## Chapter 3:Multiple Linear Regression

3.1 Description of the Data and Model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon, \epsilon,$
contains no systematic information for Y that is not already captured.
一系列定義野: Least Square: SSE $= S(\beta_0, \beta_1, \cdots, \beta_p) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p)^2 = ||\boldsymbol{y} - \boldsymbol{X\beta}||^2$, Least Square Solution: $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$. Fitted equation: $\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1} X_1 + \cdots + \widehat{\beta_p} X_p$, ordinary least squares residuals: $e_i = y_i - \hat{y}_i$. unbiased estimate of $\sigma^2$ is $\widehat{\sigma^2} = \frac{\text{SSE}}{n-p-1} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p-1}$(for no intercept model, only divide by n-p), called sum of squared residuals(RSS)

從 simple lin reg to multiple: 例如$\hat{Y} = 15.3276 + 0.7803X_1 - 0.0502X_2$, 想 fit 翻-0.0502 出黎嘅步驟係: 1. Fit Y vs X1, 2. Fit X2 vs X1, 3. Fit $e_{Y \cdot X_1} vs\ e_{X_2 \cdot X_1}$ 解釋: 第一步計到嘅 residual 係 part of Y not linearly related to X1, 第二步計到嘅係 part of X2 not linearly related to X1,第三步係 take out effect of X1 to give coeff

3.3 Scaling and Centering: Scale: for w or w/o intercept model, 兩種 scaling 方法:
1. Unit-length Scaling: $\widetilde{Z}_j = (X_j - \bar{x}_j)/L_j$, $L_j = \sqrt{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}$, same as y(just replace j by y), $\widetilde{Z}_j$has 0 mean and 1 length.$Cor(X_j, X_k) = \sum_{i=1}^{n} \tilde{z}_{ij}\tilde{z}_{ik}$
2. Standardizing: $\widetilde{X}_j = \frac{x_j - \bar{x}_j}{s_j}$, $s_j = \sqrt{\frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}{n-1}}$,
Center: only for w/ intercept, $X_j - \bar{x}_j$ to make mean as 0
Original to standardized by: $\widehat{\beta}_j = (s_y/s_j)\widehat{\theta}_j$,

3.4 Assumptions for Multi Linear Regression
1. $\varepsilon_1, \cdots, \varepsilon_n \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$2. $\mathbf{X}'\mathbf{X}$ is invertible
If assumptions hold, we have 2 results: 1. $\widehat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj})$2. $W = \text{SSE}/\sigma^2 \sim \chi^2(n-p-1)$, $\widehat{\beta}_j$ $independent$ $\widehat{\sigma^2}$ $(\widehat{\sigma^2} = \sum(y_i - \hat{y}_i)^2/(n-p-1))$
Multiple Corr. Coeff: $Cor(Y, \hat{Y}) = \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum(y_i - \bar{y})^2 \sum(\hat{y}_i - \bar{\hat{y}})^2}}$, $R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$, x% of the total variation in Y can be accounted for by the p predictors
Drawback of R^2: more variables will yield a larger R^2. so we have adjusted R^2:$R_a^2 = 1 - \frac{SSE/(n-p-1)}{\text{SST}/(n-1)}$, $meaning\ R_a^2 = 1 - \frac{n-1}{n-p-1}(1 - R^2)$,而$R_a^2$不可解作這個

3.5 Inference for Individual Regression Coefficients
$H_0$: $\beta_j = \beta_j^0$ vs $H_1$: $\beta_j \neq \beta_j^0$, $reject\ if\ |t_j| = \frac{\widehat{\beta}_j - \beta_j^0}{\text{s.e.}(\widehat{\beta}_j)} \geq t_{(n-p-1, \alpha/2)}$ $or\ p(|t_j|) \leq \alpha$
Conf. Interv.: $\widehat{\beta}_j \pm t_{(n-p-1,\alpha/2)} \times$ s.e. $(\widehat{\beta}_j)$
Even stat. not suff., a constant should always included

