9.6. Tests about a population proportion:

The general setting is the following:

p: **the population proportion;**

p_0: **the particular hypothesized value.**

The statistical description for a testing problem is

$$H_0: p = p_0 \ \ vs \ \ H_a: p \neq p_0$$

or

$$H_0: p \geq p_0 \ \ vs \ \ H_a: p < p_0$$

or

$$H_0: p \leq p_0 \ \ vs \ \ H_a: p > p_0.$$

Motivating Example:

Objective: **Determine whether a special promotion has increased the proportion.**

The statistical description for this testing problem is

$$H_0: p \leq p_0 = 0.2 \ \ vs \ \ H_a: p > p_0 = 0.2$$

with

$$n = 400, \bar{p} = \frac{100}{400} = 0.25, \alpha = 0.05.$$

General Case: Large sample $(np \geq 5, n(1-p) \geq 5)$ and level of significance α

$$z = \frac{\bar{p} - p_0}{\sigma_{\bar{p}}} = \frac{\bar{p} - p_0}{\left(\sqrt{\frac{p_0(1-p_0)}{n}}\right)}.$$

1. $H_0: p \geq p_0 \ \ vs \ \ H_a: p < p_0$:

$$reject \ \ H_0: \ z < -z_\alpha$$

$$not \ reject \ \ H_0: \ z \geq -z_\alpha$$

In addition,

$$p - value = P(Z < z).$$

2. $H_0: p \leq p_0 \ \ vs \ \ H_a: p > p_0$:

$$reject \ \ H_0: \ z > z_\alpha$$

$$not \ reject \ \ H_0: \ z \leq z_\alpha$$

In addition,

$$p - value = P(Z > z).$$

3. $H_0: p = p_0 \;\; vs \;\; H_a: p \neq p_0$:

$$reject \;\; H_0: |z| > z_{\alpha/2}$$

$$not \;\; reject \;\; H_0: |z| \leq z_{\alpha/2}$$

In addition,

$$p - value = P(|Z| > |z|).$$

.

Confidence interval approach for two-sided test:

$$p_0 \notin \left[\bar{p} - z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}, \bar{p} + z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right] \Rightarrow reject \;\; H_0.$$

Note:

$$\sigma_{\bar{P}} = \sqrt{\frac{p_0(1-p_0)}{n}}$$

Is the standard deviation of $\bar{P}$, i.e., $\sigma_{\bar{P}} = \sqrt{Var(\bar{P})}$, as $p = p_0$ (H_0 is true).

Note:

For two-sided test, the confidence interval method might not be equivalent to the other two methods!!

Motivating Example:

$$z = \frac{\bar{p} - p_0}{\left(\sqrt{\frac{p_0(1-p_0)}{n}}\right)} = \frac{0.25 - 0.2}{\left(\sqrt{\frac{0.2(1-0.2)}{400}}\right)} = 2.5 > 1.645 = z_{0.05} = z_\alpha.$$

Therefore, we reject H_0.

The other approach is to use p-value.

$$p - value = P(Z > z) = P(Z > 2.5) = 0.0062 < 0.05 = \alpha,$$

we reject H_0.

Example 10:

Consider the following hypothesis test: $H_0: p \leq 0.8 \;\; vs \;\; H_a: p > 0.8$.

A sample of 400 provided a sample proportion of 0.853.

(a) Using $\alpha = 0.05$, what is the conclusion based on classical hypothesis test?

(b) Using $\alpha = 0.001$, what is the conclusion based on p-value?

[Solution:]

$$p_0 = 0.8, \bar{p} = 0.853, z = \frac{\bar{p} - p_0}{\left(\sqrt{\frac{p_0(1-p_0)}{n}}\right)} = \frac{0.853 - 0.8}{\left(\sqrt{\frac{0.8(1-0.8)}{400}}\right)} = 2.65.$$

(a) Since

$$z = 2.65 > 1.645 = z_{0.05} = z_{\alpha},$$

we reject H_0.

(b)

$$p - value = P(Z > z) = P(Z > 2.65) = 0.004 > 0.001 = \alpha,$$

we do not reject H_0.

Example 11:

An official of a large national union claims that the fraction of women in the union is not significant different from 0.5, $H_0: p = 0.5 \ vs \ H_a: p \neq 0.5$. Using the sample information reported below,

Sample size	400
Women	168
Men	232

(a) Test the hypothesis at 5% level of significance using a critical value (classical approach).

