Research Skills

Module 9

Qualitative research

Registered Nutritionist (NSA)
Registered Teacher (SA TRB)

Quantitative Data Analysis

This issue of ***Research Bites*** looks at quantitative data analysis.

Type of variables

An item of data that can be observed or measured is called a *variable*. There are two main types of variables.

Numerical variables can be:

- *Discrete variables* - values that are separate and distinct e.g. number of GP visits; or
- *Continuous variables* - when all values are possible e.g. blood pressure, weight.

Categorical variables represent membership of a particular category. They can be:

- *Ordinal variables* - several categories where order is relevant e.g. physical activity measured as minimal, moderate or vigorous;
- *Nominal variables* - no natural order e.g. area of residence; or
- *Dichotomous variables* - only two responses e.g. yes/no.

Sometimes variables are converted for analysis e.g. age (numerical) to age groups (categorical ordinal).

Descriptive statistics

These provide basic summaries of individual observations or measures in a sample. The statistical technique used depends on the type of variable.

In summarising numerical data, some common measures are:

- *Mean* - sum of all individual counts or measures divided by the number of individuals;
- *Mode* - most frequently occurring count or measure across a group of individuals;
- *Median* - middle observation in a sample of individuals;
- *Range* - difference between the maximum and minimum observations; and
- *Standard deviation* - a measure of how much the individual data tend to deviate from the mean.

To describe the relationship between numerical variables, a common test used is the *correlation coefficient, r*. Correlations range between –1 and +1 and describe the nature and degree of association between two variables.

It is important to remember that correlations are not concerned with causality. An additional factor may underlie both variables.

In summarising categorical data, *counts* are used. They can also be expressed as *proportions* or *percentages* by dividing the count by the total number of individuals. While categorical variables may be coded using numbers, it is important not to summarise them as numerical data e.g. do not average coded numbers in Likert scales as below:

Very dissatisfied				Very satisfied
1	2	3	4	5

The relationship between categorical variables is usually presented in a contingency table and tested using *chi-square test*.

Statistical tests

There are many statistical tests available. Several authors have published flowcharts for selecting statistical tests (Peat, 2001, pp. 197-199). Two of the most common tests are the *t-test* for numerical data and the *chi-square test* for categorical data.

Inferential statistics

In inferential statistics you are trying to reach conclusions about a population based on a sample of individuals from the population.

In hypothesis testing, statistical methods are used to determine the probability of obtaining the observed effect by chance. The *p-value* of your chosen statistical test is compared to the *level of significance* (usually set at 0.05 or 0.01). For example, a *t-test* with a *p-value* of 0.03 (and *level of significance* of 0.05) indicates that the results are not due to chance and are *statistically significant.*

Confidence intervals can also be used. Based on a sample, they are estimations of a range of values (confidence interval) within which the population parameter is likely to lie. With a *level of significance* set at 95%, you can say that based on your sample, you are 95% confident that the population value lies within your confidence interval.

Source: Quantitative Data Analysis. Research Bites (Aug 2003). University of New South Wales. Available from: http://www.phcris.org.au/phcred/research_bites/research_bites_7.pdf

Inclusion and Exclusion criteria

Inclusion criteria = attributes of subjects that are essential for their selection to participate.

Inclusion criteria function remove the influence of specific confounding variables.

Exclusion criteria = responses of subjects that require their removal as subjects.

eg., failure to adhere to pre-test requirements, infection, evidence of altered training/fitness, etc.

Source: Robergs. PEP507: Research Methods. 2010. Available from: http://www.unm.edu/~rrobergs/604Lect2.pdf

Biased and Unbiased Sampling

Sample = selected subset of a population

As it is typically impractical, if not impossible, to research an entire population, we need to sample from the population

What is an unbiased sample?

One where every member of the population has an equal chance of being included in the sample.

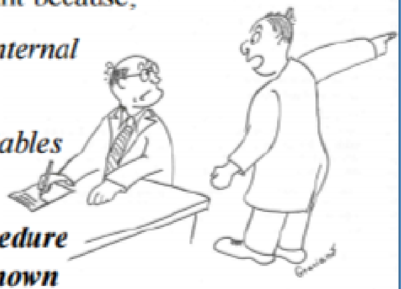
Source: Robergs. PEP507: Research Methods. 2010. Available from: http://www.unm.edu/~rrobergs/604Lect2.pdf

What type of sample should I use?

