Research Design + Statistics Tests

Aligning research design and statistical analyses

Kamil Mysiak Follow

Aug 19 · 14 min read ★

Image by PIRO4D from Pixabay

From the first day I sat in my undergraduate “Research Methods” course staring at SPSS output, I knew I found my calling. I can still recall my first research paper. Watching my the completed surveys come in, diligently cleaning the data and crossing my fingers in the hopes of significant results. Despite my results coming back not significant I knew I found my passion.

In spite of engrossing myself in the topic, I found it particularly difficult aligning research design to statistical analysis. As the terminology began to roll-in (ie. t-tests, ANOVA, effect size, IV, MANOVA, ANCOVA, regression, R^2, etc.) I grew more confused and frustrated. The sheer amount of terminology used in designing experiments, analyzing, and interpreting results can be a daunting reality.

Although to fully describe the in-depth nature of each research design along with the appropriate statistical models we would require a lengthy

textbook, I hope to provide a condensed summary. Below is a high-level summary of research design principals and appropriate statistical models used to analyze the data.

Variables

For an experiment to be considered a true experiment there needs to be some manipulation. In other words, we need to impose onto our study participants and measure the effect of that imposition. This is where the independent and dependent variables come into play.

- **Independent Variable (IV):** The variable which is being actively manipulated by the researcher. Depending on study design, we can have multiple IVs and each IV can have multiple "levels" which participants are subjected to. For example, does the type of teaching style affect students' test scores? This study's only IV would be teaching style and it would have 3 unique levels (ie. authoritarian, authoritative, permissive).
- **Dependent Variable (DV):** The variable which is hypothesized to be affected by the IV. This is also the variable which the research measures to determine what effect, if any, the IV had. In our example, students' test scores would be the only DV.

Measurement Scales

The cornerstone of any study is the collection and analysis of data. Depending on the study design, hypotheses, and data collection methods we can have four types of quantitative data. Each scale of measurement has distinct properties which determines the use of specific statistical analyses.

Categorical Variables

- **Nominal/Categorical:** Simple labeled or classified data without order or measurable meaning. Can take on string and/or numerical format. (ie. Location: Chicago, San Fransisco, Atlanta OR Gender: 1, 0).
- **Ordinal**: Labeled measurement data which can be ranked in order but we cannot numerically distinguish the intervals between each measurement. Classified data which can be sorted and ranked (ie. What is your highest level of completed education? (A) Less than HS, (B) HS, (C) Some college, (D) Bachelor's degree, (E) Graduate degree)