3.6 Test of Hypothesis in Linear Model and Prediction
Framework: $Reject\ H_0\ if\ F = \frac{[SSE(\text{RM}) - SSE(\text{FM})]/(p+1-k)}{SSE(\text{FM})/(n-p-1)} \geq F_{(p+1-k, n-p-1;\alpha)}$,meaning RM does not give as good a fit as FM.
FM has p+1 params(including $\beta_0$), SSE(FM)'s DF = n-p-1, RM has k params(including $\beta_0$), SSE(RM)'s DF = n-k
Note that $SSE(RM) \geq SSE(FM)$since additional params cant increase SSR.
1. $H_0$: $Y = \beta_0 + \epsilon\ vs\ H_1$: $FM$
$F = \frac{SSR/p}{SSE/(n-p-1)} = \frac{\text{MSR}}{\text{MSE}} = \frac{R_p^2/p}{(1-R_p^2)/(n-p-1)}$, since SSE(RM) = $\sum(y_i - \widehat{y_i})^2 = \sum(y_i - \bar{y})^2 = SST(FM)$,SSE(RM)-SSE(FM)=SST(FM)-SSE(FM)=SSR(FM)

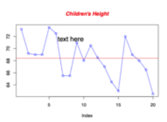
**Table 3.6** Analysis of Variance (ANOVA) Table in Multiple Regression

| Source | Sum of Squares | df | Mean Square | $F$-Test |
|---|---|---|---|---|
| Regression | SSR | $p$ | $\text{MSR} = \frac{\text{SSR}}{p}$ | $F = \frac{\text{MSR}}{\text{MSE}}$ |
| Residuals | SSE | $n-p-1$ | $\text{MSE} = \frac{\text{SSE}}{n-p-1}$ | |

2.(下列 model 假設有 6 個 params) $H_0$: $Y = \beta_0 + \beta_1 X_1 + \beta 3 X_3 + \epsilon$, $H_1$: $FM$, $F = \frac{(R_p^2 - R_q^2)/(p-q)}{(1-R_p^2)/(n-p-1)}$, FM has p params, RM has q params 注意呢到同上面唔同
特殊情況:如果 q=p-1(RM=FM-1),或者 simple lin re,$F = t_j^2 = \frac{\widehat{\beta_j}^2}{Var(\widehat{\beta_j})} < F_{(1, n-p-1;\alpha)}$
3.$H_0$: $\beta_1 = \beta_3 | (other\ \beta = 0)$, $H_1$: $Y$: $\beta_0 + \beta_1 X_1 + \beta_3 X_3 + \epsilon$
呢題做法: $RM$: $Y - 0.5X_1 - 0.5X_3 \sim 1$, $FM$: $Y \sim X_1 + X_3$, and use p=2,k=1 do F test.
4. $H_0$: $\beta_1 = \beta_3$, $other\ \beta = 0$, $H_1$: $Y$: $\beta_0 + \beta_1 X_1 + \ldots + \beta_6 X_6 + \epsilon$, RM 同上面一樣, FM 係全部 6 個 var, use p = 6, k = 1
=== Prediction ===
Prediction interval: $\widehat{y_0} \pm t_{(n-p-1,\alpha/2)}$ s.e. $(\widehat{y_0})$, s.e. $(\widehat{y_0}) = \hat{\sigma}\sqrt{1 + x_0'\left(X'X\right)^{-1} x_0}$
Confidence interval: $\widehat{\mu_0} \pm t_{(n-p-1,\alpha/2)}$ s.e. $(\widehat{\mu_0})$, s.e. $(\widehat{\mu_0}) = \hat{\sigma}\sqrt{x_0'\left(X'X\right)^{-1} x_0}$

R corner:

```
plot(x = 1:20,
     y = GaltonFamilies$childHeight[1:20],
     type="o",
     col="blue",
     xlab = "Index",
     ylab = "Height")
title(main="Children's Height", col.main="red", font.main=4)
abline(h=mean(GaltonFamilies$childHeight[1:20]),
       col="red")
text(x = 10, y = 72,                    # text draws the str
     labels = "text here",              # at the coordinates
     adj=c(1,1),
     cex = 1.5)
```




Basic: rev()reverse vector, rbind,cbind=combine vectors,t(X)transpose, matrix(c()/1:9/0,nrow=3,ncol=3),subset(df,df!='remove'),all.equal(a,b,tolerane)
Stat: confint(lm,level=0.99), coef(),resid(),fitted(),rnorm()runique()gen R.V of dist
Qqplot: qqPlot(lm), DTITS:ols_plot_dffits(lm), Cook:ols_plot_cooksd_bar(lm), PR:ols_plot_resid_pot(lm), leverage_cal: hatvales(lm),abline(lm,lwd=width)

