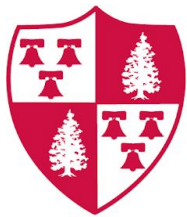


STATISTICS FOR PSYCHOLOGY



MONTCLAIR STATE

UNIVERSITY

This a “What you need to know” handbook of statistical terminology and principles that will be of use when revising introductory statistical material prior to taking an advanced statistics course. The material in this handbook is not a replacement for professional training in statistics, but rather a guideline and quick summary of some statistical concepts.

By Noel D’Souza
Montclair State University



Table of Contents

- Introduction
- Chapter 1 - Measures of Central Tendency
- Chapter 2 - Measurement
- Chapter 3 - Z-scores and Standardized Scores
- Chapter 4 - Null Hypothesis Significance Testing
- Chapter 5 - T-Tests
- Chapter 6 - Linear Regression



Introduction

This handbook is a simple guide to basic statistical concepts. It will include necessary equations and computations as well as a link to helpful videos that can be found on YouTube. Some concepts may seem fairly ‘dry and boring’ to remember, but if you begin studying this subject in a systematic way, you will find that things will start to piece together and make more sense. I would suggest that having a textbook nearby might be useful for study as opposed to merely relying on these notes. There are a few steps to studying statistics that I would recommend. First would be to write down concepts summed-up basically in one-word or short sentences. Second, you could take those terms and connect them to a definition, an example and a reason for WHY it is used or important. If it is a numerical concept, my suggestion would be to practice the calculation. Third, you could then take the connections that you created and then apply it to certain practice problems which are found in textbooks.



Chapter 1 - Central Tendency

1.1 What are measures of Central Tendency?

Measures of central tendency tell us the middle scores/values or the central point in a given data set. These measures are most commonly the mean, median and mode. These measures are the 'bread and butter' of statistics and play a role in almost all statistical concepts.

The **mean** is the average score - the sum of all scores in a dataset divided by the number of scores there are in that dataset.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

The **median** is the middle number if all the scores in a given dataset were to be arranged from lowest to highest.

Consider this dataset:

25, 33, 41, 15, 12, 45, 36, 22, 44

(ordering from lowest to highest): ~~12, 15, 22, 25,~~**33**, ~~36, 41, 44, 45~~

33 is the median value of this dataset.

The **mode** is the most frequent value that occurs in a given dataset.

