Set 31: Paired data

We have $X_1, \ldots, X_n$ iid arising from a population with mean μ_1, and $Y_1, \ldots, Y_n$ iid arising from a population with mean μ_2. Furthermore, assume that the data are paired such that X_i corresponds to Y_i. This natural pairing implies that there is a dependence between X_i and Y_i.

To carry out inference (testing and the construction of CI's), we define a new random variable, the difference $D_i = X_i - Y_i$. Our interest concerns the unknown parameter

$$\begin{aligned} \mathrm{E}(D_i) &= \mathrm{E}(X_i - Y_i) \\ &= \mathrm{E}(X_i) - \mathrm{E}(Y_i) \\ &= \mu_1 - \mu_2. \end{aligned}$$

Our analysis proceeds as in the single sample case based on the data $D_1, \ldots, D_n$.

Example: Suppose scores measuring jitteriness are normally distributed . We believe that scores increase after drinking coffee. Let X_i be the before drinking coffee score and let Y_i be the after drinking coffee score for the i-th individual. Based on $\alpha = 0.01$, test the hypothesis.

x_i	y_i	d_i
50	56	
60	70	
55	60	
72	70	
85	82	
78	84	
65	68	
90	88	

Example cont'd: Obtain a 95% CI for the mean difference in jitteriness scores.

Example cont'd: Suppose we have the same data but the experiment involves 16 people where 8 people were measured without having coffee and 8 other people where measured after drinking coffee. How does the analysis differ?

Pairing is a special case of *blocking* (read in text). Blocking attempts to reduce variation by grouping data that are similar, and this hopefully leads to *more sensitive* tests (ie. tests that reject H_0 more often when H_0 is false).

Example: To illustrate the above, consider five before and after measurements involving a drug where there are big differences in responses between people but there is small variation in the D_i's. Assuming normal data, we carry out a paired analysis and a non-paired analysis.

x_i	y_i	d_i
25	29	-4
46	50	-4
30	33	-3
75	78	-3
19	25	-6