§5.3 Interval Estimation

Point Estimation

Any point estimator is a random variable, whose distribution is induced by the distribution of X_1, X_2, … , X_n. For example, suppose that we use the sample mean $\bar{X} = \frac{X_1+X_2+\ldots+X_n}{n}$, as the point estimator for the mean μ of a normal distribution. We write $\hat{\mu} = \bar{X}$.

A point estimate is the realized value (that is, a number) of a point estimator $\hat{\theta}$ that is obtained when a sample is actually taken.

Interval Estimation

Confidence Interval (CI): an interval of values that we know will contain the true value of the parameter of interest. The interval is calculated with the use of information from a sample. When we have large samples, we know that the sampling distribution of the sample mean is approximately normal.

■ A **95% confidence interval** for the population mean μ when σ is known is given by

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

or

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

or

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

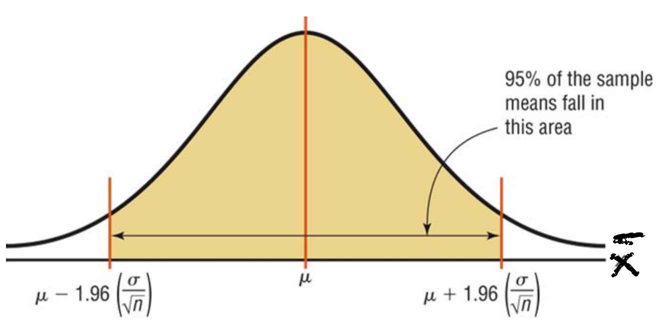
P(interval contains μ) = 0.95

$\mu = E[\bar{X}]$

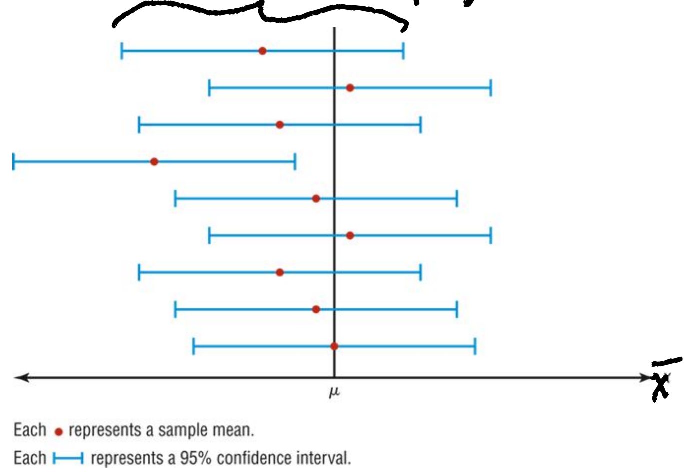
By CLT, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

What does this mean?

1. If we repeatedly draw samples from the population and
2. compute confidences intervals using the above formula for each sample then
3. 95% of the intervals constructed in this way will contain the true population parameter μ.



95% CI for μ



Interpretation: When the values of the variable are normally distributed in the population or the sample size is 30 or more (by CLT), you can be 95% confident that an interval built around a specific sample mean would contain the population mean μ.

100(1-α)% Confidence Interval for Estimating a Population Mean μ with σ Known

i.e. $\alpha = 0.07$
$\Rightarrow 93\%$ CI

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

converting z score to value in $\bar{X}$ distribution

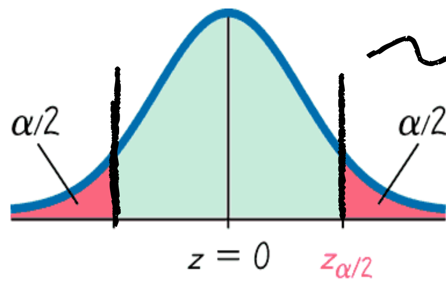
or

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

or

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$z_{\frac{\alpha}{2}}$ s.t. $P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$

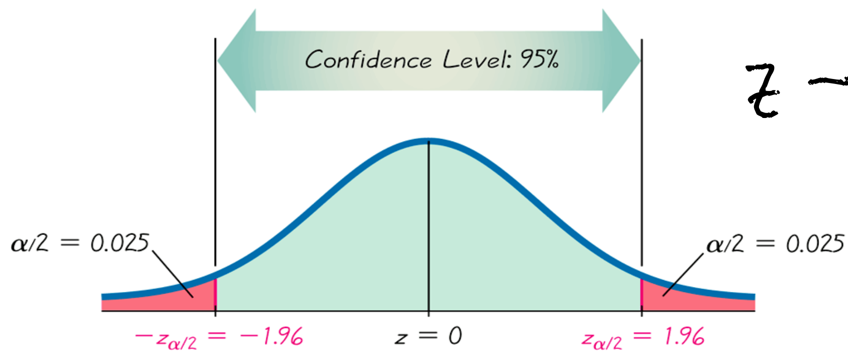


By CLT, we know that $\overline{X} \sim \text{Normal}(\mu, \sigma^2/n)$. Then,

from $Z = \frac{X - \mu}{\sigma}$

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0, 1)$$

For example, if we want 95% confidence interval, $\alpha = 0.05$.



