Simin Hojat, PhD, M.Phil

1/16/2017

Research Methodology: Quantitative Research

Introduction

Social research was about producing knowledge of the social world that was captured through the epistemological questions of how the social world can be known in regard with the issues of belief, assertions and propositions (Pascale, 2011). Newell (1990) stated perception occurs through the stimulation of sensory neurons in the form of seeing, hearing, touching, smelling, and tasting. Further research by Boeree (2003) proposed that perception entailed not only sensing things but also making the relationship between things as well (Boeree, 2003). Therefore, our perception of the world is through making relationship between natural and social phenomena.

However, it was the seventeenth century that science and scientific research became the center of attention; in fact, seventeenth century "distinguished the modern world from earlier centuries" (Russell, 1974, p. 512). The philosophers such as Descartes (1596-1650), Kepler (1571-1630), Galileo (1564 – 1642), Newton (1642-1727), and Bacon (1561 – 1626) were pre-eminent in the creation of science. Seventeenth century was a pronounced era of growth of science and technology (Gilbert, 1540-1603; Harvey, 1578-1657; Leeuwenhoek, 1632-1723). These scientific miracles transformed the outlook of educated man completely; no more witchcraft stories and no medieval thinking. Among these philosophers Francis Bacon (1561-1626) was the pioneer in the attempt at logical systematization of scientific research. There is a famous statement of Francis Bacon and that is: “Knowledge is Power”; it simply says the way that man acquire mastery over the forces of nature, is by scientific discoveries and inventions (as Cited in Russell, 1974, p. 521).

However, historically, science was resembled to nature and scientific research referred to as the study of nature. In other words, formulating, unfolding, learning, and understanding this process was embodied in what was named scientific research. Scientific research was based on a world that existed, was known to us through our senses, and had governing order that must be discovered (Singleton & Straits, 2005). Thus, the function of scientists was to describe and explain the presumed order to the world and made inferences through different disciplines with the help of deductive reasoning. Besides, with these assumptions scientists only dealt with problems that could be resolved through observations. Moreover, the empirical evidence was external to the scientists; and scientists were objective in a sense that they were “free from emotion, conjecture, or personal bias” (p.29). These assumptions were the core base of the positivist worldview and the design that was developed, further, based on that outlook and was named quantitative research.

Therefore, according to this tradition, the social research was considered the same as scientific research and took the methods established in the physical sciences. Auguste Comte (1798- 1857) was one of the first researcher who asserted that “authentic Knowledge comes from personal experience rather than from metaphysical or theological foundations” (as cited in Pascale 2011, p.13). Comte who was considered as the founder of positivism argued that scientists can discover the laws governing empirical events by relying on observable facts and relations among them. (Hawkesworth, as cited in pascale, 2011, p.13).

Thus, from this perspective, positivism presumed that social laws were found the same as natural laws and the world existed as an objective entity and was knowable in its entirety. Positivism was an ontological position that assumed the existence of a reality that could be objectively described or measured. In fact, positivism through the oncological belief that a single reality existed and the epistemological claim that this reality could be known objectively argued that there was one correct logic for scientific inquiry and that was positivism; as such, "positivism served as the methodological foundation of early social sciences" (Pascale, p.14.).

Historically, the positivist paradigm became the backbone theory for quantitative research. However, later it was challenges in a way that it was not possible to get to the absolute true knowledge (Phillips & burbules, 2000). Moreover, "we cannot be positive about our claims of knowledge when studying the behavior and action of human" (Creswell, 2014, p.7). As a result, post positivist perspective came out from the work of 19th century writers such as Comte, Mill, Durkheim, Newton, and Locke (Smith 1983, as cited in Creswell, p.7). According to post-positivists view, causes would probably determine the outcome; therefore, the researcher should identify the causes that determine the outcome. Therefore, the positivists perspective that reality can be fully known changed into the post-positivism; "reality can be known within some level of probability**"** (p.77). With the same reasoning, whereas in positivism, objectivity can be achieved in the research process, in post-positivism, objectivity is a goal that may be achieved.

Positivist research can be characterized by a quest for deterministic view which is expressed through a reductionist reasoning in a way that to reduce the ideas to small, discrete set to test. The knowledge that developed through a post-positivist lens is based on careful observation, and measurement of the objective reality that exists in the world. Finally, there are laws or theories that govern the world and they have to be tested. In other words, a researcher attempts to determine a cause and effect relationship between identifiable objects with a reductionist policy that reduces the research problem to a specific variable, hypotheses, and questions. Then, the researcher produces statistical data through measurement, observation, instrumentation, employing appropriate design and would tests the theories. Applying quantitative methodology in social research is, therefore, similar to natural science and is based on positivist approach (Bryman, as cited in Creswell, 2014).

The characteristics of positivist approach are: “preoccupation with operational definitions, objectivity, replicability, causality, and the like “(Creswell, 2014 p. 77). For example, in a survey research which is very typical in quantitative approach, the use of questionnaire separates the observed from the observer, can be replicated when the same instrument is used in different context. The quantitative methodology which carries the positivist or empiricist assumptions should also follow the epistemological argument and be pinned on a solid available theory, as an acceptable knowledge (Bryman, as cited in Creswell, 2014). In quantitative research, questions and hypotheses are related to variables of the study. Quantitative research questions inquire about the relationships among variables, while quantitative hypotheses are predictions the researcher make about the expected outcome of those relationship.

Moreover, on the purpose or goal of the study, in quantitative studies it typically starts with a generalization or general hypothesis based on existing theory and knowledge and the goal of researcher is to determine whether statistical support exists for the hypothesized relationship. Variables are measured through instrumentation; that is being measurable and numeric, even if subjective phenomena are involved in the research. This is done through surveys, tests, experiments, and quasi experiments. Finally, if a relationship is tested and proved positive repeatedly then, there is an established theory that is added to the body of knowledge and can be inferred deductively from macro level to micro, or from top –down. Besides, in regard with data collection, in quantitative research data, variables or constructs, have to be observable or measurable, even when subject of research is about feelings, beliefs, intentions, prior behavior, effects, and so on. They needs to be reduced to, or reframed as, numeric data and the conclusions and final discussions are based on statistical significance and in numerical form.

Different Designs in Quantitative Research

Research design is "the glue that holds the research projects together" (Trochim, 2001, p. 171). In other words, it shows how the research is structured and how the whole procedure of research is put together. Different strategies in quantitative research have changed with the advancement of technology and economy during the 19th and 20th centuries. These inquiries all embraced the post-positivist, empiricist paradigms among which survey research, true experiments, and the ones that are less rigorous, quasi-experiments, and correlational inquiries (Campbell and Stanley, 1963). Furthermore, there are also specific single-subject experiments (Cooper, Heron, & Heward, 1987; Neuman & McCormick, 1995), factorial designs, and repeated measures designs (Creswell, 2014). However, the most used designs in quantitative research are survey research, experimental, and quasi-experimental research.

Survey Research

In survey research, there are usually a large number of respondents, called the sample; the sample must be a true representative of population in character and behavior. The purpose in a survey research is that certain attitude, behavior, or character of the sample to be generalized to population. The methodology is based on asking a lot of questions from participants, questions can be open ended, or closed ended, direct, or indirect. Answers to be obtained via mail, email, or personal interviews. The answers are numerically coded and the data will be analyzed. Data analysis techniques depends on whether the survey's purpose is descriptive, explanatory, or combination of both; in descriptive analyses, researchers describe the concluded results from sample and generalize it to population (Creswell, 2014). Also, survey can be cross sectional or longitudinal, and the sample can be single individual, or a group, cluster, random or non- random (. (Singleton& Straits, 2005).

In survey design of inquiry, the purpose and rationale for study is to generalize from a sample to a population so that the concluding results about characteristics, behavior, and attitude of the sample might be generalized to the population. Conducting a survey research, in comparison with other methods of inquiry, is preferred for this research design because it is easier to collect data from a sample, it is more economical, and it has a quicker turn around (Creswell, 2014). Collecting data from the sample can be cross sectional or longitudinal. Cross sectional data collection is collecting data from the sample at the same time. Longitudinal way of collecting data is collecting data from sample individuals over time. Forms of data collection in survey studies like other type of studies might be "mail, telephone, internet, personal interviews or group administration" (Fowler, citing in Creswell, 2014, p.157). Instrumentation in the research is either modified or intact, an instrument developed by someone else. If it is intact the researcher must have permit to use it and the validity and reliability of scores obtained by this instrument in the past experiences must be tested.

In survey research inquirer seeks for correlational relationship between independent and dependent variables and the purpose and rationale for study is to generalize from a sample to a population so that the concluding results about characteristics, behavior, and attitude of the sample might be generalized to the population. Questions are closed-ended and should be designed on the basis of Likert scales. Since testing the hypotheses is essential in quantitative research, research is based on large number of participants and therefore collecting data is through questionnaire and interviews without direct supervision of the researcher.

Population and Sample

The objective of a researcher is to test the hypotheses and predict the expected relationship between variables in a specific population. In order to do that, researchers collect data, and get the result that has to be generalized from a sample to the population in research. It is however, impossible for a researcher to administer the research procedure on the entire population covered by the research problem. Therefore, researchers apply the testing on small number of population (a sample) and infer the results to the population. Therefore, sampling strategy is an essential part of quantitative research projects which ensure that sample is a true representative of the population in the study.

