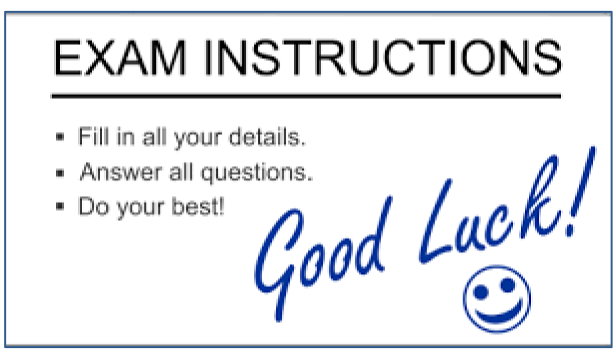
QUALIFYING EXAMINATION
TEST PREPARATION



Sponsored by the Department of Educational Leadership

Table of Contents

Quantitative Research Designs

Correlational

- Correlational research attempts to determine how related two or more variables are. This degree of relation is expressed as a correlation coefficient. This means to study whether an increase or decrease in one variable corresponds to an increase or decrease in the other variable. It is very important to note that correlation doesn't imply causation. Discuss in terms of direction (positive, negative) and strength (strong, weak).

True Experimental

- For an experiment to be classed as a true experimental design, it must fit all of the following criteria.
    - The sample groups must be assigned randomly.
    - There must be a viable control group.
    - Only one variable can be manipulated and tested. It is possible to test more than one, but such experiments and their statistical analysis tend to be cumbersome and difficult.
    - The tested subjects must be randomly assigned to either control or experimental groups.

Quasi Experimental

- Quasi experiments resemble quantitative and qualitative experiments, but lack random allocation of groups or proper controls, so firm statistical analysis can be very difficult. Quasi-experimental design involves selecting groups, upon which a variable is tested, without any random pre-selection processes. After this selection, the experiment proceeds in a very similar way to any other experiment, with a variable being compared between different groups, or over a period of time.

Causal-comparative/Ex post facto

- Causal-comparative research attempts to identify a cause-effect relationship between two or more groups. Causal-comparative studies involve comparison in contrast to correlation research which looks at relationship.

Descriptive/Survey Research Design

- Descriptive research often involves collecting information through data review, surveys, interviews, or observation. This type of research best describes the way things are. The survey research design is a very valuable tool for assessing opinions and trends.

Qualitative Research Designs

Qualitative research studies are more observational. The data gathered in these studies is verbal rather than numerical. Qualitative studies often lead to more questions than answers, and may even be used as a forerunner to quantitative research. Qualitative data can be used to assist researchers in formulating quantifiable questions and studies.

Phenomenological Studies

- These examine human experiences through the descriptions provided by the people involved. The goal is to describe the meaning that experiences hold for each subject. This type of research is used to study areas in which there is little knowledge. Respondents are asked to describe their experiences as they perceived them.

Ethnographic Studies

- Ethnography can be defined as the systematic process of observing, detailing, describing, documenting, and analyzing the lifeways or particular patterns of a culture in order to grasp the lifeways or patterns of the people in their familiar environment. These involve the collection and analysis of data about cultural groups. Ethnographers try to show how actions in one work make sense from the point of view of another world. They often interview key informants or observe subjects in order to learn about their culture, rituals, and customs.

Grounded Theory Studies

- In Grounded Theory Studies, data are collected and analyzed and then a theory is developed that is grounded in the data. Inductive and deductive reasoning are used. Purposeful sampling is used given that the researcher looks for certain subjects who will be able to shed new light on the phenomenon being studied. Diversity rather than similarity is sought in the people that are sampled. Data (interviews and observations) are gathered in a naturalistic/field setting. A process called constant comparison is used, in which data are constantly compared to data that have already been gathered. Coding is then used.

Historical Studies

- These studies relate to the identification, location, evaluation, and synthesis of data from the past. Historical research seeks not only to discover the events of the past but to relate these past happenings to the present and to the future. Data for historical research are usually found in documents or in relics and artifacts. Material may be found in libraries, archives, or in personal collections. Primary resources should be used when possible.

Case Studies

- Case studies are in-depth examinations of people, groups, or institutions. A case study may be considered quantitative or qualitative research depending on the purpose and design of the study. Hypotheses are generated from case studies but people do not use case studies to test hypotheses. Content analysis is used in evaluating data from case studies.

Action Research Studies

- Action research seeks action to improve practice and study the effects of the action that was taken. There is no goal of trying to generalize the findings. The implementation of solutions occurs as an actual part of the research process.

Mixed Method Research Design

The Triangulation Design

- This is the most common and well-known approach. The purpose of this design is to obtain different but complementary data on the same topic to best understand the research problem. The intent is to bring together the differing strengths and nonoverlapping weaknesses of quantitative methods with those of qualitative methods. This design is used when a researcher wants to directly compare and contrast quantitative statistical results with qualitative findings or to validate or expand quantitative results with qualitative data. Quantitative and qualitative methods are implemented during the same timeframe with equal weight. The researcher attempts to merge the two datasets.

