Introduction to mathematical Statistics Lecture14

P-value & Power of the Test

Example 1.

Bayport High was chosen to participate in a new curriculum program. A year later, 86 Bayport sophomores who participated in this program was randomly selected to take a SAT-I math exam. The national average on this exam was 494 with a standard deviation of 124. The 86 students averaged a score of 528. Can it be claimed at the $\alpha = 0.05$ level that the new curriculum is effective? (Another question for later, if the true mean is 514, what is the power of the test at $\alpha = 0.05$?)

Solution

Test on one population mean μ

$$n = 86,\ \bar{x} = 528,\ \sigma = 124,\ \text{large sample}$$

$$\begin{cases} H_0 : \mu = 494 \leftarrow \mu_0 \\ H_a : \mu \succ 494 \end{cases}$$

Test statistic : $Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \overset{H_0}{\sim} N(0,1)$

$$z_0 = \frac{528 - 494}{124 / \sqrt{86}} = 2.54$$

At $\alpha = 0.05,\ z_0 = 2.54 \succ Z_{0.05} = 1.645$

Therefore, we reject H_0 and conclude the new curriculum is effective.

p-value

It is the probability that we observe something at least as extreme as the sample (or test statistic) observed, given that the null hypothesis is true.

Example 1.

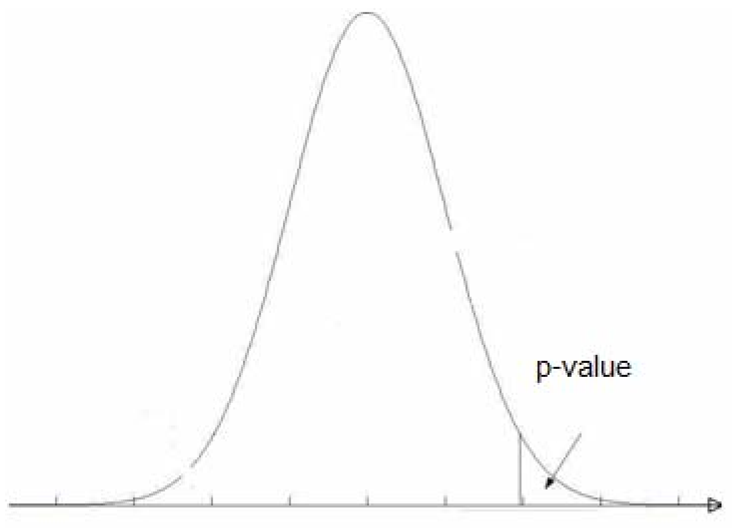
1. One-sided to the right

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \succ \mu_0 \end{cases}$$

Test statistic : $z_0 = 2.54$

$$\text{p-value} = P(Z_0 \geq z_0 \mid H_0 : \mu = \mu_0)$$
$$p-value = P(Z_0 \geq 2.54 \mid H_0 : \mu = \mu_0) \ \ (example)$$

Here $Z_0 \overset{H_0}{\sim} N(0,1)$



$z_0 = 2.54$

In summary, for $\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \succ \mu_0 \end{cases}$, the p-value for the z-test is the area under the N(0,1) curve to the right of $z_0 (= 2.54 \text{ in example})$.

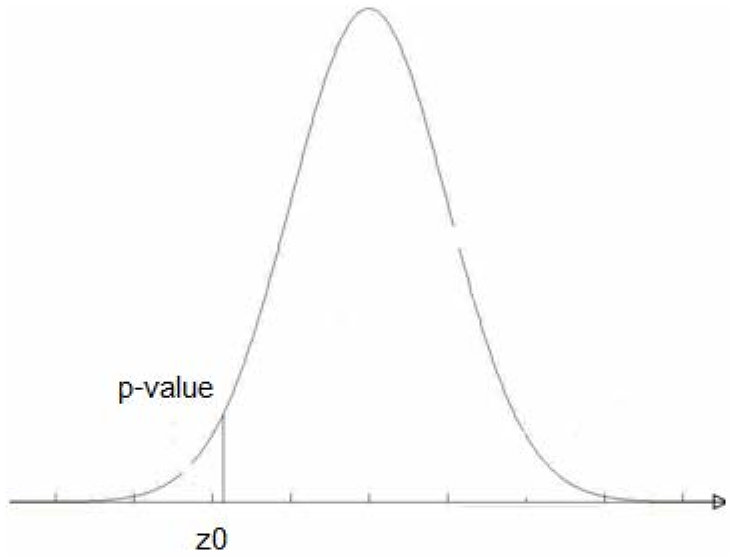
Make decision for your test using p-value. We reject H_0 in favor of H_a at the significance level α iff p-value $\prec \alpha$.

