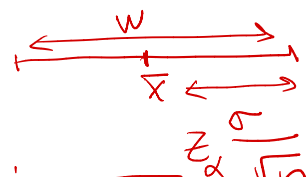
Lec 26

Question: How do we balance precision and reliability?

Answer: Specify both CL and interval width, then determine the <u>sample size</u> as needed. Suppose we have a normal population where σ is known. Then

α given & width of CI is w



We want n large enough so that: $2 z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le W$ so $\boxed{n \ge \frac{4 z_{\frac{\alpha}{2}}^2 \sigma^2}{W^2}}$

Example 81. A new operating system has been installed, and we wish to estimate the true average response time μ for the new environment. Assuming that response time is normally distributed with $\sigma = 25$ millisec, what sample size is necessary to ensure that the resulting 95% CI has a width of (at most) 10?

Solution.

$R \sim N(\mu, 25^2)$

$W = 10 \quad \alpha = 0.05 \qquad z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

$\frac{4 \times 1.96^2 \cdot 25^2}{10^2} = 96.04 \qquad \boxed{n \ge 97}$

7.2 Large-Sample Confidence Intervals for a Population Mean and Proportion

The CI for μ given in the previous section assumed that the population distribution is <u>normal</u> with the value of <u>σ known</u>. We now present a <u>large-sample CI</u> whose validity does not require these assumptions.

7.2.1 A Large-Sample Interval for μ

Let $X_1, X_2, \cdots, X_n$ be a random sample from a population having a mean μ and standard deviation σ. Provided that n is sufficiently large, the <u>CLT</u> implies that $\overline{X}$ has approximately a <u>normal</u> distribution whatever the nature of the population distribution. It then follows that <u>$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$</u> has approximately a standard normal distribution, so that

A practical difficulty with this development is that computation of the CI requires the value of σ, which will rarely be known. Consider replacing the population standard deviation σ by the sample stdev S, which gives

> If n is sufficiently large, the standardized variable
>
> $$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$
>
> has approximately a standard normal distribution. This implies that
>
> $$\bar{X} \pm z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leftarrow \text{sample stdev}$$
>
> is a large-sample confidence interval for μ with confidence level approximately $100(1 - \alpha)\%$. This formula is valid regardless of the shape of the population distribution.
> **NOTE:** $n \geq 30$ will be sufficient to justify the use of this interval.

Example 82. A random sample of 110 lightning flashes in a certain region resulted in a sample average radar echo duration of 0.81 sec and a sample standard deviation 0.34 sec. Calculate a 99% (two-sided) confidence interval for the true average echo duration μ, and interpret the resulting interval.

Solution.

$\bar{X} = 0.81 \qquad S = 0.34$

$n = 110 \geq 30$

CL 99% $\qquad \alpha = 0.01$

P161 in book Table 4.1 $\longrightarrow$ $z_{\frac{\alpha}{2}} = z_{0.005} = 2.57$

CI $\left(0.81 - 2.57 \times \frac{0.34}{\sqrt{110}},\ 0.81 + 2.57 \times \frac{0.34}{\sqrt{110}}\right)$

$= (0.7265, 0.8935)$

Table 4.1 on p161 of book;

Percentile	90	95	97.5	99	99.5	99.9	99.95
α (upper-tail area)	.1	.05	.025	.01	.005	.001	.0005
$z_\alpha = 100(1 - \alpha)$th percentile	1.28	1.645	1.96	2.33	2.58	3.08	3.27

7.2.2 A confidence Interval for a Population Proportion

Consider a population whose members can be divided into 2 separate groups. Let p be the population proportion or the true proportion, which is unknown. To estimate p, we use

$\hat{p} = \frac{X}{n}$ ______ where X = the number of people in the sample who have a given characteristic, n = sample size.

- X follows a $Bin(n, p)$ ______.
- Furthermore, if both $np \geq 10$ and $n(1-p) \geq 10$, X has approximately a $N(np, np(1-p))$ distribution.
- Since $\hat{p}$ is just X multiplied by the constant $\frac{1}{n}$, $\hat{p}$ also has approximately a $N\left(p, \frac{p(1-p)}{n}\right)$ distribution with mean

$$E(\hat{p}) = p$$

and variance

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

For μ:

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$\hat{\mu}$

- If $n > \cancel{40}$ 30, then a CI for the proportion p is

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

Since in general, for all CI:

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

Example 83. (Exercise 21 on textbook page 294) In a sample of 1000 randomly selected consumers who had opportunities to send in a rebate claim form after purchasing a product, 250 of these people said they never did so. Calculate an upper confidence bound at the 95% confidence level for the true proportion of such consumers who never apply for a rebate. Based on this bound, is there compelling evidence that the true proportion of such consumers is smaller than 1/3? Explain your reasoning.

Solution.