

Section 6.3: Estimating the Difference

Monday, March 5, 2018

8:46 AM

Between Two Populations

- important question - are discrepancies in the results for two different samples significant?

- comparison of materials furnished by different suppliers

- comparison of test scores for students in different sections of programs

- section 1 got 78% on Quiz 1

- section 2 got 65% " "

estimating the difference between two population means

from Gilles' handout:

$$\underbrace{\mu_1 - \mu_2}_{\text{estimator, the difference in population means}} = \underbrace{\bar{x}_1 - \bar{x}_2}_{\text{difference in sample means}} \pm \underbrace{MOE}_{\substack{\uparrow \\ \text{margin of error}}}$$

$$= \bar{x}_1 - \bar{x}_2 \pm Z_{\alpha/2} SE$$

where $SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Note: if σ_1 and σ_2 are unknown, but n_1 and n_2 are large (≥ 30), can approximate with s_1 and s_2 , respectively

if the two sampled populations are each normally distributed, then the difference $(\bar{x}_1 - \bar{x}_2)$ is also normally distributed, regardless of the sample size.

if the two sampled populations are not normally distributed, then $(\bar{x}_1 - \bar{x}_2)$ is approximately normally distributed if n_1 and n_2 are both greater or equal to 30 (Central Limit Theorem)

example: Police records for two different districts record the mean number of emergency calls per shift.

	region 1	region 2
sample size	100	100
mean	2.4	3.1
$s^2 = \text{variance}$	1.44	2.64

Find a 90% confidence interval for the difference in mean numbers of emergency calls per shift and interpret your result.

90% confidence

$$\begin{aligned}\mu_1 - \mu_2 &= \bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\ &= (2.9 - 3.1) \pm 1.645 \sqrt{\frac{1.44}{100} + \frac{2.69}{100}} \\ &= -0.7 \pm 0.332 \\ &= -0.7 \pm 0.3\end{aligned}$$

with 90% confidence, the difference is
-1.0 to -0.4 calls per shift

note: zero is not included in the interval

in the above example it is likely that there
is a significant difference between the
two regions at the 90% level

note: if zero is included, it is likely
that there is no significant
difference between the two samples