Hypothesis testing

Terminologies:

**NULL HYPOTHESIS** (usually denoted by \( H_0 \))
- "prior belief"

**ALTERNATIVE HYPOTHESIS** (denoted by \( H_a \))
- "alternative belief"

Typical situation: Collect data, use the data to decide whether \( H_0 \) should be rejected.

1. A **TEST STATISTIC** is a function of the sample data on which the decision is to be based.

2. A **REJECTION REGION** is the set of all test statistic values for which \( H_0 \) will be rejected.

A **TYPE I ERROR** - rejecting the null hypothesis \( H_0 \) when it is true

A **TYPE II ERROR** - not rejecting \( H_0 \) when \( H_0 \) is false.
Example pg 324 no. 9

Two companies

\( p = \) proportion of all potential subscribers who favor the 1st company over the 2nd

\[ H_0 : \quad p = 0.5 \]
\[ H_a : \quad p \neq 0.5 \]

\( X = \text{# of, in a random sample of size 25 individuals, individuals who favor the 1st company} \]
\[ x = \text{an observed value of } X \]

(a) Rejection region \( R = \{ x : x \leq 7 \text{ or } x \geq 18 \} \)

(b) Distribution of \( X \) when \( H_a \) is true (i.e. \( p = 0.5 \))

\[ P[X = x] = \left( \frac{25}{x} \right) \left( \frac{1}{2} \right)^x \left( 1 - \frac{1}{2} \right)^{25-x} \]

\[ P[\text{Type I error}] = ? = P[X \leq 7 \text{ or } X \geq 18] \]

\[ = \sum_{x=0}^{7} b(x; 25, \frac{1}{2}) + (1 - \sum_{x=0}^{17} b(x; 25, \frac{1}{2})) \]

\[ \text{Table A.1} \]
\[ 0.0220 + 1 - 0.978 = 0.044 \]

pg 738
\[ d. \: \beta(p) = P[ \text{Type II error when the true } p \text{ is } p = 0.5] \]

Note: Distribution of \( X \) when \( p = 0.3 \) is
\[
P[X=x] = \binom{25}{x} (0.3)^x (1-0.3)^{25-x}
\]

\[ \beta(0.3) = P[\text{we fail to reject } H_0 \text{ when } p = 0.3] \]
\[
= P[8 \leq X \leq 17] = P[X \leq 17] - P[X \leq 7]
\]
\[
= \sum_{x=0}^{17} b(x; 25, 0.3) - \sum_{x=0}^{7} b(x; 25, 0.3)
\]

\[
\begin{array}{c}
\text{Table A.1} \\
\hline
1.0000 \\
- 0.512 \\
\hline
= 0.488
\end{array}
\]
Tests about a population mean (section 8.2)

3 cases (as in the study of CI)

(I) normal population with known $\sigma$

(II) not necessarily normal, but $n$ large

(III) normal population, $\sigma$ not known, $n$ not necessarily large

(II) $N(\mu, \sigma) - \sigma$ known

- want to decide if we should reject the null hypothesis $H_0: \mu = \mu_0$ or not.
- we have a random sample $X_1, X_2, \ldots, X_n \sim i.i.d. N(\mu, \sigma)$

Recall: if $\mu$ is really $\mu_0$, then

$$\bar{X} = (X_1 + \cdots + X_n)/n \sim N(\mu_0, \sigma^2/n)$$

or $$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$
One sample $z$ test

**Null Hypothesis** $H_0 : \mu = \mu_0$

**Test statistic value:** $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

**Alt. hypothesis**
- $H_a : \mu > \mu_0$
  - $z > z_\alpha$
- $H_a : \mu < \mu_0$
  - $z < -z_\alpha$
- $H_a : \mu \neq \mu_0$
  - $z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$

**Rejection region for level $\alpha$ test**

Q: What's the rationale behind these?

A: The rejection regions are chosen in such a way that

$$P[\text{Type I error}] = \alpha$$

Note: $\alpha \downarrow \Rightarrow$ rejection region smaller

(makes sense, right?)
What about Type II error?

