MAT2377

Catalin Rada

Version May 28, 2009

Lectures covering 8-1; 8-2

Confidence intervals for mean when the variance is known

Recall: The observed value of an estimator (of an unknown parameter) is called point estimation. An example: the arithmetic mean of a random sample $\overline{x} = 35$ is a point estimation of the mean of the population.

Question: How precise is this estimation? The answer is going to be given in terms of: Confidence Interval.

Suppose that θ is a parameter (unknown), assume that $\hat{\Theta} = h(X_1, \ldots, X_n)$ is an estimator for θ. Now let us assume that we can find 2 statistics $L = L(X_1, \ldots, X_n)$ and $U = U(X_1, \ldots, X_n)$ such that $1 - \alpha = P(L \leq \theta \leq U)$, where $\alpha \in [0, 1]$.(In many exmaples we are going to use 0.90, 0.95, 0.99.) Let l and u be the observed values of L and U, respectively.

Def: We say that $[l, u]$ is a $100(1-\alpha)\%$ confidence interval for θ. Moreover: a) $1-\alpha$ is called the level of confidence; b) $u-l$ (the length of the confidence interval) measures the precision: shorter length implies more precision!

Def: Let $Z \sim N(0,1)$. A real number z_a is called a percentile (for Z) if: $a = P(Z > z_a) = 1 - \Phi(z_a)$.

EXP: $z_{0.01} = 2.32$, or 2.326 if you want more precision; $z_{0.05} = 1.645$; $z_{0.025} = 1.96$; $z_{0.005} = 2.576 \cong 2.58$.

ESTIMATION of μ, when σ is known:

Conditions: a) Normal population or large size of random sample(i.e., $n > 30$); b) variance of population σ^2 is known.

Under these conditions $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is normal or approximately normal! The magic is this: $1-\alpha = P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}})$. PROOF: $P(-z_{\frac{\alpha}{2}} < Z <$

$z_{\frac{\alpha}{2}}) = P(Z \leq z_{\frac{\alpha}{2}}) - P(Z \leq -z_{\frac{\alpha}{2}}) = 2P(Z \leq z_{\frac{\alpha}{2}}) - 1 = 2(1 - \frac{\alpha}{2}) - 1 = 1 - \alpha$. We used the def. of percentile as follows: $\frac{\alpha}{2} = P(Z > z_{\frac{\alpha}{2}}) = 1 - P(Z \leq z_{\frac{\alpha}{2}}) = 1 - \Phi(z_{\frac{\alpha}{2}})$, hence $\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$.

So, let us continue: $1 - \alpha = P(-z_{\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\frac{\alpha}{2}})$, so $1 - \alpha = P(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$, and hence a $100(1 - \alpha)\%$ confidence interval for μ is given by: $\overline{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$.

EXC: The weight of a population of aliens is normaly distributed with standard deviation $\sigma = 20$.

a) Find a 95% confidence interval for μ when $n = 10$ and $\overline{x} = 1000$;

b) Find a 95% confidence interval for μ when $n = 25$ and $\overline{x} = 1000$;

c) Find a 99% confidence interval for μ when $n = 10$ and $\overline{x} = 1000$;

d) Find a 99% confidence interval for μ when $n = 25$ and $\overline{x} = 1000$;

e) What can you say about the length of the CIs?

Sol: a) $(1-\alpha)100 = 95$, so $\alpha = 0.05$, but we need $z_{\frac{\alpha}{2}}$, hence we compute $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$. We know $z_{0.025} = 1.96$, hence the CI is given by: $\overline{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$, or: $1000 - 1.96\frac{20}{\sqrt{10}} \le \mu \le 1000 + 1.96\frac{20}{\sqrt{10}}$, or: $987.6 \le \mu \le 1012.4$; for part b) we just change n; so we get $1000 - 1.96\frac{20}{\sqrt{25}} \le \mu \le 1000 + 1.96\frac{20}{\sqrt{25}}$, or: $992.2 \le \mu \le 1007.8$;

c) $(1-\alpha)100 = 99$, so $\alpha = 0.01$, but we need $z_{\frac{\alpha}{2}}$, hence we compute $\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$. We know $z_{0.005} = 2.58$, hence the CI is given by: $\overline{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$, or: $1000 - 2.58\frac{20}{\sqrt{10}} \le \mu \le 1000 + 2.58\frac{20}{\sqrt{10}}$, or: $983.7 \le \mu \le 1016.3$; for part d) we just change n; so we get $1000 - 2.58\frac{20}{\sqrt{25}} \le \mu \le 1000 + 2.58\frac{20}{\sqrt{25}}$, or: $989.7 \le \mu \le 1010.3$;

e) If n increases, then the CI will get narrower; if the confidence level increases, then the CI gets wider!

