Tutorial 5: Hypothesis Testing – Answers Sheet

Rob Nicholls – nicholls@mrc-lmb.cam.ac.uk

MRC LMB Statistics Course 2014

Task 1

1.

```
shapiro.test(CO2$uptake[CO2$Plant=="Mc1"])
```

Answer : p-value = 0.3768. It is not unreasonable to assume that the CO2 uptake for Mc1 is Normally distributed.

```
shapiro.test(CO2$uptake[CO2$Plant=="Mc2"])
```

Answer : p-value = 0.1543. It is not unreasonable to assume that the CO2 uptake for Mc1 is Normally distributed.

```
shapiro.test(CO2$uptake[CO2$Plant=="Mc3"])
```

Answer : p-value = 0.001716. Reject the hypothesis that CO2 uptake is Normally distributed for Mc3.

2.

```
install.packages("outliers")
library(outliers)
grubbs.test(CO2$uptake[CO2$Plant=="Mc1"])
```

Answer : p-value = 0.1038. Grubb's test does not detect any outliers.

```
grubbs.test(CO2$uptake[CO2$Plant=="Mc2"])
```

Answer : p-value = 0.02225. Grubb's test detects an outlier (10.5).

```
dixon.test(CO2$uptake[CO2$Plant=="Mc3"])
```

Answer : p-value ≈ 0. Dixon's test detects an outlier (10.6).

Task 2

1.

```
hist(LakeHuron)
qqnorm(LakeHuron)
```

2.

```
shapiro.test(LakeHuron)
```

Answer : p-value $= 0.3271$. It is not unreasonable to assume that the data are Normally distributed.

3.

```
t.test(LakeHuron,mu=578)
```

Answer : p-value ≈ 0. We can reject the hypothesis that $\mu = 578$.

4.

95% confidence interval: $[578.7398, 579.2684]$.

5.

The only integer value that could reasonably equal the mean is 579.

Task 3

1.

```
hist(Nile)
qqnorm(Nile)
```

2.

```
shapiro.test(Nile)
```

Answer : p-value $= 0.04072$. Reject the hypothesis that the data are Normally distributed.

3.

Choosing to use either the sign test or the Wilcoxon test is acceptable. Choosing to use the sign test over the Wilcoxon test on the grounds of potential asymmetry of the distribution is a particularly acceptable choice. However, for the sake of gathering maximal information, it may be best to consider the results of both tests:

```
binom.test(sum(Nile>850),length(Nile))
wilcox.test(Nile,mu=850)
```

Answer : p-value $= 0.1933$ from the sign test, and p-value $= 0.0008844$ from the Wilcoxon test. The data are not entirely symmetric, so the assumptions for the

Wilcoxon test may not be valid. However, the lack of significance achieved by the sign test may be due to lack of power. Consequently, it is not clear whether or not the hypothesis that the median is 850 should be rejected. If type II errors are deemed 'worse' than type I errors, then it may be prudent to not reject the hypothesis in this case, due to the ambiguity, but nevertheless acknowledge the fact that evidence has been gained that may support rejection of the hypothesis. As such, further testing would be required in order to reach a solid conclusion.

```
binom.test(sum(Nile>950),length(Nile))
wilcox.test(Nile,mu=950)
```

Answer : p-value $= 0.0352$ from the sign test, and p-value $= 0.05503$ from the Wilcoxon test. The sign test rejects the hypothesis that the median is 950, whilst the Wilcoxon test does not. On the basis that the sign test makes reasonable assumptions, and achieves significance despite low power, it is reasonable to reject the hypothesis in this case. Indeed, the lack of significance achieved by the Wilcoxon test may be due to incorrect assumptions regarding symmetry of the data.

4.

```
t.test(Nile,mu=880)
```

Answer : p-value $= 0.0221$.

```
wilcox.test(Nile,mu=880)
```

Answer : p-value $= 0.08039$.

The hypothesis that $\mu = 880$ is rejected when using a t-test, but not rejected when using the Wilcoxon test.

5.

An argument could be made for using either the t-test or Wilcoxon test in this case.

On the one hand, we have established that the data cannot be considered Normally distributed, using the Shapiro-Wilk Normality test. Consequently, the assumptions for using the t-test to test for differences in the mean may be violated, and thus we should use the Wilcoxon signed-rank test in order to test for differences in the median.

On the other hand, note that the dataset comprises many observations. Consequently, whilst the data may not be Normally distributed, it may be reasonable to assume that the sampling distribution of the mean ($\bar{x}$) is Normally distributed (as a direct consequence of the Central Limit Theorem, since the number of observations is large and the data are not *too* non-Normal). Consequently, the assumptions made by the t-test may be justified.

