Sampling Distribution of a Proportion

When n*P and n*(1-P) are both ≥ 5, we can use the Z distribution as an approximation to the sampling distribution of the Proportion.

$$Z \cong \frac{P_s - P}{\sqrt{\frac{P(1-P)}{n}}}$$

P_s = sample proportion

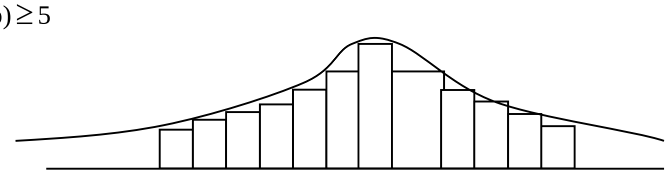
P = population proportion

(1-α)% Confidence Interval Estimator of P: $$P_s \pm Z_\alpha \sqrt{\frac{P_s(1-P_s)}{n}}$$

This formula comes from the fact that the normal distribution can be used to approximate the binomial distribution if np and n(1-p) ≥ 5.

$$\left\{\begin{array}{l} \left.\begin{array}{l}\mu = np \\ \sigma = \sqrt{np(1-p)}\end{array}\right\} \begin{array}{l} for \quad a \\ binomial\end{array} \\ Z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{np(1-p)}} = \frac{\frac{x}{n} - p}{\frac{\sqrt{np(1-p)}}{n}} = \frac{P_s - P}{\sqrt{\frac{p(1-p)}{n}}} \end{array}\right\}$$

∗ As n gets large and p→.5, then the binomial→normal.
Use when np and n(1-p) ≥ 5

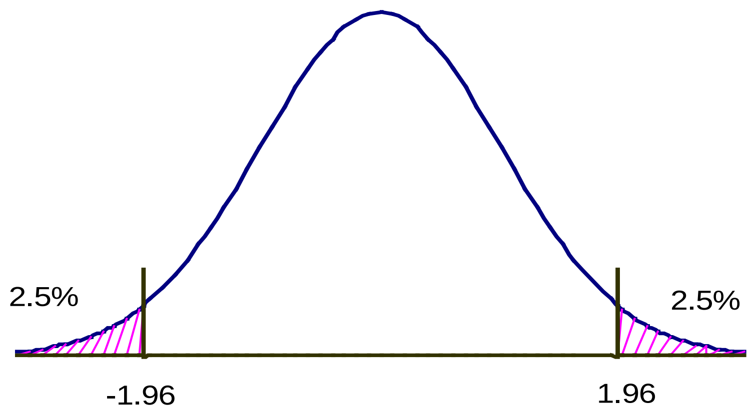


Question: Why does the formula for a CIE use P_s and the formula for Z_{calc} use P?

EXAMPLE:
A politician claims that 70% of the people in her district are Democrats. A researcher takes a sample of 100 people and finds that only 50 are Democrats. Is the politician a liar?
(a) Test at α=.05 (2-tail test)
(b) Construct a 2-tail 95% C.I.E of P

(a)

$$H_0 : P = .70$$
$$H_1 : P \neq .70$$



Z distribution

$$Z = \frac{.50 - .70}{\sqrt{\frac{(.70)(.30)}{100}}} = \frac{-.20}{.0458} = -4.36$$

REJECT H_0 $p < .05$

(b) $$.50 \pm 1.96\left(\sqrt{\frac{(.50)(.50)}{100}}\right)$$

$$.50 \pm .10$$
$$.40 \longleftrightarrow .60$$

We are 95% confidence that the interval .40 to .60 contains the true proportion P of Democrats in the district.

