# Handout 05: Two-Sample Statistics

We have fully derived the one-sample T-test. Today, we will introduce the **two-sample T-test**. The setup and final results are repeated here for easy reference. Consider observing two different random samples from two potentially different underlying distributions. We will write this as $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{G}_X$ and $Y_1, \ldots, Y_m \overset{iid}{\sim} \mathcal{G}_Y$. We will assume that both $\mathcal{G}_X$ and $\mathcal{G}_Y$ are normal and that they have a shared common (but unknown) variance $\sigma^2$. We want to produce an hypothesis test that the difference in means $\theta = \mu_X - \mu_Y$ is equal to some fixed value $\theta_0$ (typically zero). First, we define the pooled sample variance as follows:

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

Then, the following is a valid pivot:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim T(n+m-2).$$

With the null value of $\theta$ plugged in, we get a valid test statistic. The corresponding confidence inverval for the difference means is:

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

The central limit theorem can be used to extend this result to the case where the distributions are not normal. In R, we will see a variant that further extends this to the situation where the groups have different variances.