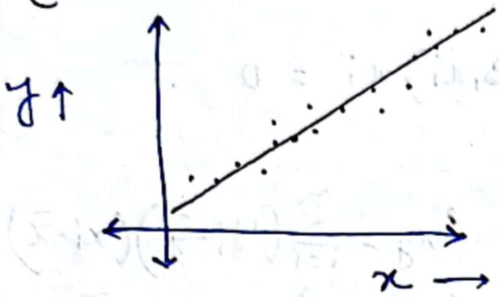
// Linear Regression & Anova //

31/07/2023

Simple Linear Regression

$\begin{cases} y : \text{Dependent variable} \\ x : \text{Regressor / Independent variable} \end{cases}$



Model

$$y = \beta_0 + \beta_1 x + \epsilon .$$

Putting the observations, $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1(1)n$.

We can replace y by $f(y)$ & x by $g(x)$.

$$f(y) = \beta_0 + \beta_1 g(x) + \epsilon .$$

[Linearity in parameters]

How to estimate β_0, β_1?

$$\underset{\sim}{\beta} = (\beta_0, \beta_1)^T$$

Define $S(\underset{\sim}{\beta}) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$

Assumptions on ϵ:

- $E(\epsilon_i) = 0$, $\forall i$
- $Var(\epsilon_i) = \sigma^2$, $\forall i$
- $cov(\epsilon_i, \epsilon_j) = 0$ if $i \neq j$

$$\underset{\underset{\sim}{\beta} \in \mathbb{R}^2}{\text{Min}} \; S(\underset{\sim}{\beta}).$$

$$\frac{\partial S(\underset{\sim}{\beta})}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial S(\underset{\sim}{\beta})}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \; .$$

$$\left\{ \begin{array}{l} \boxed{\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}}} \\ \\ \boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}} \end{array} \right. \text{, where } \left\{ \begin{array}{l} S_{xy} = \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) \\ S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \\ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \\ \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \end{array} \right.$$

$\rightarrow$ LEAST SQUARE ESTIMATORS.

Then, $\boxed{\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x}$

▶ SIGNIFICANCE OF THE REGRESSOR

$$y = \beta_0 + \beta_1 x.$$

Testing Problem $\longrightarrow$ $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$.

▶ PROPERTIES OF THE ESTIMATOR

$\longrightarrow$ $\boxed{E(\hat{\beta}_1) = ?}$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

Now, $E(y_i) = \beta_0+\beta_1 x_i \ \forall i = 1(1)n$.

$$E(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i-\bar{x}) \cdot \beta_1(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

$$= \beta_1 \cdot \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

$$= \beta_1.$$

Similarly, $\underline{E(\hat{\beta}_0) = \beta_0}$.

$$\left[\begin{aligned} E(\hat{\beta}_0) &= E(\bar{y}) - E(\hat{\beta}_1).\bar{x} \\ &= \beta_0+\beta_1\bar{x} - \beta_1\bar{x} \\ &= \beta_0 \end{aligned}\right]$$

$$\frac{\hat{\beta}_1 - E(\hat{\beta}_1)}{\sqrt{var(\hat{\beta}_1)}} \longrightarrow 0, \text{ under } H_0$$

02/08/2023

VARIANCE OF ESTIMATORS

$$var(\hat{\beta}_1) = var\left(\sum c_i(y_i-\bar{y})\right).$$

$$c_i = \frac{(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}.$$

$$\cancel{= \sum_{i=1}^{n} c_i^2 \, var(y_i-\bar{y})}.$$

$$\cancel{= \sum_{i=1}^{n} c_i^2 \left[var(y_i) + var(\bar{y}) - 2cov(y_i,\bar{y})\right]}.$$

$$\cancel{= \sum_{i=1}^{n} c_i^2 \left[\sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n}\right]}.$$

Scanned with CamScanner

$$= \cancel{\sum_{i=1}^{n} c_i^2 \sigma^2 \left(1-\frac{1}{n}\right)}$$

$$= \cancel{\sigma^2\left(1-\frac{1}{n}\right) \cdot \sum_{i=1}^{n} c_i^2}$$

$$= \cancel{\sigma^2\left(1-\frac{1}{n}\right) \cdot \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}}$$

$$= \cancel{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\left(1-\frac{1}{n}\right)}.$$

$var(\hat{\beta}_1)$

$$= var\left(\sum_{i=1}^{n} c_i(y_i-\bar{y})\right)$$

$$c_i = \frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$= \frac{var\left(\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right)}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}$$

$$= \frac{var\left(\sum_{i=1}^{n}(x_i-\bar{x})y_i\right)}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}$$

$$= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 \, var(y_i)}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}$$

$$= \sigma^2 \cdot \frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$var(\hat{\beta}_0)$

$$= var(\bar{y} - \hat{\beta}_1 \bar{x})$$

$$= var\left(\bar{y} - \bar{x} \cdot \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}\right)$$

$$\cancel{= var(\bar{y}) +} = var\left(\bar{y} - \frac{\bar{x}}{\sum(x_i-\bar{x})^2}\sum(x_i-\bar{x})y_i\right)$$

$$= var(\bar{y}) + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2} var(y_i)$$

$$- 2\, cov\left(\bar{y}, \frac{\bar{x}}{\sum(x_i-\bar{x})^2} \cdot \sum(x_i-\bar{x})y_i\right)$$

$$= \frac{\sigma^2}{n} + \sigma^2 \cdot \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2} - \frac{2}{n} \cdot \frac{\bar{x}}{\sum(x_i-\bar{x})^2} cov\left(\sum y_i, \sum y_i(x_i-\bar{x})\right)$$

$$= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right) - \frac{2}{n} \cdot \frac{\bar{x}}{\sum(x_i-\bar{x})^2} \cdot \left[\cancel{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-\bar{x})\, cov(y_i,y_j)}\right]^{0}$$

$$= \sigma^2\left(\frac{\sum x_i^2 - \cancel{n\bar{x}^2} + \cancel{n\bar{x}^2}}{n\sum(x_i-\bar{x})^2}\right)$$

$$= \sigma^2 \cdot \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Another Method of Estimation

Let's assume, we have additional information about the error terms.

Let, $\epsilon_i \sim N(0, \sigma^2)$.

Then, with $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1(1)n$.

We have,

$$L(\underset{\sim}{y}, \underset{\sim}{\beta}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}$$

Maximum Likelihood Estimators (MLE)

$$\log L = -\frac{n}{2}\log\sigma^2 - \frac{\sum(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} + c.$$

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^4} = 0$$

$$\Rightarrow \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

[In this method, we also get the estimate of σ^2 unlike method of least squares].

Multiple Regression

Model: $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon$.

Here,

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \dots + \beta_k x_{1k} + \epsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \dots + \beta_k x_{2k} + \epsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \dots + \beta_k x_{nk} + \epsilon_n \end{cases}$$

$$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}$$

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \qquad \underset{\sim}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$ Here, $\underline{\underline{p = k+1}}$

$$\therefore \boxed{\underset{\sim}{y}_{n\times 1} = X^{n\times p} \underset{\sim}{\beta} + \underset{\sim}{\epsilon}^{n\times 1}.}$$ [Here, we need $p \geq n$]

LSEs

$$\underset{\sim}{\epsilon}^T \underset{\sim}{\epsilon} = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$ [For simple LR]

Here, $\underset{\sim}{\epsilon}^T \underset{\sim}{\epsilon} = (\underset{\sim}{y} - X\underset{\sim}{\beta})^T (\underset{\sim}{y} - X\underset{\sim}{\beta}) = S(\beta).$

$$\min_{\underset{\sim}{\beta} \in \mathbb{R}^p} S(\beta)$$

Now, $S(\beta) = \underset{\sim}{y}^T \underset{\sim}{y} - \underset{\sim}{\beta}^T X^T \underset{\sim}{y} - \underset{\sim}{y}^T X \underset{\sim}{\beta} + \underset{\sim}{\beta}^T X^T X \underset{\sim}{\beta}.$

Results

<1> Let $\alpha = \underset{\sim}{y}^T A \underset{\sim}{x}$.

$$\frac{\partial \alpha}{\partial \underset{\sim}{x}} = \underset{\sim}{y}^T A, \quad \frac{\partial \alpha}{\partial \underset{\sim}{y}} = \underset{\sim}{x}^T A^T$$

<2> $\alpha = \underset{\sim}{x}^T A \underset{\sim}{x}$ $\quad \dfrac{\partial \alpha}{\partial \underset{\sim}{x}} = \underset{\sim}{x}^T (A + A^T)$

~~$\frac{\partial S(\beta)}{\partial \beta}$~~

Then, $\frac{\partial S(\beta)}{\partial \underset{\sim}{\beta}} = -\underset{\sim}{y}^T X - \underset{\sim}{y}^T X + \beta^T (X^T X + X^T X) = 0.$

From this,

$$\boxed{\hat{\beta}_{LSE} = (X^T X)^{-1} X^T \underset{\sim}{y}}$$

[Here, X is of full rank]

Now, $\frac{\partial^2 S(\beta)}{\partial \beta^T \partial \beta} = X^T X$, which is positive definite.

[Proof: For $\underset{\sim}{z} \in \mathbb{R}^p$. $\underset{\sim}{z}^T X^T X \underset{\sim}{z}$

$= (X\underset{\sim}{z})^T X \underset{\sim}{z}$

$= v^T v \geq 0$.

and $Xz = 0$.

$\Rightarrow [\underset{\sim}{1}, \underset{\sim}{x_1}, \underset{\sim}{x_2}, \dots, \underset{\sim}{x_k}] \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ \vdots \\ z_k \end{bmatrix} = 0.$

$\Rightarrow \sum_{i=0}^{k} z_i \underset{\sim}{x_i} = 0.$

$\Rightarrow z_i = 0 \ \forall i$ [As X is of full column rank]]

Properties of the estimators :-

$\longrightarrow E(\hat{\beta}) = (X^T X)^{-1} X^T E(Y).$

$= (X^T X)^{-1} X^T X \underset{\sim}{\beta}.$

$= \underset{\sim}{\beta}.$

Hence, the estimator is unbiased.

$\longrightarrow \; Var(\underset{\sim}{\epsilon}) = \sigma^2 I_n$

$$Var(\hat{\beta}) = Var((X^T X)^{-1} X^T \underset{\sim}{y})$$

$$\cancel{[redacted]} = (X^T X^{-1})^{-1} X^T \, Var(\underset{\sim}{y}) . ((X^T X)^{-1} X^T)^T$$

$$= (X^T X)^{-1} X^T . (\sigma^2 I_n) . X (X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1} \cancel{(X^T X)} (X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1}$$

04/08/2023

$$\hat{\sigma}^2_{MLE} = \frac{(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})^T (\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}})}{n} = \frac{\underset{\sim}{y}^T H \underset{\sim}{y}}{n}$$

Now, $y - X\hat{\beta} = y - X(X^T X)^{-1} X^T y$

$$= \underbrace{(I - X(X^T X)^{-1} X^T)}_{H} y$$

Properties of 'H' matrix:

① $H^T = H$

② $H^2 = H$.

[Proof: $H^2 = (I - X(X^T X)^{-1} X^T)(I - X(X^T X)^{-1} X^T)$

$$= I - X(X^T X)^{-1} X^T - X(X^T X)^{-1} X^T + X(X^T X)^{-1} X^T$$

$$= I - X(X^T X)^{-1} X^T$$

$$= H]$$

RESULT: Z is a RV. with $E(Z) = \mu$, $Var(Z) = \Sigma$.

then, $E(Z^T A Z) = trace(A\Sigma) + \mu^T A \mu$.

Now, $E(Y) = X\beta$.

$\Sigma = Var(Y) = \sigma^2 I_n$.

$\therefore E(\hat{\sigma}^2_{MLE}) = E\left(\frac{Y^T H Y}{n}\right) = \frac{1}{n}\left[trace(H \cdot \sigma^2 I_n) + \beta^T X^T H X\beta\right]$

Now, $HX = (I - X(X^TX)^{-1}X^T)X$
$= X - X = 0$.

$\sigma^2 \, tr(H) = \sigma^2 (tr(I - X(X^TX)^{-1}X^T))$
$= \sigma^2 (tr(I_n) - tr(X(X^TX)^{-1}X^T))$.
$= \sigma^2 (n - p)$ $\quad [\because tr(X(X^TX)^{-1}X^T)$
$= tr(X^TX(X^TX)^{-1})$
$= tr(I_p)$.
$= p]$

$E\left(\frac{Y^T H Y}{n}\right) = \sigma^2\left(\frac{n-p}{n}\right)$.

$\therefore E\left(\frac{Y^T H Y}{n-p}\right) = \sigma^2$.

$$\boxed{\hat{\sigma}^2 = \frac{Y^T H Y}{n-p}}$$

Error sum of squares. (SSE)

$\therefore \hat{\sigma}^2 = \frac{SSE}{n-p} = MSE$

THEOREM.

(GAUSS-MARKOV THEOREM)

Under Gauss-Markov Assumptions, $\boxed{\hat{\beta}_{MLE/LSE} = (X^TX)^{-1}X^TY}$ is "Best Unbiased Linear Estimator" of β.

[Gauss-Markov Assumptions:
↳ $E(\epsilon_i)=0$
↳ $V(\epsilon_i)=\sigma$
↳ $cov(\epsilon_i,\epsilon_j)=0 \ \forall i,j$]

LINEAR ESTIMATORS → $\mathbb{C} = \{AY+b_0 : b_0 \in \mathbb{R}^p\}$.

'BEST' is in terms of variance as —

$$\boxed{var(l^T \hat{\beta}_{LSE}) \leq var(l^T \tilde{\beta})} \quad \forall \tilde{\beta} \in \mathbb{C}, \ \forall l \in \mathbb{R}^p.$$

PROOF 1) $\tilde{\beta} \in \mathbb{C} \quad \therefore \tilde{\beta} = AY + b_0$ for some A & b_0

Choose $A = ((X^TX)^{-1}X^T + B)$

$$E(\tilde{\beta}) = ((X^TX)^{-1}X^T + B)X\beta + b_0$$
$$= \beta + BX\beta + b_0.$$

Now, to make $\tilde{\beta}$ unbiased

$BX\beta + b_0 = 0 \ \forall\beta$.

$\Rightarrow b_0 = 0 \overset{\forall\beta}{\Rightarrow} BX\beta = 0 \Rightarrow BX = 0$.

Scanned with CamScanner

$$\text{var}(\tilde{\beta}) = \text{var}\left(\left[(X^TX)^{-1}X^T + B\right]Y\right).$$

$$= \sigma^2 \left[(X^TX)^{-1}X^T + B\right]\left[X(X^TX)^{-1} + B^T\right]$$

$$= \sigma^2 \left[(X^TX)^{-1} + BX(X^TX)^{-1} + (X^TX)^{-1}X^TB^T + BB^T\right]$$

$$= \sigma^2 \left[(X^TX)^{-1} + BB^T\right]. \text{ [Putting } BX = 0]$$

$$= \sigma^2 (X^TX)^{-1} + \sigma^2 BB^T.$$

$$= \text{Var}(\hat{\beta}_{LSE}) + \sigma^2 BB^T$$

$$\text{Var}(l^T\tilde{\beta}) = \text{Var}(l^T\hat{\beta}_{LSE}) + \sigma^2 \underbrace{l^TBB^Tl}_{\geq 0}.$$

$$\geq \text{Var}(l^T\hat{\beta}_{LSE}).$$

—·—

07/08/2023

Multivariate Normal Distribution

Defn. Let, $\underset{\sim}{u} = (u_1, u_2, \ldots, u_n)'$, where $u_i \sim N(0,1)$ IID $\forall i=1(1)n$

Let, $X_{n \times n}$ is a full rank matrix and $\underset{\sim}{\mu}_{n\times 1}$ is a vector.

and $\underset{\sim}{Z}_{n\times 1} = X\underset{\sim}{u} + \underset{\sim}{\mu}$.

So, $E(\underset{\sim}{Z}) = \mu$,

$\text{var}(\underset{\sim}{Z}) = XX^T = \Sigma_{n\times n}$ (say)

CHARACTERISTIC FUNCTION

$$\boxed{\phi_Z(t) = E[\exp(it^T Z)]}$$

First, let us find CF of $\underset{\sim}{u} \sim N_n(0, I_n)$.

$$\phi_{\underset{\sim}{u}}(t) = E\left[e^{it^T \underset{\sim}{u}}\right]$$

$$= E\left[e^{i\sum_{j=1}^{n} t_j u_j}\right]$$

$$= \prod_{j=1}^{n} E\left(e^{it_j u_j}\right).$$

Now, $E\left(e^{it_j u_j}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u_j^2/2} e^{it_j u_j} \, du_j$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(u_j - it_j)^2 - \frac{t_j^2}{2}\right] du_j$$

$$= e^{-t_j^2/2} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2(u_j - it_j)^2} \, du_j$$

$$= e^{-t_j^2/2}.$$

$$\phi_{\underset{\sim}{u}}(t) = \prod_{j=1}^{n} e^{-t_j^2/2}$$

$$= e^{-\frac{\underset{\sim}{t}^T \underset{\sim}{t}}{2}} \qquad \underset{\sim}{u} \sim N(\underset{\sim}{0}, I_n).$$

$$\phi_{\underset{\sim}{Z}}(t) = E\left[\exp\{it^T z\}\right]$$

$$= E\left[\exp\left(it^T(X\underset{\sim}{u} + \underset{\sim}{\mu})\right)\right]$$

$$= \exp(it^T \underset{\sim}{\mu}) \; E\left[\exp(it^T X \underset{\sim}{u})\right]$$

$$\cancel{= e^{it\mu} \phi_{y}(t^T x)}$$

$$\cancel{= e^{it\mu} e^{-\frac{1}{2}(t^T x)^T (t^T x)}}$$

$$= e^{it^T\mu} e^{-\frac{1}{2}(x^T t)^T (x^T t)}$$

$$= e^{it^T\mu} e^{-\frac{1}{2} t^T \underbrace{x x^T}_{\Sigma} t}$$

$$= e^{it^T\mu - \frac{1}{2} t^T \Sigma t}$$

$\boxed{\text{PDF}}$ $$f(z) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(z-\mu)^T \Sigma^{-1} (z-\mu)\right]$$

Let us define, $\omega = \Sigma^{-1/2}(z-\mu)$.

$$(z-\mu)^T \Sigma^{-1} (z-\mu) = \omega^T \omega .$$

Now, $\int_{-\infty}^{\infty} f(z)\, dz$

$$= \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\omega^T\omega} \cdot |\Sigma|^{1/2}\, d\omega . \quad \left[|J| = \left| \frac{dz}{d\omega} \right| = |\Sigma|^{1/2} \right]$$

$$= \frac{1}{(2\pi)^{n/2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\sum_{i=1}^{n} \omega_i^2}\, d\omega$$

$$= \prod_{i=1}^{n} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\omega_i^2/2}\, d\omega_i \right)$$

$$= 1 .$$

Let's show that this PDF corresponds to same CF.

$$\phi_{\underset{\sim}{Z}}(t) = E\left(e^{it^T \underset{\sim}{Z}}\right)$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{it^T \underset{\sim}{z}} f(\underset{\sim}{z})\, d\underset{\sim}{z}.$$

put $\underset{\sim}{w} = \underset{\sim}{z} - \underset{\sim}{\mu}$.