Continuous or Numerical Variables

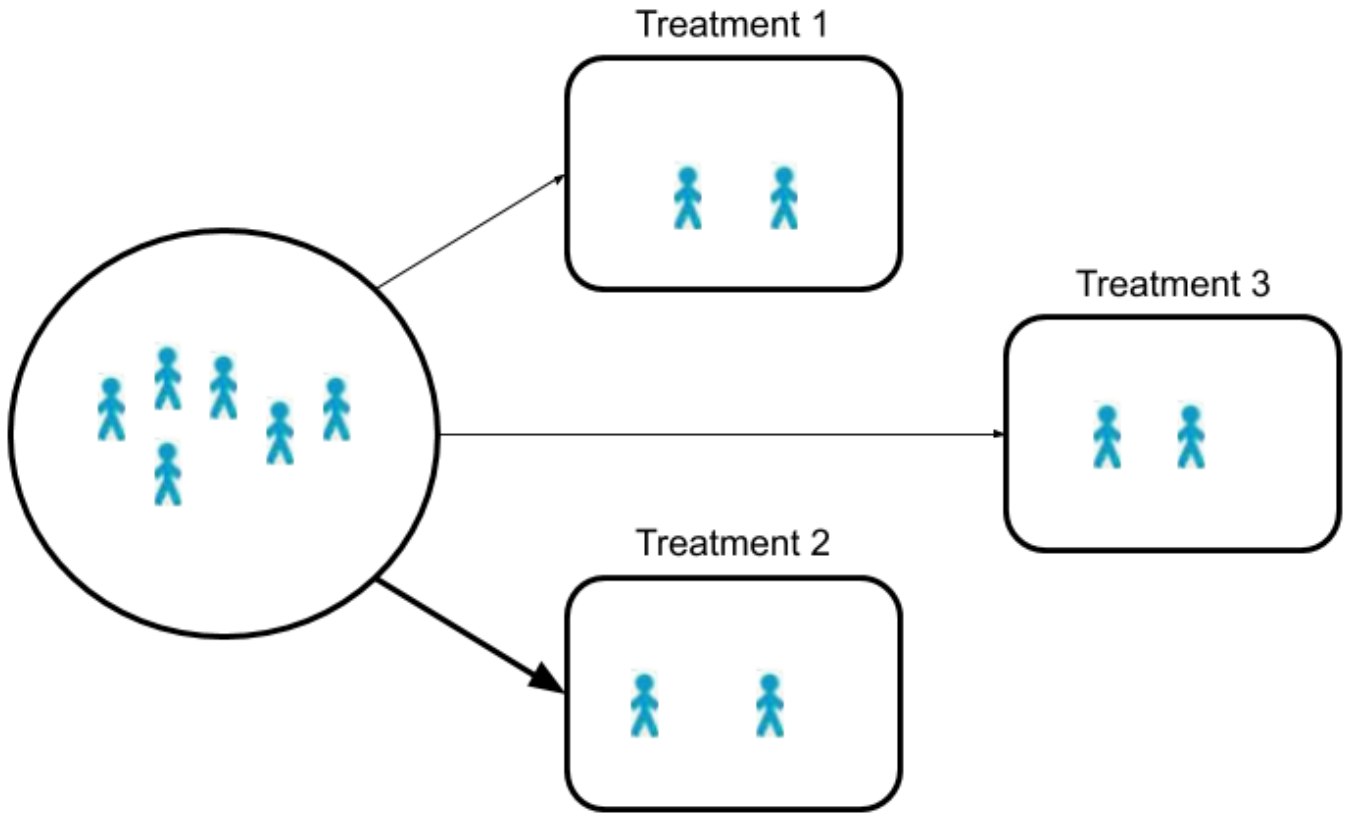
- **Interval**: Labeled and ranked numeric data where the distance between categories can be compared but the concept of zero does not exist and

distance between each value is equal (ie. Likert rating scales: strongly agree=5, agree=4, neutral=3, disagree=2, strongly disagree=1)

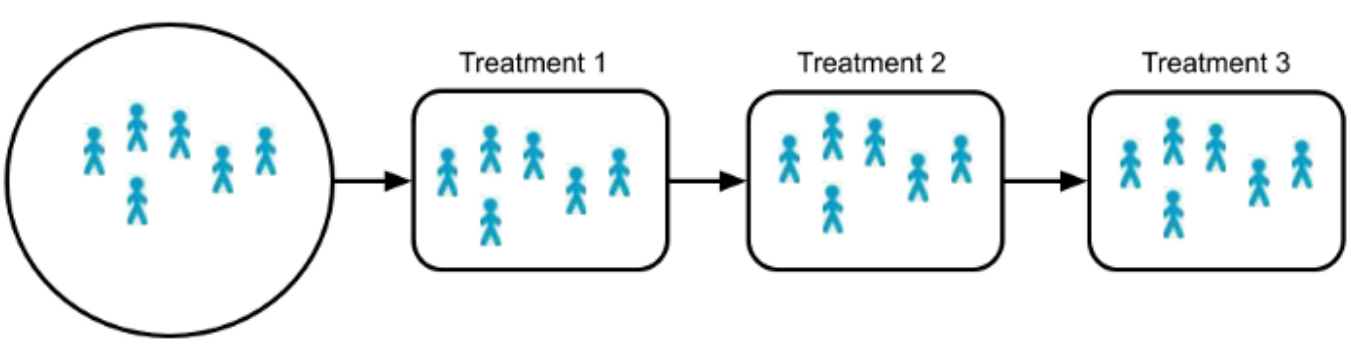
- **Ratio**: Labeled and ranked continuous data where the distance between categories can be compared, the concept of zero exists, can contain negative values (ie. What is your age? 10, 22, 23, 50, etc.)

