$$SST_p = \sum_i (y_i - \bar{Y}_{(i)})^2$$

$$RSS_p = \sum (y_i - \underline{x}_i' \hat{\beta}_{(i)})^2$$

$\hat{\underline{\beta}}_{(i)} \longrightarrow$ estimate of β when ith observation has not been used.

$$\underline{x}_i' \hat{\beta}_{(i)} = \hat{Y}_{(i)}$$

$$X = \begin{pmatrix} \underline{x}_1' \\ \underline{x}_2' \\ \vdots \\ \underline{x}_n' \end{pmatrix}$$

$$y_i = \underline{x}_i' \underline{\beta} + \epsilon_i \quad , \quad i = 1, 2, \ldots n$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i$$

$$x_i' = (1, x_{i1}, \ldots, x_{i,p-1})$$

Generalyzed Least Sq.

So for the model $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$,

$E(\underline{\epsilon}) = \underline{0}$; $V(\underline{\epsilon}) = \sigma^2 I_n$

Q: What happen if allow the ϵ_i's to be correlated.

Assume, $V(\underline{\epsilon}) = \sigma^2 V$, V: Known pd matrix

Since V is p.d. $\exists$ an $n\times n$ n-s matrix.

Say R $\ni$ $V = RR'$

$\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ X is $n\times p$ with $r(X) = p$

$R^{-1}\underline{Y} = R^{-1}X\underline{\beta} + R^{-1}\underline{\epsilon}$ $R^{-1}\underline{Y} = \underline{Z}$

$R^{-1}X = B$

$\Rightarrow \underline{Z} = B\underline{\beta} + \underline{\eta}$ $\underline{\eta} = R^{-1}\underline{\epsilon}$

B is a $n\times p$ of rank p.

$E(\underline{\eta}) = \underline{0}$; $V(\underline{\eta}) = Var(R^{-1}\underline{\epsilon}) = \sigma^2 I_n$

Minimising $\underline{\eta}'\underline{\eta}$ w.r.t. $\underline{\beta}$, the LSE of $\underline{\beta}$ for the transformed model.

$$\underline{\beta}^* = (B'B)^{-1}B'\underline{Z}$$

$$= (X'(R^{-1})'R^{-1}X)^{-1}X'(R^{-1})'R^{-1}Y$$

$$= (X'(RR')^{-1}X)^{-1}X'(RR')^{-1}\underline{Y}$$

$$= (X'V^{-1}X)^{-1}X'V^{-1}Y$$

$\longrightarrow$ GLSE of $\underline{\beta}$

$$\underline{\eta}'\underline{\eta} = \underline{\epsilon}'V^{-1}\underline{\epsilon} = (\underline{Y} - X\underline{\beta})'V^{-1}(\underline{Y} - X\underline{\beta})$$

$$= Y'V^{-1}Y - 2\beta'X'V^{-1}Y + \beta'X'V^{-1}X\beta$$

$$\frac{\partial \eta'\eta}{\partial \beta} = -2X'V^{-1}\underline{Y} + 2X'V^{-1}X\underline{\beta} = 0$$

↳ (normal equation)

$$\underline{\beta}^* = (X'V^{-1}X)^{-1} X'V^{-1}\underline{Y}$$

$$E(\beta^*) = (X'V^{-1}X)^{-1} X'V^{-1} E(\underline{Y})$$

$$= (X'V^{-1}X)^{-1} X'V^{-1}X\beta$$

$$= \underline{\beta}$$

$$Var(\beta^*) = V\left((X'V^{-1}X)^{-1} X'V^{-1}\underline{Y}\right)$$

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

$$RSS = \hat{\underline{\eta}}'\hat{\underline{\eta}} = \hat{\underline{\epsilon}}' V^{-1}\hat{\underline{\epsilon}} = (\underline{Y} - X\underline{\beta}^*)'V^{-1}(\underline{Y} - X\underline{\beta}^*)$$

GLSE → a sp. case when V is diagonal.

→ uncorrelated but non-const variance.

Example: $\underline{Y} = \underline{X}\beta + \underline{\epsilon}$, $\underline{Y}_{n\times 1} = (Y_i)$, $\underline{X}_{n\times 1} = (x_i)$

$E(\underline{\epsilon}) = 0$, $Var(\underline{\epsilon}) = \sigma^2 V$

$V = diag\{w_1^{-1}, w_2^{-1}, \ldots, w_n^{-1}\}$, $w_i > 0$

Wtd LS (WLSE) of β.

$$\underline{\eta}'\underline{\eta} = \sum_{i=1}^{n} (Y_i - x_i\beta)^2 w_i \quad \text{[wtd sum of sq]}$$

$$\frac{\partial \underline{\eta}'\underline{\eta}}{\partial \beta} = -2\sum_{i=1}^{n} (Y_i - x_i\beta)\, x_i w_i = 0$$

$$\Rightarrow \quad \beta^* = \frac{\sum_i w_i Y_i x_i}{\sum w_i x_i^2}$$

$$\mathrm{Var}(\beta^*) = \sigma^2 \left(\sum_i w_i x_i^2\right)^{-1}$$

<u>Ex:</u> $Y_i = \beta x_i + \epsilon_i, \quad i = 1, 2$ $\quad \epsilon_1 \sim N(0, \sigma^2)$, $\epsilon_2 \sim N(0, 2\sigma^2)$

and ϵ_1 & ϵ_2 are statistically indp.

If $x_1 = +1, \quad x_2 = -1$

<u>WLSE?</u> & its variance?

$$V = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\left.\begin{matrix} w_1 = 1 \\ w_2 = 1/2 \end{matrix}\right\}$$

$$\beta^* = \frac{2Y_1 - Y_2}{3}$$

$$V(\beta^*) = \frac{2}{3}\sigma^2$$

Departure from Underlying Assumption :

Basic multiple linear regression model.

$\underline{Y} = X\underline{\beta} + \underline{\epsilon}$, X is $n \times p$ of $r(X) = p$

Assumptions: ϵ

(1) are unbiased.

(2) have constant variance.

(3) are uncorrelated.

(4) are normally distributed.

(1) $\Rightarrow$ $E(\underline{\epsilon}) = \underline{0}$ $\Rightarrow$ X is the correct design matrix.

i.e. $E(\underline{Y}) = X\underline{\beta}$

(2) & (3) $\Rightarrow$ $V(\underline{\epsilon}) = \sigma^2 I$.

(3) & (4) $\Rightarrow$ the independence of ϵ_i's

- Implicitly assumed that that regression variables x_j's are not random variable but are pre-determined constant.
- If x_j's are r.v.s and are measured without error, then the regression can be taken as conditional.

$$E(Y|X) = X\underline{\beta}$$

- If ϵ is a sum of n errors from different sources then as n increases, ϵ tends to normality irrespective of the prob. dist^n of the individual errors.

Underfitting

If $E(\underline{Y}) = X\underline{\beta}$, the LSE of $\underline{\beta}$ is $\hat{\beta} = (X'X)^{-1}X'\underline{Y}$

Suppose the model is underfitted.

True model: $\underline{Y} = X\underline{\beta} + Z\underline{\gamma} + \underline{\epsilon}$

where Z is $n \times t$ of rank t and columns of Z are linearly indpt. of the columns of X.

$Z\underline{\gamma}+\underline{\epsilon} \rightarrow$ error. $\rightarrow$ ϵ is biased.

$$E(\hat{\beta}) = (X'X)^{-1}X'(X\underline{\beta}+Z\underline{\gamma})$$

$$= \underline{\beta} + (X'X)^{-1}X'Z\underline{\gamma}$$

$$= \underline{\beta} + L\underline{\gamma}$$

So, $\hat{\beta}$ is a biased estimate of $\underline{\beta}$ with biased $L\underline{\gamma}$.

$L\underline{\gamma}$ depends on the postulated as well as the true model.

L: matrix of reg. coeff^t of the omitted variables regressed on the x-variable already included in the model.

- Even if the wrong model has seen postulated, then also bias can be zero.

e.g. if the columns of Z are orthogonal to the columns of X $\Rightarrow$ $X'Z=0$ & $L\underline{\gamma}=\underline{0}$

$\because X'Z=0$ (zero correlation between a pair of x & z variable).

Example: True model, $E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2$.

postulated model $E(\underline{Y}) = \beta_0 + \beta_1 x$

use, obs of $x_1 = -1$, $x_2 = 0$, $x_3 = 1$ to estimate β_0 & β_1

$$X = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

$$L\beta_2 = \begin{pmatrix} 2/3 \\ 0 \end{pmatrix} \beta_2$$

$\left\{ \begin{array}{l} \hat{\beta}_1 \text{ is unbiased.} \\ \hat{\beta}_0 \text{ is biased. with bias } 2/3\, \beta_2 \end{array} \right.$

$Var(\underline{\epsilon}) = \sigma^2 I_n$

$Var(\hat{\underline{\beta}}) = (X'X)^{-1}\sigma^2$ $\quad \hat{\underline{\beta}} = (\beta_0, \beta_1)'$

For the true model.

$$\hat{\beta}_G = \begin{pmatrix} \hat{\beta}_{0,G} \\ \hat{\beta}_{1,G} \end{pmatrix} \qquad \underline{\delta} = (\underline{\beta}', \underline{\gamma}')$$

$$Var(\hat{\underline{\beta}}_G) = \sigma^2 \left((X'X)^{-1} + LML' \right)$$

$$M = (Z'RZ)^{-1}$$

Now, σ^2 is estimated with under fitted model.

$$s^2 = \frac{\underline{Y}'(I-P)\underline{Y}}{n-p}, \quad P = X(X'X)^{-1}X'$$

$E(s^2)$

$$s^2 = \frac{\underline{y}'(I-P)\underline{y}}{(n-p)}$$

$$E(s^2) = tr\left((I-P)\sigma^2 I\right) \cdot \frac{1}{n-p} + \frac{\underline{\gamma}'Z'(I-P)Z\underline{\gamma}}{n-p}$$

$(PX=X)$

$$= \sigma^2 + \frac{\underline{\gamma}'Z'(I-P)Z\underline{\gamma}}{(n-p)}$$

$$> \sigma^2$$

s^2 overestimate σ^2.

Underfitting on fitted model?

$$\hat{Y} = P\underline{Y}$$

$$E(\hat{\underline{Y}}) = P(X\underline{\beta} + Z\underline{\gamma}) = X\underline{\beta} + PZ\underline{\gamma}$$

$\Rightarrow$ Ignoring $Z\underline{\gamma}$ in the fitting, it behaves like using $P\underline{Z}$ instead of Z.

