10.1. Population Mean: Large Sample Case ($n \geq 30$)

Motivating example:
In the survey conducted by CJW, Inc., a mail-order firm, the satisfaction scores (1–100) of 100 customers ($n=100$) are obtained. Suppose $\sigma = 20$ is known. Also,

$$\bar{x} = 82, \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2.$$

Derivation of 95% confidence interval:

Since

$$\bar{X} \approx N(\mu, \frac{\sigma^2}{n}).$$

Thus,

$$\frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \approx Z (\sim N(0,1)).$$

Then,

$$0.95 = P(|Z| \leq 1.96) \approx P\left(\left|\frac{\bar{X} - \mu}{\sigma_{\bar{X}}}\right| \leq 1.96\right) = P\left(\left|\frac{\mu - \bar{X}}{\sigma_{\bar{X}}}\right| \leq 1.96\right) = P(-1.96 \leq \frac{\mu - \bar{X}}{\sigma_{\bar{X}}} \leq 1.96) = P(-1.96\sigma_{\bar{X}} \leq \mu - \bar{X} \leq 1.96\sigma_{\bar{X}})$$

$$= P(\bar{X} - 1.96\sigma_{\bar{X}} \leq \mu \leq \bar{X} + 1.96\sigma_{\bar{X}}).$$

⇒ There is an approximate 95% chance that the population mean $\mu$ will fall between $\bar{X} - 1.96\sigma_{\bar{X}}$ and $\bar{X} + 1.96\sigma_{\bar{X}}$, i.e.,

$$P(\mu \in [\bar{X} - 1.96\sigma_{\bar{X}}, \bar{X} + 1.96\sigma_{\bar{X}}]) \approx 0.95, \quad \mu \text{ falls in the interval }$$

$$[\bar{X} - 1.96\sigma_{\bar{X}}, \bar{X} + 1.96\sigma_{\bar{X}}] \text{ with a chance close to 0.95.}$$

Note: in the above equation,
\[
0.95 \approx P\left( \left| \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \right| \leq 1.96 \right) = P\left( \left| X - \mu \right| \leq 1.96\sigma_{\bar{X}} \right)
\]

There is an approximate 95% chance that the sample mean will provide a sampling error of \(1.96\sigma_{\bar{X}}\).

**Example (continue)**

In the above example, since \(\sigma_{\bar{X}} = 2\), thus

\[
P(\bar{X} - 1.96 \times 2 \leq \mu \leq \bar{X} + 1.96 \times 2) = P(\bar{X} - 3.92 \leq \mu \leq \bar{X} + 3.92) \approx 0.95
\]

There is an approximate 95% chance that the population mean \(\mu\) will fall between \(\bar{X} - 3.92\) and \(\bar{X} + 3.92\).

**95% confidence interval:**

Suppose the sample size is large.

- **As \(\sigma\) is known,**

  \[
  \bar{x} \pm 1.96\sigma_{\bar{x}} = \bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = \left[ \bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}} \right]
  \]

  is a 95% confidence interval estimate of the population mean \(\mu\).

- **As \(\sigma\) is unknown,**

  \[
  \bar{x} \pm 1.96s_{\bar{x}} = \bar{x} \pm 1.96\frac{s}{\sqrt{n}} = \left[ \bar{x} - 1.96\frac{s}{\sqrt{n}}, \bar{x} + 1.96\frac{s}{\sqrt{n}} \right]
  \]

  is a 95% confidence interval estimate of the population mean \(\mu\).

where \(s^2\) is the sample variance and \(s_{\bar{x}}\) is the estimate of \(\sigma_{\bar{x}}\).

**Example (continue)**

In the above example, since \(\bar{x} = 82\), \(\sigma_{\bar{x}} = 2\), and \(n = 100\), thus

\[
\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = 82 \pm 1.96\frac{20}{\sqrt{100}} = 82 \pm 3.92 \equiv [78.08, 85.92]
\]
is a 95% confidence interval estimate of the population mean $\mu$.

