TUTORIAL EXERCISE PACKAGE - 2013

MATH1015 - BIOSTATISTICS

Semester 1	Solution to Problem Set 2	2013

1. (a) The best mark = 99
   (b) The number of students = 21
   (c) The number of students scored 90 = 0
   (d) The range = 99 - 64 = 35

2. Mean=81.71429, SD=10.74310

3. Proportion=7/21=0.33

4. (a) Mean A=13.785714, Mean B=12.933333
   SD A=2.750624, SD B=4.182731
   (b) Standard score of Peter=(15-13.785714)/2.750624=0.4414582
   Standard score of Kim=(15-12.933333)/4.182731=0.4940951; Kim is better

5. D We have $n = 5$, $\sum_i x_i = 25$ and $\sum_i x_i^2 = 153$. Hence the sample variance is

$$s^2 = \frac{1}{n-1}\left[\sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n}\right] = \frac{1}{4}\left[153 - \frac{25^2}{5}\right] = 7$$

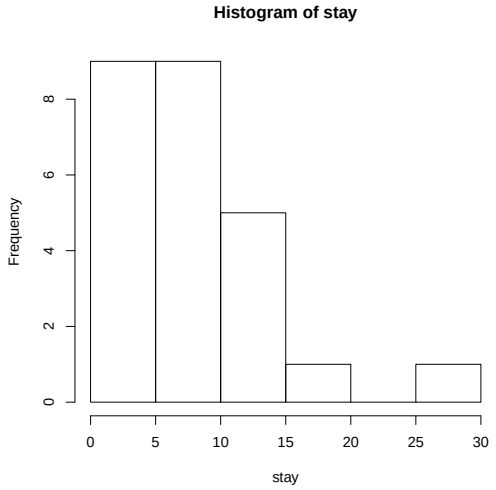
6. (a) `x = scan()` then hit "Enter" and copy and paste the data from the pdf file and hit "Enter" two more times.
   (b)
```
mean(x)
[1] 7.2375
var(x)
[1] 40.41984
sd(x)
[1] 6.35766
```
   (c)
```
stem(x)

 The decimal point is 1 digit(s) to the right of the |

  0 | 011112234444
  0 | 567
  1 | 0013334
  1 |
  2 | 13

stem(x, scale = 2)

  The decimal point is 1 digit(s) to the right of the |

  0 | 011112234444
  0 | 567
  1 | 0013334
  1 |
  2 | 13
```

```
stem(x, scale = 5)

  The decimal point is at the |

   0 | 2
   1 | 00248
   2 | 4
   3 | 478
   4 | 119
   5 | 7
   6 |
   7 | 3
   8 |
   9 | 99
  10 | 8
  11 |
  12 | 89
  13 | 0
  14 | 4
  15 |
  16 |
  17 |
  18 |
  19 |
  20 |
  21 | 0
  22 |
  23 | 0
```

The scale command makes the ratio of the height to the width of the stem-and-leaf plot approximately equal to the scale factor.

7.
```
> dat = read.table(file=url("http://www.maths.usyd.edu.au/math1015/r/hospital.txt"),skip=1)
> dat
   V1 V2 V3 V4   V5 V6 V7 V8 V9
1   1  5 30  2 99.0  8  2  2  1
2   2 10 73  2 98.0  5  2  1  1
3   3  6 40  2 99.0 12  2  2  2
4   4 11 47  2 98.2  4  2  2  2
5   5  5 25  2 98.5 11  2  2  2
6   6 14 82  1 96.8  6  1  2  2
7   7 30 60  1 99.5  8  1  1  1
8   8 11 56  2 98.6  7  2  2  1
9   9 17 43  2 98.0  7  2  2  1
10 10  3 50  1 98.0 12  2  1  2
11 11  9 59  2 97.6  7  2  1  1
12 12  3  4  1 97.8  3  2  2  2
13 13  8 22  2 99.5 11  1  2  2
14 14  8 33  2 98.4 14  1  1  2
15 15  5 20  2 98.4 11  2  1  2
16 16  5 32  1 99.0  9  2  2  2
17 17  7 36  1 99.2  6  1  2  2
18 18  4 69  1 98.0  6  2  2  2
19 19  3 47  1 97.0  5  1  2  1
20 20  7 22  1 98.2  6  2  2  2
21 21  9 11  1 98.2 10  2  2  2
22 22 11 19  1 98.6 14  1  2  2
23 23 11 67  2 97.6  4  2  2  1
24 24  9 43  2 98.6  5  2  2  2
25 25  4 41  2 98.0  5  2  2  1
> stay=dat[,2] #(a)
```

```
> stay
 [1]  5 10  6 11  5 14 30 11 17  3  9  3  8  8  5  5  7  4  3  7  9 11 11  9  4
> fivenum(stay) #(b) the 5 no. are  min Q1 median Q3 max. Median=8
[1]  3  5  8 11 30
> var(stay) #(b)
[1] 32.66667
> sd(stay) #(b)
[1] 5.715476
> hist(stay) #(c)
```

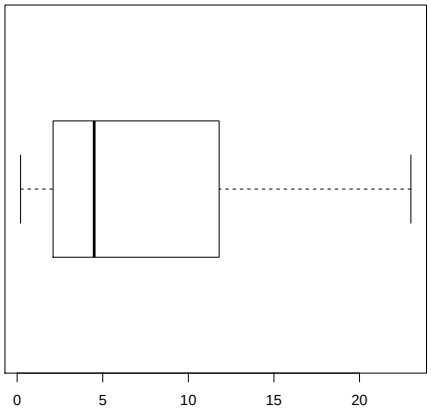


The distribution is right skewed.

Problem Set 2:

1. Median=4.50; Quartiles = 2.10, 11.80, IQR = 11.80-2.10=9.70,
   $LT = 2.10 - 1.5 \times 9.70 = -12.45$, $UT = 11.80 + 1.5 \times 9.70 = 26.35$. The data set has no outliers.

2. `boxplot(x, horizontal = T)`



3. $\bar{x} - 2s = 7.237 - 2 \times 6.35766 = -5.48$; $\bar{x} + 2s = 7.237 - 2 \times 6.35766 = 19.96$.
   The proportion is 22/24 or approx 92%.

4. The plot below was generated using the following commands:

```
x=c(7.3,1.2,4.9,5.7,13.0,1.0,3.7,0.2,10.8,1.0,2.4,12.8,1.4,14.4,1.8, 9.9,3.4,3.8,
    9.9,4.1,4.1,23.0,21.0,12.9)
hist(x,breaks=c(0,3,6,9,12,15,18,21,24),xlim=c(0,25), xaxp=c(0,24,8))
```