Simple Random Sample = *when every member of the population has an equal chance of being included in the sample.*

Random sampling is important because;

1. *Helps control threats to internal and external validity*
2. *Can control for many variables simultaneously*
3. ***It is the only control procedure that can control for unknown factors***



"Everythings been so completely randomized out there that we seem to have lost our research assistants among the subjects!"

Stratified Random Sampling = Attempts to decrease sampling errors that exist even if using simple random sampling.

When a population is first divided into strata based on a different variable (eg. Gender), and then random sampling occurs from each strata.

- the same relative representation of each strata should occur

- more than one additional stratification variable can be used (eg. age, gender, ethnicity, wealth, geographical location, political bias, hours of television/day, etc.)

Multi-Stage Sampling = *really a multiple level stratified random sample. (eg. Stratify all counties in US based on socio-economic issues, randomly select households from this list, and then randomly select household members. Used a lot in survey research)*

Sample of Convenience = *when, through convenience, sampling occurs from only a subset of the intended population.*

Volunteerism (*ad hoc sampling*) = *when sampling is based to a large extent on individuals volunteering to participate in the study.*

(due to ethical reasons mandated by human subjects review committees, this is hard to avoid)

Systematic Sampling = *When every n^{th} person is selected.*

Free Random Assignment = *using random number tables or computer generated random numbers*

Matched Random Assignment = *for smaller sample/groups sizes, subjects can be matched on certain characteristics, and then matched subjects can be randomly assigned*

Balanced Assignment = *ensuring that all group sizes, or sequences of trial orders, are equal*

Cluster Sampling = *when groups (clusters) of individuals are drawn rather than separate individuals (eg. all students of randomly chosen APS 3rd grades; pregnant women from pre-natal classes)*

Purposive Sampling = *intentionally selecting specific individuals due to their traits.*

Snowball Sampling = *when subject recruitment is aided by the first participant.*

Source: Robergs. PEP507: Research Methods. 2010. Available from: http://www.unm.edu/~rrobergs/604Lect2.pdf

Hypothesis testing

A hypothesis test is a statistical test that is used to determine whether there is enough evidence in a sample of data to infer that a certain condition is true for the entire population.

A hypothesis test examines two opposing hypotheses about a population: the null hypothesis (Ho) and the alternative hypothesis (Ha). The null hypothesis (Ho)is the statement being tested. Usually the null hypothesis (Ho) is a statement of "no effect" or "no difference". The alternative hypothesis is the statement you want to be able to conclude is true.

Based on the sample data, the test determines whether to reject the null hypothesis.

Source: http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/what-is-a-hypothesis-test/

