03/13/2023

Chapter 7 – Estimating and Confidence Intervals for Parameters

One of the main components of Statistics is to make decisions about population parameters based on sample information. This is also called statistical inference.

$\mu, \sigma, p, \ldots$

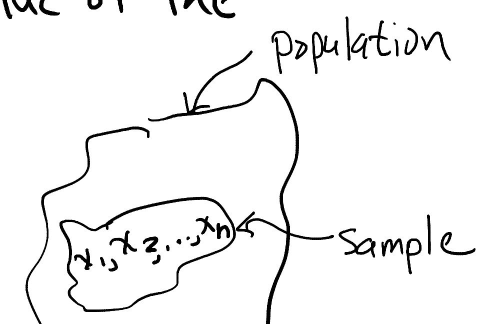
Types of Inferences

1. Estimation of population parameters
   - Point estimation (our "best guess" of the value of the parameter)
   - Interval estimation (An interval of "likely" values of the parameter)

2. Testing of statistical hypotheses concerning population parameters (e.g., $\mu, p, \ldots$)

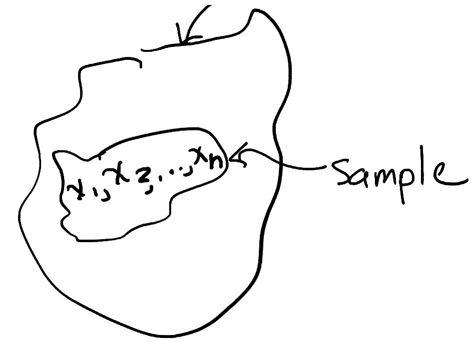
Point Estimation

Def: A point estimate of a population parameter is a single value derived from the sample which represents a plausible value of the parameter.



Ex: Estimate $\underbrace{\text{the population mean}}_{\mu}$

parameter.



Ex: Estimate the $\underbrace{\text{population mean}}_{\mu}$

$$\text{sample mean} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Ex: Estimate the $\underbrace{\text{population proportion}}_{p}$

$$\text{sample proportion} = \hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{where } x_i = 0 \text{ or } 1.$$

Def: A point estimator with mean value equal to the value of the parameter being estimated is called an unbiased estimator.

Ex: Since $\mu_{\bar{x}} = \mu$, $\bar{x}$ is an unbiased estimator of μ.

Ex: Since $\mu_{\hat{p}} = p$, $\hat{p}$ is an unbiased estimator of p.

NOTE: μ_{s^2} = mean of $s^2 = \sigma^2$

μ_s = mean of $s \neq \sigma$

$\underbrace{\text{Confidence Intervals}}_{\text{CI}}$ for a Population Proportion (p)

sample: $x_1, x_2, \ldots, x_n$ where $x_i = \begin{cases} 1, & \text{if "success"} \\ 0, & \text{if "failure"} \end{cases}$

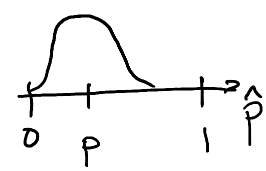
$\hat{p} = \sum_{i=1}^{n} x_i$ changes from sample-to-sample

(0, if "failure"

$\hat{p} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ changes from sample-to-sample

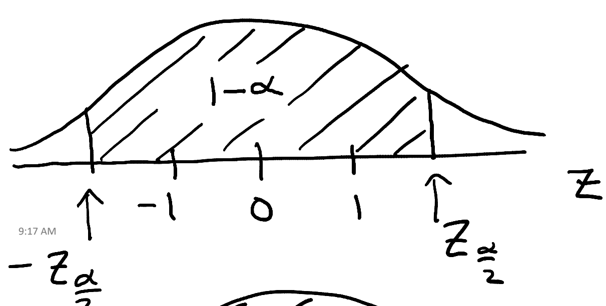
Recall that if $np \geq 5$ and $n(1-p) \geq 5$, then the CLT for $\hat{p}$ yields

$$\hat{p} \overset{approx}{\sim} N\Big(\underset{\mu_{\hat{p}}}{\underset{\shortparallel}{p}}, \underset{\sigma^2_{\hat{p}}}{\underset{\shortparallel}{\frac{p(1-p)}{n}}}\Big).$$

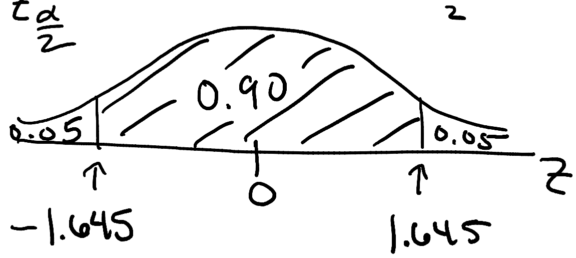


$$Z \approx \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} \text{ where } Z \sim N(0,1).$$

To construct a confidence interval (CI) for p, the endpoints are computed from the sample values. Also, one must specify a <u>confidence level</u>. Typically, values such as 90%, 95%, and 99% are used.



