Hypotheses and its Testing

DR. NILESH B. GAJJAR
M.Com. (A/c), M.Com. (Mgt.), M.Ed., M. Phil., G.Set. Ph. D.
Assistant Professor,
S. V. S. Education College, (M.Ed.) P. G. Dept.,
Nagalpur, Mehsana, Gujarat (India)
drgajjar75@yahoo.com
nbg61075@rediffmail.com
editorraijmr@yahoo.com
http://www.raijmr.com

Abstract:
A statistical hypothesis test is a method of making decisions using data from a scientific study. In statistics, a result is called statistically significant if it has been predicted as unlikely to have occurred by chance alone, according to a pre-determined threshold probability, the significance level. The phrase "test of significance" was coined by statistician Ronald Fisher. These tests are used in determining what outcomes of a study would lead to a rejection of the null hypothesis for a pre-specified level of significance; this can help to decide whether results contain enough information to cast doubt on conventional wisdom, given that conventional wisdom has been used to establish the null hypothesis. The critical region of a hypothesis test is the set of all outcomes which cause the null hypothesis to be rejected in favor of the alternative hypothesis. Statistical hypothesis testing is sometimes called confirmatory data analysis, in contrast to exploratory data analysis, which may not have pre-specified hypotheses. Statistical hypothesis testing is a key technique of frequentist inference. Here the author of this article wants to introduce the testing of hypotheses.

Keywords: *Alternative hypotheses, Null Hypotheses, One tail & two tail test, Statistics, Type I & II error*

2. Introduction

A hypothesis is a specific statement of prediction. It describes in concrete (rather than theoretical) terms what you expect will happen in your study. Not all studies have hypotheses. Sometimes a study is designed to be exploratory. There is no formal hypothesis, and perhaps the purpose of the study is to explore some area more thoroughly in order to develop some specific hypothesis or prediction that can be tested in future research. A single study may have one or many hypotheses.

Actually, whenever I talk about hypotheses, I am really thinking simultaneously about *two* hypotheses. Let's say that you predict that there will be a relationship between two variables in your study. The way we would formally set up the hypothesis test is to formulate two hypothesis statements, one that describes your prediction and one that describes all the other possible outcomes with respect to the hypothesized relationship. Your prediction is that variable A and variable B will be related (you don't care whether it's a positive or negative relationship). Then the only other possible outcome would be that variable A and variable B are *not* related. Usually, we call the hypothesis that you support (your prediction) the ***alternative*** hypothesis, and we call the hypothesis that describes the remaining possible outcomes the ***null*** hypothesis. Sometimes we use a notation like H_A or H_1 to represent the alternative hypothesis or your prediction, and H_O

or H_0 to represent the null case. You have to be careful here, though. In some studies, your prediction might very well be that there will be no difference or change. In this case, you are essentially trying to find support for the null hypothesis and you are opposed to the alternative.
If your prediction specifies a direction, and the null therefore is the no difference prediction and the prediction of the opposite direction, we call this a ***one-tailed hypothesis***. For instance, let's imagine that you are investigating the effects of a new employee training program and that you believe one of the outcomes will be that there will be *less* employee absenteeism. Your two hypotheses might be stated something like this:

The null hypothesis for this study is:
H_O: As a result of the XYZ company employee training program, there will either be no significant difference in employee absenteeism or there will be a significant *increase*. Which is tested against the alternative hypothesis.

H_A: As a result of the XYZ company employee training program, there will be a significant *decrease* in employee absenteeism. See fig. 1.

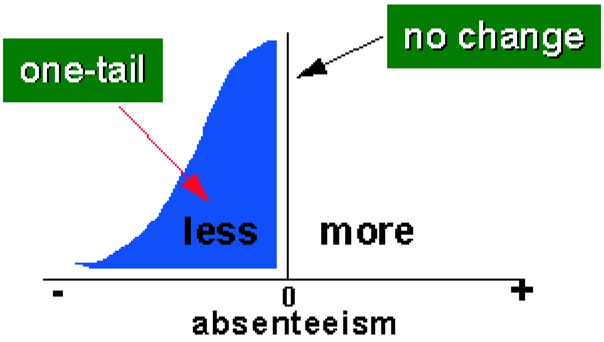

Fig. 1 Asenteeism

In the figure 1 on the left, we see this situation illustrated graphically. The alternative hypothesis -our prediction that the program will decrease absenteeism is shown there. The null must account for the other two possible conditions: no difference or an increase in absenteeism. The figure shows a hypothetical distribution of absenteeism differences. We can see that the term "one-tailed" refers to the tail of the distribution on the outcome variable.