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{it^T(\underset{\sim}{w} + \underset{\sim}{\mu})} \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \cdot \exp\left(-\frac{1}{2} \underset{\sim}{w}^T \Sigma^{-1} \underset{\sim}{w}\right) d\underset{\sim}{w}$$

$$= \exp\left(it^T \underset{\sim}{\mu} - \frac{1}{2} \underset{\sim}{t}^T \Sigma \underset{\sim}{t}\right) \quad [\text{show}]$$

$\boxed{09/08/2023}$

Theorem Let $\underset{\sim}{y} \sim N(0, I_n)$ and let A be a symmetric $n \times n$ matrix. Then $\underset{\sim}{y}^T A \underset{\sim}{y}$ has a chi-squared distribution with r d.f. iff A is idempotent & $\text{Rank}(A) = r$.

Proof If $A^2 = A$ and $\text{rank}(A) = r$.

Now, $A = P^T D P$ as A is symmetric.

$$= P^T \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} P, \quad P \text{ is orthogonal matrix}$$

Now, if $\underset{\sim}{z} = P\underset{\sim}{y}$ then $Z \sim N(0, I_n)$.

$$\underset{\sim}{y}^T A \underset{\sim}{y} = \underset{\sim}{y}^T P^T D P Y = Z^T D Z = \sum_{i=1}^{r} Z_i^2 \sim \chi^2_r$$

Conversely, given, $\underset{\sim}{y}' A \underset{\sim}{y} \sim \chi^2_r$

Since, A is symmetric $\Rightarrow A = P^T \operatorname{diag}(\lambda_1, \ldots, \lambda_n) P$.

[λ_i's are the eigen. values of A]

[P is an orthogonal matrix]

Take, $Z = P\underset{\sim}{y} \sim N(0, I_n)$

$$\phi(t) = E\left[e^{it\, y^T A y}\right] \quad [\text{CF of } \underset{\sim}{y}^T A \underset{\sim}{y}]$$

$$= E\left[e^{it\, z^T D z}\right]$$

$$= E\left[e^{it \sum_{j=1}^{n} \lambda_j z_j^2}\right]$$

$$= \prod_{j=1}^{n} E\left(e^{it\lambda_j z_j^2}\right).$$

$$= \prod_{j=1}^{n} (1 - 2it\lambda_j)^{-1/2} \quad [\text{Since, } z_j^2 \sim \chi^2_1] \quad -(*)$$

Now, we know, $\underset{\sim}{y}^T A \underset{\sim}{y} \sim \chi^2_r$.

$$\therefore \phi(t) = (1-2it)^{-r/2} \quad -(**)$$

Equating $(*)$ & $(**)$,

$$(1-2it)^{r} = \prod_{j=1}^{n} (1 - 2it\lambda_j)$$

Now, LHS is a 'r' degree polynomial in t with roots $1/2i, \ldots, 1/2i$ (r times).

and, RHS is a 'n' degree polynomial in t, so, $(n-r)$ roots are 0. [As, LHS is r degree pol.]

For the rest of the roots,

$$\cancel{1 - 2i\lambda_j t = \frac{1}{2i} \Rightarrow \lambda_j = 1}$$

$$\cancel{\text{So, } \lambda_j = \begin{cases} 1, \forall j = 1, 2, \ldots, r \text{ (WLG)} \\ 0, \forall j = r+1, \ldots, n \end{cases}}$$

$$1 - 2i \cdot \lambda_j \cdot \frac{1}{2i} = 0 \Rightarrow \lambda_j = 1.$$

$$\text{So, } \lambda_j = \begin{cases} 1, \forall j = 1, 2, \ldots, r \text{ (WLG)} \\ 0, \forall j = r+1, \ldots, n \end{cases}$$

A is a matrix with ~~...~~ eigenvalues 1 (r times) & 0 ($n-r$ times).

$\Rightarrow$ A is idempotent & $r(A) = r$ (Proved).

$\boxed{\text{THEOREM}}$ Let, $\underset{\sim}{y} \sim N(\underset{\sim}{\mu}, \Sigma)$ and suppose, $\underset{\sim}{y}, \underset{\sim}{\mu}$ & Σ are conformally partitioned. Then $\underset{\sim}{y}_1$ & $\underset{\sim}{y}_2$ are independent iff $\Sigma_{12} = 0$.

only if> Independent $\Rightarrow cov(\underset{\sim}{y}_1, \underset{\sim}{y}_2) = \Sigma_{12} = 0$.

if part>

$$\phi_{\underset{\sim}{y}}(t) = exp\left[i t^T \underset{\sim}{\mu} - \frac{1}{2} t^T \Sigma t\right]$$

$$= exp\left[i \left[t_1^T \; t_2^T\right] \begin{bmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{bmatrix} - \frac{1}{2} \left[t_1^T \; t_2^T\right] \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}\right]$$

$$= exp\left[i(t_1^T \mu_1 + t_2^T \mu_2) - \frac{1}{2}(t_1^T \Sigma_{11} t_1 + t_2^T \Sigma_{22} t_2)\right]$$

$$= exp\left[i t_1^T \mu_1 - \frac{1}{2} t_1^T \Sigma_{11} t_1\right] \cdot exp\left[i t_2^T \mu_2 - \frac{1}{2} t_2^T \Sigma_{22} t_2\right]$$

$$= \phi_{\underset{\sim}{y}_1}(t_1) \cdot \phi_{\underset{\sim}{y}_2}(t_2).$$

AB^{-} ...

Theorem > $\underset{\sim}{y} \sim N(\underset{\sim}{\mu}, \sigma^2 I)$. Let A & B are matrices such that $AB^T = 0$ then $A\underset{\sim}{y}$ & $B\underset{\sim}{y}$ are independent.

Proof > $\text{cov}(A\underset{\sim}{y}, B\underset{\sim}{y})$ $\qquad \begin{bmatrix} A\underset{\sim}{y} \\ B\underset{\sim}{y} \end{bmatrix} = \begin{bmatrix} A \\ B \end{bmatrix} \underset{\sim}{y}.$

$= \sigma^2 AB^T$

$= 0.$

-.-

11/08/2023

Theorem > Let $\underset{\sim}{y} \sim N(0, I_n)$ and let A_1 and A_2 be symmetric idempotent matrices. Then $\underset{\sim}{y}^T A_1 \underset{\sim}{y}$ & $\underset{\sim}{y}^T A_2 \underset{\sim}{y}$ are independent iff $A_1 A_2 = 0$.

$\boxed{\text{If part}}$

Proof > $A_1 A_2 = 0$

$\Rightarrow A_1 \underset{\sim}{y}$ and $A_2 \underset{\sim}{y}$ are independent [$A_2 = A_2^T$ and using above theorem]

$\underset{\sim}{y}^T A_1 \underset{\sim}{y} = y^T A_1 A_1 y$

$= (A_1 y)^T (A_1 y)$

$\underset{\sim}{y}^T A_2 \underset{\sim}{y} = y^T A_2 A_2 y = (A_2 y)^T (A_2 y)$

$\therefore \underset{\sim}{y}^T A_1 \underset{\sim}{y}$ & $\underset{\sim}{y}^T A_2 \underset{\sim}{y}$ are independent [Since, function of two independent entities are also independent]

$\boxed{\text{Only if part}}$

Given: $y^T A_1 y$ and $y^T A_2 y$ are independent.

$$y^T A_1 y + y^T A_2 y$$

$$= y^T (A_1 + A_2) y$$

$$\sim \chi^2_{r_1 + r_2} \quad \begin{bmatrix} r_1 = \text{rank}(A_1) \\ r_2 = \text{rank}(A_2) \end{bmatrix}$$

and $A_1 + A_2$ is symmetric.

$\Rightarrow A_1 + A_2$ is idempotent.

$$(A_1 + A_2)(A_1 + A_2) = A_1 + A_2$$

$$\Rightarrow A_1 + A_1 A_2 + A_2 A_1 + A_2 = A_1 + A_2$$

$$\Rightarrow A_1 A_2 + A_2 A_1 = 0$$

$$\Rightarrow (A_1 A_2 + A_2 A_1) . A_2 = 0$$

$$\Rightarrow A_1 A_2 + A_2 A_1 A_2 = 0 . \text{ — } (**)$$

$$\Rightarrow A_2 (A_1 A_2 + A_2 A_1 A_2) = 0$$

$$\Rightarrow A_2 A_1 A_2 + A_2 . A_1 A_2 = 0 .$$

$$\Rightarrow A_2 A_1 A_2 = 0$$

Now putting in $(*)$, we get.

$$A_1 A_2 = 0$$

...

Theorem ⟩ Let $\underset{\sim}{y} \sim N(0, I_n)$. Let $A^T = A$, $A^2 = A$. Let $\underset{\sim}{l} \in \mathbb{R}^n$, $\underset{\sim}{l} \neq \underset{\sim}{0}$. Then, $y^T A y$ and $l^T y$ are independent iff $Al = 0$.

Proof: WLG, we assume $\|l\| = 1$.

Define ~~####~~ $ll^T = B$.

Only if part ⟩

~~First~~ Given: $y^T A y$ & $l^T y$ are independent

Then $y^T A y$ & $y^T l l^T y = y^T B y$ are independent

$$\Rightarrow AB = 0 \Rightarrow A l l^T = 0$$

$$\Rightarrow A l l^T l = 0$$

$$\Rightarrow A l = 0 \quad [\text{Since } l^T l = 1]$$

If part ⟩

Given: $Al = 0 \Rightarrow Ay$ and $l^T y$ are independent.

$$\Rightarrow y^T A y = (Ay)^T Ay.$$

$$\Rightarrow y^T A y \ \& \ l^T y \text{ are independent.}$$

Theorem: (Cochran's) ⟩ Let $A_1, A_2, \ldots, A_K$ be $n \times n$ matrices with $\sum_{i=1}^{K} A_i = I$. Then following conditions are equivalent:

(i) $\sum_{i=1}^{k} R(A_i) = n$

(ii) $A_i^2 = A_i \ \forall\ i = 1(1)k$.

(iii) $A_i A_j = 0 \ \forall i \neq j$

Proof → (i) ⇒ (iii)

$$\text{let, } A_i = \underset{n \times r_i}{B_i} \ \underset{r_i \times n}{C_i}$$

$$\sum A_i = I$$

$$\Rightarrow \sum B_i C_i = I$$

$$\Rightarrow [B_1 \ B_2 \ \dots \ B_k] \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{bmatrix} = I.$$

$$\Rightarrow B^* C^* = I.$$

$$\Rightarrow C^* B^* = I \quad [\text{Using Problem set 1, Q 1}]$$

$$\Rightarrow \begin{bmatrix} C_1 \\ \vdots \\ C_k \end{bmatrix} [B_1 \ B_2 \dots B_k] = I$$

$$\Rightarrow \begin{bmatrix} C_1B_1 & C_1B_2 & \dots & C_1B_k \\ & C_2B_2 & & \\ & & \ddots & C_kB_k \end{bmatrix} = \begin{bmatrix} I_1 & & \\ & I_2 & 0 \\ 0 & & \ddots \ I_k \end{bmatrix}$$

$$\Rightarrow C_i B_j = 0 \ \forall \ i \neq j$$

$$\Rightarrow A_i A_j = B_i \underline{\underline{C_i B_j}} C_j$$

$$= 0 \quad \forall i \neq j$$

(iii) ⟹ (ii), Since $\sum A_i = I$

$$A_j \left(\sum A_i\right) = A_j$$

$$\Rightarrow \sum A_i A_j + A_j^2 = A_j$$

$$\Rightarrow A_j^2 = A_j.$$

[As, From (iii) $A_i A_j = 0$]

(ii) ⟹ (i) Rank(A_i) = Trace(A_i).

Now, $\sum \text{Rank}(A_i) = \sum \text{trace}(A_i)$

$= \text{trace}(\sum A_i)$

$= \text{Trace}(I_n)$

$= n.$

...

▶ **COCHRAN'S THEOREM FOR QUADRATIC FORMS**

Let $U_1, U_2, \ldots, U_N \overset{IID}{\sim} N(0,1)$. Let $B^{(1)}, B^{(2)}, \ldots, B^{(k)}$ be symmetric matrices.

Define, $r_i = \text{Rank}(B^{(i)})$ $\forall i = 1(1)k$.

and $Q_i = U^T B^{(i)}_{N\times N} U$ where $U^T = (U_1, \ldots, U_N)$.

Now, if $\sum_{i=1}^{k} B^{(i)} = I$. Then the following statements are equivalent –

(i) $r_1 + r_2 + \ldots + r_k = N$

(ii) Q_i are chi-squared RV with df 'r_i'

(iii) Q_i are independent.

Back to Regression Analysis ...

Example > Consider the model →

$$\log F = \log C - \beta \log d :$$

represents the force of gravity between two bodies distance 'd' apart.

Say, we are interested in testing:

$$H_0 : \beta = 2 \quad \text{against} \quad H_1 : \beta \neq 2$$

Example > Full Model: $Y = X\beta + \epsilon$

~~$Y_i = \beta_0 + \beta_1 X$~~.

Say, we are interested in testing to determine whether some of the regressors are significant or not.

i.e. $H_0 : \beta_r = \beta_{r+1} = \cdots = \beta_{p-1} = 0$

vs. H_1: Not H_0.

In matrix notation,

$$H_0 : \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

More generally,

$$\left. \begin{aligned} H_0 &: A\underset{\sim}{\beta} = \underset{\sim}{c} \\ \text{vs. } H_1 &: A\underset{\sim}{\beta} \neq \underset{\sim}{c} \end{aligned} \right\}$$

Here, we can use LRT:

$$LR = \frac{\underset{\sigma,\beta \in \mathcal{H}_0}{L}(\underset{\sim}{y}, \underset{\sim}{\beta}, \sigma)}{\underset{\sigma,\beta \in \mathcal{H}_0 \cup \mathcal{H}_1}{L}(\underset{\sim}{y}, \underset{\sim}{\beta}, \sigma)}.$$

(Likelihood Ratio)

In $H_0 \cup H_1$ (full parameter space)

$$L(\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \|Y - X\beta\|^2\right]$$

and $\hat{\beta} = (X'X)^{-1} X'Y$

$$\hat{\sigma}^2 = \frac{\|Y - X\hat{\beta}\|^2}{n}.$$

and, $L(\hat{\beta}, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-n/2} e^{-n/2}$

Denominator of Likelihood Ratio.

Now, for restricted optimization i.e. $\|Y - X\beta\|^2$ given $A\beta = c$.

$$S(\beta) = (Y - X\beta)^T (Y - X\beta) + \lambda(A\beta - c).$$

$$\frac{\partial S(\beta)}{\partial \beta} = -2X^TY + 2X^TX\beta + A^T\lambda = 0.$$

$$\Rightarrow \hat{\beta}_R = (X^TX)^{-1}X^TY - \frac{1}{2}(X^TX)^{-1}A^T\lambda.$$

$$\frac{\partial S(\beta)}{\partial \lambda} = A\hat{\beta}_R = c.$$

$$\Rightarrow A\left[(X^TX)^{-1}X^TY - \frac{1}{2}(X^TX)^{-1}A^T\lambda\right] = c.$$

$$\Rightarrow -\frac{1}{2}\hat{\lambda} = \left[A(X^TX)^{-1}A^T\right]^{-1}\left[c - A\hat{\beta}\right]$$

$$\hat{\beta}_R = \hat{\beta} + (X^TX)^{-1}A^T\left[A(X^TX)^{-1}A^T\right]^{-1}\left[c - A\hat{\beta}\right].$$

$$\hat{\sigma}_R^2 = \frac{\|Y - X\hat{\beta}_R\|^2}{n}$$

$$L(\hat{\beta}_R, \hat{\sigma}_R) = (2\pi \hat{\sigma}_R^2)^{-n/2} e^{-n/2} \cdot \text{[Numerator of LR]}$$

$$\therefore LR = \Lambda = \frac{L(\hat{\beta}_R, \sigma_R^2)}{L(\hat{\beta}, \hat{\sigma}^2)} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_R^2}\right)^{n/2}$$

Therefore, the test statistic is given by –

$$\frac{n-p}{p}\left(\Lambda^{-2/n} - 1\right) \sim F_{q, n-p}.$$

[$q := \text{rank}(A)$]

16/08/2023

~~let our testing problem be –~~

~~$H_0: \beta_1 = \cdots = \beta_{p-1} = 0$~~

~~vs. H_1: not H_0.~~

RESULT // $Y \sim N(\underset{\sim}{\mu}, V)$ and $U = Y^T A Y$.

Then $U \sim \chi^2_{r, \lambda}$ if (AV) is idempotent

when $r(A) = r$ & $\lambda = \underline{\mu}^T A \underline{\mu}$ [non-centrality parameter]

Theorem //

i) $RSS_R - RSS = \|\hat{Y}_R - \hat{Y}\|^2$

$$= (A\hat{\beta} - c)^T \left[A (X^T X)^{-1} A^T\right]^{-1} (A\hat{\beta} - c).$$

ii) $E[RSS_R - RSS] = \sigma^2 q + (A\beta - c)^T \left[A (X^T X)^{-1} A^T\right]^{-1} (A\beta - c)$

iii) when H_0 is true, then $F = \frac{(RSS_R - RSS)/q}{RSS/n-p} \sim F_{q, n-p}$

$$RSS := (Y - \hat{Y})^T (Y - \hat{Y}).$$
$$= (Y - X\hat{\beta})^T (Y - X\hat{\beta}).$$

$$\& \ RSS_R = (Y - \cancel{X\beta}\ \hat{Y}_R)^T (Y - \hat{Y}_R)$$
$$= (Y - X\hat{\beta}_R)^T (Y - X\hat{\beta}_R^T).$$

Then $RSS_R - RSS$.

$$= (Y - X\hat{\beta}_R)^T (Y - X\hat{\beta}_R) - (Y - X\hat{\beta})^T (Y - X\hat{\beta}).$$

[where, $\hat{\beta}_R = \hat{\beta} + (X'X)^{-1} A' (A (X'X)^{-1} A')^{-1} (C - A\hat{\beta})$

$= \hat{\beta} + B$ (say)].

Therefore, $RSS_R - RSS$

$$= (Y - X\hat{\beta} - XB)^T (Y - X\hat{\beta} - XB) - (Y - X\hat{\beta})'(Y - X\hat{\beta})$$
$$= -B^T X^T (Y - X\hat{\beta}) - (Y - X\hat{\beta})^T X\beta + B^T X^T X B.$$
$$= B^T X^T X B \quad [\text{As, } X^T (Y - X\hat{\beta}) = X^T Y - X^T X (X^T X)^{-1} X' Y = 0]$$
$$= (C - A\hat{\beta})^T (A(X^TX)^{-1}A^T)^{-1} A(X^TX)^{-1} (X^TX) \cdot (X^TX)^{-1} A^T (A(X^TX)^{-1}A^T)^{-1} (C - A\hat{\beta})$$
$$= (C - A\hat{\beta})^T (A (X^TX)^{-1} A^T)^{-1} (C - A\hat{\beta}).$$

$$= \underbrace{(A\hat{\beta} - c)}_{y^{\star T}} \ (\underbrace{A\,(X^TX)^{-1}\,A^T}_{B^\star})^{-1} \ \underbrace{(A\hat{\beta} - C)}_{y^\star}.$$

$$= \|\hat{Y} - \hat{Y}_R\|^{\prime} \text{ (Show)}.$$

...

