Lecture 29

We now study the two sample problem where the data $X_1, \ldots, X_m$ **iid** $\text{Normal}(\mu_1, \sigma_1^2)$ **is independent of** $Y_1, \ldots, Y_n$ **iid** $\text{Normal}(\mu_2, \sigma_2^2)$**. Initially, we make the unrealistic assumption that both** σ_1 **and** σ_2 **are known.**

Under the above conditions, interest lies in the unknown parameter $\mu_1 - \mu_2$**. The test statistic used in the construction of confidence intervals and hypothesis testing is**

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim \text{Normal}(0, 1)$$

Example: Suppose that fifty test scores from class A are independent of 70 test scores from class B. Assume further that the test scores are normal, $\sigma_1^2 = \sigma_2^2 = 84$, $\bar{X} = 73$ and $\bar{Y} = 59$. Is there a difference between the two classes?

Example continued: Construct a 95% confidence interval for $\mu_1 - \mu_2$.

Example continued: Suppose the question had instead been, "Is class A better than class B?"

Example continued: Suppose the question had instead been, “Is class A more than five marks better than class B?”

The significance of ”significance”:

When we reject the null hypothesis H_0, we say that the result is *statistically significant.*

Discussion points:

- **always report the p-value**
- **keep in mind that $\alpha = 0.05$ is arbitrary**
- **significance does not always mean importance**
- **p-values are related to sample size**

More on stat significance vs practical importance:

Example: Spring Birthday Confers Height Advantage - Yahoo Health News, Feb 18/98

In an Austrian study of 507,125 military recruits, it was found that the average height of those born in the spring was $1/4$ **inch more than those born in the fall.**