# 1 Forewords

Suppose you wanted to determine whether the mean level of a driver's blood alcohol exceeds the legal limit after two drinks, or whether the majority of registered voters approve of the president's performance. In both cases, you are interested in **making an inference** about how the value of a population relates to a specific numerical value. Is it less than, equal to, or greater than the specified number?
This type of inference, called a **Statistical Test of Hypothesis**.

## 1.1 What is a hypothesis?

**Definition 1** *Hypothesis means a conjecture–an assertion. A statistical hypothesis is a statistical statement about the numerical value of a population parameter.*

A test is a rule by which we would like to verify on the hypothesis.

**Example** Suppose building specifications in a certain city require that the average breaking strength of residential sewer pipe be more than 2400 pounds per foot of length (i.e., per linear foot). Each manufacturer who wants to sell pipe in that city must demonstrate that its product meets the specification (see this is the verification). Note that we are interested in throwing some lights about the mean $m$ of a population (What is population here?). However, in this example we are less interested in estimating the value of mean than we are in testing a hypothesis about its value that is, we want to decide whether the mean breaking strength of the pipe exceeds 2400 pounds per linear foot.

## 1.2 Types of hypothesis

We define two hypotheses:
(1) **The null hypothesis** represents the status quo to the party performing the sampling experiment on the hypothesis that will be assumed to be true (which has to be verified/nullified) unless the data provide convincing evidence that it is false. Statistacally, it is formulated by $H_0$.
(2) **The alternative, or research** hypothesis is the theory (statement) that contradicts the null hypothesis. This usually represents the values of a population parameter for which the researcher wants to gather evidence to support. Statistically it is formulated by $H_1$ or $H_a$.

Get back to Example, try to figure out the null and alternative hypothesis.
The null hypothesis is that the manufacturer's pipe does not meet specifications unless the tests provide convincing evidence otherwise.

Null hypothesis ($H_0$): mean$\leq$ 2400
(i.e., the manufacturer's pipe does not meet specifications)
Alternative (research) hypothesis ($H_a$): mean $\geq$ 2400
(i.e., the manufacturer's pipe meets specifications)

Now you can assign a notation of mean (For normality concept we denote mean by $\mu$). The above testing of hypothesis problem can be rewritten that

$$H_0 : \mu \leq 2400$$
$$H_a : \mu \geq 2400$$

**Remember** A hypothesis can always be made on a population characteristic(parameter). You need to verify the hypothesis on the basis of a sample collected.

Now look at the form of the alternative hypothesis and the null hypothesis. One can think of three formats of hypothesis,viz. $=, \geq, \leq$ (equal to, less than, greater than). See, when we consider equality the form of hypothesis is exact.

**Definition 2** ***Simple Hypothesis***: *It refers to the one in which all parameters associated with the distribution are stated. For instance, if the height of the students in a school is distributed normally with $\sigma^2 = 6$ and mean $\mu$(unknown) and the hypothesis that the mean stands equivalent to 70 implying $H_0 : \mu = 70$. So simple hypothesis completely specifies the population's probability distribution.*

**Definition 3** *When we consider inequality in the formulation of hypothesis the structure of hypothesis will be different which is named as Composite hypothesis. It refers to the hypothesis that does not stand to be simple. A composite hypothesis does not completely specify the probability distribution. For instance, if the height of the students in a school is distributed normally with $\sigma^2 = 6$ and mean $\mu$(unknown) and the hypothesis that the mean stands more than 70 cm implying $H_0 : \mu > 70$.*

A null hypothesis can be simple/composite, so is an alternative hypothesis.

**Remember** Generally, null hypothesis is taken as simple and alternative is taken as composite.

**Remember** An alternative might be a combination of two composite hypotheses,like for the example of height on students $H_a : \mu > 70 \, or \, \mu < 50$ or just $H_a : \mu \neq 70$.

**Definition 4** *A* ***one-tailed test*** *of hypothesis is one in which the alternative hypothesis is directional and includes the symbol $>$ or $<$.*
*A* ***two-tailed test*** *of hypothesis is one in which the alternative hypothesis does* **not** *specify departure from $H_0$ in a particular direction and is written with the symbol $\neq$.*
*Since null hypothesis is generally taken as simple(= always), so one tailed/two tailed test concept does not work on null hypothesis. Alternative hypothesis , if it is composite, might be one tailed or two tailed.*

Example: For $H_0 : \mu = 70$, alternative may be

$$H_a : \mu > 70 \,\text{one-tailed test/right tailed}$$
$$H_a : \mu < 70 \,\text{one-tailed test/left tailed}$$
$$H_a : \mu \neq 70 \,\text{two-tailed test}$$

Thus if the parametric space is denoted by $\Omega$, then $H_0$ and $H_1$ are the **disjoint** subset of $\Omega$.,i.e., $\Omega = H_0 \cup H_1$.