Say, if \( \mu \) is not \( \mu_0 \) but is \( \mu' \) \((\mu > \mu_0)\)

(\text{so we should reject } H_0 - \text{ assume } H_a: \mu > \mu_0 \)

What is \( P[\text{type II error}] \)?

\[
\frac{X - \mu_0}{\sigma/\sqrt{n}} < z_\alpha
\]

\( = P[\mu = \mu' \neq \mu_0 \text{ but the level } \alpha \text{ test fails to reject } H_0 \]

\( = \text{accepts} \)

\[
\frac{X - \mu'}{\sigma/\sqrt{n}} \sim N(0,1)
\]

\( = \Phi(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}) \)

\( \text{Note: } \mu' \uparrow \Rightarrow P[\text{type II error}] \downarrow \)

\( \sigma \uparrow \Rightarrow P[\text{type II error}] \uparrow \)

\( n \uparrow \Rightarrow P[\text{type II error}] \downarrow \)

How large should \( n \) be such that \( P[\text{type II error}] = \beta \)?

\[
Solve: \quad \Phi(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}) = \beta
\]

\( \text{or } -z_\beta = z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} \quad \text{or } n = \left[ \frac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'} \right]^2 \)
Sample size \( n \) for which a level \( \alpha \) test also has \( \beta(\mu') = \beta \) at the alternative value \( \mu' \) is

\[
\begin{align*}
\frac{1}{n} &= \begin{cases} 
\left[\frac{\sigma(\bar{X} + z\beta)}{\mu_0 - \mu'}\right]^2 & (\text{Ha: } \mu > \mu_0) \\
\left[\frac{\sigma(\bar{X} - z\beta)}{\mu_0 - \mu'}\right]^2 & (\text{Ha: } \mu < \mu_0) \\
\left[\frac{\sigma(\bar{X})}{\mu_0 - \mu'}\right]^2 & (\text{Ha: } \mu \neq \mu_0)
\end{cases}
\end{align*}
\]

Go through example 8.6 and 8.7 in class.
(iv) \( n \) large (underlying population distribution) needs not be normal

Recall \( \frac{\bar{X} - \mu}{S/\sqrt{n}} \) is approximately \( N(0, 1) \)

So can use the tests in case (ii) with the appropriate changes

(basically just change \( \sigma \) to \( S = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} \).)

Read example 8.8
Recall \[ T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \] has a t-distribution with \((n-1)\) d.o.f.

The one-sample t-test

Null hypothesis: \( H_0 : \mu = \mu_0 \)

Test statistic value: \( t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \)

Alternative hypothesis

Rejection region for α level test

<table>
<thead>
<tr>
<th>( H_a )</th>
<th>Test Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu &gt; \mu_0 )</td>
<td>( t \geq t_{\alpha, n-1} )</td>
</tr>
<tr>
<td>( \mu &lt; \mu_0 )</td>
<td>( t \leq -t_{\alpha, n-1} )</td>
</tr>
<tr>
<td>( \mu \neq \mu_0 )</td>
<td>( t \geq t_{\alpha/2, n-1} ) or ( -t_{\alpha/2, n-1} )</td>
</tr>
</tbody>
</table>

Same reasoning behind the z-test:

\[ \text{e.g.} \quad P[\text{type I error}] = \alpha \Rightarrow \text{the “tolerance”} \]

\[ H_a : \mu \neq \mu_0 \quad P[T \in \text{rejection region when } \mu = \mu_0] \]

So if we choose the rejection region to be ;
then we get what we want!

Determining $\beta$ in this case is more complicated. (Can you see why?)

See p. 335

Read Example 8.9
Tests concerning a population proportion (Sec 8.3)

Hypothesis testing:

\[ H_0: p = 0.4 = p_0 \leq \text{null value} \]

\[ H_a: p > 0.4 \]

\[ \hat{p} = \frac{ \text{count of } 'A' \text{ events} }{ \text{total events} } \]

id. Bernoulli r.v.'s \( x_1, \ldots, x_n \)

\[ \hat{p} = \frac{ \sum x_i }{ n } \]

\[ \text{VAR}[\hat{p}] = \frac{ p(1-p) }{ n } \]

Test statistic:

\[ \frac{ \hat{p} - p_0 }{ \sqrt{ p_0 (1-p_0) / n } } \]

(approximately \( N(0,1) \))

Null hypothesis: \( H_0: p = p_0 \)

Test statistic value: \( z = \frac{ \hat{p} - p_0 }{ \sqrt{ p_0 (1-p_0) / n } } \)