Residual:

$$\underline{e} = \hat{\underline{\epsilon}} = \underline{Y} - X\hat{\underline{\beta}}$$

$$E(\hat{\underline{\epsilon}}) = E(\underline{Y}) - X E(\hat{\underline{\beta}}) = X\underline{\beta} + Z\underline{\gamma} - X\underline{\beta} - XL\underline{\gamma}$$

$$= Z\underline{\gamma} - X(X'X)^{-1}X'Z\underline{\gamma}$$

$$= (I-P)Z\underline{\gamma}$$

$$Var(\hat{\underline{\epsilon}}) = Var(\underline{Y} - X\hat{\underline{\beta}}) = (I-P)\sigma^2.$$

If we use of true model:

$$\sigma^2(I-P_W)$$

$$P_W = W(W'W)^{-1}W'$$

Effect of underfitting on Predictor :

To predict the value of Y at $\underline{\omega}_0 = (\underline{x}_0' ; \underline{z}_0')'$

although x_0 is only observed a used

Then the prediction using only $\underline{x}_0$

$$\hat{\underline{Y}}_0 = \underline{x}_0' \hat{\underline{\beta}}$$

Where the correct prediction

$$\hat{Y}_{0G} = \underline{\omega}_0' \hat{\delta}_G ,$$

$$\hat{\delta}_G = \begin{pmatrix} \hat{\beta}_G \\ \hat{\gamma}_G \end{pmatrix} \longrightarrow \text{LSE of } \delta .$$

$\underline{Y} = W\underline{\delta} + \underline{\epsilon}$

Where $\underline{Y} = \omega \underline{\delta} + \epsilon$

$$E(\hat{Y}_0) = \underline{x}_0' (\underline{\beta} + L\underline{\gamma})$$

$$Var(\hat{Y}_{0G}) = (\underline{x}_0', \underline{z}_0') \, Var(\hat{\delta}_G) \, (\underline{x}_0', \underline{z}_0')'$$

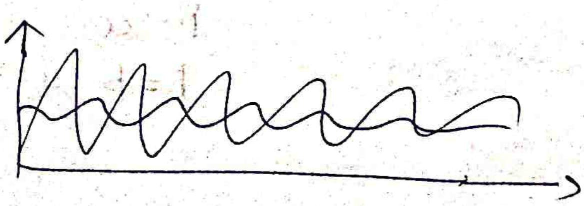
$$= \sigma^2 (\underline{x}_0', \underline{z}_0') (W'W)^{-1} (\underline{x}_0', \underline{z}_0')'$$

$$= \sigma^2 (\underline{x}_0', \underline{z}_0') \begin{pmatrix} (X'X)^{-1} + LML' & -LM \\ -ML' & M \end{pmatrix} (\underline{x}_0', \underline{z}_0')'$$

$$= \sigma^2 \underline{x}_0' (X'X)^{-1} \underline{x}_0 + \sigma^2 \underbrace{(L'x_0 - z_0)' M (L'x_0 - z_0)}_{\text{pd.}}$$

$$\Rightarrow\ \geq Var(\hat{Y}_0)$$

Since, $M = [Z'(I-P)Z]^{-1}$ is pd.



Overfitting :

Suppose the true model is $E(\underline{Y}) = X_1 \underline{\beta}_1$

Where X_1 consists of first K columns of X.

$$X = \left(\underset{n\times K}{X_1} , \underset{n\times(p-k)}{X_2} \right) \longrightarrow \underline{\beta}$$

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

$$E(\hat{\underline{\beta}}) = (X'X)^{-1} X' X_1 \beta_1 = (X'X)^{-1} X' [X_1, X_2] \begin{pmatrix} \beta_1 \\ \underline{0} \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \underline{0} \end{pmatrix}$$

So, $\hat{\underline{\beta}}_1$ (first K element of $\underline{\beta}$)

is an unbiased estimate of β_1

$$E(\hat{Y}) = E(X\hat{\beta}) = X \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix} = X_1 \underline{\beta}_1$$

→ fitted ~~model~~ value are unbiased of the model.

$Var(\hat{\underline{\beta}})$

→ $\sigma^2 (X'X)^{-1}$ → leads to inflated. expression for the variance of the elements of $\hat{\underline{\beta}}_1$

$Var(\hat{\underline{\beta}})$

$$= \left((X_1'X_1)^{-1} + LML' \right.$$

$X = X_1$; $\underline{\beta} = \beta_1$

$Z = X_2$; $\underline{\gamma} = \beta_2$

$p = \mu$

$t = p - \mu$

$$\underline{Y} = X_1 \underline{\beta}_1 + \underline{\epsilon} \qquad\qquad \underline{Y} = X\underline{\beta} + \underline{\epsilon}$$

$n \times n \qquad\qquad n \times p$

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

$$(X'X)^{-1} = \begin{pmatrix} (X_1'X_1)^{-1} + LML' & -LM \\ -ML' & M \end{pmatrix}$$

$$L = (X_1'X_1)^{-1}X_1'X_2, \qquad M = [X_2'(I-P)X_2]^{-1}$$

M and LML' are p.d.

$$\mathrm{Var}(\hat{\beta}_i) = \text{"true" } \mathrm{Var}(\hat{\beta}_i) + (LML')_{(i,i)}$$

$\hat{\beta}_i \to$ an element of $\hat{\beta}_1$.

$$\Rightarrow \quad > \text{"true" } \mathrm{Var}(\hat{\beta}_i)$$

$$(I-P)X \Rightarrow 0 \Rightarrow (I_n - P)(X_1 X_2) = 0$$

$$s^2 = \frac{Y'(I-P)Y}{(n-p)}$$

$$E[Y'(I-P)Y] = tr((I-P)\sigma^2 I) + \beta_1'X_1'(I_n-P)X_1\beta_1$$

$$= (n-p)\sigma^2 + 0$$

s^2 is still an unbiased estimator of σ^2.

Ex:

True model: $Y = \beta_0 + \beta_1 x + \epsilon$

Postulated model: $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$

Y at $x = -1$, $x_2 = 0$, $x_3 = 1$

$E(Y - X\hat{\beta}) = 0$

$Var(\underline{Y} - X\hat{\underline{\beta}}) =$

$\underline{x}_0' = (\underline{x}_{10}', x_{20}')$, $\underline{x}_0 \longrightarrow \overline{p\times 1}$

$K\times 1$

The prediction at $\underline{x}_0$ is $\hat{y}_0 = \hat{\underline{x}}_0' \hat{\underline{\beta}}$

(using overfitted model)

using $\underline{x}_{10}$, $\hat{y}_{10} = \underline{x}_{10}' \hat{\beta}_1$

Compare $\hat{y}_0$ & $\hat{y}_{10}$ w.r.t. there expected value and variance.

Incorrect Variance Matrix:

Usual assumptions : $\underline{y} = X\underline{\beta} + \underline{\epsilon}$

$E(\underline{\epsilon}) = \underline{0}$

$V(\underline{\epsilon}) = \sigma^2 I$

Instedd we assume,

$$Var(\underline{\epsilon}) = \sigma^2 V = \Sigma$$

If we use LSE, then $\hat{\beta} = (X'X)^{-1}X'\underline{y}$.

$E(\hat{\beta}) = \underline{\beta}$ is still unbiased.

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

$$\neq \sigma^2 (X'X)^{-1} \quad (\text{In general})$$

$$E(s^2) = \frac{E\left[\underline{y}'(I-P)\underline{y}\right]}{(n-p)} = \frac{\sigma^2 tr(V(I-P))}{(n-p)}$$

$$= \frac{1}{(n-p)}\left[\left[tr(I-P)\sigma^2 V\right] + \underbrace{\beta' X'(I-P)X\beta}_{=0}\right], \text{ as } PX = X$$

$$= \frac{\sigma^2}{(n-p)} tr(V(I-P))$$

& s^2 is generally a biased.

Q: If the first column of X is $\underset{\sim}{1}_n$ and $V = (1-\rho)I_n + \rho 1_n 1_n'$ $(0 \le \rho < 1)$

Then show that $E(s^2) = \sigma^2(1-\rho)$

$$V = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & & \ddots & & \vdots \\ \rho & & & \cdots & 1 \end{bmatrix}$$

$$E(s^2) = \frac{\sigma^2}{(n-p)} \, tr\left[V(I-P)\right]$$

$$V(I-P) = V - VP = V - \left[(1-\rho) I_n P + \rho \, 1_n 1_n' P\right]$$

$$= V - \left[(1-\rho) I_n P + \rho \, 1_n 1_n'\right]$$

$$tr\left[V(I-P)\right] = tr\left[V - (1-\rho) \cancel{I_n} P + \rho \, 1_n 1_n\right]$$

$$= tr(V) - (1-\rho)\, tr(P) - \rho \, tr(1_n 1_n')$$

$$= n - (1-\rho)\cdot p - \rho \cdot n$$

$$= n - p + p\rho - \rho n = (n-p)(1-\rho)$$

$$\Rightarrow E(s^2) = (1-\rho)\sigma^2$$

Outliers

Model: $\quad y_i = \underline{x}_i' \underline{\beta} + \epsilon_i \quad , i = 1, 2, \dots, n$

x_i' : ith row of the regression matrix X.

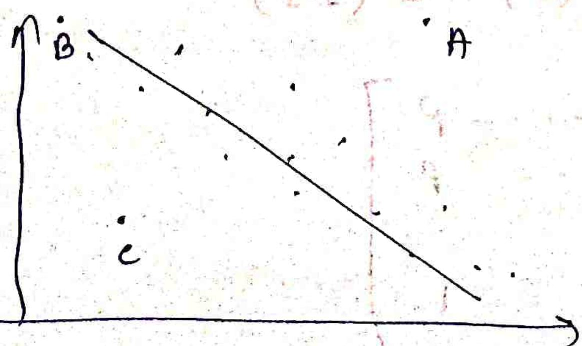
$(\underline{x}_i', y_i) \rightarrow$ ith data point in $(p+1)$ dim space.

Interested in two kinds of points.

1. those pts whose error ϵ_i is Large.
2. those pts whose x_i value is far from the bulk of the data.

① outlier, error outlier, outlier in y- direction regression outliers.

② extreeme pt, outlier in x-direction, Leverage pt, high leverage pt.



A & C $\rightarrow$ outlier (big vertical displacement from the true line).

B & C $\rightarrow$ high leverage pts.

If A were present and others were absent. we would expect A to have modest effect on the LS fit on the line to the data having ~~in the~~ it up.

The point B would have negligible effect on the fit

The point C which is both an outlier & large leverage pt will have considerable effect, much greater than that of A.

Pts like C are often referred as influential pts. Since they have big influence on the position of the fitted value.

• Fitted value : $\hat{\underline{Y}} = P\underline{Y}$ $\qquad P = ((p_{ij}))$

$$\hat{Y}_i = \sum_j p_{ij} Y_j = p_{ii} Y_i + \sum_{j \neq i} p_{ij} Y_j \qquad \text{(1)}$$

Since $P = P^2$

$$p_{ii} = \sum_{j=1}^{n} p_{ij}^2 = p_{ii}^2 + \sum_{j \neq i} p_{ij}^2 \qquad \text{(2)}$$

$\Rightarrow p_{ii} \geq p_{ii}^2 \Rightarrow p_{ii} \leq 1$

In terms of MD_i

$$p_{ii} = \frac{1}{n} + \frac{1}{(n-1)} MD_i \qquad \text{[with influential model]}$$

$$\geq 1/n$$

So, $\frac{1}{n} \leq p_{ii} \leq 1$

$$MD_i = (x_i - \bar{x})' S_{xx}^{-1} (\underline{x}_i - \underline{\bar{x}})$$

↳ ith reduced row

$$\hat{Y}_i \approx \underbrace{p_{ii}}_{\simeq 1} Y_i \approx Y_i$$

$(\underline{x}_i', Y_i)$ → outlier and highly leverage pt.

From ②, where p_{ii} is close to 1, $p_{ij}; j \neq i$ will be close to zero and then from ① this means that $\hat{Y}_i$ will be determined largely by the value of Y_i. If ith data point is both an outlier and a high leverage pt, MD_i will be large, p_{ii} will be close to 1 & $\hat{Y}_i$ will be affected by Y_i to a great extent.

A point like c can here serious effect on the fitted line.

Suppose that the pt is an outlier, so write the model $Y_i = \underline{x}_i' \underline{\beta} + \Delta_i + \epsilon_i$

Δ_i is a +ve constant.

