

§7.5 ONE-FACTOR ANOVA

Introduction: Comparison of two means (See §7.3)

Let \( X_1 \sim N(\mu_1, \sigma_1^2) \), \( X_2 \sim N(\mu_2, \sigma_2^2) \) be independent random variables ("treatments")

\[ H_0 : \mu_1 = \mu_2 \]
\[ H_a : \mu_1 \neq \mu_2 \]

We assume that \( \sigma_1 = \sigma_2 = \sigma \) is unknown.

Random samples and their means:

\[ X_{11}, X_{12}, \ldots, X_{1,n_1}, \quad \bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i} \] (the results of the first "treatment")

\[ X_{21}, X_{22}, \ldots, X_{2,n_2}, \quad \bar{X}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} X_{2i} \] (the results of the second "treatment")

The test statistic is

\[ T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t(n_1 + n_2 - 2) \]

under \( H_0 \).

\[ T^2 = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{\sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim \chi^2(1) \]

\((\bar{X}_1 - \bar{X}_2)^2\) is the variation ("variance") between "treatments"; it is close to \((\mu_1 - \mu_2)^2\) by the LLN.

\[ \frac{\sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \]

(the pooled variance) is the variation ("variance") inside the "treatments" (it is an estimator of \(\sigma^2\)). It characterizes the "randomness", the "errors".

Let us recall that under \( H_0 \)

\[ Z = \frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \sim N(0, 1) \) and that \( Z^2 \sim \chi^2(1) \). Also \( \frac{n}{\chi^2(m)} \sim F(n, m) \).

Therefore

\[ T^2 = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{\sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2}{\sigma^2(n_1 + n_2 - 2)}} \sim \frac{Z^2}{\frac{\chi^2(n_1 + n_2 - 2)}{n_1 + n_2 - 2}} \sim \frac{\chi^2(1)}{\frac{\chi^2(n_1 + n_2 - 2)}{n_1 + n_2 - 2}} \sim F(1, n_1 + n_2 - 2). \]