0.0.1 Using Statistical Computing for Inference Procedures

1. Paired t-test using `R`
2. Test for the equality of variances for two samples
3. Shapiro-Wilk Test for Normality
4. Graphical procedures for assessing normality
5. The minitab graphical summary
6. Grubb's Procedure for Determinin an Outlier

0.0.2 p-values using Statistical Software

- In every inference procedure performed using programs such as SPSS or `R`, a p-value is presented to the screen for the user to interpret.
- If the p-value is larger than a specified threshold α/k then the appropriate conclusion is a failure to reject the null hypothesis.
- Conversely, if the p-value is less than threshold, the appropriate conclusion is to reject the null hypothesis.
- In this module, we will use a significance level$\alpha = 0.05$ and almost always the procedures will be two tailed ($k = 2$). Therefore the threshold α/k will be 0.025.

0.0.3 R Statistical Computing

- `R` is a computing software for statistical analysis
- The package is available for all popular operating systems: Windows , Mac OS and Linux
- It is free!
- Everyone (knowledgeable enough) can contribute to the software by writing a package. Packages are available for download through a convenient facility
- `R` is fairly well documented and the documentation is available either from the program help menu or from the web-site.
- `R` is the top choice of statistical software among academic statisticians but also very popular in industry.
- `R` is a powerful tool not only for doing statistics but also all kind of scientific programming.

`R` is a language and environment for statistical computing and graphics. `R` provides a wide variety of statistical and graphical techniques, and is highly extensible. Among its tools one can find implemented

- linear and nonlinear modelling,
- classical statistical tests,
- time-series analysis,

- classification,
- clustering,
- ...and many more.

One of R's strengths is the ease with which well-designed publication quality plots can be produced. including mathematical symbols and formulae where needed.

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent. integrated collection of intermediate tools for data analysis, graphical facilities for data analysis and display either on-screen or on hard-copy, and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

0.0.4 R Statistical Computing

Downloading and Installing R:

- R can be downloaded from the CRAN website: http://cran.r-project.org/
- You may choose versions for windows, mac and linux.
- As per the instructions on the respective pages, you require the "base" distribution.
- Now you can download the installer for latest version of R , version 2.17.
- Select the default settings. Once you finish, the R icon should appear on your desktop.
- Clicking on this icon will start up the program.

0.0.5 Using R for Inference Procedures

1 Review of the Paired t-test.

2 Paired t-test using R

3 Test for the equality of variances for two samples

4 Shapiro-Wilk Test for Normality

5 Graphical procedures for assessing normality

6 Grubb's Procedure for Determinin an Outlier

Using Confidence Limits

- Alternatively, we can use the confidence interval to make a decision on whether or not we should reject or fail to reject the null hypothesis.
- If the null value is within the range of the confidence limits, we fail to reject the null hypothesis.
- If the null value is outside the range of the confidence limits, we reject the null hypothesis.
- Occasionally a conclusion based on this approach may differ from a conclusion based on the p-value. In such a case, remark upon this discrepancy.

0.1 The paired t-test

- Previously we have seen the paired t-test. It is used to determine whether or not there is a significant difference between paired measurements. Equivalently whether or not the case-wise differences are zero.
- The mean and standard deviation of the case-wise differences are used to determine the test statistic.
- Under the null hypothesis, the expected value of the case-wise differences is zero (i.e $H_0 : \mu_d = 0$).
- The test statistic is computed as

$$TS = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

The Paired t-test (b)

- The calculation is dependent on the case-wise differences.
- Here the case-wise differences between paired measurements (e.g. "before" and "after").
- Under the null hypothesis, the mean of case-wise differences is zero.
- As a quick example, the mean, standard deviation and sample size are presented in the next slide.

The paired t-test (c)

- Observed Mean of Case-wise differences $\bar{d} = 8.21$,
- Expected Mean of Case-wise differences under $H_0 : \mu_d = 0$,
- Standard Deviation of Case-wise differences $S_d = 7.90$,
- Sample Size $n = 14$.

$$TS = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{8.21 - 0}{\frac{7.90}{\sqrt{14}}} = 3.881$$

The paired t-test (c)

```
> CWdiff = Before - After
> mean(CWdiff)
[1] 8.214286
> sd(CWdiff)
[1] 7.904999
> length(CWdiff)
[1] 14
```

$$TS = \frac{8.21 - 0}{\frac{7.90}{\sqrt{14}}} = 3.881$$

In an exam situation, the candidate will be expected to compute this value. It will be omitted from `R` code output.