Let, $\underline{\Delta} = (0, --, 0, \Delta_i, 0, --, 0)'$

Then $\underline{Y} = X\underline{\beta} + \underline{\Delta} + \underline{\epsilon}$

$E(\hat{\underline{y}}) = E[P\underline{Y}] = P[X\underline{\beta} + \underline{\Delta}] = X\underline{\beta} + P\underline{\Delta}$

$E(\hat{Y}_i) = \underline{x}_i' \underline{\beta} + p_{ii} \Delta_i$

Robustness of F-test to Non-normality

- The sensitivity of the F test to normality depends on the numerical value of the reg. variables

- Let, $y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_{p-1} x_{ip-1} + \epsilon_i$

$H_0: \beta_0 = \beta_1 = \cdots = \beta_{p-1} = 0$

When H is true and the regression assumptions are valid. then

$$F = \frac{RSS_H - RSS}{RSS} \cdot \frac{n-p}{p-1} \sim F_{p-1, n-p}$$

(under normality)

If we relax the dist^n assumption

- Assumption - ϵ_i are independently distributed with some common but not necessarily normal dist^n,

then $F \sim F_{\nu_1, \nu_2}$ under H

Where $\nu_1 = \delta(p-1)$

$\nu_2 = \delta(n-p)$

$$\delta^{-1} = 1 + \frac{(n+1)\alpha_2}{(n-1) - 2\alpha_2}, \quad \alpha_2 = \frac{n-3}{2n(n-1)} C_x \sqrt{\nu}$$

$$\sqrt{\nu} = E\left[\frac{K_4}{K_2^2}\right], \quad K_4, K_2 \to \text{sample cumulants for } n \text{ values of } y$$

C_x - multivariate analog of $\frac{K_4}{K_2^2}$ of x var.

[**Cumulant Generating F^n**

$$K(t) = \log E(e^{tx})$$

$$= \sum_{n=1}^{\infty} K_n \frac{t^n}{n!}$$

$K_n = K^{(n)}(0) \longrightarrow$ nth order cumulant] where $\underline{\epsilon}$ and $\underline{y}$ are normally distributed, $\sqrt{\beta_1}=0$, $\delta=1$

When exp variable can be regarded as approximately normal, $C_x \approx 0$ and F test is insensitive to non-normality.

Quadratically Balanced F-tests

Let, $Y_1, Y_2, \ldots, Y_n$ be independent r.v.'s with means $\theta_1, \theta_2, \ldots, \theta_n$ respectively common variance σ^2 and common third and fourth moments (μ_3 & μ_4) about their mean.

Let, $\gamma_2 = \dfrac{\mu_4 - 3\sigma^4}{\sigma^4} \longrightarrow$ Common Kurtosis.

Theorem:

Let, P_1 & P_2 be symmetric idempotent matrix of rank f_1 and f_2, $\ni$

$E(\underline{y}'P\underline{y}) = \sigma^2 f_i$ and $P_1P_2 = 0$. If $\underline{p}_i$ is the column rank of the diagonal entries of P_i, then

ⓐ $Var(\underline{y}'P_i\underline{y}) = 2\sigma^4\left(f_i + \frac{1}{2}\gamma_2 \underline{p}_i'\underline{p}_i\right)$

ⓑ $Cov(\underline{y}'P_1\underline{y}, \underline{y}'P_2\underline{y}) = \sigma^4 \gamma_2 \underline{p}_1'\underline{p}_2$

Proof:

P_i is symmetric and identical

$tr(P_i) = rank(P_i) = f_i$

$E(\underline{y}'P_i\underline{y}) = tr(\sigma^2 P_i) + \underline{\theta}'P_i\underline{\theta}$

$E(\underline{Y}) = \underline{\theta}$

$= \sigma^2 f_i$ [statement of the thm]

$$\Rightarrow \underline{\theta}' P_i \underline{\theta} = \underline{\theta}' P_i^2 \underline{\theta} = 0$$

$$\Rightarrow P_i \underline{\theta} = 0 \quad \forall \underline{\theta}$$

[Th^m: $\quad Var(x'Ax) = (\mu_4 - 3\mu_2^2)\, \underline{a}'\underline{a} + 2\mu_2^2 \, tr(A^2)$

$\qquad + 4\mu_2 \, \theta' A \underline{\theta} + 4\mu_3 + \theta' A \underline{a}$

$x_1, x_2, \ldots, x_n \;,\; \theta_1, \ldots, \theta_n \;,\; \mu_2, \mu_3\, \mu_4$

$A_{n \times n}$]

Then replacing A by P_i

$$Var(\underline{Y}' P_i \underline{Y}) = (\mu_4 - 3\sigma^4)\, p_i' p_i + 2\sigma^4 \, tr(P_i^2)$$

$$= 2\sigma^4 \left[f_i + \frac{1}{2} \gamma_2 \underline{p}_i' \underline{p}_i \right] \qquad P_i \underline{\theta} = \underline{0}$$

(b) $P_1 P_2 = 0$ (Given)

$(P_1 + P_2)^2 = P_1 + P_2 \longrightarrow$ idempotent.

$$Var(\underline{Y}' P_1 \underline{Y} + \underline{Y}' P_2 \underline{Y}) = Var\,(\underline{Y}'(P_1 + P_2)\underline{Y}]$$

$$= 2\sigma^4 \left[tr(P_1 + P_2) + \frac{1}{2} \gamma_2 (\underline{p}_1 + \underline{p}_2)' (p_1 + p_2) \right]$$

$$= 2\sigma^4 \left[f_1 + f_2 + \frac{1}{2} \gamma_2 \{ \underline{p}_1' p_1 + 2 \underline{p}_1' \underline{p}_2 + \underline{p}_2' \underline{p}_2 \} \right]$$

$$= Var(\underline{Y}' P_1 \underline{Y}) + Var(\underline{Y}' P_2 \underline{Y}) + 2\sigma^4 \gamma_2 \underline{p}_1' \underline{p}_2$$

$$Cov(\underline{Y}' P_1 \underline{Y}, \underline{Y}' P_2 \underline{Y}) = \sigma^4 \gamma_2 \, P_1' P_2$$

Assignment-1

$$\underline{x}' L M L' \underline{x} = \underline{x}'(X'X)^{-1} \underbrace{X'Z \, M \, Z'X}\, (X'X)^{-1} \underline{x}$$

$$Z' X (X'X)^{-1} \underline{x} = 0 \quad \text{for some } \underline{x} \neq \underline{0}$$

(6)
(7)

$$\|\underline{y} - \underline{\hat{y}}_H\|^2 - \|Y - \hat{Y}\|^2 = \|\hat{Y} - \hat{Y}_H\|^2$$

$$\hat{y} - \hat{y}_H = X\hat{\beta} - X\hat{\beta}_H$$

$$= X(X'X)^{-1}X'\underline{y} - X\left[\hat{\beta} - (X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}(A\hat{\beta} - \underline{c})\right]$$

$$= X(X'X)^{-1}X'\underline{y} - X\left[\hat{\beta} - \frac{1}{2}(X'X)^{-1}A'\underline{\hat{\lambda}}_H\right]$$

$$= \frac{1}{2} X(X'X)^{-1}A'\underline{\hat{\lambda}}_H$$

$$\|\hat{y} - \hat{y}_H\|^2 = \hat{\lambda}_H' A (X'X)^{-1} X'X (X'X)^{-1} A' \hat{\lambda}_H \; \frac{1}{4}$$

$$= \underline{\hat{\lambda}}_H' A (X'X)^{-1} A' \hat{\lambda}_H \cdot \frac{1}{4}$$

$$-\frac{1}{2}\hat{\lambda}_H = \left(A(X'X)^{-1}A'\right)^{-1}\left(\underline{c} - A\hat{\beta}\right)$$

$$\mathrm{var}\left(\hat{\lambda}_H \cdot \frac{1}{2}\right) = \left(A(X'X)^{-1}A'\right)^{-1} A \cdot \mathrm{var}(\hat{\beta}) \cdot A' \left[A(X'X)^{-1}A'\right]^{-1}$$

$$= \sigma^2 \left[A(X'X)^{-1}A'\right]^{-1}$$

Theorem:

Suppose that p_1 & p_2 satisfy the conditions of the ~~text~~ last theorem. and let

$$z = \frac{1}{2}\log F$$

Where, $F = \dfrac{Y'P_1Y/f_1}{Y'P_2Y/f_2} = \left(\dfrac{s_1^2}{s_2^2}, \text{ say}\right)$.

Then for large f_1 & f_2, we have

$$E(z) \sim \frac{1}{2}[f_2^{-1} - f_1^{-1}]\Big[1+\frac{1}{2}\gamma_2 (f_1p_2 - f_2p_1)' (f_1p_2 + f_2p_1)\{f_1f_2(f_1-f_2)\}^{-1}\Big]$$

$$Var(z) \sim \frac{1}{2}(f_1^{-1}+f_2^{-1})\Big[1+\frac{1}{2}\gamma_2(f_1p_2 - f_2p_1)'(f_1p_2 - f_2p_1)\{f_1f_2(f_1+f_2)\}^{-1}\Big]$$

Proof: Using Taylor series expansion of $\log s_i^2$ about $\log \sigma^2$.

$$\log s_i^2 \sim \log\sigma^2 + \frac{s_i^2 - \sigma^2}{\sigma^2} - \frac{(s_i^2-\sigma^2)^2}{2\sigma^4}$$

Take expectation & we $E(s_i^2) = \sigma^2$

$$E(\log s_i^2) \sim \log\sigma^2 - \frac{1}{2\sigma^4} V(s_i^2 - \sigma^2)^2$$

$$= \log\sigma^2 - \frac{1}{2\sigma^4}V(s_i^2)$$

$$\cancel{\log\sigma^2 \frac{1}{2\sigma^4}}$$

$$V(s_i^2) = \frac{Var(Y'P_iY)}{f_i^2} = 2\sigma^4\left(f_i^{-1} + \frac{1}{2}\gamma_2 f_i^{-2} p_i'p_i\right).$$

$$= 2\sigma^4\left[f_i^{-1} + \frac{1}{2}\gamma_2 f_i^{-2}p_i'p_i\right]$$

$$E(Z) = \frac{1}{2}\left[E(\log s_1^2) - E(\log s_2^2)\right]$$

$$\sim \frac{1}{2}\left[\log\sigma^2 - \frac{1}{2\sigma^4}V(s_1^2) - \log\sigma^2 + \frac{1}{2\sigma^4}V(s_2^2)\right]$$

$$= \frac{1}{2}\left[-f_1^{-1} - \frac{1}{2}\gamma_2 f_1^{-2}\underline{p}_1'\underline{p}_1 + \frac{1}{2}\gamma_2 f_2^{-2}\underline{p}_2'\underline{p}_2 + f_2^{-1}\right]$$

$$= \frac{1}{2}\left[f_2^{-1} - f_1^{-1} + \frac{1}{2}\gamma_2 f_1^{-2}f_2^{-2}\left(-f_2^2\,\underline{p}_1'\underline{p}_1 + f_1^2\,\underline{p}_2'\underline{p}_2\right)\right]$$

$$= \frac{1}{2}\left[(f_2^{-1} - f_1^{-1}) + \frac{1}{2}\gamma_2 f_1^{-2}f_2^{-2}\left[-f_2^2\,\underline{p}_1'\underline{p}_1 - f_1f_2\,\underline{p}_1'\underline{p}_2 + f_1f_2\,\underline{p}_1'\underline{p}_1 + f_1^2\,\underline{p}_2'\underline{p}_2\right]\right]$$

$$= \frac{1}{2}\left[(f_2^{-1} - f_1^{-1})\left[1 + \frac{1}{2}\gamma_2\left(f_1\underline{p}_2 - f_2\underline{p}_1\right)'\left(f_1\underline{p}_2 + f_2\underline{p}_1\right) \times c\right]\right]$$

$$c = f_1^{-2}f_2^{-2}(f_2^{-1} - f_1^{-1})^{-1}$$

$$= \left\{f_1f_2(f_1 - f_2)\right\}^{-1}$$

$$Var(Z) = Var\left[\frac{1}{2}\log\frac{s_1^2}{s_2^2}\right]$$

$$= \frac{1}{4}\left[Var(\log s_1^2) + Var(\log s_2^2) - 2\,cov(\log s_1^2, \log s_2^2)\right]$$

