Wen Wang
Department of Biostatistics
University of Michigan

Jan 21, 2013

HOMEWORK — #1

Problem 1.

(a)

$$L(\beta)=\prod_{i=1}^{n}\beta^{\alpha}e^{-y_i\beta}$$
$$l(\beta)=\sum_{i=1}^{n}\alpha\log(\beta)-y_i\beta$$
$$U(\beta)=\sum_{i=1}^{n}\frac{\alpha}{\beta}-y_i$$

(b)

$$J(\beta)=\frac{n\alpha}{\beta^2}$$

Problem 2.

(a)

$$L(\theta)=\prod_{i=1}^{n}\frac{1}{\sigma}e^{-\frac{(y_i-\beta_1x_i)^2}{2\sigma^2}}$$
$$l(\theta)=-n/2\log\sigma^2-\sum_{i=1}^{n}\frac{(y_i-\beta_1x_i)^2}{2\sigma^2}$$
$$U(\theta)=\begin{pmatrix}\frac{-\beta_1\sum_{i=1}^n x_i^2+\sum_{i=1}^n x_iy_i}{\sigma^2}\\ -n/(2\sigma^2)+\sum_{i=1}^n\frac{(y_i-\beta_1x_i)^2}{2\sigma^4}\end{pmatrix}$$
$$\hat{\beta}_1=\frac{\sum_{i=1}^n x_iy_i}{\sum_{i=1}^n x_i^2}$$
$$\hat{\sigma}^2=\frac{\sum_{i=1}^n(y_i-\hat{\beta}_1x_i)^2}{n}$$

For LSE:

$$\hat{\beta}_1=\frac{\sum_{i=1}^n x_iy_i}{\sum_{i=1}^n x_i^2}$$
$$\hat{\sigma}^2=\frac{\sum_{i=1}^n(y_i-\hat{\beta}_1x_i)^2}{n-1}$$

The estimations for θ are different.

(b)

$$J(\theta)=\begin{pmatrix}\frac{\sum_{i=1}^n x_i^2}{\sigma^2} & \frac{-\beta_1\sum_{i=1}^n x_i^2+\sum_{i=1}^n x_iy_i}{\sigma^4}\\ \frac{-\beta_1\sum_{i=1}^n x_i^2+\sum_{i=1}^n x_iy_i}{\sigma^4} & -n/(2\sigma^4)+\frac{\sum_{i=1}^n(y_i-\beta_1x_i)^2}{\sigma^6}\end{pmatrix}$$
$$I(\theta)=\begin{pmatrix}\frac{\sum_{i=1}^n x_i^2}{\sigma^2} & 0\\ 0 & n/(2\sigma^4)\end{pmatrix}$$
$$Var(\hat{\theta})=\begin{pmatrix}\frac{\sigma^2}{\sum_{i=1}^n x_i^2} & 0\\ 0 & (2\sigma^4)/n\end{pmatrix}$$

Problem 3.

(a) X is the total number of males and N is the total number of infants.

$$L(\theta) = \theta^X (1-\theta)^{N-X},$$
$$l(\theta) = X \log\theta + (N-X)\log(1-\theta),$$
$$U(\theta) = \frac{X}{\theta} - \frac{N-X}{1-\theta},$$
$$I(\theta) = E\left(\frac{X}{\theta^2} + \frac{N-X}{(1-\theta)^2}\right) = \frac{N}{\theta} + \frac{N}{1-\theta}.$$