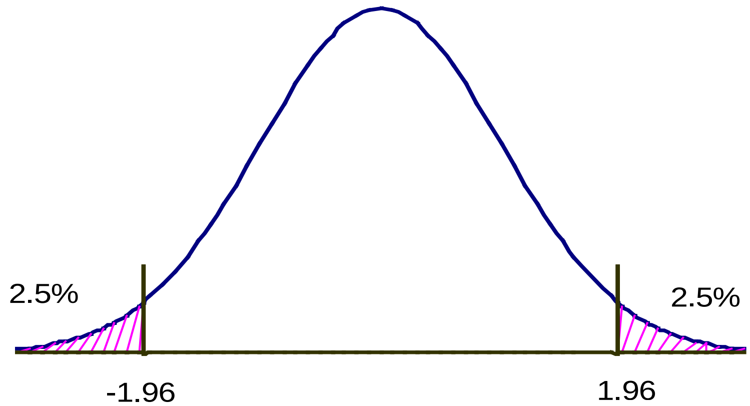
EXAMPLE:

A politician claims that exactly 90% of the American public favors the legalization of drugs. A survey of 100 people shows that only 79 are in favor of drug legalization.

(a) Test at α=.05

(b) Construct a 2-sided 95% C.I.E.

(a) $H_0 : P = .90$
$H_1 : P \neq .90$



Z distribution

$$Z = \frac{.79 - .90}{\sqrt{\frac{(.90)(.10)}{100}}} = \frac{-.11}{.03} = -3.67$$

REJECT H_0 $p < .05$

(b) $$.79 \pm 1.96\left(\sqrt{\frac{(.79)(.21)}{100}}\right) \Rightarrow .79 \pm .08$$

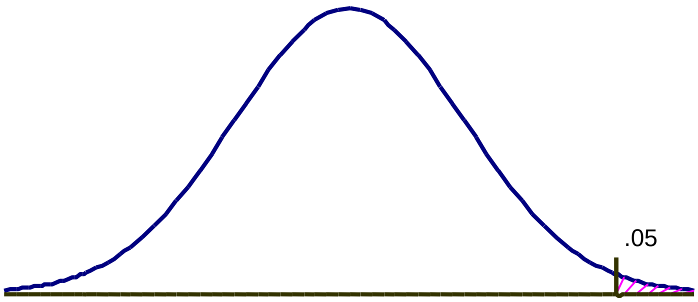
$$.71 \longleftrightarrow .87$$

EXAMPLE:
A Company claims that no more than 8% of its widgets are defective.
Sample: n = 100; 10 defectives.

(a) Test at α=.05
(b) Construct a 2-sided 95% Confidence Interval.

P_s = 10/100=.10

(a) $H_0: P \le .08$
$H_1: P > .08$



$$Z = \frac{.10 - .08}{\sqrt{\frac{(.08)(.92)}{100}}} = \frac{.020}{.027} = .74$$

Do not reject H_0 $P > .05$

(b) $.10 \pm 1.96\left(\sqrt{\frac{(.10)(.90)}{100}}\right)$

$.10 \pm .06$

$.04 \longleftrightarrow .16$

EXAMPLE: SUBSTANTIATING A CLAIM

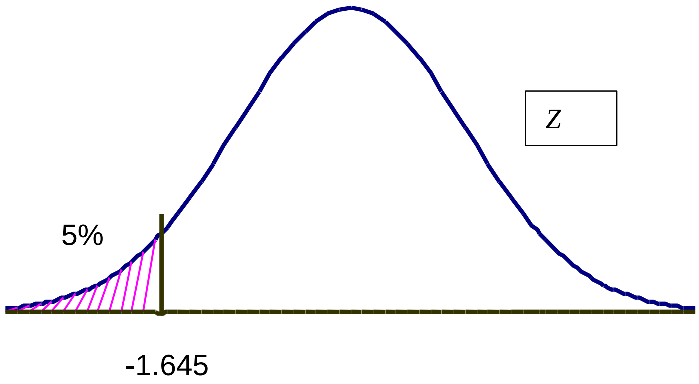
A test prep company prepares candidates for the CPA exam. The company claims that the failure rate for its students is less than 15%. The FTC requires that they substantiate this claim.

The company takes a random sample of 60 students and finds that 3 out of the 60 failed the CPA exam.

Substantiate the claim at $\alpha = .05$

H_0: P $\geq$ .15
H_1: P $<$ .15

The data:
n = 60
$P_s = .05$



$$Z = \frac{.05 - .15}{\sqrt{\frac{(.15)(.85)}{60}}} = \frac{-.10}{.0461} = -2.17$$

Conclusion: Reject H_0. The claim is substantiated.