Ignoring the third term in $\log s_i^2$

$$E(\log s_i^2) \sim \log\sigma^2$$

$$Var(\log s_i^2) \approx E\left[\log s_i^2 - \log\sigma^2\right]^2$$

$$\sim \frac{E\left[(s_i^2 - \sigma^2)^2\right]}{\sigma^4} = \frac{1}{\sigma^4}V(s_i^2)$$

$$\operatorname{Cov}(\log s_1^2, \log s_2^2) = E\left[(\log s_1^2 - \log \sigma^2)(\log s_2^2 - \log \sigma^2)\right]$$

$$\approx \frac{E\left[(s_1^2 - \sigma^2)(s_2^2 - \sigma^2)\right]}{\sigma^4}$$

$$= \frac{\operatorname{Cov}(s_1^2, s_2^2)}{\sigma^4}$$

$$\operatorname{Var}(Z) \sim \frac{1}{4\sigma^4}\left[V(s_1^2) + V(s_2^2) - 2\operatorname{cov}(s_1^2, s_2^2)\right]$$

F test for testing $A\beta = 0$

$$F = \frac{\underline{Y}'(P - P_H)\underline{Y}/q}{\underline{Y}'(I-P)\underline{Y}/(n-p)} = \frac{\underline{Y}'P_1\underline{Y}/q}{\underline{Y}'P_2\underline{Y}/(n-p)} = \frac{S_1^2}{S_2^2}$$

$P_1P_2 = (P - P_H)(I-P) = 0$

Let us relax the the distributed assumption. only assume that ϵ_i's are iid with $E(\underline{\epsilon}) = \underline{0}$, $V(\underline{\epsilon}) = \sigma^2 I_n$

Then $E(S_2^2) = \sigma^2$

When H is true, $E(S_1^2) = \sigma^2 \quad (A\hat{\beta} = \underline{c} = 0)$

$$E(RSS_H - RSS) = \sigma^2 q + (A\beta - c)'[A(X'X)^{-1}A']^{-1}(A\underline{\beta} - \underline{c})$$

(Normality assumption is not req. ~~for~~ to prove this)

- When H is true, we can use last theorem with $f_1 = q$ & $f_2 = n-p$

When ϵ_i & so Y_i are normally distributed then for large f_1 & f_2,

$Z = \frac{1}{2}\log F$ is approximately ~~as~~ normally distributed with mean & variance given by $E(Z)$ & $V(Z)$ with $\gamma_2 = 0$.

This approximation is quite good for small f_1 & f_2 (as small as 4), we consider the following proposition :-

For a moderate amount of non-normality Z is still approximately normal with mean and variance given by $E(Z)$ & $V(Z)$

On this assumption Z & therefore F will approximately independent of γ_2 if the coeff of γ_2 is zero.

i.e. $$f_1 \underline{P}_2 = f_2 \underline{P}_1$$

F is quadratically balanced if the diagonal elements P_i $(i=1,2)$ are equal. $(\underline{P}_1 = \underline{P}_2)$.

Since $tr(P_i) = f_i$, So $\underline{P}_i = f_i/n \, \underline{1}_n$

$$f_1 \underline{P}_2 = \frac{f_1 f_2}{n} \underline{1}_n = f_2 \underline{P}_1$$

So, a sufficient condition for $(f_1 \underline{P}_2 = f_2 \underline{P}_1)$ to hold is that F is quadratically balanced.

Effect of Random Explanatory Variable:

① Random exp var measured without error

Suppose we have the following the relationship between V & $U_1, \ldots U_{p-1}$.

$$V = \beta_0 + \beta_1 U_1 + \cdots + \beta_{p-1} U_{p-1}$$

→ structural relationship with U_j being observed exactly but V unknown.

$Y = V + \epsilon$ is actually observed.

Appropriate model

$$Y = \beta_0 + \beta_1 U_1 + \cdots + \beta_{p-1} U_{p-1} + \epsilon$$

as $E[Y \mid U_1, \ldots U_{p-1}] = \beta_0 + \beta_1 U_1 + \cdots + \beta_{p-1} U_{p-1}$

Fitting carry out a standard regression analysis conditionally on the values observed for the exp. variable.

We proceed as though the exp variable were fixed.

We require the usual assumption of normality, constant var & independent to hold conditionally on the V_j's

The problem of bias etc due to model misspecification will be same as for fixed exp var.

A different approach:

Suppose that the true model is

$$Y = \beta_0 + \beta_1 U_1 + \cdots + \beta_s U_s$$
$$= \beta_0 + \beta_1 U_1 + \cdots + \beta_r U_r + \delta \quad , \quad r < s$$

where U_j, $j = 1, 2, \ldots s$ are r.v. with $E(U_j) = \theta_j$

The "error" f is assumed to be due to further hidden variables $U_{r+1}, \ldots U_s$

$$Y = \beta_0 + \beta_{r+1}\theta_{r+1} + \cdots + \beta_s\theta_s + \beta_1 U_1 + \beta_r U_r + \sum_{j=r+1}^{s} \beta_j (U_j - \theta_j)$$

$$= \alpha_0 + \beta_1 U_1 + \cdots + \beta_r U_r + \epsilon$$

$E(\epsilon) = 0$

$\hookrightarrow$ the same model by ⊛.

Since r is arbitrary $E(\epsilon) = 0$, will be ~~or~~ a satisfied always irrespective of the no of regression varb. we include in the model.

② ~~Errors~~ Exp. variable measured with error

Suppose that the linear relationship is now between the expected values

$$v = E(y) = \beta_0 + \beta_1 E(X_1) + \cdots + \beta_{p-1} E(X_{p-1})$$

$$= \beta_0 + \beta_1 U_1 + \cdots + \beta_{p-1} U_{p-1}$$

$\hookrightarrow$ functional relationship

Here v & $u_1, \ldots, u_{p-1}$ are unknown & all are measured with error.

$\longrightarrow$ errors-in-variable model.

The approximate model

$$Y = \beta_0 + \beta_1 E(X_1) + \cdots + \beta_{p-1} E(X_{p-1}) + \epsilon$$

Where ϵ is assumed to be ~~incorporated~~ independent of $X_1, \dots, X_{p-1}$. Suppose n obs on the model

$$y_i = \beta_0 + \beta_1 u_{i1} + \dots + \beta_{p-1} u_{ip-1} + \epsilon_i$$

$$= \underline{x}_i' \underline{\beta} + \epsilon_i$$

$$\underline{y} = U\underline{\beta} + \epsilon_i ; \quad U = (\underline{u}_1, \dots \underline{u}_n)'$$

Suppose that the data pt $\underline{x}_i$ is measured with an unbiased error $\underline{\delta}_i$.

So, we actually observe,

$$\underline{x}_i = \underline{u}_i + \underline{\delta}_i$$

$$X = U + \Delta \quad ; \quad X = (\underline{x}_1, \dots, \underline{x}_n)'$$

$$\Delta = (\underline{\delta}_1, \underline{\delta}_2, \dots \underline{\delta}_n)' \quad ; \quad E(\Delta) = 0.$$

Assume that $\underline{\delta}_i$ are uncorrelated & have the same variance matrix.

$$E(\underline{\delta}_i, \underline{\delta}_j') = \begin{cases} D & , i = j \\ 0 & , i \neq j \end{cases}$$

Since the first element of each $\underline{u}_i$ & $\underline{x}_i$ is one. the 1st ~~column~~ element of each $\underline{\delta}_i$ is zero and so the 1st row & first column D consists of zeros.

Assume that Δ is independent of $\underline{\epsilon}$.

The LSE of $\underline{\beta}$ is $\hat{\underline{\beta}}_\Delta = (X'X)^{-1} X' \underline{y}$ → no longer unbiased.

insted of $\hat{\underline{\beta}} = (U'U)^{-1} U' \underline{y}$

since Δ is independent

$$E[\hat{\beta}_\Delta] = E_\Delta [E_Y (\hat{\beta}_\Delta \mid \Delta)]$$

$$E\left[\hat{\beta}_\Delta\right] = E_\Delta\left[E_Y(\hat{\beta}_\Delta \mid \Delta)\right]$$

$$= E_\Delta\left[(X'X)^{-1}X'U\underline{\beta}\right]$$

$$= E_\Delta\left[(X'X)^{-1}X'(X-\Delta)\underline{\beta}\right].$$

$$= \underline{\beta} - E_\Delta\left[(X'X)^{-1}X'\Delta\underline{\beta}\right] = \underline{\beta} - \underline{b}, \text{ say}$$

$$E(X'X) = E\left[U'U + \Delta'U + U'\Delta + \Delta'\Delta\right]$$

$$= U'U + E\left[\Delta'\Delta\right]$$

$$= U'U + E\left[\sum_{i=1}^{n} \delta_i\delta_i'\right]$$

$$= U'U + nD$$

$$E\left[X'\Delta\beta\right] = E\left[(U+\Delta)\Delta\right] = nD.$$

[Generate ϵ_i from $U(0, 0.1)$ ...
Generate data] ... Covariate: Calculate.
Residuals:

When n is large

$$\underline{b} \approx \left(\frac{1}{n} U'U + nD\right)^{-1} D\underline{\beta}$$

⊛ If $\lim_{n \to \infty} \left\{ \frac{1}{n} (X'X) \right\} = A$,

then $\hat{\beta}_{\Delta}$ is a ~~constant~~ consistent estimate of $(A+D)^{-1} A\underline{\beta}$

$$(A+D)^{-1} A\underline{\beta} = (A+D)^{-1} (A+D-D)\, \beta$$

$$= \beta - (A+D)^{-1} D \underline{\beta}$$

$$\hat{\underline{b}} \approx n\, (X'X)^{-1} \hat{D} \hat{\underline{\beta}}$$

$\hat{D}$ is a rough estimate of D which is available from other /previous experiment.

Example: $Y_i = \beta_0 + \beta_1 u_i + \epsilon_i$

and $X_i = u_i + \delta_i$ $\quad i = 1,2,\dots,n$

~~The pairs (δ_i, ϵ_i)~~

We assume that δ_i & ϵ_i are independently distributed with zero means and unknown variances.

σ_δ^2 & σ_ϵ^2

n no of observations (Y_i, X_i) $i = 1,2,\dots n$, but $(n+4)$ unknowns.

