**HOW TO GENERATE SIGNIFICANT RESULTS**

**by Simon Moss**

|  |
| --- |
| **Introduction** |

In quantitative studies, research candidates often need to conduct statistical tests that generate p values, such as ANOVAs or multiple regression. If none of the p values are significant, the research is unlikely to be published—and certainly not in prestigious journals. So, how you can increase the likelihood of generating significant results? That is, how you can boost the statistical power of your research? Briefly, to increase the likelihood of generating significant results, researchers need to

• maximize the range of their treatments, interventions or independent variables

• increase the reliability of their measures

• control variables that are associated with these measures

• choose the most powerful statistical tests or design

• increase the sample size

This document offers some guidelines on how to achieve these goals.

|  |
| --- |
| **Maximize the range of treatments or independent variables** |

Consider a researcher who wants to ascertain whether participants who consume Echinacea are more likely than participants who do not consume Echinacea to report satisfaction with their life. To increase the likelihood of significant results, researchers should first maximize the range of your treatments, interventions, or independent variables. The following table illustrates how you can achieve this goal.

|  |  |
| --- | --- |
| Practice | Example |
| For experiments or quasi-experiments in which you compare two or more conditions, increase the duration, frequency, or intensity of some treatment relative to the control. | The participants who consume Echinacea should receive high doses—perhaps the highest safe dose |
| For correlational designs in which you explore the association between numerical variables, consider how you can select more people, specimens, or units that are very high or very low on the independent variable | To recruit people who consume very high levels of Echinacea, perhaps distribute surveys to individuals who post on websites about Echinacea and thus must value this supplement |

|  |
| --- |
| **Increase the reliability of measures** |

Next, you should, whenever possible, utilize reliable, accurate measures. Some measures are very reliable: If you administered these measures to the same participants an hour later, you would generate the same results. If you changed the measures very slightly, such as utilized synonyms of certain words, the results would not shift noticeably. The following table illustrates how can choose more reliable measures

|  |  |
| --- | --- |
| Practice | Example |
| When utilizing validated measures, choose the option that generates a high Cronbach’s alpha—a measure of reliability—but is not too long. | A Cronbach’s alpha above 0.8, or even 0.7, is usually adequate. |
| If you write your questions or adapt existing questions, write as precisely as possible | When constructing questions, include some tangible examples. In addition, to write more precisely   * Use a thesaurus to replace ambiguous words, such as get, give, done, take, make, have, and hold, with more precise alternatives. * Always include a noun after the word “this” |

|  |
| --- |
| **Control variables that are associated with these measures** |

Researchers who want to ascertain whether Echinacea improves satisfaction with life may experience a problem: Satisfaction with life depends on so many other characteristics or circumstances. Therefore, in both participants who consume Echinacea and participants who do not consume Echinacea, satisfaction with life is so variable, as the following figure shows. Because of this variability, the effect of Echinacea is hard to discern.

To overcome this problem, researchers should

* identify other variables—characteristics or circumstances—that are related to the measures. For example, satisfaction with life may be related to income, self-esteem, and many other variables.
* measure as many of these variables as possible
* then control these variables statistically.

That is, particular statistical tests, such as ANCOVAs, multiple regression, and logistic regression, can be utilized to control other variables statistically. These tests, in essence, predict what the results would have been had everyone been average on these other variables. In this example, these statistical tests could estimate the effect of Echinacea on satisfaction with life had everyone been average on income, self-esteem, and so forth.

|  |
| --- |
| **Choose the most powerful statistical test** |

In some instances, one statistic test might generate non-significant results whereas another statistical test might generate significant results. Therefore, you need to choose the most powerful statistical test that is appropriate in the circumstances. The following table specifies a couple of more powerful alternatives to common procedures.

|  |  |
| --- | --- |
| Common procedure | More powerful alternative |
| A series of t-tests or ANOVAs, in which each t-test or ANOVA comprises the same independent variables but a different measure or dependent variable | Begin with a MANOVA instead. That is   * a MANOVA can generate a significant effect even if none of the ANOVAs are significant * this pattern is often observed when the measures or dependent variables are moderately correlated |
| A sobel test to assess mediation | Consider the Aroian or Goodmen test. To conduct these tests, simply proceed to quantpsy.org/sobel/sobel.htm |
| A Bonferroni adjustment | More powerful alternatives are available, such as the Holm procedure. For more information, read https://www.sicotests.com/psyarticle.asp?id=251 |

|  |
| --- |
| **Choose the most powerful design** |

**Crossover trials**

When you want to compare several conditions or treatments, one design that could increase power is called a cross-over trial. In essence, suppose you wanted to examine whether a paleo diet is healthier than a soup diet in 100 participants. Arguably, half the participants could observe a paleo diet for one month. The other participants could observe a soup diet for one month. Next, the health of all participants would be measured—perhaps by measuring blood pressure, for example. This design is a typical randomized control trial.

Alternatively, all participants could observe the paleo diet for one month and the soup diet for another month. In particular, half the participants could begin with the paleo diet and half the participants could begin with the soup diet. And, their health could be examined after diet. This design is called a cross-over trial.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Subgroup | March | Next day | April | Next day |
| 1 | Paleo diet | Health measured | Soup diet | Health measured |
| 2 | Soup diet | Health measured | Paleo diet | Health measured |

If you use a cross-over trial, you need about one quarter of the participants to generate significant effects. For example, 100 participants in a typical randomized control trial will generate the same power as 25 participants in a cross-over trial. Typically, researchers would conduct a mixed-model ANOVA, in which time is the repeated measures factor and subgroup—that is, a paleo diet first versus a soup diet first—is the between-subject factor. If your study comprises more than two conditions, you can utilize an analogous design called the multiple cross-over trial.

In short, to increase power, researchers could utilize a cross-over trial. The problem however is that

* cross-over trials are sometimes hard to arrange, because all participants need to complete multiple conditions over time
* one condition might affect the impact of another condition. For example, eating paleo could change the body dramatically and thus diminish the benefits of a soup diet.

Therefore, utilize cross-over designs only if

* the participants you recruit would be willing to complete multiple conditions
* completing one condition is unlikely to affect the impact of another condition.

|  |
| --- |
| **Increase the sample size** |

Obviously, if you increase the number of participants, organizations, specimens, and so forth, you are more likely to generate significant results. Nevertheless, in many instances, financial or logistical constraints limit the sample size.

|  |
| --- |
| **Rationale** |

So, why do these procedures increase the likelihood of significant results. The reason is the p values depends on two considerations: effect size and sample size. You might remember the formula to represent effect size—at least when you need to compare two conditions—is



where

* d is the symbol for effect size in these circumstances
* the numerator symbolizes the difference between the means of each condition
* the demoninator is roughly the average standard deviation in each condition

As this formula implies, to increase the effect size, you could attempt to

* increase the difference between these means—perhaps by magnifying the difference between the treatment and control
* diminish the standard deviation, perhaps by using more reliable measures—measures that are not as contaminate by random error—or by controlling other measures