## 1.3 A worked out problem

A metal lathe is checked periodically by quality control inspectors to determine whether it is producing machine bearings with a mean diameter of .5 inch. If the mean diameter of the bearings is larger or smaller than .5 inch, then the process is out of control and must be adjusted. Formulate the null and alternative hypotheses for a test to determine whether the bearing production process is out of control.

**Solution**: The hypotheses must be stated in terms of a population parameter. Here, we define $\mu$ as the true mean diameter (in inches) of all bearings produced by the metal lathe. If either $> .5$ or $< .5$, then the lathe's production process is out of control. Because the inspectors want to be able to detect either possibility (indicating that the process is in need of adjustment), these values of m represent the alternative (or research) hypothesis. Alternatively, because $\mu = .5$ represents an in-control process (the status quo), this represents the null hypothesis. Therefore, we want to conduct the two-tailed test:

$$H_0 : \mu = .5 \text{ i.e., the process is in control } H_a : \mu \neq .5 \text{ i.e., the process is out of control}$$

## 1.4 Some Exercises

1. Suppose instead that we wanted to see if girls scored significantly different than the national average score on the verbal section of the GRE, and suppose that national average was 500. Popular concept is that girls do score significantly better than the national average. Formulate null hypothesis and alternative hypothesis.
   If the popular belief is girls do significantly different than the national average, what will the null hypothesis and alternative hypothesis.
   Also state whether alternative hypothesis right tailed/left tailed/two tailed.

2. Cigarette advertisements are required by federal law to carry the following statement: "Warning: The surgeon general has determined that cigarette smoking is dangerous to your health." However, this warning is often located in inconspicuous corners of the advertisements and printed in small type. Suppose the Federal Trade Commission (FTC) claims that 80% of cigarette consumers fail to see the warning. A marketer for a large tobacco firm wants to gather evidence to show that the FTC's claim is too high, i.e., that fewer than 80% of cigarette consumers fail to see the warning.
   What is the parameter of interest here? Is it mean?
   Specify the null and alternative hypotheses for a test of the FTC's claim.

3. The effect of drugs and alcohol on the nervous system has been the subject of considerable research. Suppose a research neurologist is testing the effect of a drug on response time by injecting 100 rats with a unit dose of the drug, subjecting each rat to a neurological stimulus, and recording its response time. The neurologist knows that the mean response time for rats not injected with the drug (the "control" mean) is 1.2 seconds. She wishes to test whether the mean response time for drug-injected rats differs from 1.2 seconds. Construct $H_0$ and $H_1$.

4. The US Education Survey of Online Learning, "Going the Distance: Online Education in the United States, 2011," reported that 68% of college presidents believe that their online education courses are as good as or superior to courses that utilize traditional face-to-face instruction. Give the null hypothesis for testing the claim made by this Survey.

5. In a study investigating a link between walking and improved health (Social Science & Medicine, Apr. 2014), researchers reported that adults walked an average of 5.5 days in the past month for the purpose of health or recreation. Specify the null and alternative hypotheses for testing whether the true average number of days in the past month that adults walked for the purpose of health or recreation is lower than 5.5 days.

6. In order to determine the melting point (in degrees Celsius) of a mercury compound, an analyst recorded the readings obtained during his experiment. The analyst claimed that the average melting point of the compound was 150°C.
   Give the null and alternative hypotheses for testing whether the average melting point is less than 150°C.

## 1.5 Concept on test statistic

Get back to the breaking strength example of sewer pipe. The question is: how can the city decide when enough evidence exists to conclude that the manufacturer's pipe meets the specifications?
Because the hypotheses concern the value of the population mean $\mu$, it is reasonable to use the sample mean $\overline{x}$ to make the inference. The city will conclude that the pipe meets specifications only when the sample mean $\overline{x}$ **convincingly** indicates that the population mean exceeds 2,400 pounds per linear foo (the alternative hypothesis).

"Convincing" evidence in favor of the alternative hypothesis will exist when the value of $\overline{x}$ exceeds 2,400 by an amount that cannot be readily attributed to sampling variability. So basically we need to device one rule or function, based on sample mean (more specifically, based on sample observations,$x_1, x_2, \cdots, x_n$.) for checking the convincing evidence.

