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 $\bar{x} = \dfrac{\sum \textit{observed x values}}{n}$

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

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

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

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

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: ($np(1-p) > 5$)

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.