Hypothesis Testing

Definitions of Hypothesis

- "Any supposition which we make in order to endeavour to deduce conclusions in accordance with facts which are known to be real under the idea that if the conclusions to which the hypothesis leads are known truths, the hypothesis itself either must be or at least likely to be true." - **J.S. Mill**

- "A hypothesis is a tentative generalization the validity of which remains to be tested. In its most elementary stage the hypothesis may be any hunch, guess, imaginative idea which becomes basis for further investigation." - **Lundberg**

- "It is a shrewd guess or inference that is formulated and provisionally adopted to explain observed facts or conditions and to guide in further investigation**."** - **John W. Best**

Nature of Hypothesis

(i) Conceptual: Some kind of conceptual elements in the framework are involved in a hypothesis.

(ii) Verbal statement in a declarative form: It is a verbal expression of ideas and concepts. It is not merely mental idea but in the verbal form, the idea is ready enough for empirical verification.

(iii) It represents the tentative relationship between two or more variables.

(iv) Forward or future oriented: A hypothesis is future-oriented. It relates to the future verification not the past facts and information.

(v) Pivot of a scientific research: All research activities are designed for verification of hypothesis.

Functions of Hypothesis

(i) It is a temporary solution of a problem concerning with some truth which enables an investigator to start his research works.

(ii) It offers a basis in establishing the specifics what to study for and may provide possible solutions to the problem.

(iii) It may lead to formulate another hypothesis.

(iv) A preliminary hypothesis may take the shape of final hypothesis.

(v) Each hypothesis provides the investigator with definite statement which may be objectively tested and accepted or rejected and leads for interpreting results and drawing conclusions that is related to original purpose.

(vi) It delimits field of the investigation.

(vii) It sensitizes the researcher so that he should work selectively, and have very realistic approach to the problem.

(viii) It offers the simple means for collecting evidences for verification.

Forms of Hypothesis

(i) Question form:

A hypothesis stated as a question represents the simplest level of empirical observation. It fails to fit most definitions of hypothesis. It frequently appears in the list. There are cases of simple investigation which can be adequately implemented by raising a question, rather than dichotomizing the hypothesis forms into acceptable / reject able categories.

(ii) Declarative Statement:

A hypothesis developed as a declarative statement provides an anticipated relationship or difference between variables. Such a hypothesis developer has examined existing evidence which led him to believe that a difference may be anticipated as additional evidence. It is merely a declaration of the independent variables effect on the criterion variable.

(iii) Directional Hypothesis:

A directional hypothesis connotes an expected direction in the relationship or difference between variables. This type of hypothesis developer appears more certain of anticipated evidence. If seeking a tenable hypothesis is the general interest of the researcher, this hypothesis is less safe than the others because it reveals two possible conditions. First that the problem of seeking relationship between variables is so obvious that additional evidence is scarcely needed. Secondly, researcher has examined the variables very thoroughly and the available evidence supports the statement of a particular anticipated outcome.

(iv) Non –Directional Hypothesis or Null Hypothesis:

This hypothesis is stated in the null form which is an assertion that no relationship or no difference exists between or among the variables. Null hypothesis is a statistical hypothesis testable within the framework of probability theory. It is a non-directional form of hypothesis. There is a trend to employ or develop null hypothesis in research in most of the disciplines**.** A null hypothesis tentatively states that on the basis of evidence tested there is no difference. If the null hypothesis is rejected, there is a difference but we do not know the alternative or the differences. In this the researcher has not to anticipate or give the rational for the declaration or directional form. It does not make researcher biased or prejudiced. He may be objective about the expected outcomes of the research or findings. Actually this is a statistical hypothesis which is self- explanatory.

Null hypothesis means zero hypotheses. A researcher has not to do anything in developing it. While research hypothesis is second step in the process of reflective thinking. A null hypothesis in an appropriate form is order to accommodate the object of inquiry for extracting this information. It does not necessarily reflect the expectations of the researcher so much as the utility of the null form as the best fitted to the logic of chance in statistical knowledge or science.

It is the no difference form, i.e. there is no difference or relationship between or among variables under certain conditions.

