Lecture 8

Refining Research Ideas and Beginning to Design your Study

Elements of Research Design: Comparison/Control Groups

- Selecting a comparison group

  Between Groups Designs
  - Compare it to another group (that is similar to research group except with respect to the treatment/construct you are measuring)

  Within Group Designs
  - Can compare one group to itself over time
    (i.e., before treatment and after treatment)

- Note: qualitative/descriptive studies do not use comparison groups – they just describe…really well

Elements of Research Design: Between Groups Design

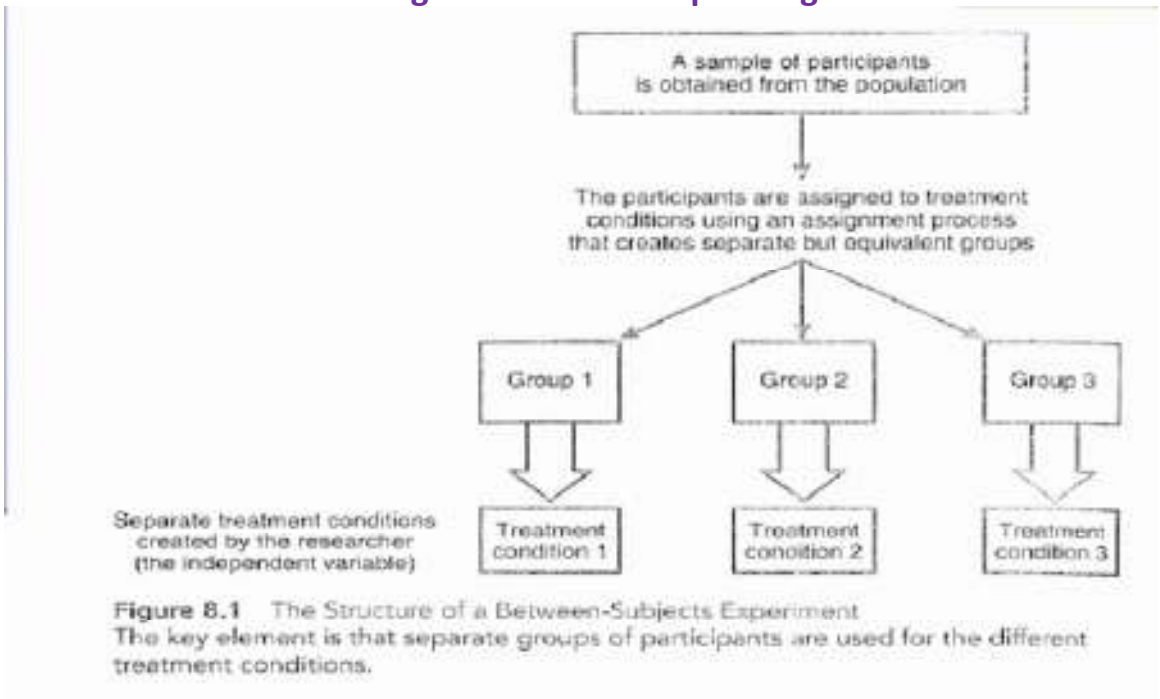

Figure 8.1 The Structure of a Between-Subjects Experiment
The key element is that separate groups of participants are used for the different treatment conditions.

Source: Gravetter, F.J., & Forzano, L. B. (2006). *Research Methods for the Behavioral Sciences* (2nd ed.). United States of America: Thomson Wadsworth.