Nachimas and Nachimas (2008) defined population as “aggregate all cases that conform to some designated set of specifications” (p. 163). This applies to population in a specific country, or a specific school and so forth. In the research projects, the specification that determines population is research problem. Thus, population for a study is what is covered by the research problem and is assigned for a particular test or study. Population in this respect refer to individuals, groups, government, organizations, and so on. Thus, the first problem researchers are facing in a research is how to look for proper type of population that accommodates their research projects. Thus, population must be “defined in terms of (1) content, (2) extent, (3) time” (P. 164).

Population in the study is the number of individuals or groups that are the subject of study. Samples are individuals that are selected by researchers to be studied. The size of the sample depends on the methodology used in the study and it is more important in quantitative research than qualitative. Since researchers take a fraction of population as sample, it is useful for researchers to have pre-sample information about population. Sample design includes form of sampling, single or multiple which is called clustering. Random sampling, that is mostly recommended for quantitative survey research, is the random selection of individuals from population so that each individual has some probability of being selected. In this case, there is a higher probability that the result can be generalized from sample to the population. It is also possible that to have different distinct groups in the population, called stratification; that is, male, female (Creswell, 2014). The researcher should verify the site of investigation and get the permit for accessibility to the site.

Type of sampling

After researchers determine the population involved in research, they have to set a sampling frame. Sample frame consists of all units of samples and might not cover all unit of analysis in the population. Sometimes some units of sample are not accessible for research; therefore, the total units in the sampling frame will be less than population of research. Sample is drawn from the sample units in the sample frame. Sample unit is a single unit is a sample frame and can be individual, event, a nation, or documents. A good sample is the one that produces the same outcome as if research is administered with all unit of analysis, or population (Nachmias & Nachmias, 2008). In this case we can claim that the sample is a true representative of population. There are two types of sampling: Probability and nonprobability sampling. In order to get the best sample, all unit of analysis must have the same chance of participation in sampling; this is called probability sampling. Therefore, only in probability sampling a sample can be a true representative of the population and only probability sampling can be used in sample design. This is because in probability sample design, the researcher can use statistical measurement and estimate the degree to which the outcome based on a sample differs from the observations from studying the entire population. However, sometimes probability sampling is not attainable due to the nature of study. In these cases, nonprobability sample design is implemented by researchers.

There are four types of probability sampling: “Simple random sampling, systematic samples, stratified samples, and cluster samples” (Nachmias & Nachmias, 2008, p.169). Simple random sampling is assigning an equal and nonzero probability of selection to each sample units in the sample frame. Nowadays, random sampling is done by computer program or tables. Systematic sampling is a procedure that the first unit is selected randomly and after that every kth unit will be selected. In this sampling every sample unit has a 1/kth probability of being selected. Although this system of design is more convenient for researchers, it will cause a bias in the sample. Stratified sampling is primarily used by researchers to make sure that different group of population are represented in the sample. Cluster sampling is often used in large scale studying and it is the least expensive sample design. Nonprobability sampling is administered by convenience, purposive, and quota sampling (Nachmias & Nachmias).

Sample Size

Sample size is the total number of sampling units that are selected by researchers for the experiment. It can be any number; therefore, it is important to know how researchers should determine the sample size. Sample size is another element that establishes whether a sample is a true representative of population. In order for researchers to determine size of a sample in a research, they must know what level of accuracy they expect to get from their research; the size of standard error that is acceptable to them. Refer to the definition of a good sample as a sample that results in closeness of sample outcome with true value of the parameter in research, the statistical measurement that relates to this quality is standard error. Researchers usually use tables or power calculators to calculate the adequate sample size for the study (Nachmias & Nachmias, 2008). The three pieces of information that researchers need to know before calculating sample size are: Statistical power, alpha, and effect size (Burkholder, 2013). Statistical power is the probability that a given statistical test will detect whether true relationship exist between variables in the population. For this to happen, a large enough sample is needed. Also, a proper sample size establishes good statistical measurement for the research and ensures that the null hypotheses will be rejected.

Data Analysis

In the final proposal of research, a report should be given about the respondents and non-respondents in the survey and to estimate if there is any response bias (Fowler, as cited in Creswell, p.162). The researcher should give information about the independent and dependent variables and the statistical procedures for testing the score of variables. If the number of participants are small and the analysis is descriptive, there is no need for further information. However, in more advanced research there should be statistical description of the research program and presenting the result in tables or figures is necessary. APA 2010, suggests that the most complete meaning of the result come from reporting extensive description. The statistical significance test reports whether the observed scores reflect a pattern other than chance. A confidence interval is a range of values that describe the level of uncertainty around an estimated observed score. Effect size identifies the strength of conclusions about group differences.

Experimental Design

Experimental design is a design commonly used in social research in which intervention is involved and is seeking causal relationship between an independent variable, treatment variable, and a dependent variable or outcome. Singleton& Straits (2005) stated that despite the fact that many see a lot of resemblance between scientific research and experimentation, much of social research is non-experimental. However, they consider this method to be the best to test the causal hypotheses. They pointed out that the key feature for experimental studies is manipulation and control through which the researcher can control over the conditions of observation. Also, by changing the conditions, testing the hypotheses, and measuring the change. In experimental design of research researchers are seeking the relationship between variables, called hypotheses. Individuals can be selected on a random or non-random basis. When individuals are not randomly selected, the procedure is called a quasi- experiment and when individuals are randomly assigned into groups, the procedure is called the true experiment. Variables need to be specified in an experiment. Contrary to survey method, in experimental research, the relationships, hypotheses, are important, therefore researchers must identify independent and dependent variables.

Experimental design is a design in which a treatment is involved and is seeking for a causal relationship between an independent variable, or treatment variable, and a dependent variable, or outcome. In this design, the researcher usually selects two groups; experiment group and comparison group. The idea of the research is to manipulate one group with the treatment and then compare the posttest result with the other group that is not manipulated. Selection for participants and sample selection either is done randomly or non-randomly. A randomized experiment generally is the strongest when your interest is in establishing a cause and effect relationship. In case of random sampling it is called true experiment and there is higher possibility that the sample represents the population; this is because, each individual has equal probability of being selected. In true experiment, there is no systematic bias, because the procedure of sampling eliminates the systematic differences among characteristics of participants that can affect the outcome (Creswell, J., 2014).

The experiment is conducted on two equivalent groups, one of which is given the 'treatment' (the independent variable) and the other 'control' group is not given the treatment. The following table displays the classic design where X is the treatment – the independent variable applied, O_1 and O_3 are the pretested two randomly selected two groups, X is the treatment, the test or the one independent variable applied to one group while O_2 and O_4 are the results from the post-test.

Table 1

The Classic Experimental Design

Group	**Pretest**	**Posttest**	**Difference**
Experimental R	$O_1 \longrightarrow$ x	--- $\longrightarrow O_2$-	$O_2 — O_1 = d_e$
Control R	$O_3 \longrightarrow$	--- $\longrightarrow O_4$-	$O_4 — O_3 = d_c$

Source: Frankfort-Nachmias & Nachmias (2008, p. 91)

There are usually two groups in experimental designs: Control group in which the researcher randomly selects and assign participants, but does not expose them to treatment and the treatment group that has to be equally assigned and selected but is exposed to the treatment. Pretest and posttest by researcher shows the effect of the treatment. The threat to validity in this research, true experiment, is minimized due to random sampling, random assignment, and including the control group. The merit in minimizing the threat to internal validity is to control the preexisting subject differences and by making sure that the events occurring within each experimental conditions are exactly the same except for the manipulated independent variable, the principle is to allow only one factor, the independent variable to vary while controlling the rest.

The purpose of experimental research is to trace a causal relationship between an independent variable, determining factor, and dependent variable, determined factor. A valid causal relationship in a research is only possible, if a researcher can control all extraneous variables and only let the independent variable vary. This is only conceivable, if a researcher can eliminate all pre-existing intrinsic factors pretest and control the events that occur during the test. One big condition for not allowing the pre-existing intrinsic factors to intervene in the treatment and distort the outcome is through the procedure of equalization or equivalency of groups. If groups are not identical from the outset, the researcher does not know whether the outcome is a result of treatment or the other extraneous element. In case extraneous influence, any inferences from the independent to dependent variable does not have validity. It is almost impossible to create identical groups, even with pairwise matching, because research groups consist of human being with different background and attitude. Therefore, we have to rely on the idea of "probabilistically equivalent or equivalent within known probabilistic ranges", (Trochim, 2001, p. 193).

According to Singleton& Straits (2005) despite the fact that many see a lot of resemblance between scientific research and experimentation, "much of social research is non-experimental" (p. 155). However, they consider this method to be the best to test the causal hypotheses. They pointed out that the key feature for experimental studies is manipulation and control, through which the researcher can control over the conditions of observation. .