Between-Groups vs. Within-Groups

- Between Groups design places study participants into different groups or conditions (ie. control or one treatment group). The participant can only belong to one group for the entire experiment. Each participant subjected to only one level on the IV.

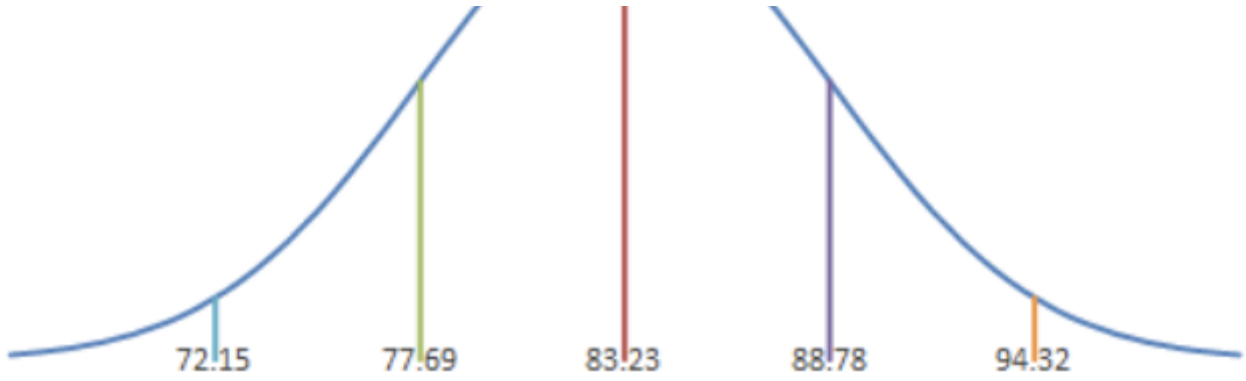


- Within Groups design places the same study participant into all the levels on the IV. Often referred to as “Repeated Measures” design.



Parametric vs. Non-Parametric

The main parameter in choosing the appropriate statistical test for your experiment is the shape of your data’s distribution. A normal or Gaussian distribution follows a bell curve we see below.



Bell Curve

- Parametric tests assume a normal distribution among your data and contain interval or ratio data.
- Non-Parametric tests do not assume anything about the distribution of your data and contain ordinal or nominal data.
- Histograms and QQ-plots are a great way to visualize the distribution of your data to determine normality

Parametric Tests

- **Pearson's r Correlation:** Measures the linear relationship between two variables. Both the IV and DV have to be ratio (ie. continuous).
- **Regression:** A regression has two main purposes. First, it is used to assess the strength of the relationship between the DV and one or multiple IVs. It can infer causation by helping us determine how much each IVs affected the DV. Prediction of values is its second purpose. There are many types of regressions depending on the type of your IVs and DVs and parametric quality of your data. Linear regression is used to predict a continuous variable (ie. job performance) with only one IV. Logistic regression is used to a predict binary outcome (ie. yes/no) with multiple IVs. Multiple regression is used to predict a continuous outcome (ie. job performance) with multiple IVs (ie. IQ test scores, personality test scores, etc.).
- **One-Sample t-test:** Compares the mean of your sample against a known or hypothesized value, typically the known population mean, to determine whether or not they are significantly different.
- **Independent Samples t-test:** *Between-groups design.* Comparing the mean DV scores of two independent groups of study participants each was exposed to only one group/level in the IV and the IV only has 2 groups/levels. DV is continuous. For example, what is the effect of teaching styles on students' test scores. Two randomly selected groups of students are selected. The first group is taught by the authoritarian teacher and the second group is taught by the authoritative teacher. We compare the mean DV (ie. test scores) scores for each class to determine if they are significantly different. In order to test for homogeneity of variance, we can use Levene's test.

