ESTIMATION

Small sample confidence intervals for means

$$\boxed{\hat{\theta} \pm z_{\alpha/2}\, \sigma_{\hat{\theta}}}$$

Small sample confidence intervals for μ

- We just studied what to do to get (at least) approximate confidence intervals for population means – this is known to work only when n is large.

$\bar{Y} \sim N(\mu, \sigma^2/n)$

- What if $Y_1, Y_2, \ldots, Y_n \sim iid\ N(\mu, \sigma^2)$, how to get a confidence interval for any sample size?

$\left\{\begin{array}{l} \text{data} \\ \text{unknown par} \end{array}\right.$

- If σ^2 is known, then $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ is a pivotal quantity.
- But if σ^2 is not known, $\frac{\bar{Y}-\mu}{s/\sqrt{n}}$ does not have a normal distribution.
- We saw earlier, however, that for normal data, $\boxed{\frac{\bar{Y}-\mu}{s/\sqrt{n}}}$ has a t_{n-1} distribution.

Small sample confidence intervals for μ

- The t_{n-1} distribution looks a bit like the $N(0,1)$ distribution but with "heavier tails" (but tails get "lighter" for larger n).
- But $\frac{\bar{Y}-\mu}{s/\sqrt{n}}$ is still a pivotal quantity. ✓ depend only on data & unknown par
  $\bar{Y}, s, n$ $\quad \mu$

$\sim t_{n-1}$

✓ dist. does not depend on unknown quantities

$$P\left(-t_{\frac{\alpha}{2},n-1} < \frac{\bar{Y}-\mu}{s/\sqrt{n}} < -t_{\frac{\alpha}{2},n-1}\right) = 1-\alpha$$

$$P\left(-t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} < \bar{Y}-\mu < t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\bar{Y} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} < \mu < \bar{Y} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

C.I.

Required assumption $Y_1, \ldots, Y_n \sim \text{iid } N(\mu, \sigma^2)$

Small sample confidence intervals for $\mu_1 - \mu_2$

- Two independent samples: $\overbrace{Y_{1,1}, Y_{1,2}, \dots, Y_{1,n_1}}^{s_1^2} \sim iid\ N(\mu_1, \sigma^2)$ and $\underbrace{Y_{2,1}, Y_{2,2}, \dots, Y_{2,n_2}}_{s_2^2} \sim iid\ N(\mu_2, \sigma^2)$.
- The variance σ^2 is common to both (but unknown).
- How to estimate σ^2?
- We have seen before that both the sample variances s_1^2 and s_2^2 are unbiased estimators of σ^2.
- A better estimator is the "pooled" estimator:
- $s_p^2 = \frac{1}{n_1+n_2-2}[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]$

Small sample confidence intervals for $\mu_1 - \mu_2$

- We have seen before that $\frac{(n_1-1)s_1^2}{\sigma^2} \sim \chi^2_{n_1-1}$ and that $\frac{(n_2-1)s_2^2}{\sigma^2} \sim \chi^2_{n_2-1}$.
- Also, the samples are independent and we know that the sum of two independent χ^2 random variables also has a χ^2 distribution.
- From this we can work out an (exact) confidence interval for $\mu_1 - \mu_2$.

$$\frac{(n_1-1)s_1^2}{\sigma^2} + \frac{(n_2-1)s_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2} \quad \text{(fact 1)}$$

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right) \quad \text{(fact 2)}$$

$$\frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1) \quad \text{(fact 2')}$$

$Z \sim N(0,1)$ and $U \sim \chi^2_\nu$ are indp. then

$$\frac{Z}{\sqrt{U/\nu}} \sim t_\nu \qquad \text{(fact 3)}$$

$$\frac{\left\{\dfrac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{\cancel{\sigma^2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right\} N(0,1)}{\left\{\sqrt{\underbrace{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{\cancel{\sigma^2}} \Big/ (n_1+n_2-2)}_{s_p^2}}\right\} \sqrt{\chi^2_\nu/\nu}} \sim t_{n_1+n_2-2}$$

P.Q. !!

$$\frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

$$P\left(-t_{\alpha/2,\, n_1+n_2+2} < \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < \right) = 1-\alpha$$

Isolate $\mu_1 - \mu_2$ in the middle

$$P\left(\bar{Y}_1 - \bar{Y}_2 - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 < \bar{Y}_1 - \bar{Y}_2 + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

$$= 1-\alpha$$

$$\bar{Y}_1 - \bar{Y}_2 \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Some notes on approximation: Confidence intervals for μ and $\mu_1 - \mu_2$

- We have seen before that for "reasonably large" sample sizes, even if not normally distributed, the "large sample" intervals can be a good approximation. (data) $\bar{Y} \sim N(\quad)$ CLT
- For normally distributed data, if the two population variances are not equal, an exact confidence interval for $\mu_1 - \mu_2$ is not possible, but an approximation is available.

small

- The t procedures we have just talked about will generally be reasonably good approximations even if the data aren't truly normal, so long as the samples aren't "too small."