Elements of Research Design: Within Group Design

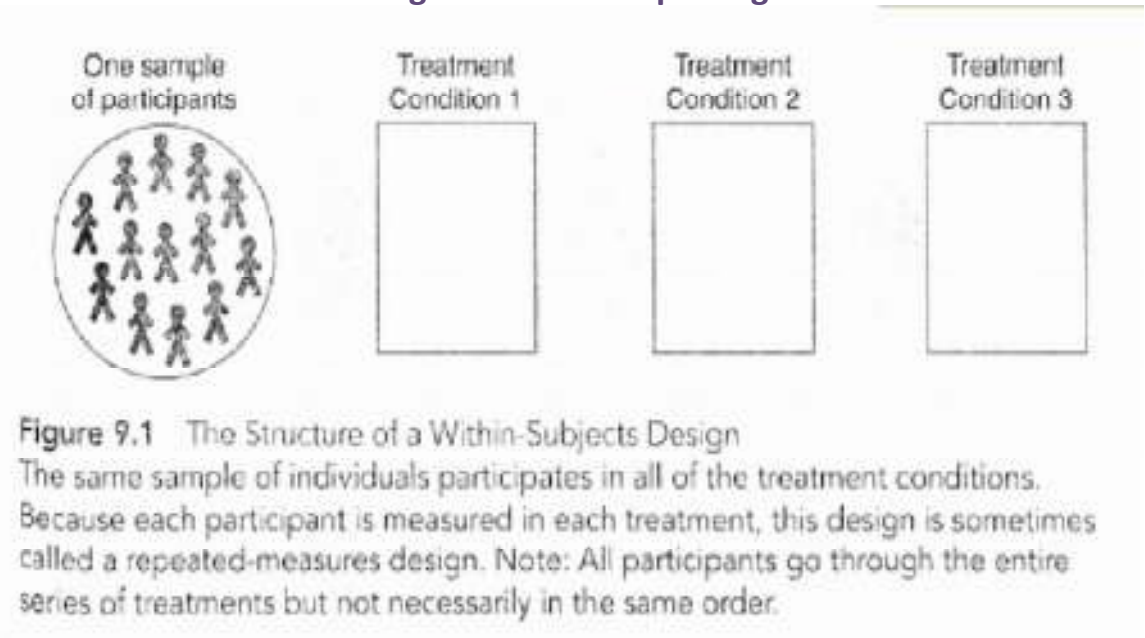

Figure 9.1 The Structure of a Within-Subjects Design
The same sample of individuals participates in all of the treatment conditions. Because each participant is measured in each treatment, this design is sometimes called a repeated-measures design. Note: All participants go through the entire series of treatments but not necessarily in the same order.

Source: Gravetter, F.J., & Forzano, L. B. (2006). *Research Methods for the Behavioral Sciences* (2nd ed.). United States of America: Thomson Wadsworth.

Elements of Research Design: One time vs. over time research

- **Cross-sectional method**
  - Same group of people are observed at one point in time
- **Longitudinal method**
  - Same group of people are observed at different points in time as they grow older

Elements of Research Design: Longitudinal Method

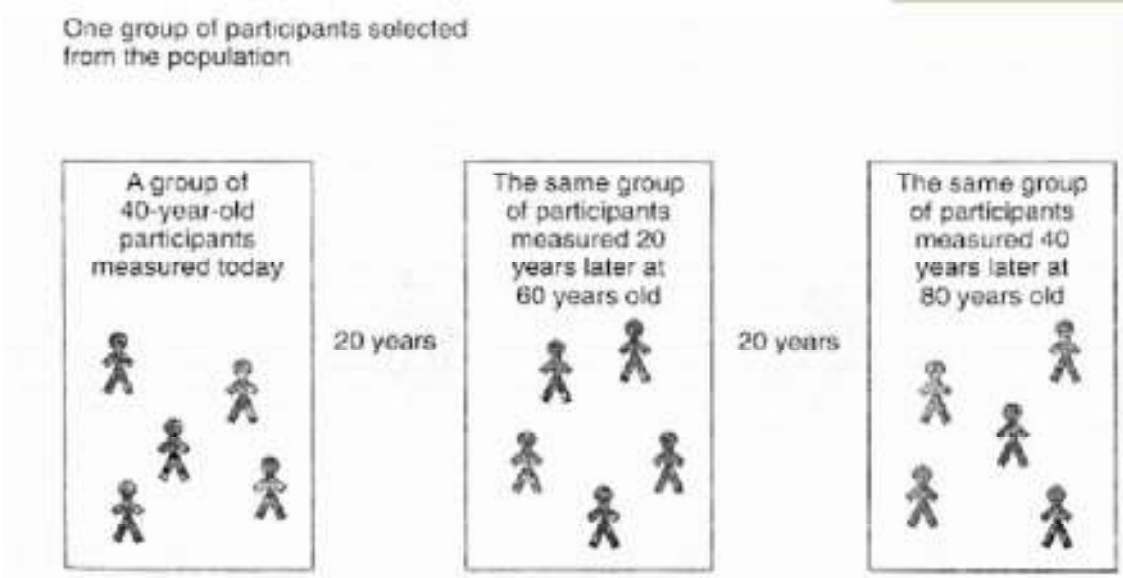

Figure 10.8 The Structure of a Longitudinal Research Design
One group of participants is measured at different times as the participants age.

