STAT371 DISCUSSION 7 October 19, 2002

1. Conditions for validity of estimation methods

   - Conditions regarding the design of the study
     - It must be reasonable to regard the data as a random sample from a large population.
     - The observations in the sample must be independent of each other.
   -
     - If n is samll, the population distribution must be approximately normal.
     - If n is large, the population distribution need not be appromately normal.

   The requirement that the data are a random sample is the most important condition.

2. Confidence interval for a population proportion

   - The 95% confidence interval for p is:

$$\tilde{p} \pm 1.96\mathrm{SE}_{\tilde{p}}$$

   where

$$\tilde{p} = \frac{y+2}{n+4}$$

   and standard error for $\tilde{p}$is

$$\mathrm{SE}_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

   - Planning a study to estimate p
     If a desired value of $\mathrm{SE}_{\tilde{p}}$ is specified, and if a rough informed guess of $\tilde{p}$ is available, than the require sample size n can be determined from the following equation:

$$\text{Desired SE} = \sqrt{\frac{(\text{Guessed}\tilde{p})(1-\text{Guessed}\tilde{p})}{n+4}}$$

3. Comparison of two independent samples

- Standard error of $(\bar{y}_1 - \bar{y}_2)$ is

$$\mathrm{SE}_{(\bar{y}_1-\bar{y}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\mathrm{SE}_1^2 + \mathrm{SE}_2^2}$$

  where$\mathrm{SE}_1 = \mathrm{SE}_{\bar{y}_1} = \frac{s_1}{\sqrt{n_1}}$ and $\mathrm{SE}_2 = \mathrm{SE}_{\bar{y}_2} = \frac{s_2}{\sqrt{n_2}}$

- The $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is constructed as

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\frac{\alpha}{2}} \mathrm{SE}_{(\bar{y}_1-\bar{y}_2)}$$

  The degree of freedom of Student's t distribution is

$$\mathrm{df} = \frac{(\mathrm{SE}_1^2 + \mathrm{SE}_2^2)^2}{\frac{\mathrm{SE}_1^4}{(n_1-1)} + \frac{\mathrm{SE}_2^4}{(n_2-1)}}$$

  where $\mathrm{SE}_1 = \frac{s_1}{\sqrt{n_1}}$ and $\mathrm{SE}_2 = \frac{s_2}{\sqrt{n_2}}$