$x_1 = 3,\ x_2 = 8,\ x_3 = 9$ 15 October

$\underset{\sim}{x}_0 = (1, 3, 8, 9)$

Example 5: Example in Multiple Linear Regression

The percent survival of a certain type of animal semen after storage was measured at various combinations of concentrations of three materials used to increase chance of survival. The data are as follows:

y (% survival)	x_1 (weight %)	x_2 (weight %)	x_3 (weight %)
25.5	1.74	5.30	10.80
31.2	6.32	5.42	9.40
25.9	6.22	8.41	7.20
38.4	10.52	4.63	8.50
18.4	1.19	11.60	9.40
26.7	1.22	5.85	9.90
26.4	4.10	6.62	8.00
25.9	6.32	8.72	9.10
32.0	4.08	4.42	8.70
25.2	4.15	7.60	9.20
39.7	10.15	4.83	9.40
35.7	1.72	3.12	7.60
26.5	1.70	5.30	8.20

$\hat{\mu}_{y|x_0} = \hat{\beta}_0 + 3\hat{\beta}_1 + 8\hat{\beta}_2 + 9\hat{\beta}_3$

$= (1\ 3\ 8\ 9)\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{pmatrix}$

$= 24.2232$

p.17 $\hat{\sigma}^2 = 4.298$

Summary statistics:

$$\sum_{i=1}^{13} y_i = 377.5 \quad \sum_{i=1}^{13} y_i^2 = 11,400.15 \quad \sum_{i=1}^{13} x_{i1} = 59.43$$

$$\sum_{i=1}^{13} x_{i2} = 81.82 \quad \sum_{i=1}^{13} x_{i3} = 115.40 \quad \sum_{i=1}^{13} x_{i1}^2 = 394.7255$$

$$\sum_{i=1}^{13} x_{i2}^2 = 576.7264 \quad \sum_{i=1}^{13} x_{i3}^2 = 1035.96 \quad \sum_{i=1}^{13} x_{i1}y_i = 1877.567$$

$$\sum_{i=1}^{13} x_{i2}y_i = 2246.661 \quad \sum_{i=1}^{13} x_{i3}y_i = 3337.78 \quad \sum_{i=1}^{13} x_{i1}x_{i2} = 360.6621$$

$$\sum_{i=1}^{13} x_{i1}x_{i3} = 522.078 \quad \sum_{i=1}^{13} x_{i2}x_{i3} = 728.31 \quad n = 13$$

Origin model

$$\begin{pmatrix} 13 & 59.43 & 81.82 & 115.40 \\ 59.43 & 394.7255 & 360.6621 & 522.078 \\ 81.82 & 360.6621 & 576.7264 & 728.31 \\ 115.40 & 522.078 & 728.31 & 1035.96 \end{pmatrix}^{-1} = \boxed{\begin{pmatrix} 8.06479 & -0.0825927 & -0.0941951 & -0.790527 \\ -0.0825927 & 0.00847982 & 0.00171669 & 0.00372002 \\ -0.0941951 & 0.00171669 & 0.0166294 & -0.00206331 \\ -0.790527 & 0.00372002 & -0.00206331 & 0.0886013 \end{pmatrix}}$$

$1 + \underset{\sim}{x}_0^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{x}_0 = 1 + (1\ 3\ 8\ 9)\left(\quad\right)\begin{pmatrix}1\\3\\8\\9\end{pmatrix} = 1 + 0.1267 = 1.1267$

Or

95% C.I. for $\mu_{y|x_0}$

$= 24.2232 \pm t_{\alpha/2, 9}\sqrt{4.298 \times (1+0.1267)}$ 2.262

$= (22.5541, 25.8923)$

Centered model

$$(\underset{\sim}{X}_c^T\underset{\sim}{X}_c)^{-1} = \begin{pmatrix} 13 & 0 & 0 & 0 \\ 0 & 123.039 & -13.3812 & -5.4775 \\ 0 & -13.3812 & 61.7639 & 2.0002 \\ 0 & -5.4775 & 2.0002 & 11.5631 \end{pmatrix}^{-1}$$

$$= \boxed{\begin{pmatrix} 0.0769231 & 0 & 0 & 0 \\ 0 & 0.00847981 & 0.00171669 & 0.00371998 \\ 0 & 0.00171669 & 0.0166294 & -0.00206338 \\ 0 & 0.00371998 & -0.00206338 & 0.0886011 \end{pmatrix}}$$

