SOCI 326
Quantitative Methods in Social Research

Session 4 – Introduction to Estimation

Lecturer: Dr. Rabiu K. B. Asante, Sociology Dept
Contact Information: rkbasante@ug.edu.gh

UNIVERSITY OF GHANA
College of Education
School of Continuing and Distance Education
2014/2015 – 2016/2017

Session Overview

- In this session we will focus on one of the first major tools used for translating a sample feature into a population feature called estimations. The procedure for estimation differs based on the kind of variable we are dealing with. At every point in time we will make reference to the variable's level of measurement. In this session we narrow this down to variables measured at the interval-ratio level.

Session Outline

The outline for this session is as follows:

- Point estimate and interval estimate
- Logic of confidence interval and the estimation process
- Calculating the confidence level, alpha level and the standard error

Topic One

POINT ESTIMATE AND INTERVAL ESTIMATE

ESTIMATION

- This a branch of inferential statistics
- It aims at estimating population values (parameters) from sample values (**statistics)**
- Population values can be estimated by using point estimates (e.g. sample means, sample proportions) constructing confidence intervals
- “Confidence Interval (CI) is a mathematical statement that says that the parameter lies within a certain range of values or interval”

…ESTIMATION

- Example: CI estimates can be stated as ‘between 28% and 36% of Ghanaian households have a toilet facility

- The interval places the population value (the percentage of Ghanaian households who own toilet facilities) to be between 28% and 36%, but does not specify an exact value

- Estimation procedures are based on sample statistics

Point Estimate

- Point estimate: Using a sample statistic to represent the population parameter.

EXAMPLE: If the 70% of my sample are females does this imply that 70% of the population are females.

- What are the possible errors with this reasoning?

Interval Estimate

- Providing a range of scores, with reference to a statistic, as a possible interval that includes the parameter.

Which will be more accurate in estimating the population parameter and why? Point estimate or interval estimate?

The foundational formulae

$$\text{Point estimate} \pm \text{Confidence level (Standard Error)}$$

- Point estimate = $\bar{X} = \text{Sample mean}$

  or Sample Proportion = $P_S = \text{Sample Proportion}$

Topic Two

CONFIDENCE LEVEL

Confidence Level

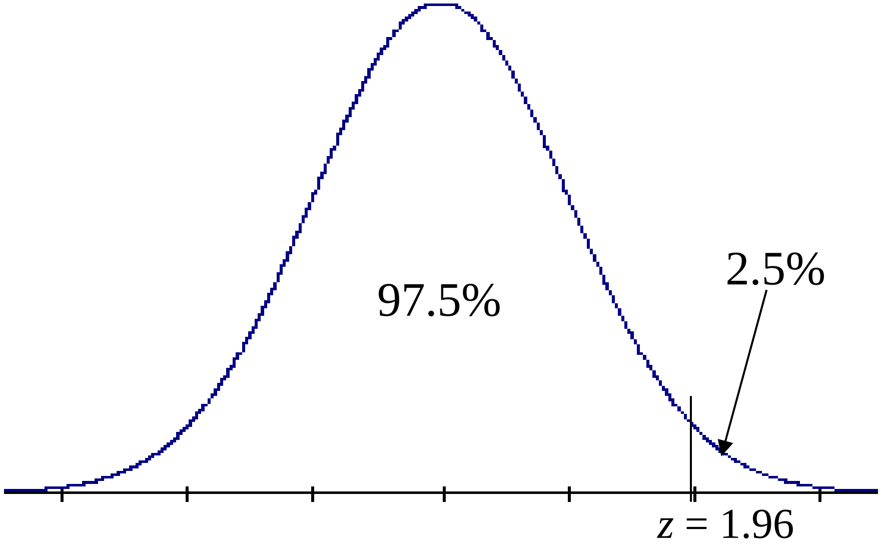
- How sure are you that the population value / parameter will fall in the range of scores
- It is often expressed in percentages:90%, 95% or 99%
- If you are 90% confident it can also mean you are 10% not sure (Error)