Source: Gravetter, F.J., & Forzano, L. B. (2006). *Research Methods for the Behavioral Sciences* (2nd ed.). UnitedStates of America: Thomson Wadsworth.

LECTURE 9

Elements of Research Design: Defining your terms

Elements of Research Design: Defining your terms

- **Independent variable** – variable that is manipulated by the researcher (or the variable that is thought to affect the outcome/dependent variable)
- **Dependent variable –** variable that is measured to assess the effects of the independent variable
- **"Operational definition"** – procedure for measuring and defining a construct (i.e., what measures will you be using)

Turning your research question into a **Research Hypothesis**

- It is the test of your idea or theory
- A hypothesis is a statement that describes or explains a relationship among variables
- It is a prediction that is derived from your research question
  - e.g. "Shared Journey staff education will improve patient satisfaction as compared to units whose staff did not receive SJ training"
  - e.g. "internalized stigma is related to increased depression"

RH & RQ

http://privatewww.essex.ac.uk/~scholp/Hypotheses05.htm

Summary

- Start thinking about who you want to participate in your study, how you will recruit them, how you will collect the data etc.
- Decide if you want a single or multiple groups of participants and if you want to collect data at one time or over time
- Start thinking about how you can minimize/eliminate confounds and bias
- Formulate a research hypothesis

Edit By : Susan

LECTURE 10

Data Collection

Defining your terms

Defining your terms

Task 1. Words:

Counting things in spoken or written products can be harder than you think. Even 'What is a word?' can be complicated (and caused a disaster to one of my PhD students when asked this in a viva once!...). Yet one cannot count words without a definition!

In this text, how many words are there?

What problems arise in deciding on the answer?

How would you prefer to resolve them in your definition of a word?

How do you think a computer would resolve them if asked to count words?

At long last I decided that I couldn't put up with the food any longer. But having taken that decision I next had to decide where else to go to eat. I long for take-away fast food. Indecisive, I first went to McDonalds then KFC.

Task 2. T units:

Often one wants to quantify the length of text or spoken utterances. One can do this in words or in sentences, but quite popular in child and learner research are also T-units. A T-unit ('terminable unit') is essentially a main clause with a non-elliptical subject, including any dependent clauses. Thus When I got home, the new TV I ordered yesterday was on the doorstep, is one T-unit, as is John got up and left the room. However, John got up and Mary left the room is two T-units because each coordinated clause has its own subject expressed, and could stand alone as a sentence.

How many T-units in each of the following fragments of child writing?

Hence what is the average length of T-unit in words of each?

And what do you think are the supposed merits of measuring length of utterances or text in T-units rather than sentences?

a) I like branded because they always win and it is fun and I like it because sometimes he gets killed. <Note: Branded is an old TV programme set in the Wild West>

b) I like to come to school and my mummy like school and my bearther we was so happe wane I come home I eat cake and daere my tea I halp my mummy. One day I wanet in m sister house The end.

Task 4. Errors:

Categorising and then counting errors in more or less naturally produced written or spoken material used to be common. But in order to constitute a proper variable a good categorisation/classification system should: be exhaustive, have mutually exclusive categories, not mix categories of different types in one set, have unambiguously defined categories, etc.

A student came up with these categories for classifying errors: do you see any problems?

How to improve the classification system?

- Grammar errors
- Vocab errors
- L1-induced errors

LECTURE 11

Data Collection

Questionnaires

Questionnaires

Data elicited in the form of people's reports <u>about</u> language or something related

Data of this sort is most used in ELT, applied linguistics and sociolinguistics: essentially subjects report <u>about</u> what they or others do, or on beliefs about or attitudes to language, language learning etc., or on non-linguistic variables you need to record (e.g. their age, years of learning English...). One type, the grammaticality judgment task, is popular in acquisition research. Reporting ranges from (a) 'think aloud' reporting, immediate retrospective reporting after a task, open interviews, or diary type of reports to (b) structured interviews, closed questionnaires or attitude rating inventories and judgment tests. The former are heavy on Data Analysis transcribing them (if spoken) and categorising what people say, and often contain material suitable for purely qualitative analysis. The latter involve more work in constructing the Materials beforehand, and the Data Analysis may be fairly automatic (and computerisable). The more open instruments of this sort are typical of ethnographic research. All might be involved in action research, or classical research usually of the nonexperimental type.

