TESTS OF STATISTICAL HYPOTHESES

Definitions :

- Statistical hypothesis : An assertion or a conjecture about the distribution of one or more random var.s.

A simple hypothesis completely specifies the distribution.
a composite hyp. does not.

$H_0 : \theta \le 100$
$H_1 : \theta > 100$ } Composite

$H_0 : \theta = 100$ } Simple

H_0 is called the null hypothesis. It shows the status quo. H_1 is the alternative hypothesis. It is usually an interesting contradiction of H_0.

Ex : A new system for launching rockets has been designed. In the old system, the prob. of success was 0.8.

$H_0 : p_{new} = 0.8$
$H_1 : p_{new} > 0.8$ Is the new system more successful?

A test is a rule which leads to the rejection or acceptance of the hyp. based on observing the values of a sample.

If the sample falls in the rejection region (or the critical region), we reject H_0. Otherwise we don't. In this case we say "we fail to reject H_0" or "we do not have sufficient evidence to reject H_0". This does not mean that we accept H_0 or that we believe that H_0 is true.

There are two types of error:

		The truth is H_0	H_a
Decision is	Reject H_0	Type I error	OK
	Do not reject H_0.	OK	Type II error

$\alpha = P(\text{Type I error}) = P(\text{reject } H_0 \mid H_0 \text{ true})$
$= P(\text{sample falls in rej. region (RR)} \mid H_0 \text{ true})$

$\beta = P(\text{Type II error}) = P(\text{do not reject } H_0 \mid H_1 \text{ true})$
$= P(\text{sample falls in } RR^c \mid H_1 \text{ true})$

What is the relationship b/w RR, α, β and n?

Ex 1: $Y_1, Y_2, \dots, Y_n \sim$ iid $\text{Bin}(1, p)$, let $n = 10$.

$H_0 : p = 0.2$
$H_1 : p = 0.3$

Let $RR_1 = \{\bar{Y} > 0.2\}$

Then, $\alpha = P(\bar{Y} > 0.2 \mid p = 0.2) = P(\Sigma Y_i > 0.2n \mid p = 0.2)$
$\Sigma Y_i \sim \text{Bin}(10, 0.2)$

$\alpha = P(\Sigma Y_i > 2) = 1 - P(\Sigma Y_i \leq 2) = 1 - 0.678 = 0.322$ High!

$\beta = P(\bar{Y} < 0.2 \mid p = 0.3) = P(\Sigma Y_i \leq 0.2n \mid p = 0.3)$
$= P(\Sigma Y_i \leq 2) = 0.383$ High!

from Binomial dist. tables.

Ex 1 (cont)

Now suppose $RR_2 = \{ \bar{Y} > 0.3 \}$

$$\alpha = P(\bar{Y} > 0.3 \mid p = 0.2) = P(\Sigma Y_i > 3) = 0.121$$

$$\beta = P(\bar{Y} \le 0.3 \mid p = 0.3) = P(\Sigma Y_i \le 3) = 0.650$$

$\Sigma Y_i \sim Bin(10, 0.3)$ $\quad$ $\Sigma Y_i \sim Bin(10, 0.2)$

Notice that when we switched from RR_1 to RR_2

$\alpha \downarrow$ but $\beta \uparrow$.

With n fixed, the cost of reducing α is increasing β or vice versa.

The only way to reduce both is to increase n. (Get more information)

Let $n = 25$ this time.

RR_2 : $\alpha = P(\bar{Y} > 0.3 \mid p = 0.2) = P(\Sigma Y_i > 7.5)$
$= 1 - 0.891 = 0.109$ $\quad$ $\Sigma Y_i \sim Bin(25, 0.2)$

$\beta = P(\bar{Y} \le 0.3 \mid p = 0.3) = P(\Sigma Y_i \le 7.5)$
$= 0.512$ $\quad$ $\Sigma Y_i \sim Bin(25, 0.3)$

Notice that as $n \uparrow$, both α and β decreases.