Chapter 2 - Simple Linear Regression

## Chapter 4: Regression Diagnostics: Detection on Model Violations

4.1 – The Standard Regression Assumptions
1.Linearity: The model that relates Y to X1,X2,…,Xp is assumed to be linear in the regression parameters $\beta_0, \beta_1, \ldots, \beta_p$, namely: $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$
Check linearity assumption: Scatter plot of Y versus X, they should be linear plot
2. Errors: The errors $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$are assumed to be independently and identically distributed(iid) normal random variables each with mean 0 and a common variance $\sigma^2$
可以拆成四個 assumptions: 2.1: Normality: error is normal distributed, 2.2: errors have mean 0. 2.3: Constant Variance/Homogeneity: errors have same variance $\sigma^2$. 2.4: Independent-errors assumptions(如果唔符合就會有 auto-correlation problem): errors are independent of each other
3. Predictors: 3.1:predictors $X_1, X_2, \ldots, X_p$ $are$ nonrandom, $the\ values\ x_{1j}, x_{2j}, \ldots, x_{nj}$ are assumed fixed or selected in advanced.
3.2. The values $x_{1j}, x_{2j}, \ldots, x_{nj}$are measured without error, 這個很難被滿足
3.3 Predictors are linear independent of each other, 為咗$(X^T X)^{-1}$ exists
4. Observations: all observations are equally reliable and have an approximately equal role in determining regression results

4.2 Various Types of Residuals: $e_i = y_i - \hat{y}_i$ , fitted value $\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} x_{i1} + \cdots + \widehat{\beta_p} x_{ip}$
or written as: $\hat{y}_i = p_{i1} y_1 + \cdots + p_{in} y_n$, $\hat{\boldsymbol{y}} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$. and P = highlighted.
Leverage value: $\mathrm{P_{ii}}$即是$P$的對角線,如果係 simple lin.reg., $p_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{\sum(x_i - \bar{x})^2}$.簡單計下得到: $Cov(e) = \sigma^2(\boldsymbol{I} - \boldsymbol{P})$, $Var(e_i) = \sigma^2(1 - p_{ii})$, 而我地想呢個$Var(e_i)$一樣, 所以需要使用 studentized residual
Internal Studentized Residual(Default, rstandard in R): $\mathrm{r_i} = \frac{\mathrm{e_i}}{\hat{\sigma}\sqrt{1 - \mathrm{p_{ii}}}}$, 呢個不符合 t-distribution
External Studentized Residual(rstudent in R):$\mathrm{r_i^*} = \frac{\mathrm{e_i}}{\hat{\sigma}_{(\mathrm{i})}\sqrt{1 - \mathrm{p_{ii}}}}$, 符合 t 分佈,df =n-p-2,因為與$\hat{\sigma}_{(\mathrm{i})}$與$\mathrm{e_i}$獨立
$\widehat{\sigma^2} = \frac{\text{SSE}}{\text{n}-\text{p}-1}$, $\hat{\sigma}^2_{(\mathrm{i})} = \frac{\text{SSE}_{(\mathrm{i})}}{\text{n}-\text{p}-2}$, which is omitting the $i$th observation, they are both unbiased estimates of $\sigma^2$.
$\mathrm{r_i}$與$\mathrm{r_i^*}$的關係是 one − to − one related: $\mathrm{r_i^*} = \mathrm{r_i}\sqrt{\frac{\text{n}-\text{p}-2}{\text{n}-\text{p}-1-\mathrm{r_i^2}}}$,