The paired t-test (e)

- Procedure is two-tailed, and you can assume a significance level of 5%.
- It is also a small sample procedure (n=14, hence df = 13).
- The Critical value is determined from statistical tables (2.1603).
- Decision Rule: Reject Null Hypothesis ($|TS| > CV$). Significant difference in measurements before and after.

The paired t-test (f)

- We consider the confidence interval. We are 95% confident that the expected value of the case-wise difference is at least 3.65.
- Here the null value (i.e. 0) is not within the range of the confidence limits.
- Therefore we reject the null hypothesis.

```
> t.test(Before,After,paired=TRUE)
...
...
95 percent confidence interval:
3.650075 12.778496
...
```

The paired t-test (f) Alternative Approach : using p-values.

- The p-values are determined from computer code. (We will use a software called `R`. Other types of software inlcude `SAS` and `SPSS`.)
- The null and alternative are as before.
- The computer software automatically generates the appropriate test statistic, and hence the corresponding p-value.

- The user then interprets the p-values. If p-value is small, reject the null hypothesis. If the p-value is large, the appropriate conclusion is a failure to reject H_0.
- The threshold for being considered small is less than α/k, (usually 0.0250). (This is a very arbitrary choice of threshold, suitable for some subject areas, not for others)

The paired t-test (g) Implementing the paired t-test using `R` for the example previously discussed.

```
> t.test(Before,After,paired=TRUE)

Paired t-test

data:  Before and After
t = 3.8881, df = 13, p-value = 0.001868
alternative hypothesis: true difference in means is not 0
95 percent confidence interval:
3.650075 12.778496
sample estimates:
mean of the differences
8.214286
```

The paired t-test (h)

- The p-value (0.001868) is less than the threshold is less than the threshold 0.0250.
- We reject the null hypothesis (mean of case-wise differences being zero, i.e. expect no difference between "before" and "after").
- We conclude that there is a difference between 'before' and 'after'.
- That is to say, we can expected a difference between two paired measurements.

The paired t-test (i)

- We could also consider the confidence interval. We are 95% confident that the expected value of the case-wise difference is at least 3.65.
- Here the null value (i.e. 0) is not within the range of the confidence limits.
- Therefore we reject the null hypothesis.

```
> t.test(Before,After,paired=TRUE)
...
...
95 percent confidence interval:
3.650075 12.778496
...
```

Using Statistical Software

1. `var.test` : Testing Equality of Variances for two samples using `R`.
2. The Shapiro-Wilk Test for Normality (Using `R`).
3. Graphical Procedures for Assessing Normality : The QQ plot
4. The Grubbs' Test for Outliers.

[fragile] **Single Sample Proportion Test (a)**

- In this procedure, we determine whether or not we are are justified in assuming that the population proportion takes a certain value.
- For example, suppose we believed that the population proportion of students with iphones or androids was 80%.
- We would write the null and alternative accordingly.

$$H_0 : \pi = 80\%$$

$$H_1 : \pi \neq 80\%$$

- The appropriate `R` command is `prop.test(x,n,p)`
- x is the number of successes, n is the sample size and p is the population proportion assumed under the null hypothesis.
- Suppose we survey 65 students, with 50 replying that they had an iphone or android.

[fragile] **Single Sample Proportion Test (b)**

```
> prop.test(50,65,0.80)

1-sample proportions test

data:  50 out of 65, null probability 0.8
X-squared = 0.2163, df = 1, p-value = 0.6418

alternative hypothesis: true p is not equal to 0.8
95 percent confidence interval:
0.6452269 0.8610191

sample estimates:
p
0.7692308
```

[fragile] **Single Sample Proportion Test (c)**

- The p-value is above the threshold. Therefore we fail to reject the null hypothesis that the population proportion (π) is 80%.

```
> shapiro.test(Y1)

        Shapiro-Wilk normality test

data:  Y1
W = 0.90493, p-value = 0.1132
```

- The observed proportion is a very straightforward calculation:

$$\hat{p} = \frac{50}{65} = 0.76923 = 76.92\%$$

- Nonetheless, you would be required to show how it was calculated.

Shapiro-Wilk Test

$\mathbf{H}_0$: The sample is drawn from a normally distributed population. $\mathbf{H}_1$: The sample is drawn from a population that is NOT normally distributed

- Compare the p-value to some pre-defined thrreshold.
- We will use a significance level of $\alpha = 0.05$.
- This test is a one-tailed test. We compare the p-value from the Shapiro Wilk Test to 0.05.
- In the case of two-tailed tests, we would compare the p-value to 0.025, i.e. $(\alpha/2)$.

Shapiro-Wilk Test

- If the p-value is less than the threshold, we reject the null hypothesis.
  (We have enough evidence to say that the population is not normally distributed.)
- If the p-value is greater than the threshold, we fail to reject the null hypothesis.
  (We dont have enough evidence to say that the population is not normally distributed.)
- Type I and Type II Errors apply to this test, just like any other test.

