ESTIMATION

Large sample confidence intervals

Large sample confidence intervals

approx

- Next, we will consider the case in which we have an estimator $\hat{\theta}$ of an unknown parameter θ, and we know that the distribution of $\hat{\theta}$ is approximately normal with mean θ and variance $\sigma_{\hat{\theta}}^2$.
- How to construct a confidence interval for θ?

Know $\hat{\theta} \,\dot{\sim}\, N(\theta, \sigma_{\hat{\theta}}^2)$

$\hat{\theta} - \theta \,\dot{\sim}\, N(0, \sigma_{\hat{\theta}}^2)$

$\dfrac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \,\dot{\sim}\, N(0, 1)$

Pivot:

$P\left(-z_{\alpha/2} < \dfrac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} < z_{\alpha/2}\right) \approx 1 - \alpha$

$P\left(-\sigma_{\hat{\theta}} z_{\alpha/2} < \hat{\theta} - \theta < \sigma_{\hat{\theta}} z_{\alpha/2}\right) \approx 1 - \alpha$

$P\left(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} < \theta < \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right) \approx 1 - \alpha$

Kinda like P.Q. (If $\sigma_{\hat{\theta}}$ were known it would be a P.Q.)

When the standard deviation is not known

- Note: Sometimes $\sigma_{\hat{\theta}}$, the standard deviation of the distribution of $\hat{\theta}$ depends on unknown parameters.
- In such a case, we will replace the unknown parameters by their estimates.

estimate of

- If the estimator $\hat{\sigma}_{\hat{\theta}}$ tends to be "close" to $\sigma_{\hat{\theta}}$, then

$$\left(\hat{\theta} - z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}, \hat{\theta} + z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}\right)$$

is an approximate $100(1-\alpha)\%$ confidence interval for θ.

$\hat{\sigma}_{\hat{\theta}}$ "standard error"

$\hat{\theta} \pm z_{\alpha/2}\, \hat{\sigma}_{\hat{\theta}}$

Four common situations

- Earlier we saw four common estimation situations:
    1. Estimating μ with $\bar{Y}$ ("one sample problem")
    2. Estimating p with $\hat{p}$ ("one sample Bernoulli problem") $\hat{p} = \#\text{success}/n$
    3. Estimating $\mu_1 - \mu_2$ with $\overline{Y_1} - \overline{Y_2}$ ("two independent samples")
    4. Estimating $p_1 - p_2$ with $\hat{p}_1 - \hat{p}_2$ ("two independent Bernoulli samples")
- By the CLT, each of these estimators has an approximate normal distribution when n is large (when n_1 and n_2 are both large).

Derivation of (approximate) confidence intervals

Situation 1: Estimating μ with $\bar{Y}$

$Var(\bar{Y}) = \frac{\sigma^2}{n}$ estimate σ with s

Approx P.Q. $\frac{\bar{Y}-\mu}{s/\sqrt{n}} \overset{\cdot}{\sim} N(0,1)$

$\bar{Y} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$

Situation 2: Estimating p with $\hat{p}$

$Var(\hat{p}) = \frac{p(1-p)}{n}$ depends on p ← estimate p with $\hat{p}$

$\hat{p} \pm z_{\alpha/2} \frac{\hat{p}(1-\hat{p})}{n}$

$Var(Y) = np(1-p)$

$Var\left(\frac{Y}{n}\right) = \frac{1}{n^2} np(1-p)$ $Y \sim binom(n,p)$

$\frac{p(1-p)}{n}$

independent samples

don't know

Situation 3: Estimating $\mu_1 - \mu_2$ with $\overline{Y_1} - \overline{Y_2}$

$$Var(\bar{Y}_1 - \bar{Y}_2) = Var(\bar{Y}_1) + Var(\bar{Y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\bar{Y}_1 - \bar{Y}_2 \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

s_1 est. σ_1

s_2 est σ_2

indep

Situation 4: Estimating $p_1 - p_2$ with $\hat{p}_1 - \hat{p}_2$

$$Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$