Department of Industrial Engineering & Operations Research

IEOR 165 (Spring 2016)

Homework 2

Due: Friday, Feb 26

Question 1. The following data indicate the gain in reading speed (G) versus the number of weeks in the program (W) of 10 students in a speed-reading program. Suppose we want to build a linear model to predict the speed gain of a student who plans to take the program for 12 weeks.

Number of Weeks (W)	Speed Gain (G) (wds/min)
2	21
3	42
8	102
11	130
4	52
9	105
6	85
5	62
7	90

a) Specify the response and the predictor, and construct the linear model.

Solution. The response is Speed Gain and the predictor is the Number of Weeks. So the linear model is

$$G = b + m \cdot W + \epsilon$$

b) Assume the variance of the gain in reading speed is proportional to the number of weeks in the program, i.e. $\text{Var}(G|W) = \text{Var}(\epsilon|W) \propto W$. Use the weighted least squares to estimate the parameters in the linear model you construct.

Solution. To avoid ambiguity, let x be our predictor and y be our response. By the assumption, $w = \frac{1}{x}$. Then following the formula of weighted least squares, we have

$$\hat{m} = \frac{\overline{wxy} - \overline{wx} \cdot \overline{wy}}{\overline{wx^2} - (\overline{wx})^2} = 12.518$$

$$\hat{b} = \frac{\overline{wy} - \hat{m} \cdot \overline{wx}}{\bar{w}} = 0.257$$

Question 2. Use the semiparametric approach to estimate the parameters for Question 1 (choose $k = 3$)

Solution. In the first step, we compute

$$
\begin{aligned}
\tilde{y}_1 &= \frac{1}{3}(21 + 42 + 52) = 38.33 \\
\tilde{y}_2 &= \frac{1}{3}(42 + 21 + 52) = 38.33 \\
\tilde{y}_3 &= \frac{1}{3}(102 + 90 + 105) = 99 \\
\tilde{y}_4 &= \frac{1}{3}(130 + 102 + 105) = 112.33 \\
\tilde{y}_5 &= \frac{1}{3}(52 + 42 + 62) = 52 \\
\tilde{y}_6 &= \frac{1}{3}(105 + 102 + 130) = 112.33 \text{ or } \frac{1}{3}(105 + 102 + 90) = 99 \\
\tilde{y}_7 &= \frac{1}{3}(85 + 90 + 62) = 79 \\
\tilde{y}_8 &= \frac{1}{3}(62 + 85 + 52) = 66.33 \\
\tilde{y}_9 &= \frac{1}{3}(90 + 85 + 102) = 92.33
\end{aligned}
$$

Then applying OLS to $(x_i, \tilde{y}_i)$ gives

$$\hat{m} = 9.78 \quad \hat{b} = 16.90 \qquad \text{or} \qquad \hat{m} = 9.22 \quad \hat{b} = 18.84$$

Question 3. Let $X_1, \ldots, X_n$ be iid from the pdf

$$f(x) = \theta x^{\theta - 1}, \quad 0 \le x \le 1,\ 0 < \theta < \infty$$

Find the MLE of θ.

Solution. The likelihood function is given as

$$L(\theta) = \prod_{i=1}^{n} \theta x_i^{\theta - 1} = \theta^n \big(\prod_{i=1}^{n} x_i\big)^{\theta - 1}$$

The log likelihood function is

$$l(\theta) = \log L(\theta) = n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log x_i$$

Set the first derivative equal to zero:

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i = 0$$

$$\Rightarrow \quad \hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \log x_i}$$

Question 4. Let $X_1, \dots, X_n$ be iid with pdf

$$f_\theta(x) = \frac{1}{\theta}, \quad 0 \le x \le \theta, \quad \theta > 0$$

Find the MLE of θ.

Solution. The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I(0 \le x_i \le \theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } \theta \ge \max\{x_1, \dots, x_n\} \\ 0 & o.w. \end{cases}$$

We can see that $\frac{1}{\theta^n}$ is strictly decreasing in θ, so we will choose $\hat{\theta} = \max\{x_1, \dots, x_n\}$.

Question 5. The chlorine residual (C) in a swimming pool at various times after being cleaned (T) is as given:

Time (hr)	Chlorine Residual (pt/million)
2	1.80
4	1.50
6	1.45
8	1.42
10	1.38
12	1.36

Assume the following relationship

$$C \approx a \exp(-bT)$$

What would you predict for the chlorine residual 15 hours after a cleaning?

Solution. Take log on both sides:

$$\log C \approx \log a - bT$$

So we derive a linear model (in approximation) with $y = \log C$ and $x = T$. Let $\beta_0 = \log a$ and $\beta_1 = -b$, we get the OLS estimates as

$$\hat{\beta}_0 = 0.56 \quad \hat{\beta}_1 = -0.024$$

So at $T = 15$,

$$\log(C(15)) = 0.56 - 0.024 \times 15 = 0.2$$

Thus $C(15) = e^{0.2} = 1.22$.