Ch8.2 One-Sample Problem

We will learn how to

- provide a point estimate of a population parameter.
  - eg, find a point estimate for μ (use $\bar{x}$)
  - find a point estimate for σ (use s)
  - both $\bar{x}$ and s^2 are unbiased estimators for μ and σ^2 respectively.

    Make sure you can explain what an unbiased estimator is [recall:

    $E(\hat{\theta}) = \theta$ means $\hat{\theta}$ is an unbiased estimator]
  - we use a hat symbol ($\hat{}$) to indicate an estimate (eg $\hat{\theta}$ is a point estimator for θ, $\hat{\mu}$ is a point estimator for μ)

What is a standard error?

When the standard deviation of a statistic is estimated from the data, the result is called the standard error of the statistic.

Eg. $SE_{\bar{X}} = \frac{s}{\sqrt{n}}$ (we use s to estimate σ)

standard error of $\bar{x}$

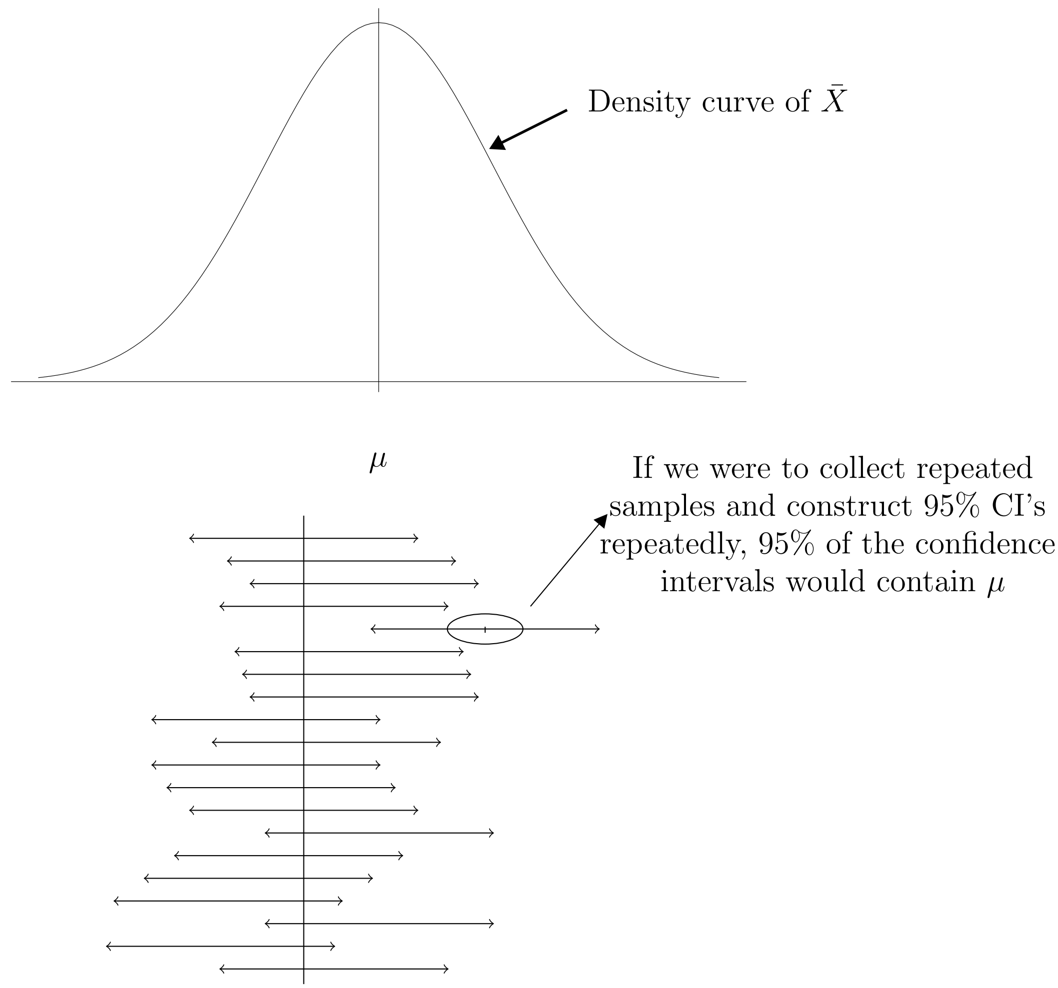
Distinguish between a *Point Estimate* and a *Confidence Interval*

A **point estimate** is a single value that has been calculated to estimate the unknown population parameter. For example, we use the **sample mean** $\bar{x}$ as a point estimate for the **population mean** μ.