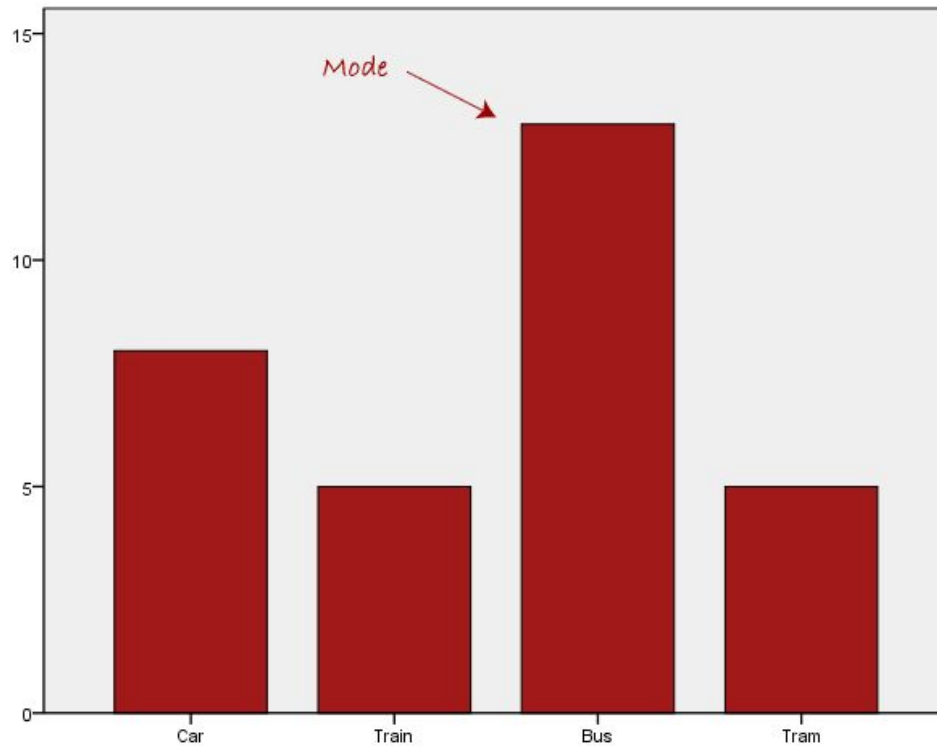


Figure 1.0

As we see in Figure 1.0, the bus is the mode of transport that is most commonly used compared to the others.

1.2 What are measures of dispersion?

Similar to measures of central tendency, we can see the measures of dispersion, which are as essential to statistics, if not more, than the measures of central tendency. The measures of dispersion tell us how scattered the data is, within a dataset. In stats terms, it tells how spread out and variable the data is. There are three measures of dispersion; the range, variance and standard deviation.



The **range** is the difference between the highest and the lowest scores/values in a dataset. The interquartile range is the difference between the top 75th percent value and the bottom 25 percent value.

The **variance** is the sum of the square of the differences between a score/value and the mean divided by the number of values in the dataset. For a sample the variance is the same as above but divided by the (number of values in the dataset - 1).

The **Standard Deviation** is the square-root of the variance. In simple terms the standard deviation is the average variation that a score/value would have from the mean.

The SD is the square root of the Variance. It is seen in the following formulas:

- For the **population**

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

σ = population standard deviation

N = the size of the population

x_i = each value from the population

μ = the population mean



- For the **sample**

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$$

Example of a Standard Deviation (for a sample) Calculation:

Observation	X (Independent Variable)	(X - Mean) "Mean = 80.6"	(X - Mean)^2
1	85	4.4	19.4
2	91	10.4	108.2
3	97	16.4	269.0
4	72	-8.6	74.0
5	68	-12.6	158.8
6	90	9.4	88.4
7	79	-1.6	2.6
8	88	7.4	54.8
9	55	-25.6	655.4
SUM TOTAL			1430.6

Table 1.0

$$SD = \sqrt{1430.6 / 8} = \mathbf{13.4}$$



Chapter 2 - Measurement

2.1 Why Use Measurement?

Measurement in psychology, is the idea of assigning numbers to constructs, facets and basic psychological ideas. The reason for assigning numbers is to get an idea as to the statistical nature of things, such as, when writing papers and analyzing the results obtained. We use numbers to show us that what we are measuring has meaning and we use our measurement to prove a hypothesis. An example in clinical psychology may be to see whether a drug improves the ability for individuals with ADHD to focus. This is then done by randomly assigning participants into two groups and then administering a placebo to one group and the drug to the other group. Whilst we cannot measure '**focus**' there are indicators we can use to measure it...or at least come close to measuring it. This could be something like continuous time spent on a particular task.

2.2 Four types of measurement

There are four types of measurement. They are incremental which means that at each level they increase in complexity and carry more meaning so to speak. These are:

1. Nominal

Nominal measures are categorical by nature. It means that the numbers used are 'dummy codes' or just represent different groups without providing any more significance to them. This simple labeling of categories is seen in gender; 1 = Female, 2= Male or in sports jersey numbers.



2. Ordinal

Ordinal measures are categorizations like nominal, except that the numbers have a specific order to them. It is used to signify labels of variables where there is a directionality that exists. From the previous example, gender does not have a directionality, but an ordinal variable like college year groups do. 1 = Freshman, 2 = Sophomore, 3 = Junior, 4 = Senior. There is some directionality to it. The Likert scales we see often in psychology, eg. 1 = Strongly Disagree, 5 = Strongly Agree, are ordinal variables.

