Confidence Intervals

Lisa Kuramoto

Centre for Clinical Epidemiology and Evaluation

Overview

Outline

- Background
- Confidence intervals (CIs)
- Examples

Learning objectives

- to understand CI construction
- to be able to name 3 factors that affect CIs
- to be able to interpret CIs found in literature

Background

Terminology

- population

  group of individuals or objects that we would like to study

- sample

  subset of a population

Background

Terminology

- population parameter

  a quantity that describes a population

- sample statistic

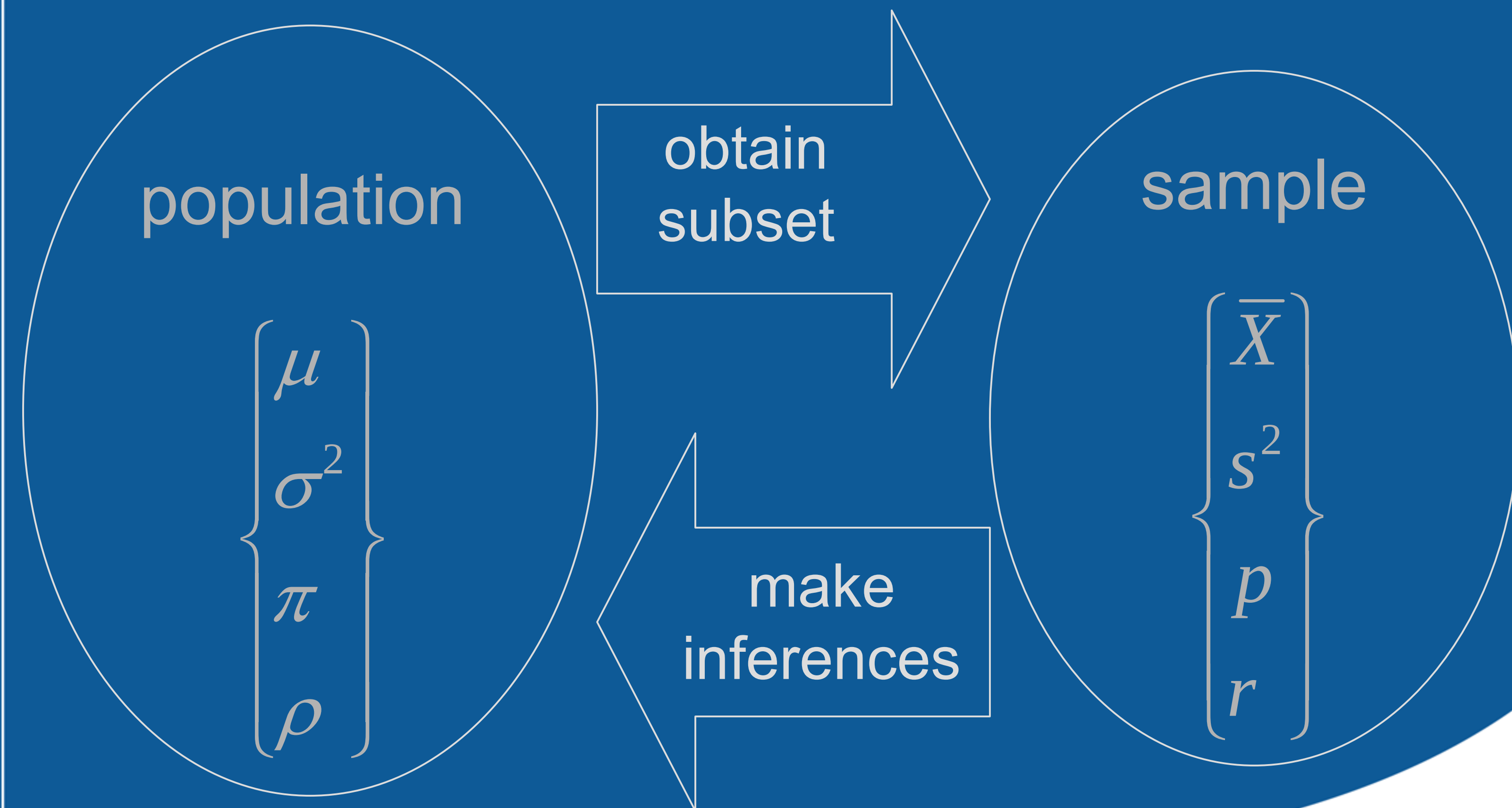
  an estimate of the population parameter

- statistical inference

  process of drawing conclusions about a population based on observations in a sample

Background

Framework for statistical inference



UBC

Vancouver Coastal Health Research Institute

Healthier lives through discovery

Background

Terminology

- point estimate

  a single value to estimate a population parameter

- interval estimate

  a range of values to estimate a population parameter

Confidence intervals

Motivation

We know:

- not practical to measure entire population, so take a random sample
- there is random error

We want to:

- summarize information on effect size and variability
- infer statistically significant results

Confidence intervals

Definitions

- confidence interval, CI

  a range of values that probably contains the population value

- confidence limits

  the values that state the boundaries of the confidence interval

Confidence intervals

Construction

- most CIs have the following form:

sample statistic +/- (critical value)x(SE of sample statistic)

"margin of error"

Confidence intervals

Construction

- the critical value represents the desired confidence level based on distribution theory
- the sample statistic is a point estimate based on data
- the SE of sample statistic is a measure of variability based on data

