TUTORIAL EXERCISE PACKAGE - 2013

MATH1015 - BIOSTATISTICS

Semester 1 **Tutorial Week 2** 2013

We encourage you to attempt the tutorial 1 and R problems before your tutorial class in week 2. Ask your tutor for any help.

Make sure you know how to find $\bar{x}$ and s using the STAT mode of your calculator. Report your numerical answers at least to two decimals.

Measures of location include $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ or $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i u_i$ for group data, Q_1, median, Q_3, mode, LT = Q_1-1.5×IQR, UT = Q_3+1.5×IQR.

Measures of spread include $s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$ or $s^2 = \frac{1}{n-1}\left(\sum_{j=1}^{k} f_j u_j^2 - n\bar{x}^2\right)$ for group data, range = max - min, IQR = Q_3-Q_1, outliers$\notin$(LT,UT).

Tutorial discussion: Q4, Q5 and Q6 marked with *

1. Use the stem-and-leaf plot of the marks of a group of students on a statistics test to answer the following questions.

6	4	5	9				
7	0	1	3	6	7	8	
8	0	2	2	5	6		
9	1	1	2	2	5	8	9

Stem unit = 10, Leaf unit = 1.

(a) What is the best mark?

(b) How many students took the test?

(c) How many students scored exactly 90?

(d) Find the range of this dataset.

2. Find the mean and standard deviation (sd) of the above sample in Q1.

3. Estimate the proportion of students who have received more than 90 marks based on the sample in Q1.

4. *Two statistics quizzes were marked out of 20 and gave the following results:
A: 9, 10, 11, 12, 12, 13, 14, 14, 15, 15, 16, 17, 17, 18
B: 5, 8, 9, 10, 10, 11, 12, 12, 14, 15, 16, 17, 18, 18, 19

(a) Find the means and standard deviations of the markes in each test using the **STAT** mode of your calculator.

(b) Peter received 15 marks in test A and Kim received 15 marks in Test B. Who is better in these two tests?

5. *(**Multiple choice**) Consider the following data on x :

x	1	6	8	4	6

The variance of x for the data is (2dp):

(a) 5.60 (b) 8.70 (c) 10.4 (d) 7.00 (e) none of the above.

Use R to answer these questions.

6. *A sample of 24 mice was used to investigate the use of iron as a dietary supplement. The percentage of iron retained were

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

(a) Enter data as a vector `x` in `R`.
Hint: open the pdf data file in the course website. Type `x = scan()` in the R session window then hit "Enter". Copy and paste the data and hit "Enter" two more times.

(b) Find the mean, variance and standard deviation of this data set in `x`. **Hint:** `mean(x); var(x); sd(x)`

(c) Produce a stem-and-leaf display of the data using `stem(x), stem(x, scale = 2)` and `stem(x, scale = 5)`. What are the differences in each plot?

7. The `hospital` data are a sample from a larger data set collected on people discharged from a selected Pennsylvania hospital as part of a retrospective chart review of antibiotic usage in hospitals. Description of the columns is given below:

```
Column  Label
1       ID no.
2       Duration of hospital stay
3       Age
4       Sex  1=male  2=female
5       First temperature following admission
6       First WBC(x1000) following admission
7       Received antibiotic 1=yes  2=no
8       Received bacterial culture 1=yes  2=no
9       Service 1=med 2=surg.
```

(a) Read the data using

```
dat=read.table(file=url("http://www.maths.usyd.edu.au/math1015/r/hospital.txt"),skip=1)
```

and set the second column to be `stay` using `stay = dat[,2]`

(b) Find the mean, median, variance and standard deviation of `stay`. **Hint:** `fivenum(stay)` produces median and quantiles.

(c) Produce a histogram and comment on its shape. **Hint:** `hist(stay)`

Semester 1 **Problem Sheet Week 2** 2013

We encourage you to do these further problems in order to understand the course material well. Lecturers and tutors are available during their office hours for help.

1. Find the IQR, LT, and UT for the data in Q6 and identify any outliers.
(Remember: LT= LQ-1.5(IQR); UQ=UQ+1.5(IQR))

2. Draw a boxplot the data in Q6. **Hint:** `boxplot(x)`

3. Again, refer to data in Q6, tutorial section. Find the propotion of observations that lie in the interval $(\bar{x} - 2s, \bar{x} + 2s)$.

4. Using `R`, draw a histogram with *exactly* 8 classes for the data in Q6.