When our prediction does *not* specify a direction, we say you have a ***two-tailed hypothesis***. For instance, let's assume you are studying a new drug treatment for depression. The drug has gone through some initial animal trials, but has not yet been tested on humans. You believe (based on theory and the previous research) that the drug will have an effect, but you are not confident enough to hypothesize a direction and say the drug will reduce depression (after all, you've seen more than enough promising drug treatments come along that eventually were shown to have severe side effects that actually worsened symptoms). In this case, you might state the two hypotheses like this:

The null hypothesis for this study is:
H_O: As a result of 300mg./day of the ABC drug, there will be no significant difference in depression. This is tested against the alternative hypothesis.

H_A: As a result of 300mg./day of the ABC drug, there will be a significant difference in depression.

The figure 2 on the right illustrates this two-tailed prediction for this case. Again, notice that the term "two-tailed" refers to the tails of the distribution for your outcome variable.

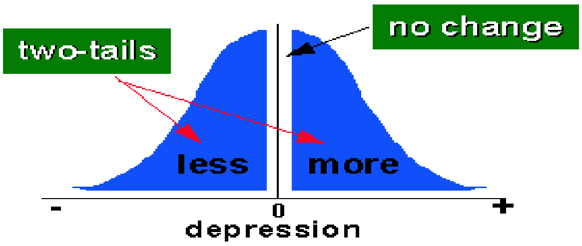

Fig.2 Depression

The important thing to remember about stating hypotheses is that you formulate your prediction (directional or not), and then you formulate a second hypothesis that is mutually exclusive of the first and incorporates all possible alternative outcomes for that case. When your study analysis is completed, the idea is that you will have to choose between the two hypotheses. If your prediction was correct, then you would (usually) reject the null hypothesis and accept the alternative. If your original prediction was not supported in the data, then you will accept the null hypothesis and reject the alternative. The logic of hypothesis testing is based on these two basic principles:

- the formulation of two mutually exclusive hypothesis statements that, together, exhaust all possible outcomes
- the testing of these so that one is necessarily accepted and the other rejected

We know it's a convoluted, awkward and formalistic way to ask research questions. But it encompasses a long tradition in statistics called the ***hypothetical-deductive model***, and sometimes we just have to do things because they're traditions. And anyway, if all of this hypothesis testing was easy enough so anybody could understand it, how do you think statisticians would stay employed?

3. Hypothesis Testing and Effect Size

3.1 One-Sample Designs

3.1.1 What is Hypothesis Testing?

A researcher uses **hypothesis testing** to support beliefs about comparisons (i.e., variables or groups). Basically, it is how we empirically test our research hypotheses for "accuracy." We NEVER prove beyond the shadow of a doubt that a comparison is true. Rather, we conclude that, based on some collected data and assumptions, the probability of the comparison being true is very high (i.e., around 95 – 99% sure).

In all hypotheses testing, the hypothesis being tested is a hypothesis about *equality*. The researcher thinks the equality hypothesis is NOT true, and by showing how the data do not fit it, the equality hypothesis can be rejected.

We call this equality hypothesis the **null hypothesis**, and its symbol is: H_0. The null hypothesis is a statement comparing two statistics (usually two means). Seldom in your comparisons do you want to show that two statistics or groups are the same? Usually your research questions will want to demonstrate that two statistics/groups are different from each other. Ex. Our Panic attack reduction drug trial research with different dosages. We want to know which group (dose) reduces attacks the best.

This *difference* hypothesis is the **alternative hypothesis**, and its symbol is: H_a or H_1. The alternative hypothesis is a statement comparing two statistics or groups, suggesting there is a difference.

A. Five Steps of Hypothesis Testing

1. Write the null hypothesis (H_0).
2. Write the alternative hypothesis (H_1).
3. Set alpha level (amount of error allowed) and determine degrees of freedom.
4. Pick & calculate the significance test that fits your design.
5. Decision Step: Accept or Reject the null.

3.2 Using the t Distribution for Hypothesis Testing

Recall from chapter 7 that any empirical distribution, through creation of a sampling distribution, will result in a normal distribution (theoretical) and share the characteristics of the normal curve. Therefore, we can use the t distribution to help us analyze the Doritos problem. We want to answer the following question:

Is it probable (likely) that a sample with a mean of 270.675 came from a population with Mu of 269.3?

If the difference is small and we conclude, "yes," then we accept the null. If the difference is too big, we conclude, "no, it isn't likely," then we reject the null and accept the alternative hypothesis.

3.3 The One-Sample t Test

For our scenario, the **one-sample t Test** is appropriate. It is used when one needs to compare a sample mean to a population mean.

A. One-Sample t Test Steps

1. H_0: $\underline{X} = M$
2. H_1: $\underline{X} \neq M$ or H1: $\underline{X} < M$ or H1: $\underline{X} > M$
3. Set alpha to a = .05 and $df = n - 1 = 7$
4. one-sample t Test
5. Decision

Because the **t-obtained** (value we calculate) is greater than the **t critical** (value you look up on t table), we reject the null and accept the alternative hypothesis. Therefore, it is unlikely that this sample of Doritos came from this population, or rather; the sample data suggest that Doritos bags weigh significantly more than they advertise.

