Name: LIU Hengche
(e.g., Chan Siu Pang)

Student Number: 57854327
(e.g., 12345678)

CS2402- Lecture 11 - In-Class Exercises

Q1. Below are the winning times in men's 10,000 m track in the Olympic Games between the years 1984 and 2000. Times are in decimal form, for example, 28.4 = 28 min and 24 s.

27.8, 27.4, 27.8, 27.1, 27.3

Let the years be the x values and the times the y values.
(a) Find the estimated regression line.
(b) Predict the winning time in the Olympic year 2088.

Solution:

9960

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$
1984	27.8	-8	0.32	-2.56	64
1988	27.4	-4	-0.08	0.32	16
1992	27.8	0	0.32	0	0
1996	27.1	4	-0.38	-1.52	16
2000	27.3	8	-0.18	-1.44	64
9960 / 1992	137.4 / 27.48			-5.2	160

$$\beta = \frac{-5.2}{160} = -0.0325$$

$$\alpha = \bar{y} - \beta\bar{x}$$

$$= 27.48 + 0.0325 \times 1992$$

$$= 27.48 + 64.74 = 92.22$$

$$\therefore \hat{y} = -0.0325 + 92.22$$

Q2. Below is a data set of 4 observations of the daily closing prices for the two U.S. stock market indices, Dow Jones Industrial Average and Nasdaq Composite Index, chosen at random from among the years of 1971 and 2003.

11.5 -42.5 -54.5 85.5

Dow: 887, 833, 821, 961

Nasdaq: 108, 86, 74, 95

17.25 -4.75 -16.7 4.25

198.375

201.875

910.15

363.375

Let X be Dow Jones Industrial Average and Y be Nasdaq Composite Index. Compute the sample correlation coefficient of X and Y. ($\bar{X} = 875.5$, $\bar{Y} = 90.75, S_x = 63.82, S_y = 14.36$)

$$\Sigma(x_i-\bar{x})(y_i-\bar{y}) = \Sigma x_i y_i - n\bar{x}\bar{y}$$

$$\frac{\Sigma \frac{1}{n-1}(x-\bar{x})(y-\bar{y})}{S_x S_y} = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

$$r = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2}\sqrt{\Sigma(y_i-\bar{y})^2}} = \frac{198.375 + 201.875 + 910.15 + 363.375}{\sqrt{11.5^2 + 42.5^2 + 54.5^2 + 85.5^2}\sqrt{17.25^2 + 4.75^2 + 16.7^2 + 4.\ldots}}$$

$$= \frac{1673.775}{\sqrt{132.25 + 1806.25 + 2970.25 + 7310.25}\sqrt{291.5625 + 22.5625 + 278.89 + 18.0625}}$$

$$= \frac{1673.775}{\sqrt{12219} \cdot \sqrt{1611.0775}} = \frac{1673.775}{110.5376 \times 24.8410} = 0.6096$$

Q3. Suppose that X is a discrete random variable with the following probability function, where $0 \leq \theta \leq 1$ is a parameter.

i	0	1	2	3
$P(X = i)$	$2\theta/3$	$\theta/3$	$2(1-\theta)/3$	$(1-\theta)/3$

The following 10 independent observations were taken from such a distribution: (3,0,2,1,3,2,1,0,2,1). What is the maximum likelihood estimate of θ?

$$\left(\frac{1-\theta}{3}\right)^2 \left(\frac{2(1-\theta)}{3}\right)^3 \left(\frac{\theta}{3}\right)^3 \left(\frac{2}{3}\theta\right)^2$$

$$(1-\theta)^2 (1-\theta)^3 \theta^3 \theta^2$$

$$(1-\theta)^5 \theta^5$$

$$\log (1-\theta)^5 + \log \theta^5$$

$$5\log(1-\theta) + 5\log\theta$$

$$-\frac{5}{1-\theta} + \frac{5}{\theta}$$

$$\frac{5}{\theta} = \frac{5}{1-\theta}$$

$$5 - 5\theta = 5\theta$$

$$\theta = 0.5$$

Q4. Suppose x is a discrete random variable with probability mass function:

$$P(x) = \theta(1-\theta)^{x-1}$$

for $x = 1, 2, \ldots$ and $0 < \theta < 1$. The n independent observations $x_1, x_2, \ldots, x_n$ are taken from such a distribution. What is the maximum likelihood estimate of θ?

$$\Sigma\ \theta(1-\theta)^{x-1}$$

$$\log\theta + (x_i - 1)\log(1-\theta)$$

$$\Sigma = \log(1-\theta)(x_1 + \cdots + x_n - n) + n\log\theta$$

$$= (x_1 + \cdots + x_n)\log(1-\theta) - n\log(1-\theta) + n\log\theta$$

$$= (x_1 + \cdots + x_n)\log(1-\theta) + n\log\frac{\theta}{1-\theta}$$

$$(x_1 + \cdots + x_n)\left(-\frac{1}{1-\theta}\right) + \frac{n}{1-\theta} + \frac{n}{\theta}$$

$$-\frac{n\mu}{1-\theta} + \frac{n}{1-\theta} + \frac{n}{\theta}$$

$$\frac{n}{\theta} + \frac{n(1-\mu)}{1-\theta} = 0$$

$$\frac{\cancel{n}}{\theta} = \frac{(\mu-1)n}{1-\theta}$$

$$1-\theta = (\mu-1)\theta$$

$$1-\theta = \mu\theta - \theta$$

$$\theta = \frac{1}{\mu}$$