**Definition 5** ***Test Statistic:*** *The test statistic is a sample statistic, computed from information provided in the sample, that the researcher uses to decide between the null and alternative hypotheses.*

In this example, test statistic would be as a function of the sample mean ($\overline{x}$) (Remember test statistic can be framed on any sample function, `not only on` sample mean).
Therefore while fixing null and alternative keep the following points in mind.

| $H_a$ | $H_0$ |
|---|---|
| Hypothesis as main purpose of study<br>Hypothesis that you want to claim to be right | Hypothesis as opposed to purpose<br>Hypothesis that you want claim to be wrong |

### 1.6 Critical region and use of test statistic on decision making

Suppose $(x_1, x_2, x_3, \cdots, x_n)$ be a sample available with the investigator. Mean of $(x_1, x_2 \cdots x_n) = \overline{x}$. Thus the nature of $\mathbf{x}$ will dictate whether $H_0$ is to be considered valid or not. We may say that for making the decision the whole sample space (which may be taken as $R^n$ and denoted by $X$). So, $X$ is partitioned into two measurable sets, say $W$ and $A$,i.e. $X = W \bigcup A$. If $x \in A$ we shall have reason to doubt the validity of $H_0$. In former case $H_0$ is said to be **accepted**; in the latter it is said to be **rejected**. The set $W$ is called the rejection region or critical region for $H_0$, while its complement, $A$, is called the **acceptance region**.

**Remember:** Acceptance of $H_0$ does not mean that $H_0$ has been proved to be true; it means only that so far as the given observations,$\mathbf{x}$, are concerned $H_0$ is valid. Also, rejection of $H_0$ means $H_0$ does not look plausible in the light of the given observations $\mathbf{x}$.

We may thus say that a test for $H_0$ is a rule for rejecting/accepting the hypothesis according as $\mathbf{x}$. Further, to each of $H_0$ there corresponds a unique critical region and conversely, to each critical region there corresponds a unique test.

### 1.7 Errors in Testing of Hypothesis

Since the decision rule for the accepting the null is made on the sample $\mathbf{X}$, errors will arise.

| | innocent defendent | guilty defendent |
|---|---|---|
| Evidence showing guilty | Incorrect verdict | correct verdict |
| Evidence showing innocent | correct verdict | incorrect verdict |

Table 1: A Court trial for a crime

You can see in the above example, there are two incorrect judgments:

1. Convict an innocent person
2. Acquit a criminal.

As similar to the Judge's example in testing of hypothesis too, two types of errors are committed. These two

| Based on $\mathbf{x} \in X$ | $H_0$ true | $H_0$ false |
|---|---|---|
| $H_0$ rejected | incorrect<br>**type I error** | correct |
| $H_0$ accepted<br>or failed to reject | correct | incorrect<br>**type II error** |

error play the key roles in testing of hypothesis theory.

### 1.7.1 Type I error

Type I error, also known as a "false positive": the error of rejecting a null hypothesis when it is actually true. In other words, this is the error of accepting an alternative hypothesis (the real hypothesis of interest) when the results can be attributed to chance. So the probability of making a type I error in a test with rejection region $W$ is $\alpha = P(\mathbf{x} \in W/H_0)$. Probability of type I error is called **size** of a test.

### 1.7.2 Type II error

Type II error, also known as a "false negative": the error of not rejecting a null hypothesis when the alternative hypothesis is the true state of nature. In other words, this is the error of failing to accept an alternative hypothesis when you don't have adequate power. So the probability of making a type II error in a test with acceptance region $A$ is $\beta = P(\mathbf{x} \in A/H_1) = 1 - P(\mathbf{x} \in W/H_1)$.

Thus the question is to address probabilities of both of the errors. But both of the errors can not be minimized simultaneously as one belongs to the complementary set to that of the other.The conventional method is to fix probability of type one error and minimizing the probability of the second one (Why the type I error?). Think a while, type I error is more influential than type II error (Look at Judge's trial example).

**Remember** Generally, probability of type I error, $\alpha$ is kept fixed at 5% or 1% level,i.e.there is only a 5 or 1 in 100 chance that the variation that we are seeing is due to chance. This fixation is called the **level of significance**.

**Remember:** Although **size** and **level of significance** bear the same mathematical logic. But conceptually, they are different. Level of significance is a fixed prob of type I error while size of a specific test varies with sample to sample.

