

7. Research and Program Evaluation

1. Research and evidence-based inquiry

Research is the systematic process of collecting and analyzing data for some purpose such as investigating a problem or answering a question. Evidence-based inquiry is the search for knowledge using empirical data which has been gathered systematically.

2. Research may be quantitative or qualitative in nature

Quantitative	Qualitative
1. Assumes social facts have a single objective reality 2. Tends to study samples or populations 3. Researchers try not to influence collection of data (instruments) 4. Statistical methods comparing and contrasting groups occurs 5. Researchers examine for causes and relationships	1. Assumes multiple realities socially constructed by individuals and group 2. Tends to study individual units-person, family, community-in naturalistic setting 3. Researchers may be primary instrument for collecting data (through observation) 4. Researchers' impressions, judgments and feelings may be used 5. Goal is to describe the nature of things

Both kinds of research are valued. One is chosen over the other because it better fits the assumptions of the researcher and the nature of the problem under investigation. Some professional journals prefer to publish one kind of research over the other.

3. Research may be:

- **Inductive:** this research begins at the real world, practical level. It tends to be descriptive, correlational, or historical and leads to the building of theory.
- **Deductive:** this research springs from theory which is already established. This research tries to determine what the relationships are between elements of the theory and may be experimental in nature.

4. Types of research -- Quantitative

A. Non-experimental designs:

- **Survey:** this may occur through questionnaires, interviews, etc. and is used to measure attitudes, perceptions, etc. (For example: Public Opinion Poll). Often the response rate of survey research is low, below 50 percent. Unless you know that the characteristics of the non-respondents are similar to the characteristics of the respondents, you must be cautious in generalizing.
- **Descriptive:** this research simply describes an existing state of events. Numbers may be used to characterize groups or individuals.

- **Comparative:** this research method investigates whether there are differences between two or more groups. There is no manipulation of conditions experienced by each group.
- **Correlational:** this research method uses the correlation coefficient to determine the degree of relationship between two or more variables or phenomena.

For example: Income level and attitude towards counseling

- **Ex-Post-Facto:** Also called causal-comparative, this research design studies possible causal relationships among variables ex post facto (after the fact). You do not manipulate any variables; the focus is on what has already happened (after the fact). You may generate 'several' reasons (causes) for the relationships you discover.

For example: Two employment agencies conduct job clubs. In one agency, the job clubs are member-initiated and run whereas in the other, employment counselors provide instruction and guidance. In examining the job placement rates of members for the past year, you find that the members of the professionally led groups had a higher placement rate.

Typical statistics used in ex-post-facto research are the t -test and analysis of variance.

B. Experimental designs:

- **True experiment:** This research is characterized by the use of experimental and control groups with random assignment to each. Experimental designs are used to determine cause-and-effect relationships.

For example: Sixty college freshmen are enrolled in an English class. Thirty are randomly assigned to a one-hour per week writing lab, the others comprise a control group. End-of-semester essay examination results are analyzed to see if the lab was associated with better writing

For experiments, there are design variations such as:

- Treatment and control group with posttest only,
- Treatment and control group with pretest and posttest,
- Two different treatment groups with control group and posttest, and \ many other combinations.
- **Quasi-experiment:** This research is similar to experimental research except that randomization of subjects to treatment and control groups is not possible. It may be that no control or comparison group is available. Results from such research will not be as unequivocal as results from a true experimental study.

For example: An elementary school has two classrooms of fourth graders. Each classroom of fourth graders is taught arithmetic by

a different method for the school year. In May, arithmetic achievement is compared for the two classrooms using scores on a national exam.

There was no random assignment of students to the two classrooms and the 'arithmetic teaching methods' were not randomly assigned to the two classes.

5. Types of research-Qualitative

Qualitative research emphasizes gathering data about naturally occurring phenomena (individual's and groups' living experiences) and events. Data collection may be in terms of words rather than numbers.

There are two principal qualitative research designs:

A. Interactive:

- **Case study** in which the case may be a program, activity, or a set of individuals who are bounded in time and place.
- **Ethnography** which is a description and interpretation of a cultural or social group or system. Data is typically collected through observation and interviewing. The issue of observer bias is important.

B. Noninteractive:

- This is **analytical research** conducted primarily through document analysis. Examples might be historical analysis (collecting and analyzing documents describing former events), biographical analysis (written or oral) and legal analysis which focuses on law and court decisions.

6. Mixed-method research designs

These designs combine quantitative and qualitative methods in the same research effort. The researcher retains the flexibility to use both types of designs. Typically, the designs are used sequentially. For example, quantitative may be gathered first and then qualitative methods are used to further explain or elaborate on the findings, using perhaps, surveys, interviews, or focus groups.