The Embedded Design

- When using the Embedded Design Model, one data set provides a supportive, secondary role in a study based primarily on the other data type. The premises of this design are that a single data set is not sufficient, that different questions need to be answered, and that each type of question requires different types of data. Researchers use this design when they need to include qualitative or quantitative data to answer a research question within a largely quantitative or qualitative study. This model mixes the different data sets at the design level, with one type of data being embedded within a methodology framed by the other data type.

The Explanatory Design

- The Explanatory Design is a two-phase mixed methods design. Its purpose is that qualitative data helps explain or build upon initial quantitative results. It can be helpful when one needs qualitative data to explain significant (or nonsignificant) results, outlier results, or surprising results. This design can also be used when a researcher wants to form groups based on quantitative results and follow up with the groups throughout subsequent qualitative research or to use quantitative participant characteristics to guide purposeful sampling for a qualitative phase. Because this design begins quantitatively, investigators typically place greater emphasis on the quantitative methods than the qualitative methods.

The Exploratory Design

- The intent of this two-phase design is that the results of the first method (qualitative) can help develop or inform the second method (quantitative). This design is based on the premise that an exploration is needed for one of several reasons: measures or instruments are not available, the variables are unknown, or there is no guiding framework or theory. This design is particularly useful when one needs to develop and test an instrument because one is not available or identify important variables to study quantitatively when the variables are unknown. It is also appropriate when a researcher wants to generalize results to different groups, to test aspects of an emergent theory, or to explore a phenomenon in depth and then measure its prevalence.

Probability Sampling Methods

Probability sampling methods*:* In this sampling technique, the researcher must guarantee that every individual has an equal opportunity for selection and this can be achieved if the researcher utilizes randomization. The advantage of using a random sample is the absence of both systematic and sampling bias. If random selection was done properly, the sample is therefore representative of the entire population.

Cluster

- Cluster random sampling is done when simple random sampling is almost impossible because of the size of the population. Just imagine doing a simple random sampling when the population in question is the entire population of the United States. In cluster sampling, the research first identifies boundaries, in case of our example; it can be states within the United States. The researcher randomly selects a number of identified areas. It is important that all areas (states) within the population be given equal chances of being selected. The researcher can either include all the individuals within the selected areas or he can randomly select subjects from the identified areas.

Multi-Stage Random Sampling

- The combining of several sampling techniques to create a more efficient or effective sample than the use of any one sampling type can achieve on its own.

Simple random

- All the researcher needs to do is assure that all the members of the population are included in the list and then randomly select the desired number of subjects.

Stratified/Proportional Random Sampling

- This is a probability sampling technique wherein the subjects are initially grouped into different classifications such as age, socioeconomic status or gender. Then, the researcher randomly selects the final list of subjects from the different strata. It is important to note that all the strata must have no overlaps.

Systematic Random Sampling

- This can be likened to an arithmetic progression wherein the difference between any two consecutive numbers is the same. The first thing you do is pick an integer that is less than the total number of the population; this will be your first subject e.g. (3). Select another integer which will be the number of individuals between subjects e.g. (5). You subjects will be patients 3, 8, 13, 18, 23, and so on.

Non-Probability Sampling Methods

Non probability sampling methods: Non-probability sampling is a sampling technique where the samples are gathered in a process that does not give all the individuals in the population equal chances of being selected. Subjects in a non-probability sample are usually selected on the basis of their accessibility or by the purposive personal judgement of the researcher.

Convenience Sampling

- With convenience sampling, the samples are selected because they are accessible to the researcher. Subjects are chosen simply because they are easy to recruit. This technique is considered easiest, cheapest and least time consuming.

Consecutive

- This technique can be considered as the best of all non-probability samples because it includes all subjects that are available that makes the sample a better representation of the entire population.

Expert Sampling

- Assembling a sample of persons with known or demonstrable experience and expertise in some area (e.g. an expert panel).

Heterogeneity Purposive Non-probability Sampling

- A way to sample a wide array of ideas. One includes a broad and diverse range of participants.

Judgmental/Purposive

- In this type of sampling, subjects are chosen to be part of the sample with a specific purpose in mind. With judgmental sampling, the researcher believes that some subjects are more fit for the research compared to other individuals.

Modal Instance Purposive Nonprobability Sampling

- Sampling from the most frequent or typical cases.

Quota

- Quota sampling is a non-probability sampling technique wherein the researcher ensures equal or proportionate representation of subjects depending on which trait is considered as basis of the quota.

Snowballing

- Snowball sampling is usually done when there is a very small population size. In this type of sampling, the researcher asks the initial subject to identify another potential subject who also meets the criteria of the research. The downside of using a snowball sample is that it is hardly representative of the population.

Research Instruments

Instruments fall into two broad categories, researcher-completed and subject-completed, distinguished by those instruments that researchers administer versus those that are completed by participants.

Types of Instruments

Researcher-completed Instruments	**Subject-completed Instruments**
Rating scales	Questionnaires
Interview schedules/guides	Self-checklists
Tally sheets	Attitude scales
Flowcharts	Personality inventories
Performance checklists	Achievement/aptitude tests
Time-and-motion logs	Projective devices
Observation forms	Sociometric devices

Validity

Validity is the extent to which an instrument measures what it is supposed to measure and performs as it is designed to perform. It is rare, if nearly impossible, that an instrument be 100% valid, so validity is generally measured in degrees. As a process, validation involves collecting and analyzing data to assess the accuracy of an instrument. There are numerous statistical tests and measures to assess the validity of quantitative instruments, which generally involves pilot testing. The remainder of this discussion focuses on external validity and content validity.

