Stat 213: Intro to Statistics 9
Confidence Interval and Sample Size

H. Kim

Fall 2007

statistical inference

- Need to have methods of statistical inference making that they are better and more reliable than just subject guesses
  - **point estimation**: a single number is calculated to estimate the population parameter, eg. $\overline{X}$
  - we would like to have point estimators, which satisfy
    1. unbiased
    2. the spread of the sampling distribution should be as small as possible.

  - If $E(\overline{X}) = \mu$, then the point estimator, $\overline{X}$, is said to be **unbiased**, otherwise it is biased estimator.

statistical inference

– **interval estimation**: two numbers (L, U) are calculated to form **an interval** within which the parameter (μ, p) is expected to lie, that is, for the mean, μ,

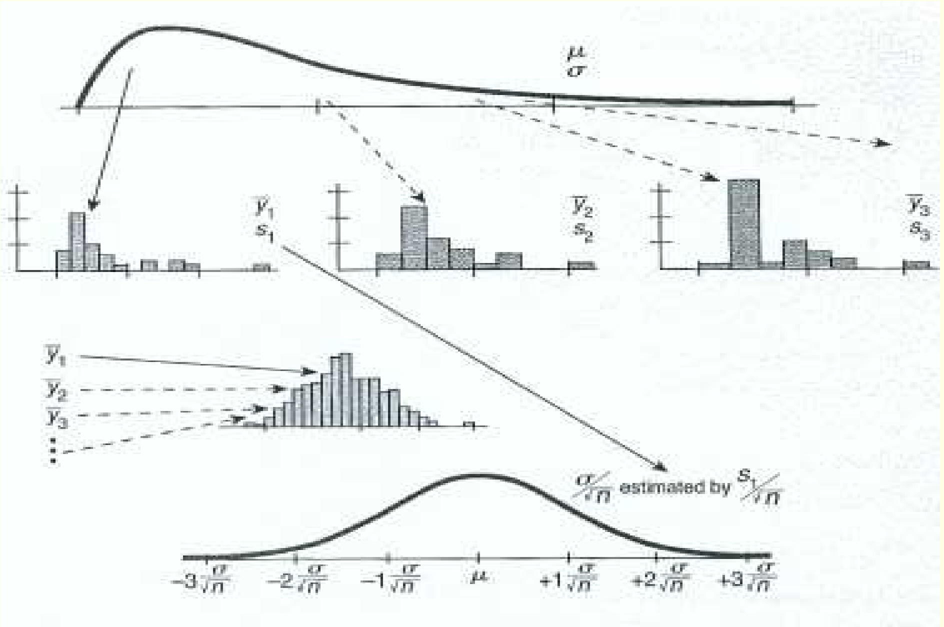
$$L < \mu < U$$

for the proportion, p,

$$L < p < U$$

$$\Big(\text{lower limit}(L), \quad \text{upper limit}(U)\Big)$$

recall: Central Limit Theorem



- Sampling distribution for $\overline{X}$ is Normal with mean μ, and standard error of $\frac{\sigma}{\sqrt{n}}$, but `still don't know` what is the true mean μ

critical values

- Because it is Normal (Empirical rule),
  - 95% of all these samples will be within $\big(\mu - 1.96\ se(\bar{X}),\ \ \mu + 1.96\ se(\bar{X})\big)$

- critical values: 1, 1.645, 1.96, 2.565
  - $z_{0.84} = 1 \quad \Longleftarrow P(Z < z_{0.84}) = 0.84$
  - $z_{0.95} = 1.645 \quad \Longleftarrow P(Z < z_{0.95}) = 0.95$
  - $z_{0.975} = 1.96 \quad \Longleftarrow P(Z < z_{0.975}) = 0.975$
  - $z_{0.995} = 2.565 \quad \Longleftarrow P(Z < z_{0.995}) = 0.995$

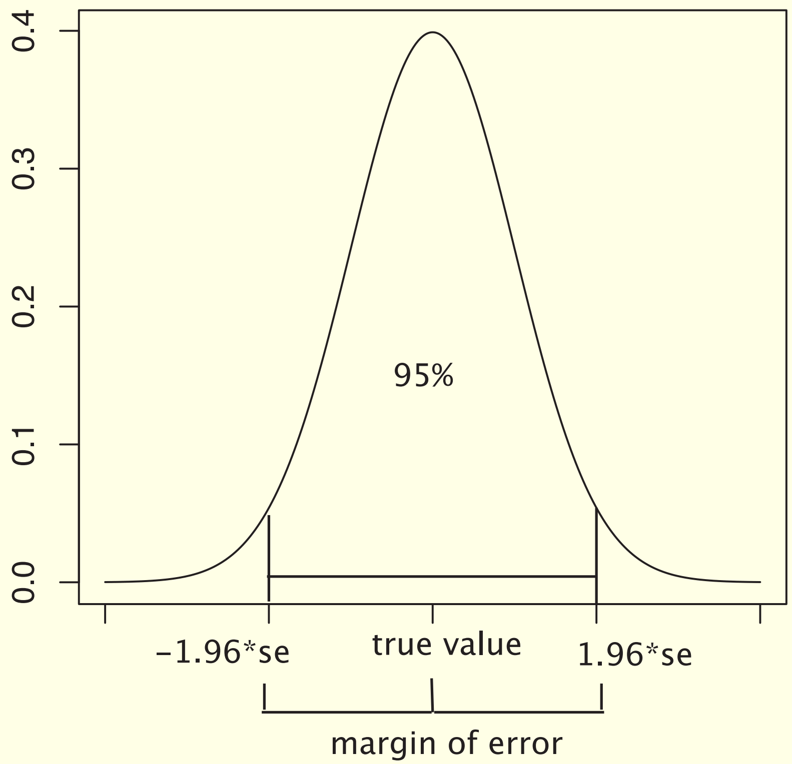
margin of error for 95% confidence

- For an unbiased estimator, the error of estimation (the distance between the an estimate and the parameter) will be less than 1.96 standard deviations with probability of 0.95.
- margin of error for the estimate of μ
  - $\pm 1.96 se(\overline{X}) = \pm 1.96 \frac{\hat{\sigma}}{\sqrt{n}}$