Other Types of Experimental Design

There are other types of experimental designs: "Pre experimental and single subject design, factorial design, and within group design" (Creswell, J, 2014, p.171). In pre experimental design, a researcher studies a single group and provides an intervention during the experiment. The design does not have a control group to compare with experimental group. In single subject design, the researcher is observing the behavior of single individual or a small group over time. In factorial design, the researcher uses two or more treatment variable to examine the independent and simultaneous effect on outcome; this is also called between group design. In within group design participants are assigned to different treatment over time or a single individual is observed over time. There are other types of research design stated by Nachmias & Nachmias (2008). These are: Cross sectional design, pre-experimental design, the Solomon four-group design, and the posttest only control group design.

Solomon Four-Group Design

In this design there are four groups, two groups with pretest, and two groups that did not have the pretest. In groups 3 and 4, one gets the X treatment, one does not, but neither experiences the pretest; thus, it minimizes all of the internal and external validity threats discussed below. The Solomon Four-Group Design can be costly or impractical to create.

Table 2

The Solomon Four-Group Design (Experimental)

	Pretest		Posttest	Differences	Comparison
R	O_1	X	O_2	$O_2 - O_1 = d_e$	$O_2 - O_5$
R	O_3		O_4	$O_4 - O_3 = d_c$	$O_4 - O_6$
R		X	O_5		
R			O_6		

Source: Frankfort-Nachmias & Nachmias (2008, p. 104)

Posttest-Only Control Group Design

Table 3

The Posttest-Only Control Group Design (Experimental)

			Posttest results
R		X	O_1
R			O_2

Source: Frankfort-Nachmias & Nachmias (2008, p. 106)

A further variation used in long term effects; in a longitudinal study. This example uses the Solomon Four design however there are two posttests.

Table 4

Delayed Effect Design (Experimental)

	Pretest		Posttest	Post 2	Differences	Comparison
R	O_1	X	O_2		$O_2 - O_1 = d_{e1}$	$d_{e1} - d_{e2}$
R	O_3		O_4		$O_4 - O_3 = d_{c1}$	$d_{c1} - d_{c2}$
R	O_5	X		O_6	$O_5 - O_6 = d_{e2}$	
R	O_7			O_8	$O_4 - O_3 = d_{c2}$	

Source: Frankfort-Nachmias & Nachmias (2008, p. 107)

Factorial Design

When an investigation requires two or more independent variables, experiment requires a factorial design in which more groups are involved and the effects of each possible combination of variables will be tested. For example, in case that we have X and Z as independent variables, we have the following chart.

Table 5

Combination Factorial Design- two independent variables (Experimental)

		Posttest	Comparison
R	$X_1 + Z_1$	O_1	$O_1 - O_2$
R	$X_1 + Z_2$	O_2	$O_1 - O_3$
R	$X_2 + Z_1$	O_3	$O_1 - O_4$
R	$X_2 + Z$	O_4	$O_2 - O_3$
			$O_2 - O_4$
			$O_3 - O_4$

Source: Frankfort-Nachmias and Nachmias (2008, p. 108)

A factorial design may be useful in external validation to generalizability. Rather than controlling for outside factors, major outside influences can be included in combination of this design (Frankfort-Nachmias & Nachmias, 2008).

Quasi-experiment Design

However, due to ethical and other factors sometimes it is not possible for a researcher to do random sampling in the experiment. Therefore, researchers have adopted other types of research designs that has no random sampling; this type of research that is the most frequently used in research is referred to as quasi experiment. In this design, there is a pretest and a posttest for two groups: experiment and control groups, but there is no randomization in the sampling selection or sampling design. The idea of the research is to manipulate one group with the treatment and then compare the posttest result with the other group that is not manipulated. The most often used type of design in quasi experiment is “nonequivalent- group design (NEGD). As the name of the design indicates in NEGD there is no equality or similarity between groups, and thus, contrary to the true experiments, its validity is dubious in reflecting the true cause and effect relationship. In order to get the valid results from this research, groups selected should be equal or similar. However, in this design, due to non- randomization, there is no guarantee that the groups are similar and this non-equivalency is the biggest challenge of this design. Thus, there is a threat of internal validity in this design due to non-randomization in selecting the groups. Therefore, we have to focus on selection threat when we are discussing threat to internal validity in this design.

As the comparison group did not change between the pretest and the posttest, selection – maturation cannot be traced in this design, unless there is a long time between pretest and posttest and groups matured at different rate which is very improbable. However, there might be a selection history threat, due to the fact that non-equivalent groups were affected differently by external events regarding the research, or some events happen for one group, and not for the other. Also, there might be selection regression which is a change due to regression to the mean. If the program group starts below the overall population pretest average, it will regress upward on the posttest. Sometimes the program group will cross over the comparison group which is called “cross over program” (Trochim, 2001, p.220). Furthermore, selection testing, selection instrumentation and selection mortality are also possibilities of threat to internal validity in this design.

According to Nachmias and Nachmias (2008), the procedure of equalization between control group and experiment group can be conducted in three ways: Matching, randomization, and control group. Matching is done when researchers have pre-knowledge about intrinsic factors in the research. Thus, they match the groups in a way that every case in one group to be exactly identical with the case in other group. Matching can be done in the form of precision or frequency distribution methods. Precision matching or pairwise matching is when for each case in experimental group another identical case is selected for control group. If pair-matching is done perfectly, then the investigator can conclude that any differences found between experimental and control group is not due matched variables, but it is because of the change in independent variable. However, it is difficult and time consuming to do precise matching when we are dealing with large numbers. In frequency distribution, on the other hand, "the experimental and control groups are made similar for each of the relevant variables separately rather than in combination" (p.99). For example, in case of age the average age of one group should be the same as average age in the other group.

In the quasi-experimental method of research, there is neither randomization, nor control (Frankfort-Nachmias & Nachmias, 2008; Tochim, 2006). However, in the quasi-experimental methods researchers employ multiple measures in the data analysis and can combine design methods to make causal inferences "by systematically combining two or more designs in a single study" (Frankfort-Nachmias & Nachmias, p. 129).

Other quasi experimental designs are the planned variation design that uses multiple stimuli to every individual or group within a study, the panels design that studies one sample group which is examined multiple times over a longitudinal study period, and the time-series design is useful when no control group can be assembled ((Frankfort-Nachmias & Nachmias, 2008).

Cross-Section Research Design

Similar attributes to the quasi-experimental are present in the cross-section design. This is usually a one-shot survey type. The cross-section is used when manipulation of an independent variable is not possible thus longitudinal attitude cannot be altered by manipulation stimuli and therefore any pre and posttest is not possible. The research is done through a random sample of participants who give their characteristics, perhaps demographic data, and then some form of attitudinal rating or feedback on specific issues or questions of importance to the study (Frankfort-Nachmias & Nachmias, 2008, p. 117). In this design the aim is to find out an attitude or opinion on some subject; the sampling is a probabilistic sampling of a population. Typically, Likert scaling or opinion choice is given to the participants and a variety of statistical analyses are used to determine the attitudinal frequency.

Pre-Experimental Research Method

In this design a single group or a single event is observed at one point in time and recorded and there is no control group, no comparison amongst groups and there is no multivariate statistics as a substitute for control. This design is the weakest of the quantitative design methods and there is no assurances that it is valid and is used primarily for the pre-testing of a hypothesis or draft exploratory study such as in a one-shot case study where a single group or event at a single point in time is observed.

Correlation and Regression

Correlation

The correlation is one of the most common and most useful statistics. A correlation is a single number that describes the degree of relationship between two variables. In other words, correlational relationship among various variables which is very common in finance and economics is defined as a group of techniques that measure the strength of association between a dependent variable and some independent variables. The purpose of correlation analysis is to examine the strength and direction of the relationships between variables. The simplest technique to present a correlational relationship between two variables is by scatter diagram; that is, a chart that portrays the relationship between two variables. Typically, the values of independent variable, or predictor, are portrayed on the horizontal axis (X axis) and the dependent variable, the predicted or estimated value, along the vertical axis (Y axis).

In order for the researcher to see how strong the relationship between two variables is, the coefficient of correlation (Pearson r) can be calculated; the value of r changes between -1 and +1 that indicates perfect negative or positive relationship between the variables and 0 indicates no relationship. Testing the significance of the correlation coefficient can also be used to determine the relation between two variables through the null hypothesis and the alternate hypothesis in a two-tailed test. Correlation does not mean causation; however, in the absence of correlation between two variables, there is no causal relationship.

Assumptions

The assumption for correlational relationship is that the relationship between the two variables can be explained as a linear relationship. However, my design includes regression analysis which is developing an equation that expresses the relationship between the independent and dependent variable. Therefore, if we establish a straight line between two variables X and Y, we have a regression equation which is:

Y' = □ + □X, where Y'' is the predicted value of Y variable for a selected X value; □ is equal to Y intercept, or value of Y when X is 0; □ is the slope of the line; it measures change in Y' for each unit change in X (regression coefficient); X is any value of independent variable that is selected. In this case, we have the following assumptions for linear regression: (a) For each value of X, there is a group of Y values that are normally distributed, (b) the means of this normal distribution of Y values all lie on the line of regression, the standard deviations of these normal distributions are equal, (c)- the Y values are statistically independent, and(d) the expected value of the error term (Y-Y') is zero.