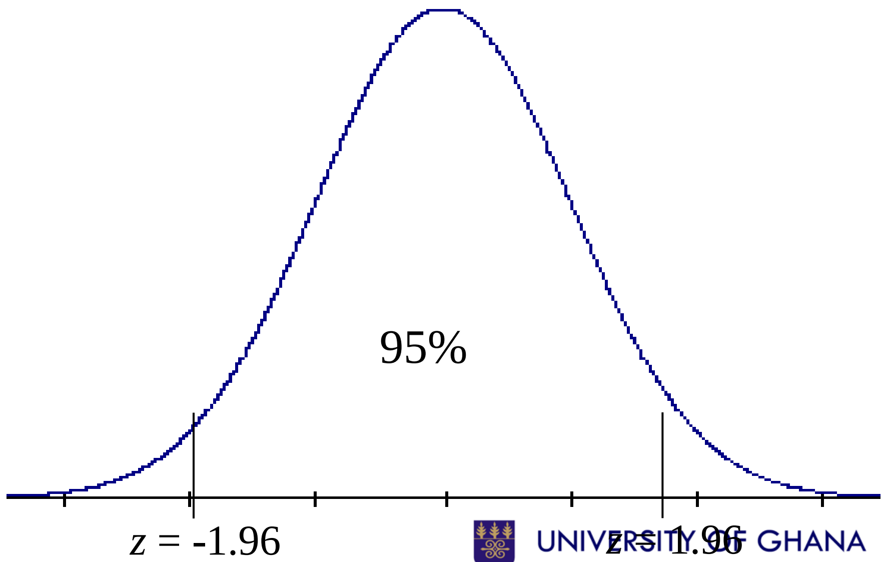
…Confidence Level

- The error is known as alpha level (α) : How unlikely is it that your population value will fall out of the range of scores
- Both the error and the confidence level is spread across both sides of the normal curve
- The error is expressed as a Z-score if your sample size is large (n =/< 100)

Finding the 95% point on a normal distribution…

- From the table we can see that when $z = 1.96$ the *p*-value, which represents the probability of being in the larger area (to the left), is 0.975.
- Therefore the area under one (small) tail of the curve is $p=0.025$.
- This means that scores greater than $z = 1.96$ occur just 2.5% of the time.
- Further (because the normal curve is symmetric) we can calculate that the area under both tails (beyond $z = 1.96$ and $z = -1.96$) is 0.05.
- In other words 95% of the area is in the middle, between $z = -1.96$ and $z = 1.96$
- And scores further from the mean than 1.96 thus only occur 5% of the time





Why 95%?

- A confidence interval need not be 95%
- However this is the generally accepted level for statistical testing. It is considered that errors occurring only 5% times are acceptable.
- Furthermore, a higher value can produce confidence intervals that may be viewed too wide (producing an unacceptable risk of Type I errors – discussed later).

Z SCORES FOR VARIOUS LEVELS OF ALPHA (α)

Confidence Level	Alpha (α)	α/2	Z Score
90%	0.10	0.0500	±1.65
95%	0.05	0.0250	±1.96
99%	0.01	0.0050	±2.58

Topic Three

STANDARD ERROR

Standard Error

- From Sampling distribution:
- Standard error is the standard deviation of the sampling distribution for sample means
- It was expressed in terms of the standard deviation of the population and the sample size

$$\frac{\sigma}{\sqrt{n}}$$

Standard error and the level of measurement

Levels of measurement and confidence interval approach to use:

- Nominal and ordinal : Estimate for Population Proportion

- Interval-ratio: Estimate for Population mean

Interval estimate for population mean (Part 1)

- Population standard deviation is known:

C.I.:

$$\bar{x} \pm Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Interval estimate for population mean (Part 2)

- When population standard deviation is unknown
- C.I. : $\bar{x} \pm Z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n-1}}\right)$

Interval estimate for Population Proportion

C.I: $$P_S \pm Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{P_U(1-P_U)}{n}}\right)$$

ESTIMATION PROCEDURES

Step 1:

- Identify your point estimate
- Know the level of measurement for your point estimate
- What kind of estimation are you required to perform

Step 2

- You have to decide on how much risk of being wrong you are willing to take
- The probability of error is called alpha (symbolized as α)
- Setting α to 0.05 also called using the 95% confidence level, means that the researcher is willing to be wrong only 5% of the time

… ESTIMATION PROCEDURES

Step 3:

- Divide the probability of error equally into the upper and lower tails of the distribution and then find the corresponding Z score
- Example:

$\alpha = 0.05$

$\alpha = 0.05/2$

$\alpha = 0.0250$

Z score = ± 1.96

… ESTIMATION PROCEDURES

Step 4:

- Construct the confidence interval by substituting the values into the formulae

Step 5:

- Interpret your results by stating the confidence level, the population value and the lower and upper limits

EXAMPLE 1

- **If we take a sample of 840 people and find that their mean score in an exam was 272 with a standard deviation of 60. We can construct a 95% confidence interval for the mean scores of the population?**

...Example

To estimate the unknown population mean (μ), use

- mean ($\bar{x}$) =272
- standard deviation of sample mean = s/√n-1 =60/ √840-1 =2.1
- Use the 95% confidence interval (Z= $\pm$1.96)

c.i. = $\bar{x} \pm Z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n-1}}\right)$

c.i. = 272$\pm$1.96(2.1)

c.i. = 272$\pm$4.1

c.i. = 272 $-$ 4.1 = 267.9

= 272 $+$ 4.1 = 276.1

Conclusion 1: I am 95% confident that the mean exam score of the population lies between 267.9 and 276.1

Conclusion 2: At 95% confidence level the mean exam score of the population is 272 (± 4.1 marks)

Sample Question

- A random sample of 100 television programs contained an average of 2.37 acts of physical violence per program. At the 99% level, what is your estimate of the population value?
  - *Sample mean* = 2.37
  - s = 0.30
  - N = 100

References

- Healey, J. F. (2010). ***Statistics: A Tool for Social Research.*** Wadsworth: CA pg 156 - 176