CHAPTERS 10. HYPOTHESIS TESTING
Some examples of hypothesis testing

1. Toss a coin 100 times and get 62 heads. Is this coin a fair coin?

2. Is the new treatment on blood pressure more effective than the old one?


<table>
<thead>
<tr>
<th>Months to Promotion</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The data is re-organized in terms of ranks, from shortest to longest.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Months to Promotion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13</td>
</tr>
<tr>
<td>Sex</td>
<td>M M M M M M M M M M M M</td>
</tr>
<tr>
<td>Rank</td>
<td>14 15 16 17 18 19 20 21 22 23 24 25 26</td>
</tr>
<tr>
<td>Sex</td>
<td>M M M M M M M M M M M F F M</td>
</tr>
</tbody>
</table>
Basic setup of hypothesis testing

Population parameters of interest $\theta$ (unknown). Samples collected from experiment or observation $\{X_1, X_2, \ldots, X_n\}$.

Hypothesis Testing.

1. Null Hypothesis and Alternative Hypothesis.

$$H_0 : \theta \in \Theta_0, \quad H_a : \theta \in \Theta_a.$$  

For example

$H_0 : \theta = 0.5, \quad H_a : \theta \neq 0.5$

$H_0 : \theta = 0.5, \quad H_a : \theta > 0.5$

$H_0 : \theta < 1, \quad H_a : \theta > 2$

2. Test statistics — a function of the samples, say $T = T(X_1, X_2, \ldots, X_n)$.

3. Rejection region (RR).

(a) When $T \in RR$, reject $H_0$ and accept $H_a$.

(b) When $T \not\in RR$, accept $H_0$. 
**Type I error, Type II error and Power of a test**

Consider the simple hypotheses

\[ H_0 : \theta = \theta_0, \quad H_a : \theta = \theta_1, \]

where \( \theta_0, \theta_1 \) are given constants.

1. **Type I error.**

\[ \alpha \doteq P(\text{Reject } H_0 | H_0 \text{ is true}) = P(T \in RR | \theta = \theta_0) \]

2. **Type II error.**

\[ \beta \doteq P(\text{Accept } H_0 | H_a \text{ is true}) = P(T \notin RR | \theta = \theta_1) \]

3. **Power.**

\[ P(\text{Reject } H_0 | H_a \text{ is true}) = 1 - \beta. \]
Example

Consider the following hypothesis testing. $X_1, X_2, \ldots, X_n$ are iid from $N(\mu, 1)$.

$$H_0 : \mu = 0, \quad H_a : \mu = 1.$$ 

Suppose $T = \bar{X}$ and the rejection region is

$$RR \doteq \{ x : x > 0.5 \}.$$ 

1. Type I error:

$$\alpha = P(\bar{X} > 0.5|\mu = 0) = P(\sqrt{n}\bar{X} > 0.5\sqrt{n}) = \Phi(-0.5\sqrt{n}).$$

2. Type II error:

$$\beta = P(\bar{X} \leq 0.5|\mu = 1) = P(\sqrt{n}[\bar{X} - 1] \leq -0.5\sqrt{n}) = \Phi(-0.5\sqrt{n}).$$

3. Power:

$$1 - \beta = 1 - \Phi(-0.5\sqrt{n}) = \Phi(0.5\sqrt{n}).$$
Remark:

1. The ideal scenario is that both $\alpha$ and $\beta$ are small. But $\alpha$ and $\beta$ are in conflict.

2. Increasing the sample size will reduce both $\alpha$ and $\beta$, and increase the power of the test.

3. Type I error is more often called the significance level of the test.

4. When $\theta_0$ and $\theta_1$ are closer, the power of the test will decrease.

5. Usually we are looking for sufficient evidence to reject $H_0$. Thus type I error is more important than the type II error. Consequently, one usually control the type I error below some pre-assigned small threshold, and then, subject to this control, look for a test which maximize the power (or minimize the type II error).
**Remark:** All the previous definitions and discussions extend to composite hypotheses

\[ H_0 : \theta \in \Theta_0, \quad H_a : \theta \in \Theta_a, \]

where

1. Type I error: for \( \theta_0 \in \Theta_0, \)
   \[ \alpha(\theta_0) = P(T \in RR|\theta = \theta_0) \]
2. Type II error: for \( \theta_a \in \Theta_a, \)
   \[ \beta(\theta_a) = P(T \notin RR|\theta = \theta_a) \]
3. \[ \text{Power} = 1 - \beta(\theta_a). \]
Testing the mean of normal distribution

Suppose $X_1, X_2, \ldots, X_n$ are iid from $N(\mu, \sigma^2)$ with $\sigma^2$ known but $\mu$ unknown. Consider the following types of one-sided tests and two-sided test.