Conventional closed questionnaires

1) Spot as many unsatisfactory features as you can in the following start of a sociolinguistic research questionnaire given to people in Wales:

Name? .....

What age category do you belong to? Under 18 years

18-21 years

21-25 years

Over 25 years

Have you ever learnt any other languages? If so, which languages?

How much do you speak Welsh at home? Often, Sometimes, Never

Do you agree that Welsh should be obligatory in schools in Wales and on official documents (e.g. income tax forms)? Yes/No

There are not enough Welsh language programs on TV. Yes/No

How many variables are being measured there? Think of more than one hypothesis one might formulate about them. How would you represent people's responses on each as a number for computer entry?

LECTURE 12

Your research variables

Your research Variables

1) How many variables are centrally involved?

We are not counting here the variables you might want to exclude the effects of… see later, just those that are central to a RQ or RH. So is this a one variable design, two variable, three variable etc. design? In the jargon: univariate, bivariate or beyond two variables it may be either factorial or multivariate (As a rough guide, it would be called factorial only where there are two or more explanatory variables in categories, see below for explanation, otherwise it would be called multivariate).

In this course we stick to two-variable designs, since understanding them properly is the key to understanding more complicated ones. In fact often a study with many variables can be broken down into a whole lot of RQs each dealt with as a two variable design. E.g. in a questionnaire you ask Taiwan senior high school learners of English their gender and also how often they use 20 different reading strategies; you also give them Nation's Levels test to check their vocab proficiency. You then potentially have a whole lot of two variable analyses (each with its own research Q or H!), involving gender in relation to each of the 20 strategies and vocab prof in relation to each of the 20 strategies (so 40 two-variable designs are analysed).

2) What roles do the central variables each play?

Often we think of one or more variables as potentially 'explaining' or 'causing' or 'affecting' or 'predicting' one or more of the others. For instance gender would be regarded as 'explaining' any differences in use of strategies we find. It would be odd to regard strategy use as somehow affecting people's gender! In the jargon, the 'explaining' variable (or variables) is perhaps most neutrally labelled the 'explanatory variable' (EV, as I prefer), but many call it the 'independent variable' (IV), or in some special design circumstances 'factor' or 'predictor'. The other variables are then 'dependent variables' (DV) or sometimes called 'response variables' etc. Sometimes there is no obvious EV - DV distinction among variables, e.g. if you are interested in the relationship between learners' grammatical proficiency and vocabulary size it is not obvious that either one is potentially affecting the other. Then regard the design as having DVs only.

There is a reason for talking in weaker terms and saying that one variable 'explains' another, or just 'is related to' it, rather than more strongly saying it 'causes' it or 'affects' it. Much language research is not experimental in the true sense, and the conventional wisdom is that it is only in a proper experiment that cause and effect can definitely be demonstrated.

3) Is this an experiment, in the strict sense?

5) What variables are or should be considered additionally to the central EVs and DVs?

These are variables that you might need to control, in the sense of 'exclude the effects of' (which I call CVs!). They may well not be mentioned in the research question/hypothesis, but are nevertheless crucial. They are things that may otherwise interfere with the results and make it hard to interpret what you discover about the central variables in the design.

You can 'control' or eliminate such variables in various ways. One is by making them constant. E.g. you choose only people in their twenties for a study comparing men and women, thus eliminating the age variation factor; for an experiment where people read two types of text (narrative and argumentative) you make all the texts at the same level of vocabulary difficulty. Another way is to randomise the variable (or, more often, claim it is as good as random, even though you have not strictly randomised it…): to eliminate age you pick men and women randomly of all ages, so hopefully you will not get a lot more older people in one group than in the other. We have already seen also the 'stratified sampling' solution to this sort of problem, where you would pick equal numbers of people of different age groups in each gender, and the use of the 'matched subjects design' which also eliminates this, if age is chosen as one of the variables to be used for matching.
If you fail to make sure relevant variables are controlled, then you may have what is called a 'confounded' design. E.g. you want to compare people's strategies depending on the rhetorical type of the text they read (narrative vs argumentative), but you use texts where the difficulty of language and unfamiliarity of topic is greater in the latter texts than the former. Then if you find a difference between text types in the strategies readers use, a critic afterwards will say 'maybe your result really shows a difference between easy and hard texts, not narrative and argumentative ones'. You will have failed to 'control' language and topic difficulty and have 'confounded' these variables with your targeted EV.