External validity is the extent to which the results of a study can be *generalized* from a sample to a population. Establishing external validity for an instrument, then, follows directly from sampling. Recall that a sample should be an accurate representation of a population, because the total population may not be available. An instrument that is externally valid helps obtain population generalizability, or the degree to which a sample represents the population.

Content validity refers to the appropriateness of the content of an instrument. In other words, do the measures (questions, observation logs, etc.) accurately assess what you want to know?

Reliability

Reliability can be thought of as consistency. Does the instrument consistently measure what it is intended to measure? It is not possible to calculate reliability; however, there are four general estimators that you may encounter in reading research:

1. *Inter-Rater/Observer Reliability*: The degree to which different raters/observers give consistent answers or estimates.
2. *Test-Retest Reliability*: The consistency of a measure evaluated over time.
3. *Parallel-Forms Reliability:* The reliability of two tests constructed the same way, from the same content.
4. *Internal Consistency Reliability:* The consistency of results across items, often measured with Cronbach’s Alpha.

Selecting your Statistical Procedures: Parametric and Nonparametric Tests

There are several main factors to consider when deciding whether you should use parametric or non-parametric statistical approaches:

1. The type of data that you are working with

You will typically use a non-parametric test when you are working with data that is nominal or ordinal.
For interval or ratio data, you may use a parametric test depending on the shape of the distribution (see below).

2. The shape of the distribution

IF the measurement scale of your data is interval or ratio, the distribution should be approximately normal to use parametric approaches.
IF you have interval or ratio data with a large sample size (> 30) you can usually use parametric approaches due to the Central Limit Theorem.

3. Variances (homoscedasticity)

Parametric tests often assume homogeneity of variances (use F test in R to test)
Non-parametric tests do NOT make assumptions about sample variances

4. Independent and random samples

Parametric tests assume that samples are independent. Non-parametric tests can be used for any type of relationships between samples.
The table below summarizes considerations for choosing a parametric versus non-parametric test:

Type of Analysis	*Parametric*	*Non-Parametric*
Assumed Distribution	Normal	No Assumed Shape
Type of Data	Interval or Ratio	Ordinal, Nominal (or Interval/ Ratio not satisfying other parametric assumptions)
Assumed Homogeneity of Variances?	YES	NO
Samples are Independent?	YES	Not necessarily.

Steps to choose the appropriate statistical method for the data and situation.

1. Identify whether a given research problem raises the question of describe, relate, or compare.

2. Identify whether the numbers in a given research problem
   - name (nominal or class A data),
   - rank (ordinal or class B data), or
   - equally space (interval/ratio or class C data).

The best procedure would seem to be to treat ordinal measurements as though they were intervals measurements, but to be constantly alert to the possibility of gross inequality of intervals.

3. Identify the number of groups, variables, or samples being described, related, or compared in a given research problem.

4. Identify whether comparison samples are related (matched on some relevant characteristic or multiple measures on the same individuals) or independent (randomly assigned).

5. Choose the appropriate statistical tool for the data and situation.

Levels of Measurement

Nominal/Categorical

In nominal measurement the numerical values just "name" the attribute uniquely. No ordering of the cases is implied. For example, jersey numbers in basketball are measures at the nominal level. A player with number 30 is not more of anything than a player with number 15, and is certainly not twice whatever number 15 is.

Ordinal

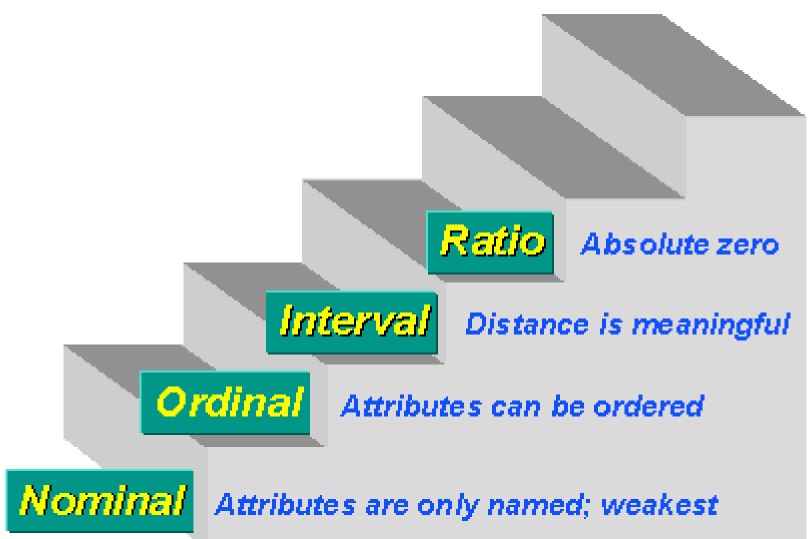
In ordinal measurement the attributes can be rank-ordered. Here, distances between attributes do not have any meaning. For example, on a survey you might code Educational Attainment as 0=less than H.S.; 1=some H.S.; 2=H.S. degree; 3=some college; 4=college degree; 5=post college. In this measure, higher numbers mean more education. But is distance from 0 to 1 same as 3 to 4? Of course not. The interval between values is not interpretable in an ordinal measure.