4.3 Graphical Methods: Before and After Fitting a Model

| 咩圖 | 咩 assumption | 點先叫好 |
|---|---|---|
| Histogram | Normality | Bell shape |
| Dot/Box Plot | No outlier in observations | No point far away |
| Pairwise Scatter Plot Matrix ,Y against X1,..Xn (plot before fit, w/corr coeff) | Linearity | 呢個圖無用,因為有可能幾個 predictor 加埋就有 linearity |
| Q-Q plot(Residual vs normal scores) Scatter $(r_{(j)}, z_{(j-0.5)/n})$ | Normality assumption, $r_i \sim N(0,1)$ | 大 sample size 下,d point 貼住條線 of y=x |
| Residual vs Predictors | Linearity/Constant Var/Uncorrelated with X | 無 nonlinear/越大越發散 |
| Residual vs Fitted | Error Mean = 0 | 在某 fitted value 的 mean(residual) ~= 0 |
| Residual vs Fitted | Constant Variance | Same Spread of residuals |
| Fitted vs Predictor + Fitted Line | No outlier when fitting line | 呢個圖無用 只可以肉眼睇分 outlier |
| Index plot of internal studentized residuals | No outlier in Y-space(Response Var) | $r_i < 2/3$ |
| Index plot of leverage | No outlier in X-space(p 個 predictors) | $p_{ii} < 2(p+1)/n$ |
| 下列方法原理=Deleting ith observation and see the change in fitted value*: | | |
| Index plot of Cook's Distance$C_i = \frac{r_i^2}{p+1} \times \frac{p_{ii}}{1-p_{ii}}$ | No influential point | $C_i < F(0.5, p+1, n-p-1)$ *Ci < 4/n (from online)* |
| Index plot of $\mathrm{DFITS_i} = r_i^*\sqrt{\frac{p_{ii}}{1-p_{ii}}}$ | No influential point | $\lvert DFITS_i \rvert < 2\sqrt{(p+1)/(n-p-1)}$ ~ 2sqrt(p/n)(online) |
| Index plot of Hadi, $H_i = \frac{p_{ii}}{1-p_{ii}} + \frac{p+1}{1-p_{ii}}\frac{d_i^2}{1-d_i^2}$ | No influential point | No threshold |
| Potential-Residual Plot, $\frac{p_{ii}}{1-p_{ii}}\ vs\ \frac{p+1}{1-p_{ii}}\frac{d_i^2}{1-d_i^2}$ | No outlier in X/Y space as well as influential point | ↖=outlier in X-space ↘=outlier in Y-space ↗=influential points |

*Masking: detect 唔到係 outlier, Swamping: detect 錯咗 as outlier
Math corner for above graphs:
Q-Q plot: $r_{(j)}$係 order statistics, $\#\{i: r_i \leq r_{(j)}\} = j$, $\Phi(z_{(j-0.5)/n}) = \frac{j-0.5}{n}$ where $\Phi(\cdot)$ is the c.d.f. of $N(0,1)$, Leverage: $\frac{\sum_{i=1}^{n} p_{ii}}{n} = tr(P) = tr(X(X^T X)^{-1} X^T) = tr((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}) = \boldsymbol{tr}(\boldsymbol{I}_{\boldsymbol{p+1}}) = \boldsymbol{p+1}$
Ordinary Residual and leverage 關係: $p_{ii} + \frac{e_i^2}{SSE} \leq 1$, Hadi 個 di=$e_i/\sqrt{\text{SSE}}$
Cook's Distance Big Formula: $\mathrm{C_i} = \frac{\sum_{\mathrm{j=1}}^{\mathrm{n}}(\hat{y}_{\mathrm{j}} - \hat{y}_{\mathrm{j(i)}})^2}{\hat{\sigma}^2(\mathrm{p}+1)}$, $\hat{\sigma}^2$obtained from LSE using all obser.
Potential Function: $p_{ii}/(1 - p_{ii})$
DFITS Big Formula: $\mathrm{DFITS}_i = \frac{\hat{y}_i - \hat{y}_{i(i)}}{\hat{\sigma}_{(i)}\sqrt{p_{ii}}}$

Cal 機 program: 1. 3x3 matrix multi & inverse, 先入 0 代表乘法, 入 1 代表 inverse. 然後由左至右，上至下輸入第一個 matrix,(inverse 嘅話 X?會出 determinant) 然後逐欄輸入第二個 matrix, 會逐欄出答案.如果就咁求 inverse,就將第二個 matrix 入做 I 就得| 2. SSReg, SSError, R^2. 先係 Reg lin 入咗然後過黎 prog 2|3.stat increase.
Σx^2、Σx、n、Σy^2、Σy 及Σxy 保持不變.|4. (or given 兩點)Σx^2=14, Σx=6, n=3, Σy^2=38, Σy=10, Σxy=23 求直線. 先去 reg lin clear stat 然後(輸入已知兩點),prog4 順住入, 然後 Svar 睇翻 a,b 係乜,條線就係 y=a+bx

Ch1 Intro

2.1 Covariance and Correlation Coefficient and Example

$SampleCov(Y,X) = Cov(X,Y) = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{n-1} = \frac{1}{n-1}(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y})$

Cov(aX+bY,cZ) = acCov(X,Z) + bcCov(Y,Z)

$Cor(Y,X) = Cor(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{y_i-\bar{y}}{s_y}\right)\left(\frac{x_i-\bar{x}}{s_x}\right) = \frac{Cov(Y,X)}{s_y s_x} =$

$\frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{\sqrt{\sum(y_i-\bar{y})^2\sum(x_i-\bar{x})^2}}, where\ s_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}.\ -1 \le Cor(Y,X) \le 1$

注意從 R^2 計 Cor(Y,X)Need examine scatter plot tgt, Cant use for prediction, only pairwise, equal imprt.