3. Interval

Interval measures, like ordinal, contain a certain directionality but they are also able to be added and subtracted. This is seen in measures such as temperature (in celsius and fahrenheit). The equal intervals between variables allows these operations to be used. The zero that we encounter in interval measures is an arbitrary zero. That means that the value of 0 is a part of the measurement and has meaning.

4. Ratio

Ratio measures are the most complex and agile of the four types. They have an absolute zero. This allows for the operations of multiplication and division of the variables in addition to the operations that an interval variable has. The presence of the absolute zero means that at zero the variable doesn't exist. Such as distance or weight being zero mean that there simply is no distance and no weight.

2.3 Discrete and Continuous Variables

Another important classification in measurement is whether a variable is continuous or discrete. A continuous variable is one where there is a range of numbers which can be a valid measure, such as height or weight measures. A discrete variable is where there



are increments between values (able to be counted). A continuous variable is one that is obtained through measurement, whilst a discrete variable is not a measured one but something qualitative.

Dichotomous variables are discrete and are seen in dummy coding such as the coding of gender (seen above), which may be done through 1's and 0's. Dichotomy means that there are two options. These are used mostly for the comparison of two different groups, for statistical analyses in SPSS or R systems.

An example would be if I wanted to 'dummy code' whether people have completed a required course or not. 1 would be 'not complete' and 2 would be 'complete'. Using this coding now I would be able to apply these values on SPSS to separate the two different groups.



Chapter 3 - Z-Scores

The title of this chapter is called Z-scores, however it is going to focus on more than just Z-scores. Using technology finding Z-scores can be done quickly using SPSS. However, the understanding behind why this standardized score is important to psychology is necessary.

3.1 What are Z-scores?

Z-scores tell us how many standard deviations higher or lower than the mean a particular value is. The mean in question is said to have a Z-score of zero. Usually a good measure to go by is that 99% of all the data is between a Z-score of -3 and +3.

The equation for calculating Z-scores is:

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean

σ = Standard Deviation

3.2 Standardized Scores

Standardized scores or 'scale scores' - as used by some professors - are commonly used for various purposes. One of the most common one's students will know of is the SAT. The scores in the SAT are standardized to reflect each person's score around a given mean. Using the Z-scores we can then center the scores around a mean through knowing the distribution of the dataset.



The following is an example of standardization:

[The population **mean** for raw scores is 21 and the **SD** is 3.5]

Person	Raw Score	Z-score
Johnny	22	0.29
Betty	16	-1.43
Alice	26	1.43
Benny	12	-2.57

Table 3.0

Say we standardize all of these scores raw scores to a mean of 80 and SD of 7:

Standardized Score = [Z-score] x [SD] + [Mean]

Applying this equation we can then convert the Z-score to a Standardized scale score, as seen in the third column of table 3.1.

Person	Z-score	Standardized
Johnny	0.29	82
Betty	-1.43	70
Alice	1.43	90
Benny	-2.57	62

Table 3.1

Standardized scores are usually rounded to the nearest whole number. When these scores are converted to **percentile ranks** - which represent the percentage of scores in the distribution that a given score is better than - they are truncated which is the removal of any decimal without any rounding.



Chapter 4 - NHST

4.1 What is NHST?

Null Hypothesis Significance Testing (NHST) would be a term that most of you would have heard of. This process uses inferences of observations by which we can test a hypothesis. There are two types of hypotheses; the null and the alternative:

The **null hypothesis** states that there is **NO DIFFERENCE** between the two groups or in the case of a treatment there is **NO EFFECT** of an intervention method. The null is represented by the symbol H_0 .

The **alternative hypothesis** states that there **IS A DIFFERENCE** seen between two groups or in the case of a treatment there **IS AN EFFECT** of an intervention method. The alternative is represented by the symbol H_1 .