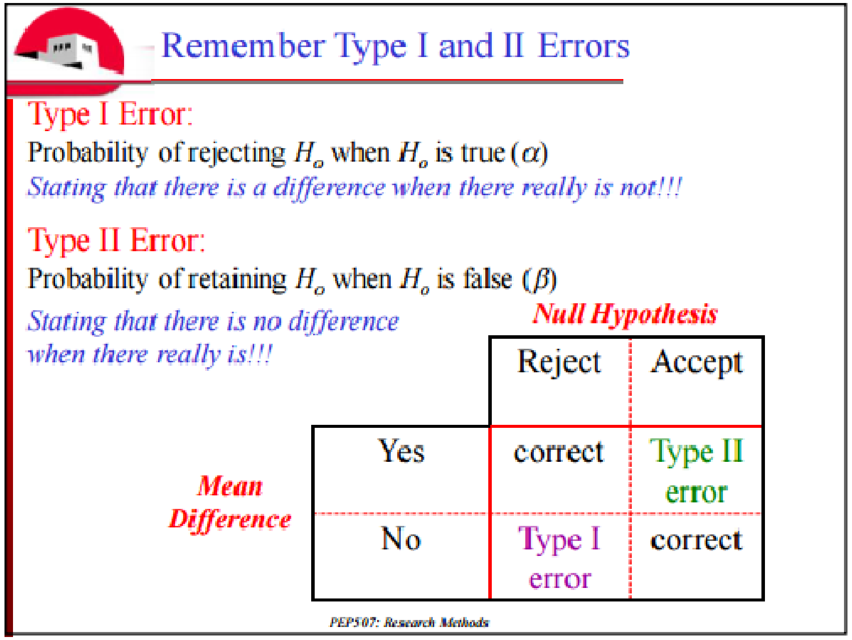
Steps in hypothesis testing

http://stattrek.com/hypothesis-test/how-to-test-hypothesis.aspx?Tutorial=AP

https://learn.bu.edu/bbcswebdav/pid-826908-dt-content-rid-2073693_1/courses/13sprgmetcj702_ol/week04/metcj702_W04S01T05_fivesteps.html

Type I errors, also known as false positives, occur when you see things that are not there.

Type II errors, or false negatives, occur when you don't see things that are there



Source: Robergs. PEP507: Research Methods. 2010. Available from: http://www.unm.edu/~rrobergs/604Lect2.pdf

Ellis, P.D. I always get confused about Type I and II errors. Can you show me something to help me remember the difference? Posted 2010. From Ellis, P/D. The essential guide to effect sizes. Available from: https://effectsizefaq.com/2010/05/31/i-always-get-confused-about-type-i-and-ii-errors-can-you-show-me-something-to-help-me-remember-the-difference/

If you are going to be testing a hypothesis that one group is different to another, e.g. males are more in favour of a particular solution than females, it is a very good idea to calculate the sample size you will need to obtain statistically significant results.

To calculate the sample size your will need. Select your probability level and statistical power level based on how big a sample size you can realistically achieve. The higher the probability and statistical power, the bigger your sample size will need to be. Next enter your anticipated effect size. If based on your literature review, you think that there will be only small differences between your groups, then set your effect size as small (0.2) and see what numbers you get. You might want to rethink what you are looking to detect and pick something for which there are likely to be bigger differences between your groups. For medium size differences set at 0.5), and for big differences set your effect size at large (0.8). Play around changing the values and see what it does to the sample size needed. To have data suitable for statistical analysis your probability level should not be less than 0.1 (90 percent) and your power level should not be less than 0.8 (80 percent).

A-priori Sample Size Calculator for Student t-Tests

This calculator will tell you the minimum required total sample size and per-group sample size for a one-tailed or two-tailed t-test study, given the probability level, the anticipated effect size, and the desired statistical power level.

Please enter the necessary parameter values, and then click 'Calculate'.

Anticipated effect size (Cohen's *d*): 0.5

Desired statistical power level: 0.8

Probability level: 0.05

Calculate!

Available from: http://www.danielsoper.com/statcalc/calculator.aspx?id=47

Statistical Power

The power of any test of statistical significance is defined as the probability that it will reject a false null hypothesis. Statistical power is inversely related to beta or the probability of making a Type II error. In short, power = 1 – β.

In plain English, statistical power is the likelihood that a study will detect an effect when there is an effect there to be detected. If statistical power is high, the probability of making a Type II error, or concluding there is no effect when, in fact, there is one, goes down.

Statistical power is affected chiefly by the size of the effect and the size of the sample used to detect it. Bigger effects are easier to detect than smaller effects, while large samples offer greater test sensitivity than small samples.

How do I calculate statistical power?

The power of any test of statistical significance will be affected by four main parameters:

the effect size
the sample size (N)
the alpha significance criterion (α)
statistical power, or the chosen or implied beta (β)

All four parameters are mathematically related. If you know any three of them you can figure out the fourth.

Why is this good to know?

If you knew prior to conducting a study that you had, at best, only a 30% chance of getting a statistically significant result, would you proceed with the study? Or would you like to know in advance the minimum sample size required to have a decent chance of detecting the effect you are studying? These are the sorts of questions that power analysis can answer.

