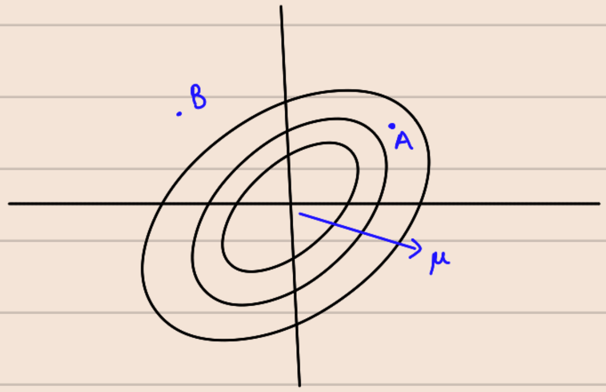
Mahalanobis distance: (MD)

→ Motivation

If $X \sim N_p(\mu, \Sigma)$ then pdf of X is

$$f_X(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$



A and B have same euclidean distance from μ but contour wise B is farther than A.

Def^n: Mahalanobis distance between point $\underline{x}$ and pop^n $N_p(\mu, \Sigma)$ is defined as

$$MD(\underline{x}, N_p(\underline{\mu}, \Sigma)) = (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})$$

Sample Mahalanobis distance:

Let $X_1, \ldots, X_n$ be a sample from $N_p(\mu, \Sigma)$

Estimator of μ, $\hat{\mu} = \bar{\underline{X}} = \frac{1}{n} \sum_{i=1}^{n} \underline{X}_i$

Estimator of Σ, $\hat{\Sigma} = S_{n-1}$

Then sample MD between point $\underline{y}$ and sample $\{X_1, \ldots, X_n\}$ is given by

$$\widehat{MD}(\underline{y}, \{x_1, \ldots, x_n\}) = (\underline{y} - \bar{\underline{X}})^T S^{-1} (\underline{y} - \bar{\underline{X}})$$

→ Mahalanobis distance between two popⁿ

$N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$

$$MD\left(N_p(\mu_1, \Sigma), N_p(\mu_2, \Sigma)\right) = (\underline{\mu}_1 - \underline{\mu}_2)^T \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$$

→ Mahalanobis distance between two popⁿ (Sample Version)

$X_1, \ldots, X_n$ iid $N_p(\mu_1, \Sigma)$

$Y_1, \ldots, Y_m$ iid $N_p(\mu_2, \Sigma)$

$$MD = (\underline{\bar{x}} - \underline{\bar{y}})^T \underbrace{\Sigma^{-1}}_{\text{pooled sample covariance}} (\underline{\bar{x}} - \underline{\bar{y}})$$

Regression:

$$\underbrace{Y}_{\text{expenditure}} = \underbrace{\beta_0}_{\text{param - I}} + \underbrace{\beta_1}_{\text{param - II}} \underbrace{X}_{\text{income}} + \underbrace{\varepsilon}_{\text{random error}}$$

Note: This topic is around section 5.2.2 in AMSA

→ $Y = \underline{X}^T \underline{\beta} + \varepsilon$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$\therefore \; Y = X\beta + \varepsilon$

Least Square :

$$\varepsilon = Y - X\beta$$

The square error loss

$$Q = \varepsilon^T \varepsilon$$
$$= \sum_{i=1}^{n} \varepsilon_i^2$$
$$= (Y - X\beta)^T (Y - X\beta)$$

Ordinary least square (OLS) estimator is

$$\hat{\beta} = \arg\min_{\beta} Q$$

By solving we get,

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T Y}$$

→ Table 5.1 ANOVA table. (AMSA)