(ii). $E(RSS_R - RSS) = E(y^{\star\,T} B^\star y^\star)$

$$y^\star = A\hat{\beta} - C = A(X^TX)^{-1}X^Ty - C$$

$$y^\star \sim N(A\beta - C, \sigma^2 A(X^TX)^{-1}A^T).$$

$$\frac{y^\star}{\sigma} \sim N\left(\frac{A\beta - C}{\sigma}, A(X^TX)^{-1}A^T\right).$$

$$E\left[\frac{RSS_R - RSS}{\sigma^2}\right] = E\left[\left(\frac{y^\star}{\sigma}\right)^T [A(X^TX)^{-1}A^T]^{-1}\left(\frac{y^\star}{\sigma}\right)\right]$$

Now, $\left(\frac{y^\star}{\sigma}\right)^T [A(X^TX)^{-1}A^T]^{-1}\left(\frac{y^\star}{\sigma}\right) \sim \chi^2_{\text{rank}(A), \lambda}$

where $\lambda = \left(\frac{AB - C}{\sigma}\right)^T [A(X^TX)^{-1}A^T]\left(\frac{AB - C}{\sigma}\right)$

$$E\left[\frac{y^{\star T} B^\star y^\star}{\sigma^2}\right]$$

$= \text{trace}\{$

(iii) $\frac{RSS_R - RSS}{\sigma^2} \sim \chi^2_{q, \lambda=0}$ (under H_0).

and, $\frac{RSS}{\sigma^2} \sim \chi^2_{n-p, \lambda=0}$ (under H_0).

$$RSS = (Y - X\hat{\beta})^T (Y - X\hat{\beta}).$$

$$= Y^T H Y. \quad \text{where } H = (I - X(X^TX)^{-1}X^T).$$

$$\frac{RSS}{\sigma^2} = \left(\frac{Y}{\sigma}\right)^T H \left(\frac{Y}{\sigma}\right).$$

Now, $AV = H.I_n = H$ is idempotent $[Y \sim N(X\beta, \sigma^2 I_n)]$

$r(H) = n-p.$

$\therefore \frac{RSS}{\sigma^2} \overset{H_0}{\sim} \chi^2_{n-p, \lambda=0}.$

Now, $(RSS_R - RSS)$ & RSS are independent (show)

[This problem will reduce to proving independence of $\hat{\beta}$ & RSS].

$$\therefore \frac{(RSS_R - RSS)/\sigma^2/q.}{(RSS)/\sigma^2/(n-p)} \sim F_{q, n-p}.$$

[Proved]

(18/08/2023)

In particulare,

$$\left.\begin{array}{l} H_0: \beta_1 = \cdots = \beta_{p-1} = 0 \\ \text{vs } H_1: H_0 \text{ is not true.} \end{array}\right\}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = [\underset{\sim}{0} \;\; I_{p-1}]$$

...

~~Note that~~,

~~Test statistic =~~ $n-p$

Example: $H_0: \underset{\sim}{a}^T \underset{\sim}{\beta} = c \quad H_1: \underset{\sim}{a}^T \underset{\sim}{\beta} \neq c.$

Here test statistic will follow $F_{1,n-p}$ distribution.

Alternatively, $a^T \hat{\beta} \sim N(a^T\beta, \sigma^2 \underset{\sim}{a}^T (X^T X)^{-1} \underset{\sim}{a})$.

$$Z = \frac{\hat{a}^T \hat{\beta} - a^T \beta}{\sqrt{\sigma^2 \underset{\sim}{a}^T (X^T X)^{-1} \underset{\sim}{a}}} \sim N(0,1)$$

Under H_0,

$$Z = \frac{a^T \hat{\beta} - c}{\sqrt{\sigma^2 \underset{\sim}{a}^T (X^T X)^{-1} a}} \sim N(0,1).$$

Here σ is unknown.

$\hat{\sigma}^2 = MSE = \frac{RSS}{n-p}$

$$\frac{RSS}{\sigma^2} \sim \chi^2_{n-p}.$$

[**Result**: $X \sim \chi^2_n$ & $Y \sim N(0,1)$ and X, Y are independent]

So replace σ^2 by $\hat{\sigma}^2$ in Z, we get

$$Z = \frac{a^T\hat{\beta} - c.}{\sqrt{MSE.\ a^T(X^TX)^{-1}a}} \sim t_{n-p}$$

we can also use F distn. as –

$$t_n^2 \equiv F_{1,n}$$

So; $$\left(\frac{a^T\hat{\beta} - a^T\beta.}{\sqrt{MSE\ a^T(X^TX)^{-1}a}}\right)^2 \sim F_{1,n-p}$$

MULTPLE CORRELATION COEFFICIENT

let's start with sample correlation X & Y.

$$r^2 = \frac{\{\Sigma(y_i-\bar{y})(x_i-\bar{x})\}^2}{\Sigma(y_i-\bar{y})^2\ \Sigma(x_i-\bar{x})^2} = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

$$\begin{aligned}
RSS &= \Sigma(y_i - \hat{y}_i)^2 \\
&= \Sigma(y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i)^2. \text{ [for usual LR]} \\
&= \Sigma(y_i - \bar{y} - \hat{\beta}_1(x_i - \bar{x}))^2. \\
&= \Sigma(y_i-\bar{y})^2 + \hat{\beta}_1^2\Sigma(x_i-\bar{x})^2 - 2\hat{\beta}_1\Sigma(y_i-\bar{y})(x_i-\bar{x})
\end{aligned}$$

$$= \Sigma(y_i-\bar{y})^2 + \frac{S_{xy}^2}{S_{xx}^2}\cdot S_{xx} - 2\cdot\frac{S_{xy}}{S_{xx}}\cdot S_{xy}$$

$$= \Sigma(y_i-\bar{y})^2 - \cancel{2}\cdot\frac{S_{xy}^2}{S_{xx}}\cdot\frac{S_{yy}}{S_{yy}}$$

$$= \Sigma(y_i-\bar{y})^2 - r^2 S_{yy}$$

$$= S_{yy}(1-r^2)$$

So, $|r| \to 1 \Rightarrow RSS \to 0$.

$$\therefore r^2 = 1 - \frac{RSS}{\Sigma(y_i-\bar{y})^2}$$

$$= 1 - \frac{RSS}{TSS}$$

$$= \frac{SSR}{TSS}.$$

~~Statistics~~ Usual Notation : R^2

$$x_1 \longrightarrow y_{11}, y_{12}, \ldots, y_{1n_1}$$

$$x_2 \longrightarrow y_{21}, y_{22}, \ldots, y_{2n_2}$$

21/8/2023.

Lack of Fit Test

A fundamental requirement for this test is, We need ~~exact~~ multiple observations of y_i for each of the covariates.

Let, y_{ij} denotes the jth observation on the response at x_i, $i=1,2,\ldots,m$
$j=1,2,\ldots,n_i$, $n = \sum_{i=1}^{m} n_i$

$SS_{Res}(RSS) = SS_{PE} + SS_{LOF}$

PE : Pure Error.
LOF : Lack of fit.

$(y_{ij}-\hat{y}_i) = (y_{ij}-\bar{y}_i) + (\bar{y}_i-\hat{y}_i)$

$\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\hat{y}_i)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 + \sum_{i=1}^{m} n_i(\bar{y}_i-\hat{y}_i)^2.$$

$$\underset{\sim}{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ \vdots \\ y_{m1} \\ \vdots \\ y_{mn_m} \end{bmatrix}$$

$$SS_{PE} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2.$$

Now, show that, $\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 \sim \chi^2_{n_i-1}$.

And these 'm' terms are independent of each other.

Hence, $SSPE = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 \sim \chi^2_{n-m}$.

Now, $SS_{LOF} \sim \chi^2_{m-2}$ (show).

Then $F_0 = \frac{SS_{LOF}/m-2}{SS_{PE}/n-m} \sim F_{m-2,n-m}$

CONFIDENCE INTERVALS & REGIONS

$H_0: \beta_j = 0$ vs. $H_1: \beta_j \neq 0$

$$t = \frac{\hat{\beta}_j - \beta_j}{\sqrt{var(\hat{\beta}_j)}}$$

Now, $P(|t| > t_n^{\alpha}) = \alpha$

$P(|t| \leq t_n^{\alpha}) = 1-\alpha$.

Scanned with CamScanner

$$\hat{\beta}_j - t_{n-p}^{\alpha/2}\sqrt{\hat{var}(\hat{\beta}_j)} \le \beta_j \le \hat{\beta}_j + t_{n-p}^{\alpha/2}\sqrt{\hat{var}(\hat{\beta}_j)}$$

Now, $\hat{var}(\hat{\beta}_j) = \hat{\sigma}^2 \underbrace{(X^TX)^{-1}_{jj}}_{c_{jj}}$

23/08/2023.

Therefore,

$$E_j = \left[\hat{\beta}_j - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 \cdot c_{jj}},\ \hat{\beta}_j + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 \cdot c_{jj}}\right]$$

Suppose E_j $(j=1,2,\dots,k)$ is the event that the j-th statement is correct and let,

$$P(E_j) = 1-\alpha_j$$

$$P\left(\bigcap_{j=1}^{k} E_j\right) = 1 - P(\cup E_j^c) \geq 1 - \sum_{j=1}^{k} P(E_j^c)$$

$$= 1 - \sum_{j=1}^{k} \alpha_j = 1 - k\alpha.$$

$$[\text{If } \alpha_j = \alpha \ \forall j]$$

This should have been $(1-\alpha)$

This is the cost of taking one parameter at a time, calculating confidence interval and then taking their joint expression, instead of taking all the parameters simultaneously.

Here are some methods to calculate simultaneous interval.

→ BONFERRONI'S T - INTERVAL

Look at the previous issue. One way to overcome is taking the lower limit as $(1-\alpha/k)$ for each of the β_j's. This is known as Bonferroni's correction

→ MAXIMUM MODULUS INTERVALS :-

Define, $T_j = \dfrac{a_j^T \hat{\beta} - a_j \beta_{H_0}}{\sqrt{\hat{var}(a_j^T \hat{\beta})}}$.

$\sim N(0,1)$ [Under H_0]

Now, $\hat{var}(a_j^T \hat{\beta}) = \hat{\sigma}^2 \underset{\sim}{a_j^T}(X^TX)^{-1} \underset{\sim}{a_j}$

Set, $1-\alpha = P\left[\max_j |T_j| \le \underbrace{U^{\alpha}_{k,n-p}} \right]$

This critical values can be calculated using simulation.

[It is also mentioned in Saber and Lee book.]

Then the CI can be given by -

$\left[\hat{\beta}_j \pm U^{\alpha}_{k,n-p} \sqrt{var(\hat{\beta}_j)} \right]$.

→ SCHEFFE'S METHOD :-

WLG, assume that first 'd' vector of the set $\{\underset{\sim}{a_1}, \underset{\sim}{a_2}, \ldots, \underset{\sim}{a_n}\}$ are linear independent and remaining (if any) are linearly dependent on the first 'd' vectors then $d \le \min\{k, p\}$.

No. of parameters of interest

Actual no. of parameters in the model.

Consider a $d \times p$ matrix of rank 'd', say, 'A'.

Define, $\hat{\phi} = A\hat{\beta}$.

Then $\dfrac{1}{\sigma^2} \cdot (\hat{\phi} - \phi)^T (A(X^TX)^{-1}A^T)^{-1}(\hat{\phi} - \phi) \sim \chi^2_d$

(Show)

and. $\frac{(n-p)\, MS_{Res}}{\sigma^2} \sim \chi^2_{n-p}$ $\longrightarrow = \frac{RSS}{n-p}$

Therefore $\boxed{\frac{(\hat{\phi}-\phi)^T [A(X'X)^{-1}A']^{-1} (\hat{\phi}-\phi)}{d \cdot MS'_{Res}} \sim F_{d,n-p}}$

25/08/2023

Then, $(1-\alpha) = P[F \leq F^{\alpha}_{d,n-p}]$

$$= P[(\hat{\phi}-\phi)^T L^{-1} (\hat{\phi}-\phi) \leq m]$$

where, $L = A(X^TX)^{-1}A^T$

& $m = MS_{Res}\, F^{\alpha}_{d,n-p}$

$$= P[b^T L^{-1} b \leq m] \quad \text{where, } b = \hat{\phi}-\phi \qquad (*)$$

RESULT If L is pd, then for any b,

$$\underset{h\neq 0}{\text{Sup/max}} \frac{(h^Tb)^2}{h^TLh} = b^TL^{-1}b$$

Using this result in $(*)$, we get,

$$P\left[\sup_{h\neq 0} \frac{(h^Tb)^2}{h^TLh} \leq m\right]$$

$$= P\left[\frac{(h^Tb)^2}{h^TLh} \leq m \;\forall h\right]$$

$$= P\left[\frac{|h^T\hat{\phi} - h^T\phi|}{\sqrt{MS_{Res}(h^T L h)}} \le \left(d\, F^{\alpha}_{d, n-p}\right)^{1/2}\right]$$

$\forall h \neq 0$

$\therefore$ $\boxed{\text{CI for } h^T\phi \text{ is given by } \left(h^T\hat{\phi} \pm \left(d\, F^{\alpha}_{d, n-p}\right)^{1/2}\sqrt{MS_{Res}(h^T L h)}\right).}$

CONFIDENCE REGIONS

We know that,

$$F = \frac{(\hat{\beta} - \beta)^T (X^T X)(\hat{\beta} - \beta)}{p \cdot \underbrace{MS_{Res}}_{(=\hat{\sigma}^2)}} \sim F_{p, n-p}$$

We need a region $\mathbb{R}$ such that

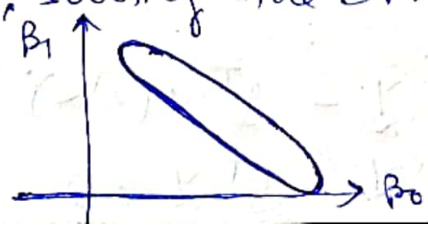
$$\mathbb{R} = \left\{\beta : F \le F^{\alpha}_{p, n-p}\right\}.$$

$\mathbb{R}$ is called simultaneous confidence region.

e.g.) $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$.

$$X^T X = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix}$$

Then, solving the eqn. we will find a region like -



26/08/2023

RESIDUAL ANALYSIS

Define, $e_i = y_i - \hat{y}_i$: Residuals.

Properties of Residuals

$\Rightarrow$ $\bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i = 0$

Proof: $\underset{\sim}{1}^T \underset{\sim}{e} = \underset{\sim}{1}^T(\underset{\sim}{y} - \hat{\underset{\sim}{y}})$

$$= \underset{\sim}{1}^T(y - X(X^TX)^{-1}X^Ty)$$

$$= \underset{\sim}{1}^T(I - X(X^TX)^{-1}X^T)\underset{\sim}{y}$$

$$(X^TX)^{-1}(X^TX) = I$$

$$X(X^TX)^{-1}(X^TX) = X$$

Now, write $X = [\underset{\sim}{1}\ X^*]$

$$\Rightarrow X(X^TX)^{-1}X^T[\underset{\sim}{1}\ X^*] = [\underset{\sim}{1}\ X^*]$$

$$\Rightarrow [X(X^TX)^{-1}X^T\underset{\sim}{1} \quad X(X^TX)^{-1}X^TX^*] = [\underset{\sim}{1}\ X^*]$$

$$\Rightarrow X(X^TX)^{-1}X^T\underset{\sim}{1} = \underset{\sim}{1}$$

$$\Rightarrow \underset{\sim}{1}^TX(X^TX)^{-1}X^T = \underset{\sim}{1}^T$$

Now,

$$\underset{\sim}{1}^T\underset{\sim}{e} = \underset{\sim}{1}^T\underset{\sim}{y} - \underset{\sim}{1}^TX(X^TX)^{-1}X^T\underset{\sim}{y}$$

$$= \underset{\sim}{1}^T\underset{\sim}{y} - \underset{\sim}{1}^T\underset{\sim}{y} = 0$$

• Sample Variance $MS_{Res} = \frac{\sum e_i^2}{n-p}$.

$$\approx \frac{\sum(e_i - \bar{e})^2}{n-p}$$

METHODS FOR SCALING RESIDUALS

• **Standardized Residuals** $d_i = \frac{e_i}{\sqrt{MS_{Res}}}$, $i = 1, 2, \ldots, n$.

If $d_i > 3$ ⇒ i-th observation is outlier.

• **Studentized Residuals** $\underset{\sim}{e} = \underset{\sim}{y} - \hat{Y} = [I - H]\underset{\sim}{y}$.

where $H = X(X^TX)^{-1}X^T$

$$\text{var}(\underset{\sim}{e}) = \sigma^2 [I - H]$$

$$\hat{\text{var}}(e_i) = \hat{\sigma}^2(1 - h_{ii}), \quad H = (h_{ij})$$

Define $r_i = \frac{e_i}{\sqrt{MS_{Res}(1 - h_{ii})}}$

• **PRESS RESIDUALS**

$e_{(i)} = y_i - \hat{y}_{(i)}$ [where $\hat{y}_{(i)}$ is the fitted value obtained from the regression line ~~while~~ without considering the i-th observation]

We have to show that, $\boxed{e_{(i)} = y_i - \hat{y}_{(i)} = \frac{e_i}{1-h_{ii}}}$

For that proof, we need a result.

Result> $(X^T X - \underset{\sim}{x}\underset{\sim}{x}^T)^{-1} = (X^T X)^{-1} + \frac{(X^T X)^{-1} \underset{\sim}{x} \underset{\sim}{x}^T (X^T X)^{-1}}{1 - \underset{\sim}{x}^T (X^T X)^{-1} \underset{\sim}{x}}$.

let's get back to ~~p~~ the proof of the original result.

Proof>. ~~Assumption~~, Let $X_{(i)}$ is the submatrix of X removing the ith row.

It is evident that, $X_{(i)}^T X_{(i)} = X^T X - x_{(i)} x_{(i)}^T$

$x_{(i)} = (1 \; x_{i1} \; x_{i2} \cdots x_{ip})$.