There are two types of alternative hypotheses. The first is used in **one-tailed** significance - which simply means that there is a **directionality (< or >)** stated in the H_1 . It could be said such as one group being better than another. The second is used in **two-tailed** significance - the H_1 has **no particular directionality (\neq)**, but just states that the two groups are different.

4.2 NHST Process

There is a process that is traditionally followed when conducting NHST. The steps below define the key elements for testing for or against a hypothesis. The use of statistical software such as SPSS or R are becoming increasingly popular compared to the old-school tables that statisticians use.



1. Defining Your Hypotheses

The first part of NHST is to define the hypothesis. Usually this refers to the alternative hypothesis, if there is a difference and what direction the difference is. The null hypothesis is the 'boring one' which just negates the alternative. Once we have set our hypotheses we can then continue the process. We begin every NHST process by ***assuming that the null is true.***

2. Power Analysis

The power analysis is then conducted which tells us about the sample size we will need for doing a one-tailed or two-tailed test. The statistical power has to be at least 0.80. The power is important because it allows us to have sample sizes that we need in order to make sure we don't miss any effect that may be there. Also, it helps us not conclude that there is an effect when there is in fact no effect.

3. Setting the Alpha

The alpha level is most commonly referred to as the Type 1 error. The precise meaning of this is; ***the probability that we reject the null hypothesis when it is true.*** In defining our hypotheses, we kept our null hypothesis as 'true' so in that sense, what the alpha level will tell us, is how likely we are to reject that null hypothesis being true. The alpha level is usually kept at 0.05 (also known as 5% error rate).

4. Collecting Data

Collecting data is a fairly self-explanatory process. There are a few complications to it where sampling of groups needs to be fair so that no other known confounds will impact the results. Avoiding biases are a key aspect of data collection.



5. Calculate Test Statistic

There are various different test statistics computed such as Z-scores, t-test statistics, F-statistic (in the case of an ANOVA). The meaning of a test statistic is; ***the probability that the null hypothesis is true.*** Using this statistic we then check them in the tables in order to determine the p-value. Alternatively, and now more commonly, we would use SPSS, which would tell us the p value (the probability of rejecting the null if it is true).

6. Conclusion

If the p value is higher than 0.05, we can conclude that more than 5% of the null is true when we have rejected it, and therefore we can conclude that we ***fail to reject the null hypothesis.*** We **CANNOT** say that we accept it, because we don't have evidence of the absence of any effect that could be.

If the p value is less than 0.05, we can conclude that less than 5% of the null is true when we reject it and therefore we say that we ***reject the null hypothesis.*** Note that this isn't the same as accepting the alternative hypothesis. The post hoc tests may confirm this finding for us (such as looking at differences in means and other factors).

Our final conclusion will then be applied to the study that we are conducting and reported in **APA format** in context of the study.

4.3 The Good, Bad and Ugly of NHST

While we see that the aforementioned process is a real gift to statistics, we need to also see the limitations of NHST. There are a few that can be seen:

First is that NHST is an 'all or nothing' method of testing a hypothesis. It uses the data collected and provides us a conclusion of significance through the p-value. Once we find something as being significant, we quickly rise to the conclusion that there is a



difference without knowing **HOW MUCH** difference is actually present. Smaller sample sizes will show us a very liberal p-value which is insensitive to the perceived difference. Whilst large sample sizes may show a non-significant value when there is actually a large difference between the groups.

Second, there is an issue that if the sampling of the two groups is not done properly - a.k.a there is a biased sample of one group - the tinkering of the observations can affect the results. This is not to say that scientists prefer to fake their way into results, but just that groups need to be sampled in a fair way.

4.4 Deadly Sins of Statistics

There are two important “deadly sins” in statistics that you will hear a lot about from professors. These are p-hacking and HARKing. Considering that these are weird words/ acronyms, they will make more sense as we discuss.