Apply $\underset{\sim}{b} \approx n(U'U + nD)^{-1} D \underline{\beta}$, $D = \text{diag}\{0, \sigma_\delta^2\} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_\delta^2 \end{bmatrix}$

$$E[\hat{\beta}_{0\Delta}] \approx \beta_0 + \frac{\bar{u}\beta_1 n \sigma_\delta^2}{\sum_i (u_i - \bar{u})^2 + n\sigma_\delta^2}$$

$$E[\hat{\beta}_{1\Delta}] \approx \beta_1 - \frac{\beta_1 n \sigma_\delta^2}{\sum (u_i - \bar{u})^2 + n\sigma_\delta^2}$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n \left(u_i + \delta_i - (\bar{u} + \bar{\delta})\right)^2$$

$$= \sum (u_i - \bar{u})^2 - 2\sum (u_i - \bar{u})(\delta_i - \bar{\delta}) + \sum (\delta_i - \bar{\delta})^2$$

$$E\left[\sum_i (x_i - \bar{x})^2\right] = \sum (u_i - \bar{u})^2 - 0 + (n-1)\sigma_\delta^2$$

So the relative bias in $\hat{\beta}_{1\Delta}$ is approximately

$$\frac{-n\sigma_\delta^2}{E\left[\sum (x_i - \bar{x})^2\right]}$$

↳ will generally be small if

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

When D is close to zero matrix, then

$$Var(\hat{\beta}_\Delta) \approx \frac{1}{n}\left\{\left(\frac{1}{n}U'U+D\right)^{-1}(\sigma^2+\underline{\beta}'\Delta\underline{\beta})\right\}$$

usual estimate of this is $\hat{V} = s^2(X'X)^{-1}$

where $(n-p)s^2 = (\underline{Y}-X\hat{\underline{\beta}}_\Delta)'(\underline{Y}-X\hat{\beta}_\Delta)$

$$= \underline{Y}'(I-P_X)\underline{Y}$$

If $\hat{V}$ an unbiased estimate of $Var(\hat{\beta}_\Delta)$?

$$E[(n-p)s^2 \mid \Delta]$$

$$= E[(n-p)\sigma^2 + \beta'U'(I_n-P_X)U\beta \mid \Delta]$$

$$= (n-p)\sigma^2 + \beta'(X'-\Delta')(I-P_X)(X-\Delta)\underline{\beta}$$

$$= (n-p)\sigma^2 + \beta'\Delta'(I_n-P_X)\Delta\underline{\beta}$$

Now for any matrix C,

$$E[\Delta'C\Delta] = \sum_i\sum_j C_{ij}E(\underline{\delta}_i\underline{\delta}_j')$$

$$= \sum C_{ii}D = D\,tr(C)$$

$$\left[\text{As } E(\delta_i\delta_j') = \begin{cases} D & \forall i=j \\ 0 & o/w \end{cases}\right.$$

$$E(\hat{V}) = E_\Delta\left[E(s^2(X'X)^{-1} \mid \Delta)\right]$$

$$\approx E_\Delta\left[\left\{\sigma^2 + \frac{1}{(n-p)}\underline{\beta}'\Delta'(I_n-P_X)\Delta\underline{\beta}\right\}(X'X)^{-1}\right]$$

$$\approx \left\{\sigma^2 + \frac{1}{(n-p)}\beta' E[\Delta'(I_n-P_X)\Delta]\underline{\beta}\right\}\{E(X'X)\}^{-1}$$

$$\approx \{\sigma^2 + \underline{\beta}'D\underline{\beta}\}(U'U+nD)^{-1}$$

$\approx \operatorname{Var}(\hat{\beta}_\Delta)$

Hence for large n and ~~sma~~ small D, $\hat{V}$ is still approximately unbiased.

Round-off Errors

we suppose that U is the correct regression matrix. $(E(Y) = U(\beta))$ & we observe $X = U + \beta\Delta$

we assume that the measurements are ~~as~~ accurate but are rounded off acc to some ~~constant~~ consistent rule.

$$x_{ij} = u_{ij} + \Delta_{ij}$$

The rounded off error Δ_{ij} can be considered as constant (unknown) & not a random variable.

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

The bias of $\hat{\beta}_\Delta$ is $E[\hat{\beta}_\Delta - \beta] = -(X'X)^{-1}X'\Delta\beta$

writing, $\Delta\beta = \sum_{j=0}^{p-1} \Delta_j \beta_j$ where Δ_j is the jth column of Δ.

So bias does not depend on β_j if $\Delta_j = 0$

$$E(s^2) = \sigma^2 + \frac{\beta'\Delta' U(I-P_X) U\beta}{n-p}$$

$$= \sigma^2 + \frac{\beta'\Delta'(I-P_X)\Delta\beta}{n-p}$$

$$\geq \sigma^2,$$

s^2 tends to overestimate σ^2.

Random explanatory variable measured with error :

Suppose we have the structural relationship as

$$Y = \beta_0 + \beta_1 U_1 + \cdots + \beta_{p-1} U_{p-1} + \epsilon$$

but the random experiment V_j are not measured with error (unbiased) So that we observe

$$X_j = U_j + \gamma_j , \quad E[\gamma_j | U_j] = 0$$

$$E[Y | U_1, \ldots, U_{p-1}] = \beta_0 + \beta_1 U_1 + \cdots + \beta_{p-1} U_{p-1}$$

$$= \beta_0 + \beta_1 E(X_1 | U_1) + \cdots + \beta_{p-1} E[X_{p-1} | U_{p-1}]$$

Collinearity :

One important assumption is that the regression matrix X is assumed to be of full rank.

In practice, the column of X could be almost linearly dependent or collinear.

This leads to $X'X$ being close to singular.

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

near collinearity will have a considerable effect on the precision of $\hat{\beta}$.

When the estimated reg. coefficients have large variances test will have low power & Confidence interval will be very large.

→ It will be difficult to decide if a variable make a significant contribution to the regression.

Straight line Regression :

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad , \quad i = 1, 2, \ldots, n .$$

The variances of the LSEs of $\hat{\beta}_0$ & $\hat{\beta}_1$.

$$V(\hat{\beta}_0) = \frac{\sigma^2 \sum x_i^2}{n S_{xx}} \quad , \quad S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} .$$

The coefficient of variation $CV_x = \frac{(S_{xx}/n)^{1/2}}{\bar{x}}$

- measures with variability of the x's relation to their average size. / mean.

- is independent of the unit used to measure x.

$$Var(\hat{\beta}_0) = \sigma^2 \left\{ S_{xx} + n\bar{x}^2 \right\}$$

$$= \frac{\sigma^2}{n} \left\{ 1 + \frac{1}{CV_x^2} \right\}$$

If CV_x is small , $Var(\hat{\beta}_0)$ will be Large.

If we consider the centered model

$$y_i = \alpha_0 + \beta(x_i - \bar{x}) + \epsilon_i \quad , \quad i = 1, 2, \ldots n$$

$\alpha_0 = \beta_0 + \beta_1 \bar{x}$, $\hat{\alpha}_0 = \bar{Y}$ with variance σ^2/n .

It is a common practice to center and scale the exp. variable.

$$x_i^* = (x_i - \bar{x}) / \sqrt{S_{xx}}$$

$$\sum_{i=1}^{n} x_i^* = 0 \quad ; \quad \sum x_i^{*2} = 1$$

→ a scale invariant exp variable which has no unit.

$$y_i = \alpha_0 + \gamma x_i^* + \epsilon_i \quad , \quad i = 1, 2, \dots, n.$$

$$\hat{\gamma} = \frac{\sum x_i^* (y_i - \bar{y})}{\sum x_i^{*2}} = \sum x_i^* y_i$$

$$Var(\hat{\gamma}) = Var\left(\sum x_i^* y_i\right) = \sum x_i^{*2} V(y_i)$$

$$= \sigma^2 . 1 = \sigma^2 .$$

<u>Two explanatory variable</u>

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i , \; i = 1, 2, \dots, n$$

$$\underline{Y} = X \underline{\beta} + \underline{\epsilon}$$

$$\underline{\beta} = (\beta_0, \beta_1, \beta_2)' \qquad X = (1_n, \underline{x}, \underline{z})$$

Let, $\underset{\sim}{w} = (\bar{x}, \bar{z})'$ $\quad \tilde{X} = (\tilde{\underline{x}}, \tilde{\underline{z}})$

$$\underset{\sim}{\tilde{x}} = \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} , \quad \underset{\sim}{\tilde{z}} = \begin{pmatrix} z_1 - \bar{z} \\ \vdots \\ z_n - \bar{z} \end{pmatrix}$$

$$(X'X) = \begin{pmatrix} n & n\bar{x} & n\bar{z} \\ n\bar{x} & \sum x_i^2 & \sum x_i z_i \\ n\bar{z} & \sum x_i z_i & \sum z_i^2 \end{pmatrix}$$

$$= n \begin{pmatrix} 1 & \underline{w}' \\ \underline{w} & S + \underline{w}\underline{w}' \end{pmatrix}$$

$$S = \tfrac{1}{n} \tilde{X}' \tilde{X}$$

$$(X'X)^{-1} = n^{-1}\begin{pmatrix} 1+\underline{\omega}'S^{-1}\underline{\omega} & -\underline{\omega}'S^{-1} \\ -S^{-1}\underline{\omega} & S^{-1} \end{pmatrix}$$

$$\mathrm{Var}[\hat{\beta}_0] = \sigma^2(1+\underline{\omega}'S^{-1}\underline{\omega})/n$$

$$\mathrm{Var}(\hat{\beta}_1) = \cancel{\sigma^2/\{S_{xx}(1-r^2)\}} = (S^{-1})_{11} = \frac{\sigma^2}{\{S_{xx}(1-r^2)\}}$$

$$\mathrm{Var}(\hat{\beta}_2) = \cancel{\sigma^2/S_{xx}(1-r^2)} = (S^{-1})_{(2,2)}$$

$$S = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix} \qquad = \sigma^2/\{S_{zz}(1-r^2)\}$$

$S_{zz} = \Sigma(z_i-\bar{z})^2$ & $r =$ correlation coeff between x & z

$$\mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2}{n}\left[1+\frac{1}{1-r^2}\left\{\frac{1}{CV_x^2}-\frac{2r}{CV_xV_y}+\frac{1}{CV_y^2}\right\}\right]$$

<u>Centered & scaled Model</u>

$$y_i = \alpha_0 + \gamma_1 x_i^* + \gamma_2 z_i^* + \epsilon_i, \quad i=1,2,\dots,n$$

$$x_i^* = \frac{x_i-\bar{x}}{S_{xx}^{1/2}}, \quad z_i^* = \frac{z_i-\bar{z}}{S_{zz}^{1/2}}$$

Write, $X_s = (\underset{\sim}{1_n}, \underline{x}^*, \underline{z}^*)$

$$X_s'X_s = \begin{pmatrix} n & \underline{0}' \\ \underline{0} & R_{xx} \end{pmatrix} \qquad R_{xx} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$$

$$\mathrm{Var}(\hat{\alpha}_0) = \sigma^2/n$$

$$\mathrm{Var}(\hat{\gamma}_1) = \sigma^2/(1-r^2) = \mathrm{var}[\hat{\gamma}_2]$$

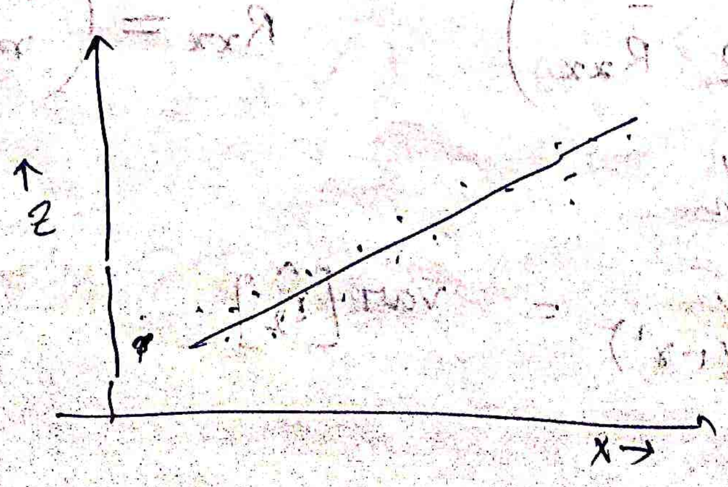
$$y_i = \alpha_0 + \gamma_1 x_i^* + \gamma_2 z_i^* + \epsilon_i \quad , \quad i = 1, 2, \ldots, n$$

$$Var(\hat{\alpha}_0) = \sigma^2/n \; ; \; Var(\hat{\gamma}_1) = \frac{\sigma^2}{1-r^2}$$

$$Var(\hat{\gamma}_2) = \frac{\sigma^2}{(1-r^2)}$$

$$r = cov(x, z)$$

- The accuracy of the LSE of scaled reg coeffs depends only on the error variance and the corrln between x & z.
- The scaled reg. coeff. cannot be estimated accurately if the corrln is ~~cols~~ close to 1.
- Then the exp variable cluster about cluster about St line.