7. Other specialized research designs and types

- **Single-subject design:** studies the effects of a program or treatment on an individual or group treated as an individual, usually after a baseline has been established.
- **Action research:** conducted in an attempt to improve services or a program. This research may often be viewed as having an evaluative function.
- **Pilot study:** a small-scale research effort often used to determine the feasibility of a large scale effort with emphasis on refining procedures and instrumentation.
- **Longitudinal research:** consists of collecting data from the same group of individuals over a period of time. This is also called a panel study. For example, this author studied the career development of school children beginning when

they were second graders and re-interviewing them every two years until they were high school seniors.

- **Cross-sectional research:** consists of collecting data from different groups at the same time and examining these differences. For example, one approach to studying career development is collecting data from each grade of students, for example from second to twelfth grade, at the same time.

Research outcomes may be measured two ways:

- **Within-subjects** -examining what changes occur within the members of a group
- **Between subjects** -examining what changes occur between two or more groups

8. Meta-analysis

This is research comparing findings across studies, i.e., the results of many studies are examined simultaneously and one or more research questions answered.

For example, a number of studies have examined what works in various counseling situations. Examining a number of these studies simultaneously can help determine what counseling techniques work for what kind of clients, with what kinds of problems, under what conditions.

9. Internal validity

Experiments are internally valid to the extent that extraneous variables have been controlled. In other words, to the extent that the treatment variable is the only one producing the observed changes, the experiment is internally valid. What are the threats (confounding variables) to internal validity? The most likely ones are:

- **Selection of subjects**-differences in the results between two groups may not be due to the treatment variable experienced by one group because the composition of the two groups are different to begin with (probably not randomly selected).
- **Instrumentation**-differences in results between two or more groups may be due to instruments which are unreliable or because the instruments are changed during the study. Or, perhaps the observers recording data become fatigued or bored and they record behaviors differentially over time.
- **Maturation**-results may be due to maturational or other changes in the subjects and not due to the treatment being applied. This is especially important if research data is gathered over a long period of time.
- **Mortality or attrition**-losing subjects during the study could lead to different results than if everyone had stayed. Subjects with the most or least amount of important characteristics to the study may be the ones dropping out. (Think of normal curve extremes.)
- **Experimenter bias**-the responses of the subjects may be influenced by the researcher. This may occur by treating some subjects differently, reinforcing different behaviors, as well as the presence of many other variables which deliberately or unintentionally influence subjects.
- **History**-these may be the extraneous incidents (e.g., national or local tragedies or other occurrences) which influence subjects in a research project. These

events might interact with the independent variable so the results which occur may be due to a combination of the event and the variable under study.

- **Statistical regression**- sometimes subjects in a study are recruited because of extreme high or low scores (e.g., self-esteem, social skills) on the dependent variable being measured. Due to statistical regression, future measures would expect these individuals to score closer to the mean score even without any intervention.

10. External validity

An experiment is externally valid to the extent that the results may be generalized to people and situations beyond the study. There are several threats to the external validity of experiments and some of these are also threats to internal validity.

A. Selection of subjects-if subjects are not randomly selected, the results may only apply to the subjects in the study. Furthermore, the results can only be generalized to people with similar characteristics. Can we generalize results of research with college students to other adults in general?

B. Ecological validity-the research has ecological validity if the results can be generalized from one setting or circumstance to another. Sometimes the circumstances, conditions or physical surroundings of the research are so unique that the results cannot be generalized beyond that study.

C. Subject reactions (Reactivity)-if any of the following are present, the results may not generalize beyond the study:

- **Hawthorne effect**-refers to the influence in performance which occurs when subjects receive attention or know they are participating in research.
- **Demand characteristics**-these are all the cues, information, knowledge, even rumors the subject has heard about the experiment which are likely to influence his or her performance.
- **Experimenter bias**-these are the changes in the subject's behavior brought about by the researcher's expectations, behaviors or attitudes. It may also be called the Rosenthal effect. He conducted research into this phenomenon and called it the Pygmalion effect which refers to the self-fulfilling expectation of doing well because it is expected.
- **Placebo**-broadly, the placebo refers to any control treatment. The placebo should be identical to the experimental treatment except for the critical item being studied (new drug, parenting skills training, reaction to violence in films, etc). Even so, control subjects may be influenced by the placebo and react in unintended ways.

D. Novelty and disruption effect -- the measured effect of the treatment on the subjects may be due to its novelty or the disruption it causes. Being selected for a research project may be exciting and energizing; as it continues it may begin to disrupt routine and one's typical schedule. When novelty and disruption wear off or stabilize, there may be no long-term effects of the treatment.

11. Levels of measurement

The level of measurement in your data determines the statistic you can use. The four levels of measurement are:

- **Nominal:** the numbers represent the variable's qualities or categories

for example: Male and female

With nominal data, you would ordinarily use a nonparametric statistic such as chi-square.