## 1.8 Examples on Type I error and Type II error

1. A large nationwide poll recently showed an unemployment rate of 10 % in US. The mayor of a local town doubts if this figure holds true in his town. So he plans of taking a sample of his residents to see if the unemployment rate is significantly different than 10% in his town. The hypothesis she uses is $H_0 : p = 0.1$ against $H_1 : p \neq 0.1$. Under which of the following conditions the mayor would commit Type I error? Also strike out the correct conclusions.

   (a) He concludes the town's unemployment rate is not 10% when it actually is.

   (b) He concludes the town's unemployment rate is not 10% when it actually is not.

   (c) He concludes the town's unemployment rate is 10% when it actually is.

   (d) He concludes the town's unemployment rate is 10% when it actually is not.

2. A large university is curious if they should build another cafeteria. They plan to survey a sample of their students to see if there is strong evidence that the proportion interested in a meal planis higher than 40% in which case they will consider building a new cafeteria. The hypothesis is $H_0 : p = .4$ against $H_0 : p > .4$. What would be the consequence of type II error in this problem?

   (a) They don't consider building a new cafeteria when they should.

   (b) They don't consider building a new cafeteria when they should not.

   (c) They consider building a new cafeteria when they should not.

   (d) They consider building a new cafeteria when they should.

## 1.9 Power of a test

Think on probability of type II error. It says on the chance of accepting a false null hypothesis. An ideal test (decision rule) would be the one which can identify a false hypothesis and reject it with high probability. This quality of a test is called power of a test.

**Definition 6** *Probability of rejecting a false null hypothesis is called power of a test. Power=* $Pr(\mathbf{x} \in W/H_1)$*. Intuitively, Power=*$1-$*Prob[type II error]=* $1 - \beta$*.*

For different values of the parameters under $H_1$, we have power function. Power closer to 1 indicates a potential test rule. So maximizing the power function would be a decision criterion in testing of hypothesis.

**Definition 7** ***Most Powerful test:*** *The critical region $W$ is the most powerful (MP) critical region of size $\alpha$ (and a corresponding test a most powerful test of level $\alpha$) for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ if for any other size $\alpha$ critical region $W_1$,*

$$P(\mathbf{x} \in W/H_1) \geq P(\mathbf{x} \in W_1/H_1).$$

Now take up the case of testing a simple null hypothesis against composite alternative hypothesis, i.e. $H_0 : \theta = \theta_0$ against $H_a : \theta > (<)\theta_0$ at the level of significance $\alpha$.

**Definition 8** *The critical region for testing $H_0 : \theta = \theta_0$ against $H_a : \theta > (<)\theta_0$ is called the uniformly most powerful (UMP) critical region of size $\alpha$ if*

$$P(\mathbf{x} \in W/H_0) = \alpha$$
$$P(\mathbf{x} \in W/H_1) \geq P(\mathbf{x} \in W_1/H_1) \forall\theta \; \textit{where}\, W_1 \; \textit{other critical region of size}\, \alpha$$

## 1.10 Steps in solving testing of hypothesis problem

1. find out the parameter of interest from the question
2. setting up $H_0$ and $H_1$ in terms of the parameter of interest
3. Find the suitable test statistic $t = t(x_1, x_2, ..x_n)$, based on which you take the decision
4. Partition the set of possible values of the test statistic $t$ into two disjoint sets $W$(critical region) and $A$ and frame the following test:
   - Reject $H_0$ if the value of $t$ falls $W$
   - Accept $H_0$ if the value of $t$ falls in $A$.
5. After framing the above test, obtain experimental sample observations, compute the appropriate test statistic

Example 1 Let X follow the following p.d.f.

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}, & 0\le x\le\theta\\ 0 & \text{o.w.}\end{cases}$$

You are testing $H_0: \theta=1$ against $H_1: \theta=2$ by means of a single observation. What would be the size of type I and p(type II error), if you choose the interval

(i) $\cdot 5 \le x$

(ii) $1 \le x \le 1\cdot 5$ as the critical regions. Which is better test?

Obtain power for each of the case.

Answer

(i) $W = \{X : x \geq \cdot 5\}$.