Using `R` for Inference Procedures

1 Paired t-test

2 Single Sample Tests for Proportions

```
> shapiro.test(Y2)

        Shapiro-Wilk normality test

data:  Y2
W = 0.96603, p-value = 0.7709

> shapiro.test(Y3)

        Shapiro-Wilk normality test

data:  Y3
W = 0.66584, p-value = 0.0001097
```

3 Two Sample Test for Proportions

4 Test for the equality of variances for two samples

5 Shapiro-Wilk Test for Normality

6 Graphical procedures for assessing normality

0.1.1 Grubbs Test for Determining an Outlier

The Grubbs test is used to determine if there are any outliers in a data set.
There is no agreed formal definition for an outlier. The definition of outlier used for this procedure is a value

that unusually distance from the rest of the values.
(For the sake of clarity , we shall call this type of outlier a **Grubbs Outlier**).
Consider the following data set: is the lowest value 4.01 an outlier?

```
6.98 8.49 7.97 6.64
8.80 8.48 5.94 6.94
6.89 7.47 7.32 4.01
```

Under the null hypothesis, there are no outliers present in the data set. We reject this hypothesis if the p-value is sufficiently small.
(Remark: This is a one-tailed test).

0.1.2 Grubbs Test for Determining an Outlier

```
> grubbs.test(x, two.sided=T)
Grubbs test for one outlier
data: x
G = 2.4093, U = 0.4243, p-value = 0.05069
alternative hypothesis: lowest value 4.01 is an outlier
```

We dont not have enough evidence to class 4.01 as an outlier. We conclude that while small by comparison to the other values, the lowest value 4.01 is not an outlier.

0.1.3 Test of Equality for Two Sample Proportions (a)

The null hypothesis is that two populations have the same proportions for a particular characteristic.

$$H_0 : \pi_1 = \pi_2$$

$$H_1 : \pi_1 \neq \pi_2$$

- The command is `prop.test(c(x1,x2),c(n1,n2))`
- $x1$ and $x2$ are the number of successes from both samples.
- $n2$ and $n2$ are the sample sizes for both groups.
- The difference in population proportions assumed under the null hypothesis is zero.
- (It is possible to specify a different null value, but we will not consider this in this module.)

0.1.4 Test of Equality for Two Sample Proportions (b)

- Consider a study where the proportion of Irish students who owned mobile devices, such as iphones and androids was compared to the corresponding proportion of French student.
- As before, 65 Irish students were interviewed, with 50 replying that they owned mobile devices.
- 90 french students were interview, with 60 responding that they owned mobile devices.
- The test of equality of proportions is implemented on the next slide.

0.1.5 Test of Equality for Two Sample Proportions (c)

Based on the p-value, we fail to reject the null hypothesis. There is not enough evidence to assume a difference in proportions. Also the expected difference assumed under the null hypothesis, i.e. 0, is contained in the confidence interval.

```
> prop.test(c(50,60),c(65,90))

2-sample test for equality of proportions

data:  c(50, 60) out of c(65, 90)
X-squared = 1.4613, df = 1, p-value = 0.2267
alternative hypothesis: two.sided
95 percent confidence interval:
-0.05202058  0.25714878
sample estimates:
prop 1    prop 2
0.7692308 0.6666667
```

0.1.6 Test of Equality for Two Sample Proportions (d)

- You would be required to compute the differences in observed proportions.
- Additionally you will given the `R` code for one sample procedures. This may or may not be relevant for answering the question.

```
> prop.test(60,90,0.80)
...
...
X-squared = 9.184, df = 1, p-value = 0.002441
alternative hypothesis: true p is not equal to 0.8
95 percent confidence interval:
0.5585219 0.7604058
sample estimates:
p
0.6666667
```

0.1.7 Test for Equality of Variance (a)

- In this procedure, we determine whether or not two populations have the same variance.
- The assumption of equal variance of two populations underpins several inference procedures. This assumption is tested by comparing the variance of samples taken from both populations.
- We will not get into too much detail about that, but concentrate on how to assess equality of variance.
- The null and alternative hypotheses are as follows:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

0.1.8 Test for Equality of Variance (b)

- When using `R` it would be convenient to consider the null and alternative in terms of variance ratios.
- Two data sets have equal variance if the variance ratio is 1.
- **Remark** : This is a two-tailed test.

$$H_0 : \sigma_1^2/\sigma_2^2 = 1$$
$$H_1 : \sigma_1^2/\sigma_2^2 \neq 1$$

0.1.9 Test for Equality of Variance(c)

You would be required to compute the test statistic for this procedure. The test statistic is the ratio of the variances for both data sets.

$$TS = \frac{s_x^2}{s_y^2}$$

The standard deviations would be provided in the `R` code.