[1]. $H_0 : \mu = \mu_0, \quad H_a : \mu > \mu_0$

[2]. $H_0 : \mu = \mu_0, \quad H_a : \mu < \mu_0$

[3]. $H_0 : \mu = \mu_0, \quad H_a : \mu \neq \mu_0$

In all three cases, the test statistics is

$$T = \bar{X} = \frac{1}{n} (X_1 + X_2 + \cdots + X_n).$$

We also assume that the type I error (significance level) is fixed to be a pre-assigned small number $\alpha$ (usually $\alpha = 0.05$).
[1]. \( H_0 : \mu = \mu_0, \quad H_a : \mu > \mu_0 \)

The rejection region is of the form

\[ RR = \{ \bar{X} > k \} \]

for some \( k \).

**Determine \( k \).**

\[ \alpha = \text{Type I error} = P(\bar{X} > k|\mu = \mu_0). \]

But \( \bar{X} \) is \( N(\mu, \sigma^2/n) \). Therefore

\[ k = \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha = \mu_0 + \sigma \bar{X} z_\alpha. \]
\[ H_0 : \mu = \mu_0, \quad H_a : \mu < \mu_0 \]

The rejection region is of the form

\[ RR = \{ \bar{X} < k \} \]

for some \( k \).

**Determine \( k \).**

\[ \alpha = \text{Type I error} = P(\bar{X} < k|\mu = \mu_0). \]

\[ k = \mu_0 - \frac{\sigma}{\sqrt{n}}z_\alpha = \mu_0 - \sigma \bar{X}z_\alpha. \]
[3]. \( H_0 : \mu = \mu_0, \quad H_a : \mu \neq \mu_0 \)

The rejection region is of the form
\[
RR = \{ \bar{X} < k_1 \} \cup \{ \bar{X} > k_2 \}
\]
for some \( k \).

**Determine \( k \).**
\[
\alpha = \text{Type I error} = P(\bar{X} < k_1 | \mu = \mu_0) + P(\bar{X} > k_2 | \mu = \mu_0).
\]

Symmetry.
\[
k_1 = \mu_0 - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} = \mu_0 - \sigma \bar{X} z_{\alpha/2}
\]
\[
k_2 = \mu_0 + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} = \mu_0 + \sigma \bar{X} z_{\alpha/2}.
\]
**Example**

1. National student exam scores are distributed as $N(500, 100^2)$. In a classroom of 25 freshmen, the mean score was 472. Is the freshmen of below average performance? (Consider different cases with the significance level $\alpha = 0.1, 0.05, 0.01$)

Reject $H_0$ at level $\alpha = 0.1$,
Accept $H_0$ at level $\alpha = 0.05, 0.01$. 
2. In a two-sided test of $H_0 : \mu = 80$ in a normal population with $\sigma = 15$, an investigator reported that “since $\bar{X} = 71.9$, the null hypothesis is rejected at 1% level.” What can we say about the sample size used?

\[ n \geq 23 \]
THE TEST FOR NORMAL DISTRIBUTIONS EASILY EXTEND TO LARGE SAMPLE TESTS WHERE THE TEST STATISTICS

\[ \hat{\theta} \text{ is approximately } N(\theta, \sigma_{\hat{\theta}}). \]

AND THE HYPOTHESES ARE

\[ H_0 : \theta = \theta_0, \quad H_a : \theta > \theta_0 \]
\[ H_0 : \theta = \theta_0, \quad H_a : \theta < \theta_0 \]
\[ H_0 : \theta = \theta_0, \quad H_a : \theta \neq \theta_0 \]
Examples

In all these examples, the significance level is assumed to be $\alpha = 0.05$.

1. Toss coin 100 times, and get 33 heads. Is this a fair coin?
2. Do indoor cats live longer than wild cats?

<table>
<thead>
<tr>
<th>Cats</th>
<th>Sample size</th>
<th>Mean age</th>
<th>Sample Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indoor</td>
<td>64</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>Wild</td>
<td>36</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>
3. In order to test if there is any significant difference between opinions of males and females on abortion, independent random samples of 100 males and 150 females were taken.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Sample size</th>
<th>Favor</th>
<th>Oppose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>100</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>Female</td>
<td>150</td>
<td>95</td>
<td>55</td>
</tr>
</tbody>
</table>
**P-value**

**P-value:** The probability of getting an outcome as extreme or more extreme than the actually observed data (under the assumption that the null hypothesis is true).

**Remark:** Given a significance level $\alpha$,

1. If P-value $\leq \alpha$, reject the null hypothesis.
2. If P-value $> \alpha$, accept the null hypothesis.
Redo all the previous examples to find P-value.
SAMPLE SIZE

Suppose the population distribution is $N(\mu, \sigma^2)$ with $\sigma^2$ known. Consider the test

$$H_0 : \mu = \mu_0, \quad H_a : \mu > \mu_0.$$ 

Pick a sample size so that the type I error is bounded by $\alpha$ and the type II error is bounded by $\beta$ when $\mu = \mu_a$.

$$n \geq \left[ \frac{(z_\alpha + z_\beta)\sigma}{\mu_a - \mu_0} \right]^2$$
Remark: The same argument extends to large sample testing. In particular, the binomial setting.

