Confidence intervals

Josemari Sarasola

Statistics for Business

Confidence interval

Point estimation and interval estimation

- To quantify parameters we can apply directly the formula of the estimator to sample data. We name this procedure a point estimation.
- Point estimation: $\mu? \to \hat{\mu} = \overline{x} = 4$
- Interval estimation: $\mu? \to \mu = 4 \pm 1 : 3 < \mu < 5$ with a 90% confidence level. ± 1 is the estimation error.
- Interval estimations give more information than point estimations.
- Notation: ϵ: estimation error; $1 - \alpha$: confidence level; n, sample size.

Interval about the population mean

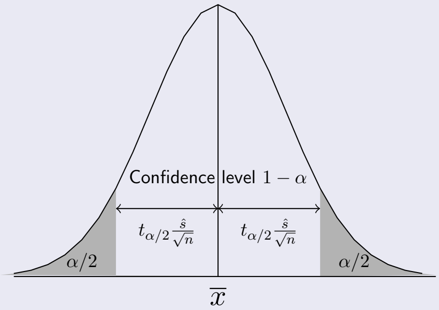
t interval

- Use with normal population, small sample size ($n < 30$) and unknown σ.
- We calculate $\overline{x}$ and $\hat{s}$.
- Sampling distribution: $\dfrac{\overline{x} - \mu}{\hat{s}/\sqrt{n}} \sim t_{n-1}$
- Interval for a $1 - \alpha$ confidence level:

$$I_{\mu,1-\alpha} : \overline{x} \pm t_{n-1,\alpha/2} \frac{\hat{s}}{\sqrt{n}}$$

Interval about the population mean

Schema of the interval



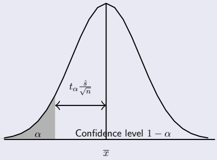
Properties of confidence intervals

They are also general properties for all kind of confidence intervals:

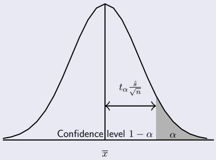
1. The bigger is n, the narrower is the interval (because we have more information)
2. The bigger is the deviation , the wider is the interval (because with larger deviations, the uncertainty in bigger).
3. The bigger is the confidence, the wider is the interval (if we want more confidence about something, we must widen its limits)

Asymmetric t intervals

- $I_{\mu,1-\alpha} : \mu > \overline{x} - t_{n-1,\alpha}\dfrac{\hat{s}}{\sqrt{n}}$



- $I_{\mu,1-\alpha} : \mu < \overline{x} + t_{n-1,\alpha}\dfrac{\hat{s}}{\sqrt{n}}$



Interval about the population mean

t interval with big samples ($n \geq 30$)

- Whenever $n \geq 30$, t_n (almost) becomes $N(0,1)$
- So, $I_{\mu,1-\alpha} : \overline{x} \pm z_{\alpha/2} \dfrac{\hat{s}}{\sqrt{n}}$

Interval about the population proportion

Population and sample proportions

- p: population proportion; e.g., proportion of voters for X party. It's generally unknown.

- $\hat{p}$: sample proportion; e.g., proportion of voters for X party in a survey. It's known: we ask 50 people, 20 say 'X party'; then $\hat{p} = \dfrac{20}{50} = 0.4$.

Interval about the population proportion

Construction of the interval

- Need big samples: $n \geq 30$.
- Sampling distribution: $\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$
- $I_{p,1-\alpha} : \hat{p} \pm z_{\alpha/2}\sqrt{\frac{pq}{n}}$
- But we dont' know either p or q. So we take their estimates: $\hat{p}$ and $\hat{q}$.
- Then,

$$I_{p,1-\alpha} : \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

The standard error

A key factor when we set a confidence interval is the standard deviation of the estimator, when the sampling distribution follows a normal distribution. We call that standard deviation the standard error.

For example, for the sample proportion the standard error is $\sqrt{\frac{pq}{n}}$.

If we want to set good (narrow) intervals, we need a small standard error. We can have a small standard error usually in two ways: clearly, taking a bigger sample size; and also choosing an estimator with a small standard deviation or standard error (in statistics, we say a efficient estimator).

Interval about the population proportion

Asymmetric intervals

- $I_{p,1-\alpha} : p > \hat{p} - z_\alpha \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$
- $I_{p,1-\alpha} : p < \hat{p} + z_\alpha \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$

Interval about the population proportion

Sample size

- Problem: calculate sample size for given ϵ error and $1 - \alpha$ confidence level in a symmetric interval.
- From the original interval:

$$\epsilon = z_{\alpha/2}\sqrt{\frac{pq}{n}} \rightarrow n = \frac{z_{\alpha/2}^2 pq}{\epsilon^2}$$

Sample size

- But we don't know p and q!
- First solution (pessimistic): the biggest needed sample size is when $p = q = 0.5$. So:

$$n = \frac{z_{\alpha/2}^2 \times 0.5 \times 0.5}{\epsilon^2} \rightarrow \frac{z_{\alpha/2}^2}{4\epsilon^2}$$

- Second solution: take a pilot sample (small and previous to the main sample) and estimate p and q, giving p_0 and q_0:

$$n = \frac{z_{\alpha/2}^2 p_0 q_0}{\epsilon^2}$$