(b)

$$U(\theta) = \frac{X}{\theta} - \frac{N-X}{1-\theta} = 0 \Rightarrow \hat\theta = \frac{X}{N} = 239/492 = 0.48577,$$
$$J(\hat\theta) = \frac{X}{\hat\theta^2} + \frac{N-X}{(1-\hat\theta)^2} = \frac{N^2}{X} + \frac{N^2}{N-X} > 0.$$
$$Var(\hat\theta) = \frac{\theta(1-\theta)}{N}$$
$$\hat{V}ar(\hat\theta) = \frac{\hat\theta(1-\hat\theta)}{N} = 0.005077$$
$$\frac{\hat\theta - \theta}{\sqrt{\theta(1-\theta)/N}} \sim N(0,1)$$
$$\left(\hat\theta - 1.96\sqrt{\hat\theta(1-\hat\theta)/N},\, \hat\theta + 1.96\sqrt{\hat\theta(1-\hat\theta)/N}\right)$$
$$= (0.4416,\, 0.5299).$$

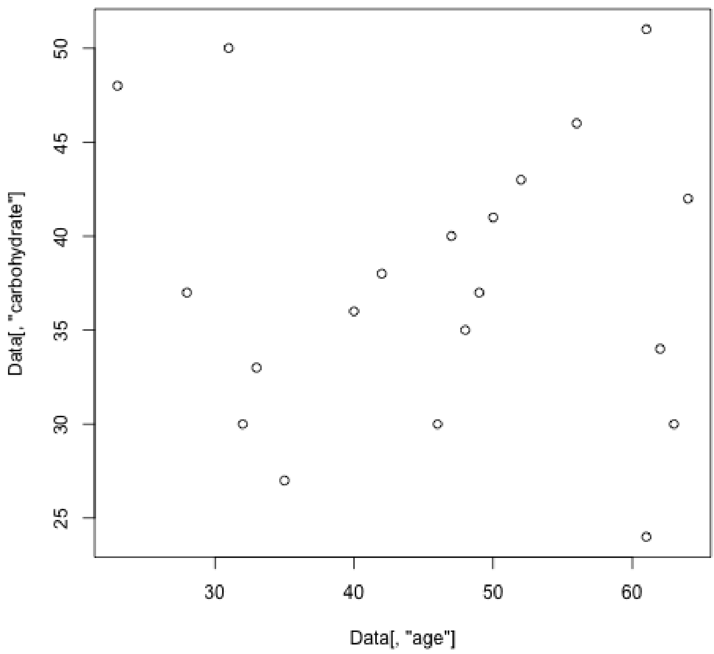
(c)

$$\frac{U(\hat\theta_H)}{\sqrt{I(\hat\theta_H)}} \sim N(0,1)$$
$$\left|\frac{U(\hat\theta_H)}{\sqrt{I(\hat\theta_H)}}\right| = 0.6312 < 1.96$$

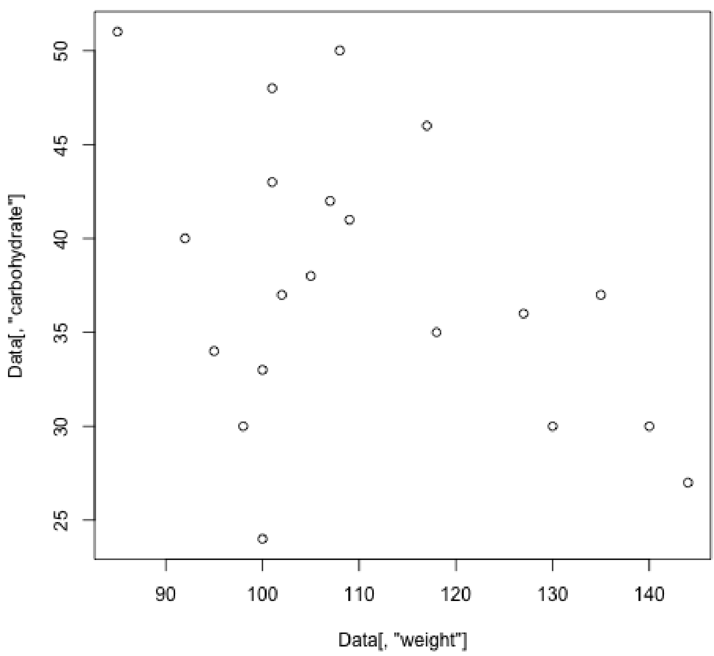
fail to reject H_0 at $\alpha = 0.05$.