Now, ~~~~ from the above result mentioned

$$(X_{(i)}^T X_{(i)})^{-1} = (X^T X)^{-1} + \frac{(X^T X)^{-1} \underset{\sim}{x_i} \underset{\sim}{x_i}^T (X^T X)^{-1}}{1 - \underset{\sim}{x_i}^T (X^T X)^{-1} \underset{\sim}{x_i}}$$

Note that, $h_{ii} = 1 - \underset{\sim}{x_i}^T (X^T X)^{-1} \underset{\sim}{x_i}$.

$$\text{Now, } e_{(i)} = y_i - \hat{y}_{(i)}$$

$$= y_i - x_i^T (X_{(i)}^T (X_{(i)})^{-1} X_{(i)}^T y_{(i)}.$$

$$= y_i - x_i^T \left[(X^T X)^{-1} + \frac{(X^T X)^{-1} x_i x_i^T (X^T X)^{-1}}{1 - h_{ii}} \right] X_{(i)}^T y_i$$

$$= y_i - x_i^T (X^T X)^{-1} X_{(i)}^T y_{(i)} + \frac{x_i^T (X^T X)^{-1} x_i x_i^T (X^T X)^{-1} X_{(i)}^T y_{(i)}}{1-h_{ii}}$$

$$= \frac{\left[(1-h_{ii}) y_i - (1-h_{ii}) x_i^T (X^T X)^{-1} X_{(i)}^T y_{(i)} - h_{ii} x_i^T (X^T X)^{-1} X_{(i)}^T y_{(i)}\right]}{1-h_{ii}}$$

$$= \frac{(1-h_{ii}) y_i - x_i^T (X^T X)^{-1} X_{(i)}^T y_{(i)}}{1-h_{ii}}$$

$$\left[\text{Since, } X^T Y = X_{(i)}^T Y_{(i)} + x_i y_i\right]$$

$$= \frac{(1-h_{ii}) y_i - x_i^T (X^T X)^{-1} (X^T y - x_i y_i)}{1-h_{ii}}$$

$$= \frac{(1-h_{ii}) y_i - x_i^T (X^T X)^{-1} X^T y + x_i^T (X^T X)^{-1} x_i y_i}{1-h_{ii}}$$

$$= \cancel{y_i - \hat{y}_i} \quad \frac{y_i - \cancel{h_{ii} y_i} - \hat{y}_i + \cancel{h_{ii} y_i}}{1-h_{ii}}$$

$$= \frac{e_i}{1-h_{ii}} \quad [\text{Proved}]$$

$$\operatorname{var}(e_{(i)}) = \frac{\sigma^2 (1-h_{ii})}{(1-h_{ii})^2} = \sigma^2 / 1-h_{ii}$$

→ Standardized Press Residuals

$$d_{(i)} = \frac{e_i / (1-h_{ii})}{\sqrt{\hat{\sigma}^2/(1-h_{ii})}}$$

28/08/2023

▣ R-Student / Externally Studentized Residuals

$$S_{(i)}^2 = \frac{(n-p)MS_{Res} - e_i^2/(1-h_{ii})}{n-p-1}$$

In other words,

$$S_{(i)}^2 = \frac{Y_{(i)}^T (I - H_{(i)}) Y_{(i)}}{n-p-1}$$

$$\left[H_{(i)} = X_{(i)} (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T \right]$$

→ this is also an estimator of σ^2.

Then $\boxed{t_i = \dfrac{e_i}{\sqrt{S_{(i)}^2 (1-h_{ii})}}}$ → Most accurate form of Residuals. We should use this unless otherwise stated.

...

➤ RESIDUAL PLOTS

① Normal Probability plot: Let $t_{[1]} < t_{[2]} < \dots < t_{[n]}$ be the R-Student residuals in increasing order.

Now, if we plot $t_{[i]}'$s against the cumulative probability $\boxed{P_i = \frac{(i-1/2)}{n}}$, $i = 1(1)n$, on the normal probability plot the resulting points should lie approximately on a straight line.

~~[illegible]~~ $\longrightarrow$ $t_{[i]}$ against $E[t_{[i]}]$
$\approx \Phi^{-1}\left((i-1/2)/n\right)$.

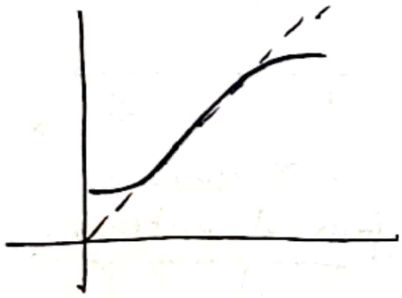
Ideally, when the errors are normally distributed, the graph should be -

► Some other cases :

(a).

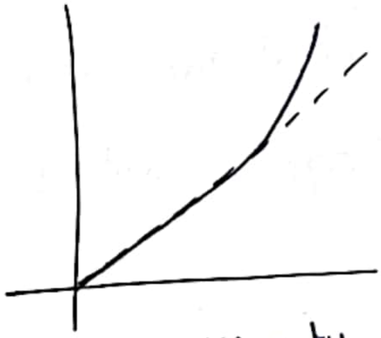
Sharp upward and downward curve at both extremes.
$\Rightarrow$ The tails of the distn. are too light to be considered normal.

(b)



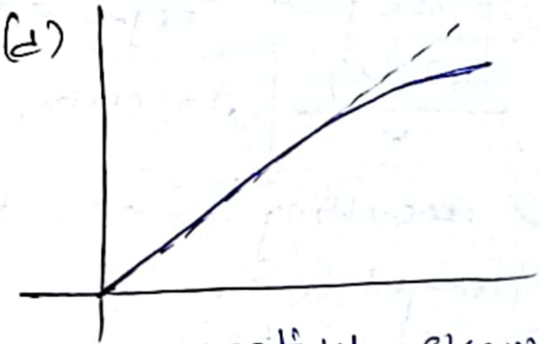
Sharp leftward and rightward curve at both extremes
$\Rightarrow$ The tails of the distn. are too heavy to be considered normal.

(c)



positively skewed

(d)

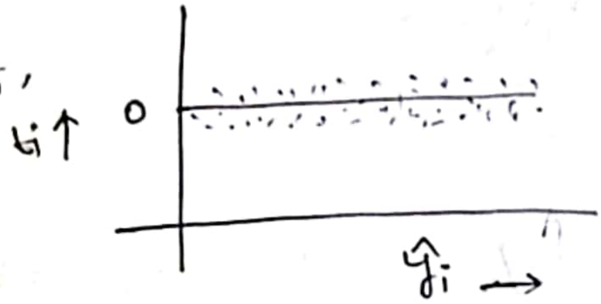


negatively skewed

- These plots with the validity of normality assumptions of error terms is the first step while analysing a data.

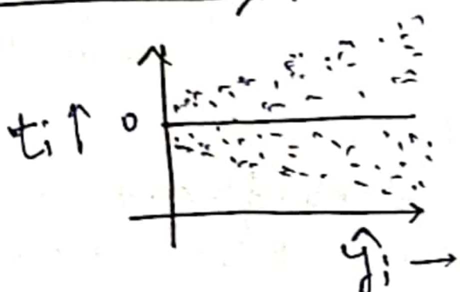
② PLOTS OF RESIDUALS AGAINST FITTED VALUES

Ideally,



Other cases

(a)

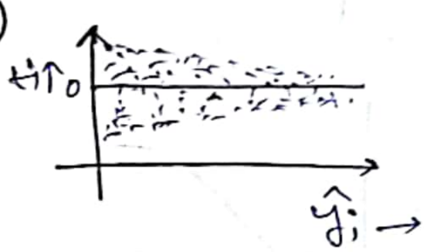


Outward funnel pattern

↓

variance is a increasing fn. of y

(b)

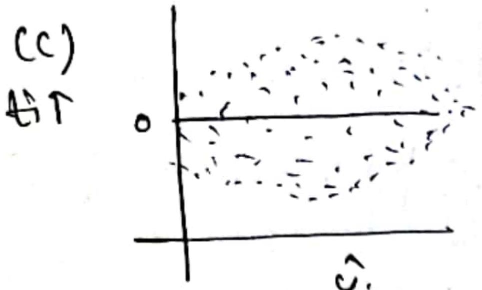


Inward ~~Outward~~ funnel patern.

↓

variance is a decreasing fn. of y.

(c)



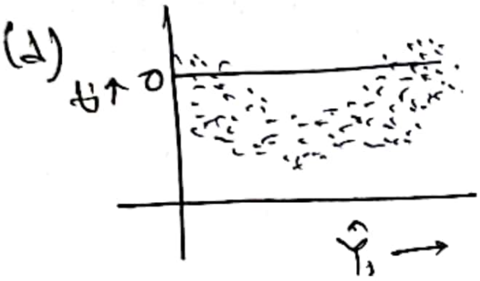
Double Bow pattern.

↓

often occurs when y is a proportion b/w 0 & 1.

e.g.) Binomial proportion has highest variance at $p = \frac{1}{2}$.

(d)



Curved pattern.

↓

indicates non-linearity.

we can think of adding higher degree terms of the covariates.

---

③ PARTIAL REGRESSION PLOTS

We can plot y with x to get an initial idea. But, in case of multivariate covariate, we can also plot y with marginal covariates. [for example, $(Y \text{ vs } X_1)$ or $(Y \text{ vs } X_2)$]

though, in presence of interaction b/w covariates, these plots may not be that much informative.

Also, we can regress y on x_j (a particular covariate) and obtain fitted values & residuals.

$$\hat{y}_i(x_j) = \hat{\theta}_0 + \hat{\theta}_1 x_{i1} + \dots + \hat{\theta}_{j-1} x_{ij-1} + \hat{\theta}_{j+1} x_{ij+1} + \dots + \hat{\theta}_p x_{ip}.$$

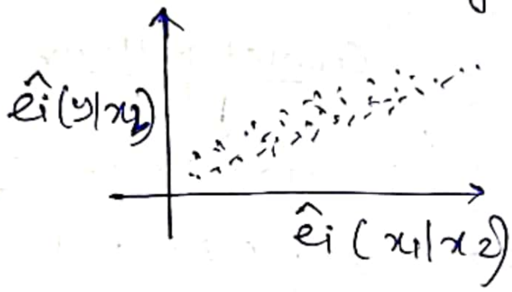
and $\underset{\sim}{\hat{e}_i(y|x_i) = y_i - \hat{y}_i(x_i^{\circ})}$.

Now, regress x_1 on x_2 [Take $p=2$ above & $j=1$] and calculate the residuals.

$$\hat{x}_{i1} = \hat{\alpha}_0 + \hat{\alpha}_1 x_{i2} \quad \text{for } i = 1(1)n.$$

$$\hat{e}_i(x_1|x_2) = x_{i1} - \hat{x}_{i1}$$

Then, if $\hat{e}_i(x_1|x_2)$ and $\hat{e}_i(y|x_2)$ are (sort of) linearly related then we say that $\underline{y \text{ \& } x_1 \text{ are linearly related}}$.



⊙ In general, say we are interested to infer the relationship of y and x_j. Now, let $e(Y|X_{(j)})$ is the residuals with j-th regressor removed and $e(x_j|X_{(j)})$ is the residuals when we regress x_j with the remaining covariates.

$\boxed{30/08/2023}$

The original model can be written as –

$$Y = X_{(j)}\beta_{(j)} + x_j\beta_j + \epsilon \quad \text{—} ⊛$$

~~Then, $\hat{\beta}_{(j)} = \ldots$~~

Define, $H_{(j)} = X_{(j)}(X_{(j)}^T X_{(j)})^{-1} X_{(j)}^T$

Since, $(I - H_{(j)})X_{(j)} = 0$, pre multiplying ⊛ with $(I - H_{(j)})$ we get,

$$(I - H_{(j)})Y = (I - H_{(j)})x_j\beta_j + (I - H_{(j)})\epsilon$$

$$\Rightarrow \left(Y - X_{(j)}\hat{\beta}_{(j)}\right) = \beta_j\left(x_j - x_{(j)}\underbrace{(x_{(j)}^T x_{(j)})^{-1} x_{(j)}^T x_j}\right) + \epsilon^*$$

$$\Rightarrow e(Y|X_{(j)}) = \beta_j e(x_j|X_{(j)}) + \epsilon^*.$$

$\hookrightarrow$ So there is actually a linear relationship b/w these terms.

$\boxed{\text{Caution}}$ These graphs might be misleading if there is presence of multicollinearity.

—·—

④ PARTIAL RESIDUAL PLOTS

In the previous page, we have seen that,

$$\hat{Y} = X_{(j)}\hat{\beta}_{(j)} + x_j\hat{\beta}_j$$

Now, $e(Y|X_{(j)}) = Y - X_{(j)}\hat{\beta}_{(j)}$

$$= Y - \hat{Y} + x_j\hat{\beta}_j$$

$$= e + x_j\hat{\beta}_j$$

$\therefore$ For ith observation,

$$e_i(Y|X_{(j)}) = x_{ij}\hat{\beta}_j + e_i$$

Plot these two terms.

—·—

PRESS STATISTICS

This is a measure to see how good is the model in terms of prediction.

$$PRESS = \sum_{i=1}^{n} (y_i - \hat{y}_{(i)})^2$$

$$= \sum_{i=1}^{n} \left(\frac{e_i}{1-h_{ii}}\right)^2 .$$

Smaller PRESS value is preferred.

➤ R^2 for prediction based on PRESS

$$R^2_{prediction} = 1 - \frac{PRESS}{TSS} .$$

TRANSFORMATION AND WEIGHTING TO CORRECT MODEL INADEQUACIES.

① VARIANCE STABILIZING TRANSFORMATION

	RELATIONSHIP OF σ^2 to $E(Y)$	TRANSFORMATION
	$\sigma^2 \propto$ const.	$y' = y$
[Poisson/ count type data]	$\sigma^2 \propto E(y)$	$y' = \sqrt{y}$ (Poisson).
[Binomial proportion]	$\sigma^2 \propto E(y)(1-E(y))$.	$y' = \sin^{-1}(\sqrt{y})$.

Scanned with CamScanner

01/09/2023

➤ BOX-COX TRANSFORMATION ON Y

A famous result of limit gives –

$$\lim_{\lambda \to 0} \frac{y^\lambda - 1}{\lambda} = \log y.$$

The transformed model ~~is~~ can be given by –

$y_i^\lambda = \underset{\sim}{x_i}'\underset{\sim}{\beta} + \epsilon_i$, But this may also ~~doesn't~~ not solve the purpose, because if λ becomes very small, it almost becomes constant.

Then another transformed model is –

$y^* = \frac{y^\lambda - 1}{\lambda}$, but this transformation is very sensitive. Small error in estimating 'λ' ~~error~~ will cause drastically large error in ~~esti~~ the actual estimation problem.

So, we will use the following improved transformation

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda \dot{y}^{n-1}}, & \text{if } \lambda \neq 0. \\ \dot{y} \ln y, & \text{if } \lambda = 0. \end{cases}$$

(where, $\dot{y} = (y_1 \cdots y_n)^{1/n}$)

So our transformed model is –

$$\underset{\sim}{y}^{(\lambda)} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}.$$

We will use MLE of λ and β for the estimation.

Note that, here, Log likelihood fn. $\propto$ RSS
$$\downarrow$$
$$[y^{(\lambda)} - X\beta]' [y^{(\lambda)} - X\beta]$$

...

➤ Box - TIDEWELL (1962).

TRANSFORMATION ON REGRESSORS

Let our original model is
$$E(y) = \beta_0 + \beta_1 x.$$

We transform the regressor using the following transformation.
$$\xi = \begin{cases} x^{\alpha}, & \alpha \neq 0 \\ \log x, & \alpha = 0 \end{cases}$$

β_0, β_1, α are unknown.

Transformed model: $E(y) = \beta_0 + \beta_1 \xi$.

We need to choose suitable α to stabilize variance.

Suppose 'α_0' is the initial guess of α. So $\xi = x^{\alpha_0}$, let us take $\alpha_0 = 1$

Then, expanding about the initial guess in a taylor series and ignoring terms higher than first order gives –

$$E(y) = \underset{\substack{\| \\ (\beta_0+\beta_1\xi_0)..}}{f(\xi_0, \beta_0, \beta_1)} + (\alpha-\alpha_0)\left\{\frac{df(\xi_0;\beta_0,\beta_1)}{d\alpha}\right\}\Bigg|_{\substack{\xi=\xi_0 \\ \alpha=\alpha_0}}$$

Now, $$\frac{df(\xi;\beta_0,\beta_1)}{d\alpha} = \frac{df(\xi,\beta_0,\beta_1)}{d\xi}\Bigg|_{\xi=\xi_0} \cdot \frac{d\xi}{d\alpha}\Bigg|_{\alpha=\alpha_0}$$

$$= \beta_1 x \log x.$$

Estimate of ~~Next~~ β_1 can be obtained from the model $\boxed{y = \beta_0 + \beta_1 x}$. —(1)

Then after taylor series expansion,

$$E(y) \approx \beta_0^{\star} + \beta_1^{\star} x + \hat{\beta}_1(\alpha-1)x\log x \quad [\hat{\beta}_1 \text{ from } (1)]$$

$$= \beta_0^{\star} + \beta_1^{\star} x + \gamma\omega \quad \left[\begin{array}{l}\gamma = \beta_1(\alpha-1) \\ \omega = x\log x\end{array}\right]$$

Then, estimate ~~the~~ using OLS to obtain $\hat{\gamma}$ —(2)

$$\alpha_1 = 1 + \frac{\hat{\gamma}}{\hat{\beta}_1}.$$

Then, repeat ~~the~~ from step (1)

– · –

06/09/2023

Generalised Least Squares

$$Y = X\beta + \epsilon \quad \text{— (1)} \quad \text{with } var(\epsilon) = \sigma^2 V$$

Since, $\sigma^2 V$ is a covariance matrix of the error. V must be non-singular and positive definite.

So, there exists an $n \times n$ non-singular symmetric matrix such that,

$$K^T K = K \cdot K = V = P^T D P.$$

$$K = P^T D^{1/2} P.$$

$$K^2 = KK = P^T D^{1/2} P P^T D^{1/2} P$$

$$= P^T D P = V.$$

Now, multiplying K^{-1} with (1),

$$K^{-1} y = K^{-1} X\beta + K^{-1}\epsilon$$

$$\cancel{\Rightarrow Z = K^{-1} y,}$$

$$\Rightarrow \boxed{Z = B\beta + g}$$

$$\left.\begin{aligned} \text{where, } Z &= K^{-1} y \\ B &= K^{-1} X \\ g &= K^{-1}\epsilon \end{aligned}\right\}$$

Now, we know,

$$\hat{\beta} = (B^T B)^{-1} B^T Z$$

$$= (X^T V^{-1} X)^{-1} X^T V^{-1} y.$$

$$var(\hat{\beta}) = \sigma^2 (B^T B)^{-1}$$

$$= \sigma^2 (X^T V^{-1} X)^{-1}$$

(show). $\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$ is BLUE.

-.-

Weighted Least Squares

Say, $Var(\epsilon_i) = \sigma^2 / w_i$

Then, $Var(y_i) = \sigma^2 / w_i$

Hence, $V = \sigma^2 \begin{bmatrix} 1/w_1 & & 0 \\ & \ddots & \\ 0 & & 1/w_n \end{bmatrix}$

Now, we want to stabilize the variance using the transformation -

$$y_i^* = y_i \sqrt{w_i}$$

Then $Var(y_i^*) = \sigma^2$.

Now, how to estimate this w_i?

Let, there are m groups in the data.

$\left.\begin{matrix} y_1 \\ y_2 \\ \vdots \end{matrix}\right\} 1 \quad \left.\begin{matrix} x_1 \\ x_2 \\ \vdots \end{matrix}\right\} 1$

$\vdots \qquad \vdots$

$y_n \} m \quad x_n \} m$.

Then for each group. ~~mean~~ sample means of x are given by - $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$

and sample variances of Y are given by - $S_1^2, S_2^2, \dots, S_m^2$.