- **Paired t-test:** *Within-groups design.* Comparing mean DV scores of randomly selected participants who were exposed to both groups/levels on the IV, the IV only has two groups/levels and there is only one IV. Typically, a pre/post test design is used. Sometimes referred to as "Dependent Samples t-test". For example, the same students have been taught by two teacher types (authoritarian, authoritative) during the course of a semester. The students' average test scores for the material taught by the first teacher is compared against the average test scores from the second teacher. This design often utilizes a pre/post-treatment measurement of the DV design. Besides histograms and QQ-plots, the Shapiro Wilk test can be used to determine normality.
- **One Way ANOVA**: *Between-groups design.* Comparing the mean DV scores of randomly selected participants each signed to only one group/level in the IV. The IV has more than 2 groups/levels and we only have one DV. For example, three randomly selected groups of students are selected. The first group is taught by the authoritarian teacher, the second group is taught by the authoritative teacher and the third group is taught by the permissive teacher. The one-way ANOVA compares the students' scores between all three teaching styles. Often referred to as "Between Groups ANOVA". The DV is a interval or ratio (ie. continuous). A One Way ANOVA is an omnibus test, as a $p<0.05$ rejects the NULL hypothesis to determine that there is a significant difference between the groups. However, to determine specifically where the difference is and its magnitude we perform ad-hoc tests (ie. Tukey, Dunnett)
- **Repeated Measures ANOVA:** *Within-group design.* Comparing the mean DV scores of participants who were all exposed to more than 2 groups/levels of the one IV and we only have one IV. For example, instead of two teachers we now that three teachers which all students are taught by. The repeated measures ANOVA is an omnibus test, as a $p<0.05$ rejects the NULL hypothesis to determine that there is a significant difference between at least two groups. A Bonferroni ad-hoc test will determine exactly where the differences lie. The DV is interval or ratio (ie. continuous) and measured n-times where $n=$number of IV levels of conditions.
- **Two-Way ANOVA:** *Between-group design.* Comparing the mean DV scores of randomly selected participants who were exposed to only one group/level on the IV and we have more than one IV. A three-way ANOVA would have three IVs. The two-way ANOVA attempts to determine the effects of more than one IVs op the DV. This is also a great test to determine if an interaction of the two IVs affects the DV. The results of a two-way ANOVA will define the main effect between each IV and the DV separately followed by the interaction effect on the DV. For example, we have randomly selected groups of students. Our first IV is teacher types (ie. authoritarian, authoritative, permissive) and our second IV is whether or not the students were taught during the day or at night. This is a $3x2=6$ factorial design because we have three teachers

types and two teaching time conditions. Each student is placed into only one of the six conditions and the mean test scores are compared to determine if there is a significant differences between the 6 groups.

- **One-Way MANOVA:** *Between-groups design.* Comparing multiple mean DV scores of randomly selected participants who were exposed to only one group/level on the IV. We have one IV but we have two or more DVs. For example, what is the effect of teaching styles (authoritarian, authoritative, permissive) on student scores and absenteeism? Each student is assigned to one of the three groups/level in the IV and measured for both their test scores and absenteeism. Teaching style is our one IV with three levels but we have two DVs, students' scores and absenteeism.

- **Two-Way MANOVA:** *Between-groups design*. Comparing multiple mean DV scores of randomly selected participants who were exposed to only one group/level on the IV and we have multiple DVs. For example, what is the effect of teaching styles (authoritarian, authoritative, permissive) and time of study (day or night) on student scores and absenteeism? This is a 3x2=6 design because we have three groups/levels on the first IV and 2 groups/level in the second IV. Each student is exposed to only one of the six groups/levels between the teaching styles IV and the time of study IV and measured for both their test scores and absenteeism. Teaching style is the IV, time of study is the second IV and we have two DVs, students' scores and absenteeism. This test will also determine the effect of the interaction between all the IVs and both DVs.

- **One-Way Repeated Measures MANOVA:** *Within-groups design.* Comparing multiple mean DV scores of participants who were exposed to more than 2 groups/levels of the one IV and we have multiple DVs. Each student is exposed to all three groups/level in the teaching styles IV and measured for both their test scores and absenteeism. Teaching style is the IV and we have two DVs, students' scores and absenteeism.

- **Two-Way Repeated Measures MANOVA:** *Within-groups design.* Each student is exposed to all 6 groups/levels (three groups/levels in the teaching styles IV and two groups/levels in the time of study IV) and measured for both their test scores and absenteeism. Teaching style is the first IV and time of study is our second IV and we have two DVs, students' scores and absenteeism.