NOTE: if an infinite number of random samples are considered and a $100(1-\alpha)\%$ CI for μ is calculated for each sample, then $100(1-\alpha)\%$ of these intervals will contain the true value of μ!

Def: The precision of a confidence interval is its length, i.e., $2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$.
Remarks: A shorter length is interpreted as an estimation more precise; it is a function of n (size) and of level of confidence; if we enlarge the confidence level, then the estimation is less precise; if we enlarge the size (n), then the estimation is more precise;

The error: $|\bar{x}-\mu|$ we commit by estimating μ by the sample mean is smaller than $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. If we want to control the error (i.e., $|\bar{x}-\mu| \leq E$), the only thing we can do is to choose the appropriate sample size: $n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$ or bigger.

EXC: 8-15/268 Here: $\sigma = 25$, $\overline{x} = 1014$, $95 = (1-\alpha)100$ implies

$\alpha = 0.05$, so $\frac{\alpha}{2} = 0.025$. The length is 6, so to get the error just divide by 2 (we are dealing with centered intervals). So $E = 3$. Since $z_{\frac{\alpha}{2}} = 1.96$ we get: $n = (\frac{1.96 \times 25}{E})^2 = 266.7$, so we choose $n = 267$.

8-2-3 One-sided confidence bounds/intervals

Def: A $100(1-\alpha)\%$ upper-confidence bound for μ is $\mu \leq \overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$

Def: A $100(1-\alpha)\%$ lower-confidence bound for μ is $\overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu$

Of course the intervals are given by: $(-\infty, \overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}}]$ and by $[\overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty)$

8-2-5 Large-sample confidence interval for μ (i.e., $n \geq 40$)

In many situations σ is not known, so we may use S to estimate σ. So the general rule is:

If $n \geq 40$, then $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ has an approximate standard normal distribution, so $\overline{x} - z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}$ is a large sample confidence interval for μ, with confidence level of approximately $100(1-\alpha)\%$. Here s is just the observed value of S.

At the blackboard we do 8-17, 8-18/page 268 and 8-1 a) and 8-2 a).

8-2 a) We have $98 = 100(1-\alpha) \Rightarrow 0.98 = 1-\alpha$, so $\alpha = 0.02$, hence $\frac{\alpha}{2} = 0.01$. We need $z_{\frac{\alpha}{2}} = z_{0.01}$. Recall the def of percentile: $0.01 = P(Z > z_{0.01}) = 1 - P(Z \leq z_{0.01})$, thus $P(Z \leq z_{0.01}) = 0.99$, looking at the table we choose 2.33.

8-1 a) Looking at the eq. in the statement we identify $z_{\frac{\alpha}{2}} = 2.14$; We need $1-\alpha$. Recall that $\frac{\alpha}{2} = P(Z > z_{\frac{\alpha}{2}}) = P(Z > 2.14)$, hence $\frac{\alpha}{2} = 1 - P(Z \leq 2.14) = 1 - 0.983823$, thus $\alpha = 2\{(1-0.983823)\} = 0.032354$, therefore $1-\alpha = 0.967646$, and if you want percentage: 96.7%.

8-17 The length is $L = 2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$. Let m the new size that does the trick! The new length of $\overline{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{m}} \leq \mu \leq \overline{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{m}}$ is $L' = 2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{m}}$. Since $\frac{L}{2} = L'$ we get that $z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{m}}$, then simplify and get $\sqrt{m} = 2\sqrt{n}$, so $m = 4n$, therefore we need to increase by $4 = 2^2$.

8-17 For size n the length is $L = 2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$. For size $2n$ the length is $L' = 2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{2n}}$. So: $L' = 2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\frac{1}{\sqrt{2}} = L\frac{1}{\sqrt{2}} = \frac{1}{1.414}L = 0.707L$, hence we reduce by $1 - 0.707 = 0.293$, or 29.3%.

Read pages 260-264! Next lecture we are going to dive into 8.3 and so on... Do not forget to bring a calculator, the textbook and the printed notes (the only objects allowed for the open book midterm!) And pen/pencil + eraser! Good luck!