Interval

In interval measurement the distance between attributes does have meaning. For example, when we measure temperature (in Fahrenheit), the distance from 30-40 is same as distance from 70-80. The interval between values is interpretable. Because of this, it makes sense to compute an average of an interval variable, where it doesn't make sense to do so for ordinal scales. But note that in interval measurement ratios don't make any sense - 80 degrees is not twice as hot as 40 degrees

Ratio Level Data

In ratio measurement there is always an absolute zero that is meaningful. This means that you can construct a meaningful fraction (or ratio) with a ratio variable. Weight is a ratio variable. In applied social research most "count" variables are ratio, for example, the number of clients in past six months. Why? Because you can have zero clients and because it is meaningful to say that "...we had twice as many clients in the past six months as we did in the previous six months."



Descriptive Statistics

All educators are involved in research and statistics to a degree. For this reason all educators should have a practical understanding of research design. Even if an educator is not involved in conducting research, he or she must be involved in the interpretation of the findings of others to remain dynamic in their teaching.

The primary use of descriptive statistics is to describe information or data through the use of numbers (create number pictures of the information). The characteristics of groups of numbers representing information or data are called descriptive statistics. Descriptive statistics are used to describe groups of numerical data such as test scores, number or hours of instruction, or the number of students enrolled in a particular course.

Descriptive statistics - Numbers which are used to describe information or data or those techniques used to calculate those numbers.

Variable (x) - A measurable characteristic. Individual measurements of a variable are called varieties, observations, or cases.

Population (X) - All subjects or objects possessing some common specified characteristic. The population in a statistical investigation is arbitrarily defined by naming its unique properties.

Parameter - A measurable characteristic of a population. A measurable quantity derived from a population, such as population mean or standard deviation.

Sample - A smaller group of subjects or objects selected from a large group (population).

Statistic - A measure obtained from a sample. It is a measurable quantity derived from a sample, such as the sample mean or standard deviation.

Frequency graph - A picture depicting the number of times an event occurred.

Bar graph or histogram - A frequency graph with number of blocks or length of bar representing the frequency of occurrence.

Frequency polygon - A modification of the bar graph with lines connecting the midpoints of the highest point on each bar.

Frequency curve - A modification of a frequency polygon with the sharp corners rounded. The area under the connecting line of the bar graph, frequency polygon, and frequency curve are equivalent and represent frequency of occurrence.

Mean (μ) or Arithmetical mean - A number having an intermediate value between several other numbers in a group from which it was derived and of which it expressed the average value. It is the simple average formed by adding the numbers together and dividing by the number of numbers in the group.

Median - The midpoint in a set of ranked numbers.

Mode - The number which occurs most often in a group of numbers.

Range - The difference in the highest score and the lowest score in a set of scores. The range is obtained by subtracting the low score from the high score R = xh - xl.

Variance - The mean of the squared deviations of individual numbers from the mean of the group of numbers; the square of the standard deviation.

Standard deviation - A measure of the deviation of individual numbers from the mean of the group of numbers. It is the mean or average deviation of those numbers from the mean of the set of numbers .

Measures of Central Tendency

Mode = Number which occurs most often

Median = Middle number - 50% above - 50% below

Mean = Average = M = $\bar{x} = \frac{\sum x}{n}$

Measures of Variation

Range = Distance from highest to lowest score

Variance = Squared standard deviation = $s^2 = \frac{\Sigma(x - \bar{x})^2}{n}$

Standard deviation = average distance of individual numbers from the mean -

$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}}$

Inferential Statistics

Descriptive Statistics - Numbers used to describe information or data or those techniques used to calculate those numbers.

Inferential Statistics - A procedure used to estimate parameters (characteristics of populations) from statistics (characteristics of samples).

Population - All subjects or objects possessing some common specified characteristic. The population in a statistical investigation is arbitrarily defined by naming its unique properties.

Parameter - A measurable characteristic of a population (μ).

Sample - A smaller group of subjects or objects selected from a large group (population).

Statistic - A measurable characteristic of a sample (x, s).

Variable - A characteristic of objects or subjects that can take on different values.

Qualitative Variables - Characteristics which vary in quality or value.

Quantitative Variables - Characteristics which vary in quantity, amount, or size.

Independent Variables - Characteristics which affect or cause the outcome of the experiment but do not measure the results.

Dependent Variables - Characteristics which measure the effects or results of the experimental treatment or independent variable.

Predictor Variables (x) - Measurable characteristics from which the criterion variable (y) can be estimated.

Hypothesis - A supposition (an educated guess) presumed to be true for the sake of subsequent testing. In educational research, hypotheses concern the existence of relationships between variables.

Statistical Hypothesis (Ho: Null Hypothesis) - States that there is no (null) relationship between the variables under analysis.