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

Where:

N	=	number of pairs of scores
Σxy	=	sum of the products of paired scores
Σx	=	sum of x scores
Σy	=	sum of y scores
Σx^2	=	sum of squared x scores
Σy^2	=	sum of squared y scores

We use the symbol **r** to stand for the correlation. Through the magic of mathematics, it turns out that r will always be between -1.0 and +1.0. If the correlation is negative, we have a negative relationship; if it's positive, the relationship is positive.

Multiple Regression

Multiple Regression is a statistical technique that allows one to assess the relationship between one dependent variable or criterion and several independent variables or predictors. Both the dependent variable and independent variables should be continuous; however, it is possible to see discrete or dichotomous variables that are dummy variables. Multiple regression is an extension of bivariate regression in which several independent variables instead of one, combined to predict a value on dependent variable. Multiple regression is a flexible techniques and it can be used with experimental, observational, as well as survey research. Multiple regression can determine the strength of the association between set of predictors and criterion. In addition, it can tell the importance of each of the independent variables to the dependent variable.

The Assumptions

The assumptions of multiple regressions are: 1- Outliers can impact the precision of results in multiple regression and must be deal with outliers prior to conducting the research.2- The ratio of cases to predictors; that is, multiple regression can be sensitive to sample size, if sample is too small, the results will not be accurate. In order to be able to accurately test for multiple correlation and regression coefficients, it is essential to have a sample size of at least greater than 104+ number of predictors in multiple regression. 3- The third assumption is multicollinearity; multiple regression is sensitive to multicollinearity which is when at least two of independent variables in the equation are too highly correlated with each other. Multicollinearity makes regression equation unreliable and it can give large standard errors in equation. 4- Assumption 4 is normality of variables; although there is no need for variables to be normally distributed, the prediction equation is enhanced if all of variables are normally distributed.

The Types of Multiple Regression

There are three types of multiple regression: 1- Standard type, simultaneous or direct multiple regression is the most widely used type of multiple regression. In this type all predictors are entered into the equation at the same time; that is, the overlapping variance refers to the overlap that is shared among the predictors.2- The second type of multiple regression is sequential in which the predictors enter the equation according to an order determined by the researcher; overlapping variance is assigned to the predictors in the order of entry into the regression equation. 3- The third type of multiple regression is statistical or stepwise multiple regression in which the order of entry for the independent variables depends on statistical criteria. The software package SPSS decides which predictor to put into the equation at each step based on statistical criteria that the researcher decides on.

The Biases in Multiple Regression

Multiple regression provides an estimate of the effect on Y of arbitrary changes, ΔX. The Multiple regression: 1- Can resolves the problem of omitted variable bias, if an omitted variable can be measured and included; it can handle nonlinear relations; the statistical inferences about causal effects are valid for the population being studied; the statistical inferences can be generalized from the population and setting studied to other populations and settings, where the "setting" refers to the legal, policy, and physical environment and related salient features. There are five threats to the internal validity of regression studies: Omitted variable bias, wrong functional form, errors-in-variables bias, sample selection bias, and simultaneous causality bias

Omitted variable bias arises if an omitted variable both is a determinant of *Y* and (ii) is correlated with at least one included regressor; the suggested Potential solutions to omitted variable bias is that if the variable can be measured, include it as a regressor in multiple regression or use panel data in which each entity (individual) is observed more than once; if the variable cannot be measured, use instrumental variables regression and run a randomized controlled experiment. Wrong functional form arises if the functional form is incorrect, for example, an interaction term is incorrectly omitted; then inferences on causal effects will be biased. In this case, use the "appropriate" nonlinear specifications in *X* (logarithms, interactions, etc.) or in case of discrete dependent variable: need an extension of multiple regression methods

Errors-in-variables bias occurs when there is measurement error in data; for example, data entry errors in administrative data, recollection errors in surveys, ambiguous questions, intentionally false response problems with surveys. In order to prevent variable bias, obtain better data, develop a specific model of the measurement error process; for example, a subsample of the data are cross-checked using administrative records and the discrepancies are analyzed and modeled. Sample selection bias is due to random sampling of the population. In some cases simple random sampling is not possible and selection process influences the availability of data. Moreover, sample selection bias induces correlation between a regressor and the error term. Simultaneous causality bias is a bias caused when X causes Y and Y causes X too. Potential solutions to simultaneous causality bias is through randomized controlled experiment.

Measurement

Measurement is an essential procedure for comparing and forecasting phenomena such as wealth, or welfare, for individuals, institutions, and nations. For example, cost of living index that compares welfare of a nation at two different times or between different countries. Due to the fact that the role of measurement is so important, measuring instruments must be reliable and viable in the sense that measuring instrument must represent the underlying concept behind them. This presented as isomorphism by Nachimas and Nachimas (2008). Furthermore, because of this importance, a researcher should have a system of measurement in mind as soon as research is conceptualized and before the process of operationalization. “Conceptualization is process of formulating and clarifying concepts in research” (Singleton & Straits, 2005, p.77) and measurement is the process of moving from the “abstracts to the concrete” (p.77). In moving from abstract to concrete, we have to identify the concepts, transform them into indicators and variables and then measure them at different level. In other words, designing the study goes along with the design for measuring the variables. Although measurement seems simple in day to day life, it needs some research and running a project when it comes to social phenomena. In many cases measuring instruments already developed by other scientist in the field of study and can be employed by researchers. However, in other cases, the researcher must develop new measuring instruments that could “convert empirical observations into the form suitable for the research problem and the research design” (Nachimas & Nachimas, 2008, P.138).

Measurement is, in fact, assigning numerals, numbers or symbols, to variables “according to a set of rules” (Nachimas & Nachimas, 2008, p.139). A numeral is a symbol of the 1,2,3….and has no specific quantitative meaning in itself; it will be given meaning when it is attached or “mapped” to phenomena, objects, or persons. Then, it becomes number and can be manipulated mathematically by researchers for description, explanation and prediction of phenomenon. Moreover, the researcher has to follow some rules in assigning numerals; the rules basically apply to the way that numbers are arrayed. Rules are important in measurement process because the quality of measures taken is decided by the rules and also, rules tie the measurement process to an empirical basis. In other words, rules in the measurement procedure is the reason for isomorphism and similarity of the structure of numerical system and the structure of the concepts being measured; in this case we can say that the number system are tied to reality and the rule defining the measurement is valid. However, many social science concepts are not directly observable and one indicator cannot describe them; a very often used example is the concept of human capital. In order to measure these type of concepts, we have to use multiple indicators and the index them to ascribe a variable; an indicator is “a single observable measure, such as single questionnaire item in a survey” (p.79).

Level of Measurements

Nachmas and Nachimas (2008) defined the concept of “isomorphism “as a similarity between the structure of numerical system used to measure and the concept being measured. Therefore there should be proximity between the measuring instrument and the empirical properties or indicators measured. Thus, scientists classified the level of measurement according to the properties of instruments: The four distinguished level are: Nominal, ordinal, interval, and ratio. Nominal level is the lowest level of measurement. Numbers or other symbols are used to measure categories of indicators, or cases. In order to assign numerical or symbolic labels to a category, it must be exhaustive and mutually exclusive. That is, they include all cases of that type and no case can be classified to more than one category (numbers in this case are only labels and no mathematical relationship exists between these numbers). Ordinal measurement is the type of measurement when relations exist among indicators, like the symbol of higher or greater. Interval measurement, in addition to nominal and ordinal property, measures the distance between variables. For example, the difference between 20 degree and 10Fahrenheit is equal to the difference between 30 and 40. In this case we can perform basic mathematical operation on them, such as addition and subtraction. Ratio measurement includes all the features of previous levels plus the features of an absolute zero point. This feature makes it possible to multiply and divide scales numbers and forms ratios.

Due to isomorphism which requires the similarity between the measuring instruments and the empirical properties measured, we can have four way of measuring the indicators. In nominal level which is the lowest level of measurement, numbers or other symbols are used to measure categories of indicators. Nominal level of measurement is obtained when a set of objects can be classified into categories that are exhaustive and mutually exclusive. That is, categories should include all cases of that type and no case can be classified to more than one category; numbers in this case are only labels and no mathematical relationship exists between these numbers.

In ordinal measurement, not only variables are classifiable, but also they exhibit some kind of relation; relations are degree of being higher, bigger, more desired, and so on. Then, observations can be arrayed as higher, lower, more or less and form a complete ranking of objects. There are some logical properties in ordinal ranking that become important when a researcher is seeking to construct hypotheses based on ordinal variables. These properties are: reflexivity, asymmetrical and transitive. An example of measurement at the ordinal level is the measurement of attitude in which the possible answers are ranked in ascending or descending orders; in this case the ordinal level of measurement is based on unidimensional transformation. It basically means that no matter how the numbers are manipulated the information obtained does not change and the way that numbers are assigned does not matter so long as consistency is maintained. The numbers assigned to ranked objects are called “rank values” (Nachimas & Nachimas, 2008, p.145). It is usually common that researchers assign rank values on the basis of the object at one extreme is assigned 1 and the rest comes after that. Although some mathematical manipulation is possible in ordinal measurement, but only those mathematical calculations are allowed that do not change the order of properties, for example, “median, range, gamma, and tau-b” (P. 146).