A **confidence interval**, on the other hand, gives us a range of possible values that is likely to contain the unknown population parameter. This range of values is generated using a set of sample data.

We can never be totally certain about what the unknown population parameter is, but we do have a level of certainty whether or not it is in our interval. This is known as the **confidence level**. The most commonly used confidence levels are 90%, 95% and 99%.

The diagram on the next page is an example of a series of 95% confidence intervals created by taking random samples from a population.



The vertical line represents the true value of the population parameter. 19 of the 20 intervals, or 95% of the intervals intersect the vertical line. This illustrates what 95% confidence means. If we were to collect repeated samples and construct 95% confidence intervals repeatedly for the population parameter, on average, 95% of the intervals generated would capture the population parameter.

(It is important to realize that when you collect data and construct a single 95% confidence interval, you are merely constructing, for example one of the 20 CI's as above. Therefore, the 95% confidence interval you constructed might NOT contain the population parameter - but then, it might also contain the population parameter. When asked what a 95% confidence interval means, many students will say something incorrect such as "the probability that the population parameter is contained in the interval is 0.95". That is incorrect!! That is because the true population parameter is either IN the interval (so probability=1), or the true population parameter is OUTSIDE the interval (in which case, probability=0). The correct interpretation is to describe the diagram on the previous page and say: "if we were to collect more samples using the same method and construct 95% confidence intervals repeatedly, 95% of the intervals so constructed would contain the population parameter".)

General notation used with CI:

① Recall confidence levels are commonly 90%, 95% and 99% (eg, it is common to construct 95% confidence intervals for population mean)

② We want to write the confidence level in the form $1-\alpha$, where α is a number between 0 and 1.

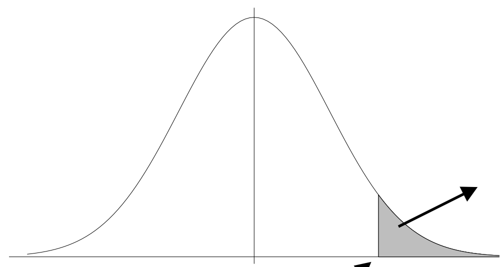
Eg, if confidence level is 95%, $\alpha = 1 - 0.95 = 0.05$

if confidence level is 90%, $\alpha = 1 - 0.9 = 0.1$ and so on

③ Z_α denotes the z-score that has area α to the right of the standard normal curve.

Eg, $Z_{0.05}$ denotes z-score that has area 0.05 to its right.

[Also, $z_{\alpha/2} = z_{0.025}$ if $\alpha = 0.05$, and $z_{0.025} = 1.96$]



area is 0.05

$Z_{0.05} = 1.645$

④ In general, $\underbrace{100(1-\alpha)\%}_{=95\% \text{ if } \alpha = 0.05}$ CI has the following form

$$statistic \ \pm z_{\alpha/2} \cdot SE_{statistic}$$

SE - standard error is the standard deviation of a statistic estimated from the data, (eg $SE_{\bar{x}} = \frac{s}{\sqrt{n}}$)

z or t (more about t in later pages)

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

⋆ ⑤ Steps to calculate a one-sample z-CI.

Very important to know

Assumptions: ① Normal population or large sample

② Simple random sample

③ σ known

population standard deviation is known

Step ①: For confidence level $1-\alpha$ (eg $\alpha = 0.05$ for 95% CI), use normal table to find $z_{\alpha/2}$

Step ②: The CI for μ is

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

Important

$z^* \cdot \frac{\sigma}{\sqrt{n}}$ is called margin of error

For 90% CI	For 95% CI	For 99% CI
$z_{0.1/2} = 1.645 \quad (= z_{0.05})$	$z_{0.025} = 1.96$	$z_{0.005} = 2.576$

How did we arrive at a formula $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$?