- **Ordinal:** the numbers represent differences in some magnitude of the variable. You can order the data, from top to bottom, from high to low.

For example: Scores on a Psychology 101 final exam can be ranked from highest to lowest.

- **Interval:** the intervals between the numbers on a scale contain the same amount of the variable throughout the scale. These intervals provide a constant unit of measurement.

For example: On a standardized test, the distance (interval) between 11 and 12 is the same as the distance between 24 and 25. Other examples are Fahrenheit or centigrade temperature scales.

- **Ratio:** the numbers are on a scale which has a true zero. In addition the numbers can be compared by ratios.

For example: Someone who weighs 200 pounds is twice as heavy as someone who weighs 100 pounds. In counseling and psychology, we cannot say someone is twice as introverted as someone else.

12. Sampling

How well sampling is conducted will determine how validly we can generalize from a sample to a population. Sampling involves the selection of a part of the population.

There are several kinds of sampling. They are:

- **Random sampling:** All the individuals in the population have an equal and independent chance of being selected.
- **Stratified sampling:** This refers to selecting in such a way that major subgroups in the population will be sampled. These subgroups may be based on ethnicity, gender, age, etc.

- **Proportional stratified sampling:** This refers to randomly selecting the same proportion of individuals for the sample as they represent proportionally in the major subgroups in the population.
- **Cluster sampling:** In this sampling, the unit is not an individual but naturally occurring groups of individuals such as classrooms, city blocks, etc. Clusters are randomly selected for study.

For example: If one-half of a population is Hispanic and one-half is white, you would randomly select your sample to be one-half Hispanic and one-half white.

- **Purposeful sampling:** In some studies, there may be no interest in generalizing findings so purposeful sampling may be carried out. Possibilities are comprehensive sampling where every case or event is selected, or there is extreme-case or typical-case selection.
- **Other samples:** Other non-random or nonprobability samples may be samples of convenience or volunteer samples. These samples cannot be counted on to yield a normal distribution of scores. However, they could yield very useful and important data.

13. Sample size

Sample size influences statistical hypothesis testing. Tables for determining appropriate sample size are available. A general rule is that 5 to 10 percent of the population is adequate.

The suggested **minimum sample sizes** for different kinds of research are:

- Correlational: 30
- Ex post facto and experimental: 15
- Survey: 100

A table of random numbers lists computer-generated random numbers used to assign to individuals. Selecting from this table assures randomness.

14. Statistical analysis may be

- **Descriptive:** sometimes called summary, these techniques are used to describe the data collected for a research sample or population, and include means, standard deviations, frequency counts, and percentages.
- or
- **Inferential:** used to make inferences from the sample to the population. The goal is to determine the probability of some event occurring. Common statistical tools used for this are the t-test and analysis of variance.
- **Parametric:** used when a sample is randomly drawn from a population and the data is normally distributed. You have para (two-sided) data that yields a bell-shaped curve. Also, you assume that the variance of the sample you are

studying is homogeneous (similar) to the variance of the population from which your sample is drawn. The scores would give you a normal curve-more or less. Examples of parametric statistics are: t-test and analysis of variance.

or

- **Nonparametric:** used when you cannot make any assumption about the shape of the curve or variance of the population scores (that is, they may not be normally distributed and variances may not be homogeneous)..

For example: You want to make a statement about a school's academic performance but you only collected scores from the "gifted and talented" students. You will not have a normal distribution of academic scores for the school.

Examples of nonparametric statistics are: Chi-square, Mann-Whitney C Test, and Wilcoxon Signed-Ranks Test, all to be reviewed later.

15. Variables

- **Independent variable:** This is the variable you manipulate or vary to see what change occurs in the dependent variable. Sometimes you group or categorize the independent variable. For example, you categorize a group of individuals by gender, male and female. Or, you categorize high school students as freshmen, sophomores, juniors, and seniors. The independent variable precedes or is antecedent to the dependent variable.
- **Dependent variable:** this is the variable you are measuring or trying to change. The value of this variable depends upon the value of the independent variable you selected.

For example: the effect of three kinds of therapeutic techniques (independent variables) on anxiety (dependent variable)

The independent variable is sometimes called: **stimulus variable, predictor variable** or **experimental variable**.

The dependent variable is sometimes called: **response variable, outcome variable** or **criterion variable**.

16. Research questions and hypotheses

Some investigators ask research questions to be answered.

For example: Is there a relationship between disciplinary practices and leadership style for men in the military?

For example: Is there a significant difference in the mean number of client contacts between public and private counseling agencies?

Some investigators **formulate hypotheses (often in the null form)** to test. The null hypothesis states there is no difference between the variables or groups measured.