Problem 4.

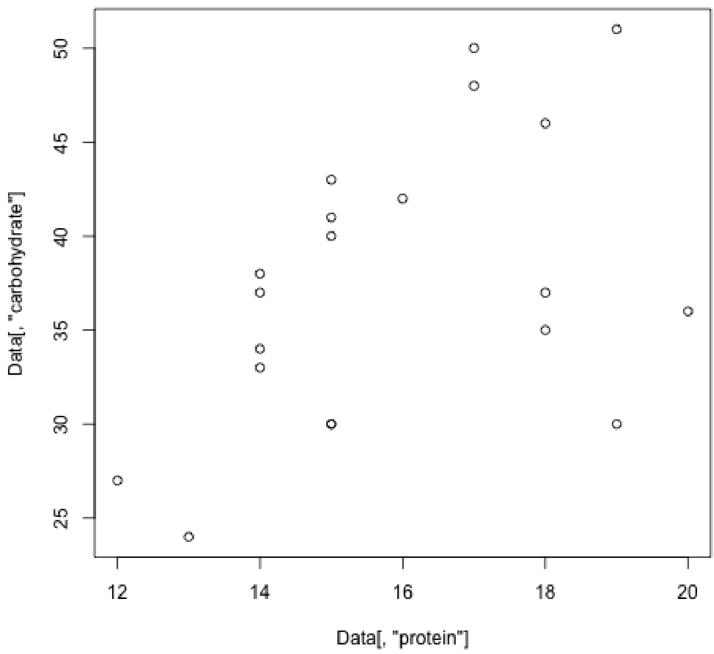
(a)



Data[, "carbohydrate"]
Data[, "age"]



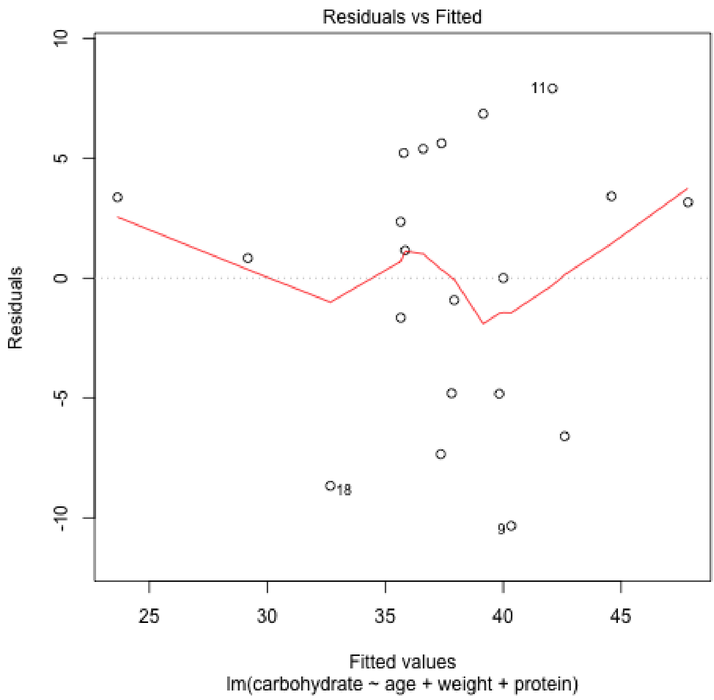
Data[, "carbohydrate"]
Data[, "weight"]