Alternative hypothesis

<table>
<thead>
<tr>
<th>Ha</th>
<th>rejection region (level ( \alpha ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p &gt; p_0 )</td>
<td>( z \geq z_\alpha )</td>
</tr>
<tr>
<td>( p &lt; p_0 )</td>
<td>( z \leq -z_\alpha )</td>
</tr>
</tbody>
</table>
| \( p \neq p_0 \) | \( z \geq z_{\alpha/2} \) or \( z \leq -z_{\alpha/2} \)

(Note: test relies on CLT, use only when

\[ n p_0 \geq 10 \quad \text{and} \quad n (1-p_0) \geq 10 \]

\[ \text{[P[Type I error] } \approx \alpha \]
Example 8.11

Hypothesis testing with $d = 0.1$

$H_0 : \ p = 0.3$

$H_a : \ p > 0.3$

$4115 \ (0.3) > 10$ , $4115 \ (0.7) > 10$ (large sample test) can be used

test statistic value

$z = (\hat{p} - 0.3) \left( \frac{\sqrt{0.3(0.7)}}{n} \right)$

$= \left( \frac{1276}{4115} - 0.3 \right) \left/ \sqrt{0.3(0.7)/4115} \right.$

$= 1.40.$

$z_{0.1} = 1.28$

since $1.40 > z_{0.1}$, the test suggests to reject $H_0$

$z_{0.08} = 1.4$

Q: what if $\alpha = 0.08$ , $z_{0.08} = \frac{\hat{p} - 0.92}{\sqrt{0.3(0.7)/4115}} = 1.4 \ldots$

Since $1.40 < z_{0.08}$, the test does not reject $H_0$
and sample size determination

What if $H_0$ is not true, i.e.

$$P = P' \neq P_0$$

test statistic

$$Z = \frac{\hat{P} - P_0}{\sqrt{P_0(1-P_0)/n}}$$

is no longer $N(0,1)$

Just like before, in order to calculate PE type II errors, we need to first understand the distribution of $Z$ when the true $P$ is $P'$.

$$\hat{P} = \frac{\prod_{i=1}^{n} x_i}{n}$$

$\chi_i$ iid Bernoulli ($P'$)

when $n$ large

so $\hat{P}$ is approximately $N(\hat{P}', \frac{P'(1-P')}{n})$

and $Z$ is approximately $N(\frac{P' - P_0}{\sqrt{P_0(1-P_0)/n}}, \frac{P'(1-P')}{P_0(1-P_p)})$

(Exercise: why?)
Alt. Hypothesis

\[ \text{Ha: } p > p_0 \]
\[ \Phi \left( \frac{p - p'}{\sqrt{p' (1-p') / n}} \right) \]

\[ \text{Ha: } p < p_0 \]
\[ 1 - \Phi \left( \frac{p_0 - p' - \frac{2\alpha}{\sqrt{p' (1-p') n}}} {\sqrt{p' (1-p') / n}} \right) \]

\[ \text{Ha: } p \neq p_0 \]
\[ \Phi \left[ \frac{p_0 - p' + \frac{2\alpha}{\sqrt{p' (1-p') n}}} {\sqrt{p' (1-p') / n}} \right] - \Phi \left[ \frac{p_0 - p' - \frac{2\alpha}{\sqrt{p' (1-p') n}}} {\sqrt{p' (1-p') / n}} \right] \]

(Exercise: derive these yourself)

If \( \alpha, \beta \) are prespecified, minimum sample size to achieve these:

\[
 n = \begin{cases} 
 \frac{\left[2 \alpha \sqrt{p_0 (1-p_0)} + 2 \beta \sqrt{p' (1-p')} \right]^2}{p' - p_0} & \text{one tailed test} \\
 \frac{\left[2 \alpha \sqrt{p_0 (1-p_0)} + 2 \beta \sqrt{p' (1-p')} \right]^2}{p' - p_0} & \text{two tailed test}
\end{cases}
\]
Example 8.12 (Sample exam question!)

A delivery company advertises

"at least 90% of packages are delivered on time"

\[ p = \text{true proportion of packages delivered on time} \]

\[ p = 0.8 \quad \text{[only the company would know this]} \]

[ secret, of course.]