For example: There are no significant differences in final academic grades (GPA) between boys and girls finishing the tenth grade

One alternative **hypothesis (directional)** states that one group's score will be significantly different from another group's score (one-tailed test).

For example: The GPA of girls will be higher than the GPA of boys finishing tenth grade.

Another alternative **hypothesis (nondirectional)** states that there will be differences between the groups but which group has higher or lower scores is not indicated (two-tailed test).

For example: The GPA of girls and boys finishing the tenth grade will be different.

17. Significance level

There is always a possibility that in regard to your research results and the null hypothesis you posed, you will make a mistake in rejecting the null hypothesis when you should not, or in accepting the null hypothesis when you should reject it. The significance level you select will determine your likelihood of making either mistake. The significance level may be set by you before data analysis at .05, .01 or some other level such as .001. The level selected is your willingness to make an error that is, rejecting the null hypothesis, when in fact there is not a significant difference between the groups. (Conventional significance levels used in research are .05, .01. and .001.) At .05, you are willing to accept the possibility of rejecting the null hypothesis in error five times out of one hundred times (if you performed the experiment 100 times).

18. The following example will be used to explain concepts in the next seven items.

For example: Assume you are the director of a large counseling agency which employs 24 therapists, twelve of whom are men and twelve are women. You are curious about the self-esteem level of clients who complete therapy and wonder if there is a relationship between gender of therapist and client self-esteem. For six months, you randomly assigned clients to all counselors. As these clients terminated, their self-esteem was measured using a standardized test. After all the clients who began therapy during that six month period of time were terminated, you began your statistical analysis.

Your null hypothesis is: There is no significant difference in mean self-esteem scores of terminated clients who had been counseled by female or male therapists.

You set a significance level at .05.

19. Type I error (alpha)

Type I or alpha error are the names we use in referring to the rejection of the null hypothesis (which states that there is no difference) when it is correct. In short, you make an error.

- **Remember:** at a significance level of .05, you will reject the null hypothesis five times out of 100 when in fact there is not a real difference. You can change the probability of Type I error by changing your significance level.

For example: In the example of female and male therapists (Item 16), committing a Type I error would mean that you would say there is a significant difference in the self-esteem level of clients when in fact the difference is not significant.

20. Type II error (beta)

Type II error refers to the failure to reject the null hypothesis when there is, in fact, a difference.

For example: In the example of female and male therapists (Item 16), committing a Type II error would mean that you would say there is not a significant difference in self-esteem of clients but in fact there is a difference. You would accept (retain) the null hypothesis (in error).

As the significance level goes down (e.g., .05 to .01), Type I error decreases but Type II error increases, i.e., the failure to reject the null hypothesis when you should. As one type of error goes up, the other goes down and vice versa.

Remember:

- Alpha (Type I) error: Reject (null hypothesis) when shouldn't.
- Beta (Type II) error: Retain (null hypothesis) when shouldn't.

21. t-test

The *t*-test is used to determine whether the mean scores of two groups are significantly different from each other. It can only be used when there are two groups (two mean scores). You would compare your obtained value of *t* (from the calculation using the formula) with the value of *t* presented in a Table of *t* Values to make that determination. Computer data analysis programs now provide this information automatically.

In the example of female and male therapists (Item 18), we would apply the *t*-test to determine whether mean self-esteem scores of clients are significantly different. It should be noted that even though statistical significance may be reached, especially with large groups, there may be no practical significance to the difference.

***t*-test**

Dependent variable: Client self-esteem

Independent variable: Gender of therapist: Male

Female

22. Analysis of variance (One Way)

There are three basic kinds of analysis of variance. These are: One-way, Factorial, and Multivariate. Let's assume that in our Item 18 example, you as the agency director, notice that you have therapists trained at three levels. These levels are:

- Master's degree - licensed mental health counselor (LMHC)
- Master's degree - licensed social worker (LSW)
- Ph.D. degree - psychologist (Ph.D.)

You are attempting to determine self-esteem differences in clients following counseling and are curious as to whether level of training of therapist might lead to different levels of self-esteem.

Since you now have three groups (the levels of training) the t-test is no longer appropriate. You would use the one-way analysis of variance -when you have only one variable (level of training) at three or more levels. Completing the calculation for the analysis of variance yields an F value and this value must be compared to the values listed in an *F Distribution Table* to determine whether significant differences are present. Computer data analysis programs now provide this information automatically.

One-Way Analysis of Variance

One Dependent Variable: Client self-esteem

One Independent Variable (Factor) (3 levels):

Therapist training: Master's -LMHC

Master's -LSW

Doctorate -Ph.D.