- Plot these S_Y^2 against $\bar{x}_k$. ~~assuming~~

If the plot indicates a linear relationship b/w. S_Y^2 and $\bar{x}_k$, go for a linear model.

$$S_Y^2 = \beta_0 + \beta_1 \bar{x}_k$$

If there is some curvature like a quadratic polynomial in the plot, go for a polynomial fit of degree 2, given by -

$$S_y^2 = \beta_0 + \beta_1 \bar{x}_k + \beta_2 \bar{x}_k^2 .$$

From here, we get the weights by inverting the least square fitted values as –

$$\cancel{\frac{1}{\hat{w}_i} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_k + \hat{\beta}_2 \bar{x}_k^2}$$

$$\frac{1}{\hat{w}_i} = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2 .$$

▶ DIAGONSTIC FOR LEVERAGE & INFLUENCE :-

◉ LEVERAGE OBSERVATION :-

It is remote in x-space from the rest of the sample, but it lies almost on the regression line plotting through the rest of the sample points.

Now, recall the centered regression model –

$$y_i = \beta_0 + \beta_1 (x_{i1} - \bar{x}_1) + \dots + \beta_{p-1} (x_{i,p-1} - \bar{x}_{p-1}) + \epsilon_i$$

Define, $H^c = X^c (X^{cT} X^c)^{-1} X^{cT}$

$$= \frac{1}{n} \underset{\sim}{1} \underset{\sim}{1}^T + \tilde{X} (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T$$

$\tilde{X}$ is obtained from X^c by excluding $\underset{\sim}{1}$.

08/09/2023

So, $\hat{h}^c_{ii} = n^{-1} + \frac{1}{n-1} \cdot \underset{\sim}{x_i}^T \left(\frac{\tilde{X}^T\tilde{X}}{n-1}\right)^{-1} \underset{\sim}{x_i}$.

$\frac{\tilde{X}^T\tilde{X}}{n-1}$: sample covaraince matrix.

$\underset{\sim}{x_i}^T (\tilde{X}^T\tilde{X})^{-1} \underset{\sim}{x_i} = MD_i$ = Mahalanobis distance between i-th reduced row of X and the avg. reduced row ($\underset{\sim}{1}$ excluded).

Result: $H = H^c + \frac{1}{n} \underset{\sim}{1}_n \underset{\sim}{1}_n^T$

Now, $\text{trace}(H) = \sum h_{ii} = p$.

$\therefore \frac{1}{n}\sum h_{ii} = p/n$.

So, we will say i-obs is leverage if $\boxed{h_{ii} > \frac{2p}{n}}$

MEASURES OF INFLUENCE

▶ Cook's Distance

$$\boxed{D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})'(X'X)(\hat{\beta}_{(i)} - \hat{\beta})}{p \cdot MS_{Res}}} \sim F_{p,\,n-p} \text{ (approximately)}$$

Now, then i-th observation is an influential point if

$$\boxed{D_i > F_{0.05;\,p,\,n-p}}$$

Scanned with CamScanner

▶ DF BETAS :

$$DFBETAS_{i,j} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S_{(i)}^2 C_{jj}}} \cdot \quad j = 0, \ldots, p-1 . \quad i = 1, \ldots, n .$$

C_{jj} : (j,j)th element of $(X^TX)^{-1}$.

Criteria: If $|DFBETAS_{i,j}| > \frac{2}{\sqrt{n}}$ for any $j = 0(1)p-1$

Then, i-th obs. will be an influential observation.

BELSLEY, KUH, WELSH [1980]

▶ DFFITS :

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{h_{ii} S_{(i)}^2}}$$

Criteria: $|(DFFITS)|_i > 2\sqrt{p/n}$.

i-th obs. is outlier.

▶ COV. RATIO :

Generalized variance is given by

$$|var(\hat{\beta})| = |\sigma^2 (X^TX)^{-1}|$$

$$COVRATIO_i = \frac{|(X_{(i)}^T X_{(i)})^{-1} S_{(i)}^2|}{|(X^TX)^{-1} MS_{Res}|}$$

~~If $COVRATIO_i > 1$, then this i-th obs. is improving the precision of estimation.~~

$\boxed{\text{Criteria}}$: If $COVRATIO > 1 + \frac{3p}{n}$ or $COVRATIO < 1 - \frac{3p}{n}$.

$\Rightarrow$ i-th observation is influential obs.

11/09/2023

▷ MULTICOLLINEARITY :-

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \epsilon$$

$$X = [\underset{\sim}{x_1}, \dots, \underset{\sim}{x_k}].$$

Then, multicollinearity happens if x_i's are linearly dependent i.e.

$$\left\{ \begin{array}{l} \exists\, (t_1, \dots, t_k) \neq 0 \text{ s.t.} \\ \sum t_i \underset{\sim}{x_j} = 0. \end{array} \right.$$

Then $(X'X)^{-1}$ doesn't exist.

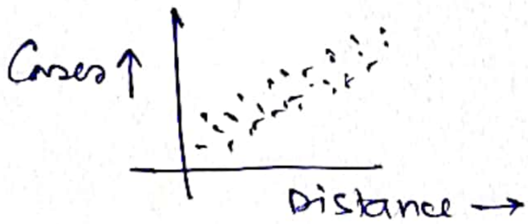
Even approximate linear dependence can make regression model dramatically inadequate.

$\boxed{\text{SOURCES}}$

→ Data Collection method :-

• When the analyst samples only a subspace of the region of regressors defined.

e.g. Consider the time delivery data. Cases & distance are the regressors.

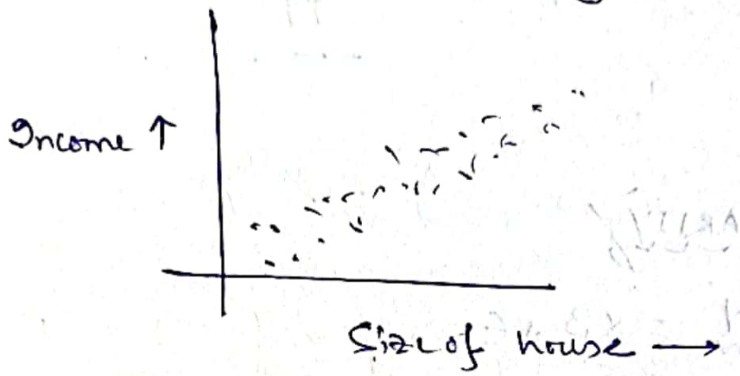


Here, the covariates are dependent.

Fact → observations with a small no. of cases also have a short distance.

→ Constraints on the Model / Population

- Consider the Electricity utility data.



This multicollinearity is hard to remove, because in reality it is almost impossible to find observation with low income but large size of house.

→ Choice of Model

- If the range of $\underset{\sim}{x}$ is small, adding a higher order term (say x^2) may result in multicollinearity.

- In this case, some subset of the regressors is usually preferred.

→ An Overdefined model

Here $n < p$

Scanned with CamScanner

$\boxed{\text{EFFECTS}}$

Result → Say $X = [\underset{\sim}{x_1} \mid \underset{\sim}{x_2}]$

Then, $X^T X = \begin{bmatrix} X_1^T \\ X_2^T \end{bmatrix} [X_1 \; X_2]$.

$$= \begin{bmatrix} X_1^T X_1 & X_1^T X_2 \\ X_2^T X_1 & X_2^T X_2 \end{bmatrix}$$

$$(X^T X)^{-1}_{11} = (X_1^T X_1)^{-1} + (X_1^T X_1)^{-1} X_1 X_2 G X_2^T X_1 (X_1^T X_1)^{-1}$$

$$(X^T X)^{-1}_{12} = -(X_1^T X_1)^{-1} X_1^T X_2 G.$$

$$(X^T X)^{-1}_{21} = -G\, X_2^T X_1 (X_1^T X_1)^{-1}$$

$$(X^T X)^{-1}_{22} = G = [X_2^T [I - H_1] X_2]^{-1}.$$

where $H_1 = X_1 (X_1^T X_1)^{-1} X_1$.

▶ $\underset{\sim\sim\sim}{\text{var}(\hat{\beta}_j) = \sigma^2 c_{jj} = \sigma^2 / 1 - R_j^2}$.

where R_j^2 is R^2 for the model

$$\boxed{\underset{\sim}{x_j} = X_{-j}\beta + \epsilon.}$$

↓

unit scaled. Such that $\sum x_{ji} = 0$

& $\underset{\sim}{x_j}^T \underset{\sim}{x_j} = 1$.

Then, $R_j^2 = \dfrac{SSR}{SST}$ where $SST = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$

$$= \underset{\sim}{x_j}^T \left[I - \tfrac{1}{n} 1 1^T\right] \underset{\sim}{x_j}$$

$$= 1.$$

$$SS_R = x_j^T \left[H - J/n \right] x_j \qquad J = \underline{1}\,\underline{1}^T$$

$$H = X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T$$

So, $$SS_R = x_j^T H x_j - \underbrace{\frac{1}{n} x_j^T \underline{1}\,\underline{1}^T x_j}_{0}$$

$$= x_j^T X_{-j} (X_{-j}^T X_{-j})^{-1} X_{-j}^T .$$

Let, WLG, $X = [X_{-j} \,\vdots\, x_j]$

Then $$X^T X = \begin{bmatrix} X_{-j}^T X_{-j} & X_{-j}^T x_j \\ x_j^T X_{-j} & x_j^T x_j \end{bmatrix}$$

Now, $$(X^T X)^{-1}_{22} = \left[x_j^T (I - H_1) x_j \right]^{-1}$$

$$= (1 - SSR)^{-1}$$

$$= (1 - R_j^2)^{-1}$$

[Since $SS_T = 1$]

$$= \frac{1}{1 - R_j^2}$$

$$\mathrm{Var}(\hat{\beta}_j) = \sigma^2 . (X^T X)^{-1}_{22} .$$

$$= \sigma^2 . \frac{1}{1 - R_j^2} .$$

13/09/2023

[UNIT SCALING] We transform x_i to $\frac{x_i - \bar{x}}{\sqrt{S_x^2}} = z_i$.

Then $\sum z_i = 0$ & $z^T z = 1$]

Remark]

$$var(\hat{\beta}_j) = \sigma^2 / 1 - R_j^2 .$$

Then, if $R_j \to 1$, $var(\hat{\beta}_j) \to \infty$. Then the C.I. for $\hat{\beta}_j$ will be very large and inference will be bad.

- $\hat{\beta}_j$ are too large in "absolute value". Consider the squared distance of $\hat{\beta}$ to β.

$$L_1^2 = (\hat{\beta} - \beta)^T (\hat{\beta} - \beta)$$

Then, $E(L_1^2) = E\left[\sum (\hat{\beta}_j - \beta_j)^2\right]$.

$$= \sum_{j=0}^{p-1} E(\hat{\beta}_j - \beta_j)^2 .$$

$$= \sum_{j=0}^{p-1} var(\hat{\beta}_j) = \sigma^2 \operatorname{trace} [X^T X]^{-1}$$

$$= \sigma^2 \sum_{j=0}^{p-1} \frac{1}{\lambda_j}$$

where, λ_j's are the eigen value of $X^T X$.

Now, if the dependence b/w covariates increases, at least one of the λ_j will tend to zero. And, as a consequence,

$$E(L_1^2) \to \infty .$$

...

DETECTION

- **EXAMINATION OF CORRELATION MATRIX**

Do the unit scaling and compute $X^T X$. This will be ~~the~~ a sample correlation matrix. Checking the ~~ab~~ elements in the matrix we can suspect presence of multicollinearity.

- **VARIANCE INFLATION FACTOR (VIF)**

$$VIF_j = C_{jj} = (1 - R_j^2)^{-1}.$$

For any j if $VIF_j > 5 \Rightarrow \underset{\sim}{x_j}$ is causing multicollinearity

- **EIGEN VALUES ANALYSIS**

Let $\lambda_1, \ldots, \lambda_p$ are eigen values of $X^T X$.

The "Condition indices" of $X^T X$ are -

$$K_j = \frac{\lambda_{max}}{\lambda_j}, \; j = 1, \ldots, p.$$

CONDITION NUMBER $\rightarrow$ $K = \dfrac{\lambda_{max}}{\lambda_{min}}$

If $CN \in (100, 1000) \Rightarrow$ moderate multicollinearity

$CN > 1000 \Rightarrow$ severe multicollinearity.

$\downarrow$

In this case we will use VIF to find the particular covariate(s) causing multicollinearity.

⊙ Eigen value analysis can be used also to identify the nature of near - linear dependendence in data.

$$X^T X = T \Lambda T$$

$T \downarrow$ eigenvectors. $\Lambda \longrightarrow \text{diag}(\lambda_1, \ldots, \lambda_p)$.

Let, $T = [\underset{\sim}{t_1}, \ldots, \underset{\sim}{t_p}]$

15/09/2023.

⊙ BELSLEY, KUH & WELSLEY (1980)

Here, instead of $X'X$, we only work with X.

Diagonalization of X is given by -.

$$X = UDT$$

where. U is $n \times p$, $T = p \times p$, $U^T U = I$, $T^T T = I$ and D is $p \times p$ diagonal matrix with non-negative diagonal elements μ_j; $j = 1, 2, \ldots, p$. Then μ_j is "singular value"s of X and, $X = UDT'$ is called "Singular value decomposition (SVD)".

Now, $X'X = TDU^T UDT^T$
$= TD^2 T^T$.

and, from Eigen decomposition,

$$X^T X = T \Lambda T^T$$

Thus, square of the singular values of X are the eigen values of $X^T X$.

Here T is the matrix of eigen vectors of $X^T X$ and U is the matrix whose columns are the eigen vectors associated with p-non zero eigen values of XX'.

⦿ Alternative representation of condition indices & condition number [using singular values] :

$\boxed{\text{Condition indices}} \longrightarrow \eta_j = \frac{\mu_{max}}{\mu_j}$

$j = 1, \ldots, p.$

and, Condition number is the largest value of condition indices.

Now,

$var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

$= \sigma^2 (T\Lambda^{-1}T)'.$

Then $var(\hat{\beta}_j) = \sigma^2 (T\Lambda^{-1}T')_{jj}$

$$T\Lambda^{-1}T' = \begin{bmatrix} t_{11} & t_{12} & \ldots & t_{1p} \\ t_{21} & t_{22} & \ldots & t_{2p} \\ & \vdots & & \\ t_{p1} & t_{p2} & \ldots & t_{pp} \end{bmatrix} \begin{bmatrix} 1/\mu_1^2 & & \\ & \ddots & \\ & & \frac{1}{\mu_p^2} \end{bmatrix} \begin{bmatrix} \cdot \\ \vdots \\ \cdot \end{bmatrix}^T$$

Then $(T\Lambda T')_{jj} = \sum_{i=1}^{p} \frac{t_{ji}^2}{\mu_i^2}$, $j = 1, 2, \ldots, p.$

$\therefore var(\hat{\beta}_j) = \sigma^2 \sum_{i=1}^{p} t_{ji}^2/\mu_i^2 .$

Therefore $\boxed{VIF_j = \sum_{i=1}^{p} t_{ji}/\mu_i^2}$

↓

Alternative representation of VIF in terms of singular values & eigenvectors of X^TX.

Scanned with CamScanner

Define, Variance decomposition proportion as -

$$\boxed{\pi_{ij} = \frac{t_{ji}^2 / \mu_i^2}{VIF_j}} \quad \left.\begin{array}{l} , j = 1, \ldots, p \\ i = 1, \ldots, p \end{array}\right\}.$$

This elements will construct a matrix Π.

Now, if any one element in a column becomes > 0.5 then, the corresponding covariate is to be suspected for causing multicollinearity.

25/09/2023

METHODS FOR DEALING WITH MULTICOLLINEARITY:

(i) VIF: If $(VIF)_j > 5$ then j-th co-variate is causing multicollinearity and we remove that.

(ii) Collecting Additional data.

(iii) Redefine your model:

Introduce some other covariates that are actually function of the original covariates.

e.g. $f(x_1, \ldots, x_p) = x_1 + x_2 + x_3$.

(iv) RIDGE REGRESSION:

Ridge Regression proposed by Hoerl and Kennard (1970). Ridge regression is the solution to the following equations:

$$(X^T X + kI)\hat{\beta} = X^T y.$$

$$\Rightarrow \hat{\beta}_R = (X^T X + kI)^{-1} X^T \underset{\sim}{y}$$

Now, MSE of $\hat{\beta}_R$ is given as -

$$MSE(\hat{\beta}_R) = E\left[(\hat{\beta}_R - \beta)^T(\hat{\beta}_R - \beta)\right].$$

$$= E\left[\{\hat{\beta}_R - E(\hat{\beta}_R) + E(\hat{\beta}_R) - \beta\}^T \{\hat{\beta}_R - E(\hat{\beta}_R) + E(\hat{\beta}_R) - \beta\}\right].$$

$$= \cdots$$

$$= E\left[\{\hat{\beta}_R - E(\hat{\beta}_R)\}^T\{\hat{\beta}_R - E(\hat{\beta}_R)\} + \{E(\hat{\beta}_R - \beta\}^T\{E(\hat{\beta}_R) - \beta\}\right]$$

$$= \sum_{j=0}^{p-1} E\{\hat{\beta}_j - E(\hat{\beta}_{Rj})\}^2 + \{Bias(\hat{\beta}_R)\}^2$$

$$= \sum_{j=0}^{p-1} var(\hat{\beta}_{Rj}) + \{Bias(\hat{\beta}_R)\}^2.$$

$$\sum_{j=0}^{p-1} var(\hat{\beta}_{Rj}) = trace(var(\hat{\beta}_R)).$$

$$= \sigma^2 trace\left[(X^TX + kI)^{-1}X^TX(X^TX + kI)^{-1}\right] \quad \leftarrow (1)$$

$$E(\hat{\beta}_R) - \beta = (X^TX + kI)^{-1}X^TX\beta - \beta.$$

$$= ((X^TX + kI)^{-1}X^TX - I)\beta.$$

$$= ((X^TX + kI)^{-1}(X^TX + kI - kI) - I)\beta.$$

$$= [\cancel{I} - k(X^TX + kI)^{-1} - \cancel{I}]\beta$$

$$= -k(X'X + kI)^{-1}\beta.$$

Therefore, $MSE(\hat{\beta_R}) = \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j+k)^2} + k^2 \beta^T (X^TX+kI)^{-2} \beta.$

[where $\{\lambda_1, \dots, \lambda_p\}$ are the eigen values of X^TX]

Note that if $k>0$, the bias in $\hat{\beta}_R$ increases with k. However, the variance decreases as k increases.

In using Ridge regression we would like to choose a value of k such that the reduction in the variance term is greater than the increase in the bias (squared).

Hoerl and Kennard have suggested that an appropriate value of k may be determined by inspection of "trace plot" or "Ridge trace".

PRINCIPAL COMPONENT REGRESSION

$y = X\beta + \epsilon$ can be rewritten as -

$y = Z\alpha + \epsilon$

where $Z = XT$, $\alpha = T'\beta$, $T'X'XT = Z'Z = \Lambda$.

$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ [λ_i's are the eigen values of X^TX]

T is the matrix of eigen vectors.

we get, ~~$var(\hat{\alpha}) = \sigma^2 (Z^TZ)^{-1}$~~ ~~$= \sigma^2 (\Lambda)^{-1}$~~

$$\hat{\alpha} = (Z^TZ)^{-1} Z^T Y = \Lambda^{-1} Z^T Y.$$

and $var(\hat{\alpha}) = \sigma^2 (Z^TZ)^{-1} = \sigma^2 (\Lambda^{-1})$

27/09/2023.

