Hypothesis Testing II: Two Sample t Tests

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

OCTOBER 11, 2016

Comparing Two Samples

- Thus far, we've tested whether data is consistent with mean μ
- What if we want to test whether two samples are from the same distribution

Comparing Two Samples

- Thus far, we've tested whether data is consistent with mean μ
- What if we want to test whether two samples are from the same distribution
- Two-Sample t-test

Two-Sample (unpooled)

- Two samples $X_1 = \{x_{1,1}, x_{1,2} \dots x_{1,N_1}\}$ and $X_2 = \{x_{2,1}, x_{2,2} \dots x_{2,N_2}\}$
- Doesn't assume that variance is the same for both samples (unpooled)
- Compute mean and sample variance for sample 1 $(\bar{x_1}, s_1^2)$ and sample 2 $(\bar{x_2}, s_2^2)$

Test Statistic

- T-statistic

$$T = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{1}$$

- Plug into t-distrubtion with

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}} \tag{2}$$

- Intuition: Difference between $\bar{x}_1$ and $\bar{x}_2$ has variance that's an interpolation between the two samples

Test Statistic

- T-statistic

$$T = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{1}$$

- Plug into t-distrubtion with

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}} \tag{2}$$

- Intuition: Difference between $\bar{x}_1$ and $\bar{x}_2$ has variance that's an interpolation between the two samples
- Two-tailed vs. one-tailed distinction still applies

ν Example

$$s_1^2 = 1, s_2^2 = 2, n_1 = 4, n_2 = 8$$

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}} \tag{3}$$

$$\tag{4}$$

ν Example

$$s_1^2 = 1, s_2^2 = 2, n_1 = 4, n_2 = 8$$

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}} \tag{3}$$

$$= \frac{\left(\frac{1}{4} + \frac{2}{8}\right)^2}{\frac{1}{3}\left(\frac{1}{4}\right)^2 + \frac{1}{7}\left(\frac{2}{8}\right)^2} \tag{4}$$

$$\tag{5}$$

ν Example

$$s_1^2 = 1, s_2^2 = 2, n_1 = 4, n_2 = 8$$

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}} \tag{3}$$

$$= \frac{\left(\frac{1}{4} + \frac{2}{8}\right)^2}{\frac{1}{3}\left(\frac{1}{4}\right)^2 + \frac{1}{7}\left(\frac{2}{8}\right)^2} \tag{4}$$

$$= \frac{\frac{1}{4}}{\left(\frac{1}{4}\right)^2\left[\frac{1}{3} + \frac{1}{7}\right]} = \frac{4}{\frac{10}{21}} = \frac{42}{5} \tag{5}$$