23. Analysis of variance (Factorial)

You would use the factorial analysis of variance (ANOVA) to simultaneously determine whether mean scores on two or more variables (factors) differ significantly from each other and whether the factors interact significantly with each other. In our example in Item 18, we are comparing self-esteem scores of clients who have been counseled by female and male counselors. Let's assume we want to compare self-esteem scores with a second counselor variable as well, kind of training. Our two independent variables then are (1) gender (two kinds-female and male) and (2) kind of training. Let's stay with the same three kinds of training we used in Item 22. **The one dependent variable is the measure of self-esteem.**

This is a 2 X 3 ANOVA design. By completing this calculation, one could determine if there are any significant differences in client self-esteem based on therapist gender, or kind of therapist training, or whether there is a significant interaction between gender and training. This is called a **factorial analysis of variance** because the effects of two or more factors (independent variables) are being measured.

Factorial Analysis of Variance

One Dependent Variable: Client self-esteem

Two or more Independent Variables (Factors):

1. Gender of Therapist: (2 kinds): Male

Female

2. Therapist Training:(3 levels): Master's -LMHC Master's -LSW Doctorate -Ph.D.

24. Analysis of variance (Multivariate) (MANOVA)

When your analysis involves more than one dependent variable, you cannot use Factorial ANOVA. The appropriate statistic is multivariate analysis of variance (MANOVA). If you were measuring something else about the clients in addition to self-esteem that would be more than one dependent variable. In the continuing example (See Item 23 - ANOVA), we had two independent variables-gender and kind of training. Our single dependent variable was a measure of self-esteem. Let's assume that we also want to measure the locus of control (internal vs. external orientation) of clients when they conclude therapy. We have each client complete a locus of control scale in addition to the self-esteem inventory.

Multivariate Analysis of Variance Two Dependent Variables: Client Self-esteem Locus of Control Two or more Independent Variables (Factors): 1. Gender of Therapist: (2 kinds): Male Female 2. Therapist Training: (3 levels): Master's-LMHC Master's-LSW Doctorate-Ph.D.

Whenever you have two or more dependent variables, you cannot use ANOVA. You must use multivariate analysis of variance (MANOVA).

25. Analysis of covariance (ANCOVA)

Analysis of covariance is used similarly to analysis of variance except that the influence of one or more independent variables on the dependent variable is controlled. This means that initial group differences are adjusted statistically on one or more variables that are related to the dependent variable.

For example: In our continuing example (Item 18 and onward), we are looking at the impact of therapists' gender on clients' self-esteem. Let's assume that we do not have random assignment of clients to therapists. In fact, let's assume those most female clients are assigned to female counselors and most male clients are seen by male counselors. On the pretest of self-esteem, we find that women clients score an average of five points higher than male clients. In ANCOVA, we make use of such information which is obviously related to our dependent measure, self-esteem. If we were measuring the impact of therapist gender on clients' self-esteem, women therapists would have started out

five points higher on client self-esteem. So we statistically adjust client self-esteem scores (for example, add points to the males pretest scores and subtract points from the female pretest scores) before making the comparison.

In this case, the covariate is the systematic difference in self-esteem scores on the pretest (women clients score an average of five points higher than men clients). Consequently, we conduct an analysis of covariance -adjust the scores. Then we do the analysis of variance.

26. Post hoc or multiple comparisons tests

If your analysis of variance yields a significant F value, you still will not know which particular pair of mean scores is significantly different from each other. You must apply one of the post hoc (after the fact) tests to determine whether different particular group means or combinations of group means are significantly different. The most commonly used post hoc or multiple comparison tests are:

- **Scheffe's** (most conservative)
- **Tukey's HSD** (Honestly Significant Difference)
- **Newman-Keuls**
- **Duncan's new multiple range test**

Your computer data analysis program will conduct these automatically.

27. Nonparametric tests

When you cannot assume that your distribution of scores is normally distributed (resembles a normal curve) or that the variance of your sample is similar to the variance of the population (homogeneity) you must use nonparametric statistics.

Some examples are:

- **Mann-Whitney U test:** when you collect data from two samples that are independent from each other and the scores are not normally distributed.
- **Wilcoxon signed-rank test:** when you have scores for two samples and these scores are correlated (that is, you matched them or got two scores for each individual - repeated measures). However, the scores do not approximate a normal distribution.
- **Kruskal-Wallis test:** when you have more than two mean scores on a single variable. This is a nonparametric one-way analysis of variance.

28. Chi-square

Chi-square is another nonparametric test and is used when you have nominal data (groups or categories). This statistic is used to determine whether two distributions differ significantly.

For example, you might be curious whether more boys than girls wear jeans to high school. The null hypothesis would be: There are no significant differences in the number

of boys and girls who wear jeans to high school. You would use the Observed frequency of jean-wearing with the Expected frequency.

