Thursday, November 06, 2014 2:28 PM

Confidence Interval

$$P\left(\underline{\theta} \in \left[\underline{\underline{L}}(x) .. \underline{\underline{U}}(x) \right] \right) \geq 1-\alpha$$

Pivot Exact Pivot

~~A~~symptotical Pivot

$X_i \sim$ ~~N~~Bernoulli (p) $\hat{p} = \frac{\Sigma x_i}{n}$

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim Z.$$ CI: $\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

7.22. VST Variance Stabalization Transform

$X_i \sim$ Bernoulli (p)

$\hat{p} = \frac{\Sigma X_i}{n}$ $E\hat{p} = p.$ $V(\hat{p}) = \frac{p(1-p)}{n}$

Assume that X has the mean μ. and $V(X) = V(\mu)$,

then $g(x) = \int_0^x \frac{1}{\sqrt{V(\mu)}} d\mu$

Let $Y = g(x)$.

$X = \hat{p}$ $V(\hat{p}) = \boxed{\frac{p(1-p)}{n}}$

$$g(x) = \int_0^x \frac{1}{\sqrt{\frac{p(1-p)}{n}}} dp = \sqrt{n} \int_0^x \frac{1}{\sqrt{p(1-p)}} dp$$

$$= \sqrt{n} \ 2 \arcsin \sqrt{x}.$$

[illegible]

$= \sqrt{n} \cdot 2 \arcsin\sqrt{x}$.

$$2\left(\arcsin\sqrt{x}\right)' = 2\cdot\frac{1}{\sqrt{1-(\sqrt{x})^2}}\cdot\frac{1}{2}x^{-\frac{1}{2}} = 2\cdot\frac{1}{2}\cdot\frac{1}{\sqrt{x(1-x)}} = \frac{1}{\sqrt{x(1-x)}}$$

VST. Let $Y = g(\hat{p}) = 2\sqrt{n}\arcsin\sqrt{\hat{p}}$

Delta Method

Thm: For a sequence of r.v. W_n if

$$a_n(W_n - b) \xrightarrow{\mathcal{D}} N(0,1)$$

Then

$$\frac{a_n\left(g(W_n) - g(b)\right)}{|g'(b)|} \xrightarrow{\mathcal{D}} N(0,1)$$

a_n

// For the Bernoulli:

$$\sqrt{\frac{n}{p(1-p)}}(\hat{p} - p) \xrightarrow{\mathcal{D}} N(0,1) \qquad \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \to N(0,1)$$

$$\frac{\sqrt{\frac{n}{p(1-p)}}\left(2\sqrt{n}\arcsin\sqrt{\hat{p}} - 2\sqrt{n}\arcsin\sqrt{p}\right)}{\sqrt{\frac{n}{p(1-p)}}} \to N(0,1)$$

$$g(p) = 2\sqrt{n}\arcsin\sqrt{p} \qquad g'(p) = 2\sqrt{n}\frac{1}{\sqrt{1-(\sqrt{p})^2}}\cdot\frac{1}{2}\frac{1}{\sqrt{p}} = \frac{\sqrt{n}}{\sqrt{p(1-p)}}$$

$$2\sqrt{n}\left(\arcsin\sqrt{\hat{p}} - \arcsin\sqrt{p}\right) \to N(0,1)$$

$$P\left(-z_{\frac{\alpha}{2}} < 2\sqrt{n}\left(\arcsin\sqrt{\hat{p}} - \arcsin\sqrt{p}\right) < z_{\frac{\alpha}{2}}\right) \geq 1-\alpha$$

CI for $\arcsin\sqrt{p}$ is $\arcsin\sqrt{\hat{p}} \pm z_{\frac{\alpha}{2}}\cdot\frac{1}{2\sqrt{n}}$

CI for p is. $\left[\sin^2\left(\arcsin\sqrt{\hat{p}} - z_{\frac{\alpha}{2}}\frac{1}{2\sqrt{n}}\right),\ \sin^2\left(\arcsin\sqrt{\hat{p}} + z_{\frac{\alpha}{2}}\frac{1}{2\sqrt{n}}\right)\right]$

Example 7.2.3. CI = (0.38. 0.44) based on VST

For discrete r.v. mean and variance are ~~in~~dependent.