Suppose $X_1, X_2, \ldots, X_n$ are iid Bernoulli with $P(X_i = 1) = p = 1 - P(X_i = 0)$. Consider the test

$$H_0 : p = p_0, \quad H_a : p > p_0.$$ 

Pick a sample size so that the type I error is bounded by $\alpha$ and the type II error is bounded by $\beta$ when $p = p_a$.

$$n \geq \left[ \frac{z_\alpha \sqrt{p_0(1 - p_0)} + z_\beta \sqrt{p_a(1 - p_a)}}{p_a - p_0} \right]^2$$
Examples

1. Suppose it is required to test population mean

\[ H_0 : \mu = 5, \quad H_a : \mu > 5 \]

at level \( \alpha = 0.05 \) such that type II error is at most 0.05 when true \( \mu = 6 \). How large should the sample be when \( \sigma = 4 \).
2. How many tosses of a coin should be made in order to test

\[ H_0 : p = 0.5, \quad H_a : p > 0.5 \]

at level \( \alpha = 0.5 \) and when true \( p = 0.6 \) type II error is 0.1?
Suppose \( \{X_1, X_2, \ldots, X_n\} \) are iid samples with common density \( f(x; \theta) \). Consider the following simple hypotheses.

\[
H_0 : \theta = \theta_0, \quad H_a : \theta = \theta_a.
\]

**Question:** Among all the possible rejection regions \( RR \) such that the type I error satisfies

\[
P(RR|\theta = \theta_0) \leq \alpha
\]

with \( \alpha \) pre-specified, which \( RR \) gives the maximal power (or minimal type II error)?
**Neyman-Pearson Lemma.**

Define for each $k$

$$RR_k = \left\{ (x_1, x_2, \ldots, x_n) : \frac{f(x_1, \theta_0)f(x_2, \theta_0) \cdots f(x_n, \theta_0)}{f(x_1, \theta_0)f(x_2, \theta_a) \cdots f(x_n, \theta_a)} \leq k \right\}.$$ 

Suppose there is a $k^*$ such that

$$P(RR_{k^*} | \theta = \theta_0) = \alpha,$$

then $RR_{k^*}$ attains the maximal power among all tests whose type I error are bounded by $\alpha$. 

1. Consider the following test for density $f$.

$$H_0 : f(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{elsewhere} \end{cases}, \quad H_a : f(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0, & \text{elsewhere} \end{cases}.$$ 

Find the most powerful test at significance level $\alpha$ based on a single observation.
2. Suppose $X_1, X_2, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ with $\sigma^2$ known. We wish to test

$$H_0 : \mu = 0, \quad H_\alpha : \mu = \theta \quad (\theta < 0)$$

Find the most powerful test at significance level $\alpha$. 

Remark: What can we say about the test

\[ H_0 : \mu = 0, \quad H_a : \mu < 0. \]

Remark: What can we say about the test

\[ H_0 : \mu = 0, \quad H_a : \mu \neq 0. \]
3. Let $X$ has density

$$f(x, \theta) = \begin{cases} 2\theta x + 2(1 - \theta)(1 - x), & 0 < x < 1, \\ 0, & \text{elsewhere} \end{cases}$$

Consider test

$$H_0 : \theta = 0, \quad H_a : \theta = 1$$

with significance level $\alpha$. 
Likelihood ratio test

The general form of likelihood ratio test

\[ H_0 : \theta \in \Theta_0, \quad H_a : \theta \in \Theta_a \]

- The test statistics

\[
\lambda = \frac{\max_{\theta \in \Theta_0} L(x_1, x_2, \ldots, x_n; \theta)}{\max_{\theta \in \Theta} L(x_1, x_2, \ldots, x_n; \theta)}
\]

where \( \Theta = \Theta_0 \cup \Theta_a \).

- Rejection region \( \{ \lambda \leq k \} \) for some \( k \).

Remark: \( \Theta_0 \) and \( \Theta_a \) may contain nuisance parameters. And \( 0 \leq \lambda \leq 1 \).
EXEMPLARY

1. Suppose $Y_1, Y_2, \ldots, Y_n$ are iid samples from Bernoulli with parameter $p$.

$$H_0 : p = p_0, \quad H_a : p > p_0.$$
2. Suppose $Y_1, Y_2, \ldots, Y_n$ are iid samples from $N(\mu, \sigma^2)$. $\mu$ and $\sigma^2$ are both unknown. We want to test

$$H_0 : \mu = \mu_0, \quad H_a : \mu > \mu_0.$$ 

Find the appropriate likelihood ratio test.
**Large sample distribution of $\lambda$**

**Theorem:** When $n$ is large, the distribution $-2\ln(\lambda)$ under $H_0$ is approximately $\chi^2$ with degree of freedom equal

$$\text{number of free parameters in } \Theta - \text{number of free parameters in } \Theta_0.$$

The Rejection region with significance level $\alpha$ is just

$$\{ -2\ln(\lambda) \geq \chi^2_\alpha \}.$$