	Yes	No
Boys		
Girls		

This is called a contingency table. Is the wearing of jeans contingent on the gender of the student?

29. Solomon four-group design

This design examines the effect of any pretest used on the experimental treatment.

For example: Assume your subjects take a pretest measuring their commitment to establishing career goals. This pretest could sensitize them to 'the need' for career goals and interact with whatever 'treatment' (e.g., completion of aptitude and interest inventories) was planned.

The design:

Group	Pretest	Treatment	Posttest
A	Yes	Yes	Yes
B	Yes	No	Yes
C	No	Yes	Yes
D	No	No	Yes

This design allows you to determine whether the pretest by itself made a difference (Group B), whether the treatment by itself made a difference (Group C), whether a combination of pretest and treatment made a difference (Group A), or whether nothing made a difference (Group D).

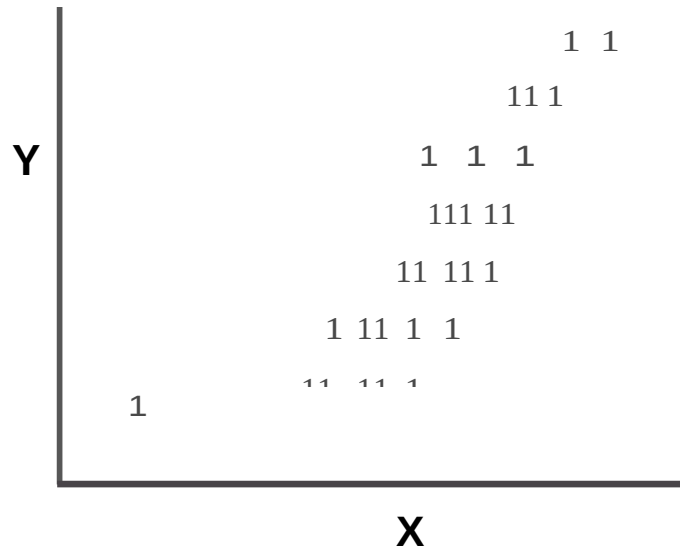
30. Multiple regression

This is the use of the correlation coefficient to determine the strength of the relationship of predictor (independent) variables on a criterion (dependent) variable. Multiple regression adds together the predictive power of several independent variables (predictors).

For example: Predictor variables such as high school GPA, class rank, and ACT scores may be used to predict the criterion (outcome) variable, which could be end-of-college freshman year GPA.

31. Scatterplot (scattergram)

This is a graphic representation of the relationship between two variables for a group of individuals. Each 1on the graph is an individual who receives a score on the X test (e.g., anxiety) and a score on the Y test (e.g., hostility). The following statement might represent the data: As individuals get more anxious about the counselor exam, their hostility increases (positive relationship).



32. Factor analysis

This is a statistical method using the correlation coefficient to determine whether a set of variables can be reduced to a smaller number of factors. For example: A factor analysis of all the items on a long inventory with 15 scales may uncover only four or five factors independent of each other underlying the scales. Thus many of the constructs of the 15 scales overlap with each other.

33. Likert scale

This is a widely used technique for measuring attitudes or opinions. It allows for several response choices, e.g., a five (or seven or nine) point scale.

34. Definitions

- **Biserial correlation**-an appropriate correlation coefficient to use when one variable yields continuous data and the other yields data that is dichotomous.
- **Cross-sectional**-studying or measuring characteristics of several groups at the same time, versus Longitudinal -- studying or measuring characteristics of a group over a period of time.
- **Degrees of freedom**-the number of observations that are free to vary.
- **Double-blind technique**-this occurs when neither the researcher nor the subject knows who is getting the active substance or the placebo.
- **Halo effect**-this is the tendency for the observer (researcher or data collector) to form an early impression of the person being observed and then letting this impression influence observations or ratings of that individual. Halo effect may be positive or negative.
- **Heteroscedasticity**-one end of a distribution of scores has more variability than the other end resulting in a fan-like appearance.

- **Homoscedasticity**-there is an equal distribution of scores throughout the range of scores, i.e., around a line of best fit
- **Inter-rater reliability**-in qualitative research, the reliability calculated by correlating the responses of several raters.
- **Observer bias**-this is the tendency of researchers to see, hear and remember what they want to.
- **Pilot study**-this is a preliminary trial or test of the research techniques and measures.
- **Placebo**-originally meant an inert drug (typically, a sugar pill); now it also means a control treatment that gives subjects the same amount and kind of attention as experimental group subjects get (this reduces Hawthorne and Rosenthal effects).
- **Rank-order correlation (Spearman rho)**-used when the values of the variables are reported in rank form rather than continuous.