2.2 Simple Linear Regression Model and Parameter Estimation

Model: $Y = \beta_0 + \beta_1 X + \epsilon$, $\epsilon$ contains no systematic information for determining Y that is not already captured in X.Y is primary importance.要有 linearity 假設

Each observation: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$SSE = S(\beta_0,\beta_1) = \sum_{i=1}^{n}\varepsilon_i^2 = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$, least squares solution

$\widehat{\beta_1} = Sxy/Sxx = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\sum_{i=1}^{n}x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}, \widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}.$

Least square reg. line: $\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1}X$

Fitted value: $\widehat{y_i} = \widehat{\beta_0} + \widehat{\beta_1}x_i$, Ordinary least squares residual: $e_i = y_i - \widehat{y_i}$

Beta1 的另一個表示方式: $\widehat{\beta_1} = \frac{Cov(Y,X)}{Var(X)} = Cor(Y,X)\frac{s_y}{s_x}$

2.3 Test of Hypothesis and Confidence Intervals

Core Assumption: For every fixed value of X, the error $\epsilon$'s are independent normal random variables with mean 0 and variance $\sigma^2$ $\varepsilon_1,\cdots,\varepsilon_n \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2)$

Unb. Est. and Distribution: $\widehat{\beta_0} \sim N\left(\beta_0, \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]\right), \widehat{\beta_1} \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$

Unb. Est. of $\sigma^2$: $\widehat{\sigma^2} = \frac{\sum e_i^2}{n-2} = \frac{\sum(y_i-\widehat{y_i})^2}{n-2} = \frac{\text{SSE}}{n-2}$, SSE=Sum of Squares of residuals,df=n-2

t-test: $H_0: \beta_1 = \beta_1^0$ vs $H_1: \beta_1 \ne \beta_1^0$, $t_1 = \frac{\widehat{\beta_1}-\beta_1^0}{\text{s.e.}(\widehat{\beta_1})}$, $df = n-2$, reject $H_0$if $|t_1| \ge t_{(n-2,\alpha/2)}$or $2\cdot P(T_{n-2} \ge |t_1|) \le \alpha$, meaning data does not support $\widehat{\beta_1}$to be $\beta_1^0$

Testing $\beta_1 = 0$ can be written as $t_1 = \frac{Cor(Y,X)\sqrt{n-2}}{\sqrt{1-[Cor(Y,X)]^2}}$

Confidence Interval: $\widehat{\beta_{0/1}} \pm t_{(n-2,\alpha/2)} \times \text{ s.e. }(\widehat{\beta_{0/1}})$

CI 的解讀: If we were to take repeated samples of the same size at the same values of X and construct, for example, 95% confidence intervals for the slope parameter for each sample, then 95% of these intervals would be expected to contain the true value of the slope.

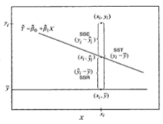
2.4 Prediction

Single response: $\widehat{y_0} \pm t_{(n-2,\alpha/2)}$ s.e. $(\widehat{y_0})$, where s.e. $(\widehat{y_0}) = \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$

Mean response: $\widehat{\mu_0} \pm t_{(n-2,\alpha/2)}$ s.e. $(\widehat{\mu_0})$, where s.e. $(\widehat{\mu_0}) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum(x_i-x)^2}}$

2.5 Measuring the Quality of Fit

1. t-test on $\widehat{\beta_1}$requires linearity & normality of $\epsilon$, but it is objective. 2. Cor(Y,X) subjective but only require linearity




We have $Cor(Y,\hat{Y}) = \frac{\sum(y_i-\bar{y})(\widehat{y_i}-\bar{\hat{y}})}{\sqrt{\sum(y_i-\bar{y})^2\sum(\widehat{y_i}-\bar{\hat{y}})^2}} = |Cor(Y,X)|$

