Estimation with Unknown Variance

Introduction and Overview

Prof Hans Georg Schaathun

Høgskolen i Ålesund

12th February 2014

Confidence interval for the mean

$$\bar{X} - z_{\alpha/2} \cdot \sigma_{\bar{X}} \leq \mu \leq \bar{X} + z_{\alpha/2} \cdot \sigma_{\bar{X}}$$

$$\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

- Two assumptions:
  1. σ is known
  2. Large n (i.e. we can use $z_{\alpha/2}$ from the normal distribution)
- What if σ is unknown?
- Can we estimate σ?
  1. Different distributions require different estimators.

HØGSKOLEN I ÅLESUND
Aalesund University College

The Sample Standard Deviation

- Sample variance: $s^2 = \sum \frac{(x_i - \bar{x})^2}{n-1}$
- Sample standard deviation: $s = \sqrt{\sum \frac{(x_i - \bar{x})^2}{n-1}}$
- s is a good estimator for σ
- For large n, replace $\sigma \mapsto s$

Confidence Interval for μ when n is large:

$$\bar{X} - z_{\alpha/2} \cdot s/\sqrt{n} \leq \mu \leq \bar{X} + z_{\alpha/2} \cdot s/\sqrt{n}$$

The Binomial Proportion

- $X \sim B(n, p)$
- We can view p as a population mean
  1. $X = X_1 + X_2 + X_3 + \ldots + X_n$, where $X_i \in \{0, 1\}$
  2. The point estimator is $\hat{p} = X/n = \bar{X}$
- But s is a not a good estimator for σ

Standard Deviation of the Binomial Distribution

1. $X \sim B(n, p)$
2. Recall $\mathrm{var}(X) = n \cdot p(1 - p)$
   - $\mathrm{std.dev.}(X) = \sqrt{n \cdot p(1 - p)}$
3. $\mathrm{std.dev.}(\hat{p}) = \mathrm{std.dev.}\left(\frac{X}{n}\right) = \sqrt{\frac{p(1-p)}{n}}$
4. But p is unknown – what do we do?

Standard Deviation of the Binomial Distribution

- $X \sim B(n, p)$

$$\hat{p} - z_{\alpha/2} \cdot \sigma_{\hat{p}} \leq p \leq \hat{p} + z_{\alpha/2} \cdot \sigma_{\hat{p}}$$

1. $\hat{p} = x/n$
2. $z_{\alpha/2}$ is ok when n is large
3. We need to estimate $\sigma_{\hat{p}}$

Confidence Interval for the Binomial Proportion p:

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Confidence intervals for large samples

$$\hat{\theta} - z_{\alpha/2} \cdot \sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$$

1. Estimate μ in any distribution

$$\bar{X} - z_{\alpha_2} \cdot s/\sqrt{n} \leq \mu \leq \bar{X} + z_{\alpha_2} \cdot s/\sqrt{n}$$

2. Estimate p in the binomial distribution

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$

3. To come:
   - Handling small samples with Student's t-distribution