Statistics I, Fall 2012
Suggested Solution for Homework 02

Ling-Chieh Kung
Department of Information Management
National Taiwan University

1. (a) The frequency distribution is listed in Table 1 and the histogram is depicted in Figure 1.

Class	Frequency
$[20, 30)$	8
$[30, 40)$	26
$[40, 50)$	31
$[50, 60)$	67
$[60, 70)$	32
$[70, 80)$	28
$[80, 90)$	6
$[90, 100)$	2

Table 1: The frequency distribution for Problem 1a.

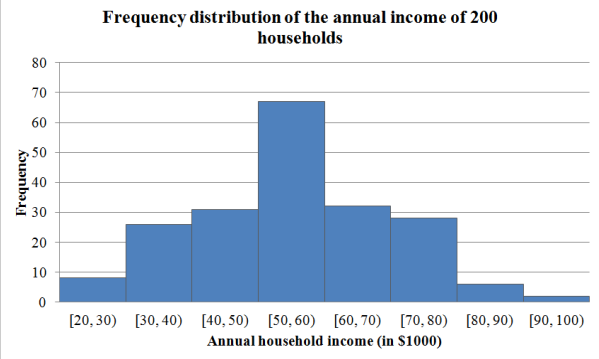

Figure 1: The histogram for Problem 1a.

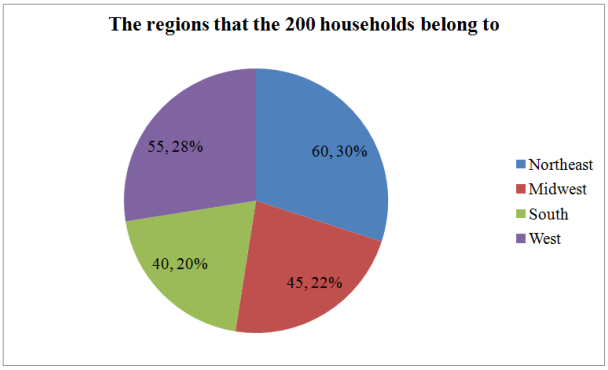

Figure 2: The pie chart for Problem 1c.

(b) Unimodal.

(c) The pie chart is depicted in Figure 2.

(d) The bar chart is depicted in Figure 3

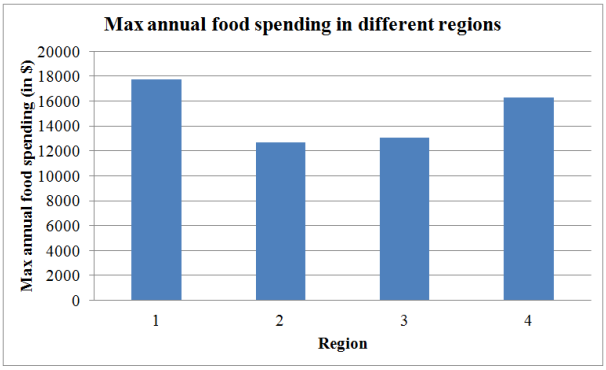

Figure 3: The bar chart for Problem 1d.

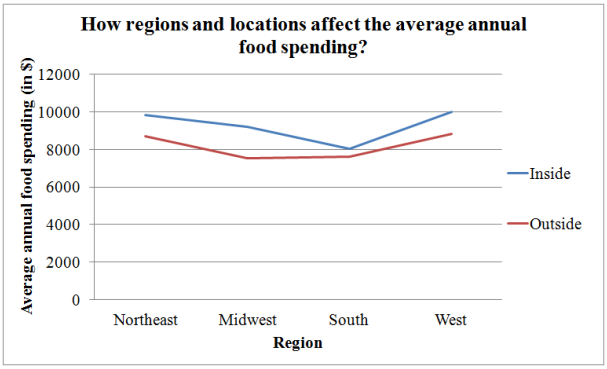

Figure 4: The frequency polygons for Problem 1f.

(e) It is not reasonable to draw a pie chart for the several maximum values as these values are not proportions to an aggregate value. Therefore, a pie chart is not suitable for the data in Part (d).

(f) What we need to do is to illustrate four comparisons (one for each region) in one figure. One good way of doing so is to draw two frequency polygons in one figure, as depicted in Figure 4. As we may see from the figure, it is clear that the average household food spending inside the metropolitan area dominates that outside the metropolitan area in all the four regions.

(g) The scatter plot of income and household food spending is depicted in Figure 5. As we may see, there is approximately a linear relationship between these two variables. Moreover, when a household has a higher income level, it tends to spend more on food, which is reasonable.

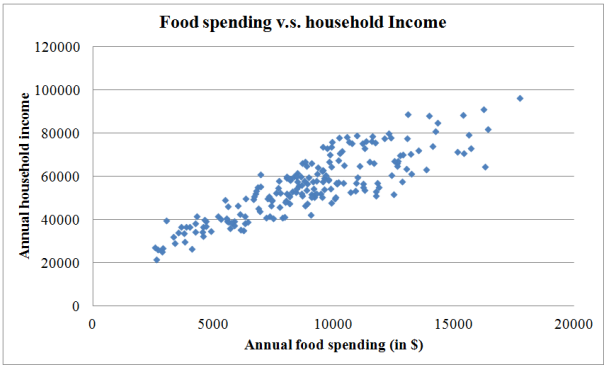

Figure 5: The scatter plot for Problem 1g.