α	confidence level $(1-\alpha)$	z-critical value $(z_{\alpha/2})$
0.10	0.90	1.645
0.05	0.95	1.96
0.01	0.99	2.575



$P(-1.645 < Z < 1.645) = 0.90$

03/20/2023

$1 - \alpha = P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}})$

$$1-\alpha = P\left(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}\right)$$

$$= P\left(-z_{\frac{\alpha}{2}} \le \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \le z_{\frac{\alpha}{2}}\right)$$

$$\vdots$$

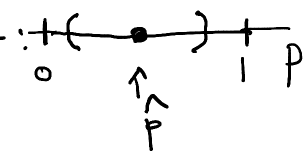
$$= P\Bigg(\underbrace{\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \le p \le \underbrace{\hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}\Bigg)$$

endpoints of CI for p

In summary, an approximate $100(1-\alpha)\%$ CI for p is

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Aside:



Remark: Need to check $n\hat{p} \ge 5$ and $n(1-\hat{p}) \ge 5$ before using above CI.

Ex: CI for p where p = proportion of individuals who have deficient magnesium levels.

Construct a 95% CI for p.

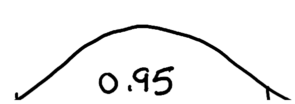
$n = 1200$ $\quad \sum_{i=1}^{n} X_i = 540$ $\quad \hat{p} = \frac{540}{1200} = 0.45$

Check: $n\hat{p} \ge 5$? ✓

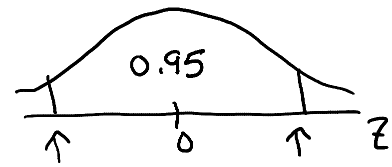
$n(1-\hat{p}) \ge 5$? ✓

We can use CLT to construct 95% CI for p.

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$



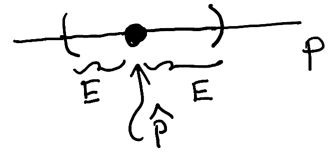
$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$



$$0.45 \pm 1.96 \sqrt{\frac{(0.45)(0.55)}{1200}}$$

$$0.45 \pm \underbrace{0.028}_{}$$

E = margin of error



An approximate 95% CI for p is

$$(0.428\ ,\ 0.478)$$

Define E = margin of error

$$= z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

It follows that

$$n = \left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2 \hat{p}(1-\hat{p})$$

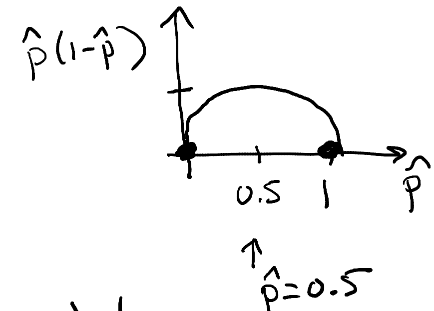
<u>Remark</u>: The sample size required to estimate p to within the distance E with 95% confidence is

$$n = \left(\frac{1.96}{E}\right)^2 \hat{p}(1-\hat{p})$$

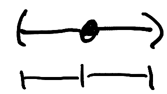
<u>NOTES</u>. 1. Use prior information to find value of $\hat{p}$.

2. If no prior information is known about $\hat{p}$, consider replacing $\hat{p}$ with the value 0.5.

$\hat{p} = 0.5$ produces a "conservative" value of n.



Ex: How many subjects/people must be sampled to be 95% confident that the margin of error of $\hat{p}$ when estimating p is no more than 0.1?



$$E = 0.1$$

$$Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

$$n = \left(\frac{1.96}{0.1}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.1}\right)^2 (0.5)(0.5) = 96.04$$

so use $n = 97$.

What if $E = 0.01$?

$$n = \left(\frac{1.96}{0.01}\right)^2 (0.5)(0.5) = 9{,}604.$$