Lec 25

7 Statistical Intervals Based on a Single Sample

Consider a population with an unknown parameter μ, the mean.

- A point estimation for μ is a single value that can be considered as a sensible estimate for μ.
- The point estimate is obtained by taking a representative sample and use the corresponding statistics, $\overline{x}$.
- Because of sampling variability, it is virtually never the case where $\overline{x} = \mu$.
- The statistics $\overline{x}$ does not give any information about how close $\overline{x}$ is to μ. Thus, we need to consider the variation around μ.
- An alternative way to a point estimate for μ is to report an entire interval of plausible values, called confidence interval
- A confidence interval reports a range of values where μ is likely to fall.
- A confidence interval depends on α, where $(1-\alpha)100\%$ is the confidence level, which is a measure of the degree of reliability of the interval. A confidence level of 95% implies that 95% of all samples would give an interval that includes μ.

7.1 Basic Properties of Confidence Intervals

(for 95% confidence level $\alpha = 0.05$)

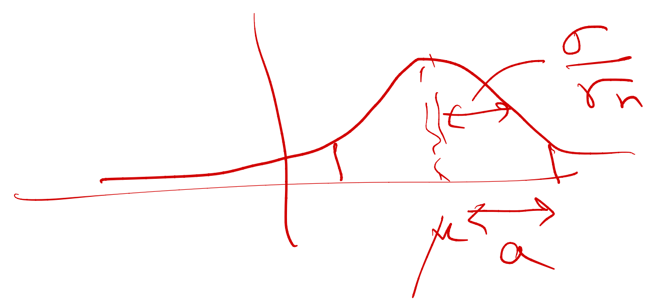
Suppose $X_1, \cdots, X_n$ is a random sample of size n. Suppose that the unknown parameter of interest is the population mean μ.

1. If population is normally distributed or n large
2. The value of pop. stdev σ is known (for now)

Then,

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

pdf for $\overline{X}$



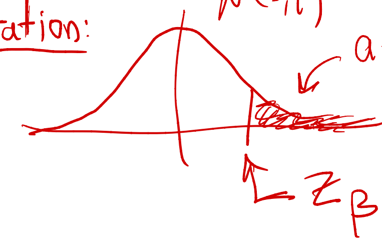
Find a s.t.

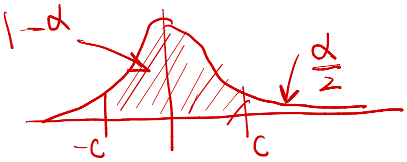
$$P(\mu - a \le \overline{X} \le \mu + a) = 1 - \alpha$$

Confidence interval then: $(\overline{X} - a, \overline{X} + a)$ $a = ?$

$$P\left(\underbrace{\frac{-a}{\sigma/\sqrt{n}}}_{-c} \le Z \le \underbrace{\frac{a}{\sigma/\sqrt{n}}}_{c}\right) = 1-\alpha$$

Notation:

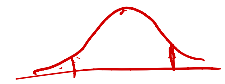




$\boxed{c = z_{\alpha/2}}$

either given w/ the table or:

95% conf. level $\alpha = 0.05$ $\quad z_{0.025} = 1.96$



0.975 | 0.025

look up in table to get 1.96

Note: σ and n are known, but $\overline{X}$ is unknown since we did not take the sample yet.

This interval is random because the two endpoints of the interval involve a random variable. The interval's width is $2 \cdot (1.96) \cdot \sigma/\sqrt{n}$, a fixed number; only the location of the interval (its midpoint $\overline{X}$) is random.

In general, for all CI:

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

$$\text{estimate} \pm \text{critical value} \cdot \sigma \text{ estimate}$$

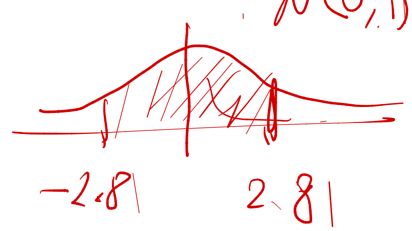
95% conf. level

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

Example 78. (Exercises 1 on textbook page 284) Consider a normal population distribution with the value of σ known.

Question: What is the confidence level for the interval $\overline{x} \pm 2.81\sigma/\sqrt{n}$?

99.5%



from table → 0.9975 $\quad$ 0.0025

0.0025 $\quad$ 0.995

Example 79. Consider a normal population distribution with the value of $\sigma = 3$.

Question: What is the 95% for the population mean, μ, when $n = 25$ and $\overline{x} = 58.3$?

$\alpha = 0.05 \rightarrow z_{0.025}$

$$58.3 \pm 1.96 \times \frac{3}{\sqrt{25}}$$

σ

n

$(57.124, 59.476)$

Example 80. A sample of $n = 31$ trained typists was selected, and the preferred keyboard height was determined for each typist. The resulting sample average preferred height was $\overline{x} = 80.0$ cm. Assuming that the preferred height is normally distributed with $\sigma = 2.0$ cm, obtain the 95% confidence interval for μ, the true average preferred height for the population of all experienced typists.

Solution.

NOTE: How to interpret a CI?

In $(1-\alpha)100\%$ of the samples, the population mean will be inside the CI.

Choosing a level of confidence:

- The precision of the CI refers to the width, w, of the interval. The more __precise__ a CI is, the __smaller__ its width. Because the smaller width implies the interval identifies fewer value μ.
- The reliability of a CI refers to its CL. The more __reliable__ the CI is, the __closer__ you are to the population mean.
- As the CL __↘__, the width of the interval also __↘__. So less precision implies more reliable.
- As the CL __↗__, the width of the interval also __↗__. So more precision implies less reliable.