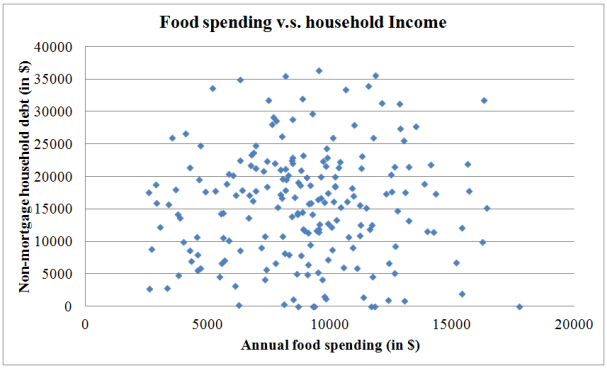

Figure 6: The scatter plot for Problem 1h.

(h) The scatter plot of non-mortgage debts and household food spending is depicted in Figure 6. As we may see, there is no obvious relationship between these two variables. We may suggest that the level of non-mortgage debts does not affect how much a household would spend on food.

2. (a) The mode is \$0, the mean is \$15,604.16, the variance is 73,677,143.95 square dollars, and the standard deviation is \$8,583.54. Note that because these data form a sample, in calculating the variance and standard deviation we need to use $n - 1 = 199$ as the denominator. If you use MS Excel, the functions to use are MODE(), AVERAGE(), VAR(), and STDEV().

(b) In calculating the first quartile, note that $\frac{1}{4}(200) = 50$ is an integer, so the first quartile is the average of the 50th and 51st term in the ascending order. Therefore, we have

$$Q_1 = \frac{9018 + 9250}{2} = 9134$$

and the first quartile is \$9,134. Similarly, the third quartile is the average of the 150th and 151st term:

$$Q_3 = \frac{21238 + 21323}{2} = 21280.5.$$

The third quartile is \$21,280.5. The interguartile range is $21280.5 - 9134 = \$12{,}146.5$.

(c) For each value of k, we find $\mu \pm k\sigma$ and the proportion of values within the range $[\mu - k\sigma, \mu + k\sigma]$. This proportion is then shown to be larger than $1 - \frac{1}{k^2}$, the bound provided by Chebyshev's theorem. The results are listed in Table 2.

3. Table 3 summarizes the relevant calculations. According to the table, the mean absolute deviation is 4.69 and the variance is 29. The standard deviation is $\sqrt{29} = 5.39$.

4. (a) The first-order derivative of σ^2 with respect to x_j is

$$\frac{\partial}{\partial x_j}\sigma^2 = \frac{\partial}{\partial x_j}\sum_{i=1}^{N}\frac{(x_i - \mu)^2}{N} = \frac{1}{N}\frac{\partial}{\partial x_j}(x_j - \mu)^2 = \frac{2}{N}(x_j - \mu).$$

k	$[\mu - k\sigma, \mu + k\sigma]$	Number of values in the range	Proportion of values in the range	$1 - \frac{1}{k^2}$
1.5	$[2728.85, 28479.47]$	168	0.84	0.56
2	$[-1562.92, 32771.24]$	193	0.965	0.75

Table 2: Verification of Chebyshev's theorem for Problem 2c.

x_i	$x_i - \mu$	$\lvert x_i - \mu \rvert$	$(x_i - \mu)^2$
6	-8	8	64
10	-4	4	16
12	-2	2	4
15	1	1	1
19	5	5	25
22	8	8	64
Average		4.67	29

Table 3: Calculations for the MAD and variance for Problem 3.

(b) The first-order derivative of σ^2 with respect to μ is

$$\begin{aligned}\frac{\partial}{\partial \mu}\sigma^2 = \frac{\partial}{\partial \mu}\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{N} &= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial}{\partial \mu}(x_i-\mu)^2\right] \\ &= \frac{1}{N}\sum_{i=1}^{N}\Big[2(x_i-\mu)(-1)\Big] = -\frac{2}{N}\sum_{i=1}^{N}(x_i-\mu).\end{aligned}$$

The second-order derivative of σ^2 with respect to μ is

$$\frac{\partial^2}{\partial \mu^2}\sigma^2 = \frac{\partial}{\partial \mu}\left[-\frac{2}{N}\sum_{i=1}^{N}(x_i-\mu)\right] = -\frac{2}{N}\sum_{i=1}^{N}\frac{\partial}{\partial \mu}(x_i-\mu) = -\frac{2}{N}\sum_{i=1}^{N}(-1) = 2.$$

(c) The first-order derivative of σ with respect to x_j is

$$\begin{aligned}\frac{\partial}{\partial x_j}\sigma = \frac{\partial}{\partial x_j}\sqrt{\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{N}} &= \frac{1}{2}\left[\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{N}\right]^{-\frac{1}{2}}\left[\frac{\partial}{\partial x_j}\sigma^2\right] \\ &= \frac{1}{2\sigma}\left[\frac{2}{N}(x_j-\mu)\right] = \frac{1}{N\sigma}(x_j-\mu).\end{aligned}$$

In your answer, you may also replace the σ in the denominator by $\sqrt{\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{N}}$.

5. First, note that

$$\begin{aligned}\sum_{i=1}^{N}(x_i-\mu)^2 &= \sum_{i=1}^{N}(x_i^2 - 2\mu x_i + \mu^2) = \sum_{i=1}^{N}x_i^2 - 2\mu\sum_{i=1}^{N}x_i + \mu^2\sum_{i=1}^{N}1 \\ &= \sum_{i=1}^{N}x_i^2 - 2N\mu^2 + N\mu^2 = \sum_{i=1}^{N}x_i^2 - N\mu^2,\end{aligned}$$

where in the third equality we have applied $\mu = \frac{\sum_{i=1}^{N}x_i}{N}$. We then have

$$\sigma^2 = \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{N} = \frac{1}{N}\left(\sum_{i=1}^{N}x_i^2 - N\mu^2\right) = \frac{\sum_{i=1}^{N}x_i^2}{N} - \mu^2$$