In much language research ideal control is not possible. In theory, it is only in experiments that it could be fully achieved. E.g. suppose you study learner behaviour going on in classes in a school taught by two different means (which could be either naturally occurring means or ones you experimentally impose). You will typically have to use existing classes ('intact groups') rather than take students and randomly assign them to the two method groups. Hence you cannot control whether, say, more proficient students get into one group than another. The best you can do here is to at least record as much as you can about the subjects in the two classes with a little background questionnaire. Then you can afterwards use the information about proficiency, for example, to help interpret the findings, and maybe analyse the data with the effects of prof statistically taken into account and discounted (by treating the offending variable as a ‘covariate’ in the analysis, but that is an advanced topic). Obviously the 'alternative' research paradigms do not lend themselves to control and rely heavily on delicate interpretation by the researcher of how all the uncontrolled factors might have affected what is observed.

LECTURE 13

Results

RESULTS IN GENERAL: THREE STATISTICAL THINGS TO DO WITH RESULTS

(a) Presentation. Mainly presentation consists of making easy to understand tables, and especially graphs of various sorts, to go in the main text and show the key features of the results (e.g. histograms, bar charts, scatterplots, line graphs of various sorts). For these tables and/or graphs, frequencies of people falling in a category may be converted to %, etc., for easy understanding, and often what will be presented are descriptive statistics derived from the data (see b), rather than scores or whatever of each case separately.

(b) Descriptive statistics. These are figures you (get the computer to) calculate from a lot of specific figures which arise from data. Essentially they summarise certain facts just about the specific cases you studied. Hence they are referred to as 'statistical measures' based on 'observed' data, sometimes referred to as O (=observed) figures for short (cf. 'statistical tests' in c which go beyond just what has been observed about samples). Mainly they are of one of the following types, depending on what kind of thing about your people/words/etc. they measure:

-- (b1) Measures of centrality. These in some way indicate the one score or category that you might choose to represent a whole set of scores or categorisations for one group of cases on one variable. These are mostly familiar measures from everyday life. One example is the "average" score of a set of interval scores (technically the Mean). Another, where you have cases that have been put in categories, is the category that the greatest proportion of people chose or fell in

-- (b2) Measures of variation. These summarise how far the individual scores were closely spread round some central measure, how far they were widely spread. In a way they measure how closely the scores (or people who scored the scores) "agreed" within a group, on a scale running upwards from 0. The higher the figure, the greater the variation. Examples of such measures are the Standard Deviation (and related notions Variance and Error) for scores, Index of Commonality for categories.

(b3) Measures of difference. These summarise the amount of difference between pairs of samples or groups measured, or between scores the same group obtained in different conditions, usually by a figure that is the 'difference between two means', or the 'difference between two percentages' (percentage difference). Again such figures normally run upwards from 0 (= no difference) to any size.

(**b4) Measures of relationship.** These quantify the amount of relationship between two (or more) variables as measured in the same group of people or whatever. They are usually on a scale 0-1 (in some instances they run from -1 through 0 to +1). I.e. if such a measure comes out near 1 (or -1 where relevant), that indicates that those cases that scored a particular value on one variable also tended to score a particular value on the other. E.g. those who scored high on motivation also scored high on proficiency. If it comes out near 0, that indicates that cases that scored a particular way on one variable scored all over the other variable, and vice versa. Examples are the Pearson 'r' Correlation Coefficient, the Spearman 'rho' Correlation Coefficient, Kendall's W, the 'phi' Correlation Coefficient, Kruskal's 'gamma'. (Remember that relationship and difference are really the same thing looked at from different points of view. If there is a difference between men and women - the two values of the gender variable - in attitude to RP accent, then there is a relationship between the variables gender and attitude to RP accent. It is just that for technical reasons sometimes statistics approaches the matter more via measuring difference, sometimes via measuring relationship).

If you are only interested in the particular cases or groups of cases you measured in themselves (e.g. because they are the whole population of interest), then (a) and (b) probably provide the answer to any questions or hypotheses

you had about them. But usually in research you have not measured everyone/thing of interest directly, but only samples, and wish to generalise, hence inferential statistics are also needed.

(c) Inferential statistics. These in some way enable you to generalise from the specific sample(s) you measured, and the descriptive measures of them (O's), to a wider 'population' that you sampled (if that is of interest to you, of course). Most descriptive statistical measures have associated inferential statistics.