95% prediction interval

$= (19.2459, 29.2)$

p.35

$\Rightarrow \hat{\beta}_0 = 39.1574,\ \hat{\beta}_1 = 1.0161,\ \hat{\beta}_2 = -1.8616,\ \hat{\beta}_3 = -0.3433.$

Centered model : $y_i = \beta_0' + \beta_1(\underset{\uparrow\, 3}{x_{i1}} - \bar{x}_1) + \beta_2(\underset{\uparrow\, 8}{x_{i2}} - \bar{x}_2) + \beta_3(\underset{\uparrow\, 9}{x_{i3}} - \bar{x}_3) + e_i$

$\underset{\sim}{x}_0^T = (1, 3-\bar{x}_1, 8-\bar{x}_2, 9-\bar{x}_3)$

$1 + \underset{\sim}{x}_0^T(\underset{\sim}{X}_c^T\underset{\sim}{X}_c)^{-1}\underset{\sim}{x}_0 = 1 + (1, 3-\bar{x}_1, 8-\bar{x}_2, 9-\bar{x}_3)\left(\quad\right)\begin{pmatrix}1\\3-\bar{x}_1\\8-\bar{x}_2\\9-\bar{x}_3\end{pmatrix} = 1 + 0.1267$

$= 1.1267$

Coefficient of determination

$$0 \le R^2 = \frac{\text{Reg S.S.}}{\text{total S.S.}} \le 1$$

Total S.S. = Reg S.S. + Res S.S.

$$= \frac{\text{Total S.S.} - \text{Res S.S.}}{\text{total S.S.}}$$

$$= 1 - \frac{\text{Res S.S.}}{\text{total S.S.}}$$

— linear ~~st~~ relationship

For $p = 1$

$$R^2 = \frac{\text{Reg S.S}}{\text{total S.S.}} = \frac{\hat{\beta}_1 S_{x_1 y}}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$\hat{\beta}_1 = \frac{S_{x_1 y}}{S_{x_1 x_1}}$$

$$\text{Res S.S.} = \underset{\sim}{Y}^T \underset{\sim}{Y} - \hat{\underset{\sim}{\beta}}^T \underset{\sim}{X}^T \underset{\sim}{Y} = S_{yy} - \underbrace{(\hat{\beta}_1 S_{x_1 y} + \hat{\beta}_p S_{x_p y})}_{\text{Reg S.S.}}$$

$$= \frac{\frac{S_{x_1 y}}{S_{x_1 x_1}} S_{x_1 y}}{S_{yy}} = \frac{S_{x_1 y}^2}{S_{x_1 x_1} S_{yy}}$$

$$= \left(\boxed{\frac{S_{x_1 y}}{\sqrt{S_{x_1 x_1}} \sqrt{S_{yy}}}} \right)^2$$

r - simple correlation coeff.

$-1 \le r \le 1$

For any p $\quad \sqrt{R^2}$

- $p = 1$
- $p > 1$ multiple correlation coeff.

Example in p.17 (multiple linear regression)

$$R^2 = 1 - \frac{38.68}{438.17} = 0.9117$$

91.17 % of variation in y can be explained by the linear regression model simple linear regression

— β_1 = regression coeff. of X_1

or slope of linear regression

$Y \sim N$

$X_1 \sim N$

population simple correlation coefficient

$$\rho = \frac{\operatorname{cov}(X_1, Y)}{\sqrt{Var(X_1)}\sqrt{Var(Y)}} \qquad X, Y \text{ r.v.}$$

$$\underline{f(y, x_1)} = \underline{f(y \mid x_1)}\,\boxed{f(x_1)}$$

$$- \quad \hat{\beta}_1 = \frac{S_{x_1 y}}{S_{x_1 x_1}} \qquad\qquad \hat{\rho} = r = \frac{S_{x_1 y}}{\sqrt{S_{x_1 x_1}}\sqrt{S_{yy}}}$$