Source: Ellis, P.D. How do I calculate statistical power? Posted 2010. From Ellis, P/D. The essential guide to effect sizes. Available from: https://effectsizefaq.com/2010/05/31/how-do-i-calculate-statistical-power/

Simple sample size and power calculator

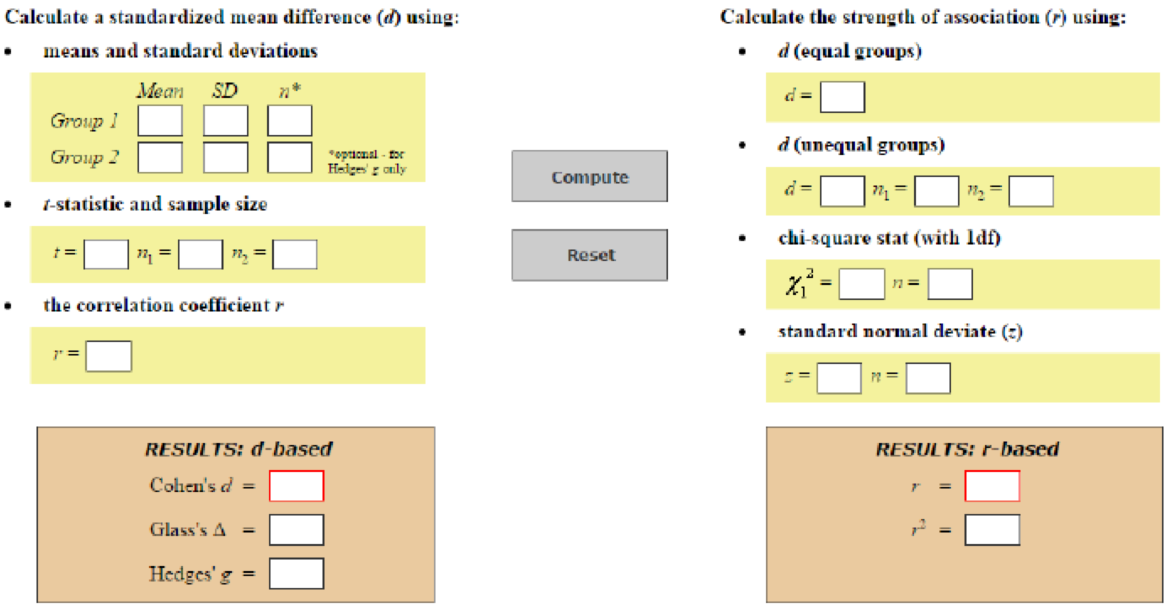
http://biomath.info/power/ or http://www.danielsoper.com/statcalc/calculator.aspx?id=47

What is an effect size?

An effect is the result of something. It is an outcome, a result, a reaction, a change in Y brought about by a change in X.

An effect size refers to the magnitude of the result as it occurs, or would be found, in nature, or in a population. Although effects can be observed in the artificial setting of a laboratory or a sample, effect sizes exist in the real world.

Effect Size Calculators



The standardized mean difference

You subtract the mean of one group from the mean of the other then divide the result by either the standard deviation of the control group (giving you a Glass's Δ) or by the pooled standard deviation of both groups (giving you Cohen's d or Hedges' g depending on the pooling equation used). OR use the calculator at http://www.polyu.edu.hk/mm/effectsizefaqs/calculator/calculator.html or http://www.danielsoper.com/statcalc/calculator.aspx?id=48

Sources: https://effectsizefaq.com/2010/05/31/what-is-an-effect-size/ and https://effectsizefaq.com/category/effect-size-calculators/ and http://www.polyu.edu.hk/mm/effectsizefaqs/calculator/calculator.html

Once you have your data you can calculate the effect size for any differences you observe between groups.

Effect Size (Cohen's d) Calculator for a Student t-Test

This calculator will tell you the (two-tailed) effect size for a Student t-test (i.e., Cohen's d), given the mean and standard deviation for two independent samples of equal size.

Please enter the necessary parameter values, and then click 'Calculate'.

Mean (group 1):	5.5
Mean (group 2):	5.1
Standard deviation (group 1):	0.5
Standard deviation (group 2):	0.5

Calculate!

Source: http://www.danielsoper.com/statcalc/calculator.aspx?id=48

A-priori Sample Size Calculator for Student t-Tests

This calculator will tell you the minimum required total sample size and per-group sample size for a one-tailed or two-tailed t-test study, given the probability level, the anticipated effect size, and the desired statistical power level.

Please enter the necessary parameter values, and then click 'Calculate'.