35. Counseling program evaluation

The **emphasis on accountability** in the human services field accelerated in the 1970s and continues today. The push for accountability comes primarily from funders including Health Maintenance Organizations, insurance companies, government funding sources at all levels, and others. This accountability in human services reflects what the **business world adopted many years ago as the 'bottom line**. There is an acute need to demonstrate the **efficacy of counseling** in general and the effectiveness of specific theories, techniques, and approaches in particular. The emphasis on short-term therapy, often between six and twelve sessions, argues for research and evaluation to determine what works well for what kinds of problems with what clients under what circumstances. **Counseling program evaluation requires that goals and measurable objectives** be specified in advance. Without those, evaluation data has little relevance. The effectiveness of counseling techniques and processes often occurs on an individual client basis. **Evaluation** is the systematic collection of evidence of the worth of a program, process, or technique.

Two kinds of evaluation are:

- **Formative evaluation:** This is ongoing, process evaluation to measure the effectiveness of a technique or part of a program. Formative evaluation tries to determine how well a new technique, process or treatment works.
- **Summative evaluation:** This is summary or product evaluation designed to measure the effectiveness of a program, usually conducted at the end of a cycle such as a school year, fiscal year, etc. Summative evaluation is conducted to see how well agency or program goals have been met. Usually a product (document) is generated so this is 'product' evaluation versus 'process' evaluation (formative).

36. Ethical issues in research

- **Confidentiality:** no one should have access to the data except for the researcher and research assistants if any. Release of research data to others is only ethical with the consent of the subject.
- **Deception:** deception may be justifiable if no risk to subjects is involved. Such research should be followed by debriefing of the subjects.
- **Informed consent:** subjects should be informed of the research they will be participating in and give their consent.

Significance of research results should outweigh the potential benefits denied the control group. Research must be approved by a 'human subjects committee' or Institutional Review Board when conducted within institutions or agencies when federal funding is involved.

37. Writing research

Most counseling and psychological journals require that the writing style presented by the Publication Manual of the American Psychological Association be used.

Sexist language is to be avoided. Never submit a professional research manuscript to more than one journal for publication consideration at a time.

Research and Program Evaluation Study Questions

1. A school counselor is interested in determining whether seniors who have jobs receive lower grades and national test scores than those who do not work. Let's assume that in May of the school year he finds that approximately 250 seniors worked this past school year and 250 did not. He compares their GPA, class rank information and ACT scores. If all these indicators were lower for the seniors who worked, the counselor would be tempted to say that having a job causes the lower indicators. He may be wrong to say that because this study is

- A. a post hoc investigation.
- B. an experiment.
- C. a survey.
- D. a descriptive study.

2 A counselor educator at a large university studied different counseling modalities for students who came to the counseling center with relationship/self-esteem problems. There was a waiting list to accommodate all the students. As a researcher, he was curious whether men or women gained self-esteem differentially whether they were seen in individual counseling, group counseling, or small psychoeducational class (18 to 20 students). There was also a possibility of an interaction between gender and the kind of service they received. The appropriate statistical technique to use in this situation was

- A. one-way analysis of variance.
- B. factorial analysis of variance.
- C. t-test.
- D. multivariate analysis of variance.

3 In the statement: 'Birds of a feather flock together,' the dependent variable is

- A. birds.
- B. feather.
- C. Together.
- D. flock.

4. A doctoral student is curious about how bilingual kindergarten children interact verbally with monolingual children of a different racial group. She decides to study this for her doctoral research. A research method which may be most appropriate to consider is

- A. ethnographic.
- B. experimental.
- C. analogue.
- D. survey.

5. A high school career counselor had 24 seniors sign up for a three-session job seeking skills workshop. He randomly assigned 12 students to the job seeking group and 12 students to a control group. The counselor developed some new materials and techniques to be used by the 'experimental' group. In designing this study, the counselor was also concerned about

- A. providing a placebo experience to the control group.
- B. the Hawthorne effect.
- C. confidentiality.
- D. the halo effect.

6. A middle school principal is curious whether different math-teaching techniques result in different (and better) math learning. She has three classrooms of fourth grade math students and assigns different math-teaching techniques to the three teachers to use for the school year. All students completed the same math exam in September as a pretest and will posttest in May. Upon examining math pretest scores, the principal finds that the Classroom A mean math score is 72, Classroom B mean is 88, and Classroom C mean is 73. In May, after posttest data are collected, an appropriate statistical procedure for the principal to apply first is

- A. multiple regression.
- B. analysis of covariance.
- C. *t*-test.
- D. analysis of variance.

7. With the pressure for accountability, many mental health agencies ask their clients to complete evaluation forms following counseling. In assessing this self-report data, it is important to be aware of

- A. scale deficiencies.
- B. 'halo ' effect.

- C. normative comparisons.
- D. interrater reliability.