Research Hypothesis (Ha: Alternative Hypothesis) - A positive statement of the null hypothesis. It states that there is a relationship between the variables under analysis.

Probability (p) - The chance of something happening under certain conditions. In other words, it is the likelihood of the occurrence of any particular form of an event, estimated as the ratio of the number of ways in which that form might occur to the whole number of ways in which the event might occur in any other form.

Example: If an event can happen in "s" ways and fail to happen in "f" ways, and if each of these s + f ways is equally likely to occur, the probability of success in a single trial is $p = s / (s+f)$

Statistical Significance

When a statistical test reveals that the probability is rare that a set of observed sample data is attributable to chance alone, this result is labeled as statistically significant. If two groups are so different that only one time in 1000 would we find such a difference by chance alone, the difference would be statistically significant. By statistically significant, it is meant that the observed phenomenon represents a significant departure from what might be expected by chance alone.

The level of significance (alpha) is the probability of a Type I error that an investigator is willing to risk in rejecting a null hypothesis. Generally, it refers to the probability of the event occurring due to chance. If alpha = .01, it is likely that one time out of a hundred the event could occur due to chance. If you lower the significance level from .05 to .01, you decrease the probability of rejecting a true hypothesis but increase the probability of accepting a false hypothesis. A Type II error (beta) occurs when an investigator fails to accept the alternative hypothesis when in fact the alternative hypothesis was true. In other words, the null hypothesis was accepted when it was not true.

Correlational Design

Steps for Hypothesis Testing

1. <u>State the null hypothesis</u>.
Ho = There is a lack of significant relationship between the pounds of nitrogen (x) applied to a crop of corn and the yield of bushels of corn (y).

2. <u>Choose a significance level</u> based on confidence sought, typically .05 or .01.

3. <u>Calculate the degrees of freedom</u>. They are calculated as the number of pairs of measures minus two. For our example, our df = 3. The correlation table of values is entered through the df on the left and the significance level (.05) on top to give a value of .878.

4. <u>Compare calculated value to table value</u>.
If the calculated value is equal to or greater than the table value, we reject the null hypothesis.
If the calculated value is less than the table value, we accept the null hypothesis.

calculated value => table value = reject null
calculated value < table value = accept null

5. <u>Accept or reject null hypothesis</u>.
Therefore, in our example we compare .99 to .878 and find our calculated value (.99) to be greater than our table value of .878. We reject the null hypothesis and conclude that in 95% of the cases this relationship would be the result of the experimental conditions rather than chance factors. Approximately five percent of the time a relationship of this magnitude could result from chance factors.

Assumptions: (Same for Correlation and Regression)

1. Representative Sample (Random)
2. Normal Population
3. Interval Measures
4. Linearity (Measures approximate a straight line)
5. Homoscedasticity (Equal variances)

According to Popham (1973, p. 80), "multiple correlation describes the degree of relationship between a variable and two or more variables considered simultaneously. . . . partial correlation allows the statistician to describe the relationship between two variables after controlling or partialing out the confounding relationship of another variable(s)."

Four Special Correlation Methods and Variable Relationships Assessed

Method	Variable Relationship
Point biserial coefficient (rpb)	Continuous versus dichotomous
Biserial coefficient (rb)	Continuous versus dichotomized
Phi coefficient (θ)	Dichotomous versus dichotomous
Tetrachoric coefficient (rt)	Dichotomized versus dichotomized

Regression Analysis

A prediction of the levels of one variable when another is held constant at several levels.

- Based on average variation remaining constant over time due to the tendency in nature for extreme scores to move toward the mean.
- Used to predict for individuals on the basis of information gained from a previous sample of similar individuals.

Regression Equation

= estimated y and is the value on the y axis across from the point on the regression line for the predictor x value. (Sometimes represented by y´.) This is the estimated value of the criterion variable given a value of the predictor variable.

a = the intercept point of the regression line and the y axis. It is calculated through the equation ; therefore, the means of both variables in the sample and the value of b must be known before a can be calculated.

b = the slope of the regression line and is calculated by this formula: If the Pearson Product Moment Correlation has been calculated, all the components of this equation are already known.

x = an arbitrarily chosen value of the predictor variable for which the corresponding value of the criterion variable is desired.

Assumptions: (Same for correlation and regression)

1. Representative sample (Random)
2. Normal distribution for population
3. Interval measures
4. Linearity (Measures approximately a straight line)
5. Homoscedasticity (Equal variances)

Simple linear regression predicts the value of one variable from the value of one other variable. Multiple regression predicts the value of one variable from the values of two or more variables. Using two or more predictor variables usually lowers the standard error of the estimate and makes more accurate prediction possible. In our example if we could add soil type or fertility, rainfall, temperature, and other variables known to affect corn yield, we could greatly increase the accuracy of our prediction.

One caution: Due to the assumption of linearity, we must be careful about predicting beyond our data. If we predict beyond the information that we have known, we have no assurance that it remains linear or in a straight line. It might begin to curve and thus negate all our predictions in this region. Also, we must remember that the variables we are predicting must be like those upon which the regression equation was built or our prediction has no basis.