---

**Q:** Based on a level 0.01 test with \( n = 225 \) packages, how likely can the test successfully tell that the company is lying in the ad.?

\[ H_0 : p = 0.9 \quad \text{(the company's claim)} \]

\[ H_a : p < 0.9 \]

\[ \beta (0.8) = P \left[ \text{level 0.01 test [fails] to reject } H_0 \right] \]

\[ = 1 - \Phi \left[ \frac{0.9 - 0.8 - \frac{2.33}{0.01} \sqrt{(0.9)(0.1)/225}} {\sqrt{(0.8)(0.2)/225}} \right] \]

\[ = 1 - \Phi (2.00) = 0.0228 \]

So \( P \left[ \text{level 0.01 test with } n = 225 \text{ can reject } H_0 \right] = 1 - \beta (0.8) \]

\[ = 0.9772. \]
If we want $\alpha = \beta = 0.01$

$2\alpha > 2\beta = 2.23$

$n = \left[ \frac{2.23 \sqrt{(0.93)(0.1)}}{0.8 - 0.9} + 2.23 \sqrt{(0.8)(0.2)} \right]^2$

$\approx 266.$

\underline{Small sample test for $p$ :}

use binomial distribution

(instead of a CLT approx.)

\underline{see pg 342}
P-values (section 8.4)

Consider the following scenario:

- $\mu = 10$ = true average time to initial relief of pain of a best-selling pain reliever

- the company considers introducing a newly developed reliever $\mu = \text{average time for the new pain reliever}$

$H_0 : \mu = 10 \quad H_a : \mu < 10$

Company tells customers

"We reject $H_0$ (i.e. new reliever is better) based on a hypothesis testing analysis with $\alpha = 0.10$.

A customer satisfied with the best-selling reliever would view a Type I error serious, he/she (contemplating a switch to the new reliever) would question the validity of the company’s claim; in particular, he/she may want to test the hypothesis with $\alpha = 0.05, 0.01$, or even lower.
Problem: The company's statement prevents an individual decision maker from reaching a conclusion at such a level (of \( \alpha \)).

Related, may be the data gathered by the company problem gives a \( z \) value that barely falls into the rejection region. \( \alpha = 0.1 \), rejection region

\( H_0 \) is rejected (using \( \alpha = 0.1 \)), but very marginally, and the customers do not know it from the statement.

Can the company say something to the public which is more informative?

There is a nice solution to this, all you need is to understand a simple fact:

[Read Ex. 8.5 and 8.15]
Example

\[ H_0 : \mu = 1.5 \quad \text{(or known in this case) \quad \sigma = 0.20} \]

\[ H_a : \mu > 1.5 \]

Test statistic:

\[ \frac{\bar{x} - 1.5}{(0.2)/\sqrt{32}} \]

\[ = 2.10 \]

\[ \bar{x} = \frac{x_1 + \cdots + x_{32}}{32} = 1.5742 \]

<table>
<thead>
<tr>
<th>Level of significance ( \alpha )</th>
<th>Rejection region</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>( z \geq 1.645 )</td>
<td>reject ( H_0 )</td>
</tr>
<tr>
<td>0.025</td>
<td>( z \geq 1.96 )</td>
<td>reject ( H_0 )</td>
</tr>
<tr>
<td>0.01</td>
<td>( z \geq 2.33 )</td>
<td>do not reject ( H_0 )</td>
</tr>
<tr>
<td>0.005</td>
<td>( z \geq 2.58 )</td>
<td>do not reject ( H_0 )</td>
</tr>
</tbody>
</table>

**KEY OBSERVATION:** There is a critical \( \alpha \) value, \( \alpha^* \), at which we switch from reject to do not reject, called “P-value”. The p-value is the smallest level at which \( H_0 \) can be rejected.

\( \alpha \geq \text{P-value} \Rightarrow \text{reject } H_0 \text{ at level } \alpha \)

\( \alpha < \text{P-value} \Rightarrow \text{do not reject } H_0 \text{ at level } \alpha \)
For a $z$-test ($z$ = test statistic value)

$$
P = \begin{cases} 
1 - \Phi(z) & \text{for an upper-tailed test} \\
\Phi(z) & \text{for a lower-tailed test} \\
2(1 - \Phi(|z|)) & \text{for a two-tailed test}
\end{cases}
$$

Read Ex 8.17

For a $t$-test, ($\Phi$ cdf of $N(0,1)$)

- Change the $\Phi( )$ above to the c.d.f. of the $t$-distribution with $(n-1)$ d.o.f. ($n$ = sample size.)

Read Ex 8.18