8. A mental health counselor had data on three personality tests for over 300 former clients. The tests measured over 30 variables and the counselor was interested to see if these tests might be measuring some common factors. To do this, she conducted a

- A. factor analysis.
- B. content analysis.
- C. factorial analysis of variance.
- D. item analysis.

9. A random sample of 100 students from each class at a large university was sent a computer message by the Division of Student Services. The message title simply said "Important Message from Student Services" and asked for acknowledgment of receipt. The number of hours it took the students to respond during the next seven days was measured for each of the four samples. An appropriate statistical analysis of the differences between the four classes was

- A. one-way analysis of variance.
- B. t-test for groups.
- C. multivariate analysis of variance.
- D. analysis of covariance.

10. There are three truck driving schools in the metropolitan area and each uses a different curriculum. A state accrediting agency plans to examine the effectiveness of the three curricula used for training drivers. The dependent variable in this investigation is

- A. the methods of instruction.
- B. the instructors used.
- C. the dropout rate.
- D. the sample size.

11. The developers of many psychological tests assume that the traits they are measuring can be parceled out in equal amounts up and down the scale, i.e., the numbers representing the variable have the same quantity. This is an example of _____ measurement.

- A. ordinal
- B. nominal
- C. ratio
- D. interval

12. Employing a mixed-method research design means

- A. doing a meta-analysis.
- B. using evidence-based inquiry.

- C. combining quantitative and qualitative methods.
- D. doing an experiment.

13. Antonio is investigating the differences in social skills between homeless veterans and a sample of veterans who are not homeless. As he gets to the point of data collection, he realizes that he will be limited to assessing only those veterans of each group who volunteer to participate. Because of that, when it comes time for data analysis, Antonio will probably have to use

- A. regression statistics.
- B. nonparametric statistics.
- C. parametric statistics.
- D. factor analysis.

14. A counselor has set up an elaborate experimental study which includes gathering pretest and posttest data. However, she is concerned that the pretest itself might influence or interact with the experimental treatment. To deal with this contingency, she decided to use

- A. a chi-square design.
- B. multiple regression.
- C. Duncan's multiple range test.
- D. Solomon four-group design.

15. Jefferson High School implemented a School-to-Work program which is now in its sixth year. As a condition for continued funding, the school must submit an evaluation as to the program's impact and whether the program's goals and objectives have been met. This is an example of

- A. experimental research.
- B. formative evaluation.
- C. parametric research.
- D. summative evaluation.

16. A researcher designs an experimental study on a college campus which focuses on sexual harassment and date rape. An experimental and a control group of 50 freshmen were randomly selected. The experimental group spent one hour for each of six weeks in discussion groups in September and October. In April, the researcher designed a survey to measure any differences in knowledge and behavior between the two groups pertaining to sexual harassment and date rape. Unfortunately, many students were not available to complete the instrument. Potential internal validity issues in this study include

- A. attrition and instrumentation.
- B. selection of subjects and experimenter bias.
- C. mortality and maturation.
- D. Hawthorne effect and experimenter bias.

17. In deciding to use parametric statistics such as *t*-test or analysis of variance, data from which one of the following samples would appear to be appropriate?

- A. Random sample.
- B. Sample of convenience.
- C. Skewed sample.
- D. Sample with a bi-modal distribution.

18. A researcher was investigating the possibility of attitude changes toward HIV positive individuals. To conduct this study, he had pairs of students discuss certain topics with one of the students being a researcher confederate who claimed to have the HIV virus. The researcher believed that deception in this experiment was necessary. Ethically,

- A. the researcher should have found enough HIV positive individuals to serve as confederates.
- B. the researcher should debrief the unaware subjects and provide an opportunity for them to desensitize if they revealed troubling attitudes or behaviors.
- C. the researcher should be reported to the appropriate ethical standards committee.
- D. the Human Subjects Committee should not have approved the research

19. There are a number of confounding variables which threaten the internal and external validity of an experiment. Which of the following is a threat to just one kind of validity?

- A. Instrumentation.
- B. Ecological validity.
- C. Selection of subjects.
- D. Experimenter bias.

20. To demonstrate a cause-and-effect relationship in research, an experiment must be conducted. A 'true experiment' must have all of the following **EXCEPT**

- A. random assignment of subjects.
- B. sample sizes of at least 50.
- C. experimental group.
- D. control group.

Research and Program Evaluation Question Answers

- | | |
|------|------|
| 1. A | 11.D |
| 2. B | 12.C |
| 3. D | 13.B |
| 4. A | 14.D |
| 5. A | 15.D |
| 6. B | 16.A |
| 7. B | 17.A |
| 8. A | 18.B |
| 9. A | 19.B |
| 10.C | 20.B |