2. One-sided to the left

For $\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \prec \mu_0 \end{cases}$

Suppose we are using the z-test : $Z_0 = z_0 \leftarrow sample$

Then, $\text{p-value} = P(Z_0 \le z_0 \mid H_0)$



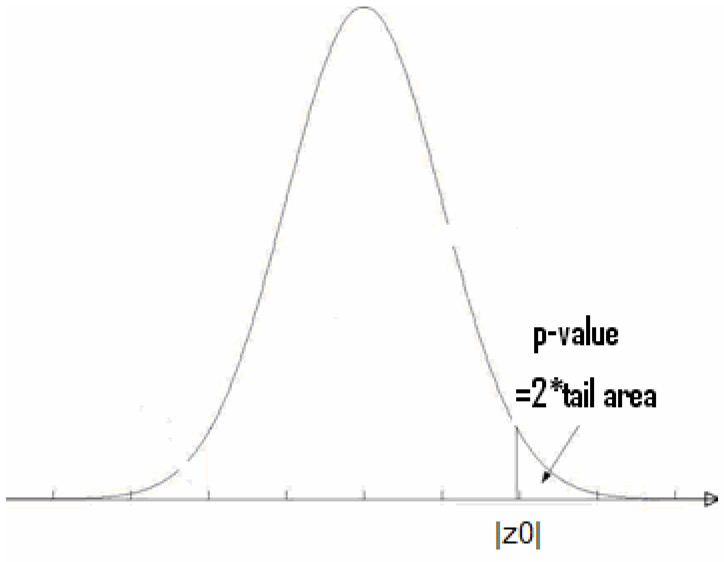
p-value is the area to the LEFT of z_0.

3. Two-sided

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \ne \mu_0 \end{cases}$$

Suppose that we are using the z-test.

$$\text{p-value} = P(|Z_0| \ge |z_0| \mid H_0) = 2 \times P(Z_0 \ge |z_0| \mid H_0)$$



$$\text{p-value} = 2 \times \text{the right tail area to } |z_0|$$

Power of the test

		Truth H_0	Truth H_a
Decision	H_0	**Correct**	Type II error
	H_a	Type I error	**Correct**

$$power = P(\text{Reject } H_0 \mid H_a)$$

Example 1. Bayport High example

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu = \mu_a \succ \mu_0 \end{cases}$$

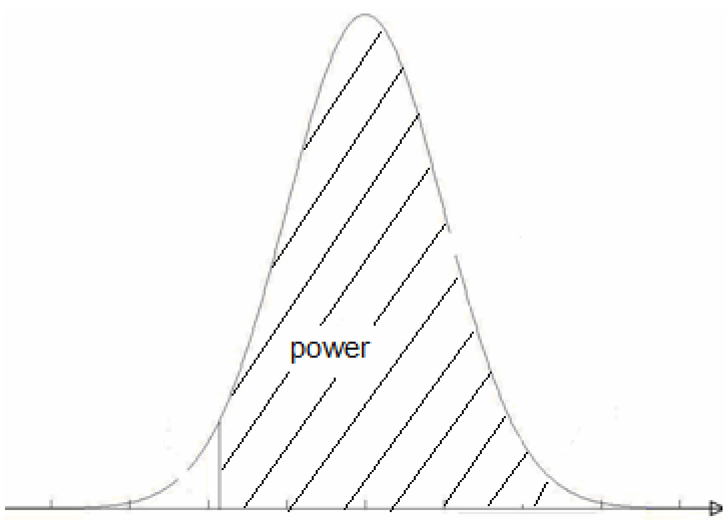
(μ_a is a specified value.)

z-test : $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \overset{H_0}{\sim} N(0,1)$

$$\begin{aligned} power &= P(\text{Reject } H_0 \mid H_a : \mu = \mu_a) \\ &= P(Z_0 \ge Z_\alpha \mid H_a : \mu = \mu_a) \end{aligned}$$