- **ANCOVA**: Used to control the impact of one or more variables on the DV. For example, what is the effect of teaching style on student test scores while controlling for age? In other words, we know that test scores are affected by dozens if not hundreds of variables but we cannot test them all. Therefore, we form hypotheses about a few specific variables such as teaching style but if we want to remove any influence that age might have on test scores, we use an ANCOVA. In order to remove the influence of age, we needed to have recorded the age of each study participant.

Non-Parametric Tests

- **Spearman's Rank Correlation:** A measure of the monotonic relationship between two variables which can be ordinal or ratio. A monotonic relationship says the variables tend to move in the same or opposite direction but not necessarily at the same rate.
- **Mann-Whitney Test:** *Between-groups design* and non-parametric version of the independent samples t-test. The DV is either interval or ratio (ie. continuous).
- **Wilcoxon Signed-Rank Test**: *Within-groups design* and non-parametric version of the dependent samples t-test or sometimes called the "Paired t-test". The DV is either interval or ratio (ie. continuous).
- **Kruskal-Wallis Test:** *Between-groups design* and non-parametric version of the one-way ANOVA. The DV is either ordinal or ratio (ie. continuous). The Kruskal-Wallis test is an omnibus test, the Mann-Whitney U test will help to identify where the significant differences exist.
- **Friedman Test:** *Within-groups design* and non-parametric version of the repeated measures ANOVA. To determine where the differences exist we have to run the Wilcoxon signed-rank test on all the unique combinations of IV levels.
- **Chi-Squared Test:** *Between-groups design.* Using a frequency table of observations between two categorical variables is used to determine the association between the variables. For example, we want to determine if there is an association between gender of data scientists and whether they use python or R. We would produce a table (ie. crosstabs) of the frequency of males and females who use python or R. If the "Pearson Chi-Square" p-value is less than 0.05 we can reject the NULL hypothesis and conclude that there is an association between gender and use of python and R. Used to evaluate the association not causation between categorical variables.

Quantitative vs Qualitative Research

From the most macro-perspective, we have two distinct research categories; quantitative and qualitative. In this article, we are going to focus on quantitative research but it is important to understand the differences between the two.

- Quantitative research deals with numerical data which is collected via assessments, analyzed using statistical methods for comparisons of experimental groups and inferences. ***Confirm/Test using numbers***
- Qualitative research follows an exploratory approach and hopes to explore ideas, theories, and hypotheses. Its data takes the shape of observations, one-on-one interviews, focus groups, open-ended survey

items, historical records, and case-studies (ie. mainly words not numbers). Data analysis attempts to summarize, categorize, and interpret themes instead of causal inferences. ***Understand using words***

Quantitative Research Designs

In general, there are 4 types of quantitative research design:

- Descriptive
- Comparison
- Quasi-Experimental
- Experimental

Descriptive Research Design

- Descriptive or exploratory (ie. data mining) in nature, the descriptive design aims to describe a phenomenon in terms of frequencies, trends, categories/comparisons, and demographics. It hopes to provide glimpses into *what, when, where, and how* but provides very weak evidence of causation (ie. why).
- The researcher can control the sampling strategy of study participants
- Does not contain variable manipulation (ie. independent & dependent variables)

Methods

- Descriptive research relies on asking participants questions pertaining to the research topic.
- Surveys or questionnaires (Likert, frequency, pair comparisons, star ratings, behaviorally anchored rating)
- Observations in the collection of numerical data (ie. how long, how many, what date, age, gender, weight, etc.)

Analyses

- Descriptive statistics such as counts, percentages, mean/mode/median, standard deviations
- Use of charts and visualizations to communicate the insights

Sample Research Questions:

- How has the employee engagement changed over the past 2 years?
- What are the main reasons for voluntary turnover?
- Did employees prefer training X to training Y?

- What are the average salary differences between high and low performers?
- Which recruiting sources bring quality candidates and most hires?
- How prevalent is disengagement among our service workforce?

. . .

Comparison Research Design

- Sometimes referred to as “Correlational” research.
- Aims to define the statistical relationship between two or more constructs. (ie. how does a change in one variable relate to change in other variables?)
- The relationships among the variables are historic in nature as they might change in the future.
- Much like descriptive design, it too provides trends and patterns.
- Although stronger than descriptive design in providing causal evidence, overall still weak.
- Absolutely no direct variable manipulation (ie. dependent and independent variables).