Another more advanced way of measuring categories is interval level of measurement. In the interval level of measurement, not only ranking a set of observations is possible, but also the exact distance between each of the observations is known to the researcher; examples of variables measured by interval level are income, intelligence, and so on. To be more precise at the interval level the distances between observations are isomorphic to the structure of arithmetic used with the associated values. Therefore, the difference between 20 degree and 10 degree of Fahrenheit is equal to the difference between 30 and 40. In this case a researcher can perform basic mathematical operation such as addition and subtraction. In more advanced stage, variables that contain zero point can be measured at the ratio level of measurement; these variables include: weight and length. In order to have a ratio level of measurement, variables should possess four properties: Equivalency, ranked, known distance of any interval, zero point. Due to these qualities this level of measurement allows highly powerful statistical manipulations, and therefore is mainly used in physical science. As a rule, the higher level measurement can be transformed into lower level measurement, but the lower level cannot be transformed into higher level of measurement.

As a rule, the higher level measurement can be transformed into lower level measurement, but the lower level cannot be transformed into higher level of measurement. Variables that can be measured at a ratio level can also be measured at interval, ordinal and nominal level. For example, Income and price are variables that can be measured in ratio form, therefore can be transformed into other levels; income of 20,000 is twice as much as 10,000; 20,000 is higher than 10,000; the difference between 20,000 and 10,000 is the same as the difference between 20,000 and 30,000. But you cannot claim the same for category of nationality or gender. In social studies there are some variables that cannot be measured by single indicators. These are called composite variables and “have several empirical properties” (Nachimas & Nachimas, 2008, p. 414) and researchers need to construct multi indicator instruments to measure them. These variables are most common in social and behavioral sciences and cover the topics such as attitude or self- esteem. The most often used method for measuring composite values are through construction of indexes and scales. These instruments are composite measures that are constructed by combining two or more indicators.

Scale and Indexing

In social studies there are some variables that have “several empirical properties” and cannot be measured by single indicators (Nachimas & Nachimas, 2008, p. 414). Thus, researchers need to construct more sophisticated, multi indicator, instruments to measure them. These variables that are also called composite variables and are most common in social and behavioral sciences and cover the topics such as attitude or self- esteem. The most often used method for measuring composite values are through construction of indexes and scaling methods. These instruments are composite measures that are constructed by combining two or more indicators and therefore are suitable for measuring composite variables. Indexes and scales are preferred instruments for measuring instruments. This is because: (1) Researchers deal with a single score instead of several variables and this makes it easier for them to analyze and manipulate the data mathematically; (2) Researchers can do more precise statistical analysis; (3) using indexes increase the reliability of measurement itself, in a composite measure that uses several indicators errors will cancel each other out. Thus, researchers use multiple items scales and indexes to increase reliability and precision of their measurement due to theoretical, methodological and practical reasons (Nachimas & Nachimas, 2008, P.415).

Since indexes and scales combine several indicators into a composite measure, they might be a better overall representation of the concept and a more reliable measurement tool. This is because in a composite measure that uses several indicators errors will cancel each other out. Also, using multiple scales and indexes will increase the reliability and precision of measurement. In addition, using indexes and scales enables researchers to have access to a single score instead of dealing with several variables and this makes it easy for them to administer statistical analyses (Nachimas & Nachimas, 2008). Whereas scales and indexes both are one composite variable produce by several indicators, their role and purpose in social science research is different; scaling play a more important role in the research of social science. This is because administrating tests of validity and reliability of measurement is possible when researchers use scaling measurement. This property of scaling is due to the fact that most scales that are used in social science incorporate the principle of unidimensional in their construction. According to this principle, the items comprising a scale reflect a single dimension and can be placed in a continuum presumed to apply only to one concept.

Scales can have any number of dimensions, they can be either unidimensional or multidimensional; dimension is like a number line and one number line is called unidimensional (Trochim, 2008). The major unidimensional scaling methods are: “Thurstone, or Equal

Apperaring Interval Scaling, Likert or summative scaling, and Guttman or Cumulative Scaling " (P.136). The most common used method of scaling in social research is Likert scale. Likert scale which is usually used for measuring attitudes can be constructed in six steps: Compile a possible large number of items (questions), select randomly a number of participants and get their replies to the questions, calculate the total score for each participant, calculate the discriminative power of the items, select the scale items, and test the reliability of the scale (Nachmias& Nachmias, 2008).

The important point in this scale is that the answers which are usually four, five, or seven must comprise a continuum and values will be given to each answer. These values express the relative weight and direction of the responses. After collecting the responses from participants, total score for each participant will be selected. Then items will be selected for final scale on the basis of discriminative power (DP) for each item; the highest DP items will be selected for final scale. Researchers can have an easy selection of items through running computer statistical program such as "bivariate correlation coefficient (Pearson's r)" and "Cronbach's alpha" (p.425). The Pearson's r indicates how closely linked each item is with other items or the entire scale. Researchers will choose items the correlate strongly with one another. Cronbach's alpha which estimates the average of all possible split-half reliability coefficient indicates how the scale "hang together"; a high alpha more than .70 is an acceptable level.

In scales like any other measurement study is intended for a population that could be individuals, groups, or institutions. Population and the sample that is selected for scaling should be able to connect with the object and the underlying concept of research. Also, statements and items should be selected in a way to be meaningful to the respondents (Gorden, as cited in Nachmias & Nachmias, 2008). There are some popular scales such as Likert that are commonly used in the attitude studies for different population. However, a researcher must be careful in applying the scale and test examples of one population to another; scales must be tested for reliability and validity in relation to the population they are designed for.

Indexes can be constructed by combining two or more items or indicators. Researchers have to consider four major issues in the construction of indexes: The purpose for which the index is constructed, the sources of data for index, the base of comparison, and methods for aggregating and weighting data (Nachimas & Nachimas, 2008). The most common use of index in social sciences is “Attitude Indexes”. To construct an attitude indexes, researchers prepare a set of questions, selected a priori Numerical values (e.g., 0 to 4 or 1 to 5) are assigned arbitrary to the items question responses. The value assigned by the respondents is added to obtain total scores. The scores are then interpreted as the indicators of the respondent’s’ attitude. There are some ambiguities about this indexing which result in doubt whether this is a reliable index?

Likert scaling is a method designed to measure attitude. To construct a Likert scale researchers usually follow six steps: Compiling a list of possible scale items, administering these items to a random sample of respondents, computing a total score for each respondent, determining the discriminative power(DP) for each items, selecting the scale items, and testing the scale reliability (Nachimas & Nachimas , 2008). In the first step, a researcher compiles a series of items, as much as it is possible, that express a wide range of attitude, from extremely positive to extremely negative. The respondent is then requested to check one of the five offered fixed alternative expressions, such as “strongly agree”, agree” neither agree nor disagree and strongly disagree which comprise a continuum of responses. Then values are assigned to these five – point continuum, either ascending like 0,1,2,3,4, and 5, or descending like5,4,3,2,and 1.

These values express the relative weights and direction of responses and determine the favorableness or favorableness of the items.

In the fourth step the researcher must determine a basis for the selection of items for the final scale. The researcher's objective is to find items that consistently distinguish respondents who are high on the attitude continuum from those who are low. Researchers can employ two methods to do this: First is the internal consistency method which is performed by correlating each item with the total score and retaining those with the highest correlation. Second is the item analysis in which the researcher subjects each item to a measurement of its ability to differentiate the highs from the lows. In order to calculate the DP, researchers will calculate total score for each respondent and rank the scores, usually from lowest to highest. Then, they calculate the upper and lower quartile; DP for each item is the difference between the weighted means of the score for the upper and lower quartile. Then the researcher will select the items with higher DP for the final test.

Validity and Reliability

Validity refers to the congruence between an operation definition and the concept it is supposed to measure. The question of validity is: Whether the operational definition truly reflects validity. However, measuring variables in social science is difficult, because in many cases, the measurement is through the index of multi-indicators and researchers cannot be sure that measurement has truly defined the concepts. Reliability, on the other hand, deals with consistency and stability. Whether reapplication of research operation in similar conditions gives the same results. A reliable measurement might not be valid, but a very unreliable one is definitely not valid; you can be measuring something very reliably something other than what you intended to measure.

Validity in an experiment comes from the fact that whether the experiment can claim with confidence that what has caused the effects was truly the independent variables rather some intervention or other variables. There are two types of threat to validity of an experiment: Internal and external. If the effect is caused by independent variables not some *extraneous* variable, it has internal validity, otherwise the effects are “confounded” (Singleton& Straits, 2005, P.187). The threats to internal validity are; history, length of experiment, maturation, and any psychological or physical changes within subject that affects the tests. Moreover, changes in what is being measured, instrumentation, unwanted changes in characteristics of measuring instrument, or in the measurement procedure; statistical regression, the tendency for extreme scores to move closer to the mean are other causes of the treat to internal validity (P.187). Singleton& Straits signified five basic requirement of true experiment in order to have *internal validity*. These are: Random assignment, meaning that each subject has an equal chance of being in the observation to avoid that the results of the experiment have occurred by chance, manipulation of the independent variable; measurement of the dependent variable; at least one comparison or control group; the constancy of conditions across groups (P.187). External validity threat is caused by applying results in new setting, people, or samples. Creswell (2014) stated the types of threat to external validity are: “interaction of selection and treatment, interaction of setting and treatment, and Interaction of history and treatment” (P.176). Trochim (2001) explain that in order to have a causal relationship we must have: “Temporal precedence, co-variation of cause and effect, and no plausible alternative explanations” (p.173). Even if we have the first two in a causal relationship it is not certain that our intervention has caused the effect. These alternative explanations are threat to internal validity and the researcher has to rule them out to get a valid causal relationship.