For example, $\alpha = 0.05$

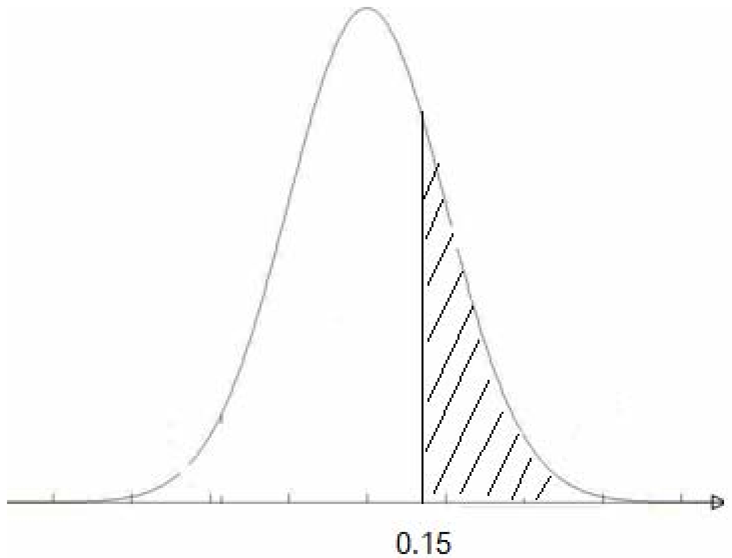
$$\begin{aligned} power &= P(\text{Reject } H_0 \mid H_a : \mu = \mu_a) \\ &= P(Z_0 \ge Z_\alpha \mid H_a : \mu = \mu_a) \\ &= P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \ge Z_\alpha \mid H_a : \mu = \mu_a\right) \\ &= P\left(\frac{\overline{X} - \mu_a}{\sigma/\sqrt{n}} + \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} \ge Z_\alpha \mid H_a : \mu = \mu_a\right) \\ &= P\left(Z \ge Z_\alpha - \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} \mid H_a : \mu = \mu_a\right) \end{aligned}$$



$$Z_\alpha - \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}}$$

In this example, $\sigma = 124, n = 86, \mu_a = 514, \mu_0 = 494$

$$power = P(Z \ge 1.64 - \frac{514-494}{124/\sqrt{86}} \mid H_a : \mu = 514)$$
$$= P(Z \ge 0.15 \mid H_a : \mu = 514)$$
$$= 0.44$$



Note: The power and p-value for the other two scenarios of the test on one population mean follows in exactly the same approach – For Scenario 1, normal population with population variance known, replace the approximate Z-test with the exact Z-test (no dot); for Scenario 3, normal population with population variance known, replace the approximate Z-test with the exact T-test (with n-1 degrees of freedom).

For:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \prec \mu_0 \end{cases}$$

$Power = P(reject\ H_0|H_a) = P(Z_0 \le -Z_\alpha|H_a)$

For

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_a : \mu \ne \mu_0 \end{cases}$$

$Power = P(reject\ H_0|H_a) = P(|Z_0| \ge Z_{\alpha/2}|H_a)$

Example 2.

John Pauzke, president of Cereal's Unlimited Inc, wants to be very certain that the mean weight μ of packages satisfies the package label weight of 16 ounces. The packages are filled by a machine that in set to fill each package to a specified weight. However, the machine has random variability measured by σ^2. John checked 1 box at random and found it to weight 17 oz. He is concerned the production line is filling more than 16 oz on the average. George Williams, quality control manager, advises him to examine a random sample of 25 packages of cereal. From his past experience, George knew that the weight of the packages follows a normal distribution with standard deviation 0.4 oz. At the significance level $\alpha = 0.05$,

(1) What is the decision rule (rejection region) in terms of the sample mean $\overline{X}$?
(2) What is the power of the test when $\mu =$ 16.2 oz?
(3) How many packages of cereal should be sampled if we wish to achieve a power of 85% when $\mu = 16.2$ oz?
(4) What sample size do we need to be 95% sure that the discrepancy between the sample mean and the population mean is within 0.5 oz?

Solution:

Inference on one population mean, the population is normal with σ =0.4 oz

n=25, α=0.05,

(1)

$$\begin{cases} H_0 : \mu = \mu_0 = 16 \\ H_a : \mu = \mu_a > 16 \end{cases} \Leftrightarrow \begin{cases} H_0 : \mu \le 16 \\ H_a : \mu > 16 \end{cases}$$

$$\text{Test Statistic : } Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \overset{H_0}{\sim} N(0,1)$$

We reject H_0 at $\alpha = 0.05$ iff

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq Z_\alpha \Rightarrow \overline{X} \geq \mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}} = 16 + 1.645 \times \frac{0.4}{\sqrt{25}} = 16.1316 \text{ (oz)}$$

(2)

Def: Power=P(reject** $H_0 \mid H_a$**)=1- P(Type II error)

Test Statistic : $Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \overset{H_0}{\sim} N(0,1)$

We reject H_0 at $\alpha = 0.05$ iff $Z_0 \geq Z_\alpha$

Power = $P(\text{Reject } H_0 \mid H_a)$

$$= P(Z_0 \geq Z_\alpha \mid \mu = \mu_a) \quad \textit{Here } Z_0 \overset{H_a}{\sim} N(\frac{\mu_a - \mu_0}{\sigma/\sqrt{n}}, 1)$$

$$= P(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq Z_\alpha \mid \mu = \mu_a)$$

$$= P(\frac{\overline{X} - \mu_a + \mu_a - \mu_0}{\sigma/\sqrt{n}} \geq Z_\alpha \mid \mu = \mu_a)$$

$$= P(\frac{\overline{X} - \mu_a}{\sigma/\sqrt{n}} \geq Z_\alpha - \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} \mid \mu = \mu_a)$$

