useful in very simple studies involving a few questions and a limited number of responses. Most studies involve large numbers of respondents and many items to be analyzed and these generally rely on computer tabulation. There are many packed programmes for studies in the social sciences.

A note on the use of summary statistics.

Researchers frequently use summary statistics to present survey findings. The most commonly used summary statistics include:

- ✓ measurers of central tendency (mean, median and mode)
- ✓ Measures of dispersion (variance, standard deviation, range, interquartile range).
- ✓ Measures of shape (skewness and kurtosis)
- ✓ We can also use percentages

These are all summary statistics that are only substitutes for more detailed data. They enable the researchers to generalize about the sample of study objects. It should be noted these summary statistics are only helpful if they accurately represented the sample.

One can also use some useful techniques for displaying the data. These include frequency tables, bar charts, and pie chart etc.

11.1 **hypothesis testing**

There are two approaches to hypothesis testing. The more established is the classical or sampling theory approach; the second is known as the Bayesian approach. Classical statistics are found in all of the major statistics books and are widely used in research applications. This approach represents an objective view of probability in which the decision-making rests totally on an analysis of available sampling data. A hypothesis is established, it is rejected of fails to be rejected, based on the sample data collected.

In classical test of significance, two kinds of hypothesis are used:

(i) The null hypothesis debited $H_{O;}$ is a statement that no difference exists between the parameter and the statistic being compared.

(ii) The alternative hypothesis denoted H_1 is the logical opposite of the null hypothesis.

The alternative hypothesis – denoted (H_1) may take several forms, depending on the objective of the researchers. The H_1 may be of the 'not the same' form (nondirectional) A second variety may be of the 'greater than' or less the' form directional)

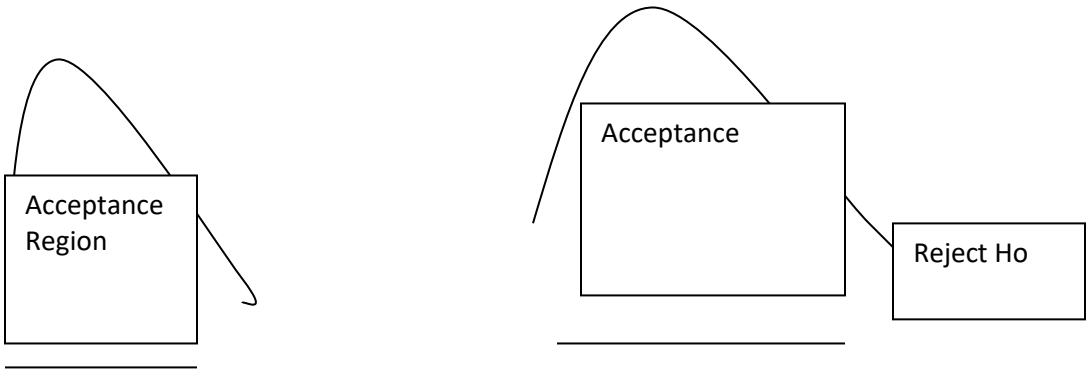
These types of alternative hypothesis correspond with two-tailed and one tailed tests.

- A two –tailed test, or non-directional test, considers two possibilities. The average could be 'more than' or it could be less than. To test hypothesis, the region of rejection is divided into two tails of the distribution.
- A one –tailed test, or directional test, places the entire probability of an unlikely outcome into the tail specified by the alternative hypothesis.

In figure 11.1 the first diagram represents a no directional hypothesis, and the second is a directional hypothesis of the 'greater than' variety. Hypothesis may be expressed in the following form;

Null	H_o u (mu)	= 50 days
Alternative	H_1: u	=days (not the same case)
Or H_1: u		>50 days (greater than case)
Or	H1: u	<0 days (less than the case

Figure 11.1 one and two tailed tests at the 5% level



Note

A type 1 error is committed when a true hypothesis is rejected
A type 11 error is committed when one fails to reject a false null hypothesis.

Statistical Testing Procedures:

(i) State the null hypothesis.
(ii) Choose the statistical test.
(iii) Select the desired level of significance. 0.05 0.01.
(iv) Compute the calculated difference value.
(v) Obtain the critical test value, i.e., t, z, or x2.
(vi) Make the decision. For most tests if the calculated value is larger than critical value, we reject the null hypothesis and conclude that the alternative hypothesis is supported (although it is by no means proved).

11.2 A Note on Test of Significance

There are two general classes of significance tests, parametric and nonparametric.

- Parametric tests are more powerful because their data are derived from interval and ratio measurements.
- Nonparametric tests are used to test hypothesis with nominal and ordinal data.

Parametric Tests

The Z or test is used to determine the statistical significance between a sample distribution mean and a parameter. The Z distribution and T distribution differ (the latter is used for large samples, i.e., greater than 30) But when sample sizes approach 120, the t and Z distributions are virtually identical.

One- Sample Tests:
These are used when we have a single sample and wish to test the hypothesis that it comes from a specified population. In this case, we encounter such as these:

a) Is there a difference between observed frequencies and the frequencies we would expect, based on some theory?
b) Is there a difference between observed and expected proportion?
c) Is it reasonable to conclude that a sample is drawn from a population with some specified distribution (normal, poison, and so forth)
d) Is there significant difference between some measure of central tendency (x) and its population parameter (u)?

For example
A machine fills packets with spice which are supposed to have a mean weight of 40 grams. A random sample of 36 packets is taken and the mean weight is found to be 42.2 grams with a standard deviation of 6 grams. It is required to conduct by following the six-step procedure recommended earlier.

1. Null Hypothesis H_0: u= 40 grams
H_1: u=40 grams (a two-tail test)

X= 42.4 grams (given)
s =6 grams | (given)
n=36 grams (given)

Where

x=sample mean
u=population mean
s=standard deviation
n=sample size (greater than 30)
2. Statistical test: Select the Z test because sample size is greater than 30.
3. Significance level: x 0.05 and n=36
4. Calculated value: first compute standard error= Sx=S/n=6/36=1

Z= x-u/sx= 42.4-40=2.24

5. Critical test. At the 5% significance level the Z value must be within ±1.96, i.e., (obtain From normal tables).