Stat 381 4/10/2015

Summary Thus Far

	100(1-α)% Conf. Interval
Single Population mean μ n large, σ known	$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
n small or σ unknown	$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ $n-1$ d.f.

Difference in pop means

$\mu_2 - \mu_1$

σ_1, σ_2 known, n_1, n_2 large	$\bar{x}_2 - \bar{x}_1 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
σ_1, σ_2 unknown but assumed equal	$\bar{x}_2 - \bar{x}_1 \pm t_{\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ $S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$ $n_1 + n_2 - 2$ d.f.

What if we don't assume $\sigma_1 = \sigma_2$

Basically we use $\bar{x}_2 - \bar{x}_1 \pm t_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

but we need to use correct d.f. for the t-distribution.

$$df = \nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}$$

round to nearest whole number.

Say we want to estimate the # $\begin{matrix}\text{Unpopped}\\\text{popcorn}\end{matrix}$ kernels difference between Or. R. and Pop Secret popped at 2:30

Sample $n_1 = 14$ bags of Or. R $\bar{x}_1 = 32.7$ $s_1 = 7.7$

$n_2 = 19$ bags Pop Secr. $\bar{x}_2 = 29.3$ $s_2 = 10.5$

$\boxed{95\%\ CI}$

$\sigma_1 = \sigma_2$ assumed
(Pooled = Yes)

$(-3.372, 10.172)$
$df = 31$

$(-3.062, 9.8616)$
$df = 30.999 \approx 31$

interval estimate $\mu_1 - \mu_2$

---

estimation of population proportion "p" in say a Bernoulli experiment.

Sample $X_1, \dots, X_n$ from Bern (p)

we want to estimate for p, it's reasonable to use $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$ sample proportion.

- $E(X_i) = p$ so $E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n} \cdot E(\Sigma X_i)$
  $= \frac{1}{n} \Sigma E(X_i)$
  $= \frac{1}{n} \cdot np = p$

  $Var(X_i) = pq$
- Further $Var(\hat{p}) = Var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} \sum_i Var(X_i)$
  $= \frac{1}{n^2} \cdot npq = \frac{pq}{n}$
- CLT: $\hat{p}$ approx Normal

~~To find~~ $\hat{P} \sim N\left(p,\ \text{stdev} = \sqrt{\frac{pq}{n}}\right)$

use $\hat{p}$ and $\hat{q}$ to estimate $\sqrt{\frac{pq}{n}}$

use $\sqrt{\frac{\hat{p}\hat{q}}{n}}$

so $$\frac{\hat{P} - p}{\sqrt{\frac{\hat{p}\hat{q}}{n}}} \sim N(0,1)$$

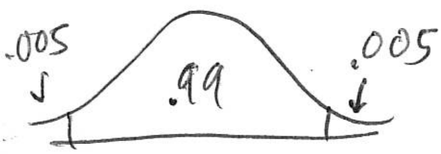
100(1-α)% CI

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

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

Say we want to estimate proportion of Chicagoans who ride bikes.
Survey of 1000 random people,
and .32 sample proportion say they do.

a 99% CI for p



$$.32 \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$.32 \pm 2.58\sqrt{\frac{(.32)(.68)}{1000}}$$

$$(.282\ ,\ .358)$$

If we have 2 populations and want to estimate $p_1 - p_2$

What is the S.E. of $(\hat{p}_1 - \hat{p}_2)$

$$SD(\hat{p}_1 - \hat{p}_2) = \sqrt{Var(\hat{p}_1) + Var(\hat{p}_2)}$$ ← assuming independent samples.

$$= \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

See if residents of Wicker park prefer Rom ~~to~~ more than residents of Hyde Park.

Sample 72 from W.P., 43 say "yes" Pop. 1

89 from H.P. 32 say "yes" Pop 2.

estimate $P_{WP} - P_{HP}$

95% CI is

(.0867, .3885)

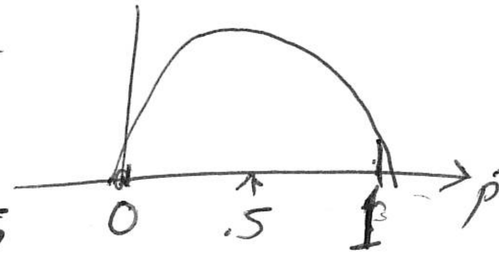
Say we desire a margin of error of ε when estimating "p"

ie $\underbrace{z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}}_{\text{marg. of error}} \leq \varepsilon \Rightarrow \frac{z_{\frac{\alpha}{2}}^2 \hat{p}\hat{q}}{\varepsilon^2} \leq n$

Find how big can $\hat{p}\hat{q}$ be?

Maximize $\hat{p}(1-\hat{p}) = \hat{p} - \hat{p}^2$

$\hat{p} - \hat{p}^2 \leq .5 - .5^2 = .25$

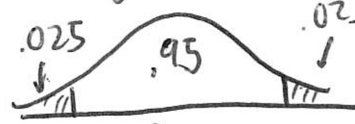


Use .5 as worst case scenario for $\hat{p}$

ie choose n such that

$$\frac{z_{\frac{\alpha}{2}}^2}{4\varepsilon^2} = \frac{z_{\frac{\alpha}{2}}^2 (.5)(.5)}{\varepsilon^2} \leq n$$

Say candidate wants to make a 95% CI for his support with a margin of error of 1% or .01



needs $n \geq \dfrac{z_{\frac{\alpha}{2}}^2 (.5)(.5)}{\varepsilon^2} = \dfrac{1.96^2}{4(.01)^2} = 9604$

if he already sampled 1000 people to get $\hat{p} = .32$

now revise $n \geq \dfrac{1.96^2 (.32)(.68)}{.01^2} = 8359.3$