# Statistics and Data Analysis, Fall 2014
# Suggested Solution for Homework 1

Instructor: Ling-Chieh Kung

Department of Information Management

National Taiwan University
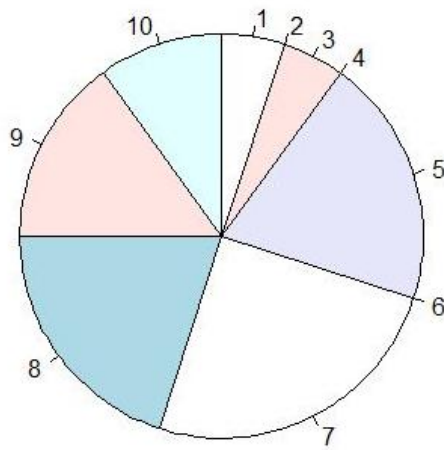
1. (a) –41.1304. (b) 5.607902. (c) 20.89949.

2. (a) True. (b) False. (c) True.

3. (a)   5'10".

    (b)   5'11.2"   5'8.8"   5'10"   6'1.6"   5'8.8"   5'10"   5'6"   5'11.2"

          5'10.8"   5'10"   6'0"   6'0.8"   5'10.8"   6'1.2"   6'0"   5'11.2"

          5'11.6"   5'4.8"   5'8"   5'8.4".

    (c)   1, 2, 3, 5, 6, 7, 8, 9, 10, 13, 16, 17, 18, 19, 20.

    (d)   173.6.

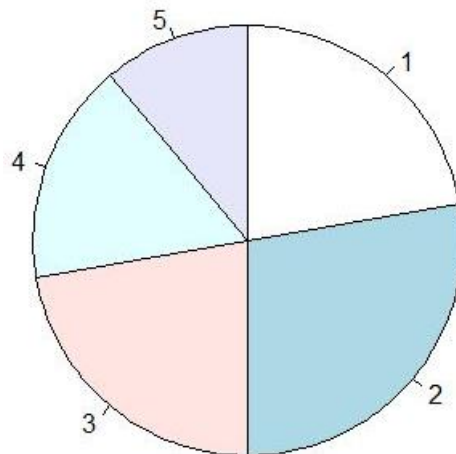4. (a)   Class [175, 180) has the highest frequency, which is 9.

Histogram of teamHeight
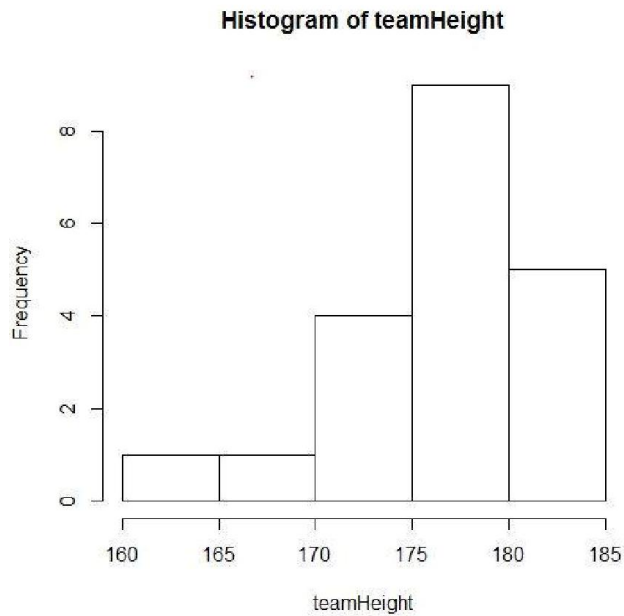
(b) Class [175, 177.5) has the highest frequency, which is 5.

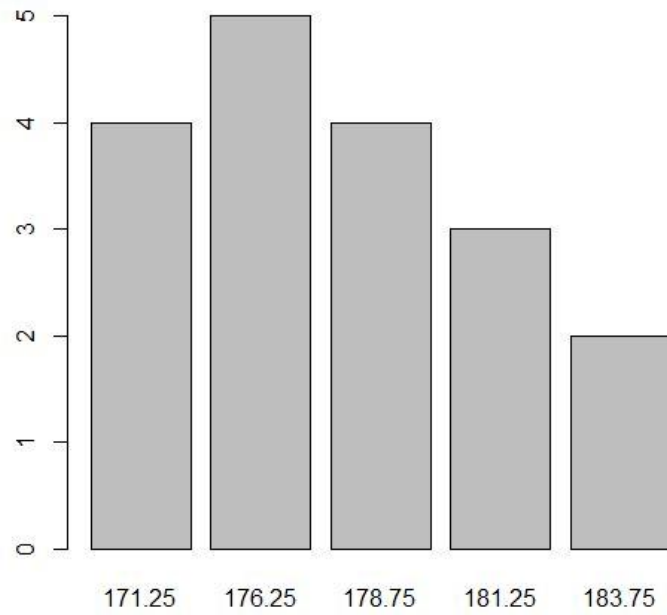**Histogram of teamHeight**
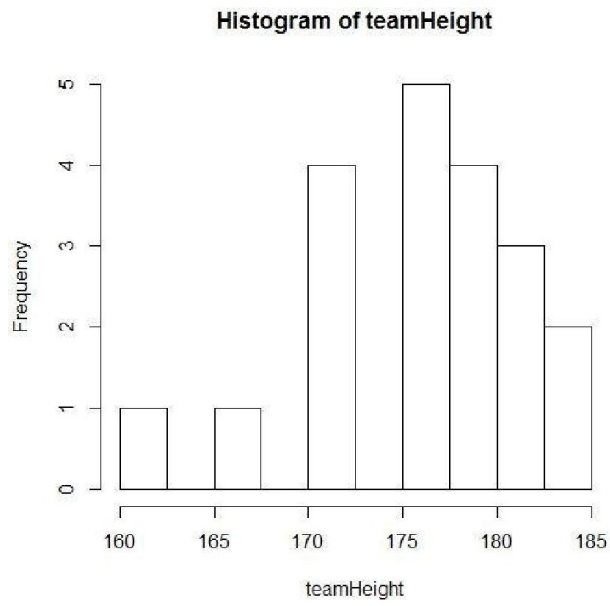
5. (a)