(b) Test the hypothesis at 10% level of significance using a p-value.

(c) Using a confidence interval, test the hypothesis at the 5% level of significance.

[Solution:]

$$p_0 = 0.5, n = 400, \overline{p} = \frac{168}{400} = 0.42,$$

$$z = \frac{\overline{p} - p_0}{\left(\sqrt{\frac{p_0(1-p_0)}{n}}\right)} = \frac{0.42 - 0.5}{\left(\sqrt{\frac{0.5(1-0.5)}{400}}\right)} = -3.2.$$

(a)

$$|z| = |-3.2| = 3.2 > 1.96 = z_{0.025} = z_{\alpha/2} \Rightarrow reject \ H_0.$$

(b)

$$p - value = P(|Z| > |z|) = P(|Z| > 3.2) \approx 0 < 0.1 = \alpha \Rightarrow reject \ H_0.$$

(c) A 95% $(\alpha = 0.05)$ confidence interval of p is

$$\overline{p} \pm z_{\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = 0.42 \pm 1.96\sqrt{\frac{0.42(1-0.42)}{400}} = [0.371, 0.468].$$

Since

$$p_0 = 0.5 \notin [0.371, 0.468],$$

we reject H_0.

Example 12:

Consider the following hypothesis test, $H_0: p \geq 0.75$ vs $H_a: p < 0.75$..

A sample of 300 is selected. Use $\alpha = 0.05$, test the hypothesis based on p-value as:

(a) $\overline{p} = 0.68$.

(b) $\overline{p} = 0.77$.

[Solution:]

(a)

$$p_0 = 0.75, n = 300, \overline{p} = 0.68,$$

$$z = \frac{\overline{p} - p_0}{\left(\sqrt{\frac{p_0(1-p_0)}{n}}\right)} = \frac{0.68 - 0.75}{\left(\sqrt{\frac{0.75(1-0.75)}{300}}\right)} = -2.8.$$

Then

$$p - value = P(Z < z) = P(Z < -2.8) = 0.0026 < 0.05 = \alpha,$$

we reject H_0.

(b)

$$\overline{p} = 0.77, z = \frac{\overline{p} - p_0}{\left(\sqrt{\frac{p_0(1-p_0)}{n}}\right)} = \frac{0.77 - 0.75}{\left(\sqrt{\frac{0.75(1-0.75)}{300}}\right)} = 0.8.$$

Then

$$p - value = P(Z < z) = P(Z < 0.8) = 0.7881 > 0.05 = \alpha,$$

we do not reject H_0.

Chapters 8 and 9: Summary

$(1-\alpha)\cdot 100\%$ confidence intervals:

$$(point\ \ estimate) \pm \left[\left(z_{\alpha/2}, t_{n-1,\alpha/2}\right)\cdot(standard\ \ deviation\ \ (error)\ \ of\ \ point\ \ estimate)\right]$$

Hypothesis testing:

$$\text{test statistic} = \frac{point\ \ estimate - (\mu_0, p_0)}{standard\ \ deviation\ \ (error)\ \ of\ \ point\ \ estimate\ \ under\ \ H_0}$$

Summary table:

	Sample Mean μ		Sample Proportion p
	$n \geq 30$	$n < 30$, normal population	
Test Statistic	$z = \frac{\bar{x}-\mu_0}{\sigma_{\bar{X}}}$ or $z = \frac{\bar{x}-\mu_0}{s_{\bar{X}}}$	$t = \frac{\bar{x}-\mu_0}{s_{\bar{X}}}$	$z = \frac{\bar{p}-p_0}{\sigma_{\bar{P}}}$
Classical approach (critical values)	$-z_\alpha, z_\alpha, z_{\alpha/2}$	$-t_{n-1,\alpha}, t_{n-1,\alpha}, t_{n-1,\alpha/2}$	$-z_\alpha, z_\alpha, z_{\alpha/2}$
Distribution (p − value)	Z	$T(n-1)$	Z
C. I.	$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{X}}$ or $\bar{x} \pm z_{\alpha/2}s_{\bar{X}}$	$\bar{x} \pm t_{n-1,\alpha/2}s_{\bar{X}}$	$\bar{p} \pm z_{\alpha/2}s_{\bar{P}}$