Section 12, 11/18/19

t-Statistic (page 421)

Motivation

Suppose we have measurements of interest from two groups, n measurements $X_1, \ldots, X_n$ from a control group and m measurements $Y_1, \ldots, Y_m$ from a treatment group. The effect size is the difference $\mu_X - \mu_Y$ and may let us make inferences on the effect of the treatment. The natural estimate of this is $\bar{X} - \bar{Y}$. If we assume that the $X's$ and $Y's$ are independent and follow normal distributions $N(\mu_X, \sigma^2)$ and $N(\mu_Y, \sigma^2)$ respectively, then clearly

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

If we knew σ^2, then we could construct a confidence interval using our knowledge of the normal distribution. However, we typically don't know σ^2 and so must estimate it. We combine information form the sample variances and define the **pooled sample variance** as

$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{m+n-2}$$

This is the weighted average of the sample variances. This leads us to the following theorem.

Theorem A

Suppose $X_1, \ldots, X_n \sim N(\mu_X, \sigma^2)$ and $Y_1, \ldots, Y_m \sim N(\mu_Y, \sigma^2)$ are independent of each other. Then the t-statistic

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

follows a $t-$distribution with $m+n-2$ degrees of freedom.

Proof

We know that a $t-$distribution is the ratio of a normal distribution and the square root of an independent chi-squared distribution divided by its degrees of freedom.

First note that $(n-1)s_X^2/\sigma^2$ and $(m-1)s_Y^2/\sigma^2$ are chi-squared random variables with $n-1$ and $m-1$ degrees of freedom. They are also independent as the $X's$ and $Y's$ are. Thus their sum is also a chi squared distribution with $m+n-2$ degrees of freedom. Thus

$$U = \frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

is distributed according to a standard normal and

$$V^2 = \left(\frac{(n-1)s_X^2}{\sigma^2} + \frac{(m-1)s_Y^2}{\sigma^2}\right)\frac{1}{m+n-2}$$

is a chi-squared distribution divided by its degrees of freedom. So, $t = U/\sqrt{V}$ follows a $t-$distribution.