There are two types of errors in the operational measurement: systematic error and random error. Systematic measurement error results from factors that systematically influence either the process of measurement or the concept being measured. Two types of measurement errors in this category are: reactive measurement effect and social desirability effect. Measurement random errors are the result of transitory upswing or downswing in the health and mood of subjects; temporary variation in the administration or coding; momentarily investigator fatigue. Because they are unsystematic, random errors tend to cancel each other out with repeated measurement thus they do not bias the measure in a particular direction.

Internal Validity

Internal validity refers to whether your premise or proposition, and then your inference or conclusion has logical truth. In order to establish validity in quantitative research, the assignment of the persons within the groups should be random and equivalent participants in the studied groups must be matched for those factors that are "known to the investigators prior to the research operation" (Frankfort-Nachmias & Nachmias, 2008, p. 98). However, in case of quasi-experimental design randomization may not take place, but to establish internal validity, the researcher has to control for history, maturation, and mortality biases (Frankfort-Nachmias & Nachmias). In other words, "Internal validity is the approximate truth about inferences regarding causal relationship" (Trochim, p.172). It simply means that only the intervention has caused the outcome in the research and nothing else. Internal validity has meaning in causal relationship if r researcher can prove that the only variable that caused the outcome was the independent variable and nothing else.

In other words, internal validity is the approximate truth about inferences regarding causal relationship" (Trochim, 2001, p.172). It simply means that only the intervention has caused the outcome in the research and nothing else. It is of primary concern for studies that assess the effect of social programs or interventions. The key question of internal validity is whether observed changes can be attributed to the treatment in research and not to other possible causes; sometimes described as alternative explanations for outcome, intrinsic factors or strenuous variables. Trochim (2001) explain that in order to have a causal relationship we must have: “Temporal precedence, co-variation of cause and effect, and no plausible alternative explanations” (p.173). Even if we have the first two in a causal relationship it is not certain that our intervention has caused the effect. These alternative explanations are threat to internal validity and the researcher has to rule them out to get a valid causal relationship.

Internal validity has meaning in causal relationship if researcher can prove that the only variable that caused the outcome was the independent variable and nothing else. Trochim (2001) explain that in order to have a causal relationship we must have: “Temporal precedence, co-variation of cause and effect, and no plausible alternative explanations” (p.173). Even if we have the first two in a causal relationship it is not certain that our intervention has caused the effect. These alternative explanations are threat to internal validity and the researcher has to rule them out to get a valid causal relationship.

Internal threat to validity is all that threatens the researcher’s ability to draw correct inferences from the data about population. On the other hand, external threat to validity is when researcher draws incorrect inferences from the sample data to other past and future situations. Threat to external validity impedes the generalization of the research findings. Cronbach (2000) has framed the issues related to external validity in two ways: generalizing from the finding of research to the cause and effect that they represent; generalizing from the categorizing represented in a study to different categories (cited in Cook). In regression analysis, the internal validity is whether the statistical inferences about relationship are valid for the population being studied. There are a few threats to this validity such as omitted variable bias, errors in variable bias, sample selection bias. External validity is when the statistical inferences can be generalized from the population and setting studied to other populations and settings.

There are different types of threat to internal validity such as: History, maturation, testing, instrumental, mortality, and regression threat. History threat happens when some external events occurred during the research operation and affects the outcome in the research. This threat can be ruled out by adding control groups in the operation; control group is a group comparable with the program group, but is not engaged in treatment. Maturation threat is due to events that typically inspires in participants' life. Testing threat is when testing is done twice and after pretest participants are more aware of the problem. Instrumental threat is when change of instrument happens during the operation and affects the outcome. Mortality threat is when due to circumstances participants dropping out of the study, therefore the combination of sample in the group changes and thereby changes the outcome. Regression threat is a "statistical phenomenon" that explains when the pretest scores are lower than average, it will increase and moves towards the mean even if no treatment occurs (Trochim, 2001, p. 177).

Threats to internal validity can be reduced through randomization and including control group in the research design. Randomization will eliminate the effects of intrinsic factors in the experiment. Researchers also control the influence of intrinsic factors by using a control group from which they withhold the experimental stimulus. The two groups, experiment and control, should be identical either through matching or randomizing and also the two groups should be exposed to the same event during the experiment. By using a control group, the researcher controls most of the intrinsic factors that could threaten the validity of experiment. History does not become a rival hypothesis because the control and experiment groups have been exposed to the same events during the experiment. Similarly, maturation is neutralized because the two groups undergo the same change. Including control group does not necessarily avoid the mortality problem because one group might drop more cases than other. Researchers can also avoid the influence of instrument change by using control group if instruments are unreliable, this reflected in both groups. The reactive effect if present is also reflected in both groups. A good design is a design that minimizes the threat to internal validity. The researcher has the flexibility to minimize the threat to internal validity by changing the research design.

Validity and Measurement

Validity of a measurement is defined by whether the measuring instrument produces the same observation as the researcher's intention. The same definition applies to scale type of measurement and validity of scales can be verified through content, empirical and construct validity; scale should cover all the attributes of the concept as much as possible. The reliability of the scale can be tested by the three common ways of estimating reliability: Test-retest method, the parallel forms technique, and the split half method. In test-retest method, the researcher administers the measuring instrument to the same groups or persons at two different times and then computes the correlation between the two sets of observations. In parallel forms technique a researcher develops two parallel version of a measuring instrument and administers both forms to the same group and then correlate the two sets of scores to obtain estimate of reliability. Split-half method researchers in here divide the measuring instrument and the resulting scores will be correlated. The same principle applies to the scale type of measuring.

In any measurement, there is the possibility that error occur in the measurement. There are several common sources of measurement errors such as: (1) The influence of associated attribute other than what the researcher intended to measure; (2) change in the conditions such as health or mood of the researcher; (3) differences in the research setting; (4) differences in the administration of the measuring instrument, processing, and coding; (5) different interpretation of results by different people (Nachimas & Nachimas, 2008). These errors are the reasons that some measures lack validity and reliability. Validity is the similarity between what is measured and what is intended to measure by a researcher. The problem of validity is more pronounced in social sciences because of the indirect measurement; therefore, the researcher must always provide supporting evidence that a measuring instrument does in fact measure the variable that it appears to be measuring. There are three types of validity: Content validity, empirical Validity, and construct validity.

Content validity implies that all the attributes of the concept is covered by the measurement instrument and nothing relevant to the phenomenon under investigation is left out. There are two common types of content validity: Face validity and Sampling validity. In face validity a researcher gives a subjective evaluation of the appropriateness of measuring instrument for measuring the concept that the researcher wishes to measure. After a researcher develop a new questionnaire, the researcher is required to assess the face validity of the questions; to prove that questions are capable of catching all attributes of the variable under investigation.in order to achieve this, a researcher should compare the questions in the questionnaire of the study with questions in the other questionnaire in the subject and also the researcher must seek consultation from specialists in the field. Researchers must be reasonably sure that the questionnaire has face validity; i.e., questions “capture all the elements of the phenomenon” (Nachimas & Nachimas, 2008, P. 150). In sampling validity, a researcher must be sure that a given population is adequately sampled by a measuring instrument. In other words, there must be a large enough number of items so that the researcher could select a good amount of item samples for the research instrument.

Empirical validity is the relationship between a measuring instrument and the outcome obtained from using the measurement. It is presumed by scientist that a valid measuring instrument results in similarity between the outcome of research and the relationship that exists among variables in the real world. In predictive validity, on the other hand, researchers assess the results they expect to obtain from research against some other “external criterion”, or simply criterion. Researchers, then compare these two outcomes and computes the degree of correlation, or correlation coefficient, between the results of a given instrument and an external criterion. For example, for a research on the intelligence test to be valid, the researcher first by obtains test scores of a group of college students, and then adopts the grade point averages that these students achieved in their first year of college as “criterion” and then computes correlation coefficient between these two sets of measurement. The results is correlation coefficient which is named the “predictive validity coefficient” (P.151).

Another type of testing validity is through construct validity; this type of testing is based on the idea that the measuring instrument should relate with the theoretical framework of the study. This test will assure the researcher that the measuring instrument fits well with the concept and the theoretical assumptions of the study. In order for researchers to provide the construct validity of a measuring instrument, they must show that these relationships can be recognized and measured by research instruments.

