ProbStats L14b

Confidence Intervals

Margin of Error

March 30, 2023

Sample $\overset{RV}{X_1}, X_2, \ldots X_n \underset{iid}{\sim} f(\theta)$

Estimator $\overset{R.V.}{\hat{\theta}} = T(X_1, \ldots X_n)$

unbiased $E[\hat{\theta}] = \theta$ ← const.

Confidence Interval $100(1-\alpha)\%$

e.g. $\alpha = 0.05$ ↳ 95% conf int

RVs

$$\Pr(L_n \le \theta \le R_n) = 1 - \alpha$$

$L_n = \bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

$R_n = \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

1 std

$z_{0.025} = 1.96$

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, I)$$

$\text{stdev}(\bar{X}_n)$

$\sigma = \text{stdev}(X_i)$

Table

α	$Z_{\alpha/2}$
0.05	1.96
0.01	2.58
0.005	$\Rightarrow$ 3...

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

$$\frac{\sigma}{\sqrt{n}} Z_n = \bar{X} - \mu$$

$$\mu = \bar{X} - \frac{\sigma}{\sqrt{n}} Z_n$$

Random Variables

not scalar quantities

L_n R_n

Repeat 20 times for 95% conf inter

l_{n_1} l_{n_2} θ r_n r_{n_2}

Understand mean Snowpack > 700 ft in Wasatch

Sample n locations $\boxed{n=40}$

$X_1, \ldots, X_n$

We know variance $\boxed{\sigma^2 = 36 \text{ inches}^2}$??

3600 in^2

Compute $\boxed{\bar{X}_n = 620 \text{ inches}}$

95% confidence interval. average snowpack μ

$\alpha = 0.05$

$z_{\alpha/2} = z_{0.025} = 1.96$

[618, 622] inches

$r_n = \bar{X}_n + z_{0.025} \cdot \frac{\sigma}{\sqrt{n}}$

$= 620 + (1.96)(60/\sqrt{40})$

≈ 622 inches

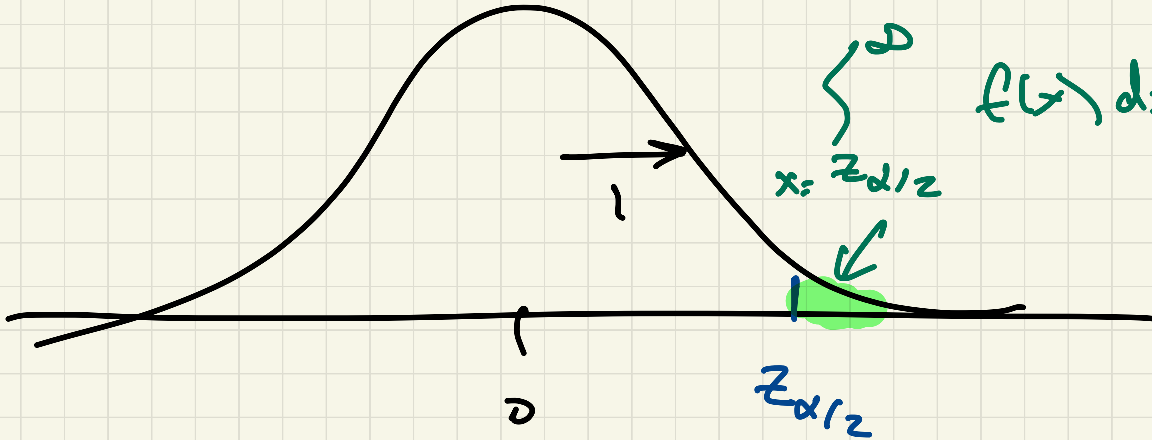
630

$\ell_n = \bar{X}_n - z_{0.025} \cdot \frac{\sigma}{\sqrt{n}}$

610

$= 620 - (1.96)(60/\sqrt{40}) \approx 618$ inches

$$Z \sim N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = f(x)$$



$$\int_{x=z_{\alpha/2}}^{\infty} f(x)\,dx = \frac{\alpha}{2}$$

Margin of Error in Polls

Ask n random likely voters

$$\left.\begin{array}{ll} 1 & \text{if vote for Cox} \\ 0 & \text{if not} \end{array}\right\}$$ Bernoulli R.V.

$X_i \sim Ber(p)$

How many voters said yes for Cox

$$S = \sum_{i=1}^{n} X_i$$ Binomial

$$S \sim Bin(n, p)$$

$$E[S] = np$$

$$Var[S] = np(1-p)$$

$$\left(\bar{X}_n = \right) \frac{S}{n} \approx N\left(p, \frac{p(1-p)}{n}\right)$$

95% confidence interval $\quad N\left(p, \frac{(1-p)\cdot p}{n}\right)$

$$L_n = \bar{X}_n - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \qquad p(1-p) \le 0.25, \quad p\in[0,1]$$

$$= \bar{X}_n - \underbrace{1.96\sqrt{\frac{0.25}{n}}}_{\text{margin of error}} \quad (100\,\%$$

$n=100$

$$1.96\cdot\frac{(0.5)}{\sqrt{100}} = (1.96)\cdot\frac{(0.5)}{10} = 0.098\ (100\%)$$

$$= 9.8\%$$

---

margin-of-error

3% $\Longrightarrow n=1067$

Confidence Intervals

pair statistics from sample unknown parame

RVs L_n, R_n $X_1, \dots, X_n \sim f(\theta)$

$$\Pr(L_n \le \theta \le R_n) = 1 - \alpha$$

$(1-\alpha)100\%$ confidence interval

Confidence Interval for mean

$$L_n = \bar{X}_n - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$R_n = \bar{X}_n + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

		Interval
increase	α	↘
"	n	↘
"	σ	↗