General Case:

Multiple Regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{ip-1} + \epsilon_i, \quad i = 1, 2, \ldots, n$$

In centered form,

$$y_i = \alpha_0 + \gamma_1 x_{i1}^* + \cdots + \gamma_{p-1} x_{ip-1}^* + \epsilon_i$$

$$\underline{y} = X_s \begin{pmatrix} \alpha_0 \\ \underline{\gamma} \end{pmatrix} + \underline{\epsilon} \qquad X_s = (\underline{1}_n, X^*)$$

$$X_s' X_s = \begin{bmatrix} n & \underline{0}' \\ \underline{0} & R_{xx} \end{bmatrix}$$

R_{xx} - Correl^n matrix of $x_1, \ldots, x_{p-1}$

$$R_{xx} = \begin{bmatrix} 1 & r' \\ r & R_{22} \end{bmatrix} \quad [\text{Partition of } R_{xx}]$$

where r is $\underline{r} = (r_{12}, r_{13}, \ldots, r_{1p-1})'$

r_{1j} = Sample correlation coeff between x_1 & x_j

R_{22}: matrix formed by ~~detet~~ deleting first row and first column from R_{xx}

$$r_{1j} = \underline{x}^{*(1)\prime} \underline{x}^{*(j)}, \quad j = 2, \ldots, p-1$$

$$Var(\hat{\underline{\gamma}}) = \sigma^2 R_{xx}^{-1}$$

The (1,1) th element of R_{xx}^{-1}

$$(R_{xx}^{-1})_{(1,1)} = (1 - \underline{r}' R_{22}^{-1} \underline{r})^{-1}$$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} C_{11}^{-1} & -C_{11}^{-1} C_{12} \\ -C_{21} C_{11}^{-1} & A_{22}^{-1} + C_{21} C_{11}^{-1} C_{12} \end{bmatrix}$$

$$C_{11} = A_{11} - A_{12} A_{22}^{-1} A_{21} \; ; \quad C_{21} = A_{22}^{-1} A_{21}, \quad C_{12} = A_{12} A_{22}^{-1}$$

$$\mathrm{Var}(\hat{\gamma}_1) = \sigma^2 (1 - r' R_{22}^{-1} r)^{-1}$$

$\underset{\sim}{x}^{*(j)}$: jth column of X^*

$X^{*(j)}$: X^* with jth column removed.

Consider the formal regression of $\underset{\sim}{x}^{*(j)}$ on the column of $X^{*(j)}$ including an intercept term.

& find the coeff of determination R_j^2 of this regression

x_{ij}^* for y_i & use $\Sigma x_{ij}^* = 0$, $\Sigma x_{ij}^{*2} = 1$

$$\Sigma (y_i - \bar{y})^2 = \Sigma x_{ij}^{*2} = 1. \qquad [\bar{Y} = 0\ ??]$$

$$R_j^2 = 1 - \frac{RSS_j}{\Sigma (y_i - \bar{y})^2} = 1 - RSS_j$$

where RSS_j is the residual sum of sq. of ~~this~~ vector $x^*(j)$ and matrix $X^{*(j)}$

$$RSS = Y'(I-P)Y$$

$$= \Sigma (y_i - \bar{y})^2 - y' \tilde{X} (\tilde{X}'\tilde{X})^{-1} \tilde{X}' y$$

Substituting $\underset{\sim}{x}^{*(j)}$ for $\underline{Y}$, $X^{*(j)}$ for $\tilde{X}$ & use $\Sigma (y_i - \bar{y})^2 = 1$

$$RSS_j = 1 - \underset{\sim}{x}^{*(j)'} X^{*(j)} (X^{*(j)'} X^{*(j)})^{-1} X^{*(j)'} x^{*(j)}$$

$$= 1 - \underset{\sim}{x}^{*(j)'} P_j \underset{\sim}{x}^{*(j)} \qquad P_j = X^{*(j)} (X^{*(j)'} X^{*(j)})^{-1} X^{*(j)'}$$

When $j = 1$, $X^{*(1)'} \underset{\sim}{x}^{*(1)} = \underline{r}$

$$X^{*(1)'} X^{*(1)} = R_{22}$$

$$RSS_1 = 1 - \underline{r}' R_{22}^{-1} \underline{r}$$

$$\Rightarrow R_1^2 = 1 - RSS_1$$

$$\Rightarrow R_1^2 = \underline{r}' R_{22}^{-1} \underline{r}$$

So, $\quad Var(\hat{\gamma}_1) = \sigma^2 (1 - \underline{r}' R_{22}^{-1} \underline{r})^{-1}$

$$= \sigma^2 / (1 - R_1^2)$$

Similar formula holds for the regression coefficients.

$$Var(\hat{\gamma}_j) = \frac{\sigma^2}{(1-R_j^2)}$$

Note that $\quad x^{*(j)\prime} x^{*(j)} = 1$ & $(I_n - P_j)$ is symmetric and idempotent.

$$1 - R_j^2 = RSS_j - 1 - x^{*(j)\prime} P_j \underline{x}^{*(j)}$$

$$= \underset{\sim}{x}^{*(j)\prime} (I_n - P_j) \underset{\sim}{x}^{*(j)}$$

$$= \| (I_n - P_j) \underset{\sim}{x}^{*(j)} \|^2$$

$(1 - R_j^2)$ measures how close $\underset{\sim}{x}^{*(j)}$ is to close the subspace $\mathcal{C}(x^{*(j)})$, since it is the squared length of residual vector where $\underline{x}^{*(j)}$ is projected onto $\mathcal{C}(x^{*(j)})$

<u>Variance Inflation Factors (VIF)</u>

$$Var(\hat{\gamma}_j) = \frac{\sigma^2}{1 - R_j^2}$$

Since R_j^2 is a squared correlation,

$$0 \leq R_j^2 \leq 1$$

$Var(\hat{\gamma}_j) \geq \sigma^2$ with equality iff $R_j^2 = 0$

This occurs when $\underset{\sim}{x}^{*(j)}$ is orthogonal to the other collumns of X^*. The term $(1 - R_j^2)^{-1}$ is called the j-th variance inflation factor or VIF_j

So, $VIF_j = Var[\hat{\gamma}_j] / \sigma^2$

or, $VIF_j = \| (I_n - P_j) x^{*(j)} \|^{-2}$

In terms of original X, using $\hat{\gamma}_j = \hat{\beta}_j s_j$

$$VIF_j = \frac{Var(\hat{\gamma}_j)/\sigma^2}{}$$

$$= Var(\hat{\beta}_j s_j)/\sigma^2$$

$$= s_j^2 \left[(X'X)^{-1}\right]_{(j+1,\, j+1)}$$

Variances and eigenvalues

A is a $n \times n$ sym matrix

then $\exists$ $\perp$ matrix $T = (\underline{t}_1, \dots, \underline{t}_n)$ $\ni$

$$A = T' \Lambda T$$

$$\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$$

λ_i: eigen values of A

Consider spectral decomposition of R_{xx}

$R_{xx} = T' \Lambda T$,

$T = (t_{ij})$ is orthogonal

$\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_{p-1}\}$

λ_i: eigen values of R_{xx} are positive

So, $R_{xx}^{-1} = T \Lambda^{-1} T'$

and $Var(\hat{\gamma}_j) = \sigma^2 R_{xx}^{-1}{}_{(j,j)}$

$$= \sigma^2 \sum_{l=1}^{p-1} t_{jl}^2 \lambda_l^{-1}$$

Since T is $\perp$, $\sum t_{ji}^2 = 1$

So, $|t_{ji}| \leq 1$ for j & i

Condition Number

The condition number if the matrix X is the ratio of the largest and smallest singular values of X and is written as

$$K(x) = \left(\frac{\lambda_{max}}{\lambda_{min}}\right)$$

[Singular value = sq root of the eigen value of $X'X$]

λ_{max} & λ_{min} are the largest & smallest eigen values of $X'X$.

Collinearity & prediction

Collinearity the estimation of the regression coefficients much more than prediction.

Suppose we want to predict Y, the response at

$$\underset{\sim}{x_0} = (x_{01}, x_{02}, \ldots, x_{0p-1})'$$

The predictor $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_c' \underset{\sim}{x_0}$

$$\underline{\hat{\beta}_c} = (\hat{\beta}_1, \ldots, \hat{\beta}_{p-1})$$

$$\hat{Y} = \bar{Y} + \hat{\beta}_c' (\underline{x_0} - \underline{\bar{x}}), \quad \bar{x} = (\bar{x}_1, \ldots, \bar{x}_{p-1})$$

$$\underline{\hat{\beta}_c} = (\tilde{X}'\tilde{X})^{-1} \tilde{X} y, \quad \tilde{X} \rightarrow \text{centered } X$$

(without $\underset{\sim}{1_n}$)

$$\mathrm{Cov}(\bar{Y}, \hat{\beta}_c) = \mathrm{Cov}\left(\frac{1}{n} \underset{\sim}{1_n}' \underline{Y}, (\tilde{X}'\tilde{X})^{-1} \tilde{X}' \underline{Y}\right)$$

$$= 0$$

$(\because 1_n' \tilde{X} = 0)$

$$\mathrm{Var}(\hat{y}) = \mathrm{Var}\left(\bar{y} + \hat{\beta}_c' (\underline{x}_0 - \underline{\bar{x}})\right)$$

$$= \mathrm{Var}(\bar{y}) + \mathrm{var}\left(\hat{\beta}_c' (\underline{x}_0 - \underline{\bar{x}})\right)$$

$$= \sigma^2/n + (\underline{x}_0 - \underline{\bar{x}})\sigma^2 . (\tilde{X}'\tilde{X})^{-1} (\underset{\sim}{x}_0 - \underline{\bar{x}})$$

$$= \sigma^2 \left[1/n + (\underset{\sim}{x}_0 - \underline{\bar{x}})' (\tilde{X}'\tilde{X})^{-1} (\underline{x}_0 - \underline{\bar{x}}) \right]$$

#

If we are predicting the response at x_0, the variance of the predictor depends on how outlying $\underset{\sim}{x}_0$ is.
Prediction at pts close to $\underline{\bar{x}}$ have small error.

<u>Departuries from assumptions : Diagonosis and remedies</u>

① Different form of residuals
- Hat matrix (projection matrix) diagonalog's

② Most form of misspecification : linear model is fitted where as $E(Y) = \mu(x)$ (unbiased)
where $\mu(x)$ is a non-linear f^n of x

③ $\mathrm{Var}(y|x) = \sigma^2$
suppose $V(y|x) = \omega(x)$, $\rightarrow$ not a constant f^n

④ If the var f^n is constant, the error ϵ may fail to be independent/ uncorrelated. ~~observation.~~
↳ systematic way, successive or sequentially
→ Durbin - Watson Test.

⑤ If the errors are not normally distributed.
(~~part~~ joint dist^n of the exp. variables are approximitly normal)

⑥ ...

⑥ Outliers.

Hat matrix diagonals :

Model: $\underline{y} = X\underline{\beta} + \underline{\epsilon}$, $X^{n\times p}$ of rank p.

$$\underline{e} = (I_n - P)\underline{y} = (I_n - P)\underline{\epsilon} \qquad [PX = X]$$

The element of the fitted regression $\hat{\underline{y}} = X\hat{\beta}$.

$$\hat{\underline{y}} = P\underline{y}$$

P is called the Hat matrix since it ~~H~~ transforms the responses into the fitted values. and denoted by H.

$$E(\underline{e}) = E[(I_n - H)y] = 0$$

$$Var(\underline{e}) = Var[(I_n - H)\underline{y}] = \sigma^2 \cdot (I_n - H)$$

$$E(\hat{\underline{y}}) = H\,E(\underline{y}) = X\underline{\beta}$$

$$V(\hat{\underline{y}}) = H\,var(\underline{y})\,H' = \sigma^2 H$$

$$cov(\underline{e}, \hat{\underline{y}}) = Cov((I_n - H)\underline{y}, HY)$$
$$= 0$$

$\Rightarrow$ independence of $\underline{e}$ and $\hat{\underline{y}}$ under the normality assumption.