Anticipated effect size (Cohen's *d*): 0.8

Desired statistical power level: 0.9

Probability level: 0.1

Calculate!

Minimum total sample size (one-tailed hypothesis): 42

Minimum sample size per group (one-tailed hypothesis): 21

Minimum total sample size (two-tailed hypothesis): 56

Minimum sample size per group (two-tailed hypothesis): 28

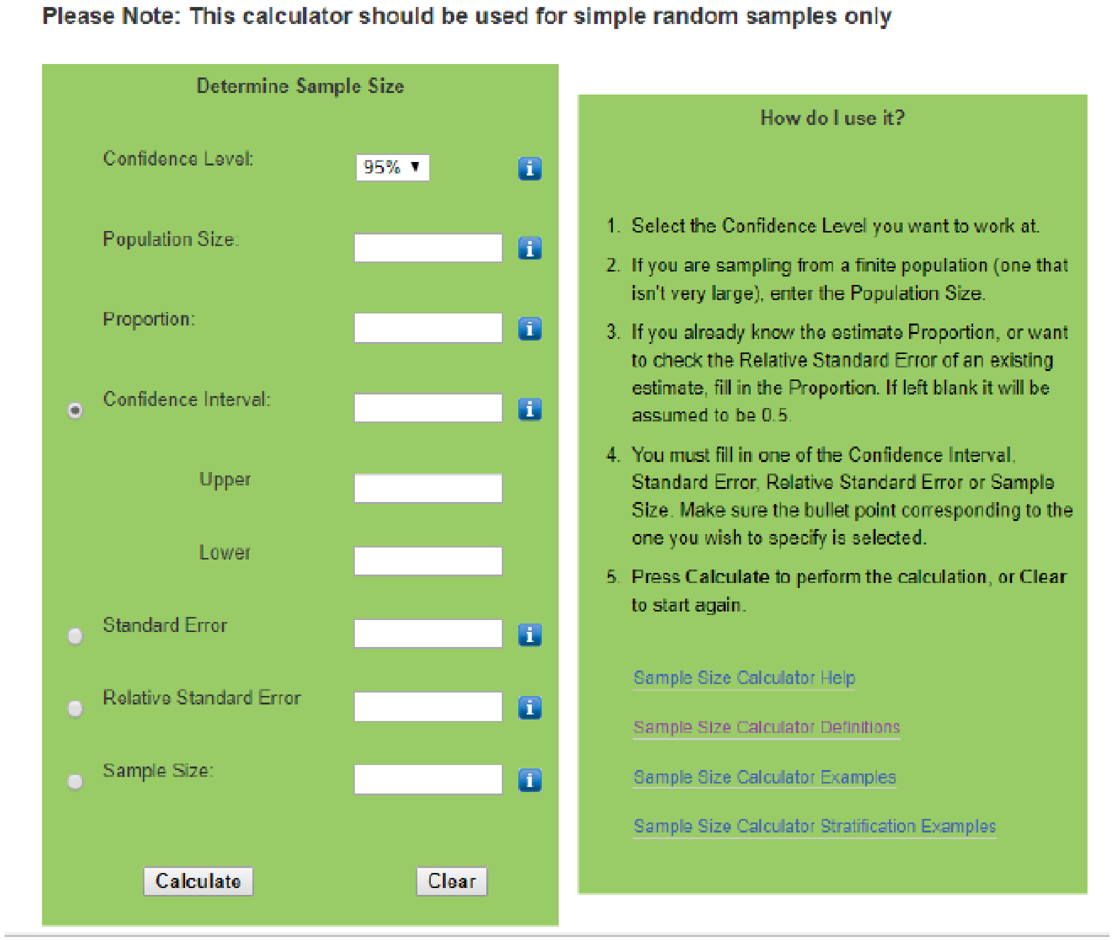
If effect size is unknown, its okay to estimate if it is likely to be small, medium or large, based on previously published literature, or if no literature is available, your observations.

Assumptions

Large effect size of 0.8 (score in one group will exceed 79 percent of scores in the comparison group, same a Z scores of a normal distribution)
Statistical power of 0.9 or 90 percent
Probability level of 0.1 or 90 percent

http://www.danielsoper.com/statcalc/calculator.aspx?id=47

A simple sample size calculator for random samples, provided by the Australian Bureau of Statistics



https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Sample+Size+Calculator

What is the difference between statistical and substantive significance?

