Hypothesis Testing

Asking all sorts of questions

- Do the daily number of homicides in India follow a Poisson distribution?
- Has the unemployment rate in India changed in the last quarter?
- Does taking Moderna's COVID-19 vaccine help in reducing the chances of contracting COVID-19?
- Does the Higgs Boson exist?

What is a Hypothesis?

A hypothesis can be defined as a proposed explanation for a phenomenon. It is not the absolute truth but a provisional working assumption.

In statistics, a hypothesis is considered to be a particular assumption about a set of parameters of a population distribution.

It is called a hypothesis because it is not known whether or not it is true.

What is a Hypothesis Test?

A hypothesis test is a standard procedure for testing a claim about a property of a population.

Rare Event Rule for Inferential Statistics

If, under a given assumption, the probability of a particular observed event is exceptionally small, we conclude that the assumption is probably not correct.

Example

ProCare Industries, Ltd., once provided a product called “Gender Choice,” which, according to advertising claims, allowed couples to “increase your chances of having a boy up to 85%, a girl up to 80%.” Gender Choice was available in blue packages for couples wanting a baby boy and (you guessed it) pink packages for couples wanting a baby girl. Suppose we conduct an experiment with 100 couples who want to have baby girls, and they all follow the Gender Choice “easy-to-use in-home system” described in the pink package. For the purpose of testing the claim of an increased likelihood for girls, we will assume that Gender Choice has no effect.

Using common sense and no formal statistical methods, what should we conclude about the assumption of no effect from Gender Choice if 100 couples using Gender Choice have 100 babies consisting of a) 52 girls?; b) 97 girls?

Example - Continued

We normally expect around 50 girls in 100 births. The result of 52 girls is close to 50, so we should not conclude that the Gender Choice product is effective. If the 100 couples used no special method of gender selection, the result of 52 girls could easily occur by chance. The assumption of no effect from Gender Choice appears to be correct. There isn't sufficient evidence to say that Gender Choice is effective.

The result of 97 girls in 100 births is extremely unlikely to occur by chance. We could explain the occurrence of 97 girls in one of two ways : Either an extremely rare event has occurred by chance, or Gender Choice is effective. The extremely low probability of getting 97 girls is strong evidence against the assumption that Gender Choice has no effect. It does appear to be effective.

Components of a formal hypothesis test

- Given a claim, identify the null hypothesis and the alternative hypothesis, and express them both in symbolic form.
- Given a claim and sample data, calculate the value of the test statistic.
- Given a significance level, identify the critical value(s).
- Given a value of the test statistic, identify the P-value.
- State the conclusion of a hypothesis test in simple, non-technical terms.

Null Hypothesis : H_o

Based on the questions we listed, the null hypotheses might be :

- The daily number of homicides in India do follow a Poisson distribution.
- The unemployment rate in India has remained unchanged over the last quarter.
- Moderna's COVID-19 vaccine does not reduce the chances of contracting COVID-19
- The Higgs Boson does not exist.

Null Hypothesis : H_o

- The null hypothesis (denoted by H_o) is a statement that the value of a population parameter (such as proportion, mean, or standard deviation) is equal to some claimed value.

- We test the null hypothesis directly.

- Either reject H_o or fail to reject H_o.

- The null hypothesis is what we are willing to assume is the case until proven otherwise. We can never claim that the null hypothesis has been actually proved.

Alternative Hypothesis : H_A

- The alternative hypothesis (denoted by H_1 or H_a or H_A) is the statement that the statistic has a value that somehow differs from the null hypothesis.

- The symbolic form of the alternative hypothesis must use one of these symbols: $\neq$, $<$, $>$.

How to form your claim or hypothesis?

If you are conducting a study and want to use a hypothesis test to support your claim, the claim must be worded so that it becomes the alternative hypothesis.

How to form your claim or hypothesis?

Step 1 : Identify the specific claim or hypothesis to be tested and express it in symbolic form

Step 2 : Give the symbolic form that must be true when the original claim is false

Step 3 : Of the two symbolic expressions obtained so far, let the alternative hypothesis H_A be the one not containing equality so that H_A uses the symbol $<$ or $>$ or $\neq$. Let the null hypothesis H_o be the symbolic expression that the statistic equals the fixed value being considered

Identify the null and alternative hypothesis

The proportion of drivers who admit to running red lights is greater than 0.5.

Step 1 : We express the given claim as $p > 0.5$.

Step 2 : We see that if $p > 0.5$ is false, then $p \leq 0.5$ must be true.

Step 3 : We let the alternative hypothesis H_A be $p > 0.5$, and we let H_o be $p = 0.5$.

Identify the null and alternative hypothesis

The mean height of professional basketball players is at most 7 ft.

Step 1 : We express "a mean of at most 7 ft" in symbols as $\mu \leq 7$.
Step 2 : We see that if $\mu \leq 7$ is false, then $\mu > 7$ must be true.
Step 3 : We let the alternative hypothesis H_A be $\mu > 7$, and we let H_0 be $\mu = 7$.

Identify the null and alternative hypothesis

The standard deviation of IQ scores of actors is equal to 15.