In effect then, the input to inferential

- **the level of certainty is** about what inferential stats tells you that you will be satisfied with. No inferential stats give you 100% certainty of anything. I.e. statistics can never tell you that, based on the difference between 3rd graders and 4th graders you found in your samples, it is 100% certain that there is a difference between 3rd and 4th graders in the populations your samples represent. You have to choose to be satisfied with something less than 100%. 95% is commonly taken as adequate in language research: this is the same as choosing the .05 (or 5%) level of significance as the one you will be satisfied with. (Statistics actually works with the chances of being wrong about a difference rather than being correct, hence 5% not 95%). If you adopt that level, then if a statistical test comes up with a significance of less than .05 for some difference or relationship you are interested in, then that is the same as saying that there is a 95% or more certainty that there is a population difference/relationship, not just one in the sample. So you will take it that a difference or relationship is proved to be real in the population(s) as well as the sample(s). If you adopted .01 as the threshold then you would only be satisfied if the test came out with a significance smaller than that (You would be demanding 99% or more certainty).

Significance tests. These deal with hypotheses about 'differences' or 'relationships', which is why it was a good idea to think in these terms when formulating hypotheses and planning what to do in the first place - before actually starting gathering data. They tell us if a difference or relationship we have observed in samples is strong enough to indicate a 'real' difference/ relationship in the populations sampled or not.

Suppose you are comparing the attitudes of men and women to RP. You find an observed difference between the results for two samples (one of men and one of women) - i.e. the sample difference between the two average scores for attitude to RP English is not zero. So clearly the samples are, descriptively, different, but what can you say about the hypothesis about the populations of men and women that you sampled (since it is this "large-scale" hypothesis that you are really interested in)? Common sense says that you could get small differences between samples of men and women without there being any real population difference between men and women, just because samples from populations don't exactly reflect those populations in microcosm. Something called 'sampling error' always comes in. What you want (though you may not realise it!) is to be told a probability: you need to know the probability that you would get a difference the size of your observed one between samples if there were no population difference. If the probability is remote (say 5% or less ($p<.05$) - the common threshold chosen), then you will conclude that your samples are evidence for a population difference and will say that the difference is, technically, 'significant'. But if the probability is reasonably large (bigger than 5%, $p>.05$ say), then it is not safe to regard the "no difference" hypothesis as rejectable. The main bit of information you get from any significance test is therefore a probability, which may be referred to as p or sig.

http://privatewww.essex.ac.uk/~scholp/onevardesc.htm

LECTURE 14

Revision & Final Exam

Revision & final Exam

A hypothesis is:

- A hypothesis is a statement that describes or explains a relationship among variables
- A hypothesis is a statement about your research
- A hypothesis is a statement about the problems in your research
- A hypothesis is a statement about the outcome of your research

The independent variable is:

- the variable that is thought to affect the dependent variable
- the variable that is thought to affect the hypothesis
- the variable that is thought to affect the results
- the variable that is thought to affect the abstract

Research is:

Looking for knowledge only
Looking for data only
Looking for new ideas and findings
Looking for previous studies

An Abstract is:

A summary of the whole thing
A summary of the whole results
A summary of the whole literature review
A summary of the whole methodology

A good classical report will consist of:

Abstract- methodology- results-introduction
Abstract-literature review- results-introduction
Abstract-introduction-literature review-methodology-results
Abstract-results-introduction-literature review

In the introduction:

You introduce the results
You introduce the study and its significance
You introduce all previous studies and a critique for them
You introduce all the methods and instruments you used

In the literature review:

You talk about the results
You talk about the study and its significance
You talk about all previous studies and a critique for them
You talk about all the procedures used

Plagiarism is:
Representing other authors' language and ideas as your own original work
Representing your own language and ideas as your own original work
Representing other authors' language and ideas as their own original work
Representing other authors' language and ideas as a plagiarised work.

The dependent variable is
The variable that is affected by the independent variable
The variable that is dependent on the hypothesis
The variable that is affected by the abstract
The variable that is affected by the results

The significant difference has to be at the level of:
P= 50
P=.05
P=.50
P=0.50

If you have one variable in your research, then it is:
Multivariate
Univariate
Bivariate
factorial

We use questionnaires in research as a:
tool to collect data
tool to analyse data
tool to generate results
tool to design research