Review on $\sum$ and $\prod$:

1, $\sum_{i=1}^{n} y_i = y_1 + \cdots + y_n$

2, $\sum_{i=1}^{n} c\, y_i = c \sum y_i \Rightarrow \sum_{i=1}^{n} c = nc$

3, For $i = 1, \ldots, n$, $\sum_{j=1}^{m} y_{ij} = y_{i\cdot}$.

For $j = 1, \ldots, m$, $\sum_{i=1}^{n} y_{ij} = y_{\cdot j}$

4, $\sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij} = \sum_{i=1}^{n} y_{i\cdot} = \sum_{j=1}^{m} y_{\cdot j} = y_{\cdot\cdot}$

5, $\sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j = \sum_{i=1}^{n} x_i \sum_{j=1}^{m} y_j$

* Convention : $\sum_i \sum_j \sum_k y_{ijk} = \sum_{ijk} y_{ijk} = \sum y_{ijk}$

6, $\prod_{i=1}^{n} y_i = y_1 \cdots y_n$

7, $\prod c y_i = c^n \prod y_i \Rightarrow \prod_{1}^{n} c = c^n$

8, $\prod_{i=1}^{n} x_i y_i = \prod_{1}^{n} x_i \prod_{1}^{n} y_i$

What is "Multivariate Stats" ?

We have multiple variables and we simultaneously measure them for each single observation (subject, person, ...)

- We typically have n observations and for each observation we observe the values of p variables.
- Most of the time in real situations variables are not indep. and hence we have dependent observations.

Subject	Variables of interest
Students	(Math, history, Music, art, physics)
Patient	(Height, weight, blood pressure)

Some of the techniques we might see in the multivariate data analysis are as follows:

① Hotelling's T^2 - test: Comparing two multivariate Samples.

② MANOVA (Multivariate ANOVA): Comparing more than two multivariate Samples.

③ Discriminate Analysis : We have two multivariate populations and a new given observation.

Goal : which pop. shall we assign this new observation?

④ Cluster Anal. : We have a bunch of observations which comes from g groups. (cluster)

Goal : How to assign obs. to the groups.

⑤ Principal Component Anal. (PCA) :

⑥ Factor Anal.

⑦ Canonical Correlation Anal.

Since we have n obs. and p var's, it seems natural to store the observed value of each observations as an element of some arrays or matrices. That's why we need to spend some time studying matrix algebra.

- Difference between Multiple (linear) regression and Multivariate (linear) regression?

A multiple reg. has more than one predictor X, while a multivariate reg has more than one response variable y.

Therefore a Multivariate Multiple regression has multiple X's to predict multiple y's.

- In a Multiple regression set up we have:

Response = Predictors + error

Response: Single var. (Univariate)

Predictors: Multiple variables

But the predictors are treated as non-random. (under experimenter's control)

- Often the time, the are Covariates among the predictors which are not directly under the experiment's control.

In the ANOVA: we have multiple treatment groups (Say, various drugs) and a response (say, Cholesterol level)