**General confidence interval:**

**Definition of $\frac{Z_{\alpha/2}}{2}$:**

Let $Z$ be the standard normal random variable. Then,

$$P\left(Z > z_{\alpha/2}\right) = \frac{\alpha}{2}.$$

As $\alpha \leq 0.5$,

$$P\left(|Z| \leq z_{\alpha/2}\right) = 1 - \alpha.$$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$1 - \alpha$</th>
<th>$\alpha/2$</th>
<th>$z_{\alpha/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>0.05</td>
<td>$z_{0.05} = 1.64$</td>
</tr>
<tr>
<td>0.05</td>
<td>0.95</td>
<td>0.025</td>
<td>$z_{0.025} = 1.96$</td>
</tr>
<tr>
<td>0.01</td>
<td>0.99</td>
<td>0.005</td>
<td>$z_{0.005} = 2.576$</td>
</tr>
</tbody>
</table>
Derivation of $(1-\alpha) \times 100\%$ confidence interval:

As the sample size is large,

\[
1-\alpha = P\left( |Z| \leq z_{\alpha/2} \right) \approx P\left( \left| \frac{\bar{X} - \mu}{\sigma_x} \right| \leq z_{\alpha/2} \right) = P\left( \frac{\mu - \bar{X}}{\sigma_x} \leq z_{\alpha/2} \right)
\]

\[
= P\left( -z_{\alpha/2} \leq \frac{\mu - \bar{X}}{\sigma_x} \leq z_{\alpha/2} \right) = P\left( -z_{\alpha/2} \sigma_x \leq \mu - \bar{X} \leq z_{\alpha/2} \sigma_x \right)
\]

\[
= P\left( \bar{X} - z_{\alpha/2} \sigma_x \leq \mu \leq \bar{X} + z_{\alpha/2} \sigma_x \right)
\]

\[\Rightarrow\] There is an approximate $(1-\alpha) \times 100\%$ chance that the population mean $\mu$ will fall between $\bar{X} - z_{\alpha/2} \sigma_x$ and $\bar{X} + z_{\alpha/2} \sigma_x$, i.e.,

\[
P\left( u \in \left[ \bar{X} - z_{\alpha/2} \sigma_x, \bar{X} + z_{\alpha/2} \sigma_x \right] \right) \approx 1 - \alpha , \mu \text{ falls in the interval}
\]

\[
\left[ \bar{X} - z_{\alpha/2} \sigma_x, \bar{X} + z_{\alpha/2} \sigma_x \right] \text{ with a chance close to } 1 - \alpha .
\]

**Note:** As $\alpha = 0.05$, the above derivations are exactly the same as the ones for 95% confidence interval estimate.

**Motivating Example (continue)**

As $\alpha = 0.1,$

\[
P(\bar{X} - z_{0.1/2} \sigma_x \leq \mu \leq \bar{X} + z_{0.1/2} \sigma_x) = P(\bar{X} - z_{0.05} \sigma_x \leq \mu \leq \bar{X} + z_{0.05} \sigma_x)
\]

\[
= P(\bar{X} - 3.28 \leq \mu \leq \bar{X} + 3.28) \approx 1 - 0.1 = 0.9
\]

There is an approximate 90% chance that the population mean $\mu$ will fall between

\[
\bar{X} - 3.28 \text{ and } \bar{X} + 3.28 .
\]

$(1-\alpha) \times 100\%$ confidence interval:

Suppose the sample size is large.

- As $\sigma$ is known,
$\bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}} = \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \equiv \left[ \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$ is a $(1 - \alpha)\times100\%$ confidence interval estimate of the population mean $\mu$.

As $\sigma$ is unknown,

$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} = \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \equiv \left[ \bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \right]$ is a $(1 - \alpha)\times100\%$ confidence interval estimate of the population mean $\mu$.

**Motivating Example (continue)**

As $\alpha = 0.1$,

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 82 \pm z_{0.05} \cdot \frac{20}{\sqrt{100}} = 82 \pm 1.64 \cdot \frac{20}{\sqrt{100}} = [78.72, 85.28]$$

is a 90% confidence interval estimate of the population mean $\mu$.

**Note:** in the CJW, Inc. example, the 95% confidence interval is wider than the 90% confidence interval. Intuitively, if we want to make sure that we will make less mistakes, we should speak vaguely (wider confidence interval). For instance, if we want to get a 100% confidence interval (for sure), the interval $(\infty, \infty)$ would make us not make any mistake.

**Note:** the length of the confidence interval is $2z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ or $2z_{\alpha/2} \frac{s}{\sqrt{n}}$.

Therefore, a larger sample size $n$ will provide a narrow interval and a greater precision.
Example A:
A random sample of 81 workers at a company showed that they work an average of 100 hours per month with a standard deviation of 27 hours. Compute a 95% confidence interval for the mean hours per month all workers at the company work.

[solution:]
As \( \alpha = 0.05 \),

\[
\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 100 \pm z_{0.025} \frac{27}{\sqrt{81}} = 100 \pm 1.96 \cdot \frac{27}{9} = [94.12, 105.88]
\]

is a 95% confidence interval estimate of the population mean \( \mu \).

Online Exercise:

*Exercise 10.1.1*
*Exercise 10.1.2*