Methods:

- Surveys and questionnaires (Likert, frequency, pair comparisons, star ratings, behaviorally anchored rating)
- Observations in the collection of numerical data (ie. how long, how many, what date, age, gender, weight, etc.)

Analyses:

- Correlations calculate the relationship between the variables.
- Correlations: positive, negative, zero, curvilinear, multiple correlation, partial correlation
- Pearson r Correlation: the most widely used correlation among linearly related constructs which are both continuous.
- Spearman’s Rank & Kendall Tau Rank Correlations: similar to Pearson’s correlation but more robust to non-linear relationships between variables

Sample Research Questions:

- What is the relationship between job performance and cognitive ability?
- Are younger employees more or less likely to quit?

- Do higher scores on training evaluation scores relate to better job performance?

. . .

Quasi-Experimental Design

- Often called “natural experiments” as they are utilized in natural settings.
- They typically involved some sort of treatment (IV) which is hypothesized to influence the DV.
- Independent variable/s are present in study design but very often have not been manipulated by the researcher.
- The effects of the independent variable/s on the dependent variable/s are measured by the researcher.
- Research and control groups of participants are present but have formed naturally and, therefore, random assignment is not present.
- Possible differences between research and control groups introduce unknown confounding variables which might moderate or mediate the results.

Methods:

- Non-Equivalent Groups Design: between-groups design (ie. participants are only assigned to one research group, experiment, or control) where participants are not randomly assigned to the groups.
- Pretest-Posttest Design: The dependent variable is measured pre-treatment to obtain a baseline and then measured again post-treatment to determine the effect the independent variable had on the dependent variable. The amount of time (ie. confounding variable) between pre and post-tests has enormous effects on our degree of confidence that the independent variable affected the dependent variable (ie. **history**: events occurred between pre/post-test; **maturation**: natural growth/learning of research subjects).

Analyses:

- Which statistical tests one uses depends on the design of the study (number of IVs, levels in the IVs, number of DVs, between-groups or within-groups, etc).
- t-tests, ANOVAs, regressions, correlation

Sample Research Questions:

- Did the introduction of the Affordable Care Act affect employee absenteeism?

- Are turnover rates higher for males or females?

. . .

True Experiment

- Some would argue we can never truly determine causality but true experiment designs come the closest to establishing causality.
- The main differentiator of true experiments is the random assignment of study participants into research groups. Random assignment is the most thorough way of establishing equal participant groups. This helps us to impose control over possible confounding variables which might influence the DV.
- Depending on study design, a true experiment can have multiple IVs all of which can have multiple levels.
- Depending on the study design, a true experiment can have multiple DVs.
- A control group of participants is used which does not receive the actual treatment.
- Depending on student design, confederates, or researchers posing to be participants are used.

Methods:

- Pre-Test Post-Test Control Group Design: All study participants are given a pretest to establish a baseline score on the DV. The experiment imposes the IV/manipulation onto the participants. Finally, a post-test is given to determine if the IV had any effect on the DV.
- Post-Test Only Control Group Design: Similar to the design above but omitting the pre-test. The extent of the effect of the IV on the DV is determined by comparing the DV score between the treatment group(s) with the control group.
- Solomon Four Group Design: A combination of the designs listed above. Group 1 of the participants are given a pretest/intervention/posttest. Group 2 is given a pretest and posttest but not the intervention. Group 3 receives the intervention and posttest but not the pretest. Finally, group 4 only receives the posttest.

Analyses:

- Which statistical tests one uses depends on the design of the study (number of IVs, levels in the IVs, number of DVs, between-groups or within-groups, etc).
- t-tests, ANOVAs, regression, correlations, etc.

Sample Research Questions:

- How do various levels of salary increases affect employee satisfaction?
- Does drug X decrease level of depression?

Summary

We can certainly extend this article into infinity as the fields of statistics and research design are quite extensive. I do hope this was a helpful summary of statistical analyses and research design.

Research Methods Statistical Analysis Parametric Anova Quantitative Analysis