Step 1 : We express the given claim as = 15.
Step 2 : We see that if σ = 15 is false, then σ ≠ 15 must be true.
Step 3 : We let the alternative hypothesis H_A be σ ≠ 15 and we let H_0 be σ = 15.

Test Statistic

The test statistic is a value used in making a decision about the null hypothesis, and is found by converting the sample statistic to a score with the assumption that the null hypothesis is true.

Test Statistic - Formula

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Test statistic for proportions

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

Test statistic for mean

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Test statistic for variance

Test Statistic - Example

Problem : A survey of n = 880 randomly selected adult drivers showed that 56% (or p = 0.56) of those respondents admitted to running red lights. Find the value of the test statistic for the claim that the majority of all adult drivers admit to running red lights.

The preceding example showed that the given claim results in the following null and alternative hypotheses: H_0 : p = 0.5 and H_A : p > 0.5. Because we work under the assumption that the null hypothesis is true with p = 0.5, we get the following test statistic:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.56 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{880}}} = 3.56$$

Significance Level

The significance level (denoted by $\boldsymbol{\alpha}$) defines how much evidence we require to reject H_0 in favor of H_A

Critical Region

The critical region (or rejection region) is the set of all values of the test statistic that cause us to reject the null hypothesis.

Critical Value

A critical value is any value that separates the critical region (where we reject the null hypothesis) from the values of the test statistic that do not lead to rejection of the null hypothesis. The critical values depend on the nature of the null hypothesis, the sampling distribution that applies, and the significance level $\boldsymbol{\alpha}$.

Two-tailed, Right-tailed, Left-tailed Tests

The tails in a distribution are the extreme regions bounded by critical values.

- Two-tailed test : $H_0 : =$, $H_A : \neq$
- Right tailed test : $H_0 : =$, $H_A : >$
- Left tailed test : $H_0 : =$, $H_A : <$

P-value

The P-value (or p-value or probability value) is the probability of getting a value of the test statistic that is at least as extreme as the one representing the sample data, assuming that the null hypothesis is true. The null hypothesis is rejected if the P-value is very small, such as 0.05 or less.

If a P-value is small enough, then we say the results are statistically significant

Conclusions in Hypothesis Testing based on P-value

We always test the null hypothesis. The initial conclusion will always be one of the following:

- ❑ Reject the null hypothesis - if the P-value ≤ α (where α is the significance level, such as 0.05).
- ❑ Fail to reject the null hypothesis - if the P-value > α

Example 1

An article distributed by the Associated Press included these results from a nationwide survey: Of 880 randomly selected drivers, 56% admitted that they run red lights. The claim is that the majority of all Americans run red lights. That is, $p > 0.5$. The sample data are $n = 880$, and $p = 0.56$.

$H_0 : p = 0.5$, $H_A : p > 0.5$, $\alpha = 0.05$

We see that for values of $z = 3.50$ and higher, we use 0.9999 for the cumulative area to the left of the test statistic.

The P-value is $1 - 0.9999 = 0.0001$. Since the P-value of 0.0001 is less than the significance level of = 0.05, we reject the null hypothesis. There is sufficient evidence to support the claim.

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.56 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{880}}} = 3.56$$

Example 2

We have a sample of 106 body temperatures having a mean of 98.20°F. Assume that the sample is a simple random sample and that the population standard deviation is known to be 0.62°F. Use a 0.05 significance level to test the common belief that the mean body temperature of healthy adults is equal to 98.6°F.

$H_0 : \mu = 98.6$, $H_A : \mu \neq 98.6$, $\alpha = 0.05$, $\bar{x} = 98.2$, $\sigma = 0.62$

This is a two-tailed test and the test statistic is to the left of the center, so the P-value is twice the area to the left of $z = -6.64$. Using norm.cdf , the area to the left of $z = -6.64$ is 0.0001, so the P-value is 2(0.0001) = 0.0002.

Because the P-value of 0.0002 is less than the significance level of = 0.05, we reject the null hypothesis. There is sufficient evidence to conclude that the mean body temperature of healthy adults differs from 98.6°F.

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}} = \frac{98.2 - 98.6}{\frac{0.62}{\sqrt{106}}} = -6.64$$

Type - I error

A Type I error is the mistake of rejecting the null hypothesis when it is true.

The symbol $\boldsymbol{\alpha}$ (alpha) is used to represent the probability of a type I error.

Type - II error

A Type II error is the mistake of failing to reject the null hypothesis when it is false.

The symbol β (beta) is used to represent the probability of a type II error.

Type I and Type II errors

	True State of Nature	
	The null hypothesis is true	The null hypothesis is false
Decision — We decide to reject the null hypothesis	Type I error (rejecting a null hypothesis when it is true)	Correct Decision
Decision — We fail to reject the null hypothesis	Correct Decision	Type II error (failing to reject a false null hypothesis)

Power of a hypothesis test

The power of a hypothesis test is the probability ($1 - \beta$) of rejecting a false null hypothesis, which is computed by using a particular significance level $\boldsymbol{\alpha}$ and a particular value of the population parameter that is an alternative to the value assumed true in the null hypothesis. That is, the power of the hypothesis test is the probability of supporting an alternative hypothesis that is true.