Lecture 7

Xiaoping Shi
Department of Statistics, University of Toronto
xpshi@utstat.toronto.edu

July 27, 2013

Assume that a sample, $X_1, \cdots, X_n$, is drawn from a normal distribution that has mean μ_X and variance σ^2, and that an independent sample, $Y_1, \cdots, Y_m$, is drawn from another normal distribution that has mean μ_Y and the same variance, σ^2.

If we think of the X's as having received a treatment and the Y's as being the control group, the effect of the treatment is characterized by the difference

$$\mu_X - \mu_Y$$

This lecture is concerned with the following topics.

- MLE of $\mu_X - \mu_Y$.
- Confidence interval for $\mu_X - \mu_Y$.
- Hypothesis tests: $H_0 : \mu_X = \mu_Y$ v.s. $H_1 : \mu_X \neq \mu_Y$, $H_2 : \mu_X > \mu_Y$, and $H_3 : \mu_X < \mu_Y$.

$$L(\mu_X, \mu_Y, \sigma^2) = \frac{\exp\left[-\frac{\sum_{i=1}^{n}(X_i-\mu_X)^2}{2\sigma^2}\right]}{(\sqrt{2\pi})^n\sigma^n} \frac{\exp\left[-\frac{\sum_{i=1}^{m}(Y_i-\mu_Y)^2}{2\sigma^2}\right]}{(\sqrt{2\pi})^m\sigma^m}$$

Setting the first derivatives $\partial L/\partial\mu_X$ and $\partial L/\partial\mu_X$ equal to zeros, we obtain

$$\hat{\mu_X} = \bar{X}, \quad \hat{\mu_Y} = \bar{Y}$$

So the MLE of $\mu_X - \mu_Y$ is $\bar{X} - \bar{Y}$.

Question: What's the MLE of σ^2?

If σ^2 were known, confidence interval for $\mu_x - \mu_y$ could be based on

$$\bar{X} - \bar{Y} \sim N\left[\mu_X - \mu_Y, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right]$$

The $1 - \alpha$ confidence interval would be of the form

$$(\bar{X} - \bar{Y}) \pm z(\alpha/2)\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}$$

If σ^2 were unknown, confidence interval for $\mu_x - \mu_y$ could be based on

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{s_p^2\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

where $s_p^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \bar{Y})^2}{n+m-2}$ called **pooled sample variance** (unbiased). The $1 - \alpha$ confidence interval would be

$$(\bar{X} - \bar{Y}) \pm t_{n+m-2}(\alpha/2)s_p\sqrt{\frac{1}{n} + \frac{1}{m}}$$

The test statistic would be

$$t = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Given the significance level α, the rejection regions are

$$\text{for } H_1, |t| > t_{n+m-2}(\alpha/2)$$

$$\text{for } H_2, t > t_{n+m-2}(\alpha)$$

$$\text{for } H_3, t < -t_{n+m-2}(\alpha)$$

Example 1 Two methods, A and B, were used in a a determination of the latent heat of fusion of ice. The investigators wanted to find out by how much the methods differed.
$X = (79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04, 79.97, 80.05, 80.03, 80.02, 80.00, 80.02)$ $(n = 13)$
$Y = (80.02, 79.94, 79.98, 79.97, 79.97, 80.03, 79.95, 79.97)$ $(m = 8)$.

The point estimate $\bar{X} - \bar{Y} = 0.04$ for $\mu_X - \mu_Y$.

1-95% confidence interval is

$$(\bar{X} - \bar{Y}) \pm t_{19}(0.025) s_p \sqrt{1/13 + 1/8}$$

where $s_p^2 = 0.0007178$

For testing

$$H_0 : \mu_X = \mu_Y, \quad \text{v.s.} \quad H_1 : \mu_X \neq \mu_Y$$

the test statistic is

$$t = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{1/13 + 1/8}} = 3.33$$

Given the significance level $\alpha = 0.01$, the rejection region for the the alternative is

$$|t| > t_{19}(0.005) = 2.861$$

Hence, we reject H_0.