MODEL SELECTION :-

⊙ CONSEQUENCE OF MODEL SPECIFICATION :-

Assume that there are k regressors, $\underset{\sim}{x_1}, \ldots, \underset{\sim}{x_k}$. The full model is –

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon} \quad \text{—— (1)}.$$

Now, let, 'r' be the no. of regressors that are deleted from (1). Then the reduced/subset model will contain $(k-r)$ of the original regressors.

Let, $p-1 = k-r$.

Subset/Reduced model: $\underset{\sim}{Y} = X_p \underset{\sim}{\beta_p} + \underset{\sim}{\epsilon}$ – (2).

Let $\hat{\beta}^*$ is the LSE of the full model. Therefore,

$$\hat{\beta}^* = (X^TX)^{-1}X^T\underset{\sim}{Y}.$$

Let, $\hat{\sigma}^{*2}$ is an unbiased estimator of σ^2.

$$\hat{\sigma}^{*2} = \frac{Y^T[I-H]Y}{n-k-1} \quad ; H = X(X^TX)^{-1}X^T.$$

LSE of the subset model :-

$$\hat{\beta}_p = (X_p^T X_p)^{-1} X_p^T \underset{\sim}{Y}.$$

$\hat{\sigma}^2$ is an estimate of σ^2.

where, ~~Here~~, $\hat{\sigma}^2 = \frac{Y^T[I-H_p]Y}{n-p}$. $\quad ; H_p = X_p(X_p^TX_p)^{-1}X_p^T$

(a) $E(\hat{\beta}_p) = (X_p^T X_p)^{-1} X_p^T (X\beta)$. $\quad \left[\begin{array}{l} X = [X_p \,\vdots\, X_r] \\ \beta = \begin{bmatrix} \beta_p \\ \beta_r \end{bmatrix} \end{array} \right]$

$$= (X_p^T X_p)^{-1} X_p^T (X_p\beta_p + X_r\beta_r)$$

$$= \beta_p + (X_p^TX_p)^{-1}X_p^T X_r \beta_r = \beta_p + A\beta_r.$$

$A = (X_p^T X_p)^{-1} X_p^T X_r$ is called '<u>alias matrix</u>'.

$\hat{\beta}_p$ will be unbiased if $\underline{X_p^T X_r = 0 \text{ or } \beta_r = 0}$.

(b) $var(\hat{\beta}_p) = \sigma^2 (X_p^T X_p)^{-1}$.

Also, $var(\hat{\beta}^*) = \sigma^2 (X^T X)^{-1}$.

Again, partition the matrix as —

$$X = [X_p \vdots X_r].$$

$$X^T X = \begin{bmatrix} X_p^T \\ X_r^T \end{bmatrix} [X_p \; X_r] = \begin{bmatrix} X_p^T X_p & X_p^T X_r \\ X_r^T X_p & X_r^T X_r \end{bmatrix}$$

[A result] If A and $D - CA^{-1}B$ are non singular,

then $\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BM^{-1}CA^{-1} & -A^{-1}BM^{-1} \\ -M^{-1}CA^{-1} & M^{-1} \end{bmatrix}$

where $M \cancel{\text{is}} = D - CA^{-1}B$.]

~~Here, $M = X_r^T X_r - X_r^T X_p X_p^T (X_p^T)^{-1} X_p X_r = 0$~~

Here, $M = X_r^T X_r - X_r^T X_p (X_p^T X_p)^{-1} X_p^T X_r$.

$= X_r^T [I - H_p] X_r$.

$\longrightarrow$ [This can be shown as a p.d. matrix]

$var(\hat{\beta}_p^*) = \sigma^2 (X_p^T X_p)^{-1} + [(X_p^T X_p)^{-1} X_p^T X_r M^{-1}]$
$[X_r^T X_p (X_p^T X_p)^{-1}]$.

[Show, the 2nd term is p.d.]

Therefore, $var(\hat{\beta}_p^*) - var(\hat{\beta}_p)$ is p.s.d.

The variance of the least square estimates of the parameters in the full model are greater or equal to the variance of the corresponding parameters in the subset model.

ⓒ. We know, $\hat{\sigma}^2 = \dfrac{y^T [I - H_p] y}{n-p}$.

$$E(\hat{\sigma}^2)(n-p) = E\left[y^T [I - H_p] y \right]$$

$$= E(y^T)[I - H_p] E(y) + \text{trace}\left[[I - H_p] \sigma^2 I \right]$$

$$= \sigma^2 (n-p) + [X_p \beta_p + X_r \beta_r]^{-1} [I - H_p] [X_p \beta_p + X_r \beta_r].$$

$$= \sigma^2 (n-p) + \beta_r^T X_r^T [I - H_p] X_r \beta_r.$$

$$E(\hat{\sigma}^2) = \sigma^2 + \underbrace{\frac{\beta_r^T X_r^T [I - H_p] X_r \beta_r}{n-p}}_{\text{upward biased}}.$$

[and this can be written as $u^T A u$ $(u = X_r \beta_r,\ A = I - H_p)$.

$= u^T A . A u$ $[A \text{ is } A^2 = A]$

$= u^T A^T A u$ $[A^T = A]$

$= (Au)^T (Au)$.

$= v^T v = \sum v_i^2 \geq 0$.

29/09/2023.

Subset Selection

→ Selecting the set of regressors based on some criteria.

① Based on coefficient of Multiple determination (R^2)

Suppose the full model consist of $\overline{k-1}$ regressors and the reduced model consists of p-regressors.

For example, $k-1 = 4, p = 2$.
Then we select all possible choices –

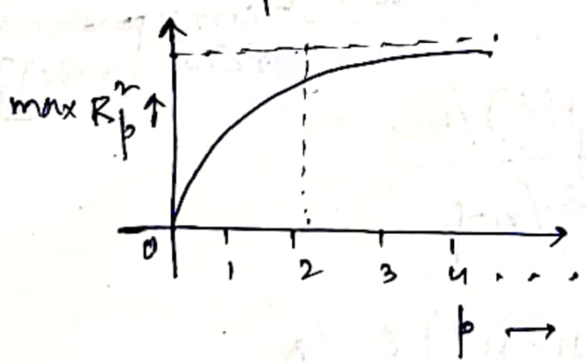
$$\{x_1, x_2\}, \{x_1, x_3\}, \ldots, \{x_3, x_4\}$$

and the pair which gives us the ~~largest value~~ largest value of R^2, gets selected.

Also, for inter-p choice, we know R^2 is an increasing function of p. So, ultimately we end up including all the regressors.

So, a stopping criteria would be choosing p, after which there is not significant amount of increase in the value of R^2.

So, if ($\max R_p^2$) gives me the maximum of of R^2 with different combination of regressors and with same p, the graph should look like.



and we choose after which we don't have any significant improvement.

(ii) BASED ON ADJUSTED R^2 :-

Theorem

Let $H_0: \underset{\sim}{\beta_r} = \underset{\sim}{0}$ vs $H_1: \underset{\sim}{\beta_r} \neq \underset{\sim}{0}$.

Then, $F = \dfrac{(RSS_r - RSS)/r}{RSS/n-p}$.

where RSS_r is the RSS of reduced model.

Then (i) R^2_{adj} (full model) $\propto f(F)$.

(ii) R^2_{adj} (reduced model) $\propto f(F)$.

(iii) $R^2_{adj}(f.m.) \geq R^2_{adj}(r.m.)$ iff $F \geq 1$.

Proof :-

$RSS = Y^T(I-H)Y$, where, $H = X(X'X)^{-1}X'$.

$= (Y - X\beta)^T(I-H)(Y-X\beta)$.

$= \underset{\sim}{\epsilon}^T(I-H)\underset{\sim}{\epsilon}$. where $\underset{\sim}{\epsilon} \sim N(0, \sigma^2 I)$.

$RSS_r = Y^T(I-H^*)Y$

$[H^* = X^*(X^{*T}X^*)^{-1}X^*]$

$= \underset{\sim}{\epsilon}^T(I-H^*)\underset{\sim}{\epsilon}$ [* denotes corresponding to reduced model]

Then, $F = \dfrac{(RSS_r - RSS)/r}{RSS/n-p}$

$= \dfrac{\underset{\sim}{\epsilon}^T(H-H^*)\underset{\sim}{\epsilon}/r}{RSS/n-p}$.

$= \dfrac{\underset{\sim}{\epsilon}^T(H-H^*)\underset{\sim}{\epsilon}/r}{\underset{\sim}{\epsilon}^T(I-H)\underset{\sim}{\epsilon}/n-p}$.

Now, to validate the independencies of numerator & denominator we need to show,

$$(I-H)(H-H^*) = 0.$$

$$\Rightarrow \cancel{HH} \; H - H^* - H^2 + HH^* = 0$$

$$\Rightarrow \underline{HH^* = H^*}. \quad -(i).$$

Also to show, $(H-H^*)$ to be idempotent,

$$(H-H^*)(H-H^*) = (H-H^*).$$

$$\Rightarrow H^2 - HH^* - H^*H + H^{*2} = H - H^*$$

$$\Rightarrow \underline{H - HH^* - H^*H + H^* = H - H^*} \; -(ii).$$

Careful inspection would reveal that to show (i) & (ii), we only need to show,

$$\left.\begin{array}{l} \text{(a) } HH^* = H^* \\ \text{(b) } H^*H = H^* \end{array}\right\}.$$

$$H = X(X^TX)^{-1}X'$$

Let's partition the X matrix into —

$$X = [X^* Z].$$

Then $X^TX = \begin{bmatrix} X^{*T} \\ Z^T \end{bmatrix} [X^* Z].$

$$= \begin{pmatrix} X^{*T}X & X^{*T}Z \\ Z^TX^* & Z^TZ \end{pmatrix}.$$

Hence, $H = (X^* Z) \begin{pmatrix} X^{*T}X & X^{*T}Z \\ Z^TX^* & Z^TZ \end{pmatrix}^{-1} \begin{pmatrix} X^{*T} \\ Z^T \end{pmatrix}.$

$$= [X^* \vdots Z] \begin{bmatrix} (X^{*T}X^*)^{-1} + (X^{*T}X^*)^{-1} X^{*T} Z M^{-1} Z^T X^* (X^{*T}X^*)^{-1} & -(X^{*T}X^*)^{-1} X^{*T} Z M^{-1} \\ -M^{-1} Z^T X^* (X^{*T}X^*)^{-1} & M^{-1} \end{bmatrix} \begin{bmatrix} X^{*T} \\ Z^T \end{bmatrix}.$$

$$[M = Z^T Z - (Z^T X^*)(X^{*T}X)^{-1}(X^{*T}Z).$$

$$= H^* + H^* Z M^{-1} Z^T - H^* Z M^{-1} Z^T - Z M^{-1} Z^T H^* + Z M^{-1} Z^T.$$

Then, $HH^* = H^*H^* - ZM^{-1}Z^T H^{*2} + ZM^{-1}Z^T H^*$

$= H^*$ (Proved.) $[H^{*2} = H^*]$.

Similarly, $H^*H = H^*$.

04/10/2023

We know that,

$$F = \frac{(RSS_r - RSS)/r}{RSS/(n-p)} \sim F_{r, n-p}.$$

$$= \frac{\left(\frac{RSS_r}{SST} - \frac{RSS}{SST}\right)/r}{\left(\frac{RSS}{SST}\right)/n-p}$$

$$= \frac{(1-R_r^2)-(1-R^2)}{(1-R^2)} \cdot \frac{n-p}{r} = \frac{R^2 - R_r^2}{1-R^2} \cdot \frac{n-p}{r} \quad \text{—(1)}$$

Now, $R^2_{adj} = 1 - \frac{RSS}{SST} \cdot \frac{n-1}{n-p}$

$$= 1 - (1-R^2) \cdot \frac{n-1}{n-p} \quad \text{—(2)}.$$

From (1) & (2), we can show that –

$$F \propto f(R^2_{adj}, R^2_{adj,r}, n-p, r)$$

Also, iff $F > 1$, $R^2_{adj} \geq R^2_{adj,r}$.

RESIDUAL MEAN SQUARE

$$MS_{Res}(p) = \frac{SS_{Res}(p)}{n-p}.$$

We can develop similar criterion for selecting subsets, involving MS_{Res} instead of R^2_{adj}, as –

$$R^2_{adj,p} = 1-(1-R^2_p)\frac{n-1}{n-p}.$$

$$= 1 - \frac{SS_{Res}(p)}{SS_T} \cdot \frac{n-1}{n-p}$$

$$= 1 - MS_{Res}(p) \cdot \frac{n-1}{SS_T}$$

So, the requirement of maximizing adj R^2 is equivalent to minimizing $MS_{Res}(p)$.

MELLOW C_p STATISTIC

Mellow has proposed a criterion which is related to mean square error of a ~~filtered~~ fitted value, given by.

$$E[\hat{y}_i - E(y_i)]^2 = \{E(y_i) - E(\hat{y}_i)\}^2 + var(\hat{y}_i)$$

Here, (i) $E(y_i)$ is the expected response from the ~~regressor~~ true model.

(ii) $E(\hat{y}_i)$ is the expected value of $\hat{y}_i$ from p-term model.

[$\longrightarrow$ $(p-1)$ regressors + 1 due to intercept.]

Let, the total Bias for a p-term model be-,

$$SS_B(p) = \sum_{i=1}^{n} [E(y_i) - E(\hat{y}_i)]^2.$$

Define the standardized total mean square as-

$$\Gamma(p) = \frac{1}{\sigma^2} \{SS_B(p) + \sum Var(\hat{y}_i)\} \quad -(1)$$

It can be shown that,

$$\sum_{i=1}^{n} Var(\hat{y}_i) = trace\,(Var(\hat{y})) = p\sigma^2 \quad -(2)$$

The expected value of the residual sum of squares from a p-term model is-

$$E(SS_{Res}(p)) = SS_B(p) + (n-p)\sigma^2.$$

(It can be shown) $-(3)$

~~Suppose that $\hat{\sigma}^2$ is a good estimator estimate of σ^2. Then replacing $E(SS_{Res}(p))$ by the~~

From ② & ③, ① reduces to -

$$\Gamma p = \frac{1}{\sigma^2}[E(SS_{Res}(p)) - (n-p)\sigma^2 + p\sigma^2]$$

$$= \frac{E(SS_{Res}(p))}{\sigma^2} - (n+2)p.$$

Suppose that, $\hat{\sigma}^2$ is a good estimate of σ^2, then replacing $E(SS_{Res}(p))$ by the observed value of $SS_{Res}(p)$. This will produce a reasonable estimate of Γ_p.

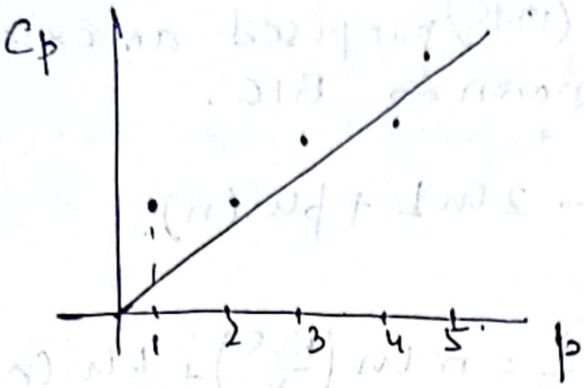
$$C_p = \frac{SS_{Res}(p)}{\hat{\sigma}^2} - (n+2)p.$$

If p-term model has negligible Bias, then $SS_B(p) = 0$, consequently,

$$E(SS_{Res}(p)) = (n-p)\sigma^2.$$

Then, $E(C_p \mid Bias = 0) = \frac{(n-p)\sigma^2}{\sigma^2} - (n+2)p = p.$

~~A plot of C_p vs p will look like~~



If C_p value is close to p, ~~we~~ for some p, we should choose that p. If there are more than one C_p values which are close to p, then the one with less p will be chosen.

⊳ THE AKAIKE INFORMATION CRITERION

06/10/2023.

Akaike proposed an information criterion, AIC based on maximizing the expected entropy of the model.

$$\boxed{AIC = -2\log(L) + 2p}$$

where, L := ~~log~~ likelihood function for a special model.

p := no. of parameters.

e.g.)
$$L = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{(Y-X\beta)^T(Y-X\beta)}{2\sigma^2}\right]$$

$$\hat{L} \propto \left(\frac{RSS}{n}\right)^{-n/2} e^{-n/2}. \quad \left[\hat{\sigma}^2 = \frac{RSS}{n} = \frac{(Y-X\hat{\beta})^T(Y-X\hat{\beta})}{n}, \quad \hat{\beta} = (X'X)^{-1}X'Y\right]$$

$$AIC = n\log\left(\frac{RSS}{n}\right) + 2p.$$

Bayesian Information Criterion

Schwartz (1978) proposed an extension of AIC, which is known as BIC.

$$BIC = -2\ln L + p\ln(n).$$

For a MLR,

$$BIC = n\ln\left(\frac{RSS}{n}\right) + p\ln(n)$$

▶ Another criteria may be given using PRESS STATISTICS. The model with less value of PRESS statistic should be preferred.

- **Stepwise Procedures**

1) **Forward Selection**

$\boxed{\text{Step 1}}$ The first regressor to enter is the one that has "largest simple correlation" with y. ~~and~~ Not only that, check the corresponding F-values.

If F value of that regressor > F_{IN} (userdefined) then only choose that.

[Note that in R, $F_{IN} = t^2_{0.25/2,\, n-p}$, by default].

$\boxed{\text{Step-2}}$. $F_j = \dfrac{SS_R(x_j \mid x_4)}{MS_{Res}(x_j, x_4)}$ [Say, x_4 was chosen in the first step].

Partial - F statistic.

for $j = 1, \ldots, p-1$. $j \neq 4$.

$$SS_R(x_j \mid x_4) = SS_R(x_j, x_4) - SS_R(x_4).$$

...

▷▷ BACKWARD ELIMINATION ≫ $\boxed{09/10/2023}$

Step-I ≫ We begin with a full model including all 'k' regressors.

Step-II ≫ Choose a cut-off value F_{OUT} (equivalently t_{OUT}).

Here, $F_j = \dfrac{SS_R(x_j \mid x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k)}{MS_{Res}(\text{full model})}$.

Partial F-Statistic.

$(t_j = \sqrt{F_j})$

where, $SS_R(x_j \mid x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k)$.

$= SS_R(\text{full}) - SS_R(x_j \text{ Removed})$

Step-III ≫ Choose $\min\limits_{j \in \{1, \ldots, k\}} F_j = F_{j^*} > F_{OUT}$ [in R, $t_{0.1/2, n-p}$ $t_{OUT} = F_{OUT}$]

$\Rightarrow x_{j^*}$ is remained in the model.

Step-IV ≫ Repeat step 1 to 3 until.

$F_j > F_{OUT} \; \forall j$

▷▷ STEPWISE REGRESSION

The ~~1/~~

First, we perform the forward selection method to select the first two regressors, then do the backward selection to decide whether to keep both the regressor or not.

...