$\mu - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

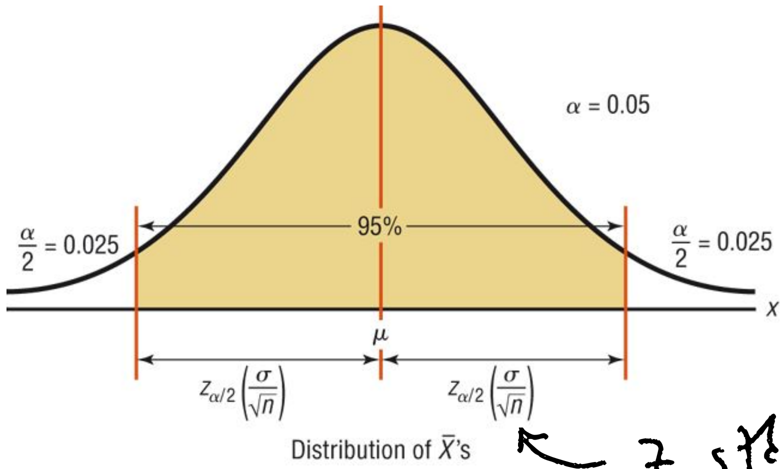
By calculator, $-z_{\alpha/2} = \textbf{invNorm}(0.025, 0, 1) \approx -1.96$, or $z_{\alpha/2} = \textbf{invNorm}(0.975, 0, 1) \approx 1.96$. (We only need one of them)

Then,

$$P\left(-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha = 95\%$$

$\mu \pm z_{\frac{\alpha}{2}}$ std deviations

$$\mu = \overline{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$



z std deviations from the mean

■ A **confidence level** is the probability $1 - \alpha$ (often expressed as the equivalent percentage value) that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times. (The confidence level is also called degree of confidence, or the confidence coefficient.)

■ α = **level of significance**

■ $\frac{\sigma}{\sqrt{n}}$ is the **standard error** of the mean.

■ A **critical value** is the number on the borderline separating sample statistics that are likely to occur from those that are unlikely to occur. The number $\boldsymbol{z_{\alpha/2}}$ is a critical value that is a z score with the property that it separates an area of $\alpha/2$ in the right tail of the standard normal distribution.

$$|\mu - \bar{x}| < z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

■ The **margin of error**, denoted by E, is the maximum likely difference between the observed sample mean $\bar{X}$ and the true value of the population mean μ. $E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

■ Confidence Interval for Estimating a Population Mean μ with σ Known: $\bar{x} - E < \mu < \bar{x} + E$

■ $2E$ is the **length** (or **width**) of the confidence interval.

■ **Calculator:** STAT >> TESTS >> #7: ZInterval

Deriving confidence interval

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

$$= P\left(-z_{\frac{\alpha}{2}} \le \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \le z_{\frac{\alpha}{2}}\right)$$

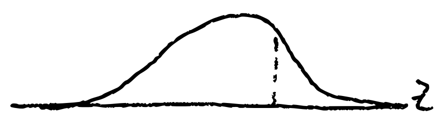
$$= P\left(-z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \bar{x}-\mu \le z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \ge \mu-\bar{x} \ge -z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Common choices for α

$\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$



$z_{\frac{\alpha}{2}}$

$z_{\frac{\alpha}{2}} = 1.96, \quad \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

$\alpha = 0.01, \quad \frac{\alpha}{2} = 0.005$

$z_{\frac{\alpha}{2}} = 2.576$

$\bar{x} \pm 2.576 \frac{\sigma}{\sqrt{n}}$

$\alpha = 0.10 \Rightarrow \frac{\alpha}{2} = 0.05$

$z_{\frac{\alpha}{2}} = 1.645, \quad \bar{x} \pm 1.645 \frac{\sigma}{\sqrt{n}}$

Example 1. People have died in boat and aircraft accidents because an obsolete estimate of the mean weight of men was used. In recent decades, the mean weight of men has increased considerably, so we need to update our estimate of that mean so that boats, aircrafts, elevators, and other such devices do not become dangerously overloaded. Using the weights of men from a certain research, we obtain these sample statistics for the simple random sample: n = 40 and $\overline{X}$ = 172.55 lb. Research from several other sources suggest that the population of weights of men has a standard deviation given by a standard deviation = 26 lb. Construct a 95% confidence interval estimate of the mean weight of all men.

