Inference: Consider the cause-effect situation:

$$X \to Y$$

where the link function between X and Y is $\mu(\cdot)$. The goal of statistical inference is to determine the link function $\mu(\cdot)$ based on data:

$$(X_1, Y_1), ..., (X_n, Y_n).$$

The most common choice of μ is the conditional mean of Y given X, $E(Y|X)$. Another possibility is the $Med(Y|X)$.

A common assumption on the form of this expectation is that the relationship between X and Y is linear, i.e. $\mu(X) = XB$.

We consider the Multivariate General Linear Model.

Background on vector and Matrix differentiation:

Derivative with respect to $\mathbf{X}$, an $n \times p$ matrix is:

$$\partial f(\mathbf{X})/\partial \mathbf{X} = (\partial f(\mathbf{X})/\partial x_{ij},$$

Results on Differentiation:

1) $\partial \mathbf{a}'\mathbf{x}/\partial \mathbf{x} = \mathbf{a}$
2) a) $\partial \mathbf{x}'\mathbf{x}/\partial \mathbf{x} = 2\mathbf{x}$ and b) $\partial \mathbf{x}'\mathbf{A}\mathbf{x}/\partial \mathbf{x} = (\mathbf{A}' + \mathbf{A})\mathbf{x}$
3) $\partial \mathbf{x}'\mathbf{A}\mathbf{y}/\partial \mathbf{x} = \mathbf{A}\mathbf{y}$
4) Vector $\mathbf{x}$ minimizing $(\mathbf{y} - \mathbf{A}\mathbf{x})'(\mathbf{y} - \mathbf{A}\mathbf{x})$ satisfies:

$$\mathbf{A}'\mathbf{A}\mathbf{x} = \mathbf{A}'\mathbf{y}$$

Setup for inference in the general linear model:

Let $(\mathbf{x_1}, \mathbf{y_1}), ..., (\mathbf{x_n}, \mathbf{y_n})$ be a random sample from the distribution $\mathbf{x}, \mathbf{y}$, where $\mathbf{x} \in \Re^q$ and $\mathbf{y} \in \Re^p$. Let $\mathbf{Y}' = (\mathbf{y_1}, ..., \mathbf{y_n})$ and $\mathbf{X}' = (\mathbf{x_1}, ..., \mathbf{x_n})$. Let $\mathbf{B}$ be a $q \times p$ matrix of (non-random) parameters. Then the GLM is:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E},$$

where $\mathbf{E}$ is an $n \times p$ data matrix of errors from $N_p(\mathbf{0}, \mathbf{\Sigma})$. I.e., $\mathbf{e_i} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$. This implies that $\mathbf{y_i} \sim N_p(\mathbf{B}'\mathbf{x_i}, \mathbf{\Sigma})$.

Two main methods to Estimate $\mathbf{B}$:

1) Least Squares

$\hat{\mathbf{B}}$ is defined to be the value of $\mathbf{B}$ that minimizes:

$$|(\mathbf{Y} - \mathbf{X}\mathbf{B})'(\mathbf{Y} - \mathbf{X}\mathbf{B})|.$$

Result: $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and $E(\hat{\mathbf{B}}) = \mathbf{B}$.

Sketch of Proof:

a) Show that $(\mathbf{Y} - \mathbf{X}\mathbf{B})'(\mathbf{Y} - \mathbf{X}\mathbf{B}) =$

$$(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) + (\hat{\mathbf{B}}' - \mathbf{B}')\mathbf{X}'\mathbf{X}(\hat{\mathbf{B}} - \mathbf{B})$$

b) Use $|\mathbf{A}| \geq |\mathbf{B}|$ when $\mathbf{A} - \mathbf{B}$ is p.s.d.

2) Maximum Likelihood Recall in univariate setting:

Joint density function, $f(\mathbf{x};\theta)$, is viewed as a function of x given parameter values θ. The likelihood function $L(\mathbf{x};\theta)$ is viewed as a function of the parameter θ given the data $\mathbf{x}$. Similar idea in the multivariate setting.