Lec 2/3

Friday, February 3, 2017 14:57

Recall:

	σ^2 known	σ^2 unknown
norm. pop	$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ approx, $\bar{X} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$ exact
non-normal.	Same, but approx	$\bar{X} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$ approx ($\downarrow$ or $Z_{\frac{\alpha}{2}}$)

when n is large, these work well.

other ways to construct CI:

$$P\left(\mu < \bar{X} + Z_{\alpha} \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha, \quad \text{so} \quad \left(-\infty, \bar{X} + Z_{\alpha}\frac{\sigma}{\sqrt{n}}\right) \text{ is a } (1-\alpha)\times 100\% \text{ CI for } \mu.$$

Method used above is quite general; not limited to mean parameter.

↳ Pivotal method.

e.g.

$$\underbrace{Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}} \sim N(0,1)$$

Transformation of data and μ to get a dist. that does not depend on μ.

Comparing averages of 2 populations. § 11.3

Estimating difference in mean:

$\bar{X}_1 - \bar{X}_2$ where $\bar{X}_1$ is sample mean of pop 1, etc.

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \qquad \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

- $\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$ (provided pop 1, pop 2 independent).

∴ can construct a confidence interval

So can construct a confidence interval...

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

So construct CI for difference in means in the same way.

When distributions are not normal, this CI is still valid for large n_1, n_2.

Ex: stability of a filler machine. 95% CI for diff 2 days.

day 1: 100 boxes are sampled, $\bar{X}_1 = 1.15$ lbs.

day 2: 50 boxes are sampled $\bar{X}_2 = 1.05$ lbs

assume $\sigma = 0.17$.

CI: $0.1 \pm 1.96 \sqrt{\frac{0.17^2}{100} + \frac{0.17^2}{50}}$

$= (0.0403, 0.1577)$ lbs.

So we are 95% confident that true diff $\in$ CI.

What if σ^2 is unknown?

use pooled sample variance: $S_0^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ (ex: $E(S_0^2) = \sigma^2$).

$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_0^2 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ is T dist. w/ df $n_1 + n_2 - 2$.

Ex: