Notations:

Y_{ij}: response measurement for subject i at time t_{ij}

Y_i: n by 1 vector of response measurements for subject i

X_{ijk}: explanatory variable k observed at time t_{ij} (sometimes we use $X_{ij} = t_{ij}$)

X_{ij}: p by 1 vector of explanatory variables at time t_{ij}

X_i: n by p matrix of covariates for subject i

Properties of $\bar{Y}$ and S:

Define $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ as an n by 1 vector and $S = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(Y_i - \bar{Y})^T$ as an n by n matrix

where Y_i is a n by 1 vector of independent and identically distributed samples from a multivariate distribution with mean μ (n by 1) and covariance matrix Σ (n by n). Then $\bar{Y}$ and S are unbiased for μ and Σ, respectively.

In particular, $E(\bar{Y}) = \mu$ and $E(S) = \Sigma$:

(1) $E(\bar{Y}) = \frac{1}{N}\sum_{i=1}^{N} E(Y_i) = \mu$, and $V(\bar{Y}) = \frac{1}{N^2}\sum_{i=1}^{N} V(Y_i) = \frac{1}{N}\Sigma$

(2) First, $E\left(Y_{ij}Y_{ik}\right) = Cov\left(Y_{ij}, Y_{ik}\right) + E\left(Y_{ij}\right)E(Y_{ik}) = \Sigma_{jk} + \mu_j\mu_k$. Therefore, we have $E\left(Y_iY_i^T\right) = \Sigma + \mu\mu^T$.

Second, we have $E(Y_i\bar{Y}^T) = \frac{1}{N}E\left(Y_i\sum_{k=1}^{N}{Y_k}^T\right) = \frac{1}{N}E\left(Y_iY_i^T + (N-1)Y_iY_k^T\right) = \frac{1}{N}E(\Sigma + \mu\mu^T + (N-1)\mu\mu^T) = \frac{1}{N}E(\Sigma + N\mu\mu^T)$

Third, we have $E(\bar{Y}\bar{Y}^T) = \frac{1}{N^T}E\left((\sum_{k=1}^{N}Y_i)(\sum_{k=1}^{N}{Y_k}^T)\right) = \frac{1}{N^2}E\left(NY_iY_i^T + N(N-1)Y_iY_k^T\right) = \frac{1}{N^2}E(\mathrm{N}\Sigma + N\mu\mu^T + N(N-1)\mu\mu^T) = \frac{1}{N}E(\Sigma + N\mu\mu^T)$

Therefore, we have

$$E(S) = \frac{N}{N-1}E\left((Y_i - \bar{Y})(Y_i - \bar{Y})^T\right)$$
$$= \frac{N}{N-1}\left(\Sigma + \mu\mu^T - 2\left(\frac{1}{N}\Sigma + \mu\mu^T\right) + \left(\frac{1}{N}\Sigma + \mu\mu^T\right)\right) = \Sigma$$

Multivariate normal distribution:

If Y (n by 1) ~ MVN (μ, Σ), then the density function for Y is given by

$$f(Y) = (2\pi)^{-n/2}|\Sigma|^{-1/2}e^{-\frac{1}{2}(Y-\mu)^T\Sigma^{-1}(Y-\mu)}$$

Because $\int f(Y)dY = \int_{y_1}\int_{y_2}\cdots\int_{y_n} f(Y)dy_n \ldots dy_2dy_1 = 1$

Therefore $\int e^{-\frac{1}{2}(Y-\mu)^T\Sigma^{-1}(Y-\mu)}dY = (2\pi)^{n/2}|\Sigma|^{1/2}$

One can verify that $E(Y) = \mu$, and $V(Y) = \Sigma$.

Malhalanobis Distance:

-- In the univariate setting, standardized residuals are often used to measure the distance of value from the mean

-- In the multivariate setting, we define

$$D^2 = (Y - \mu)^T \Sigma^{-1} (Y - \mu)$$

And the observed value

$$O_i^2 = (Y_i - \bar{Y})^T \mathrm{S}^{-1} (Y - \bar{Y})$$

Transformation of MVN random variables:

When $Y \sim MVN\ (\mu, \Sigma)$, then $Z = AY + C$ also follows a MVN distribution, where A is an r by n matrix; C is an r by 1 vector; Z is an r by 1 vector.

$$Z \sim MVN\ (A\mu + C, A\Sigma A^T)$$

Also,

$$D^2 = (Y - \mu)^T \Sigma^{-1} (Y - \mu) \sim \chi_n^2$$

The above holds because $Y \sim MVN\ (\mu, \Sigma)$ -> $Y - \mu \sim MVN\ (0, \Sigma)$ -> $\Sigma^{-\frac{1}{2}}(Y - \mu) \sim MVN\ (0, \mathrm{I})$ -> D^2 is a summation of the square of n random variables that follow a standard normal distribution.

Marginals of MVN:

If $Y \sim MVN\ (\mu, \Sigma)$, then $Y_i \sim N(\mu_i, \Sigma_{ii})$.

Joint MVN implies marginal normality. But the reverse is not true.

Conditional of MVN:

We are interested in split the Y vector into part: $Y = (Y_1, Y_2)$, where Y1 is a q by 1 vector and Y2 is an r by 1 vector, q+r=n.

Then the conditional distribution of Y1|Y2 is also a MVN:

$$Y_1|Y_2 \sim MVN(\mu_{1|2}, \Sigma_{1|2})$$

Where:

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Distribution of sample mean and variance

If $Y \sim MVN\ (\mu, \Sigma)$, then

$$\bar{Y} \sim MVN\ (\mu, \Sigma/N)$$

$$(N-1)S \sim W_n(\Sigma, N-1)$$

Where a random n by n matrix W is said to follow an n-dimensional Wishart distribution $W_n(\Sigma, N)$, with N degree of freedom and parameter Σ, if W can be represented as $W = \sum_j X_j X_j^T$, where Xj, j=1, …, N, are i.i.d. from MVN (0, Σ). Provided that N>n, the density function of W is given by

$$f(W;\ \Sigma, N) = \frac{|W|^{1/2(N-n-1)} e^{tr(-1/2\Sigma^{-1}W)}}{2^{\frac{nN}{2}} |\Sigma|^{\frac{N}{2}} \Gamma_n(\frac{N}{2})}$$

For n=1, with $\Sigma = \sigma^2$, then $W_1(\sigma^2, N) = \sigma^2 \chi_N^2$.

In addition, $E(W) = N\Sigma$.