11/10/2023

Now to propose the ultimate model, we need some further analysis.

① ANALYSIS OF MODEL COEFFICIENTS & PREDICTED VALUES

ⓐ The coefficients in the final model should be studied to determine if they are "stable" and if their "signs" and "magnitudes" are reasonable.

ⓑ Previous experience, theoretical considerations or an analytical model can often be used to validate ⓐ.

ⓒ Coefficients with unexpected signs or that are too large in magnitude often indicate either an inappropriate mode or poor estimates.

ⓓ The "VIFs" are other multiplicity diagonstics are important guide to validate the model.

ⓔ The predicted response values $\hat{y}$ also indicate which model to choose.

---

② Collect fresh data >

Based on the previously collected data, say we have some proposed models.

We collect some fresh data and perform some predictive analysis using the models.

Calculate $R^2_{prediction} = 1 - \frac{\sum (y_i^* - \hat{y}_i^*)^2}{\sum (y_i^* - \bar{y}^*)^2}$ for each of the model

If the value is close to 1, then the model is performing all.

Also, we can compare the average squared prediction error $= \frac{1}{n_{new}} \sum_{i=1}^{n_{new}} (y_i^* - \hat{y}_i^*)^2$ of the models.

③ DATA SPLITTING >

Divide the data set into training & testing dataset.

How to split ...

~~Accurate~~

DUPLEX Algorithm for data splitting.

— Proposed by Snee (1978).

1D A method to check whether the splitting is well enough or not,

if $\left[\frac{|X_E^T X_E|}{|X_P^T X_P|}\right]^{1/p} \approx 1.$

then the splitting is ~~good~~.

...—

X_E : ~~Es~~ Design matrix of Estimation dataset

X_P : Design matrix of Prediction dataset

13/10/2023.

⊙ There may be the situation that such that ~~the~~ some of the regressors are qualitative. In that situation, we introduce some <u>Dummy variables</u>.

<u>e.g</u> 7. Suppose that a mechanical engineer wishes to relate the effective life of a cutting tool (y) used on a lathe to the lathe speed in relatives per minute (x_2) and type of cutting tool used.

Types of cutting tool used are given by A & B.

~~Bo~~ Then the corresponding regressor ~~cat~~ is defined as –

$$x_2 = \begin{cases} 0, & \text{if tool type A is used.} \\ 1, & \text{if tool type B is used.} \end{cases}$$

This is a <u>dummy variable</u>.

~~So, for type A, +~~

Let the full model is given by –

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon.$$

Then, for type A, the mode becomes –.

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$

and, for type B, the model is –

$$y = \beta_0 + \beta_1 x_1 + \beta_2 + \epsilon$$

$$= (\beta_0 + \beta_2) + \beta_1 x_1 + \epsilon.$$

So, basically, if we want to test whether there is significant difference in the performance of two types of machines we may test –

$$\underline{H_0 : \beta_2 = 0,\ H_1 : \beta_2 \neq 0.}$$

Now, when we consider qualitative regressors, we should include all possible interaction terms in the model.

In the last example, the modified model may be -

$$\boxed{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 x_2) + \epsilon}$$

Interaction term.

Then for type A the model is -

$$y = \beta_0 + \beta_1 x_1 + \epsilon.$$

& For type B, the model is -

$$y = \beta_0 + \beta_1 x_1 + \beta_2 + \beta_3 x_1 + \epsilon.$$

$$= (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_1 + \epsilon.$$

And here that testing problem will become -

$$H_0: \beta_2 = \beta_3 = 0 \text{ vs. } H_1: \text{not } H_0.$$

...

Now suppose, there is a qualitative variable having 3 categories. Then the model will become -

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon.$$

where

	x_2	x_3
A	0	0
B	1	0
C	0	1

☒ In general, if we have a qualitative variable with 'k' levels, we need 'k-1' dummy variables to be introduced in the model.

☒ Note that, we can put any values instead of '0' & '1'. One can show that, this ~~variable~~ translation won't affect any result.

:) **ONE - WAY ANOVA** → Analysis of Variance

[Note that, we have only one month left for the end sem when we are starting this... :D]

Here we are considering ~~BF~~ balanced ANOVA only, i.e. no. of sample is same for each of the treatment group.

Let us consider an example consisting of 4 categories [trt. groups]

Define y_{ij} : the response of jth person nested in treatment group 'i'.

$$\begin{cases} i = 1(1)4 \\ j = 1(1)n_i \end{cases}$$

Data: $\begin{cases} y_{11} & y_{21} & & y_{31} & y_{41} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ y_{1n_1} & y_{2n_2} & & y_{3n_3} & y_{4n_4} \end{cases}$

The Model is given by —

$$\boxed{y_{ij} = \mu + \alpha_i + \epsilon_{ij}} = \mu_i + \epsilon_{ij} \quad [\text{where, } \mu_i = \mu + \alpha_i]$$

μ : parameter associated with the overall mean

α_i : parameter associated with ith trt. effect.

Here to check the differential effect of ~~the~~ different treatment groups, the corresponding hypothesis testing problem is given by –

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$

vs. H_1 : H_0 is not true.

$\Rightarrow \exists$ α_i for which $\alpha_i \neq 0$.

Define, $x_{1j} = \begin{cases} 1, & \text{if obs. comes from trt. 1} \\ 0, & \text{o/w.} \end{cases}$

$x_{2j} = \begin{cases} 1, & \text{if obs. comes from trt. 2} \\ 0, & \text{o/w.} \end{cases}$

The model becomes –

$$y_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \epsilon_{ij}$$

Trt 1 $\quad y_{ij} = \beta_0 + \beta_1 + \epsilon_{ij}$

Trt 2 $\quad y_{ij} = \beta_0 + \beta_2 + \epsilon_{ij}$

Trt 3 $\quad y_{ij} = \beta_0 + \epsilon_{ij}$

16/10/2023.

So, if we have 10 obs. per trt. group, the Y data becomes.

$$
\begin{aligned}
y_{11} &= \beta_0 + \beta_1 + \epsilon_{11} \\
y_{12} &= \beta_0 + \beta_1 + \epsilon_{11} \\
&\vdots \\
y_{1,10} &= \beta_0 + \beta_1 + \epsilon_{1,10} \\
y_{2,1} &= \beta_0 + \beta_2 + \epsilon_{2,1} \\
&\vdots \\
y_{2,10} &= \beta_0 + \beta_2 + \epsilon_{2,10} \\
y_{3,1} &= \beta_0 + \epsilon_{3,1} \\
&\vdots \\
y_{3,10} &= \beta_0 + \epsilon_{10}
\end{aligned}
\longrightarrow
\begin{bmatrix} \underset{\sim}{1}_{10} & \underset{\sim}{1}_{10} & \underset{\sim}{0}_{10} \\ \underset{\sim}{1}_{10} & \underset{\sim}{0}_{10} & \underset{\sim}{1}_{10} \\ \underset{\sim}{1}_{10} & \underset{\sim}{0}_{10} & \underset{\sim}{0}_{10} \end{bmatrix}
$$

$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \epsilon_{ij}$

	A	B	C
x_1	1	0	0
x_2	0	1	0

In general,

$$\underset{\sim}{Y} = X \underset{\sim}{\beta} + \underset{\sim}{\epsilon}$$

where
$$X = \begin{bmatrix} \underset{\sim}{1}_n & \underset{\sim}{1}_n & \underset{\sim}{0}_n & \cdots & \underset{\sim}{0}_n \\ \underset{\sim}{1}_n & \underset{\sim}{0}_n & \underset{\sim}{1}_n & \cdots & \underset{\sim}{0}_n \\ \vdots & \vdots & \vdots & & \vdots \\ & & & & \underset{\sim}{1}_n \\ \underset{\sim}{1}_n & \underset{\sim}{0}_n & \underset{\sim}{0}_n & \cdots & \underset{\sim}{0}_n \end{bmatrix}$$

$$\underset{\sim}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{bmatrix} \quad \therefore \hat{\beta} = (X^T X)^{-1} X^T Y$$

Now,
$$X^T X = \left[\begin{array}{c|cccc} nk & n & n & \cdots & n \\ \hline n & n & 0 & \cdots & 0 \\ n & 0 & n & \cdots & 0 \\ \vdots & & & \ddots & \\ n & 0 & 0 & \cdots & n \end{array}\right]$$

$$= n \begin{bmatrix} k & \underset{\sim}{1}_{k-1}^T \\ \underset{\sim}{1}_{k-1} & I_{k-1} \end{bmatrix}$$

To invert this we need a formula:

$$\begin{bmatrix} a & v^T \\ v & I \end{bmatrix}^{-1} = \begin{bmatrix} (a - v^T v)^{-1} & -(a - v^T v)^{-1} v^T \\ -v(a - v^T v) & I + v(a - v^T v)^{-1} v^T \end{bmatrix}$$

Here, $(X^TX)^{-1} = \frac{1}{n}\begin{bmatrix} 1 & -1'_{k-1} \\ -1_{k-1} & I + 1_{k-1}1'_{k-1} \end{bmatrix}$

$$X^T\underline{Y} = \begin{bmatrix} 1_n^T & 1_n^T & \cdots & 1_n^T \\ 1_n^T & 0 & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 \cdots & 1_n & 0 \end{bmatrix}\begin{bmatrix} \underset{\sim}{y}_1 \\ \underset{\sim}{y}_2 \\ \vdots \\ \underset{\sim}{y}_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij} \\ \sum_{j=1}^{n} y_{1j} \\ \vdots \\ \sum_{j=1}^{n} y_{k-1,j} \end{bmatrix}$$

$$\therefore \hat{\beta} = (X^TX)^{-1}(X^T\underline{Y}) = \frac{1}{n}\begin{bmatrix} \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij} - \sum_{j=1}^{n} y_{1j} - \cdots - \sum_{j=1}^{n} y_{k-1,j} \\ -\sum\sum y_{ij} + 2\sum y_{1j} + \sum y_{2j} + \cdots + \sum y_{k-1,j} \\ -\sum\sum y_{ij} + \sum y_{1j} + 2\sum y_{2j} + \cdots + \sum y_{k-1,j} \\ \vdots \\ -\sum\sum y_{ij} + \sum y_{1j} + \cdots + \sum y_{k-2,j} + 2\sum y_{k-1,j} \end{bmatrix}$$

$$= \frac{1}{n}\begin{bmatrix} \sum_{j=1}^{n} y_{k,j} \\ \sum_{j=1}^{n} y_{1j} - \sum_{j=1}^{n} y_{k,j} \\ \vdots \\ \sum_{j=1}^{n} y_{k-1,j} - \sum_{j=1}^{n} y_{k,j} \end{bmatrix} = \begin{bmatrix} \bar{y}_k \\ \bar{y}_1 - \bar{y}_k \\ \vdots \\ \bar{y}_{k-1} - \bar{y}_k \end{bmatrix}$$

$\left[\because \bar{y}_i = \frac{1}{n}\sum_{j=1}^{n} y_{ij}\right]$ $i = 1(1)k$.

Now, we know,

$$SS_R = (\hat{y} - \bar{y}_{..})^T(\hat{y} - \bar{y}_{..})$$

$$SS_{Res} = (y - \hat{y})^T(y - \hat{y})$$

$$\hat{y} = X\hat{\beta} = \begin{bmatrix} \bar{y}_1 1_n \\ \bar{y}_2 1_n \\ \vdots \\ \bar{y}_k 1_n \end{bmatrix} = \begin{bmatrix} 1_n\left(\frac{1_n^T}{n}\underset{\sim}{y}_1\right) \\ \vdots \\ 1_n\left(\frac{1_n^T}{n}\underset{\sim}{y}_k\right) \end{bmatrix} \quad [J_n = 1_n 1_n^T]$$

$$= \begin{bmatrix} \frac{1}{n}J_n & & 0 \\ & \frac{1}{n}J_n & \\ & & \ddots \\ 0 & & \frac{1}{n}J_n \end{bmatrix}\underset{\sim}{y}$$

$$= J^* \underset{\sim}{y}$$

18/10/2023

The F-statistic is given by -

$$F = MS_R / MS_{Res}.$$

$$SS_{Res} = (y-\hat{y})^T(y-\hat{y}) = y^T(I-J_\star)y.$$

$$\left[\text{As } \hat{y} = J_\star y \quad \text{where } J_\star = \begin{bmatrix} \frac{1}{n}J_n & & 0 \\ & \frac{1}{n}J_n & \\ 0 & \cdots & \frac{1}{n}J_n \end{bmatrix}\right]$$

$$\underset{\sim}{y} \sim MVN\left[\begin{pmatrix}(\mu+\alpha_1)1_n \\ \vdots \\ (\mu+\alpha_k)1_n\end{pmatrix}, \sigma^2 I\right]$$

$$\begin{bmatrix}(\mu+\alpha_1)1_n \\ \vdots \\ (\mu+\alpha_k)1_n\end{bmatrix}'(I-J_\star)\begin{bmatrix}(\mu+\alpha_1)\underline{1}_n \\ (\mu+\alpha_2)\underline{1}_n \\ \vdots \\ (\mu+\alpha_k)\underline{1}_n\end{bmatrix} = 0 \text{ [Show]}$$

Hence, $\frac{SS_{Res}}{\sigma^2} \sim \chi^2_{nk-k}$
$\equiv k(n-1)$.

...

$$SS_R = (\hat{y} - 1_{nk}\bar{y}_{..})^T(\hat{y} - 1_{nk}\bar{y}_{..}).$$

$$\hat{y} = J_\star y, \quad \bar{y}_{..} = \frac{1}{nk}1_{nk}^T\underset{\sim}{y}.$$

$$= \left(J_\star y - 1_{nk}\left(\frac{1}{nk}1_{nk}^T y\right)\right)^T(\underline{\text{Same term.}})$$

$$= \left(J_\star y - \frac{1}{nk}J_{nk}y\right)^T(\underline{\text{same}})$$

$$= \underline{y}^T\left(J_\star - \frac{J_{nk}}{nk}\right)^T\left(J_\star - \frac{J_{nk}}{nk}\right)\underline{y}.$$

Scanned with CamScanner

অদ্ভুত বাবা

অদ্ভুত বাবা ✓ this is better

Now, $\left(J_{*}-\frac{J_{nk}}{nk}\right)\left(J_{*}-\frac{J_{nk}}{nk}\right)$.

$$= J_{*}J_{*} - \frac{1}{nk} J_{nk} J_{*} - \frac{1}{nk} J_{*} J_{nk} + \frac{1}{n^2k^2} J_{nk} \cdot J_{nk}$$

$$= J_{*} - \frac{1}{nk} J_{nk} - \frac{1}{nk} J_{nk} + \frac{1}{n^2k^2} \cdot nk J_{nk}$$

$$= J_{*} - \frac{1}{nk} J_{nk}.$$

Therefore $\frac{SS_R}{\sigma^2} = \frac{1}{\sigma^2}\left[y^T\left(J_{*}-\frac{1}{nk}J_{nk}\right)y\right]$

$$\sim \chi^2_{k-1}$$

[check the non-centrality parameter]

Now, SS_R & SS_{Res} are independent [show]

$$F = \frac{MS_R}{MS_{Res}} \sim F_{k-1,\,k(n-1)}$$

...

30/10/2023.

EXPERIMENTAL DESIGN (ED):- [And now we have 2 weeks left for the end sem. All the Best !!

→ **EXPERIMENTAL UNITS (EU)** :- For conducting an E, the experimental material is divided into smaller parts and each part is referred to as an EU. The EU is randomly assigned to treatment.

→ **SAMPLING UNITS (SU)** :- The object that is measured in a 'E' is called SU. SU may be different from EU.

→ **TREATMENT** :- Different procedures/objects which are to be compared.

→ **FACTOR** :- A factor is a variable defining categorization.

→ **REPLICATION** :-

T_1	T_2	...	T_k
○	○	....	○
n_1 times	n_2 times		n_k times.

▶ PRINCIPLES OF ED :-

(i) RANDOMIZATION.

(ii) REPLICATION.

(iii) LOCAL CONTROL.

▶ COMPLETELY RANDOMISED DESIGN (CRD) :-

It is the simplest design. Suppose there are 'k' treatments to be compared.

→ All EUs are considered same & no division or grouping among them exist.

→ In CRD, the 'V' treatments are randomly allotted to the whole set of EUs.

→ Design is entirely flexible in the sense that any no. of treatments or replications may be used.

MODEL ⟩

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad i=1(1)k, \; j=1(1)n_i$$

The testing of hypothesis problem is given by -

$H_0: \alpha_1 = \dots = \alpha_k = 0$.

vs. H_1: Not H_0. condition: $\sum n_i \alpha_i = 0$.

We are following least square technique.

$$S(\mu, \underline{\alpha}) = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \epsilon_{ij}^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\mu-\alpha_i)^2$$

$$\frac{\partial S}{\partial \mu} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\mu-\alpha_i) = 0 \Rightarrow \hat{\mu} = \bar{y}$$

$$\frac{\partial S}{\partial \alpha_i} = \sum_{j=1}^{n_i}(y_{ij}-\mu-\alpha_i) = 0$$

$$\Rightarrow \hat{\alpha}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}).$$

$$= \bar{y}_{i.} - \bar{y}$$

→ PARTICIPATION OF SUM OF SQUARES ⟩

$$y_{ij} = \bar{y} + (\bar{y}_{i.} - \bar{y}) + (y_{ij} - \bar{y}_{i.})$$

$$(y_{ij} - \bar{y}) = (y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y})$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2 + \sum_{i=1}^{k} n_i(\bar{y}_{i.}-\bar{y})^2$$

↳ TSS = SSR + SSE

Then the F- statistic will be —.

$$F = \frac{SSR/df_R}{SSE/df_E} \sim F_{df_R, df_E}$$

$$\begin{cases} df_E = \sum_{i=1}^{k} n_i - k \\ df_R = k-1. \end{cases}$$

01/11/2023

Randomised Block Design (RBD)

let's take an example:

Suppose there are three methods of teaching A, B & C.

Suppose there are three categories of students having different IQ levels, i.e. the population is not homogeneous.

IQ 1-5	A	C	B
IQ 5-10	C	A	B
IQ 11-15	B	A	C

- When the EUs are not homogeneous, and there are 'k' treatments to be compared, then it may be possible to
  1. group the EUs into 'blocks' of size k units.
  2. Blocks are constructed such that the EUs within a block are relatively homogeneous and resemble to each other more closely than the units in different blocks.
  3. If there are 'b' blocks, we say that blocks are at 'b' levels. Similarly if there are 'k' treatments, we say that treatments are at 'k' levels.

Treatment (Factor A)

Blocks (Factor B)

	1	2	...	j	...	k	Block Totals
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1k}	$B_1 = \sum_{j=1}^{k} y_{1j}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2k}	$B_2 = \sum_{j=1}^{k} y_{2j}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ik}	$B_i = \sum_{j=1}^{k} y_{ij}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
b	y_{b1}	y_{b2}	...	y_{bj}	...	y_{bk}	$B_b = \sum_{j=1}^{k} y_{bj}$
Treatment Totals	$T_1 = \sum_{i=1}^{b} y_{i1}$	$T_2 = \sum_{i=1}^{b} y_{i2}$	....	$T_j = \sum_{i=1}^{b} y_{ij}$	...	$T_k = \sum_{i=1}^{b} y_{ik}$	

e.g.) Suppose we have 7 treatments $(=k)$ & 4 blocks $(=b)$.