Statistical tests of significance are used to accept and reject the null hypothesis. If it is rejected, the general hypothesis is accepted.

Non-directional hypothesis is known as null hypothesis because it 'nullifies' the positive argument of the findings or non-directional statement of the generalization. It is also termed as statistical or zero hypothesis because it denies the existence of any systematic principles apart from the effect of chance. It assumes that none or zero difference exists between the two population means or the treatments.

Difficulties in the Formulation of Useful Hypothesis

Moving from the operational to the conceptual level and vice –versa is a critical ingredient of the research to demonstration process. The following are the difficulties in the formulation of hypothesis:

1. Absence of knowledge of a clear theoretical framework.

2. Lack of ability to make use of the theoretical framework logically.

3. Lack of acquaintance with available research technique resulting in failure to be able to phrase the hypothesis properly

Hypothesis Test concerning mean

This lesson explains how to conduct a hypothesis test of a mean, when the following conditions are met:

- The sampling method is simple random sampling.
- The sampling distribution is normal or nearly normal.

Generally, the sampling distribution will be approximately normally distributed if any of the following conditions apply.

- The population distribution is normal.
- The population distribution is symmetric, unmoral, without outliers, and the sample size is 15 or less.

- The population distribution is moderately skewed, unimodal, without outliers, and the sample size is between 16 and 40.
- The sample size is greater than 40, without outliers.

This approach consists of four steps:

1. State the Hypotheses

Every hypothesis test requires the analyst to state a null hypothesis and an alternative hypothesis. The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false; and vice versa.

The table below shows three sets of hypotheses. Each makes a statement about how the population mean μ is related to a specified value M. (In the table, the symbol ≠ means " not equal to ".)

Set	Null hypothesis	Alternative hypothesis	Number of tails
1	$\mu = M$	$\mu \neq M$	2
2	$\mu > M$	$\mu < M$	1
3	$\mu < M$	$\mu > M$	1

The first set of hypotheses (Set 1) is an example of a two-tailed test, since an extreme value on either side of the sampling distribution would cause a researcher to reject the null hypothesis. The other two sets of hypotheses (Sets 2 and 3) are one-tailed tests, since an extreme value on only one side of the sampling distribution would cause a researcher to reject the null hypothesis.

2. Formulate an Analysis Plan

The analysis plan describes how to use sample data to accept or reject the null hypothesis. It should specify the following elements.

- Significance level. Often, researchers choose significance levels equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used.
- Test method. Use the one-sample t-test to determine whether the hypothesized mean differs significantly from the observed sample mean.

3. Analyze Sample Data

Using sample data, conduct a one-sample t-test. This involves finding the standard error, degrees of freedom, test statistic, and the P-value associated with the test statistic.

- Standard error. Compute the standard error (SE) of the sampling distribution.

SE = s * sqrt{ (1/n) * [(N - n) / (N - 1)] }

Where s is the standard deviation of the sample, N is the population size, and n is the sample size. When the population size is much larger (at least 20 times larger) than the sample size, the standard error can be approximated by:

SE = s / sqrt(n)

- Degrees of freedom. The degrees of freedom (DF) is equal to the sample size (n) minus one. Thus, DF = n - 1.
- Test statistic. The test statistic is a t statistic (t) defined by the following equation.

t = (x - μ) / SE

Where x is the sample mean, μ is the hypothesized population mean in the null hypothesis, and SE is the standard error.

- P-value. The P-value is the probability of observing a sample statistic as extreme as the test statistic. Since the test statistic is a t statistic, use the t Distribution Calculator to assess the probability associated with the t statistic, given the degrees of freedom computed above. (See sample problems at the end of this lesson for examples of how this is done.)

4. Interpret Results

If the sample findings are unlikely, given the null hypothesis, the researcher rejects the null hypothesis. Typically, this involves comparing the P-value to the significance level, and rejecting the null hypothesis when the P-value is less than the significance level.