If $H = ((h_{ij}))$, then h_{ii} , the diagonal elements are called the hat matrix diagonals and sometimes denoted by h_i , so $Var(e_i) = \sigma^2(1 - h_i)$

$\rightarrow$ When the model is correct, then also the variances of the residuals depends on the hat matrix diagonals.

$$Cov(e_i, e_j) = -h_{ij}\sigma^2$$

The correlation between e_i & e_j, $\rho_{ij} = \dfrac{-h_{ij}}{\sqrt{(1-h_i)(1-h_j)}}$

$\rightarrow$ depends only on the elements of X.

Sum of sq due to regression.

$$= \hat{\beta}' X' \underline{Y} = \hat{\underline{Y}}' \underline{Y} = \underline{Y}' H' \underline{Y}$$

$$= \underline{Y}' H^2 \underline{Y}$$

$$= \hat{\underline{Y}}' \hat{\underline{Y}}$$

RSS

$$(\underline{Y} - X\hat{\beta})'(\underline{Y} - X\hat{\beta})$$

$$(\underline{Y} - H\underline{Y})'(\underline{Y} - H\underline{Y})$$

The average $V(\hat{Y}_i)$ over all data points $= \frac{1}{n} \sum_{i=1}^{n} var(\hat{Y}_i)$

$$= \frac{\sigma^2}{n} tr(H)$$

$$= \frac{\sigma^2}{n} p$$

$H \underline{1}_n = \underline{1}_n$, when the model contains an intercept term.

$\Rightarrow$ every row of H adds upto 1 and every column of H as H is symmetric.

$$\underline{1}_n' H, \quad \underline{1}_n' H \underline{1}_n = n$$

The diagonal entries of H are often called leverage as $\hat{Y}_i = h_{ii} Y_i + \sum_{j \neq i} h_{ij} y_j$

- indicates how heavily Y_i contributes $\hat{Y}_i$.

Internally Student typed Residuals

When the model is correct, the variances of the residuals depend on h_{ii}. So we scale the residuals to have approximately unit variance.

$\longrightarrow$ leads to internally student typed residuals.

$$r_i = \frac{e_i}{s(1-h_{ii})^{1/2}}, \quad s^2 = \underline{e}'\underline{e}/(n-p)$$

usual elements of σ^2.

8: $\frac{r^2}{(n-p)}$ has a Beta $[\frac{1}{2}, \frac{1}{2}(n-p-1)]$ distⁿ

r_i's are identically distributed.

Since the residuals and hence the estimate of σ^2, s^2 can be affected by outliers, the externally student typed residuals are defined.

$$t_i = \frac{e_i}{s(i)\,(1-h_i)^{1/2}}$$

$s(i)$: same as s which is calculated in the usual way form $(n-1)$ data pts that remain after deleting the ith pt.

- Results is an estimate of σ^2 that will not be affected of the ith point is an outlier.

The relationship between $s(i)^2$ & s^2

Theorem: Let, $\hat{\beta}$ and $\hat{\beta}(i)$ denote the LSE of β with and without the ith obs included in the data. Then

$$\hat{\beta} - \hat{\beta}(i) = \frac{(X'X)^{-1}\underline{x}_i e_i}{1-h_i}, \quad \underline{x}_i \to \text{ith row of } X.$$

Proof:

Let, $X(i)$ denote the regression matrix X when the ith row is deleted.

$$X(i)'X(i) = X'X - \underline{x}_i\underline{x}_i'$$

$$X' = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix}$$

$$(X(i)'X(i))^{-1}$$

$$= (X'X - \underline{x}_i\underline{x}_i')^{-1}$$

$$X'X = (x_1 \cdots x_n)\begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix}$$

$$\left[(A - \underline{u}\underline{v}')^{-1} = A^{-1} + \frac{A^{-1}\underline{u}\underline{v}'A^{-1}}{1-\underline{v}'A^{-1}\underline{u}}\right]$$

$$= (X'X)^{-1} + \frac{(X'X)^{-1} \underline{x}_i \underline{x}_i' (X'X)^{-1}}{1 - \underline{x}_i' (X'X)^{-1} \underline{x}_i}$$

$$= (X'X)^{-1} + \frac{(X'X)^{-1} \underline{x}_i \underline{x}_i' (X'X)^{-1}}{1 - h_i}$$

$\underline{Y}(i)$ → same as $\underline{Y}$ when ith element is deleted.

$$\underline{\hat{\beta}}(i) = (x(i)' x(i))^{-1} x(i)' \underline{Y}$$

$$= \left[(X'X)^{-1} + \frac{(X'X)^{-1} \underline{x}_i \underline{x}_i' (X'X)^{-1}}{1 - h_i} \right] \left[X'\underline{Y} - \underline{x}_i y_i \right]$$

$$= \underline{\hat{\beta}} + \frac{(X'X)^{-1} \underline{x}_i \underline{x}_i' (X'X)^{-1} X'\underline{Y}}{1 - h_i} - (X'X)^{-1} \underline{x}_i y_i - \frac{(X'X)^{-1} x_i x_i'\ldots}{1 - h_i}$$

$$= \underline{\hat{\beta}} + \frac{(X'X)^{-1} \underline{x}_i}{1 - h_i} \left[\underline{x}_i' \underline{\hat{\beta}} - y_i (1 - h_i) - h_i y_i \right]$$

$$= \underline{\hat{\beta}} + \frac{(X'X)^{-1} \underline{x}_i}{1 - h_i} \left[y_i - \underline{x}_i' \underline{\hat{\beta}} \right]$$

$$= \underline{\hat{\beta}} - \frac{(X'X)^{-1} x_i e_i}{1 - h_i}$$

$$(n-p-1) s(i)^2 = \sum_{k \neq i} \left[y_k - \underline{x}_k' \underline{\hat{\beta}}(i) \right]^2$$

$$= \sum_{k \neq i} \left[y_k - \underline{x}_k' \left(\underline{\hat{\beta}} - \frac{(X'X)^{-1} \underline{x}_i e_i}{1 - h_i} \right) \right]^2$$

$$= \sum_{k \neq i} \left[e_k + \frac{\underline{x}_k' (X'X)^{-1} \underline{x}_i e_i}{1 - h_i} \right]^2$$

$$= \sum_{k \neq i} \left[e_k + \frac{h_{ki} e_i}{1 - h_{ii}} \right]^2$$

$$= \sum_{K=1}^{n} \left[e_K + \frac{h_{Ki}\, e_i}{1 - h_i} \right]^2 - \frac{e_i^2}{(1-h_i)^2}$$

$$\left[e_i + \frac{h_{ii}\, e_i}{1 - h_{ii}} \right]^2$$

He

H is symmetric & $H\underline{e} = \underline{0}$ & $H^2 = H$

$$\sum_{K} h_{Ki}\, e_K = 0 \quad ; \quad \sum_{K=1}^{n} h_{Ki}^2 = h_{ii}$$

Using this,

$$(n-p-1)\, s(i)^2 = \sum_{K} e_K^2 + \frac{e_i^2}{1-h_i} \sum h_{Ki}^2$$

$$+ \frac{2e_i}{1-h_i} \sum_{K} h_{Ki}\, e_K$$

$$- \frac{e_i^2}{(1-h_i)^2}$$

$$= (n-p)\, s^2 + \frac{e_i^2 h_{ii}}{(1-h_i)^2} - \frac{e_i^2}{(1-h_i)^2}$$

$$= (n-p)\, s^2 - \frac{e_i^2}{1-h_i}$$

$t_i \sim t_{n-p-1}$ dist^n

$$\frac{e_i^2}{1-h_{ii}} = r_i^2 s^2$$

$$t^2 = \frac{e_i^2}{s(i)^2 (1-h_i)} \quad \frac{(n-p-1)}{(n-p-1)}$$

$$= \frac{e_i^2 (n-p-1)}{(1-h_i)\left[(n-p)s^2 - \frac{e_i^2}{1-h_i}\right]}$$

$$= \frac{e_i^2}{s^2(1-h_{ii})} \cdot \frac{n-p-1}{n-p-r_i^2}$$

$$= \frac{r_i^2 (n-p-1)}{n-p-r_i^2}$$

$$= \frac{A}{1-A}(n-p-1) \quad ; \quad A = \frac{r_i^2}{n-p}$$

$$\sim \text{Beta}(1/2, 1/2(n-p-1))$$

If $A \sim \text{Beta}(1/2\alpha, 1/2\beta)$ then

$\beta A\{\alpha(1-A)\}^{-1}$ has an $F_{\alpha,\beta}$ distⁿ

Take $\alpha = 1$, $\beta = n-p-1$

$t_i^2 \sim F_{1,n-p-1}$ distⁿ $\Rightarrow t_i \sim t_{n-p-1}$

- The hat matrix diagonals can be interpreted as a measure of distance in $(p-1)$ dimensional space.

$$h_{ii} = n^{-1} + \frac{1}{(n-1)} MD_i \quad ;$$

$$MD_i = (\underset{\sim}{x}_i - \bar{x})' S^{-1} (x_i - \bar{x})$$

$\underset{\sim}{x}_i$: ith reduced row of reg matrix X.

Example: Model: $y_i = \beta x_i + \epsilon_i$; $i = 1, 2, \ldots, n$.

$$h_{ii} = \frac{x_i^2}{\sum x_k^2}$$

Model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

$$h_{ii} = 1/n + \frac{\sum (x_i - \bar{x})^2}{\sum (x_k - \bar{x})^2}$$

Exercise: Show that for R_{xx}, $\lambda_{max} \geq 1$ and $VIF_j \leq K^2$, where K is the condition no

Non-constant variance:

Model: $y_i = \underline{x}_i \underline{\beta} + \epsilon_i$; $i = 1, 2, \ldots, n$

$\epsilon_i \sim$ indp $N(0, \cdot)$

$Var(\epsilon_i) = \sigma_i^2$ instead of $V(\epsilon_i) = \sigma^2$; $i = 1, 2, \ldots, n$

σ_i^2 may depend on either on mean $E[Y_i] = \underline{x}_i \underline{\beta}$ and possibly have other parameters or on a vector explanatory variables.

LSE of β may not be efficient of σ_i^2 are not equal.

Assume that $\sigma_i^2 = w(\underline{z}_i, \underline{\lambda})$

$\underline{z}_i$: a vector of known expl variables for the ith obs.

w: a variance f^n with the property that for some $\underline{\lambda}_0$, $w(\underline{z}_0, \underline{\lambda}_0)$ is independent of $\underline{z}_0$

The form of w is known but not the value of $\underline{\lambda}$.

$w(\underline{z}, \underline{\lambda}) = \exp\{\underline{z}' \underline{\lambda}\}$

The residuals $\underline{e}$ from a LS fit contain the informations about the variances. If $\Sigma = diag\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2\}$

then $V(\underline{e}) = Var[(I-H)\underline{Y}]$

$= Var[(I-H)\underline{\epsilon}]$

$= (I-H)\Sigma(I-H)'$

$= (I-H)\Sigma(I-H)$

So, $var(e_i) = (1-h_{ii})^2\sigma_i^2 + \sum_{k\neq i} h_{ik}^2 \sigma_k^2$

$h_{ik} << h_{ii}$ for $k \neq i$

If i th obs is a high-leaverage pt, $Var(e_i)$ will be small

Note $E[e_i] = 0$

$V(e_i) = E(e_i^2)$

Define ; $b_i = \frac{e_i^2}{1-h_i}$ → useful in graphical display.