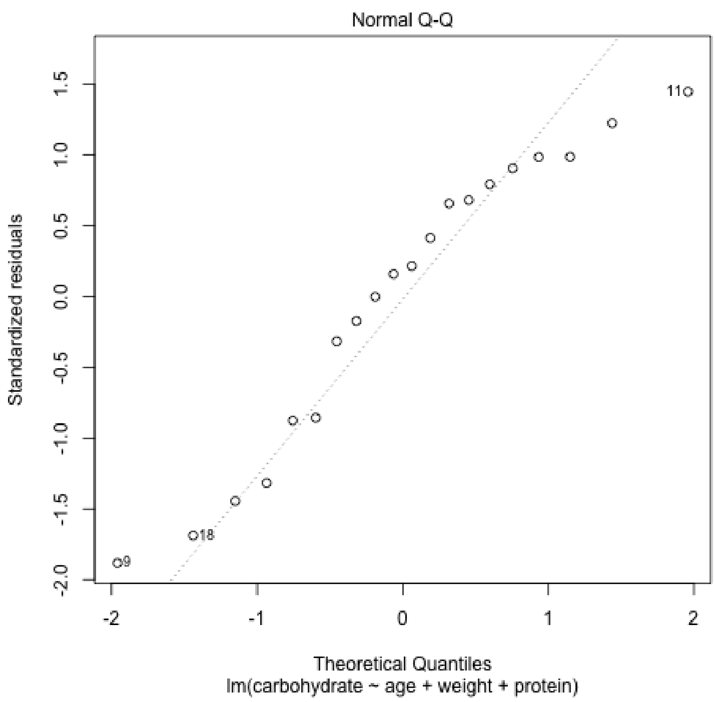


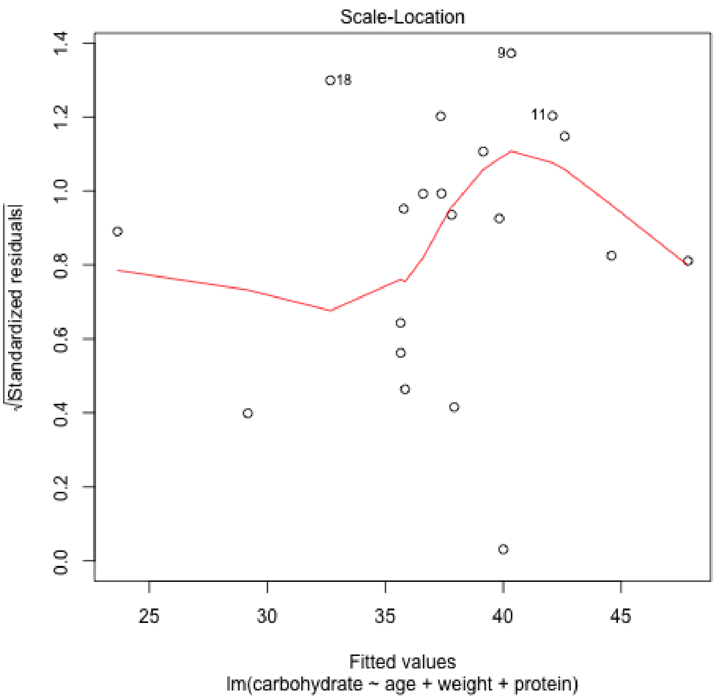
Age's linear relation with carbohydrate seems vague because it is hard to tell whether they are correlated positively or negatively. For weight and protein, they seem to have linear relation with carbohydrate.