SST=Total Sum of Squared deviation in Y from $\bar{Y}$,

$= \sum(y_i-\bar{y})^2 = \sum y_i^2 - 2\bar{y}\sum y_i + n\bar{y}^2$= SSR+SSE

SSR=Sum of Squares due to regression = $\sum(\widehat{y_i}-\bar{y})^2 = \sum_{i=1}^{n}(\widehat{\beta_0} + \widehat{\beta_1}x_i - \widehat{\beta_0} - \widehat{\beta_1}\bar{x})^2 = \widehat{\beta_1^2}S_{xx}$

SSE=Sum of squared residuals(errors) = RSS = residual sum of squares $= \sum(y_i - \widehat{y_i})^2$

Class discussion:$\hat{Y} = (\widehat{y_1},\cdots,\widehat{y_n})'$ is orthogonal to $\hat{Y} - Y = (\widehat{y_1}-y_1,\cdots,\widehat{y_n}-y_n)'$

So $\sum_{i=1}^{n}\widehat{y_i}(\widehat{y_i}-y_i) = 0.$

Goodness of fit Index: $R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}} = [Cor(Y,X)]^2 = [Cor(Y,\hat{Y})]^2$

It is the proportion of the total variability in the response variable Y that is accounted for by the predictor variable X.

2.6 Regression Line Through Origin

$Y = \beta_1 X + \epsilon, \widehat{\beta_1} = \frac{\sum y_i x_i}{\sum x_i^2}, \widehat{y_i} = \widehat{\beta_1}x_i, e_i = y_i - \widehat{y_i}$, s.e. $(\widehat{\beta_1}) = \frac{\hat{\sigma}}{\sqrt{\sum x_i^2}}$,where $\hat{\sigma} = \sqrt{\frac{\sum e_i^2}{n-1}} = \sqrt{\frac{\text{SSE}}{n-1}}$, SST=SSR+SSE 變做$\sum_{i=1}^{n}y_i^2 = \sum_{i=1}^{n}\widehat{y_i}^2 + \sum_{i=1}^{n}e_i^2$

R^2 會變做$\frac{\sum\widehat{y_i}^2}{\sum y_i^2} = 1 - \frac{\sum e_i^2}{\sum y_i^2}$,佢嘅意思變做 proportion of the variation in Y that is accounted for by the predictor variable X (without adjustment of Y)

The residuals may not necessarily add up to 0.Consider a case that $\bar{x} = 0$ but $\bar{y} \ne 0, \sum_{i=1}^{n}e_i = \sum_{i=1}^{n}(y_i - \widehat{\beta_1}x_i) = \sum_{i=1}^{n}\left(y_i - \frac{\sum_{j=1}^{n}x_j y_j}{\sum_{j=1}^{n}x_j^2}x_i\right) = \sum_{i=1}^{n}y_i - \frac{\sum_{j=1}^{n}x_j y_j}{\sum_{j=1}^{n}x_j^2}\sum_{i=1}^{n}x_i \ne 0$

Uses of regression

- Prediction: modelling existing observations or forecasting new data
- Decision making: make policy/ decisions by the relationship among variables
- Exploring associations: summarizing how well one variable predicts the outcome

Steps in Regression analysis

Statement of problem → Selection of potentially relevant variables → Data collection → Model specification → Choice of fitting method → Model fitting → Model validation and criticism → Using the chosen model for the soln of the posed problem

Probability distributions

$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

$X \sim Bernoulli(p) \leftrightarrow P(X=1) = p, P(X=0) = 1-p$

$X \sim Bin(n,p) \leftrightarrow P(X=k) = nCk\, p^k(1-p)^k, k = 0,1,\ldots n$

$X \sim Poisson(\lambda) \leftrightarrow P(X=k) = \frac{e^{\lambda}\lambda^k}{k!}$

$X \sim Normal(\mu,\sigma^2) \leftrightarrow f(x) = \frac{1}{\sqrt{2\pi\sigma}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}, for\ all\ x$

Linear transformation of normal is also normal:

$X \sim N(\mu,\sigma^2) \rightarrow aX + b \sim N(a\mu + b, a^2\sigma^2)$

The 3σ rule:

$P(|X-\mu| > 3\sigma) \approx 0.0026$

Standard Normal: N(0,1)

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2};$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy.$$

- Find the least squares estimates of $\beta_0$ and $\beta_1$ in terms of $y_i$
- Find the $Var(\hat{\beta}_0)$ and $Var(\hat{\beta}_1)$.