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

7.2.3: Bootstrap Confidence Interval.

population mean / median

$X_i \overset{ind}{\sim}$ Normal $\qquad \hat{p} \quad X_i \sim$ Bernoulli(p)

Nonparametric Confidence Interval

If we have all the data,

Sample 1000 students. $\longmapsto \bar{x}_1$

Sample 1000 students $\longmapsto \bar{x}_1$

10,000 times $\longmapsto \bar{x}_{10,000}$

CI. $[x_{(250)},\ x_{(9750)}] \longmapsto 95\%$

Bradley Efron Bootstrap Treat the data as the full population

We sample with replacement

For each Bootstrap sample, Calculate the mean $\hat{\mu}_j^B$

Replicate $B = \underline{10,000}$ times

① $\hat{\mu}_1^B, \hat{\mu}_2^B, \hat{\mu}_3^B, \ldots \hat{\mu}_B$

② $V^B(\hat{\mu})$ = sample variance of $\hat{\mu}_j^B$

③a $\hat{\mu} \pm \underline{Z_{\frac{\alpha}{2}} \sqrt{V^B(\hat{\mu})}}$

③b $\left[\hat{\mu}_{(\frac{\alpha}{2} \times B)} \; ; \; \hat{\mu}_{(1-\frac{\alpha}{2}) \times B} \right]$

$\Big\}$ $(1-\alpha)$ - CI.

$\hat{\mu} = \bar{X}$

<u>Example</u> (Law School Data)

<u>LSAT</u> GPA

What is the correlation? What is the CI for the correlation?

$$\hat{\rho} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2 \cdot \Sigma (y_i - \bar{y})^2}}$$

Fisher <u>Transformation</u>.

$$\xi = 0.5 \log \frac{1+\rho}{1-\rho} \qquad \hat{\xi} = 0.5 \log \frac{1+\hat{\rho}}{1-\hat{\rho}}$$

$$\hat{\xi} \sim N(\xi, V(\hat{\xi}))$$

7.3. CI for two populations

Independent samples $\mu_1 - \mu_2$

(i) Equal Variances ($\sigma_1^2 = \sigma_2^2$)

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

CI $\mu_1-\mu_2$ is

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\frac{\alpha}{2}, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

(ii) Unequal variance.

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

CI for $\mu_1-\mu_2$ is

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\frac{\alpha}{2}, df} \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

Example: Microarray. 1st and 2nd gene

7.3.2 Variances CI for $\frac{\sigma_1^2}{\sigma_2^2}$

$$\frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} \sim \frac{\chi^2_{n_2-1}/n_2-1}{\chi^2_{n_1-1}/n_1-1} \sim F_{n_2-1, n_1-1}$$

$$F_{\frac{\alpha}{2}, n_1-1, n_2-1} < \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} < F_{1-\frac{\alpha}{2}, n_2-1, n_1-1}$$

$$\frac{S_1^2}{S_2^2} F_{\frac{\alpha}{2}, n_2-1, n_1-1} < \frac{\sigma_1^2}{\sigma_2^2} \cdot \cancel{\frac{S_2^2}{S_1^2}} < \frac{S_1^2}{S_2^2} F_{1-\frac{\alpha}{2}, n_2-1, n_1-1}$$

[, —]

$$F_{\frac{\alpha}{2}, n_2-1, n_1-1} = \frac{1}{F_{1-\frac{\alpha}{2}, n_1-1, n_2-1}}$$

7.3.3 Two sample proportions

One sample. CI: $\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

population 1 and 2 $\quad n_1, x_1 \quad \hat{p}_1 = \frac{x_1}{n_1}$

$n_2, x_2 \quad \hat{p}_2 = \frac{x_2}{n_2}$

p_1 :

p_2 :

CI for $p_1 - p_2$ is

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

$$2\sqrt{n}\arcsin\sqrt{p_1} - 2\sqrt{n}\arcsin\sqrt{p_2}$$

$p_1 - p_2$?

Agresti and Caffo : $\tilde{p}_i = \frac{Y_i + 1}{n_i + 2}$

Then the CI for $p_1 - p_2$ is

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

Example 7.3.3. (Exit Poll)

Sample Size Calculation

$1-\alpha$

Length / Width | $MOE = \frac{\text{Length}}{2}$

$(1-\alpha) \nearrow$, Length MOE $\nearrow$

Length / MOE $\downarrow$ $(1-\alpha) \downarrow$

Enough Sample size

CI for the mean known σ^2, $(1-\alpha)$, MOE

$$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = MOE \qquad n \geq \frac{\sigma^2 (Z_{1-\frac{\alpha}{2}})^2}{(MOE)^2}$$

Sample size for proportion:

$$\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+2}}$$

$$Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\tilde{p})}{n+2}} = MOE$$

$$Z_{\frac{\alpha}{2}}^2 \frac{\hat{p}(1-\tilde{p})}{n+2} = MOE^2 \Rightarrow n+2 = \frac{\tilde{p}(1-\tilde{p}) \cdot Z_{\frac{\alpha}{2}}^2}{MOE^2}$$

Worst Case: $\hat{p} = 1-\tilde{p} = \frac{1}{2}$

$$n \geq \frac{\frac{1}{4} Z_{\frac{\alpha}{2}}^2}{MOE^2} - 2 = \frac{Z_{\frac{\alpha}{2}}^2}{4MOE^2} - 2.$$