Hypothesis concerning proportions

This lesson explains how to conduct a hypothesis test of a proportion, when the following conditions are met:

- The sampling method is simple random sampling.
- Each sample point can result in just two possible outcomes. We call one of these outcomes a success and the other, a failure.
- The sample includes at least 10 successes and 10 failures.
- The population size is at least 20 times as big as the sample size.

This approach consists of four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results.

1. State the Hypotheses

Every hypothesis test requires the analyst to state a null hypothesis and an alternative hypothesis. The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false; and vice versa.

2. Formulate an Analysis Plan

The analysis plan describes how to use sample data to accept or reject the null hypothesis. It should specify the following elements.

- Significance level. Often, researchers choose significance levels equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used.

- Test method. Use the one-sample z-test to determine whether the hypothesized population proportion differs significantly from the observed sample proportion.

3. Analyze Sample Data

Using sample data, find the test statistic and its associated P-Value.

- Standard deviation. Compute the standard deviation (σ) of the sampling distribution.

$$\sigma = \text{sqrt}[\ P * (\ 1 - P\ ) / n\ ]$$

Where P is the hypothesized value of population proportion in the null hypothesis, and n is the sample size.

- Test statistic. The test statistic is a z-score (z) defined by the following equation.

$$z = (p - P) / \sigma$$

Where P is the hypothesized value of population proportion in the null hypothesis, p is the sample proportion, and σ is the standard deviation of the sampling distribution.

- P-value. The P-value is the probability of observing a sample statistic as extreme as the test statistic. Since the test statistic is a z-score, use the Normal Distribution Calculator to assess the probability associated with the z-score. (See sample problems at the end of this lesson for examples of how this is done.)

4. Interpret Results

If the sample findings are unlikely, given the null hypothesis, the researcher rejects the null hypothesis. Typically, this involves comparing the P-value to the significance level, and rejecting the null hypothesis when the P-value is less than the significance level.

ANOVA Test

An **ANOVA** test is a way to find out if survey or experiment results are significant. In other words, they help you to figure out if you need to reject the null hypothesis or accept the alternate hypothesis.

Basically, you're testing groups to see if there's a difference between them**.** Examples of when you might want to test different groups:

- A group of psychiatric patients are trying three different therapies: counseling, medication and biofeedback. You want to see if one therapy is better than the others.
- A manufacturer has two different processes to make light bulbs. They want to know if one process is better than the other.
- Students from different colleges take the same exam. You want to see if one college outperforms the other.

What Does "One-Way" or "Two-Way Mean?

One-way or **two-way** refers to the number of independent variables (IVs) in your Analysis of Variance test.

- One-way has one independent variable (with 2 levels). For example: brand of cereal,
- Two-way has two independent variables (it can have multiple levels). For example: brand of cereal, calories.

What are "Groups" or "Levels"?

Groups or levels are different groups within the same independent variable. In the above example, your levels for "brand of cereal" might be Lucky Charms, Raisin Bran, Cornflakes — a total of three levels. Your levels for "Calories" might be: sweetened, unsweetened — a total of two levels.

Let's say you are studying if an alcoholic support group and individual counseling combined is the most effective treatment for lowering alcohol consumption. You might split the study participants into three groups or levels:

- Medication only,
- Medication and counseling,
- Counseling only.

Your dependent variable would be the number of alcoholic beverages consumed per day.

If your groups or levels have a hierarchical structure (each level has unique subgroups), then use a nested ANOVA for the analysis.

What Does “Replication” Mean?

It’s whether you are replicating (i.e. duplicating) your test(s) with multiple groups. With a two way ANOVA with replication , you have two groups and individuals within that group are doing more than one thing (i.e. two groups of students from two colleges taking two tests). If you only have one group taking two tests, you would use without replication.

Types of Tests

There are two main types: one-way and two-way. Two-way tests can be with or without replication.

- One-way ANOVA between groups: used when you want to test **two groups** to see if there’s a difference between them.
- Two way ANOVA without replication: used when you have **one group** and you’re **double-testing** that same group. For example, you’re testing one set of individuals before and after they take a medication to see if it works or not.
- Two way ANOVA with replication: **Two groups**, and the members of those groups are **doing more than one thing**. For example, two groups of patients from different hospitals trying two different therapies.