95% chance that $\mu \in \left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$

$$\mu \in \left(172.55 - 1.96\left(\frac{26}{\sqrt{40}}\right), 172.55 + 1.96\left(\frac{26}{\sqrt{40}}\right)\right)$$

$$\boxed{(164.49, 180.61)}$$

Example 2. A commonly used IQ test is scaled to have a mean of 100 and a standard deviation of σ =15. A school counselor was curious about the average IQ of the students in her school and took a random sample of fifty students' IQ scores. The average of these was $\overline{X}$ =107.9. Find the 95% confidence interval for the student IQ in the school.

$$\left(107.9 - 1.96\frac{\sigma}{\sqrt{n}}, 107.9 + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

$$= \left(107.9 - 1.96\left(\frac{15}{\sqrt{50}}\right), 107.9 + 1.96\left(\frac{15}{\sqrt{50}}\right)\right)$$

$$= (103.742, 112.058)$$

Finding a Sample Size for Estimating a Population Mean

$$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{E}\right)^2$$

μ = population mean
σ = population standard deviation
$\overline{X}$ = sample mean
E = margin of error
$z_{\alpha/2}$ = z score separating an area of $\alpha/2$ in the right tail of the standard normal distribution

sample size $\mathbf{n} \geq \left[\frac{\mathbf{Z_{\alpha/2} \cdot \sigma}}{\mathbf{E}}\right]^2$ round up to whole #

Example 3. Assume that we want to estimate the mean IQ score for the population of statistics students. How many statistics students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean? σ = 15.

$$\alpha = 0.05 \Rightarrow z_{\frac{\alpha}{2}} = \pm 1.96, \quad \sigma = 15, \ E = 3$$

$$n \geq \left[\frac{1.96 \cdot 15}{3}\right]^2 = 96.04 \Rightarrow \boxed{n = 97}$$

Example 4. Suppose a sample of size n is to be drawn from a normal distribution where σ is known to be 14.3. How large does n have to be guarantee that the length of the 95% confidence interval for μ will be less than 3.06?

$$\sigma = 14.3, \quad 2E < 3.06 \Rightarrow E < 1.53, \quad \alpha = 0.05 \Rightarrow z_{\frac{\alpha}{2}} = 1.96$$

$$n \geq \left[\frac{1.96 \cdot 14.3}{1.53}\right]^2 = 335.58$$

$$\Rightarrow \boxed{n = 336}$$

Example 5. An institute wants to estimate the household income in a country. The incomes are normally distributed with standard deviation \$26, 000. The institute take a survey of 2416 households randomly. The average household income in the survey is \$56, 000.

(1) Find a 95% confidence interval for the average household income in the country.

$n = 2416$, $\sigma = 26{,}000$, $\bar{x} = 56{,}000$

$\alpha = 0.95 \Rightarrow z_{\frac{\alpha}{2}} = 1.96$,

$$z_{\frac{\alpha}{2}} - \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow \boxed{54963 < \mu < 57037}$$

(2) How large does the sample size have to be to guarantee that the length of the 95% confidence interval for μ will be less than \$1000.

$E < 500$, $z_{\frac{\alpha}{2}} = 1.96$, $\sigma = 26{,}000$,

$$n \geq \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{E}\right)^2 = \left(\frac{1.96 \cdot 26000}{500}\right)^2 = 10387.68$$

$$\Rightarrow \boxed{n = 10388}$$

Example 6. According to stats.nba.com, in NBA 2019-2020 season, Boston Celtics got an average 113.4 points in the first 61 games. Suppose the number of points is normally distributed with population standard deviation = 10.5. Find a 95% confidence interval for the average number of points for Boston Celtics.

