Capítulo 7

Confidence intervals

7.1. Introduction

Definition 7.1.1 (Confidence interval)**.** *Let $X_1, \ldots, X_n, \ldots$ be a sequence of independent identically distributed random variables with the same distribution as $X \sim F \in \mathcal{F}_M$. Let $\theta = \Phi(F)$ be the parameter of interest. Consider two statistics*

$$A_n = A_n(X_1, \ldots, X_n),\ B_n = B_n(X_1, \ldots, X_n)$$

We say that the interval (A_n, B_n) is a $(1-\alpha)$ confidence interval for θ if

$$\Pr(A_n \leq \theta \leq B_n) = 1 - \alpha$$

for all $F \in \mathcal{F}_M$. We write

$$\mathrm{CI}_{1-\alpha}(\theta) \equiv (A_n, B_n).$$

Example 7.1.1.

Confidence interval for μ when $X \sim N(\mu, \sigma^2)$ and σ^2 is known.

- Complete derivation of the confidence interval

$$\mathrm{CI}_{1-\alpha}(\mu) \equiv \left(\bar{X}_n \mp z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right).$$

- Meaning of $\mathrm{CI}_{1-\alpha}(\mu) = (A_n, B_n)$, a random interval.
- Meaning of $\mathrm{CI}_{1-\alpha}(\mu) = (a_n, b_n)$, a realization of the random interval (A_n, B_n).

□

7.2. Confidence intervals: A “general” method

Assume that there exist an estimator $\hat{\theta}$ of the parameter θ such that

$$\hat{\theta} \sim N(\theta, \text{s.e.}(\hat{\theta}))$$

with known standard error s.e.$(\hat{\theta})$. Then

$$\left(\hat{\theta} \mp z_{\alpha/2}\, \text{s.e.}(\hat{\theta})\right)$$

is a $(1-\alpha)$ confidence interval for θ.

Example 7.2.1.

Confidence interval for the difference of means in two independent samples, normal case, known variances:
Assume that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, and that σ_X^2 and σ_Y^2 are known. Let $\theta = \mu_X - \mu_Y$ be the parameter of interest.

Let $X_1, \ldots, X_n$ and X, $Y_1, \ldots, Y_m$ be two independent simple random samples of X and Y, respectively. Then

$$\bar{X}_n \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right), \bar{Y}_m \sim N\left(\mu_Y, \frac{\sigma_Y^2}{m}\right)$$

and $\bar{X}_n$ and $\bar{Y}_m$ are independent. Therefore

$$\bar{X}_n - \bar{Y}_m \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right).$$

Applying the general method,

$$\text{CI}_{1-\alpha}(\mu_X - \mu_Y) \equiv \left(\bar{X}_n - \bar{Y}_m \mp z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right).$$

□

7.3. Formulas for confidence intervals

- Confidence level $1 - \alpha$, level of significance α.
- $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, sample mean
- $S^2 = \frac{1}{n-1}\sum_{i=1}(X_i - \bar{X})^2$, sample variance.
- $S_p^2 = \frac{(n_X-1)S_X^2+(n_Y-1)S_Y^2}{n+m-2}$, pooled sample variance of two samples.
- z_a is the $(1-a)$–quantile of $N(0,1)$
- t_a is the $(1-a)$–quantile of t–Student

Assumptions	Parameter	Statistic	Distribution	$(1-\alpha)$ confidence interval
$X \sim N(\mu, \sigma^2)$ σ^2 known	Mean μ	$\bar{X}$	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$
$X \sim N(\mu, \sigma^2)$ σ^2 unknown	Mean μ	$\bar{X}$	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t$–Student, $n-1$ df	$\bar{X} \pm t_{\alpha/2}S/\sqrt{n}$
$X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, σ_X^2, σ_Y^2 known, two independent samples	Difference of means $\mu_X - \mu_Y$	$\bar{X} - \bar{Y}$	$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}} \sim N(0,1)$	$\bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$
$X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, $\sigma_X^2 = \sigma_Y^2$ unknown, two independent samples	Difference of means $\mu_X - \mu_Y$	$\bar{X} - \bar{Y}$	$\frac{\bar{X}-\bar{Y}}{S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}} \sim t$–Student, $n_x + n_y - 2$ df	$\bar{X} - \bar{Y} \pm t_{\alpha/2}S_p\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$
X r.v. $\mathbb{E}(X) = \mu$ $\mathrm{Var}(X) = \sigma^2$ unknown, large sample size n	Mean μ	$\bar{X}$	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \underset{\text{Ass}}{\sim} N(0,1)$	$\bar{X} \pm z_{\alpha/2}S/\sqrt{n}$
X, Y r.v. $\mathbb{E}(X) = \mu_X$, $\mathbb{E}(Y) = \mu_Y$ $\mathrm{V}(X) = \sigma_X^2$, $\mathrm{V}(Y) = \sigma_Y^2$ unknown, two independent samples of large sizes n_X, n_Y	Difference of means $\mu_X - \mu_Y$	$\bar{X} - \bar{Y}$	$\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}}} \underset{\text{Ass}}{\sim} N(0,1)$	$\bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}$
$X \sim \mathrm{B}(n, p)$	Proportion p	$\hat{p} = \frac{X}{n}$	$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
$X \sim \mathrm{B}(n_X, p_X)$, $Y \sim \mathrm{B}(n_Y, p_Y)$ large values of n_X, n_Y	Difference of proportions $p_X - p_Y$	$\hat{p}_X - \hat{p}_Y$ $= \frac{X}{n_X} - \frac{Y}{n_Y}$	$\frac{\hat{p}_X-\hat{p}_Y}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X}+\frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}} \sim N(0,1)$	$\hat{p}_X - \hat{p_Y} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}$
$X \sim N(\mu, \sigma^2)$	Variance σ^2	S^2	$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$	$\left[\frac{(n-1)S^2}{\chi_{\alpha/2}^2}, \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2}\right]$
$X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$	Quotient of variances σ_X^2/σ_Y^2	$\frac{S_X^2}{S_Y^2}$	$\frac{S_X^2/\sigma^2 X}{S_Y^2/\sigma_Y^2} \sim F(n_x - 1, n_Y - 1)$	$\left[\frac{S_X^2/S_Y^2}{F_{\alpha/2}(n_x-1,n_Y-1)}, \frac{S_X^2/S_Y^2}{F_{1-\alpha/2}(n_X-1,n_Y-1)}\right]$

Cuadro 7.1: Summary of statistics and confidence intervals.