Confidence intervals

100(1-α)% CI for mean

- critical value: $z_{1-\alpha/2}$, 100(1-α/2)th percentile of standard normal distribution
- sample statistic: $\bar{x}$, sample average
- SE of sample statistic: $\hat{\sigma}/\sqrt{n}$, where $\hat{\sigma}$ is sample standard deviation and n is sample size

Vancouver Coastal Health Research Institute
Healthier lives through discovery

Confidence intervals

100(1-α)% CI for mean

$$\left(\bar{x} - z_{1-\alpha/2} \,\hat{\sigma}/\sqrt{n},\ \bar{x} + z_{1-\alpha/2} \,\hat{\sigma}/\sqrt{n} \right)$$

Confidence intervals

100(1-α)% CI for proportion (large sample size)

- critical value: $z_{1-\alpha/2}$, 100(1-α/2)th percentile of standard normal distribution
- sample statistic: $\hat{p}$, sample proportion
- SE of sample statistic: $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$,

and n is sample size

Confidence intervals

100(1-α)% CI for proportion (large sample size)

$$\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

UBC

Vancouver Coastal Health Research Institute
Healthier lives through discovery

Example: scenario

Construction example

Suppose that you would like to know the effect of a newly developed drug (drug A) and a current drug (drug B) on systolic blood pressure (sbp).

You would like to know the effect of drug A.

What information do we need?

Example

population

Q: What is the population parameter?

A: mean sbp, μ

Example

sample

Q: What is the sample statistic?

A: The average sbp among drug A and drug B patients was 107 mmHg and 125 mmHg, respectively.

So, $\overline{X} = 107$ mmHg.

Example

confidence level & critical value

Q: What is desired confidence level?

A: 95% CI, so $1-\alpha = 0.95$ and $\alpha = 0.05$

Q: What is the critical value?

A: $z_{1-\alpha/2} = 1.96$

Example

variability

Q: What was the variability of the data?

A: The estimated standard deviation of sbp among drug A and drug B patients was 19 mmHg and 20 mmHg, respectively.

Q: How many patients were sampled?

A: You took a random sample of 35 patients on drug A and 35 patients on drug B.

So, $n = 35$.

Example

variability

Q: What is the SE of the sample statistic?

A: $\hat{\sigma}/\sqrt{n} = 19/\sqrt{35} \approx 3.21\text{mmHg}$

Example

estimated CI

A 95% CI for the mean sbp of patients on drug A is

$$\left(\bar{x} - z_{1-\alpha/2} \times \hat{\sigma}/\sqrt{n}, \bar{x} + z_{1-\alpha/2} \times \hat{\sigma}/\sqrt{n}\right)$$

(107 - 1.96 x 3.21, 107 + 1.96 x 3.21) mmHg

= (100.7, 113.3) mmHg

Example: follow-up

Exercise

What is a 95% CI for mean sbp of patients on drug B?

Example: follow-up

estimated CI

A 95% CI for the mean sbp of patients on drug B is

$$\left(\bar{x} - z_{1-\alpha/2} \times \hat{\sigma}/\sqrt{n}\,,\, \bar{x} + z_{1-\alpha/2} \times \hat{\sigma}/\sqrt{n}\right)$$

$(125 - 1.96 \times 20/\sqrt{35},\ 125 + 1.96 \times 20/\sqrt{35})$ mmHg

= (118.4, 131.6) mmHg

Example

Graphing CIs

drug A

drug B

CIs and influencing factors

Factors that affect the width of a CI are:

- desired confidence level, 1-α
- sample size, n
- variability or standard deviation, σ

CIs and influencing factors

Desired confidence level, 1-α

- intuition: a higher confidence level without improving data quality means a larger margin of error
- as the desired confidence level increases, the confidence interval width increases , given all other quantities remain fixed

CIs and influencing factors

Sample size, n

- intuition: a larger sample size means more information, which implies better inference
- as the sample size increases, the confidence interval width decreases, given all other quantities remain fixed

CIs and influencing factors

Variability/Standard deviation, σ

- intuition: more variability or larger spread means more difficult to estimate population value without large amounts of data
- as the variability/standard deviation increases, the CI width increases , given all other quantities remain fixed

Interpretation

Thought experiment

- Imagine taking many samples of equal size and constructing 95% CIs

Interpretation

Thought experiment observations

- the population value is fixed
- some CIs contain the population value and some do not
- about 95% of the CIs contain the population value

Interpretation

Proper interpretation of 95% CI

- the probability that the CI contains the population value is 0.95
- "Our estimate is *sample estimate*. This result is accurate to within *margin of error*, 19 times out of 20."

Improper interpretation of 95% CI

- the probability that the population value lies within the CI is 0.95

Interpretation

Notes

- assume measurements are not biased (ie. no systematic error)
- statistical significance does not imply clinical significance

Literature

Berry, et al. (2003).

- population parameter?
- confidence interval?
- width of confidence intervals?
- overlap of confidence intervals?

References

- Berry, MJ, et al. (2003). A Randomized, Controlled Trial Comparing Long-term and Short-term Exercise in Patients with Chronic Obstructive Pulmonary Disease. Journal of Cardiopulmonary Rehabilitation. 23:60-68.
- Norman, GR and Streiner, DL. (2005). Biostatistics: The Bare Essentials 2E. Hamilton: BC Decker.

UBC

Vancouver Coastal Health Research Institute
Healthier lives through discovery