Statistical significance reflects the improbability of findings drawn from samples given certain assumptions about the null hypothesis.

Substantive significance is concerned with meaning, as in, what do the findings say about population effects themselves?

Researchers typically estimate population effects by examining representative samples. Although researchers may invest considerable effort in minimizing measurement and sampling error and thereby producing more accurate effect size estimates, ultimately the goal is a better understanding of real world effects. This distinction between real world effects and researchers' sample-based estimates of those effects is critical to understanding the difference between statistical and substantive significance.

The statistical significance of any test result is determined by gauging the probability of getting a result at least this large if there was no underlying effect. The outcome of any test is a conditional probability or p value. If the p value falls below a conventionally accepted threshold (say .05), we might judge the result to be statistically significant.

The substantive significance of a result, in contrast, has nothing to do with the p value and everything to do with the estimated effect size. Only when we know whether we're dealing with a large or trivial sized effect, will we be able to interpret its meaning and so speak to the substantive significance of our results. Note, though, that while the size of an effect size will be correlated with its importance, there will be plenty of occasions when even small effects may be judged important.

Source: https://effectsizefaq.com/category/statistical-significance/

What does a p value represent?

A common misperception is that p = .05 means there is a 5% probability of obtaining the observed result by chance. The correct interpretation is that there is a 5% probability of getting a result this large (or larger) if the effect size equals zero.

A p value is the answer to the question: if the null hypothesis were true, how likely is this result? A low p says “highly unlikely”, making the null improbable and therefore rejectable.

In substantive terms, a p value really tells us very little.

p < 0.001 means a less than 1% probability of getting a result this large (or larger) if the effect size equals zero. **Traditionally a highly statistically significant result.**

p = 0.05 means a 5% probability of getting a result this large (or larger) if the effect size equals zero. **Traditionally a statistically significant result** (p = 0.05 is the usual level used).

p = 0.10 means a 10% probability of getting a result this large (or larger) if the effect size equals zero. **Traditionally the lowest level of statistical significance usually accepted.**

Source: https://effectsizefaq.com/category/p-values/

Why is it a dumb idea to interpret results by looking at p values?

It is a common (but wickedly bad) practice to make judgments about a research result by looking at ***p* values**. Even in top journals you'll sometimes see the following decision rules applied:

- if $p \geq .10$, then the result is interpreted as providing "no support" for a hypothesis
- if $.05 \leq p < .10$, then this is interpreted as providing "marginal support"
- if $p < .05$, then this is interpreted as evidence "supporting" the hypothesis
- if $p < .01$ or .001, this is sometimes interpreted as "strong support" or "strong confirmation" or "strong evidence"

What's wrong with this? Everything! A *p* value is a confounded index (because it reflects both the size of the underlying effect and the size of the sample) so it should not be used to make judgments about effects of interest. The bigger the sample size, the more likely the result will be statistically significant, regardless of other factors. Conversely, as N goes down, p must go up. The smaller the sample, the less likely the result will be statistically significant.

Imagine that we had hypothesized that X has a positive effect on Y. We collect some data, run a test and get the following result:

$N = 70{,}000$, $r = .01$, $p < .01$

Looking at the very low *p* value of this result we might conclude that this test revealed good evidence in support of our hypothesis. But we would be wrong. We have confused statistical with substantive significance.

Look closely at the numbers again. Note how the effect size estimate (*r*) is tiny, virtually zero. We have in all likelihood detected nothing of significance, just a little fluff on the proverbial lens.

So how is it that the *p* value is so low in this case? Because this is an overpowered test. The sample size (*N*) is off the scale. **It's the effect size that you need to be looking at.**

Smart researchers understand that p values should never be used to inform judgments about real world effects.

Sources: https://effectsizefaq.com/2010/05/30/why-is-it-a-dumb-idea-to-interpret-results-by-looking-at-p-values/ and https://effectsizefaq.com/category/p-values/ and https://effectsizefaq.com/2010/05/30/why-do-you-say-a-p-value-is-a-confounded-index/