Conclusion Validity

A threat to conclusion validity is when the researcher reaches an incorrect conclusion about a relationship in the observations. Here, there are two type of errors about relationships: (1) Conclude that there is no relationship when in fact there is; (2) there is a relationship when in fact there is not. In most cases the researcher miss the relationship that exists; perhaps, because, it is hard to find relationships in the data at all when data is not as big or frequent; “we tend to have more problems finding the needle in the haystack than seeing things that aren't there” (Trochim, 2006).

In the first type error, no relationship is found while there is a relationship: This happens because of “the tiny needle and too much hay” or “signal-to-noise ratio problem” (Trochim, 2006). The "signal" is the relationship and the "noise" consists of all of the factors that make it hard to see the relationship. The number of noises that create is a threat to conclusion validity are: low reliability of measures due to many factors such as poor question wording, bad instrument design or layout, illegibility of field notes, and so on and so forth; noise that is caused by random irrelevancies in the setting such as for example, in a classroom context, the traffic outside the room, disturbances in the hallway, and countless other irrelevant events can distract the researcher or the participants. Moreover, the types of participants in the study can make it harder to see relationships, because of random heterogeneity of respondents. Moreover, another important factor in conclusion validity is the statistical power and that is related to the amount of information that researchers collect and the amount of risk they are willing to take in making a decision about whether a relationship exists. Low statistical power causes a threat to conclusion validity.

On the other hand, there are occasions that the researcher finds a relationship when there is not one; this is due to the fact that researchers can play with the data long enough to get the results that support or corroborate their hypotheses; in other words, they are "fishing" for a specific result and they get it through analyzing the data repeatedly under slightly differing conditions or assumptions. Statistically, the researchers set an arbitrary value known as the level of significance to decide whether their result is credible or could be considered a "fluke; therefore, by changing the level of significance the researcher gets different results. More importantly there are errors due to the variety of assumptions that the researcher makes about the nature of the data, the procedures that the researcher use to conduct the analysis, and the match between these two. In quantitative research this is referred to as the violated assumptions of statistical tests (Trochime, 2006); for example, the assumptions that the data are distributed normally, the population from which they are drawn would be distributed according to a "normal" or "bell-shaped" curve, and so on. If that assumption is not true for your data and you use that statistical test, you are likely to get an incorrect estimate of the true relationship. And, it's not always possible to predict what type of error you might make -- seeing a relationship that isn't there or missing one that is.

Bibliography

APA *Publication Manual of the American Psychological Association* sixth edition. American Psychological Association, 750 First st. NE Washington, D.C. 20002

Aristotle. (350 BC). *Metaphysics*. Retrieved from http://classics.mit.edu//Aristotle/metaphysics.html

Benninga, S. (2000). *Financial modeling.* Cambrige, MA: The MIT Press

.

Bernanke, Ben S., and Kenneth N. Kuttner (2003). "What explains the Stock Market's reaction to Federal Reserve Policy?" Mimeo, Board of Governors and Federal Reserve Bank of New York. Bomfim, Antulio N. (2001). "

Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of Business, 45*, 444-455.

Bohacek, R., &, Mendizabal, H., R. (2007). Credit Markets and the propagation of monetary policy shocks. Journal of Money, Credit and Banking, 39(6), 1429-1455 URL: http://www.jstor.org/stable/4494304 Accessed: 20/ 01/2013 13:37

Bory, M. M. (2011). Testing multi-factor asset pricing models in the Visegrad Countries. *Journal of Economics and Finance*, 61(2), 118-139.

Brealy, R. A., Myers, S. C., & Marcus, A. J. (2004). *Fundamentals of corporate finance* New York: McGraw-Hill Irwin.

Buchner, A., Faul, F., & Erdfelder, E. (n.d.) *G*Power*. Retrieved January 29, 2010, from http://www.gpower.hhu.de

Burns, W. E. (2011). *Science in World History*. Pearson Educated inc. 1 Lake St. N.J. 07458.

Campbell, D.T. and Stanley, J.C. (1963, 1966). *Experimental and Quasi-Experimental Designs for Research*. Chicago, Ill.: Rand McNally.

Canuto, O. 2011a. “How Complementary Are Monetary Policy and Prudential Regulation?” *Economic Premise* 60, World Bank,October.

Chan, L., Y. Hamao, & Lakonishok, J. (1991) “Fundamentals and Stock Returns in Japan,” *Journal of Finance*, vol. 46, p. 1739-1764

Charmoz, K. (2006). *Constructing Grounded theory: A practical guide through qualitative analysis*. Sage publication, London.

Chen, N.-F., R. Roll, and S.A. Ross (1986). Economic Forces and the Stock Market. Journal of Business 59: 383-403.

Chen, S., S. (2007). Does monetary policy have asymmetric effects on stock returns. Journal of Money, Credit and Banking, 39(2), 667-688. URL: http://www.jstor.org/stable/4494266 .Accessed: 20/01/2013 13:31

Chiarella, C., Dieci, R., He, X. Z., & Li, K. (2013). An evolutionary CAPM under heterogeneous beliefs. *Ann Finance*, *9*, 185–215.doi:10.1007/s10436-012-0215

Cochrane, J. (2001). *Asset pricing*. Princeton, NJ: Princeton University Press.

Cochrane, J. H., and Piazzesi, M. (2002), "The Fed and interest rates: A high frequency identification," *American Economic Review Papers and Proceed ings*, 92, 90-101.

Creswell, J.W. (2013). *Qualitative inquiry and research design choosing among five traditions*. Sage publication, Thousand Oak CA 91320

Creswell, J.W. (2014). *Research Design, Qualitative, Quantitative, and Mixed Methods Approaches***.** Sage Publication, inc. Thousand Oak CA 91320

Glaser, B. G., & Strauss, A. L. (1967) *The Discovery of Grounded Theory*. Hawthorne, NY:

DEMPSEY, M. (2013).The CAPM: A case of elegance is for tailors? *ABACUS, 49,* 82-87. doi:10.1111/j.1467-6281.2012.00389.x

***www.KansasCityFed.org*.**

Denzin, N. K. & Lincoln, Y. S. (2013). *Collecting and Interpreting Qualitative Materials.* Sage Publications inc. 2455 Teller Road Thousand Oaks, California 91320

Dewalt, K. M. & Dewalt, B. R. (2002). *Participant observation: a guide for fieldworker*s. Altamira Press, Rowman & Littlefield Publishers, UK.

Doh, T. & Connolly, M. (2013). Has the effect of monetary policy announcements on asset prices changed? *Federal Reserve Bank of Kansas City.*

www.KansasCityFed.org

Döpke, J., Hartmann, D., & Pierdzioch, C. (2006). Forecasting stock market volatility with macroeconomic variables in real time. *Banking and Financial Studies*, *1,* 1-25.

Ehrmann, Michael, and Marcel Fratzscher (2003). "Monetary Policy Announcements and Money Markets:A TransatlanticP erspective."I nternationalF inance 6, 309-328.

Ehrmann, M., & Fratzscher, M. (2004). Taking stock: Monetary policy transmission to equity markets. *Journal of Money, Credit and Banking, 36*(4) 719-737. URL: http://www.jstor.org/stable/3839039 .Accessed: 20/01/2013 13:32Your use 123–143.

Fama, E. F. & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*, 427-464.

Fama, E. F. & French, K. R. (1995). Size and book-to-market factors in earnings and returns. *The Journal of Finance, 50*, 131-155.

Fama, E. F. and French, K. R. (FF) (2006). The value premium and the CAPM. *Journal of Finance 61*, 2163–85.

Fama, E. F. and French, K. R (2008). Dissecting anomalies. *Journal of Finance, 63,* 1653–1678.

Febrian, E., & Herwany, A. (2010). The performance of asset pricing models before, during, and after an emerging market financial crisis: Evidence from Indonesia. *The International Journal of Business and Finance Research, 4*(1), 85-97.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). London: Sage Publication, inc. Thousand Oak CA 91320

Frankfort-Nachmias, C., & Nachmias, D. (2008). *Research methods in the social sciences (7th ed.*). New York, NY: Worth.

Gabaix, X. (2011). Disasterization: A simple way to fix the asset pricing properties of macroeconomic models. *American Economic Review, 101(3),* 406–409

Goodfreind, M. (2009). “Central Banking in the credit turmoil: An Assessment of Federal Reserve practice.” *Carnegie Mellon University and NBER.*

Gharghori, P., Lee, R. & Veeraraghavan, M. (2009) “Anomalies and Stock Returns: Australian Evidence,” *Accounting and Finance, 49*(3), 555-576.

Gitman, L. J. (2000). *Principles of managerial finance.* New York: Addison Wesley

Gordon, M. J. (1989). Corporate finance under MM theorems. *Financial Management*, *2*, 19-28.

Graham, J. R. & Harvey, C. R. (2001). The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics,* 60, 187-243.

Graham, J. R. & Harvey, C. R. (2002). How do CFOs make capital budgeting and capital structure decisions? *Journal of Applied Corporate Finance*, 15, 1-28.

Groat, L. & Wang, D. (2013). *Architectural Research Methods*. John Wiley and Sons, inc. Hoboken, NewJersey

Gürkaynak, R., S., Sack, B., P., & Swanson, E., T.(2007). Market-based measures of monetary policy expectations. *Journal of Business & Economic Statistics, 25*(2), 201-212.