Model : $Y = \beta_0 + \beta_1 X_1 + e$

Assume

$$\begin{pmatrix} X_1 \\ Y \end{pmatrix} \sim MN\left(\begin{pmatrix} \mu_{x_1} \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_{x_1}^2 & \boxed{\beta_1 \sigma_{x_1}^2} \\ \beta_1 \sigma_{x_1}^2 & \sigma_y^2 \end{pmatrix} \right)$$

$\operatorname{cov}(X_1, Y)$ $= \rho\,\sigma_{x_1}\sigma_y$

$$\begin{aligned} \operatorname{cov}(X_1, Y) &= \operatorname{cov}(X_1, \beta_0 + \beta_1 X_1 + e) \\ &= \beta_1 Var(X_1) \quad \text{— assume } X_1 \text{ \& } e \text{ are indep.} \\ &= \beta_1 \sigma_{x_1}^2 \end{aligned}$$

$$\Rightarrow \beta_1 \sigma_{x_1}^2 = \rho\, \sigma_{x_1} \sigma_y$$

$$\Rightarrow \boxed{\rho} = \frac{\boxed{\beta_1}\sigma_{x_1}}{\sigma_y} \qquad \text{If } \beta_1 = 0 \Rightarrow \rho = 0$$

measures the linear relationship between X_1 & Y

$H_0 : \beta_1 = 0$

From linear regression model

$$\Rightarrow t = \frac{\hat{\beta}_1 - 0}{\text{s.e. of } \hat{\beta}_1} = \frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{S_{x_1 x_1}}}$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{x_1 x_1}}$$

$$\frac{S_{x_1 y}}{S_{x_1 x_1}} = \hat{\beta}_1$$

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\text{Res S.S.}}{n-2} \\ &= \frac{1}{n-2}\left[S_{yy} - \hat{\beta}_1 S_{x_1 y} \right] \\ &= \frac{1}{n-2}\left(S_{yy} - \frac{S_{x_1 y}^2}{S_{x_1 x_1}} \right) \end{aligned}$$

$$t = \frac{\dfrac{S_{x_1 y}}{S_{x_1 x_1}}}{\sqrt{\boxed{\left(S_{yy} - \dfrac{S_{x_1 y}^2}{S_{x_1 x_1}}\right)}\Big/(n-2)} \cdot \sqrt{\dfrac{1}{S_{x_1 x_1}}}}$$

$$\Rightarrow S_{yy}\left(1 - \boxed{\frac{S_{x_1 y}^2}{S_{x_1 x_1} S_{yy}}}\right) = S_{yy}(1 - r^2)$$

$$\| \quad r^2$$

$$= \frac{\dfrac{S_{x_1 y}}{\sqrt{S_{x_1 x_1}}}}{\sqrt{S_{yy}(1-r^2)/(n-2)}} = \boxed{\frac{S_{x_1 y}}{\sqrt{S_{x_1 x_1}}\sqrt{S_{yy}}}}\frac{1}{\sqrt{1-r^2}} * \sqrt{n-2} \qquad (\# = r)$$

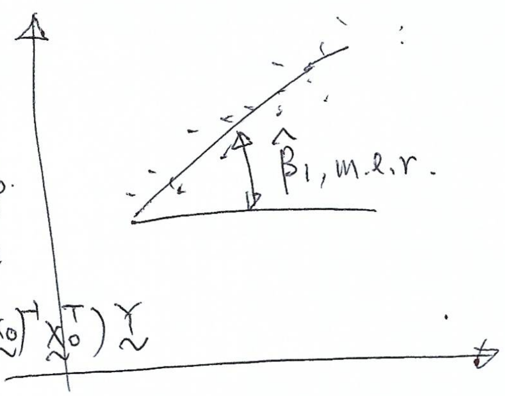
$$= \frac{\sqrt{n-2}\, r}{\sqrt{1-r^2}} \sim t_{(n-2)}$$

$$\Rightarrow \text{Reject } H_0: \rho = 0 \text{ if } \frac{\sqrt{n-2}\,|r|}{\sqrt{1-r^2}} > t_{\alpha/2}(n-2)$$

Section 7 Added variable plot (partial regression plot) (p.39)