The following steps provide the background for the construction of a $100(1-\alpha)\%$ CI for a population mean μ:

We know $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Hence, $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$\Rightarrow P(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1-\alpha$

$\Rightarrow P(-z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = 1-\alpha$

$\Rightarrow P(-\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = 1-\alpha$

$$\Rightarrow P(\underbrace{\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}) = 1-\alpha$$

this provides a formula for a $100(1-\alpha)\%$ CI for μ: $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Example:

Cadmium, a heavy metal, is toxic in animals. Mushrooms, however are able to absorb and accumulate cadmium at high concentrations. The Czech and Slovak governments have set a safety limit for cadmium in dry vegetables at 0.5 parts per million (ppm). Below are cadmium levels of a random sample of the edible mushroom Boletus pinicola:

0.24 0.59 0.62 0.16 0.77 1.33 0.92 0.19 0.33 0.25 0.59 0.32

Find and interpret a 95% confidence interval for the mean cadmium level of all Boletus pinicola mushrooms. Assume a population standard deviation of cadmium levels in Boletus pinicola mushrooms of 0.37 ppm.

<u>Solution:</u>

Make sure you get into the habit of checking assumptions:

① You should state that normal population is assumed (in reality, you construct a normal probability plot of your sample to see if it is plausible that the population in indeed Normally distributed - we won't do this step in STAT 251 but am pointing it out here so you know)

② Simple random sample assumed

③ For this question, σ is known ($\sigma = 0.37$)

④ $\bar{x} = \frac{0.24+0.59+\cdots+0.32}{12} = 0.52583$

$n = 12$

$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{4.6819 - \frac{(6.31)^2}{12}}{11} = 0.1239901515$

$\therefore s = 0.352122353$

Let μ be the population mean cadmium level.

95% confidence interval for μ is

1.96

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

$$0.52583 \pm 1.96(\frac{0.37}{\sqrt{12}})$$

$$(0.3165, 0.7352)$$

How to interpret this confidence interval?

We are 95% confident that the interval from 0.3165 to 0.7352 captures the true population mean.

Example

Assume that the helium porosity (in percentage) of coal samples taken from any particular seam is Normally distributed with true standard deviation 0.75.

(a) Compute the 95% CI for the true average porosity of a certain seam if the average porosity for 20 specimens from the seam was 4.85.

(b) Compute a 98% CI for true average porosity of another seam based on 16 specimens with a sample average porosity of 4.56.

(c) How large a sample size is necessary if the width of the 95% interval is to be 0.40?

(d) What sample size is necessary for the margin of error to be 0.2 with 99% confidence?

Solution:

Note that population is normally distribution, σ is known.

(a) 95% CI is:

$$\bar{x} \pm \underbrace{z}_{1.96} \cdot \frac{\sigma}{\sqrt{n}}$$

$$4.85 \pm 1.96 \cdot \frac{0.75}{\sqrt{20}} = 4.85 \pm 0.33 = (4.52, 5.18)$$

(b) 98% CI is

$$\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}} = 4.56 \pm \frac{(2.33)(0.75)}{\sqrt{16}} = (4.12, 5.00)$$

$z_{\alpha/2} = z_{0.01} = 2.33$

(c) Width is $2 \cdot z \cdot \frac{\sigma}{\sqrt{n}}$

$$2 \cdot z \cdot \frac{\sigma}{\sqrt{n}} = 0.4$$

$$2 \cdot 1.96 \times \frac{0.75}{\sqrt{n}} = 0.4$$

$$n = (\frac{2 \times 1.96 \times 0.75}{0.4})^2 = 54.02$$

$$n = 55 \quad (\textit{we round } \underline{\textit{up}})$$

(d) Margin of error$= \frac{z \cdot \sigma}{\sqrt{n}} = 2.575 \cdot \frac{0.75}{\sqrt{n}}$

Set $2.575 \times \frac{0.75}{\sqrt{n}} = 0.2$

$$n = (\frac{2.575 \times 0.75}{0.2})^2 = 93.24$$

$$\textit{so} \quad n = 94 \quad (\textit{we round up})$$