One Way ANOVA

A one way ANOVA is used to compare two means from two independent (unrelated) groups using the F-distribution. The null hypothesis for the test is that the two means are equal. Therefore, a significant result means that the two means are unequal.

Two Way ANOVA

A Two Way ANOVA is an extension of the One Way ANOVA. With a One Way, you have one independent variable affecting a dependent variable. With a Two Way ANOVA, there are two independents. Use a two way ANOVA when you have one measurement variable (i.e. a quantitative variable) and two nominal variables. In other words, if your experiment has a quantitative outcome and you have two categorical explanatory variables, a two way ANOVA is appropriate.

Chi- square Test

There are basically two types of random variables and they yield two types of data: numerical and categorical. A chi square (X^2) statistic is used to investigate whether distributions of categorical variables differ from one another. Basically categorical variable yield data in the categories and numerical variables yield data in numerical form. Responses to such questions as "What is your major?" or Do you own a car?" are categorical because they yield data such as "biology" or "no." In contrast, responses to such questions as "How tall are you?" or "What is your G.P.A.?" are numerical. Numerical data can be either discrete or continuous. The table below may help you see the differences between these two variables.

Data Type	Question Type	Possible Responses
Categorical	What is your sex?	male or female
Numerical	Disrete- How many cars do you own?	two or three
Numerical	Continuous - How tall are you?	72 inches

Notice that discrete data arise fom a counting process, while continuous data arise from a measuring process.

The Chi Square statistic compares the tallies or counts of categorical responses between two (or more) independent groups. (note: Chi square tests can only be used on actual numbers and not on percentages, proportions, means, etc.)

2 x 2 Contingency Table

There are several types of chi square tests depending on the way the data was collected and the hypothesis being tested. We'll begin with the simplest case: a 2 x 2 contingency table. If we set the 2 x 2 table to the general notation shown below in Table 1, using the letters a, b, c, and d to denote the contents of the cells, then we would have the following table:

Table 1. General notation for a 2 x 2 contingency table.

Variable 1

Variable 2	Data type 1	Data type 2	Totals
Category 1	a	b	a + b
Category 2	c	d	c + d
Total	a + c	b + d	a + b + c + d = N

For a 2 x 2 contingency table the Chi Square statistic is calculated by the formula:

$$\chi^2 = \frac{(ad - bc)^2(a + b + c + d)}{(a + b)(c + d)(b + d)(a + c)}$$

Note: notice that the four components of the denominator are the four totals from the table columns and rows.

Suppose you conducted a drug trial on a group of animals and you hypothesized that the animals receiving the drug would show increased heart rates compared to those that did not receive the drug. You conduct the study and collect the following data:

Ho: The proportion of animals whose heart rate increased is independent of drug treatment.

Ha: The proportion of animals whose heart rate increased is associated with drug treatment.

Table 2. Hypothetical drug trial results.

	Heart Rate Increased	No Heart Rate Increase	Total
Treated	36	14	50
Not treated	30	25	55
Total	66	39	105

Applying the formula above we get:

Chi square = $105[(36)(25) - (14)(30)]^2 / (50)(55)(39)(66) = 3.418$

Before we can proceed we eed to know how many degrees of freedom we have. When a comparison is made between one sample and another, a simple rule is that the degrees of freedom equal (number of columns minus one) x (number of rows minus one) not counting the totals for rows or columns. For our data this gives (2-1) x (2-1) = 1.

We now have our chi square statistic ($x^2 = 3.418$), our predetermined alpha level of significance (0.05), and our degrees of freedom (df = 1). Entering the Chi square distribution table with 1 degree of freedom and reading along the row we find our value of x^2 (3.418) lies between 2.706 and 3.841. The corresponding probability is between the 0.10 and 0.05 probability levels. That means that the p-value is above 0.05 (it is actually 0.065). Since a p-value of 0.65 is greater than the conventionally accepted significance level of 0.05 (i.e. $p > 0.05$) we fail to reject the null hypothesis. In other words, there is no statistically significant difference in the proportion of animals whose heart rate increased.

