Handout 05: Two-Samples

Consider extending our current setup for a statistical sample to the case where we have two samples, where each random observation is mutually independent from all the others:

$$X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{G}_X, \quad \mathbb{E}[X_i] = \mu_X, \quad \text{Var}[X_i] = \sigma_X^2$$
$$Y_1, \ldots, Y_m \overset{iid}{\sim} \mathcal{G}_Y, \quad \mathbb{E}[Y_i] = \mu_Y, \quad \text{Var}[Y_i] = \sigma_Y^2$$

We will have the normal definitions of the sample means ($\bar{X}$ and $\bar{Y}$) and the sample variances (S_X^2 and S_Y^2). The **pooled variance** S_p^2 is defined as a combination of the sample variance of X and Y:

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

We will study its properties on today's worksheet. One of our key questions will how to build a point estimator and confidence interval for the difference in the means of the two distributions: $\mu_X - \mu_Y$.

F-distribution

There is one more special distribution that we make significant usage of in statistics that, like the T-distribution, is particularly designed to serve as a pivot. If we have two independent random variables C_1 and C_2 that have chi-squared distributions of k_1 and k_2 degrees of freedome. Then, the following:

$$F = \frac{C_1/k_1}{C_2/k_2}$$

Has an F-distribution with k_1 and k_2 degrees of freedom (yes, it has two different degrees of freedom). We will use this today and then for several seemingly quite different tasks throughout the next several weeks.