T-Tests

We are called on many times to determine if the mean performance of two groups are significantly different. Those two groups might be students, cattle, plants, or other objects. When attempting to determine if the difference between two means is greater than that expected from chance, the "t" test may be the needed statistical technique. If the data is from a normal population and at least ordinal in nature, then we are surer that this is the technique to use. If you wish to generalize to a population, then the samples must be representative.

"t" is the difference between two sample means measured in terms of the standard error of those means, or "t" is a comparison between two groups means which takes into account the differences in group variation and group size of the two groups. The statistical hypothesis for the "t" test is stated as the null hypothesis concerning differences. There is no significant difference in achievement between group 1 and group 2 on the welding test.

Separate variance formula

Use the separate variance formula if:

If $n_1 = n_2$ and $s_1^2 = s_2^2$	$df = n_1 + n_2 - 2$
If $n_1 = n_2$ and $s_1^2 \neq s_2^2$	$df = n_1 - 1$ or $n_2 - 1$
If $n_1 \neq n_2$ and $s_1^2 \neq s_2^2$	$df =$ average of $n_1 - 1$ and $n_2 - 1$

where $\bar{x}_1 =$ mean of sample 1

$\bar{x}_2 =$ mean of sample 2

$n_1 =$ number of subjects in sample 1

$n_2 =$ number of subjects in sample 2

$$s_1^2 = \text{variance of sample 1} = \frac{\Sigma(x_1 - \bar{x}_1)^2}{n_1}$$

$$s_2^2 = \text{variance of sample 2} = \frac{\Sigma(x_2 - \bar{x}_2)^2}{n_2}$$

Pooled Variance Formula

Use the pooled variance formula if:

If $n_1 = n_2$ and $s_1^2 = s_2^2$	$df = n_1 + n_2 - 2$
If $n_1 \neq n_2$ and $s_1^2 = s_2^2$	$df = n_1 + n_2 - 2$
If $n_1 = n_2$ and $s_1^2 \neq s_2^2$	$df = n_1 - 1$ or $n_2 - 1$

Correlated Data Formula

If the samples are related (two measures from the same subject or matched pairs), the correlated data formula is used.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} - 2r\left[\frac{s_1}{\sqrt{n_1}}\right]\left[\frac{s_2}{\sqrt{n_2}}\right]}}$$

In choosing the correct formula, it is fairly easy to determine if the sample sizes are equal. The number of subjects are either the same or they are not.

However, to determine if the variances are homogeneous, use the formula F = s^2 (largest) / s^2 (smallest). We compare the calculated F value to the F table value at the .05 or .01 level of significance with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

If the calculated values >= table value, then the variances are not equal; if the calculated value < table value, then the variances are equal.

Assumptions

1. Representative sample (Random)

2. Normal distribution for population

3. At least ordinal measures

Analysis of Variance (ANOVA)

The purpose of the Analysis of Variance (ANOVA) technique is to test for significant differences among two or more groups.

In single classification ANOVA, you are trying to find out if there is any relationship between a dependent variable (such as student achievement) and several classifications of one independent variable (such as different instructional materials).

In multiple classification ANOVA, you are trying to find out the relationship between one dependent variable (such as student achievement) and classifications of two or more independent variables (such as several methods of instruction and different instructional materials).

Therefore, the factor determining whether to use single or multiple classification ANOVA is the number of independent variables.

Since the variance (or its square root, the standard deviation) is really an average distance of the raw scores in a distribution of numbers from the mean of that distribution, this functional relationship between the variance and the mean can be used to determine mean differences by analyzing variances.

In essence, the ANOVA method is to calculate the variances of each subgroup being compared. The average variance of these subgroups is then compared to the variance of the total group (created by artificially combining the subgroups). If the average variance of the subgroups is about the same as the variance of the total group, then no significant difference exists among the means of the subgroups. However, if the average variance of the subgroups is smaller than the variance of the total group, then the means of the subgroups are significantly different.

The first step in computing ANOVA is to calculate the sums of squares of the deviations of the observations from their mean (hereafter referred to as sums of squares or SS) for each of the separate groups being compared and add them together to form the within groups SS. Next, compute the SS for the total group made by combining the subgroups. Subtract the within group sum of squares from the total group sum of squares to derive the among group sum of squares. Divide the among and within sums of squares by their degrees of freedom to obtain their mean squares (variances), then divide the among mean square by the within mean square to obtain the calculated F value. Finally, determine if the calculated F value is sufficiently large to reject the null hypothesis. If the calculated F is >= table value at the chosen level of significance, the null hypothesis is rejected; if the calculated value is < table value, then null hypothesis is accepted.

The among group mean square or variance and the within group mean square or variance determine the size of F.

The hypothesis being tested is: There are no significant differences among the means of achievement of the groups being taught by the three different methods.
The degrees of freedom for the different sums of squares

Among group df equals the number of groups minus one (k - l)

Within groups df equals the number of groups times the number within each group minus one k(n - l)

Total group df equals the total number of subjects minus 1 (kn - l) and can be used as a cross check since among df plus within df must equal total df.