Sol: (a) $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{3}(i-2)y_i}{\sum_{i=1}^{3}(i-2)^2} = \frac{1}{2}(y_3-y_1)$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = \frac{1}{3}(y_1+y_2+y_3) - (y_3-y_1) = \frac{4}{3}y_1 + \frac{1}{3}y_2 - \frac{2}{3}y_3$

(b) $Var(\hat{\beta}_0) = Var(\frac{4}{3}y_1 + \frac{1}{3}y_2 - \frac{2}{3}y_3) = \frac{16}{9}Var(y_1) + \frac{1}{9}Var(y_2) + \frac{4}{9}Var(y_3) = \frac{16}{9}\sigma^2 + \frac{2}{9}\sigma^2 + \frac{12}{9}\sigma^2 = \frac{10}{3}\sigma^2$

$Var(\hat{\beta}_1) = \frac{1}{4}(3\sigma^2 + \sigma^2) = \sigma^2$

$X \sim Unif(a,b) \leftrightarrow f(x) = \frac{1}{b-a}, \forall x \in (a,b)$

$P(X > t | X > s) = P(X > t - s)$

$X \sim Exp(\lambda) \leftrightarrow f(x) = \begin{cases}\lambda e^{-\lambda x}, if\ x \ge 0 \\ 0, otherwise\end{cases}$

$X \sim \Gamma(\alpha,\lambda) \leftrightarrow f(x) = \begin{cases}\frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, if\ x \ge 0 \\ 0, otherwise\end{cases}$

$Gamma\ function, \Gamma(\alpha) = \int_0^{\infty} e^{-y}y^{\alpha-1}dy$

$for\ integer\ n, \Gamma(n) = (n-1)!$

$X \sim \chi^2(k) \leftrightarrow \Gamma(\frac{k}{2},\frac{1}{2})$

Chi-squared distribution is the sum of squares of i.i.d. standard normal variables

$X \sim t_v \leftrightarrow \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)}(1 + \frac{x^2}{v})^{\frac{-(v+1)}{2}}, \forall x$

$If\ X_1,\ldots,X_n\ are\ i.i.d.\ N(\mu,\sigma^2), then\ \frac{X-\mu}{S/\sqrt{n}} \sim t_{n-1}$

$In\ general, \frac{N(0,1)}{\sqrt{\frac{\chi^2(k)}{k}}} \sim t_k$

(c) $y_i = \beta_0' + \beta_1 x_{i1}' + \beta_2 x_{i2}' + \epsilon_i \quad i=1,\ldots,n$

$x_{i1}' = x_{i1} - \bar{x}_1, \quad x_{i2}' = x_{i2} - \bar{x}_2$

$X_c = [1_n, x_1', x_2']$

Column vector of all 1's of length n

$X_c^T X_c = \begin{bmatrix} 1_n^T \\ x_1'^T \\ x_2'^T \end{bmatrix}[1_n\ x_1'\ x_2'] = \begin{bmatrix} 1_n^T 1_n & 1_n^T x_1' & 1_n^T x_2' \\ 1_n^T x_1' & \sum x_{i1}'^2 & \sum x_{i1}' x_{i2}' \\ 1_n^T x_2' & \sum x_{i1}' x_{i2}' & \sum x_{i2}'^2 \end{bmatrix}$

$= \begin{bmatrix} n & 0 & 0 \\ 0 & S_{x_1x_1} & S_{x_1x_2} \\ 0 & S_{x_1x_2} & S_{x_2x_2} \end{bmatrix}$

$X \sim F(d_1,d_2) \leftrightarrow f(x) = \frac{\left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}}}{B\left(\frac{d_1}{2},\frac{d_2}{2}\right)}x^{\frac{d_1}{2}-1}(1 + \frac{d_1}{d_2}x)^{-(d_1+d_2)/2}$

$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$

$If\ X_1 \sim \chi^2(d_1)\ and\ X_2 \sim \chi^2(d_2)\ are\ independent\ then\ \frac{X_1/d_1}{X_2/d_2} \sim F(d_1,d_2)$

$If\ X_1 \sim \Gamma(\alpha_1,\lambda_1)\ and\ X_2 \sim \Gamma(\alpha_2,\lambda_2)\ are\ independent\ then\ \frac{\lambda_1 X_1/\alpha_1}{\lambda_2 X_2/\alpha_2} \sim F(2\alpha_1, 2\alpha_2)$