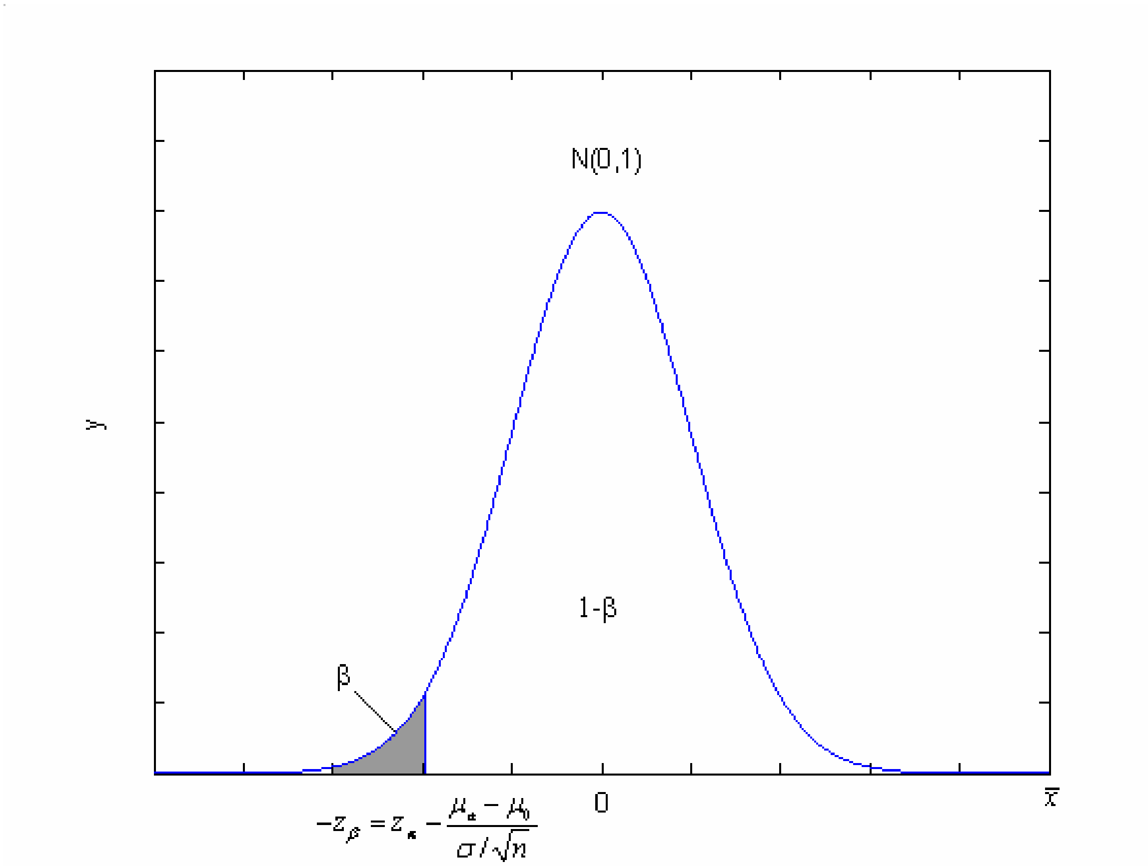
$$= P(Z \geq 1.645 - \frac{16.2 - 16}{0.4/\sqrt{25}} \mid \mu = 16.2) = P(Z \geq -0.855) \doteq 0.805$$

(3) Sample size calculation based on the Type I and Type II error rates (α and β)

Power = 1-$\beta = P(reject\ H_0 \mid H_a)$

$= P(Z_0 \geq Z_\alpha \mid \mu = \mu_a)$

$= P(Z \geq Z_\alpha - \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} \mid \mu = \mu_a)$



$\Rightarrow \quad \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} = Z_\alpha + Z_\beta \quad \Rightarrow \quad n = [\frac{(Z_\alpha + Z_\beta)\sigma}{\mu_a - \mu_0}]^2$

For the given example, we have $\alpha = 0.05, \beta = 1 - 0.85 = 0.15, \mu_0 = 16, \mu_a = 16.2, \sigma = 0.4$

$$\therefore \quad n = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2} = \frac{(1.645 + 1.04)^2 (0.4)^2}{(16.2 - 16)^2} \doteq 29$$

We will have a quiz this Friday on test of one-population mean. It will include all materials covered till now (including today's lecture.)