Assumptions

1. Representative Sample (Random)
2. Normal Distribution for the Populations
3. Interval Measures
4. Homoscedasticity
5. Independent Observations

Chi-Square

Purpose of Chi Square

The Chi Square (X2) test is undoubtedly the most important and most used member of the nonparametric family of statistical tests. Chi Square is employed to test the difference between an actual sample and another hypothetical or previously established distribution such as that which may be expected due to chance or probability. Chi Square can also be used to test differences between two or more actual samples.

Basic Computational Equation

$$X^2 = \Sigma \frac{(\text{Observed frequency} - \text{Expected frequency})^2}{\text{Expected frequency}}$$

One-Way Classification

The One-Way Classification (or sometimes referred to as the Single Sample Chi Square Test) is one of the most frequently reported nonparametric tests in journal articles. The test is used when a researcher is interested in the number of responses, objects, or people that fall in two or more categories. This procedure is sometimes called a goodness-of-fit statistic. Goodness-of-fit refers to whether a significant difference exists between an observed number and an expected number of responses, people or objects falling in each category designated by the researcher. The expected number is what the researcher expects by chance or according to some null hypothesis.

Two-Way Classification

The two-way Chi Square is a convenient technique for determining the significance of the difference between the frequencies of occurrence in two or more categories with two or more groups. For example, we may see if there is any difference in the number of freshmen, sophomores, juniors, or seniors in regards to their preference for spectator sports (football, basketball, or baseball). This is called a two-way classification since we would need two bits of information from the students in the sample, their class and their sports preference.

Degrees of Freedom

A value of X2 cannot be evaluated unless the number of degrees of freedom associated with it is known. The number of degrees of freedom associated with any X2 may be easily computed.

If there is one independent variable, df = r - 1 where r is the number of levels of the independent variable.

If there are two independent variables, df = (r - l) (s - l) where r and s are the number of levels of the first and second independent variables, respectively.

If there are three independent variables, df = (r - l) (s - 1) (t - 1) where r, s, and t are the number of levels of the first, second, and third independent variables, respectively.

Assumptions

Even though a nonparametric statistic does not require a normally distributed population, there still are some restrictions regarding its use.

1. Representative sample (Random)

2. The data must be in frequency form (nominal data) or greater.

3. The individual observations must be independent of each other.

4. Sample size must be adequate. In a 2 x 2 table, Chi Square should not be used if n is less than 20. In a larger table, no expected value should be less than 1, and not more than 20% of the variables can have expected values of less than 5.

5. Distribution basis must be decided on before the data is collected.

6. The sum of the observed frequencies must equal the sum of the expected frequencies.

Practice Questions

1. A pilot experiment, designed to test the effectiveness of a new approach to spelling proficiency, has been conducted over a semester-long period in a junior high school. Two different classes of 29 seventh grade pupils participated in the study. Students in Group X received their formal spelling instruction from Mr. Jordan, an experienced teacher. Mr. Jordan employed conventional methods of spelling instruction such as are commonly used in upper elementary grades. While he met with Group X five times each week for a period of one hour, only one hour per week was devoted to spelling instruction. Group Y also met with Mr. Jordan daily, but while they devoted an hour per week to spelling instruction, all formal spelling teaching was carried out through the use of newly developed programmed spelling lessons that were presented to the student through the use of self-instruction "teaching machines." Since all of the students involved in the pilot study had been pre-selected for purposes of the research, subjects in the two groups had been matched in pairs with respect to intellectual ability and pre-experimental performance on an orally administered spelling quiz of 150 words. At the conclusion of the semester's experiment, the spelling quiz was readministered to both groups. Which statistical procedures should be selected to test whether there is a significant difference between Mr. Jordan's conventional instruction techniques and the programmed method of teaching spelling as determined by post-experiment performance on the spelling quiz?

Question ______________________ Measurement Level ______________________

Number of Samples ______________________ Related or Independent ______________________

Statistic ______________________

2. A vocational instructor wishes to arrive at a representative performance for his class on an assigned task. In addition to the representative performance he also wishes to know the variability of the performance of the individuals in the class. The performance was assessed by a 100 point objective test and the scores were spread fairly evenly. What statistics should be used to analyze this data?

Question ______________________ Measurement Level ______________________

Number of Samples ______________________ Related or Independent ______________________

Statistic ______________________

3. Three new methods of teaching spelling are being contrasted with a conventional method by only one instructor. Four groups of fifteen pupils, randomly drawn from a large number of potential subjects, have been equated with respect to pre-experiment spelling scores, intelligence test scores, and previous scholastic performance in school. If one wished to determine significant differences in the scores of the four groups (three experimental, one control) on a post-experiment spelling test, which of the statistical procedures would be most appropriate?

Question ______________________ Measurement Level ______________________

Number of Samples ______________________ Related or Independent ______________________

Statistic ______________________

4. A high school guidance counselor wants to determine the association between number of extracurricular activities participated in by students and their scholastic achievement. This counselor tallied the number of activities each student participated in and then calculated their grade point averages. What statistics should the counselor use to analyze this data?