$\alpha = Pr\{\text{type I error}\} = \text{size} = Pr\{X\in W / H_0\}$

$= Pr\{X \geq \cdot 5 \mid \theta=1\}$

$= Pr\{\cdot 5 \le x \le \theta \mid \theta=1\}$

$= \int_{\cdot 5}^{1} 1\, dx = \boxed{\cdot 5}$ [see when $\theta=1$ $f(x;\theta)=1, 0\le x\le 1$]

similarly, $\beta = Pr\{X\in A / H_1\}$

$= Pr\{X < \cdot 5 \mid \theta=2\}$

$= \int_{0}^{0\cdot 5} \frac{1}{2}\, dx = \boxed{0\cdot 25}$

power $= P\{x\in W/H_1\} = 1-\underset{(\beta)}{\cdot 25} = \boxed{\cdot 75}$ [you can find power by direct integration Also]

(ii) $W = \{X : 1 \le X \le 1\cdot 5\}$

$\alpha = Pr\{\text{type I error}\} = \int_{1}^{1\cdot 5} 1\cdot dx = \boxed{0}$ since under $H_0$, $f(x)=0$ for $1\le x\le 1\cdot 5$.

$\beta = Pr\{X\in A \mid \theta=2\}$ when $\theta=2$ $f(x)=\frac{1}{2}$ for $0\le X\le 2$

$= \int_{0}^{1}\frac{1}{2}dx + \int_{1\cdot 5}^{2}\frac{1}{2}dx.$

$= \frac{1}{2} + \frac{\cdot 5}{2} = \cdot 5 + \cdot 25 = \boxed{\cdot 75}$

Power $= 1-\beta = 1-\cdot 75 = \boxed{\cdot 25}$

(i) gives the better test as the power is higher.

Example 2

If $X \geq 1$ be critical region for testing $H_0: \theta = 2$ against the $H_1: \theta = 1$ on the basis of the single observation from the population

$$f(x;\theta) = \theta e^{-\theta x}, \quad 0 \leq x < \infty.$$

obtain $\alpha$ and $\beta$.

Answer

$$\alpha = Pr\{X \in W / H_0\}.$$
$$= Pr\{X \geq 1 / \theta = 2\}$$
$$= \int_1^{\infty} 2e^{-2x} dx = 2\left[\frac{e^{-2x}}{-2}\right]_1^{\infty} = \frac{1}{e^2}$$

$$\beta = P\{x \in A / H_1\}$$
$$= P\{X < 1 / \theta = 1\}$$
$$= \int_0^1 e^{-x} dx = \frac{e^{-x}}{-1}\Big|_0^1 = (1 - e^{-1}) = \frac{e-1}{e}.$$

Example 3 Let $p$ be the probability that a coin will fall head in a single toss in order to test $H_0: p = \frac{1}{2}$ against $H_1: p = 3/4$. The coin is tossed 5 times and $H_0$ is rejected if more than 3 heads are obtained. Find probability of type I error and that of type II error.

Answer. $H_0: p = \frac{1}{2}$ ag. $H_1: p = 3/4$

The random variable $x$ denotes the number of heads in $n$. tosses of a coin then so that.

$$X \sim Bin(n, p)$$
$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \binom{5}{x} p^x (1-p)^{5-x}$$

The critical region is $W = \{x : x \geq 4\}$

$$A = \{x : x \leq 3\}$$

$$\alpha = P[X \geq 4 / H_0] = P(X = 4 / p = \tfrac{1}{2}) + P(X = 5 / p = \tfrac{1}{2})$$
$$= 3/16$$

$$\beta = P[X \in A / H_0] = P(X = 3 / p = \tfrac{3}{4}) + P(X = 2 / p = \tfrac{3}{4})$$
$$+ P(X = 1 / p = 3/4)$$
$$+ P(X = 0 / p = \tfrac{3}{4})$$

$= \frac{47}{128}.$

Power of the test $= 1-\beta = \frac{81}{128}.$

H.W. An urn contains 6 marbles of which $\theta$ are white and others black. In order to test the null hypothesis $H_0: \theta = 3$ against the alternative $H_1: \theta = 4$, two marbles are drawn at random (without replacement) and $H_0$ is rejected if both the marbles are white, otherwise $H_0$ is accepted. Find the probabilities of committing type I and type II errors.

H.W. It is desired to test the hypothesis $H_0: \theta = 0$ against $H_1: \theta > 0$, by observing a random variable $x$ which is uniformly distributed on $[\theta, \theta+1]$. Given only one observation sketch the power function of the test whose critical region is defined by $[x > c]$. What value of $c$ would you choose?

H.W. $f(x;\theta) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta. \\ 0 & \text{otherwise} \end{cases}$

Find the $\alpha$ and $\beta$ for $H_0: \theta = 1.5$ against $H_1: \theta = 2.5$ if the rejection region is $.8 \le x$.