Model : $y_i = \beta_0 + \boxed{\beta_1} x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$

y' = residual of y on the other indep. variables except x_1



x_1' = residual of x_1 on the other indep. variables (x_2, x_3)

$$\hat{\underset{\sim}{e}}_0 = (\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T \underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)\underset{\sim}{Y}$$

$$\underset{\sim}{Y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{e}$$

$$\underset{\sim}{X} = \left(\underbrace{\begin{matrix} x_{11} \\ \vdots \\ x_{n1} \end{matrix}}_{\underset{\sim}{x}_1} \;\; \underbrace{\begin{matrix} 1 & x_{12} & x_{13} \\ \vdots & \vdots & \vdots \\ 1 & x_{n2} & x_{n3} \end{matrix}}_{\underset{\sim}{X}_0}\right) \qquad \underset{\sim}{\beta} = \begin{pmatrix} \beta_1 \\ \beta_0 \\ \beta_2 \\ \beta_3 \end{pmatrix} \quad \begin{matrix} \\ - \beta_* \end{matrix}$$

Note that :
$$\begin{aligned}\hat{\underset{\sim}{e}} &= \underset{\sim}{Y} - \hat{\underset{\sim}{Y}} \\ &= \underset{\sim}{Y} - \underset{\sim}{X}\hat{\underset{\sim}{\beta}} \\ &= \underset{\sim}{Y} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\underset{\sim}{Y} \\ &= (\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T)\underset{\sim}{Y}\end{aligned}$$

y' = residual of y on the other indep. variables except x_1

$$\ni \hat{\underset{\sim}{e}}_0 = (\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)\underset{\sim}{Y} \leftarrow$$

$$(\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)\underset{\sim}{X}_0$$

$$=(\underset{\sim}{X}_0 - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T\underset{\sim}{X}_0) = \underset{\sim}{0} \quad *$$

x_1' = residual of x_1 on the other indep. variables, x_1 & x_2

$$\text{Model} = x_1 = \gamma_0 + \gamma_1 x_2 + \gamma_2 x_3 + e_1$$

$$\hat{\underset{\sim}{e}}_1 = (\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)\underset{\sim}{X}_1 \leftarrow$$

question 1: Is $\hat{\beta}_1$ the regression coefficient of $\underset{y'}{\hat{e}_0}$ on $\underset{x_1'}{\hat{e}_1}$?

$$\hat{\underset{\sim}{e}}_0 = (\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)(\beta_1\underset{\sim}{X}_1 + \underset{\sim}{X}_0\underset{\sim}{\beta} + \underset{\sim}{e})$$

$$= \beta_1\underbrace{\boxed{(\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)\underset{\sim}{X}_1}}_{\hat{\underset{\sim}{e}}_1} + \underbrace{\boxed{(\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)\underset{\sim}{X}_0}}_{= 0}\underset{\sim}{\beta} + (\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)\underset{\sim}{e}$$

$$= \boxed{\beta_1}\hat{\underset{\sim}{e}}_1 + (\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)\underset{\sim}{e}$$

question 2: $\boxed{\text{slope in the plot of } \hat{e}_0 \text{ on } \hat{e}_1} = \boxed{\hat{\beta}_{1,m.l.e}}$?

$$y' = \hat{e}_0 = (\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)\underset{\sim}{Y}$$

$$\begin{pmatrix} A & B \\ B^T & D \end{pmatrix}^{-1} = \begin{pmatrix} & \\ & \end{pmatrix}$$

$$x_1' = \hat{e}_1 = (\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)\underset{\sim}{X}_1$$

$$(\underset{\sim}{X}_1^T\underbrace{\boxed{(\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)^T(\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)}}_{= (\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)}\underset{\sim}{X}_1)^{-1}$$

