Statistics and Data Analysis, Fall 2014
Suggested Solution for the Midterm Exam

Instructor: Ling-Chieh Kung
Department of Information Management
National Taiwan University

1. (a) The histogram is in Figure 1. The class having the highest frequency is $[175, 185)$, whose frequency is 12.
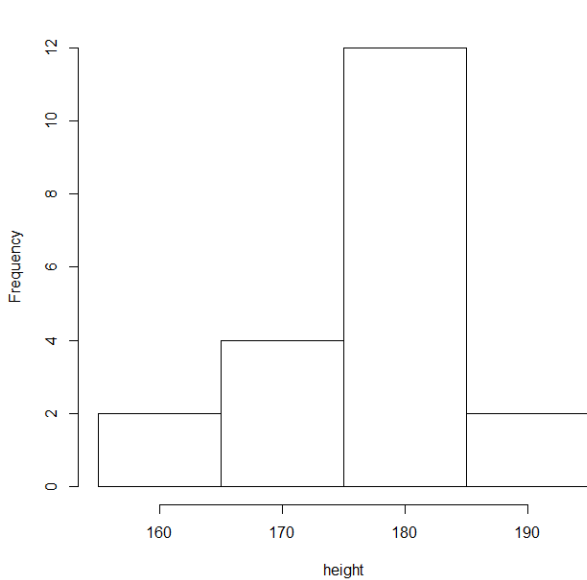

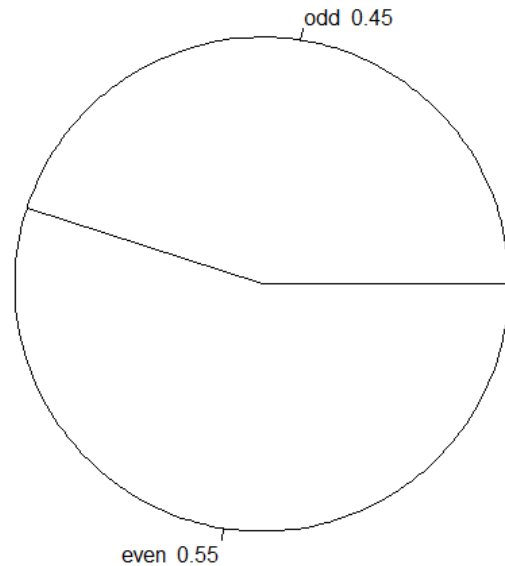Figure 1: Histogram for Problem 1a.

Figure 2: Pie chart for Problem 1b.

   (b) The pie chart is in Figure 2.

   (c) The mean is 176.15, the median is 177.5, and the mode is 180 (the midpoint of the highest-frequency class).

2. (a) The mean is 10784, the median is 8866, and the sample variance is 43681808.

   (b) The mean is 5200, the median is 2926, and the sample variance is 29327873.

   (c) The sample coefficients of variation of mill is $\frac{\sqrt{43681808}}{10784} = 0.6128$. The sample coefficients of variation of fresh food is $\frac{\sqrt{29327873}}{5200} = 1.0414$. The variable of fresh food sales has higher variability.

   (d) The correlation coefficient is 0.1292, which means the two variables are weakly positively correlated.

3. The expected earning is 4.5, which is less than the ticket price. Therefore, we should not buy it.

4. (a) 0, 1, 2, and 3.

   (b) $\Pr(X = 0) = 0.8^3 = 0.512$, $\Pr(X = 1) = 0.8^2 \times 0.2 \times 3 = 0.384$, $\Pr(X = 2) = 0.8 \times 0.2^2 \times 3 = 0.096$, and $\Pr(X = 3) = 0.2^3 = 0.008$.

   (c) The distribution is skewed to the right, i.e., positively skewed.

5. (a) The bar chart is in Figure 3.

   (b) The expected value, variance, and standard deviation of X are 6.125, 2.609, and 1.615, respectively.
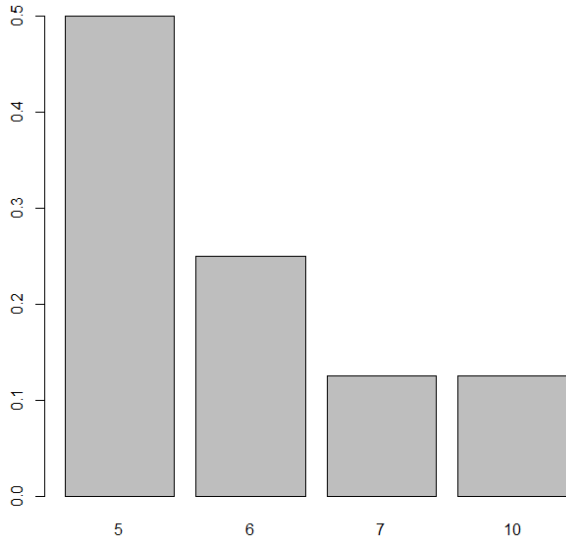
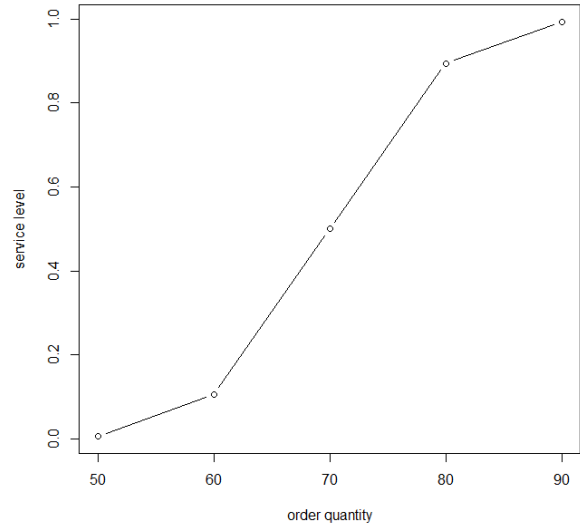6. (a) 0.7734.

Figure 3: Bar chart for Problem 5a.

Figure 4: Scatter plot for Problem 7c.

   (b) 0.4714.

   (c) 0.

7. (a) $\Pr(X \geq 80) = 0.1056$.

   (b) $\Pr(X \leq q) \geq 0.9$ requires $q \geq 80.2524$. Therefore, we should order $q = 81$ units.

   (c) The service levels are 0.0062, 0.1056, 0.5, 0.8944, and 0.9938, respectively. The scatter plot is in Figure 4.

8. (a) $\overline{X} \sim \mathrm{ND}(\mu, 0.1)$.

   (b) $\Pr(\overline{X} < 5.8) = 0.0228$ if $\mu = 6$.

   (c) $\Pr(\overline{X} > a) = 0.05$ requires $a = 6.1645$ if $\mu = 6$.

9. (a) False. Increasing the sample size will make the distribution more centralized. At the same confidence level, we will get a smaller confidence interval.

   (b) True. Increasing the confidence level will enlarge the cover of distribution. and enlarge the confidence interval.

   (c) True. When the sample size is large (above 30), no matter the population distribution is normal or not, we can use the z distribution.

   (d) False. In this case, we should use nonparametric methods.

2