P-hacking is something that researchers could easily fall into the trap of. While it may not seem to be a direct violation of research, it still goes against the basic principles. The repeating of studies multiple times so that the p-value is significant and only reporting the significant p-value is called p-hacking. This is wrong because the significance would appear to be by chance in which the two groups would have data that is very different and not because they are actually different.

HARKing means ‘hypothesising after the results are known’. In research, usually the hypotheses would be set first followed by data collection. Throughout the process, the hypotheses originally stated would still stand regardless of the data received. What ‘HARKers’ do is look at the data that they collected and change their hypothesis as a result in order for their papers to be of scientific significance and to be published.



Chapter 5 - T-Tests

5.1 Introduction to T-tests

T-tests are very popular measures of finding the difference between means. They are used to find the differences between two groups. Using the t-test statistic, we are able to see whether the difference is statistically significant. The following formulas are used for calculating t:

Independent Samples t-test:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$$

Paired-Samples t-test:

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}}$$

There are two kinds of t-tests; the independent samples t-test - also known as the independent samples t-test - and the paired t-test (which means that the two groups are linked).



5.2 Independent Samples t-test

The independent samples t-test involves two groups of independent variables that are not related to each other. An example of this would be, looking at two classes (in two different schools) in the third grade of elementary school of scores on a reading test. Each class is an independent group that is not linked to or dependent on the other. Using the mean of the two groups we can calculate the t statistic. The standard deviation of each group is adjusted based on their size (N) and converted to a pooled SD.

5.3 Paired Samples t-test

The paired samples t-test is used in situations where the two groups are connected to each other. For example, if we wanted to see the performance of participants before training vs after training. These are usually used to see if interventions have an effect. The paired aspect of the t-test is displayed when we find the difference between each observed score between the two groups and compare that mean with zero - our null hypothesis that there is no difference between the groups. The standard deviation is calculated on the difference as well (which means there's only one SD). The t statistic is calculated using the formula seen above. After looking at the t-statistic and looking through the tables where degrees of freedom (N-1) and the set p-value match up we can see whether there is significance or not. Using software, such as SPSS and R also provide us with outcomes and calculated p-values without the need for looking at tables.



5.4 Examples

Independent Samples:

Suppose you are a psychology professor and you're measuring the test anxiety of two classes in a freshman psychology course. You take a sample of 10 students from each class and want to see whether there is a difference in test anxiety between the two classes. You do this by sending out a questionnaire in a Likert response format, where the higher the score means the student is more anxious. The scores are seen in the table below:

Class A	Class B
4.86	4.49
3.91	3.5
2.55	4.1
2.82	1.19
3.31	4
2.11	3.12
4.23	4.55
2.2	3.97
4.67	2.88
3.2	2.9

Table 5.0

What is the t-statistic and the subsequent conclusion you would make?

Mean of each group:

Class A = 3.386, Class B = 3.47



Standard Deviation of each group (calculated from equations in chapter 1):

Class A = 1.00716, Class B = 0.99458

$SD\ Pooled = \sqrt{[(1.00716)^2/10 + (0.99458)^2/10]}$

$SD\ Pooled = 0.4476$

$T = (Mean\ class\ B - Mean\ of\ class\ A)/SD\ pooled$

$T = (3.47 - 3.386)/ 0.4476 = \mathbf{0.187}$

A t-statistic of 0.187 is considered to be low. The p-value for that t-statistic is 0.853 which is enough of a conclusion to say we fail to reject the null hypothesis. This means that we do not have evidence to suggest that Class A students suffer different amounts of test anxiety than Class B.

Paired Samples:

Consider that we're measuring the effect of a bootcamp training on weight loss, as measured by waist circumference (in inches). I have a class of 120 from whom I take a random sample of 5 people with their pre-bootcamp and post-bootcamp waist sizes listed below:

Pre-Bootcamp	Post-Bootcamp	Difference (pre - post)
42.3	38.4	3.9
43.3	40.1	3.2
36.8	35.4	1.4
52.2	45.6	6.6
48.8	43.1	5.7

Table 5.1

What is the t-statistic and the subsequent conclusion you would make?