(b)

(c)

6. (a) There is an extreme value which is between 110000 and 120000.

Histogram of ws$Fresh

(b)   230.

(c)

Histogram of ws$Fresh[extreme * -1]
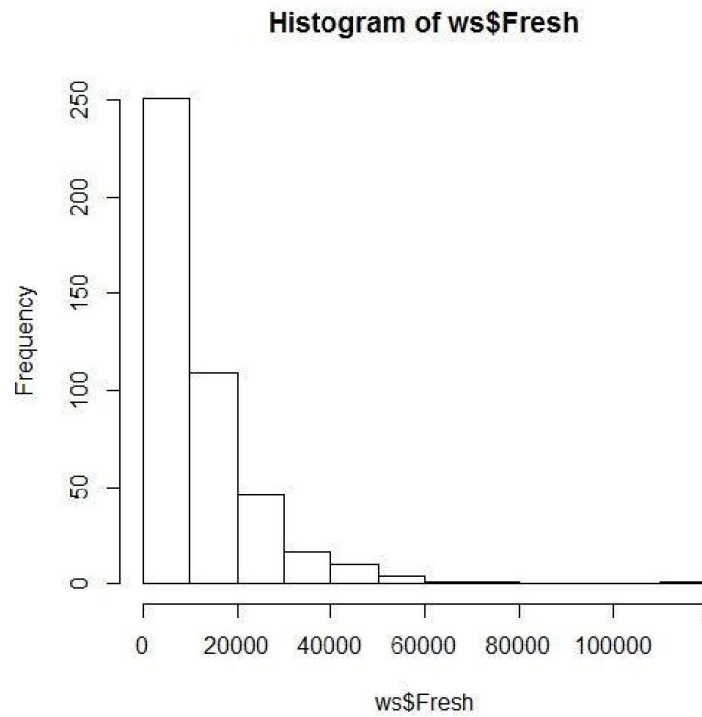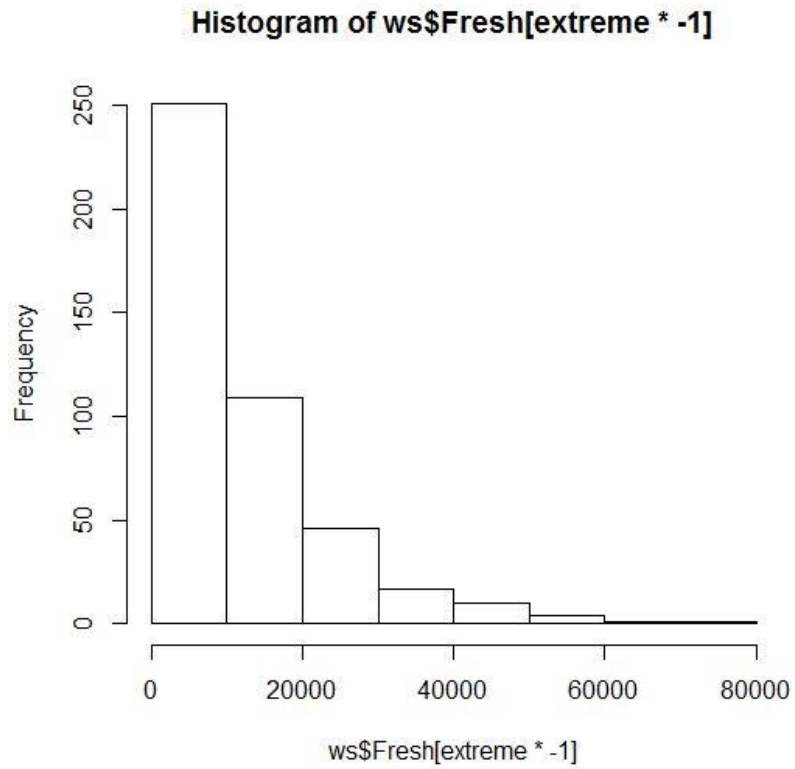
7. (a) False.