Gwilym, R. A., (2013). The monetary policy implications of behavioral asset bubbles. *Southern Economic Journal, 80*(1), 252-270. doi:10.4284/0038-4038-2011.242

Harjoat S., Bhamra, H., S., Kuehn, L., A., & Strebulaev, I., A. (2010).The aggregate dynamics of capital structure and macroeconomic risk. *The Review of Financial Studies, 23*(12). doi:10.1093/rfs/hhq075

Haugen, R. A. (2001). *Modern investment theory*. Upper Saddle River, NJ: Prentice Hall.

Humpe, A., & Macmillan, P. (2007). Can macroeconomic variables explain long term stock market movements? A comparison of the US and Japan. CENTRE FOR *DYNAMIC MACROECONOMIC ANALYSIS WORKING PAPER SERIES*, 1-15.

Janesick, V. J. (2011). *Stretching exercises for qualitative researchers (3rd ed.*). Thousand Oaks, CA. Sage Publications.

Kauppi, H., and P. Saikkonen (2008). "Predicting U.S. Recessions with Dynamic Binary Response Models. "*Review of Economics and Statistics*, *90*, 777–791.

Kim, S.-W. & Lee, B.S.(2008). "Stock Returns, Asymmetric Volatility, Risk Aversion, and Business Cycle: Some New Evidence." *Economic Inquiry*, 46), 131–148.

Kolb, R. W. (1997). *Futures, options, and swaps*. Oxford UK: Blackwell publishers

Kolozsi, P., P. (2012). Monetary policy, interest groups, financial crisis. *JEl*, 35-55.

Krueger, J. T., and Kuttner, K. N. (1996), "The Fed Funds Futures Rate as a Pre dictor of Federal Reserve Policy," Journal of Futures Markets, 16, 865-879.

Kumar, R. (2014). *Research Methodology*. Sage Publication, inc. Thousand Oak CA 91320

Lakonishok, J., Shleifer, A. and Vishny, R. W. (1994) Contrarian investment, extrapolation and risk. *Journal of Finance*, *49*, 1541–78.

Levy, H. (2010). The CAPM is alive and well: A review and synthesis. *European Financial Management*, *16*(1) 43–71. doi:10.1111/j.1468-036X.2009.00530.x

Levy, H. (2012). Two paradigms and Nobel Prizes in economics: A contradiction or coexistence? *European Financial Management*, *18*(2), 163–182.
doi: 10.1111/j.1468-036X.2011.00617.x

Liew, J. and Vassalou, M. (2000). Can book-to-market, size and momentum be risk factors that predict economic growth? *Journal of Financial Economics*, *27*, 221–45.

Lind, D. A., Marchal, W., G. & Mason, R., D. (2002). *Statistical Techniques in Business & Economics.* McGraw-Hill, Irwin

Linter, j. (1965). The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, *47*, 13-37.

Magni, C. A. (2010). CAPM and capital budgeting: Present/future, equilibrium/disequilibrium, decision/valuation. *Journal of Financial Economics, 8*(1 & 2), 7-32.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, *7*, 77-91.

Markowitz, H. M. (1999). The early history of portfolio theory: 1600-1960. *Financial Analysts Journal*, *55*, 5-16.

Maxwell, J. (2013). *Qualitative Research Design,* Sage publication, Thousand Oak CA 91320

Maysami, R., C., Howe, L., C, & Hamzah, M., A. (2004). Relationship between macroeconomic variables and stock market indices: Cointegration evidence from stock exchange of Singapore's All-S Sector Indices. *Jurnal Pengurusan 24,* 47-77

Miles, M. B., Huberman, A. M., &Seldana, J. (2014) *Qualitative Data Analysis: A method Source Book.* Sage Publications inc. 2455 Teller Road Thousand Oaks, California 91320

Modigliani, F. & Miller, M. H. (1958). The cost of capital, corporation finance, and the theory of investment. *The American Economic Review*, 49, 261-297.

Modigliani, F. & Miller, M. H. (1961). Dividend policy, growth, and valuation of shares. *Journal of Business*, 34, 411-433.

Modigliani, F. & Miller, M. H. (1963). Corporate income taxes and the cost of capital: A correction. *American Economic Review*, 53, 433-43.

Mossin, Jan. (1966). Equilibrium in a capital asset market. *Econometrica*, *34*, 768-784.

Newman, I., Ridenour, C., S., Newman, C., & DeMarco, G., M., P. (2003). A typology of research purposes and its relationship to mixed methods. *Journal of Mixed* Nyberg, H. (2010a) "Dynamic Probit Models and Financial Variables in Recession Forecasting." *Journal of Forecasting*, 29 (2010a), 215–230.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Nyberg, H. (2010b). "QR-GARCH-M Model for Risk-Return Tradeoff in U.S. Stock Returns and Business Cycles." HECER Discussion Paper 294, University of Helsinki

Nyberg, H. (2011)."Forecasting the Direction of the U.S. Stock Market with Dynamic Binary Probit Models."*International Journal of Forecasting*, 27, 561–578.

Nyberg, H. (2012). Risk-Return Tradeoff in U.S. Stock Returns over the Business Cycle *JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS 47*(1) 137–158. doi:10.1017/S0022109011000615

Methods Research, *1*(3), 167-187.

Onwuegbuzie, A.J., Leech, N.L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative Research Methodologies. *Social Research Methodology*, 18 (5) PP.375-387

Pascale, C. M. (2011).*Cartographies of knowledge: Exploring Qualitative Epistemologies*. Sage Publications Inc. 2455 Teller Road Thousand Oaks, California 91320

Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods (3rd Ed).* Thousand Oaks, CA, Sage Publications.

Privitera, G.j. 2015). Statistics for the Behavioral Sciences. Sage Publication, inc. Thousand Oak CA 91320

Rao, S., V., D., N., & Ramachandran, V., S. (2010). Stock Returns and Macro Variables: The Indian evidence. *International Journal of Business Insights &* Reilly, F. K. & Brown, K. C. (1997). *Investment analysis and portfolio management.* New York: The Dryden Press.

Ross, S. A., Westerfield, R. W., & Jordan, B. D. (2001). *Essentials of corporate finance*. New York: McGraw-Hill Irwin.

Rudebusch, G. D. & Williams J. C (2009) . "Forecasting Recessions: The Puzzle of the enduring power of the Yield Curve." *Journal of Business and Economic Statistics*, 27), 492–503.

Rosenberg, A. (2001). *Philosophy of Science*. 29 West 35th Street, New York, NY 10001.

Russell, B. (1961). *History of Western Philosophy*. George Allen & Unwin LTD.

Samuel B. Green, S. B. & Salkind, N. J. (n.d). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*, Sixth Edition. Prentice Hall, Pearson Education inc.

Sangmi, M., U., D., & Hassan, M. M. (2013). Macroeconomic variables on stock market interactions: The Indian experience. *Journal of Business and Management, 11*(3), 15-28.

Schonlau, M., Fricker, R.D., & Elliott, M.N. (2002). *Conducting research surveys via e-mail and the web*. Washington, DC: RAND.

Sehgal, S., & Balakrishnan, A. (2013). Robustness of Fama-French three factor model: Further evidence for Indian stock market. *SAGE Publications 17*(2) 119–127. doi:10.1177/0972262912483526

Silverman, D. (2000). *Doing Qualitative Research – A Practical Handbook*. London: SAGE.

Singleton, R.A.,& Straits, B.C., (2005). *Approaches to Social Research*. Oxford University Press, 198 Madison Ave. NewYork, NewYork, 10016

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*, 425-442.

Spyroua, S., & Kassimatis, K. (2009). Time-variation in the value premium and the CAPM: Evidence from European markets. *Applied Financial* Economics, *19*, 1899–1914

Stake, R.E. (1995). *The art of case study research.* Sage Publications inc. 2455 Teller Road Thousand Oaks, California 91320

Strauss, A. L., & Corbin, J. (1998) *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 2nd edition. Thousand Oaks, CA: SAGE

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson.

Tashakkori,A., & Teddlie, C., (2010). *Methods of mixed methods in Social and Behavioral Reseach.* Sage Publications inc. Thousand Oaks, 2455 Teller Road, California 91320

Trochim, W. M. (2006). *The Research Methods Knowledge Base, (2nd Ed).* Retrieved from http://www.socialresearchmethods.net/kb/ (version current as of 06/03/11)

Trochim, W.M.K. (2006). *Factorial designs*. Retrieved from http://www.socialresearchmethods.net/kb/expfact.php

Trochim, W.M.K. (2006). *Time in research*. Retrieved from http://www.socialresearchmethods.net/kb/timedim.php

Trochim, W.M.K. (2006). *Covariance designs*. Retrieved from http://www.socialresearchmethods.net/kb/expcov.php

Yalçın,A., & Ersahin, N.(2011). Does the conditional CAPM work? Evidence from the Istanbul stock exchange. *Emerging Markets Finance & Trade, 47*(4), 28–48. doi:10.2753/REE1540-496X470402

Yin, R.K. (2009). *Case study research: design and methods (4th ed.)*: Sage Inc. Thousand Oaks, 2455 Teller Road, California 91320