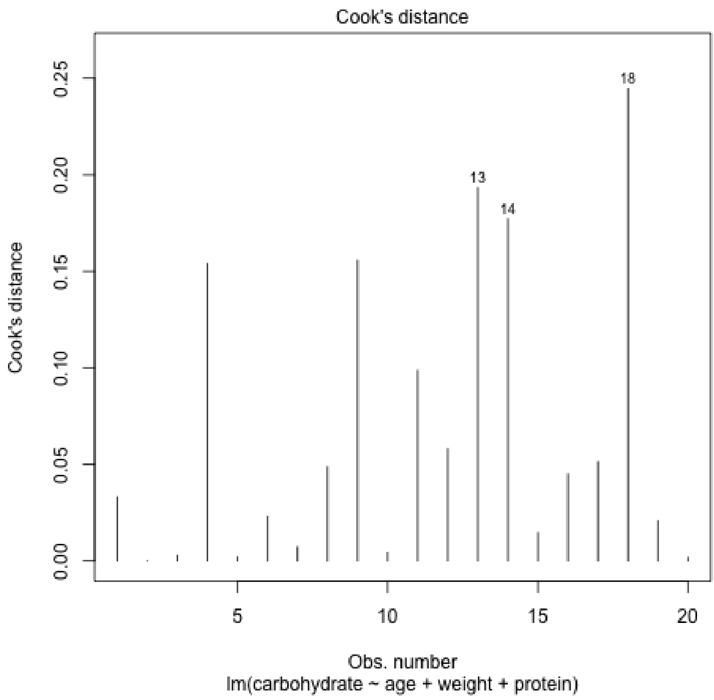
(b)

```
Call:
lm(formula = carbohydrate ~ age + weight + protein, data = Data)

Residuals:
     Min       1Q   Median       3Q      Max
-10.3424  -4.8203   0.9897   3.8553   7.9087

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 36.96006   13.07128   2.828  0.01213 *
age         -0.11368    0.10933  -1.040  0.31389
weight      -0.22802    0.08329  -2.738  0.01460 *
protein      1.95771    0.63489   3.084  0.00712 **
---

Residual standard error: 5.956 on 16 degrees of freedom
Multiple R-squared:  0.4805,    Adjusted R-squared:  0.3831
F-statistic: 4.934 on 3 and 16 DF,  p-value: 0.01297
```


Residuals vs Fitted
Residuals
10
5
0
-5
-10
11
18
9
25
30
35
40
45
Fitted values
lm(carbohydrate ~ age + weight + protein)


Normal Q-Q
Standardized residuals
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
11
18
9
-2
-1
0
1
2
Theoretical Quantiles
lm(carbohydrate ~ age + weight + protein)





The two tails of the qqplot deviate from the straight line, so the normal assumption is not appropriate. The standardized residual - fitted value plot shows a non-constant trend, however, the sample size is too small and the fitted values are too concentrated to confirm this trend. The linear trend seems unreliable.

(c)

```
Call:
lm(formula = carbohydrate ~ age + protein, data = Data)

Residuals:
     Min       1Q   Median       3Q      Max
-11.2692  -5.9968   0.9902   5.7952   9.5474
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.08848   12.16239   1.241   0.2316
age         -0.09167    0.12818  -0.715   0.4842
protein      1.68189    0.73693   2.282   0.0356 *
---

Residual standard error: 7.002 on 17 degrees of freedom
Multiple R-squared:  0.2372,    Adjusted R-squared:  0.1475
F-statistic: 2.643 on 2 and 17 DF,  p-value: 0.1001


Call:
lm(formula = carbohydrate ~ protein, data = Data)

Residuals:
     Min       1Q   Median       3Q      Max
-12.4979  -5.9829   0.9019   4.8870  10.6620

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.4787    11.4435   1.090   0.2899
protein       1.5800     0.7131   2.216   0.0399 *
---

Residual standard error: 6.907 on 18 degrees of freedom
Multiple R-squared:  0.2143,    Adjusted R-squared:  0.1706
F-statistic: 4.909 on 1 and 18 DF,  p-value: 0.03986

Analysis of Variance Table

Model 1: carbohydrate ~ age + protein
Model 2: carbohydrate ~ protein
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     17 833.57
2     18 858.65 -1   -25.079 0.5115 0.4842
```

According to the analysis of variance table, we fail to reject the null hypothesis $\beta_{age} = 0$ at $\alpha = 0.05$.
According to Table 6.5, we have an F-test: $\frac{38.359}{35.489} = 1.08087 \sim F_{1,16}$, p-value is 0.31396>0.05, fail to reject the null hypothesis $\beta_{age} = 0$ at $\alpha = 0.05$.

Submitted by Wen Wang on Jan 21, 2013.