## Hypothesis Testing: practical exercises

### Practical exercise 1: One sample t-test (see R script lower below)

* Using the farm\_data2 we generate a null hypothesis that the mean FarmIncome across all farmers in the sample (n = 300) is 2,000 versus the alternative hypothesis that it is something different.
* Across the 300 farms, the average FarmIncome = 1,206 with a standard deviation = 1,041.
* The difference between our null hypothesis and the sample mean = 1,100 – 1,206 = -106 (can go back to Ben’s example in eSMS 4.1 if want to show it visually or draw it on board).
* We want to know the probability (p-value) of getting a sample mean at least 106 greater or less than 1,100.

One Sample t-test

data: data$FarmIncome

t = 1.7711, df = 299, p-value = 0.07756 (so a little above p=0.05)

alternative hypothesis: true mean is not equal to 1100

95 percent confidence interval:1088.172 1324.682

sample estimates:mean of x 1206.427

* p-value = 0.078 which means there is a probability of 0.078 or a 7.8% chance of any sample mean being at least 106 higher or lower than 1,100. This is larger than p=0.05 (although note not so much higher) which means we don’t have sufficient evidence at the 5% level to reject the null hypothesis that the mean farm income = 1,100.

### Practical exercise 2: Two sample t-test (see R script lower down)

* Using the farm\_data2 we generate a null hypothesis that the mean FarmIncome for large farms is the same as for small farms.
* The average FarmIncome for large farms is 2,713 (s.d. = 723) and for small farms is 601 (s.d. = 166), so the difference is 2,112.
* The difference between our null hypothesis (income is the same) and the alternative hypothesis (income is different) = 2,112.
* We want to know the probability (p-value) there is no difference between the average farm income for large and small farms.

Welch Two Sample t-test

data: large\_farm\_income and small\_farm\_income

t = 26.825, df = 88.637, p-value < 2.2e-16 (p-value much smaller than 0.05)

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:1955.585 2268.485

sample estimates:mean of x mean of y

2713.0116 600.9766

* p-value < 2.2e-16 which is less than p=0.001 (the common way of writing would be p<0.001) which means there is a probability of less than 0.001 or a less than 0.1% chance of the difference in farm income being zero (0). This is much smaller than p=0.05 means we don’t have strong evidence to reject the null hypothesis that large and small farms have the same mean farm income.

## R Script

# Load necessary libraries

library**(**readxl**)**

# Load the data

data **<-** read\_excel**(**"farm\_data2.xlsx", sheet **=** "Sheet1", range **=** "A1:C301"**)**

# Conduct a 1-sample t-test, with null hypothesis mean = 1,100

t\_test **<-** t.test**(**data**$**FarmIncome, mu **=** 1100**)**

# Display the results

print**(**t\_test**)**

# Separate the data into two groups: Large and Small farms

large\_farm\_income **<-** data**$**FarmIncome**[**data**$**FarmSize **==** "large"**]**

small\_farm\_income **<-** data**$**FarmIncome**[**data**$**FarmSize **==** "small"**]**

# Conduct the two-sample t-test (assuming unequal variances - Welch's t-test)

t\_test2 **<-** t.test**(**large\_farm\_income, small\_farm\_income, var.equal **=** **FALSE)**

# Print the results

print**(**t\_test2**)**