Ultimately, in such a case the decision regarding whether or not the hypothesis should be rejected should depend on context. Specifically, the consequences of a type I versus a type II error should be considered. If incorrect rejection of the hypothesis (i.e. type I error) would have catastrophic consequences then the default would be to decide that the results are inconclusive, thus fail to reject the hypothesis – this would amount to using the Wilcoxon test in this case. By default, this is considered more prudent than choosing to reject a hypothesis in the absence of unambiguously conclusive evidence (i.e. significance achieved from a test performed with reasonable assumptions). Of course, failing to reject a hypothesis always leaves room for further investigation (e.g. collecting more data or performing other statistical tests) whereas rejecting a hypothesis essentially leaves little room for subsequent investigation (since hypothesis rejection essentially amounts to statistical contradiction).

In summary, an argument could be made for using either the t-test or Wilcoxon test in this case. However, it might be more prudent to use the Wilcoxon test, failing to reject the hypothesis, but at the same time acknowledging that this conclusion is borderline/inconclusive. Consequently, further investigation would be required in order to draw more solid conclusions.

Task 4

1.

```
hist(CO2$uptake)
qqnorm(CO2$uptake)
```

2.

```
shapiro.test(CO2$uptake)
```

Answer : p-value = 0.0007908. Reject the hypothesis that the data are Normally distributed.

3.

```
x = CO2[CO2$conc<300,]
boxplot(x$uptake~x$Treatment)

for(i in levels(x$Treatment)){
print(shapiro.test(x$uptake[x$Treatment==i]))
}
```

Answer : p-value = 0.4138 for nonchilled; p-value = 0.07255 for chilled. Cannot reject the hypothesis that the data are Normally distributed for either treatment.

```
var.test(x$uptake~x$Treatment)
```

Answer : p-value = 0.8637. Cannot reject the hypothesis that the variances are equal.

```
t.test(x$uptake~x$Treatment,var.equal=TRUE,alternative="greater")
```

Answer : p-value = 0.05681. Cannot reject the hypothesis. Therefore, cannot conclude that the average CO2 uptake for nonchilled plants is significantly greater than for chilled plants, for concentrations less than 300mL/L.

4.

```
x = CO2[CO2$conc>300,]
boxplot(x$uptake~x$Treatment)

for(i in levels(x$Treatment)){
print(shapiro.test(x$uptake[x$Treatment==i]))
}
```

Answer : p-value = 0.04062 for nonchilled; p-value = 0.003608 for chilled. Reject the hypothesis that the data are Normally distributed for both treatments (note that we cannot assume that both distributions are Normal, even when applying the Bonferroni correction).

```
wilcox.test(x$uptake~x$Treatment,alternative="greater")
```

Answer : p-value = 0.006669. Reject the hypothesis. Therefore, conclude that the average CO2 uptake for nonchilled plants is significantly greater than for chilled plants, for concentrations greater than 300mL/L.

5.

```
x = CO2[CO2$conc>400,]
boxplot(x$uptake~x$Type)

for(i in levels(x$Type)){
print(shapiro.test(x$uptake[x$Type==i]))
}
```

Answer : p-value = 0.9025 for Quebec; p-value = 0.1593 for Mississippi. Cannot reject the hypothesis that the data are Normally distributed for either type.

```
var.test(x$uptake~x$Type)
```

Answer : p-value = 0.002558. Reject the hypothesis that the variances are equal.

```
t.test(x$uptake~x$Type)
```

Answer : p-value ≈ 0. Reject the hypothesis. Therefore, conclude that the average CO2 uptake is significantly different in the plants from Quebec and Mississippi, for concentrations greater than 400mL/L.

6.

The conclusions that:

1. the average CO2 uptake for nonchilled plants is significantly greater than for chilled plants, for concentrations greater than 300mL/L, and

2. the average CO2 uptake is significantly different in the plants from Quebec and Mississippi, for concentrations greater than 400mL/L,

are believable. However, the conclusion that the average CO2 uptake for nonchilled plants cannot be considered significantly greater than for chilled plants, for concentrations less than 300mL/L, is less convincing. This result is inconclusive, and should be further investigated by (1) obtaining more data; or (2) finding other methods with which to test the hypothesis.

Task 5

1.

```
x = CO2[CO2$conc>=675,]
boxplot(x$uptake~x$conc)
```

From looking at the box plots, it does not appear that the distributions are very dissimilar.

2.

```
for(i in levels(factor(x$conc))){
print(shapiro.test(x$uptake[x$conc==i]))
}
```

Answer : p-value = 0.4104 for 675 mL/L; p-value = 0.2211 for 1000 mL/L. Cannot reject the hypothesis that the data are Normally distributed for either concentration.

3.

```
var.test(x$uptake~x$conc)
```

Answer : p-value = 0.7756. Cannot reject the hypothesis that the variances are equal.

4.

```
t.test(x$uptake~x$conc,var.equal=TRUE,alternative="less")
```

Answer : p-value = 0.3462. Cannot reject the hypothesis. Therefore, cannot conclude that CO2 uptake is substantially larger for concentrations of 1000 mL/L than for 675 mL/L.

5.

```
t.test(x$uptake~x$conc,paired=TRUE,alternative="less")
```

Answer : p-value = 0.002496. Reject the hypothesis. Therefore, can conclude that CO2 uptake is substantially larger for concentrations of 1000 mL/L than for 675 mL/L.