Since it is a paired samples *t*-test, we first combine the two groups by finding the difference between each person's pre and post-bootcamp waist size (seen in column 3 of table 5.2).

Mean of Difference = 4.16

Standard Deviation (*s*) = 2.0574 (calculated through using the equation in chapter 1)

Since we are comparing the mean difference to the null - which is zero, our calculations of the *t*-statistic will be as follows:

$$T = (\text{Mean of difference} - 0) / (s/\sqrt{N})$$

$$T = (4.16 - 0) / (2.0574/\sqrt{5})$$

$$T = \mathbf{4.52}$$

A *t*-statistic of 4.52 would result in a *p*-value of **0.0107** from which we can conclude that there is a significant difference between the pre and post-bootcamp waist sizes from the data that we collected. In stats terms that would mean that we have enough evidence to reject the null hypothesis.

The values *p*-values can be found by finding them in tables using *t*-statistics and degrees of freedom - which is calculated by, ***df* = *N* - 1**.

This YouTube video should help you understand *T*-tests with an example of a paired samples *t*-test:

<https://www.youtube.com/watch?v=AGh66ZPpOSQ>



Chapter 6 - Linear Regression

When we think of the word linear, the word ‘line’ and the well-known equation ‘ $y = mx + c$ ’ are the first things we think of....or maybe that’s just me. In any case though, that generic equation is the basis of the linear regression model.

6.1 The Linear Regression

In a linear regression we try to develop a model for our data in a way that it makes sense. This is done when there are many variables that play a role to feed into one dependent variable. Usually the equation of a linear regression is seen as follows:

$$\text{Outcome (dependent variable)} = b_0 + b_1X_1 + b_2X_2 + \text{error}$$

As we see above, we want to try to reduce the amount of unexplained error. The model in a linear regression would use different independent variables (X_1, X_2, \dots) in which help in explaining the outcome. This is particularly important when we have more than one independent variable. The error will feed into b_0 and become a part of that “intercept”. This model is helpful in exploratory statistics, predicting models or forecasting. A very prevalent example is the modelling done for covid 19 which may not be linear but modelling nonetheless.

6.2 An Example

Considering that I am an I/O psychology student let’s look at the big construct ‘job satisfaction’:

Let’s assume that we are finding want to test three predictors of job satisfaction. We know that “love of the job” is one factor, “camaraderie” with other employees is another and finally “Promotion opportunities” is the third. We collect the information



from 5 people in an organization as follows, using a 5 point likert type questionnaires that we set:

Person	Love of Job	Camaraderie	Promotion Opportunities	Job Satisfaction
Billy	4.2	4.5	4.4	4.7
Bobby	3.5	2.8	2.7	3.2
Brenda	2.4	3.1	2.0	2.6
Betty	3.7	3.9	3.6	4.1
Brent	2.9	4.0	3.3	3.7

Table 6.0

Computing the linear regression into SPSS or R would result in a model being formed using this data. If the p value is less than 0.05 there is a likelihood that the variable is a predictor.

In this example, there didn't seem to be any significant values that could be put into a model. Although if I removed camaraderie from the model, it would make the 'promotion opportunities' predictor significant. This is a good example of predictors having effects that would prevent another predictor from explaining the outcome.

In this case the final equation of the model came to:

$$\text{Job Satisfaction} = 0.31 + 0.19*\text{Love_job} + 0.19*\text{camaraderie} + 0.63*\text{Promotion}$$

This YouTube video should help you to understand linear regressions:

<https://www.youtube.com/watch?v=WWqE7YHR4Jc>



References

Field, A. (2017). *Discovering statistics using IBM SPSS statistics*, 5th edition.
Thousand Oaks, CA: Sage Publications.