What would happen if the number of control animals whose heart rate increased dropped to 29 instead of 30 and, consequently, the number of controls whose hear rate did not increase changed from 25 to 26? Try it. Notice that the new x^2 value is 4.125 and this value exceeds the table value of 3.841 (at 1 degree of freedom and an alpha level of 0.05). This means that $p < 0.05$ (it is now0.04) and we reject the null hypothesis in favor of the alternative hypothesis - the heart rate of animals is different between the treatment groups. When $p < 0.05$ we generally refer to this as a significant difference.

Table 3. Chi Square distribution table.

probability level (alpha)

Df	0.5	0.10	0.05	0.02	0.01	0.001
1	0.455	2.706	3.841	5.412	6.635	10.827
2	1.386	4.605	5.991	7.824	9.210	13.815
3	2.366	6.251	7.815	9.837	11.345	16.268
4	3.357	7.779	9.488	11.668	13.277	18.465
5	4.351	9.236	11.070	13.388	15.086	20.517

To make the chi square calculations a bit easier, plug your observed and expected values into the following applet. Click on the cell and then enter the value. Click the compute button on the lower right corner to see the chi square value printed in the lower left hand coner.

Non-Parametric Tests

A non-parametric test is a hypothesis test that does not make any assumptions about the distribution of the samples.

The Mann-Whitney Test

The Mann-Whitney test, also known as the Wilcoxon rank sum test or the Wilcoxon-Mann-Whitney test, tests the hypothesis that two samples were drawn from the same distribution. The test relies only on the relative ranks of the observations in the combined sample. It does not rely on any properties of the distributions.

The null hypothesis is that the samples were drawn from the same distribution. Location is the dominant factor in the comparison, but if the two distributions have different shapes, this may also affect the result. The alternative hypothesis for the two-tailed test (the default) is that the two samples were drawn from different distributions.

For the one-tailed test, the alternative hypothesis is roughly that the population from which the first sample was drawn has a median that is either less than (lower tailed) or greater than (upper tailed) the median of the population from which the second sample was drawn.

The Mann-Whitney test is implemented by the Mann Whitney Test class. It has three constructors. The first constructor takes no arguments. All test parameters must be provided by setting the properties of the Mann Whitney Test object. The samples the test is to be applied to must be specified by setting the Sample1 and Sample2 properties. The second constructor takes two vector arguments objects that represent the samples the test is to be applied to.

The third constructor also takes two arguments. The first is once again a Vector< T> that contains the observations from the two samples combined. The second argument must be a Categorical Vector with two levels that specifies which sample the observations in the first variable belong to.

The distribution of the Mann-Whitney statistic U can be computed exactly. For larger samples, an approximation in terms of the normal distribution can be used. If the observations don't have distinct ranks (i.e. there are ties), then the exact calculations are not available. The exact test is used by default if there are no ties, and if the product of the two sample sizes is less than or equal to 1600. You can override the default behaviour by setting the Exactness property. Note that the 'exact' calculation gives incorrect results if ties are present.

The runs Test

The runs test (also called Wald–Wolfowitz test) is a test of randomness. It compares the lengths of runs of the same value in a sample to what would be expected in a random sample. In numerical data, it uses the runs of values that are above or below a cut point. The test relies only on the sequence of runs of the same value. It does not rely on any properties of the distributions.

The null hypothesis is that the samples were drawn randomly. The alternative hypothesis is that the samples were not drawn randomly. For a one-tailed test, the alternative hypothesis is that the samples tend to occur in groups (lower tailed) or that the samples tend to alternate (upper tailed).

The runs test is implemented by the Runs Test class. It has three constructors. The first constructor takes no arguments. All test parameters must be provided by setting the properties of the Mann Whitney Test object. The samples the test is to be applied to must be specified by setting the sample 1 and Sample2 properties. The second constructor takes two vector arguments objects that represent the samples the test is to be applied to.

The third constructor also takes two arguments. The first is once again a Vector< T> that contains the observations from the two samples combined. The second argument must be a Categorical Vector< T> with two levels that specifies which sample the observations in the first variable belong to.