Definition :

The power function is the function, defined for all parameter values under consideration, that yields the probability of rejecting H_0. The value of the power function at a parameter point is called the power of the test at that point. Let's call the parameter of interest θ.

$$\text{Power}(\theta_1) = 1 - \beta(\theta_1) = 1 - P(\text{do not reject } H_0 \mid \theta = \theta_1)$$
$$= P(\text{reject } H_0 \mid \theta = \theta_1)$$

The power of the test at $\theta = \theta_1$

Q : What is the power of this test when H_0 is true? $(H_0 : \theta = \theta_0)$

A : $\text{Power}(\theta_0) = 1 - \beta(\theta_0)$

$= 1 - P(\text{do not reject } H_0 \mid \theta = \theta_0)$

$= P(\text{reject } H_0 \mid \theta = \theta_0) = \alpha$

α is called the <u>significance level</u> of the test.

In general, the significance level of a test is the maximum value of the power function of the test when H_0 is true.

Ex 2 : $n = 25$. $X_1, X_2, \dots, X_{25} \sim \text{iid } N(\theta, 100)$.

$H_0 : \theta \leq 75$

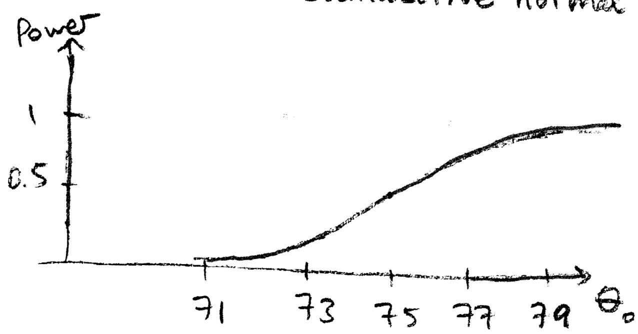
$H_1 : \theta > 75$

$RR_1 = \{\bar{X} > 75\}$

$\text{Power}(\theta_0) = P(\bar{X} > 75 \mid \theta = \theta_0)$ $\qquad \bar{X} \sim N(\theta, \underbrace{4}_{\sigma^2/n})$

$$= P\left(\frac{\bar{X} - \theta_0}{2} > \frac{75 - \theta_0}{2}\right) = 1 - \underbrace{\Phi}_{\text{cumulative normal distn.}}\left(\frac{75 - \theta_0}{2}\right)$$

Power(75) = 0.5
Power(73) = 0.159
Power(77) = 0.841
Power(79) = 0.977.



$\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$

$= P(\bar{X} > 75 \mid \theta = 75) = 0.5$

This is a very high prob. of Type I error!

Ex 2 (cont)

$RR_2 = \{\bar{X} > 78\}$

$$\text{Power}(\theta_0) = P(\bar{X} > 78 \mid \theta = \theta_0) = 1 - \Phi\left(\frac{78 - \theta_0}{2}\right)$$

Power(73) = 0.006

Power(75) = 0.067 → $\alpha = 0.067$, quite low.

So the prob. of rejecting when H_0 is true is quite low.

Power(77) = 0.309

Power(78) = 0.5

Power(79) = 0.691

However, the prob. of rejecting when H_0 is false (and $\theta = 77$) is only 0.309.

$\Rightarrow$ Prob. of not rejecting $H_0 : \theta \leq 75$ when in fact $\theta = 77$ is $1 - 0.309 = 0.691$.

$\beta(\theta = 77) = 0.691$. High Type II error!

