Large sample confidence intervals

Let $\hat{\theta}$ be an estimator of θ. Let $\sigma^2_{\hat{\theta}}(\theta) = Var_\theta(\hat{\theta})$ be the variance of $\hat{\theta}$. For many estimators, $Z = \dfrac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}(\theta)}$ has a asymptotically normal distribution. Since θ is unknown is estimated by $\hat{\sigma}_{\hat{\theta}} = \hat{\sigma}_{\hat{\theta}}(\hat{\theta})$. This can be used to obtain large sample confidence intervals.

$$\lim_{n \to \infty} P\left(-z_{\frac{\alpha}{2}} \leq \frac{\hat{\theta} - \theta}{\hat{\sigma}_{\hat{\theta}}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

An asymptotic $1-\alpha$ confidence interval for θ is $\hat{\theta} \pm z_{\frac{\alpha}{2}} \hat{\sigma}$

$$\hat{\theta} - z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{\theta}}$$

Binomial confidence intervals

Let Y be a r.v. with a binomial distribution with parameters n and p. n is known, p is unknown.

$E[Y] = np \qquad Var(Y) = np(1-p)$

By the CLT, $\dfrac{Y-np}{\sqrt{np(1-p)}}$ has an asymptotic normal distribution

To estimate p, we use $\hat{p} = \frac{Y}{n}$, $E[\hat{p}] = p$, $Var(\hat{p}) = \frac{p(1-p)}{n}$

We estimate $\sigma_{\hat{p}}$ by $\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

By the CLT $\dfrac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ has an asymptotic normal distribution

So, $\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ is an asymptotic $1-\alpha$ confidence interval for p.

8.45 A New York Times / CBS survey of 224 children of ages 9 to 17 was conducted shortly after the January 28, 1986 space shuttle disaster. Two thirds of the children said that they would like to travel in space

(a) Construct a 90% confidence interval for the proportion of children ages 9 to 17 in 1986 who would like to experience space travel

(b) Do you think that "most" children in this age group think that they would like to experience space travel? Why?

(a) Y = no of children out of sample of 224, who would like to experience space travel

$Y \overset{d}{\sim} \text{Bin}(n = 224, p)$, p unknown.

$\hat{p} = \frac{Y}{n} = \frac{2}{3}$ $\qquad \hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{1}{224} \cdot \frac{2}{3} \cdot \frac{1}{3}} = \sqrt{\frac{1}{1008}}$

A 90% confidence interval for p is

$\hat{p} \pm z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{p}} = \frac{2}{3} \pm (1.645) \frac{1}{\sqrt{1008}} = 0.6666 \pm 0.0518$

$= [0.6148, 0.7185]$

(b) Yes, because the confidence interval is above 0.50

8.42 "Athletes at major colleges graduated, on the whole, at virtually the same rate as other students", according the NCAA. Suppose that, in a new poll of 500 athletes at major colleges the number graduating was 268.

Find a 98% confidence interval for p, the proportion of athletes at major colleges who graduate

$\hat{p} = \frac{Y}{n} = \frac{268}{500} = 0.536$, $\sigma_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{n} = \frac{1}{500} \frac{268}{500}\left(1-\frac{268}{500}\right)$

$= 0.0004974$ $\quad \hat{\sigma}_{\hat{p}} = 0.0223$

$\hat{p} \pm z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{p}} = 0.536 \pm 2.326(0.0223) = (0.4841, 0.5879)$

Binomial confidence intervals - Two samples

Assume that

$Y_1 \sim Bin(n_1, p_1)$, n_1 known, p_1 unknown

$Y_2 \sim Bin(n_2, p_2)$, n_2 known, p_2 unknown

Y_1 and Y_2 are independent r.v's

We would like to estimate $\theta = p_1 - p_2$

Let $\hat{p}_1 = \frac{Y_1}{n_1}$ $\hat{p}_2 = \frac{Y_2}{n_2}$, $\hat{\theta} = \hat{p}_1 - \hat{p}_2$

$\hat{\theta} = \hat{p}_1 - \hat{p}_2$ estimates $p_1 - p_2$

$E[\hat{p}_1 - \hat{p}_2] = p_1 - p_2$

$Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} = \sigma^2_{\hat{\theta}}$

We estimate $\sigma^2_{\hat{\theta}}$ by

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

An asymptotic $1-\alpha$ confidence interval for $p_1 - p_2$ is

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

8.51 For a comparison of the rates of defectives produced by two assembly lines, independent random samples of 100 items are selected from each line. Line A yielded 18 defectives in the sample, and line B yielded 12 defectives. Find a 98% confidence interval for the true difference in proportions of defectives for the two lines. Is there evidence here to suggest that one line produces a higher proportion of defectives than the other?

Y_1 = no. of defectives in Line A with probability of defective p_1

Y_2 = no. of defectives in line B with probability of defective p_2

To estimate $p_1 - p_2$ we use $\frac{Y_1}{n_1} - \frac{Y_2}{n_2} = \frac{18}{100} - \frac{12}{100} = 0.06$

$$V(p_1-p_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} = \frac{0.18(1-0.18)}{100} + \frac{0.12(1-0.12)}{100}$$

$= 0.002532$

$z_{\frac{\alpha}{2}} = z_{0.01} = 2.33$ The confidence interval is

$0.06 \pm 2.33\sqrt{0.002532} = 0.06 \pm 0.1172 = (-0.0572, 0.1772)$

No, there is no evidence to suggest one line produces a higher proportion of defectives than the the other