MA4605 Chemometrics Lab B

Editing Data and Paired t -Test

10 Samples were measured with two measurement methods; UV spectrometric assay and Near-infrared reflectance spectroscopy.

The value are tabulated as follows (Source: Recommended Text)

Table 3.1 Example of paired data

Batch	UV spectrometric assay	Near-infrared reflectance spectroscopy
1	84.63	83.15
2	84.38	83.72
3	84.08	83.84
4	84.41	84.20
5	83.82	83.92
6	83.55	84.16
7	83.92	84.02
8	83.69	83.60
9	84.06	84.13
10	84.03	84.24

(Trafford, A. D., Jee, R. D., Moffat, A. C. and Graham, P. 1999. *Analyst* 124: 163)

mean $\mu_d = 0$. In order to test the null hypothesis, we test whether $\bar{d}$ differs significantly from 0 using the statistic t.

An incomplete data set for both sets of measurements can be constructed as follows:

```
UVSA = c(84.63,84.38,84.08,84.41,83.82)
NIRS = c(83.15,83.72,83.84,84.20,83.92,84.16, 84.02)
```

We will use the spreadsheet interface to edit the values. Use the ***R*** command `data.entry()`. Simple type in the rest of the values, then close the spreadsheet window.

```
data.entry(UVSA)
data.entry(NIRS)
```

Compute the mean and standard deviation of the both vectors. (Write your answers in the submission sheet)

```
mean(UVSA)
sd(UVSA)
```

Using the command `t.test()` to compute the confidence interval of the mean of both data sets, at 95% and 99% confidence levels.

For this, try using the following code segments.

```
t.test(UVSA)
t.test(UVSA, conf.level = 0.95)
t.test(UVSA, conf.level = 0.99)
t.test(NIRS)
t.test(NIRS, conf.level = 0.95)
t.test(NIRS, conf.level = 0.99)
```

N.B. We disregard p-values for the time being.

Compute the case-wise differences (i.e. the difference in measurements for each case).

```
CWdiff = UVSA -NIRS
```

Compute the mean, standard deviation and 95% confidence interval for the case-wise differences. How would you interpret these results? (Write your answers in the submission sheet)

Perform a paired t-test on the data sets (we assume that the required assumptions hold).

```
t.test(UVSA, NIRS, paired= TRUE)
```

What is the conclusion of this test? (You can write your answer is terms of whether you think there is a measurement bias or not)

Testing for Normality.

In statistics, a Q-Q plot ("Q" stands for quantile) is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. If the two distributions being compared are similar, the points in the Q-Q plot will approximately lie on the line y = x. A norm Q-Q plot compares the sample distribution against a normal distribution.

```
> attach(iris)
> qqnorm(Sepal.Width)
> qqline(Sepal.Width)
```

In your submission sheet, draw a small sketch of this Q-Qplot.

Shapiro Wilk Test for Normality

This is a hypothesis tests with the null hypothesis that the data comes from a normal distribution. Hence if the p-value is below the significance threshold (typically 0.05), then the null hypothesis is rejected and the alternative hypothesis is accepted. Here the alternative hypothesis is that the data does not come from a normal distribution.

Use the `shapiro.test()` to test whether the column values of iris are normally distributed.

```
> attach(iris)
> shapiro.test(Sepal.Width)
> shapiro.test(Petal.Width)
> shapiro.test(Sepal.Length)
> shapiro.test(Petal.Length)
```

Write down the p-values for each variable, and provide a brief interpretation of the outcome.