$RR_3 = \{\bar{X} > c\}$, $\bar{X} \sim N(\theta, 100/n)$

Find c and n such that

Power(75) = 0.159 and Power(77) = 0.841.

$$\text{Power}(75) = P(\bar{X} > c \mid \theta = 75) = 1 - \Phi\left(\frac{c - 75}{10/\sqrt{n}}\right) = 0.159$$

So, we have

$$1 - \Phi\left(\frac{c-75}{10/\sqrt{n}}\right) = 0.159, \quad 1 - \Phi\left(\frac{c-77}{10/\sqrt{n}}\right) = 0.841$$

From tables, $\dfrac{c-75}{10/\sqrt{n}} = 1$, $\dfrac{c-77}{10/\sqrt{n}} = -1$ $\Big\}$ $n = 100$, $c = 76$

With these values, Power(73) = 0.001

Power(79) = 0.999

Ex 3 : a) $H_0 : \theta = 30000$ $\quad X \sim N(\theta, \sigma^2 = 5000^2)$

$H_1 : \theta > 30000$

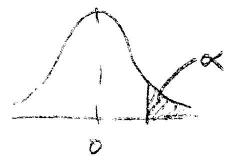
$$\alpha = P(\text{ reject } H_0 \mid H_0 \text{ true}) = P(\bar{X} > c \mid \theta = 30000)$$

$$= P\left(\frac{\bar{X} - \theta}{\sigma/\sqrt{n}} > \frac{c - 30{,}000}{\sigma/\sqrt{n}} = c_1 \right)$$

If we want a test with a significance level of $\alpha = 0.05$,

then $P(Z > c_1) = 0.05 \Rightarrow c_1 = 1.645$

$$\Rightarrow c = (1.645)\, \sigma/\sqrt{n} + 30000$$



(one-sided test)

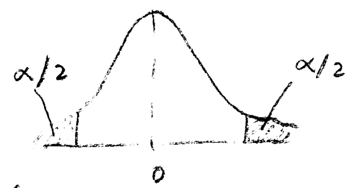
b) $H_0 : \theta = 30000$

$H_1 : \theta \neq 30000$ → reject H_0 if $\bar{X}$ is too small or too large.

$$\alpha = 0.05 = P\left(\left| \frac{\bar{X} - \overset{\to 30000}{\theta}}{\sigma/\sqrt{n}} \right| > 1.96 \right)$$

$\Rightarrow$ Reject if $\bar{X} > (1.96)\, \sigma/\sqrt{n} + 30000$ or

$\bar{X} < -(1.96)\, \sigma/\sqrt{n} + 30000$



(two-sided test)

Do not reject if $30000 - (1.96)\sigma/\sqrt{n} < \bar{X} < 30000 + (1.96)\, \sigma/\sqrt{n}$

But this looks similar to a 95% CI for μ of a normal distribution.

$$\theta - (1.96)\, \sigma/\sqrt{n} < \bar{X} < \theta + (1.96)\, \sigma/\sqrt{n}$$

$$\Rightarrow \quad \bar{X} - (1.96)\sigma/\sqrt{n} < \theta < \bar{X} + (1.96)\, \sigma/\sqrt{n}$$

** We can use the statistics that we used to construct CI's to test hypotheses.

Q : What is the "p-value" of a test ?

A : The p-value is the observed tail probability of a statistic being at least as extreme as the particular observed value when H_0 is true.

In other words, it is the smallest α value at which we can reject H_0.

Ex : Suppose we have a one-sided test for μ of a N distn. Suppose our critical (CR) region (rejection region) implies that we can reject if $z > 1.645$.

This means that we have chosen $\alpha = 0.05$, since $P(z > 1.645) = 0.05$.

Suppose we pick a random sample and compute $\bar{X}$.

What happens if $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = z = 1.75$? We reject H_0, with 5% signif. level.

We could have chosen $\alpha = 0.045$ here, since

$P(z > 1.70) = 0.045$. We still reject H_0, because 1.70 is in CR.

Is $\alpha = 0.045$ the smallest value of α that we can have with this sample ? No !

$P(z > 1.75) = 0.04$, so we can reduce α to 4%.

This is the smallest significance level that we can have with this sample.