Note: In interpreting, use the terms of the experiment and give the direction of the difference (more vs. less). The positive (+) or negative (-) sign on the t value doesn't matter when using the t table.

3.4 Types of Errors in Decision Making

For each decision we make, we run the risk of making an error. From reviewing our above example, you may have concluded that if we want to be more certain, we could use a more *stringent* alpha level (i.e., .01 instead of .05). So we are only 1/100 times likely to be wrong than 5/100 times.

Unfortunately, when we decide to reject null, we always run the risk of making what we call a **Type I Error**. A Type I error is when our decision is to reject the null when in fact, the null is true.

Using a more stringent alpha level (.01) decreases the likelihood we will make a Type I error, but it also INCREASES the likelihood we will make a **Type II Error.** A Type II error occurs when we decide to accept the null when in fact, the alternative is true. The Decision table is mentioned in following table 1.

Table 1 Type I and Type II Errors

	True Situation In Population		
Decision made Based upon Sample Data	**Our Decision**	**H_0 True**	**H_0 False**
	H_0 False (reject)	Type I Error	Correct Decision
	H_0 True (accept)	Correct Decision	Type II Error

Note: Probability of making a Type I error = alpha. The probability of making a Type II error is symbolized, b. They are inversely related.

A. Meaning of *p* in *p* < .05

Every statistical test has its own sampling distribution (like the t test). The *p* is always the probability of the statistical test value (e.g., t-obtained) when null is true. If t-test gives you *p* < .05, the sample results you obtained actually occur fewer than 5 times in 100 *when the null is true.* So 95/100 times, the null is false.

The ***p*** is the probability of the data obtained (t-obtained), if the null hypothesis is true.

4. One and Two-tailed Tests

Researchers should set alpha level and decide if they are computing a one-tail or two-tail test before data are gathered. These are **a priori** predictions (before the fact). By choosing a one- or two-tailed test, one is picking from three possible alternative hypotheses:

1. $H_1 = \underline{X}$ ne M (two-tailed)
2. $H_1 = \underline{X} < M$ (one-tailed)
3. $H_1 = \underline{X} > M$ (one-tailed)

You decision is based on your research design and question. Fig. 8.2

5. Effect Size Index

Remember effect size? It allows you to answer, "How much difference is there?" Or rather, "How meaningful is this difference?"
Whereas the test statistics (like the t-test) tells you if a significant difference exist or not, it tells you nothing about the degree or magnitude of the difference.
Here is the effect size (definitional) formula for a one-sample t-test: $d = |\underline{X} - m| / s$
The guidelines for interpreting the effect size remain.

6. Using the t Distribution to test a Correlation Coefficient

The researcher may also use the t-test and t distribution table to help you evaluate a correlation between two variables. You test the statistical significance of the difference between the population correlation (r) and the sample correlation (r).

The population parameter for the Pearson correlation is rho (r). It looks a lot like lower-case p for probability.

1. H_0: r = .00
2. H_1: r ne .00 (two-tailed)

The formula for a t-test on a correlation is: t = (r) VN - 2 / 1 - r2

7. Setting Alpha Level

An alpha level of .05 is arbitrary. It was set as a standard by scientists. As a researcher it is recommended you use a = .05 or lower (more stringent).

References

1. Bakan, David (1966). The test of significance in psychological research Psychological bulletin 66 (6): 423–437.
2. Fisher, R. A. (1925). Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd, 1925, p.43.
3. Fisher, Sir Ronald A. (1956). Mathematics of a Lady Tasting Tea. In James Roy Newman. The World of Mathematics, volume 3 [Design of Experiments]. Courier Dover Publications. ISBN 978-0-486-41151-4. Originally from Fisher's book Design of Experiments.
4. Hinkelmann, Klaus and Kempthorne, Oscar (2008). Design and Analysis of Experiments. I and II (Second Ed.). Wiley. ISBN 978-0-470-38551-7.
5. Kaye, David H., Freedman, David A. (2011). Reference Guide on Statistics. Reference Manual on Scientific Evidence (3rd ed.). Eagan, MN Washington, D.C: West National Academies Press. p. 259. ISBN 978-0-309-21421-6.
6. Lehmann, E.L., Romano, Joseph P. (2005). Testing Statistical Hypotheses. (3rd Ed.). New York: Springer. ISBN 0-387-98864-5.
7. Montgomery, Douglas (2009). Design and analysis of experiments. Hoboken, NJ: Wiley. ISBN 978-0-470-12866-4.
8. Richard J. Larsen, Donna Fox Stroup (1976). Statistics in the Real World: a book of examples. Macmillan. ISBN 978-0023677205.
9. Schervish, M. (1996). Theory of Statistics. p. 218. Springer ISBN 0-387-94546-6
10. Triola, Mario (2001). Elementary statistics (8th Ed.). Boston: Addison-Wesley. p. 388.ISBN 0-201-61477-4.