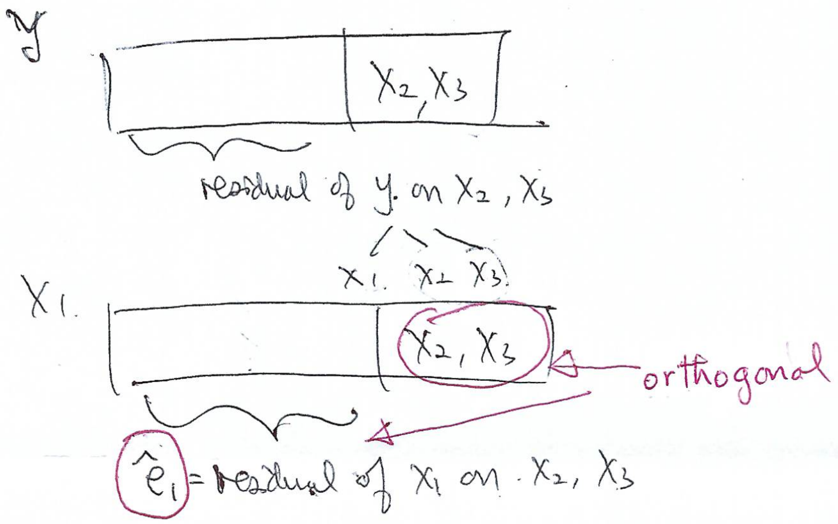
$$(\underset{\sim}{X}_1^T(\underset{\sim}{I} - \underset{\sim}{X}_0(\underset{\sim}{X}_0^T\underset{\sim}{X}_0)^{-1}\underset{\sim}{X}_0^T)\underset{\sim}{Y})$$

$\hat{\beta}_{1,mle} \leftarrow$ $\left(\begin{array}{c|c} A & B \\ \hline B^T & D \end{array}\right)^{-1} = \begin{pmatrix} A^{-1}+FE^{-1}F^T & -FE^{-1} \\ -E^{-1}F^T & E^{-1} \end{pmatrix}$

where $E = D - B^T A^{-1} B$

$F = A^{-1} B$

$$\underset{\sim}{X}^T \underset{\sim}{X} = \left(\begin{array}{c|c} \underbrace{\underset{\sim}{X}_0^T \underset{\sim}{X}_0}_{A} & \underbrace{\underset{\sim}{X}_0^T \underset{\sim}{X}_1}_{B} \\ \hline \underbrace{\underset{\sim}{X}_1^T \underset{\sim}{X}_0}_{B^T} & \underbrace{\underset{\sim}{X}_1^T \underset{\sim}{X}_1}_{D} \end{array}\right) \qquad \underset{\sim}{X}^T \underset{\sim}{Y} = \begin{pmatrix} \underset{\sim}{X}_0^T \underset{\sim}{Y} \\ \hline \underset{\sim}{X}_1^T \underset{\sim}{Y} \end{pmatrix}$$



$\hat{\beta}_{1,mle}$ = increment of residual of y on X_2, X_3 (or decreases) when residual of X_1 on X_2, X_3 increases by one-unit.

= increment of y if X_1 increases (decreases) one unit adjusted by X_2, X_3

- controlling the effect of X_2, X_3
- removing the effect of X_2, X_3

simple linear regression between X_1 & y

$\Downarrow$

simple correlation coefficient

$$r = \frac{S_{x_1 y}}{\sqrt{S_{x_1x_1}}\sqrt{S_{yy}}}$$

simple linear regression between y' and X_1'

$\Downarrow$

~~partial~~ simple correlation coeff. between y' and X_1'

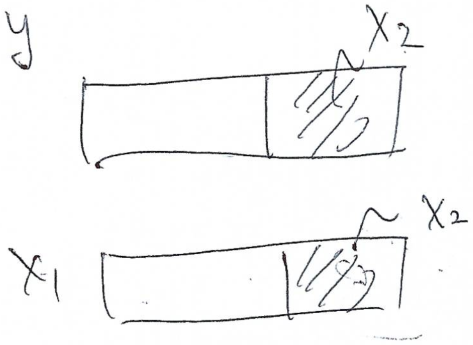
$$r = \frac{S_{x_1' y'}}{\sqrt{S_{x_1'x_1'}}\sqrt{S_{y'y'}}}$$

$$= \frac{r_{x_1y} - \overset{=0}{\boxed{r_{x_1x_2}}}\,\overset{=0}{\boxed{r_{x_2y}}}}{\sqrt{1-r_{x_2y}^2}\sqrt{1-r_{x_1x_2}^2}}$$

simple ~~[illegible]~~ ~~regress~~ correlation coefficient

$\Rightarrow$ simple correlation coeff of X_1 & y

Called as partial correlation coefficient between X_1 & y



multiple correlation coeff.

measure the linear ~~reg~~ relationship between y and $\underset{\sim}{X} = (X_1, \ldots, X_p)$

partial correlation coeff.

measure the linear relationship between y and X_1 after adjusted by the other indep. variables

simple correlation coeff

measure the linear relationship between y & X_1