(b) Channel 1.

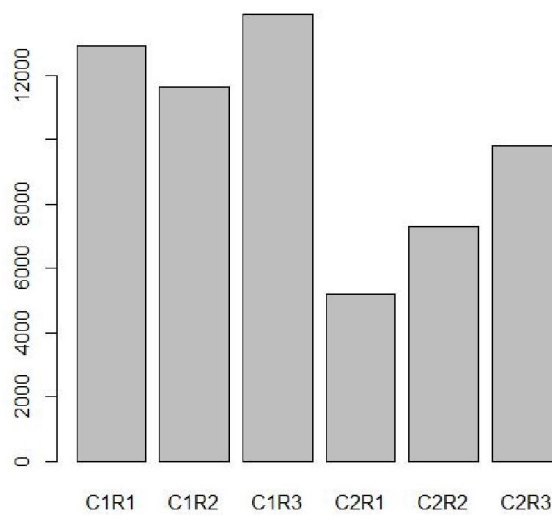(c) Channel = 1 and Region = 3.

(d)

8. (a) Variance: 2545.477.   Standard deviation: 50.45272.

(b) Variance: 281.5379.   Standard deviation: 16.77909.

(c) Sample coefficients of variation for price: 0.1905674. Sample coefficients of variation for size: 0.2270001.

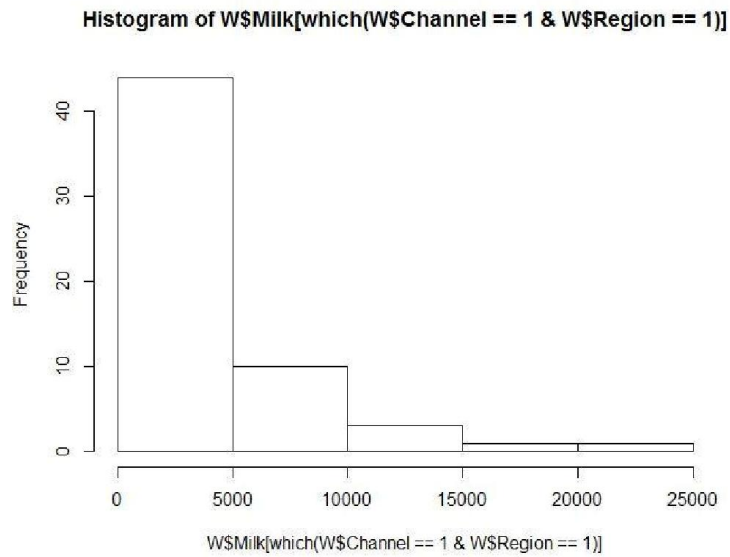The variable **size** has the higher variability.

4

(d)   Variance: 30.7275.   Standard deviation: 5.543239.

9.   (a)   Mean: 3870

Median: 2280

Sample variances: 18475565

(b)

Histogram of W$Milk[which(W$Channel == 1 & W$Region == 1)]

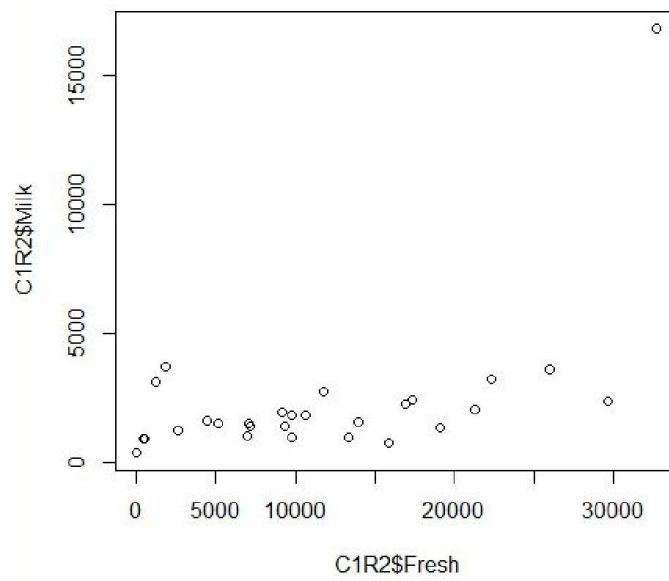(c)   Channel 1 Region 1: –0.03010351.

Channel 1 Region 2: 0.5380095.

Channel 1 Region 3: 0.2912192.

Channel 2 Region 1: 0.1291734.

Channel 2 Region 2: –0.2059987.

Channel 2 Region 3: 0.2761697.

(d)   Scatter plot for Channel 1 Region 2, which has the highest correlation coefficient.

Scatter plot for Channel 2 Region 2, which has the lowest correlation coefficient.