    if σ: unknown and $n \geq 30$, then sample standard deviation can be used to approximate σ
- margin of error for the estimate of $\hat{p}$
  - $\pm 1.96 se(\hat{p}) = \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

    if p: unknown and $np > 5, n(1-p) > 5$, then sample proportion can be used to approximate p.

margin of error for 95% confidence



95% `confidence interval for` μ

- From $\bar{X}$'s point of view, there is a 95% chance that μ is no more than $1.96se(\bar{X})$ away from $\bar{X}$, i.e.

$$P[-1.96se(\bar{X}) < \overline{X} - \mu < 1.96se(\bar{X})] \quad = \quad 0.95$$

$$P[-1.96 < \frac{\overline{X} - \mu}{se(\bar{X})} < 1.96] = 0.95$$

  but still don't know the true value

- the best we can do is an `interval`:

$$\big(\bar{X} - 1.96se(\bar{X}), \;\; \bar{X} + 1.96se(\bar{X})\big)$$

$(1-\alpha)100\%$ `confidence interval for` μ `or` p

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

$$\left(\bar{X} - z_{1-\alpha/2} se(\bar{X}),\ \ \bar{X} + z_{1-\alpha/2} se(\bar{X})\right)$$

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

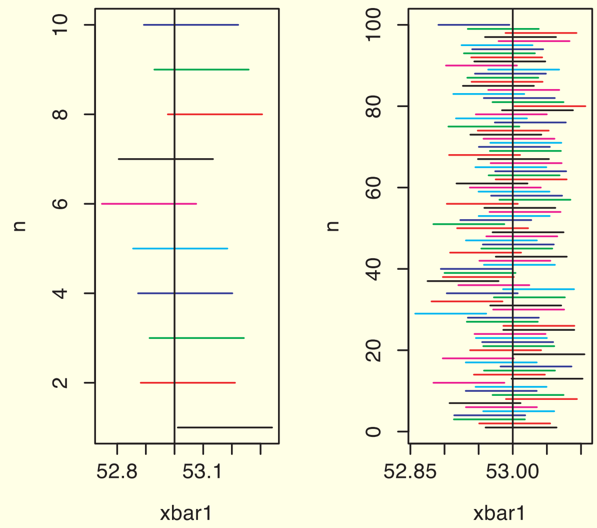
$$\left(\bar{X} - z_{1-\alpha/2} se(\hat{p}),\ \ \bar{X} + z_{1-\alpha/2} se(\hat{p})\right)$$

- `confidence coefficient` (level of confidence): $1-\alpha$

- If $\alpha = 0.05$,
  - $1-\alpha = 0.95$ and $1-\alpha/2 = 0.975$
  - $z_{1-\alpha/2} = z_{0.975} = 1.96 \qquad \Longleftarrow P(Z \leq z_{0.975}) = 0.975$

interpretation

- 95% confidence interval for μ:
  $\left(\bar{X} - 1.96 se(\bar{X}),\ \ \bar{X} + 1.96 se(\bar{X})\right) = \left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}},\ \ \bar{X} - 1.96\frac{\sigma}{\sqrt{n}}\right)$:
  "in repeated sampling from a population, we are sure that about 95% of the intervals would contain the true population mean (μ)"



remarks

- 95% confidence interval for μ:

$$\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \quad \bar{X} - 1.96\frac{\sigma}{\sqrt{n}}\right)$$

- 99% confidence interval for μ:

$$\left(\bar{X} - 2.565\frac{\sigma}{\sqrt{n}}, \quad \bar{X} - 2.565\frac{\sigma}{\sqrt{n}}\right)$$

  – the larger the confidence coefficient, the wider the confidence interval

  – the larger the sample size, the smaller the margin of error, the narrower the confidence interval

example

- A manufactory process produced ball bearings whose diameters are known to have standard deviation of 0.12 mm. Based on a random sample of 64 observations with a mean of 10.023. Find a 99% confidence interval for the population mean diameter of ball bearing produced by this process.

example

- A random sample of 985 were polled during a phonathon. Of those surveyed, 569 indicated that they intended to vote for the republican candidate in the upcoming election. Construct 90% confidence interval for p, the proportion of likely voters in the population who intend to vote for the republican. Based on this information, can you conclude that the candidate will win the election ?

sample size

- we should have some idea of how large a margin of error we need to be able to draw conclusions or detect a different we want to see

  1. determine the `standard error` (se) of the point estimator
  2. choose a `margin of error` (K) and a `confidence coefficient` $(1-\alpha)$
  3. solve $z_{1-\alpha/2} se(\overline{X}) = K$ for n
     * for the mean, μ: $z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}} = K$
     * for the proportion, p: $z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = K$

example

- The standard deviation of the population is known as 5.4711, and the associated standard error, at the 90 % level of confidence, is equal to 1.8. What is the minimum sample size ?

example

- How large a sample is needed to make a 99% confidence interval for the true population of bus riders in Calgary, with a maximum margin of error of 10% ?