MDSC 643.01, Biostatistics I
Brant, Fall 2002

INTERVAL ESTIMATION FOR MEANS AND PROPORTIONS BASED ON LARGE SAMPLES

So far have examined sampling behaviour of $\bar{x} = \frac{\sum observed\ x\ values}{n}$

- approximately normal, typical deviation from μ estimated by $SE(\bar{x}) = \frac{s}{\sqrt{n}}$

Another common estimate is estimated proportion , $\hat{p} = \frac{X}{n}$,

X = #S's in sample or series of n

- estimate of true proportion, p, in population
- $S.E.(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

e.g.

<u>Based on CLT and Normal Approximation to Binomial (special case):</u>

Sampling distribution for $\bar{x}$ (assuming random sampling) will be approximately normal for n large ($n \geq 30$?).

Sampling distribution for $\hat{p}$ (assuming random sampling) will be approximately normal for n large, true p not too close to 0 or 1: ($n\,p(1-p) > 5$)

Interval estimate: a range of "plausible values" (based on data)

e.g. $\bar{x} \pm SE(\bar{x})$ or $\hat{p} \pm SE(\hat{p})$

Question: How do we assess plausibility??

- consider behaviour in hypothetical repeated sampling.
- math development or computer simulation yields coverage rate about 68%.
- widening to $\bar{x} \pm 2SE(\bar{x})$ provides 95% coverage.

Confidence Interval for μ (Large sample)

Since $z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ is approximately normal coverage rate can be controlled at a proportion $1-\alpha$ using $\bar{x} \pm z_{1-\alpha/2} SE(\bar{x})$

where $z_{1-\alpha/2}$ cuts off a proportion $1-\alpha/2$ in the right hand tail.

Confidence Interval for p (Large sample)

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

e.g.

Confidence intervals for Differences.

Often wish to consider difference in two means, say $\mu_1 - \mu_2$ or difference in two proportions, $p_1 - p_2$ Intuitively, we use observed differences, $\bar{x}_1 - \bar{x}_2$ or $\hat{p}_1 - \hat{p}_2$

e.g.

Standard errors for differences:

Given two independent estimates $estimate_1$ and $estimate_2$ with standard errors $SE(estimate_1)$ and $SE(estimate_2)$ the standard error for the difference $estimate_1 - estimate_2$ is

$$\sqrt{[SE(estimate_1)]^2 + [SE(estimate_2)]^2}$$

e.g.

If two independent estimates are normally distributed then so is their difference. Thus

$$(estimate_1 - estimate_2) \pm z_{1-\alpha/2} \; SE(estimate_1 - estimate_2)$$

provides a (large sample) confidence interval with coverage proportion $1-\alpha$.

e.g.