Lecture 25

niceguy

March 15, 2023

1 Tolerance Limits

For a normal IID sample, with sample mean

$$\overline{x} = \frac{1}{n}\sum_i x_i$$

and sample variance

$$s^2 = \frac{1}{n-1}\sum_i (x_i - \overline{x})^2$$

Then the tolerance limits are in the form

$$\overline{x} \pm ks$$

We choose k so that $(1-\alpha)$ of the population is within $[\overline{x}-ks, \overline{x}+ks]$. This does not shrink with n

$$\lim_{n\to\infty} P(\overline{X} - kS \le X \le X + kS) = 1 - \alpha$$

Example 1.1 (Heights)**.** The confidence interval gets narrower as $n \to \infty$, but the tolerance limits don't really go anywhere.

2 Two Samples

Suppose with have two samples with n_1, n_2, means μ_1, μ_2, variances σ_1^2, σ_2^2. Then we want to estimate $\mu_1 - \mu_2$. Then consider the statistic $\overline{X_1} - \overline{X_2}$. By the Central Limit Theorem, both are approximately normal, hence

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 + \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

has a normal disstribution. This can be used to make a confidence interval for $\mu_1 - \mu_2$.

Example 2.1 (Swimmers)**.** For $n_1 = 40, n_2 = 43, \overline{x_1} = 60, \overline{x_2} = 58, s_1^2 = 1, s_2^2 = 2$. Then since $n > 30$ in both cases, we can use the Central Limit Theorem, taking the sample variance as true variance,

$$z = \frac{2 - (\mu_1 + \mu_2)}{0.27}$$

Now

$$\begin{aligned} 0.95 &= P(-z_{0.025} \leq Z \leq z_{0.025}) \\ &= P(\overline{X_1} - \overline{X_2} - 0.27z_{0.0025} \leq \mu_1 - \mu_2 \leq \overline{X_1} - \overline{X_2} + 0.27z_{0.0025}) \\ &= P(1.47 \leq \mu_1 - \mu_2 \leq 2.53) \end{aligned}$$

For some sample $X_1, \ldots, X_n$, and a statistic W with a probability density function $f(w)$, we define

$$w_{\frac{\alpha}{2}} = -F^{-1}\left(\frac{\alpha}{2}\right)$$

3 Two Samples with unknown Variance

If $n_1, n_2 < 30$ with unknown $\sigma_1 = \sigma_2$. We use a pooled estimate of variance

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Then for the statistics

$$T = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Then T has t distribution, and we can make a confidence interval as usual.