Examples for 09/22/2015

The (normal) multiple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{p-1} x_{i\,p-1} + e_i, \qquad i = 1, 2, \ldots, n,$$

where e_i's are independent $\text{Normal}(0, \sigma^2)$ (i.i.d. $\text{Normal}(0, \sigma^2)$).

$\beta_0, \beta_1, \beta_2, \ldots, \beta_{p-1}$ and σ^2 are unknown model parameters.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1\,p-1} \\ 1 & x_{21} & x_{22} & \ldots & x_{2\,p-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{n\,p-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \ldots \\ e_n \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e} \qquad\qquad \hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \qquad \mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}.$$

---

1. Consider the following data set:

x_1	x_2	y
0	1	11
11	5	15
11	4	13
7	3	14
4	1	0
10	4	19
5	4	16
8	2	8

Consider the model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i, \qquad i = 1, \ldots, 8.$$

where e_i's are i.i.d. $N(0, \sigma_e^2)$.

Then $\mathbf{X}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} 8 & 56 & 24 \\ 56 & 496 & 200 \\ 24 & 200 & 88 \end{bmatrix}$, $\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \begin{bmatrix} 96 \\ 740 \\ 336 \end{bmatrix}$,

$$\mathrm{C} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{bmatrix} 0.7125 & -0.025 & -0.1375 \\ -0.025 & 0.025 & -0.05 \\ -0.1375 & -0.05 & 0.1625 \end{bmatrix} = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix}.$$

a) Obtain the least-squares estimates $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$.

$\mathrm{SYY} = \sum(y - \overline{y})^2 = 240$, $\mathrm{SSResid} = \sum(y - \hat{y})^2 = 76.4$,

b) Perform the significance of the regression test at a 5% level of significance.

c) Test $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$ at $\alpha = 0.10$. Find the p-value.

d) Test $H_0: \beta_2 = 0$ vs. $H_a: \beta_2 \neq 0$ at $\alpha = 0.05$. Find the p-value.

e) Construct a 90% prediction interval for the value of Y at $x_{01} = 2$ and $x_{02} = 3$.

f) Construct a 90% confidence interval for the mean response at $x_{01} = 8$ and $x_{02} = 5$.

```
> x1 = c( 0,11,11, 7, 4,10, 5, 8)
> x2 = c( 1, 5, 4, 3, 1, 4, 4, 2)
> y =  c(11,15,13,14, 0,19,16, 8)
>
> fit = lm(y ~ x1 + x2)
>
> summary(fit)

Call:
lm(formula = y ~ x1 + x2)

Residuals:
   1    2    3    4    5    6    7    8 
 2.9 -3.0 -0.6  2.0 -5.3  4.7 -1.8  1.1 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)   3.7000     3.2995   1.121   0.3131  
x1           -0.7000     0.6181  -1.133   0.3088  
x2            4.4000     1.5758   2.792   0.0383 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 3.909 on 5 degrees of freedom
Multiple R-Squared: 0.6817,     Adjusted R-squared: 0.5543 
F-statistic: 5.353 on 2 and 5 DF,  p-value: 0.05717 

> predict.lm(fit,data.frame(x1=2,x2=3),interval=c("prediction"),level=0.90)
      fit      lwr      upr
[1,] 15.5 5.080037 25.91996
> predict.lm(fit,data.frame(x1=8,x2=5),interval=c("confidence"),level=0.90)
      fit      lwr      upr
[1,] 20.1 13.99869 26.20131
```