Treatments.

Blocks							
1	T_1	T_3	T_4	T_5	T_6	T_2	T_7
2	T_2	T_4	T_6	T_7	T_5	T_3	T_1
3	T_1	T_3	T_5	T_7	T_2	T_4	T_6
4	T_1	T_2	T_6	T_7	T_3	T_4	T_5

Model.

$$y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

$1 \le i \le b$

$1 \le j \le k$.

β_i: Block Effect

τ_j: Treatment Effect

$$\left\{ \begin{array}{l} \sum \beta_i = 0 \\ \sum \tau_j = 0 \end{array} \right.$$

03/11/2023

ANALYSIS:

$$SS_T = SS_{Bl} + SS_{Tr} + SSE.$$

$$\sum_{i=1}^{b}\sum_{j=1}^{k}(y_{ij}-\bar{y})^2 = k\sum_{i=1}^{b}(\bar{y}_{i.}-\bar{y})^2 + b(\bar{y}_{.j}-\bar{y})^2$$
$$+ \sum_{i=1}^{b}\sum_{j=1}^{k}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j})^2$$

$H_{0B}: \beta_1 = \dots = \beta_b = 0.$

$$F_b = \frac{MS_{Bl}}{MSE} = \frac{SS_{Bl}/b-1}{SSE/(b-1)(k-1)}$$

$$\sim F_{b-1,(b-1)(k-1)}.$$

ANOVA TABLE:

Source of variation	df	SS	MS	F
Treatment	$k-1$	SS_{Trt}	MS_{Trt}	F_{Trt}
Block	$b-1$	SS_{Bl}	MS_{Bl}	F_{Bl}.
Error	$(b-1)(k-1)$	SSE	MSE.	

$H_{0\,Tr}: \tau_1 = \dots = \tau_k = 0$

$$F_{Tr} = \frac{MS_{Tr}}{MSE} = \frac{SS_{Tr}/k-1}{SSE\,(b-1)(k-1)}$$

$$\sim F_{k-1,(b-1)(k-1)}$$

Latin Square Design (LSD)

Let's start with an example —

Say, 4 Petrol Brands are there – A, B, C & D.
and 4 ~~types~~ brands of car → 1, 2, 3, 4.
Also there are 4 types of drivers → a, b, c, d.

DRIVERS \ CARS	1	2	3	4
a	A	B	C	D
b	B	C	D	A
c	C	D	A	B
d	D	A	B	C

Model
$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}.$$

α_i : main effect of rows (drivers)

β_j : " " " columns (cars)

τ_k : " " " treatments (Brands).

with $\sum \alpha_i = 0$, $\sum \beta_j = 0$, $\sum \tau_k = 0$

$$\hat{\mu} = \bar{y}$$

$$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}, \quad i = 1(1)k \quad \left[\bar{y}_{i..} = \frac{1}{k^2}\sum_{j=1}^{k}\sum_{k=1}^{k} y_{ijk}\right]$$

$$\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}, \quad j = 1(1)k \quad \left[\bar{y}_{.j.} = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{k=1}^{k} y_{ijk}\right]$$

$$\hat{\tau}_k = \bar{y}_{..k} - \bar{y}, \quad k = 1(1)k \quad \left[\bar{y}_{..k} = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k} y_{ijk}\right]$$

$$TSS = SS_{Row} + SS_{Col} + SS_{Trt} + SSE.$$

$$SS_{Row} = k\sum_{i=1}^{k}(\bar{y}_{i..} - \bar{y})^2 \longrightarrow df: (k-1)$$

$$SS_{Col} = k\sum_{j=1}^{k}(\bar{y}_{.j.} - \bar{y})^2 \longrightarrow df: (k-1)$$

$$SS_{Trt} = k\sum_{k=1}^{k}(\bar{y}_{..k} - \bar{y})^2 \longrightarrow df: (k-1)$$

$H_{0R}: \alpha_1 = \dots = \alpha_k = 0.$

$$F_R = \frac{SS_R/(k-1)}{SSE/(k-1)(k-2)} \sim F_{k-1,(k-1)(k-2)}.$$

$H_{0C}: \beta_1 = \dots = \beta_k = 0$

$$F_C = \frac{SS_C/(k-1)}{SSE/(k-1)(k-2)} \sim \text{"}$$

$H_{0Trt}: \tau_1 = \dots = \tau_k = 0.$

$$F_{trt} = \frac{SS_{Trt}/(k-1)}{SSE/(k-1)(k-2)} \sim \text{"}.$$

04/11/2023.

GENERALISED LINEAR MODELS. (GLM)

[ONE WEEK TO GO...!! :')]

EXPONENTIAL FAMILY OF DISTRIBUTIONS. (EFD)

Let us have a sample $(Y_1, Y_2, \ldots, Y_n)$ such that,

$$Y_i \sim EFD(\theta_i).$$

The PDF/PMF is –

$$f(y_i ; \theta_i) = \exp\left[\{y_i\theta_i - b(\theta_i)\}/a(\phi) + c(y_i ; \phi)\right].$$

where a, b, c are known functions.

θ_i : Natural parameters.

ϕ : Dispersion parameter.

Let, $L_i = \log f(y_i ; \theta_i)$

$$= [y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i ; \phi).$$

RESULT Under some regularity.

$$E\left[\frac{\partial L_i}{\partial \theta_i}\right] = 0 \ \& \ E\left[\frac{\partial^2 L_i}{\partial \theta_i^2}\right] = -E\left[\frac{\partial L_i}{\partial \theta_i}\right]^2$$

Proof → $\int f(y_i, \theta_i)\, dy_i = 1.$

$$\Rightarrow \int \frac{\partial}{\partial \theta_i} f(y_i, \theta_i)\, dy_i = 0.$$

$$\Rightarrow \int \frac{\partial \log f(y_i, \theta_i)}{\partial \theta_i} f(y_i, \theta_i)\, dy_i = 0 \quad -(1)$$

$$\Rightarrow E\left(\frac{\partial L_i}{\partial \theta_i}\right) = 0. \quad [L_i = \log f(y_i, \theta_i)]$$

Take partial derivative w.r.t. θ_i of (1),

$$\int \left[\frac{\partial \log f_i}{\partial \theta_i} \cdot \frac{\partial f_i}{\partial \theta_i} + f_i \frac{\partial^2 \log f_i}{\partial \theta_i^2}\right] dy_i = 0$$

$$\Rightarrow \int \left[\frac{\partial \log f_i}{\partial \theta_i} \cdot \frac{\partial \log f_i}{\partial \theta_i} \cdot f_i + f_i \frac{\partial^2 \log f_i}{\partial \theta_i^2}\right] dy_i = 0$$

$$\Rightarrow E\left(\frac{\partial L_i}{\partial \theta_i}\right)^2 = -E\left(\frac{\partial^2 L_i}{\partial \theta_i^2}\right). \text{ [Proved]}$$

Mean of Y_i.

$$\frac{\partial L_i}{\partial \theta_i} = [Y_i - b'(\theta_i)] / a(\phi).$$

$$E\left(\frac{\partial L_i}{\partial \theta_i}\right) = E\left(\frac{Y_i - b'(\theta_i)}{a(\phi)}\right)$$

$$\Rightarrow \boxed{E(Y_i) = b'(\theta_i)}$$

Scanned with CamScanner

Again, $$\frac{\partial^2 L_i}{\partial \theta_i^2} = \frac{-b''(\theta_i)}{a(\phi)}.$$

$$E\left(\frac{\partial^2 L_i}{\partial \theta_i^2}\right) = -\frac{b''(\theta_i)}{a(\phi)}.$$

$$\Rightarrow -E\left(\frac{\partial L_i}{\partial \theta_i}\right) = -\frac{b''(\theta_i)}{a(\phi)}$$

$$E\left[\left\{(y_i - b'(\theta_i))/a(\phi)\right\}^2\right] = \frac{b''(\theta_i)}{a(\phi)}$$

$$\Rightarrow E\left[\frac{\{y_i - E(y_i)\}^2}{[a(\phi)]^2}\right] = \frac{b''(\theta_i)}{a(\phi)}.$$

$$\Rightarrow \operatorname{var}(y_i) = \frac{b''(\theta_i)}{a(\phi)}$$

e.g.) POISSON DISTRIBUTION:

$$f(y_i, \mu_i) = \frac{e^{-\mu_i} \cdot \mu_i^{y_i}}{y_i!}.$$

$$= \exp[-\mu_i + y_i \log \mu_i - \log(y_i!)]$$

$$= \exp[y_i \log \mu_i - \mu_i - \log(y_i!)].$$

Comparing we get -
$$\begin{cases} \theta_i = \log \mu_i \\ b(\theta_i) = \mu_i = e^{\theta_i} \\ a(\phi) = 1 \end{cases}$$

So,
$$\begin{cases} E(y_i) = b'(\theta_i) \\ \qquad = e^{\theta_i} = \mu_i \\ \operatorname{var}(y_i) = \frac{e^{\theta_i}}{1} = \mu_i \end{cases}$$

$\underline{\text{example}}$ 7 $\quad Y_i \sim Bin(n_i, \pi_i)$

$$f(y_i; n_i, \pi_i) = \binom{n_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{n_i - y_i}$$

$$\cancel{= \exp\left[\log\binom{n_i}{y_i} \log n_i! - \log\right.}$$

$$= \exp\left[\log\binom{n_i}{y_i} + y_i \log \pi_i + (n_i - y_i)\log(1-\pi_i)\right]$$

$$= \exp\left[\log\binom{n_i}{y_i} + y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + n_i \log(1-\pi_i)\right]$$

$$= \exp\left[y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + n_i \log(1-\pi_i)\right] + \exp\left[\log\binom{n_i}{y_i}\right]$$

$$\theta_i = \log \frac{\pi_i}{1-\pi_i}$$

$$b(\theta_i) = -n_i \log(1-\pi_i) \; ; \; a(\phi) = 1$$

$$\pi_i = \frac{e^{\theta_i}}{1+e^{\theta_i}}$$

$$b'(\theta_i) = n_i \cdot \frac{1}{1+e^{\theta_i}} \cdot e^{\theta_i}$$

$$= n_i \pi_i$$

$$\therefore E(Y_i) = n_i \pi_i$$

$$b''(\theta_i) = n_i \left[\frac{(1+e^{\theta_i})e^{\theta_i} - e^{\theta_i} e^{\theta_i}}{(1+e^{\theta_i})^2}\right]$$

$$= n_i \frac{e^{\theta_i}}{(1+e^{\theta_i})^2}.$$

$$= n_i \pi_i (1-\pi_i).$$

$$var(Y_i) = n_i \pi_i (1-\pi_i).$$

06/11/2023.

Example: $\boxed{\text{GAUSSIAN DISTRIBUTION}}$

$$Y_i \sim N(\mu_i; \sigma^2).$$

$$f(y_i; \mu_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i-\mu_i)^2}{2\sigma^2}\right]$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{y_i^2}{2\sigma^2} - \frac{\mu_i^2}{2\sigma^2} + \frac{y_i\mu_i}{\sigma^2}\right]$$

$$= \exp\left[y_i.\frac{\mu_i}{\sigma^2} - \frac{\mu_i^2}{2\sigma^2} - \frac{y_i^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right]$$

$$\theta_i = \frac{\mu_i}{\sigma^2}, \quad b(\theta_i) = \frac{\mu_i^2}{2\sigma^2} = \frac{1}{2}\theta_i^2\sigma^2.$$

$$b'(\theta_i) = \theta_i\sigma^2. \quad b''(\theta_i) = \sigma^2$$

$$a(\phi) = 1$$

Then
$$E(Y_i) = b'(\theta_i) = \theta_i\sigma^2 = \mu_i$$

$$V(Y_i) = \frac{b''(\theta_i)}{a(\phi)} = \sigma^2.$$

▷ LINK FUNCTION // Let $(x_{i1}, \ldots, x_{ip})$ denote the values of explanatory / regressor variables for observation 'i'.

The systematic components of a GLM relates parameters to these variables using a "linear predictor" defined as.

$$\eta_i = \sum_{j=1}^{p} \beta_j x_{ij}, \text{ for } i=1(1)n.$$

$$\Rightarrow \underset{\sim}{\eta} = X^{n\times p} \underset{\sim}{\beta} \quad \left[\underset{\sim}{\eta} = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_n \end{pmatrix}, \underset{\sim}{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \right]$$

The linear predictor η_i links to the mean of Y_i, i.e. $E(Y_i)$ by a link function $g(\cdot)$.

$$\eta_i = g(\mu_i) = \sum_{j=1}^{p} \beta_j x_{ij}$$

Likelihood.

For $i=1(1)n$.

The log likelihood is -.

$$L(\underset{\sim}{\beta}) = \sum_{i=1}^{n} L_i = \sum \log f(y_i, \theta_i, \phi)$$

$$= \sum_{i=1}^{n} \frac{y_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi)$$

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum \frac{\partial L_i}{\partial \beta_j} = 0, \; j=1(1)p$$

$$\frac{\partial L_i}{\partial \beta_j} = \frac{\partial L_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j}$$

$$\frac{\partial L_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)}$$

$$\frac{\partial \mu_i}{\partial \theta_i} = \frac{\partial b'(\theta_i)}{\partial \theta_i} = b''(\theta_i) . = \frac{var(Y_i)}{a(\phi)}.$$

$$\eta_i = \sum \beta_j x_{ij} \qquad \frac{\partial \eta_i}{\partial \beta_j} = x_{ij}.$$

$$\Rightarrow \frac{y_i - b'(\theta_i)}{a(\phi)} \cdot \frac{a(\phi)}{var(Y_i)} \cdot x_{ij} \cdot \frac{\partial \mu_i}{\partial \eta_i} = 0.$$

$$\Rightarrow \sum_{i=1}^{n} \frac{y_i - \mu_i}{var(Y_i)} \cdot x_{ij} \cdot \frac{\partial \mu_i}{\partial \eta_i} = 0.$$

$j = 1(1)p$

--o--

08/11/2023

COVARIANCE / INFORMATION MATRIX

$$(I)_{(h,j)} = E\left(\frac{\partial^2 l(\beta)}{\partial \beta_h \partial \beta_j}\right). \qquad \begin{array}{l} h = 0(1)p-1 \\ j = 0(1)p-1. \end{array}$$

$$= -E\left[\left(\frac{\partial l(\beta)}{\partial \beta_h}\right)\left(\frac{\partial l(\beta)}{\partial \beta_j}\right)\right]$$

$$\frac{\partial^2 l(\beta)}{\partial \beta_h \partial \beta_j} = -\frac{x_{ih} x_{ij}}{var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \longrightarrow \text{check...}$$

In matrix form.

$$\mathbb{I} = X^T W X$$

where,

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & & & \\ x_{n1} & \cdots & & x_{np} \end{bmatrix}_{n\times p} \quad \text{and } W = diag\left[\frac{1}{Var(y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right]$$

$$\beta_{\sim} = [\beta_0 \ \beta_1 \ldots \beta_{p-1}]'$$

$$\eta(\underset{\sim}{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{p-1} x_{p-1}$$

Logistic Regression

Here, $g(\pi_i) = g(\mu_i) = \eta_i$

$$g(x) = \log\left(\frac{x}{1-x}\right)$$

Then, $\mu_i = e^{\eta_i} / 1+e^{\eta_i} = \pi_i$

The score equations are given by —.

$j = 0(1)p-1$.

$$\sum_{i=1}^{n} \frac{y_i - \pi_i}{\pi_i(1-\pi_i)} \cdot \pi_i(1-\pi_i) \, x_{ij} = 0 .$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \pi_i) x_{ij} = 0 .$$

Now, $Var(\hat{\beta}_{MLE}) = I^{-1}$

Therefore, $\widehat{Var}(\hat{\beta}_{MLE}) = (X^T W X)^{-1} = (X^T \hat{W} X)^{-1}$

$$\Rightarrow \widehat{Var}(\hat{\beta}_j) = (\hat{I}_{jj}^{-1})$$

Here, $\hat{\pi}_i = e^{\hat{\eta}_i} / 1+e^{\hat{\eta}_i}$ where $\hat{\eta}_i = \sum x_{ij} \beta_j$

Testing problem...

$$\left.\begin{array}{l} H_0: \beta_j = 0 \\ H_1: \beta_j \neq 0 \end{array}\right\}$$

$\boxed{\text{Test statistic}}$

$$T_j = \hat{\beta_j} / SE(\hat{\beta_j}) \sim N(0,1).$$

This statistic is called $\boxed{\text{Wald statistic}}$

...

10/11/2023

Poisson Regression

$$\eta_i = \sum_{j=1}^{k} \beta_j x_{ij}, \quad i = 1(1)n.$$

$$f(y_i; \beta) = \frac{e^{-\mu_i} \cdot \mu_i^{y_i}}{y_i!}, \quad i = 1(1)n.$$

$$g(\mu_i) = \eta_i \qquad \mu_i = e^{\eta_i}$$

$$L(\beta) = \prod_{i=1}^{n} \frac{\exp[-\mu_i] \cdot \mu_i^{y_i}}{y_i!}$$

$$l(\beta) = \log(\beta) = \sum_{i=1}^{n} (-\mu_i + y_i \log \mu_i - \log y_i!).$$

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} \left(-\frac{\partial \mu_i}{\partial \beta_j} + \frac{y_i}{\mu_i} \cdot \frac{\partial \mu_i}{\partial \beta_j} \right) = 0.$$

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial e^{\eta_i}}{\partial \beta_j} = \frac{\partial e^{\sum \beta_j x_{ij}}}{\partial \beta_j} = e^{\sum \beta_j x_{ij}} \cdot x_{ij}$$

$$= \mu_i \cdot x_{ij}$$

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} \left\{ -\mu_i x_{ij} + \frac{y_i}{\mu_i} \cdot \mu_i x_{ij} \right\} = 0$$

$$= \sum_{i=1}^{n} x_{ij} (y_i - \mu_i) = 0 \quad j = 1(1)p.$$

$$\frac{\partial^2 l}{\partial \beta_n \partial \beta_j} = \sum x_{ij} \left(-\frac{\partial \mu_i}{\partial \beta_n} \right)$$

$$= -\sum x_{ij} \, x_{in} \mu_i$$

$$\text{Var}(\hat{\beta}) \approx \underset{\downarrow}{X^T W X} \quad \text{where } w_i = \mu_i.$$

$$\text{diag}(w_1, \ldots, w_n).$$

The score equation is given by -

$$\sum \left(\frac{y_i - \mu_i}{\text{Var}(Y_i)} \right) x_{ij} \frac{\partial \mu_i}{\partial \eta_i} = 0.$$

$$\text{Var}(Y_i) = \mu_i$$

$$\frac{\partial \mu_i}{\partial \eta_i} = e^{\eta_i} = \mu_i$$

and $W_i = \text{diag}(w)$, $i = 1(1)n$

$$= \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \cdot \frac{1}{\text{Var}(Y_i)}.$$

$$= (\mu_i)^2 \cdot \frac{1}{\mu_i}$$

$$= \mu_i$$

Then, we can again form a Wald statistic for the testing of hypothesis.