# PSTAT 120B HW 5

Outline

- Relative efficiency is the comparing of estimators based on variance. If both estimators are unbiased, $\hat{\theta}_1$ will be relatively more efficient than $\hat{\theta}_2$ if $V(\hat{\theta}_2) > V(\hat{\theta}_1)$. We can use the ratio $V(\hat{\theta}_2)/V(\hat{\theta}_1)$ to define the relative efficiency of 2 unbiased estimators, where $\hat{\theta}_1, \hat{\theta}_2$ are two unbiased estimators for the same parameter $\theta$.

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = V(\hat{\theta}_2)/V(\hat{\theta}_1)$$

- Consistent estimator is an estimator that converges to its target parameter and becomes accurate with an infinite amount of data. The estimator $\hat{\theta}_n$ is a consistent estimator of $\theta$ if, for any positive number $\varepsilon$,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \hat{\theta}| \le \varepsilon) = 1 = \lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0$$

- Sufficient Statistic is a statistic that summarizes all the information in a sample about a target parameter
- For discrete random variables, if sample observations $y_1, y_2, y_n$ were corresponding to RV. $Y_1, Y_2, Y_n$ whose distribution depends on a parameter $\theta$. Then, if $Y_1, Y_2, Y_n$ are discrete random variables, the likelihood of the sample, $L(y_1, y_2 \dots y_n | \theta)$ is defined to be the joint probability of $y_1, y_2 \dots y_n$ that is $L(y_1, y_2, \dots y_n | \theta) = p(y_1|\theta) \times p(y_2|\theta) \times \dots p(y_n|\theta)$. If $Y_1, Y_2, \dots Y_n$ are continuous random variables, the likelihood is defined to be the joint density evaluated at $y_1, y_2, y_n$ such that $L(y_1, y_2 \dots y_n|\theta) = f(y_1|\theta) \times f(y_2|\theta) \times \dots f(y_n|\theta)$. If we let $U$ be a statistic based on the random sample, then $U$ is a sufficient statistic. Likelihood is the probability of observing the event when the value of the parameter is $\theta$.

1) $Y \sim f(y;\theta)$ where $f$ is a gamma $(c, \frac{\theta}{c})$ density

$$f(y,\theta) = \frac{1}{\Gamma(c)\,(\frac{\theta}{c})^c}\, y^{c-1} \exp\left(-\frac{cy}{\theta}\right)$$

(a) $\frac{cy}{\theta} = A$

$$A \Rightarrow f_A(a) = f_y\left(y = \frac{\theta a}{c}\right)\left|\frac{dy}{da}\right|$$

$$\frac{(\theta t/c)^{c-1}}{\Gamma(c)(\theta/c)^c}\, e^{-a} \left|\frac{d}{da}\frac{\theta t}{c}\right|$$

$$= \frac{1}{\Gamma(c)}\, e^{-a} a^{c-1}$$

(b) (Presumably applied on pivot)

$$P(B_L < A < B_U) = 1-\alpha$$
$$P(A < B_U) - P(A < B_L) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2}$$
$$P(A < B_U) = 1 - \alpha/2$$
$$P[b_L < A < b_U) = P[b_L < \frac{cy}{\theta} < b_U]$$
$$= \left(\frac{cy}{F(1-\alpha/2)^{-1}}, \frac{cy}{F(1-\alpha/2)^{-1}}\right)$$

2. Sample mean = 28.8

var = 64

$\Phi(1.96) = .025$

$\Phi(-2.58) = .005$

a) $\bar{Y} = 28.8 \frac{kg}{m^2}$ $\sigma^2 = 64$

$n = 134$

b) Target parameter is mean BMI of American adults. The data shows the adult American population, the mean is the parameter of interest. The sample mean is the estimator for the whole population.

c) $100(1-\alpha)\%$ CI

$\bar{y} - \underbrace{Z_{\alpha/2}}\frac{\sigma}{\sqrt{n}} ; \bar{y} + Z_{\alpha/2}\frac{\sigma}{n}$

critical value at $\alpha/2$ prob.

✓ – data should be normally distributed

✓ – sampled elements are independent

d) $1-\alpha = .95$

95% CI

$\frac{\alpha}{2} = .025$, $s = 8$, $n = 134$

$Z_{\alpha/2} = Z_{0.025}$

$= 1.96$

$\Rightarrow [27.44, 30.15]$

e) $1-\alpha = .99$

$\frac{\alpha}{2} = .005$

$Z_{\alpha/2} = 2.58$

$\Rightarrow [28.8 - 2.58 \times \frac{8}{\sqrt{134}}, 28.8 - 2.58 \times \frac{8}{\sqrt{134}}]$

$\Rightarrow [27.02, 30.58]$

f) 95% $\Rightarrow$ 95% chance lies b/t 27.44 & 30.15

99% $\Rightarrow$ 99% chance lies b/t 27.02 & 30.58

g) – The width of the CI is larger when the confidence is higher

– Greater confidence when less precise

h) Yes.

$Z_{\alpha/2} = \frac{1}{2}(29.98 - 25)\frac{\sqrt{134}}{\sqrt{64}}$

$\approx 3.61$

$\Phi(3.61)$

$\Rightarrow$ 99.98% chance they are overweight