Estimation

LECTURE 8

Objectives

- Define terms.
- Distinguish between point and interval estimation.
- Calculate the confidence intervals means.
- Explain confidence intervals.
- Describe how confidence intervals of means can be narrowed (increase certainty).

Overview

- We typically estimate parameters, because we usually do not know them.
  - Why not?
- Two types of estimates that are commonly used.
  - Point estimates
  - Interval estimates

Point estimate

- A number obtained from the sample that approximates a particular parameter.
  - The sample mean ($\bar{x}$) is a point estimate of the population mean (μ).
- We want to known how reliable this estimate is. We can do this by using an interval estimate.

Interval estimate

- This is an interval that is defined by 2 numbers (lower and upper limits). We expect the interval to contain the value of the parameter.
  - The advantage of the interval estimate is that it shows how accurately the parameter is being estimated.
- These interval estimates are called confidence intervals.

Example

- We want to determine the effect of a new substance added to the diet of infants on their weight gain.
- The population mean is not known so it has to be estimated.

Example

- 32 infants were given the substance
- x=311.9 g
  - The sample mean is a point estimate of the population mean
  - It would be better to have an interval estimate.
    - It is unlikely that the population mean is exactly the same as the sample mean.

Example

$$\mu = \bar{x} \pm Z\sigma_{\bar{x}}$$

Example

- The weight gains of infants have been well studied for many years and we can assume that the population standard deviation is 180 g.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{180}{\sqrt{32}} = 31.8g$$

Example

- Assume that we want to be 95% (0.95) certain that the population mean lies within the interval, a 95% confidence interval.
- Table A
  - 100% - 95% = 5% (0.05)
  - Because table A only works on half of the distribution: 0.05/2 = 0.025 (two-tailed -- more on this later)

Example

- Because table A works on the portion of the distribution between the mean and another value: 0.50 - 0.025 = 0.475
- Find the Z-score (standard deviation units) equivalent to 0.4750 --- 1.96.

Example

$$\mu = \bar{x} \pm Z\sigma_{\bar{x}}$$

$$\mu = 311.89 \pm 1.96(31.8)$$

- Lower limit = 311.89-(1.96)*(31.8)=249.5 g
- Upper limit = 311.89+(1.96)*(31.8)=374.2 g
- 95% confidence interval is 250-374 g

Interpreting Confidence Intervals

- Using the results obtained above.
- Typically this is interpreted as there is a 95% chance that the true population mean (µ) is between 249.5 -374.2 g.
  - This is not correct – though it is often stated.
- Correct interpretation: If we took many samples of the population and calculated confidence intervals in a similar manner, 95% of the calculated confidence intervals would contain the true population mean(µ).

Confidence intervals

- In general, narrow confidence intervals are more desirable than wide ones (there is a smaller range of uncertainty).

Confidence intervals

- How can you reduce the width of the interval?
  - Reduce the confidence level (e.g. 99% to 95% or 95% to 90%), this may not be desirable
    - Most investigators do not want to use a level lower than 90% (too great a chance of missing the population mean)
  - Reduce the standard deviation.
    - We ordinarily can't do this if we are sampling from a population in nature.
    - By controlling the environment in a lab or in a greenhouse, we might be able to reduce the standard deviation.
  - Increase the sample size (most common way to do it)
    - The greater the sample size, the smaller the standard error.

The population standard deviation is not known

- Usually we do not know the population standard deviation. So we estimate the population standard error by calculating from the sample standard deviation.

$$s_x = \frac{s}{\sqrt{n}}$$

The population standard deviation is not known

- The resulting quantity is based on the t-distribution (as opposed to the normal distribution or Z-score).

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

The population standard deviation is not known

- We can transpose the formula so

$$\mu = \bar{x} \pm t_{\alpha,2,\nu 1} s_{\bar{x}}$$

The population standard deviation is not known

- What does this notation actually mean?

$$t_{\alpha,2,\nu 1}$$

Example

- Example 6