When $\sigma_i^2 = \sigma^2 \ \forall i$; then

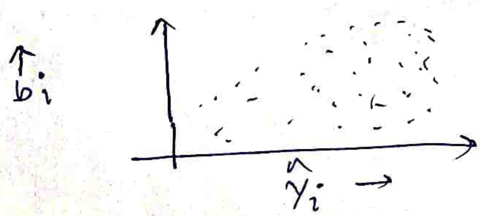
$E(b_i) = \frac{1}{1-h_i} V(e_i) = \frac{1}{1-h_{ii}}\left[(1-h_{ii})^2\sigma^2 + \sum_k h_{ik}^2\sigma^2\right]$

$= \sigma^2$ [As I-H is idempotent]

Note: Even if the variances are unequal, then fitted values $\hat{y}_i = \underline{x}_i \hat{\underline{\beta}}$ have expectation $E[Y_i]$

The plot of b_i vs $\hat{y}_i$ (or equivalently r_i^2)

Many times obs will large means have large variances.

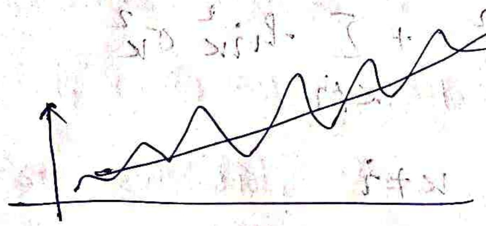


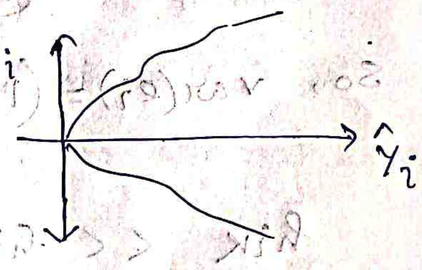
— wedged shaped.

A smoother process through the plot may reveal the relationship between the means & the variances

Alternatively the raw residuals can be plotted v/s the fitted values.

- The fan shaped pattern indicates variances increasing will the mean





Formal Test of $H_0: \underline{\lambda} = \underline{\lambda}_0$

Under normality assumption, the log-likelihood f^n $l(\underline{\beta}, \underline{\lambda})$

$$l(\underline{\beta}; \underline{\lambda}) = c - \frac{1}{2}\left\{ \ln|\Sigma| + (\underline{Y} - X\underline{\beta})' \Sigma^{-1} (\underline{Y} - X\underline{\beta}) \right\}$$

$$\cancel{= c - \frac{1}{2}\{}$$

$$= c - \frac{1}{2}\left\{ \sum_{i=1}^{n} \log w_i + \sum_{i=1}^{n} \frac{(y_i - x_i'\underline{\beta})^2}{w_i} \right\}$$

$\Sigma = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$

$\sigma_i^2 = w(\underline{z}_i, \underline{\lambda})$

$\quad\;\; = w_i$

c is a constant.

The score functions $\frac{\partial l}{\partial \underline{\beta}} = X' \Sigma^{-1} (\underline{Y} - X\underline{\beta})$

$$\frac{\partial l}{\partial \underline{\lambda}} = -\frac{1}{2}\left[\sum_{i=1}^{n} \left\{ \frac{1}{w_i} - \frac{(y_i - x_i'\beta)^2}{w_i^2} \right\} \frac{\partial w_i}{\partial \underline{\lambda}} \right]$$

The MLE of $\underline{\beta}$ and $\underline{\lambda}$ can be obtained using an algorithm.

~~Set~~ Step-1 Put $\underline{\lambda} = \lambda_0$

Step-2 Compute $\hat{\underline{\beta}} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}\underline{Y}$ using a weighted

Step-3 Solve $\left.\frac{\partial f}{\partial \underline{\lambda}}\right|_{\underline{\beta}=\hat{\underline{\beta}}} = 0$ for $\underline{\lambda}$

Step-4 Repeat steps II and III until convergence.

(*) Say $\hat{\lambda}$, then find $\hat{\Sigma}_1 = \text{diag}\{w(z_1, \hat{\underline{\lambda}}), \ldots w(z_n, \hat{\underline{\lambda}})\}$

(Step-II) $\hat{\beta}_2 = (X'\hat{\Sigma}_1^{-1}X)^{-1} X'\hat{\Sigma}_1^{-1}\underline{Y}$

(Step-III) : $\left.\frac{\partial f}{\partial \underline{\lambda}}\right|_{\hat{\underline{\beta}} = \hat{\beta}_2} = 0$

Special case:

$w(\underline{z}, \underline{\lambda}) = \exp(\underline{z}'\underline{\lambda})$; $\underline{z}' = (1, z_1, \ldots, z_k)$

; $w_i = \exp\{\underline{z}_i', \underline{\lambda})$

$$\frac{\partial w_i}{\partial \underline{\lambda}} = \exp(\underline{z}_i'\underline{\lambda})\, \underline{z}_i = w_i \underline{z}_i$$

$$\frac{\partial l}{\partial \underline{\lambda}} = -\frac{1}{2}\left[\sum_{i=1}^{n}\left\{\frac{1}{w_i} - \frac{(y_i - x_i'\underline{\beta})^2}{w_i^2}\right\} w_i \underline{z}_i\right]$$

$$= -\frac{1}{2}\left[\sum_{i=1}^{n}\left\{1 - \frac{\epsilon_i^2}{w_i}\right\}\underline{z}_i\right]$$

Take, $d_i = \frac{\epsilon_i^2}{w_i}$; $\underline{d} = (d_1, \ldots d_n)'$

$\underset{n\times(k+1)}{\underline{Z}} \longrightarrow$ matrix whoose ith row is $\underline{z}_i'$

$$\frac{\partial l}{\partial \underline{\lambda}} = \frac{1}{2} z' (\underline{d} - 1_n)$$

$$\operatorname{Var}\left(\frac{\partial l}{\partial \underline{\beta}} \right) = x' \Sigma^{-1} \operatorname{var}(\underline{\epsilon}) \Sigma^{-1} x = (x' \Sigma^{-1} x)$$

$$\operatorname{Var}\left(\frac{\partial l}{\partial \underline{\lambda}} \right) = \operatorname{Var}\left[\tfrac{1}{2} z' (\underline{d} - 1_n) \right]$$

$$= \frac{1}{4} \underline{z}' V(\underline{d}) \underline{z}$$

ϵ_i has variance ω_i, $d_i = \epsilon_i^2 / \omega_i$

d_i's are ~~independent~~ iid as χ_1^2

$$\operatorname{var}(d_i) = 2$$

$$\operatorname{Var}(\underline{d}) = 2I$$

$$\operatorname{Var}\left[\frac{\partial l}{\partial \underline{\lambda}} \right] = \frac{1}{2} \underline{z}' \underline{z}$$

Note, $E[\epsilon_i \epsilon_j^2] = 0$

$$E[\epsilon_j^3] = 0$$

$$\operatorname{Cov}(\underline{d}, \underline{\epsilon}) = 0$$

$$\operatorname{Cov}\left[\frac{\partial l}{\partial \underline{\beta}}, \frac{\partial l}{\partial \underline{\lambda}} \right] = \operatorname{Cov}\left[x' \Sigma^{-1} \underline{\epsilon}, \tfrac{1}{2} z' (\underline{d} - I_n) \right]$$

$$= \frac{1}{2} x' \Sigma^{-1} \operatorname{cov}(\underline{\epsilon}, \underline{d}) \underline{z} = \underline{0}$$

$$I(\underline{\beta}, \underline{\lambda}) = \begin{pmatrix} x \Sigma^{-1} x & 0 \\ 0 & \frac{1}{2} z' z \end{pmatrix}$$

Newton Raphson-Method :

$l(\underline{\gamma})$: The find the MLE of $\underline{\gamma}$ is given by the iterative process.

$$\gamma^{(m+1)} = \underline{\gamma}^{(m)} - \left[\left. \frac{\partial^2 l}{\partial \underline{\gamma} \partial \underline{\gamma}'} \right|_{\underline{\gamma}^{(m)}} \right]^{-1} \left[\left. \left\{ \frac{\partial l}{\partial \underline{\gamma}} \right\} \right|_{\gamma^{(m)}} \right]$$

$\gamma^{(m)} \rightarrow$ mth step estimator of γ.

$\frac{\partial l}{\partial \underline{\gamma}} \rightarrow$ first order derivative vector.

$\frac{\partial^2 l}{\partial \gamma \partial \underline{\gamma}'} \rightarrow$ ~~sect~~ second order derivative matrix.
(Hessian matrix)

<u>Fisher Scoring</u>: When the Hessian matrix is replaced by its expected value

$$\underline{\gamma}^{(m+1)} = \underline{\gamma}^{(m)} - \left\{ E\left[\frac{\partial^2 l}{\partial \underline{\gamma} \partial \underline{\gamma}'} \right]_{\gamma^{(m)}} \right\}^{-1} \left[\left. \frac{\partial l}{\partial \gamma} \right|_{\underline{\gamma}^{(m)}} \right]$$

Because of the relationship

$$-E\left[\frac{\partial^2 l}{\partial \gamma \partial \gamma'} \right] = E\left[\frac{\partial l}{\partial \underline{\gamma}} \cdot \frac{\partial l}{\partial \underline{\gamma}'} \right]$$

This is a more likely to be pd matrix.
holds under fairly general condition.

To solve the likelihood equation, we use Fisher scoring. The up dating equation

$$\underline{\beta}_{(m+1)} = \left(X' \Sigma_{(m+1)}^{-1} X \right)^{-1} X' \Sigma_m \underline{Y}$$

$$\underline{\lambda}_{(m+1)} = \underline{\lambda}_{(m)} + (Z'Z)^{-1} Z' (\underline{d} - \underline{1}_n)$$

$$\Sigma_{(m+1)} = \text{diag} \{ \omega(z_1, \lambda_{(m)}), \dots, \omega(z_n', \lambda_{(m)}) \}$$
$\exp(z_i', \lambda_{(m)})$

$$\underline{\gamma} = \begin{pmatrix} \underline{\beta} \\ \underline{\lambda} \end{pmatrix}; \quad E\left[\frac{\partial^2 l}{\partial \gamma \partial \gamma'} \right] = \begin{bmatrix} X' \Sigma^{-1} X & 0 \\ 0 & \frac{1}{2} Z'Z \end{bmatrix}$$

$$\frac{\partial l}{\partial \lambda} =$$

(*) can also be written as

$$Z'Z\, \lambda_{(m+1)} = Z'\left(\underline{d} - \underline{1}_n + Z\,\underline{\lambda}_{(m)}\right)$$

↳ normal eq for a formal regression of $\left(\underline{d} - \underline{1}_n + Z\,\lambda_{(m)}\right)$.

Testing of $H: \underline{\lambda} = \underline{\lambda}_0$

For the LR test, consider the maximization of the loglikelihood $l(\underline{\beta}) = c - \frac{1}{2}\left\{\sum_{i=1}^{n} \log w_i + \sum_{i=1}^{n} \frac{(y_i - x_i'\beta)^2}{w_i}\right\}$

When $\underline{\lambda} = \underline{\lambda}_0$, $w(z_i, \underline{\lambda}) = \sigma^2$ and this loglikelihood is just the regression likelihood.

$$l_R(\underline{\beta}, \sigma^2) = c - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}(\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})$$

↳ is maximized at β is the $\hat{\beta}_{OLS} = (X'X)^{-1}X'\underline{Y}$;

$$\hat{\sigma}^2 = \cancel{\frac{1}{n}\|Y - X\hat{\beta}\|^2}$$

$$= \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

So, the maximum value of the loglikelihood under the null hypothesis

$$L_R(\hat{\beta}, \hat{\sigma}^2) = c - \frac{1}{2}\left(n + n\log\hat{\sigma}^2\right)$$

& $l(\hat{\underline{\beta}}, \hat{\underline{\lambda}}) \rightarrow$ unrestricted max^m of the log likelihood f^n where $\hat{\underline{\beta}}$ & $\hat{\underline{\lambda}}$ has been calculated using the least algorithm.