8.8 Small-sample CIs for μ and $\mu_1 - \mu_2$

- Suppose $Y_1, \ldots, Y_n \overset{\text{indep}}{\sim} N(\mu, \sigma^2)$ with both μ and σ^2 unknown, where n is not large.

- The random quantity $\dfrac{\bar{Y} - \mu}{S/\sqrt{n}}$ no longer has an approximate standard normal distribution.

- If n is small, then S may not be near the true σ with high probability.

- Recall from Chapter 7: When $Y_1, \ldots, Y_n$ are independent $N(\mu, \sigma^2)$,

$$T = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim$$

- Hence T is still a pivotal quantity.

- The t_{n-1} density resembles the $N(0,1)$ density, but with thicker tails (more variability).

Picture:

Since

Simple algebra shows:

Two-Sample Situation

- Suppose we have two independent samples $Y_{11}, Y_{12}, \ldots, Y_{1n_1} \overset{indep}{\sim} N(\mu_1, \sigma^2)$ and $Y_{21}, Y_{22}, \ldots, Y_{2n_2} \overset{indep}{\sim} N(\mu_2, \sigma^2)$ where n_1, n_2 are not large and the common variance σ^2 is unknown. We will derive a CI for $\mu_1 - \mu_2$.
- To estimate the common σ^2, we use a weighted average of the two sample variances S_1^2 and S_2^2:

Since

is the sum of independent $\chi^2_{n_1-1}$ and $\chi^2_{n_2-1}$ r.v.'s:

Thus let

<u>Note:</u> When the sample size(s) are large enough that the degrees of freedom exceed, say, 30, the t-intervals may be replaced by the (approximate) z-intervals.

<u>Note:</u> If the two population variances σ_1^2 and σ_2^2 cannot be assumed to be equal, an <u>approximate</u> t-interval for $\mu_1 - \mu_2$ may be obtained (not covered here).

Example: Redo the Exercise from Section 8.6, but assume $n_1 = 6$ (not 60) and $n_2 = 9$ (not 90). Also assume the data come from two populations that are normally distributed with equal population variances.

- If $Y_1, \ldots, Y_n$ do not come from a normal population, the t-procedures are <u>robust</u> and can be used for approximate CIs, as long as the sample size is not too small (say, $n > 15$) and no skewness or outliers are evident in the sample data.

8.9 Confidence Intervals for Variances

- Suppose $Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with μ and σ^2 unknown. We seek a $100(1-\alpha)\%$ CI for σ^2. Recall that:

- Clearly, taking square roots of the endpoints yields a $100(1-\alpha)\%$ CI for σ.

- Suppose we have two independent samples $Y_{11}, \dots, Y_{1n_1} \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$ and $Y_{21}, \dots, Y_{2n_2} \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$.

  Recall:

Example 1 (a): Suppose the maturation times for a flower species are $N(\mu, \sigma^2)$. If a random sample of $n = 13$ seeds yielded $s^2 = 10.7$, then a 90% CI for σ^2 is:

Exercise: If a second sample of seeds (from a different normal population) yielded $n_2 = 9$ and $s_2^2 = 4.59$, show that a 98% CI for σ_2^2/σ_1^2 is $(0.095, 2.43)$.

- Note: The CIs for variances are not robust and thus not valid for non-normal populations.