Question ______________________ Measurement Level ______________________

Number of Samples ______________________ Related or Independent ______________________

Statistic ______________________

5. An educational experimenter has studied the influence of a number of instructional variables on certain aspects of students' attitudes toward school and study activities. Using a newly developed attitude inventory, the researcher has sampled a large number of pupils, attempting to establish norms for the inventory. From these early studies, he is convinced that the attitude scores yielded by the inventory are not normally distributed in the student population. Rather, samples of the numerical scores acquired through use of the inventory are always negatively skewed. The experimenter thus reasons that analyses of such data should not be conducted with those parametric techniques which require relatively normal population distributions.

He is currently conducting an experiment in which two groups of six pupils have been exposed to different instructional procedures. All twelve pupils were randomly selected an randomly assigned to one of the two groups. At the conclusion of the instructional period, both groups were given the attitude inventory. Which of the statistical techniques should be used to test whether the attitude scores of the two groups are significantly different?

Question ______________________ Measurement Level ______________________

Number of Samples ______________________ Related or Independent ______________________

Statistic ______________________

6. A school administrator studies the results of a year's tryout of team teaching. Though achievement standards have been as high as in previous years, he is puzzled by results from a post-course questionnaire which was administered to the 95 pupils who took part in the team teaching project. In particular, he is perplexed by responses to question seven:

(7) Which type of teaching situation do you prefer? Number Checking

Team teaching with more students per group 12

Team teaching with fewer students per group 57

Conventional teaching with one teacher per 25-35 pupils 26

Of the 26 students who chose conventional teaching, all but 5 were boys. The administrator wonders whether the fact that all 3 teachers constituting the teaching team were women might be important. Specifically, he speculates that there may be a significant difference between the responses to question seven of the 55 boys and 40 girls in the group. Which of the statistics should be used?

Question ______________________ Measurement Level ______________________

Number of Samples ______________________ Related or Independent ______________________

Statistic ______________________

7. A welding instructor rated his students' practice pads on a scale of one to ten with one representing a very poor set of beads and ten representing an excellent set. He rated one hundred of these pads and would now like to know what a representative rating would be. What statistic should he use?

Question ______________________ Measurement Level ______________________

Number of Samples ______________________ Related or Independent ______________________

Statistic ______________________

8. A teacher wished to demonstrate that two forms of an achievement test she has developed are "parallel," or relatively equivalent. If so, she can administer one form as a pre-test and the second as a post-test in a study of certain new instructional procedures she wishes to introduce to her class. The actual experiment she wishes to conduct involves a comparison of the achievement of three matched groups of pupils who have been taught a series of concepts by three relatively distinct instructional procedures. The results of the experiment will assist the teacher in deciding which of the instructional procedures to use with future classes.

Anticipating this project, the teacher prepared a large number of test items during the preceding semester and randomly divided them into two sets of items which represented two forms of the test. Having instructed her class by using combination of the three new instructional procedures, she administered the two forms to 110 pupils in four sections (classes) of the same course at the close of the instructional period. She notes with pleasure that the means for the two test forms are practically identical. Which techniques can the teacher use to test the statistical significance?

Question ______________________ Measurement Level ______________________

Number of Samples ______________________ Related or Independent ______________________

Statistic ______________________

9. Consider the case of an educational investigator who wished to test whether there are any significant differences between the achievement scores of three samples of students taught by three rather distinctive varieties of lecture instruction. The first lecture approach features a multitude of rhetorical but unanswered questions designed to stimulate the students to think about the content of the lecture. The second lecture approach embodies many rhetorical questions, which the lecturer himself quickly answers. The final lecture technique has few, if any, rhetorical questions.

The investigator also wishes to see if the student's level of intelligence interacts in any significant fashion with the performance under the three varieties of lecture techniques. He gathers post-test achievement scores for three sets of subjects who have been exposed to the different lecture methods. Intelligence test data are also secured for each subject so that the total sample of 300 students (100 taught by each method) can be classified as "above average intelligence," "average intelligence," or "below average intelligence."

In order to test the null hypotheses (1) that there are no significant achievement differences between students taught by the three lecture methods and (2) that there is no significant interaction of intelligence level and lecture method on achievement, which of the statistical techniques should the researcher select?

Question ______________________ Measurement Level ______________________

Number of Samples ______________________ Related or Independent ______________________

Statistic ______________________

10. A high school counselor is faced with the task of setting up a systematic procedure for predicting the likely success of his school's seniors at a nearby private university. In previous years, students with a grade point average lower than B+, regardless of intelligence or any other factors, have been advised not to enroll in the University on the grounds that they would probably fail to maintain university grade standards. Both the admission officials at the University and the administrators at the school have requested that the counseling office develop a more precise scheme for predicting potential University success. Certain of the high school's graduates, each with C averages throughout high school, have entered the University (against advisement) and maintained excellent scholastic records.

The counselor is given complete access to records at the University, as well as the high school, for all the students who were graduated from the high school and entered the University. Records were quite complete for the past twelve years. Included in the University records were college grades and entrance examination scores. The high-school records have high-school grades, scores on a standardized intelligence test administered during the ninth grade, and certain personal data such as sex, church affiliation, etc. Which of the statistical devices should the counselor use to develop the prediction scheme which will permit more precise advisement of his high school's university aspirants?

Question _____________________ Measurement Level _____________________

Number of Samples _____________________ Related or Independent _____________________

Statistic _____________________

Research Designs Flowchart