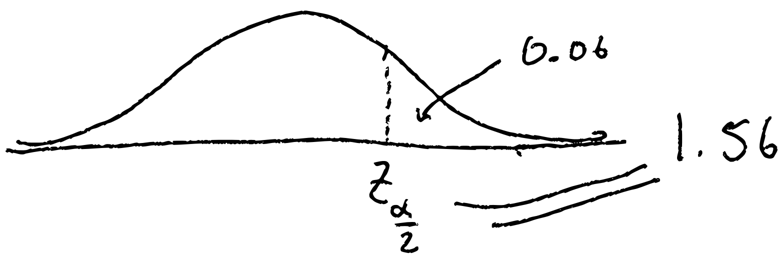
$n = 61$, $\bar{x} = 113.4$, $\sigma = 10.5$, $\alpha = 0.05 \Rightarrow z_{\frac{\alpha}{2}} = 1.96$

$$\bar{x} - 1.96\left(\frac{10.5}{\sqrt{61}}\right) < \mu < \bar{x} + 1.96\left(\frac{10.5}{\sqrt{61}}\right)$$

$$\Rightarrow \boxed{110.77 < \mu < 116.03}$$

88% CI:

$\alpha = 0.12 \Rightarrow \frac{\alpha}{2} = 0.06$



C.V $= \pm 1.56$

$$\bar{X} - 1.56 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.56 \frac{\sigma}{\sqrt{n}}$$

Estimating a Population Proportion

- The sample proportion is the best point estimate of the population proportion.
- We can use a sample proportion to construct a confidence interval to estimate the true value of a population proportion
- We should know how to find the sample size necessary to estimate a population proportion.

■ Notation for Proportions

X = binomial RV
p = population proportion, $q = 1 - p$
$\hat{p} = \frac{X}{n} =$ sample proportion of x successes in the sample of size n
$\hat{q} = 1 - \hat{p}$ = sample proportion of failures in a sample of size n

■ Recall! A point estimate is a single value (or point) used to approximate a population parameter.

The sample proportion $\hat{p}$ is the best point estimate of the population proportion p.

$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n} \cdot np = p$

$\hat{p} \sim N(p, \frac{pq}{n})$

$Var\left(\frac{X}{n}\right) = \frac{1}{n^2} \cdot npq = \frac{pq}{n}$

Confidence Interval for Estimating a Population Proportion p

$$\hat{p} - z_{\alpha/2}\underbrace{\sqrt{\frac{\hat{p}\hat{q}}{n}}}_{\sigma} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

since $\sigma^2 = \frac{pq}{n} \Rightarrow \sigma = \sqrt{\frac{pq}{n}}$

■ **Calculator:** STAT >> TESTS >> A: **1-PropZInt**

■ The margin of error E is also called the **maximum error of the estimate** and can be found by multiplying the critical value and the standard deviation of the sample proportions.

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Sample size n for estimating a population proportion p

1. When an estimate $\hat{p}$ is known, $n = \frac{(Z_{\alpha/2})^2 \hat{p}\,\hat{q}}{E^2}$
2. When **no** estimate $\hat{p}$ is known, $n = \frac{(Z_{\alpha/2})^2 (0.25)}{E^2}$

Example 6. A Pew Research Center poll of 1501 randomly selected U.S. adults showed that 70% of the respondents believe in global warming. The sample results are $n = 1501$, and $\hat{p} = 0.70$.

a. Find the margin of error E that corresponds to a 95% confidence level.

$n = 1501 \qquad \alpha = 0.05 \Rightarrow z_{\frac{\alpha}{2}} = 1.96$

$\hat{p} = 0.70$

$\hat{q} = 0.30$

$$E = 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{0.7 \cdot 0.3}{1501}} = 0.023$$

b. Find the 95% confidence interval estimate of the population proportion p.

$$\hat{p} - 0.023 < p < \hat{p} + 0.023$$

$$\boxed{0.677 < p < 0.723}$$

Example 7. In a poll of 738 cell phone users, 266 stated that they had been walked into by someone using their cell phone. Estimate the true proportion of cell phone users who have had someone walk into them with 99% confidence.

$\hat{p} = \frac{266}{738} = 0.36$

$\hat{q} = 0.64$

$n = 738$

$\alpha = 0.01 \Rightarrow z_{\frac{\alpha}{2}} = 2.58$

$$\hat{p} - 2.58\sqrt{\frac{0.36 \cdot 0.64}{738}} < p < \hat{p} + 2.58\sqrt{}$$

$$\boxed{0.3149 < p < 0.4060}$$

Example 8. People use electronic devices every day. Some people even have smartphone addiction. A health research institute estimate the percent of college students using smartphone more than 4 hours per day. How many college students must be surveyed in order to be 95% confident that the sample percentage is within 2% of the true percentage?

$E = 0.02$

$z_{\frac{\alpha}{2}} = 1.96$

$$n = \frac{(z_{\frac{\alpha}{2}})^2 (0.25)}{(E)^2} \leftarrow 0.25 = \max(p(1-p))$$

$$\boxed{= 2401}$$