$(X_c^T X_c)^{-1} = \begin{bmatrix} \frac{1}{n} & 0 & 0 \\ 0 & \multicolumn{2}{c}{\begin{pmatrix} S_{x_1x_1} & S_{x_1x_2} \\ S_{x_1x_2} & S_{x_2x_2} \end{pmatrix}^{-1}} \\ 0 & & \end{bmatrix}$

$X_c^T y = \begin{bmatrix} \sum y_i \\ \sum x_{i1} y_i \\ \sum x_{i2} y_i \end{bmatrix}$

$\hat{\beta} = (X_c^T X_c)^{-1} X_c^T y = \begin{pmatrix} 11.1 \\ 2.8606 \\ 1.3197 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0' \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2 = 3.76888$

Proof: Cov(ybar, beta1head) = 0:

Proof: $d_i = \frac{(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad \sum x_i = n\bar{x}$

$Cov(\bar{y},\hat{\beta}_1) = Cov\left(\frac{1}{n}\sum_{i=1}^{n}y_i, \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right) = Cov\left(\frac{1}{n}\sum_{i=1}^{n}y_i, \sum_{i=1}^{n}d_i y_i\right)$

$= \frac{1}{n}\sum_{i=1}^{n}Cov(y_i, d_i y_i) = \frac{1}{n}\sum_{i=1}^{n}d_i \underbrace{Cov(y_i,y_i)}_{Var(y_i)=\sigma^2} = \frac{\sigma^2}{n}\sum_{i=1}^{n}d_i = 0$

$Cov(e_i,\hat{\beta}_1) = Cov(y_i - \hat{y}_i, \hat{\beta}_1) = Cov(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i), \hat{\beta}_1)$

$= Cov(y_i - (\bar{y} - \hat{\beta}_1\bar{x}) - \hat{\beta}_1 x_i, \hat{\beta}_1) = Cov(y_i - (\bar{y} + \hat{\beta}_1(x_i-\bar{x})), \hat{\beta}_1)$

$= Cov(y_i - (\bar{y} + d_i\sum_{j=1}^{n}(x_j-\bar{x})y_j), \sum_{i=1}^{n}d_i y_i)$

$= \underbrace{Cov(y_i, \sum_{i=1}^{n}d_i y_i)}_{d_i Cov(y_i,y_i)} - \underbrace{Cov(\bar{y}, \sum_{i=1}^{n}d_i y_i)}_{=0} - \underbrace{Cov(d_i\sum_{j=1}^{n}(x_j-\bar{x})y_j, \sum_{j=1}^{n}d_j y_j)}_{d_i\sum_{j=1}^{n}(x_j-\bar{x})d_j Cov(y_j,y_j)}$

$= d_i\sigma^2$

$d_i\sigma^2\sum_{j=1}^{n}\frac{(x_j-\bar{x})(x_j-\bar{x})}{\sum_{j=1}^{n}(x_j-\bar{x})^2}$

$= d_i\sigma^2$

this one

$= d_i\sigma^2 - d_i\sigma^2 = 0$

Tutorial 4 比一堆 statistics 計 beta1,2hat, ai)beta0=2, 首先 create 一個新嘅 YNew=y-2,然後計 X'X, X'YNew, (X'X)^-1, 要留意 beta0=2 就變 nointercept model 唔洗加一行 1 比佢, 最尾 betahat=(X'X)-1X'YNew

aii). Find SSE: e=(I-H)YNew, H =X(X'X)^(-1)X. 而 SSE=||e||^2=yNew^T(I-H)(I-H)yNew 最尾會變成 Sum(yi-2)^2-betahat'*X'*YNew,代數即可.b).2beta1=beta2.model: $y_i = \beta_0 + \beta_1(x_{i1} + 2x_{i2}) + \epsilon_i$ ,$\overline{x'} = \overline{x_1} + 2\overline{x_2}$,beta1=Sx'y/Sx'x'=硬爆變翻 xi only,=$\frac{S_{x_1y}+2S_{x_2y}}{S_{x_1x_1}+4S_{x_2x_2}+4S_{x_1x_2}} = -0.842850, \widehat{\beta_2} = 2\widehat{\beta_1}, \widehat{\beta_0} = \bar{y} - \widehat{\beta_1}x_1 + \widehat{\beta_\iota x_2}$. (c). $\beta_0, \beta_1, \beta_2$ are all unknown.