- Sample standard deviation for data set $x = 3.40$
- Sample standard deviation for data set $y = 4.63$

To compute the test statistic.

$$TS = \frac{3.40^2}{4.63^2} = \frac{11.56}{21.43} = 0.5394$$

0.1.10 Variance Test (d)

```
> var.test(x,y)

F test to compare two variances

data:  x and y
F = 0.5394, num df = 9, denom df = 8, p-value = 0.3764
alternative hypothesis:
true ratio of variances is not equal to 1
95 percent confidence interval:
0.1237892 2.2125056
sample estimates:
ratio of variances
0.5393782
```

0.1.11 Variance Test (e)

- The p-value is 0.3764 (top right), above the threshold level of 0.0250.
- We fail to reject the null hypothesis.
- There is not enough evidence to say there is a difference in variance between the two populations.
- We can assume that there is no significant difference in sample variances. Therefore we can assume that both populations have equal variance.
- Additionally the 95% confidence interval (0.1237, 2.2125) contains the null values i.e. 1.

Variance Test (e)

- The p-value is 0.3764 (top right), above the threshold level of 0.0250.
- We fail to reject the null hypothesis.
- We can assume that there is no significant difference in sample variances. Therefore we can assume that both populations have equal variance.
- Additionally the 95% confidence interval (0.1237, 2.2125) contains the null values i.e. 1.

Shapiro-Wilk Test(a)

- We will often be required to determine whether or not a data set is normally distributed.
- Again, this assumption underpins many statistical models.
- The null hypothesis is that the data set is normally distributed.
- The alternative hypothesis is that the data set is not normally distributed.
- One procedure for testing these hypotheses is the Shapiro-Wilk test, implemented in `R` using the command `shapiro.test()`.

- (Remark: You will not be required to compute the test statistic for this test.)

Shapiro Wilk Test(b) For the data set used previously; x and y, we use the Shapiro-Wilk test to determine that both data sets are normally distributed.

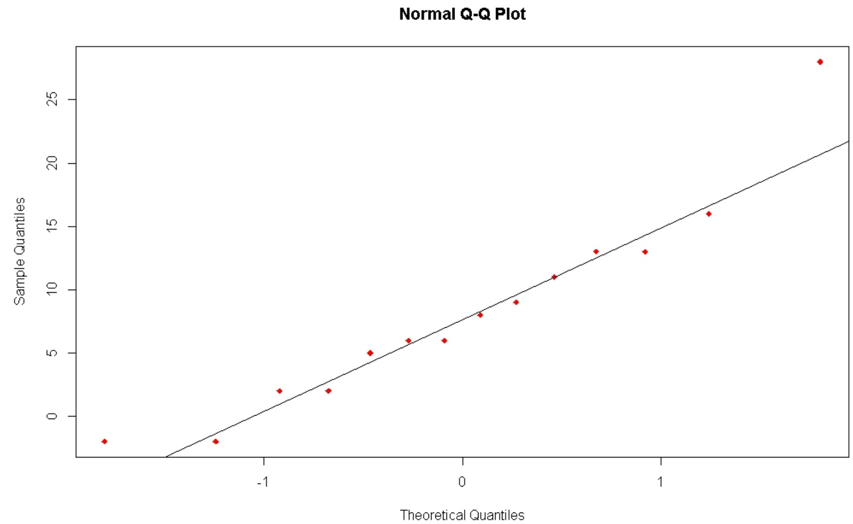
```
> shapiro.test(x)

Shapiro-Wilk normality test

data:  x
W = 0.9474, p-value = 0.6378

> shapiro.test(y)

Shapiro-Wilk normality test

data:  y
W = 0.9347, p-value = 0.5273
```

Graphical Procedures for assessing Normality

- The normal probability (Q-Q) plot is a very useful tool for determining whether or not a data set is normally distributed.
- Interpretation is simple. If the points follow the trendline (provided by the second line of `R` code `qqline`).
- One should expect minor deviations. Numerous major deviations would lead the analyst to conclude that the data set is not normally distributed.
- The Q-Q plot is best used in conjunction with a formal procedure such as the Shapiro-Wilk test.

```
>qqnorm(CWdiff)
>qqline(CWdiff)
```

Graphical Procedures for Assessing Normality



- Previously we have seen the paired $t-$test. It is used to determine whether or not there is a significant difference between paired measurements.

The paired t-test with R

- Equivalently whether or not the case-wise differences are zero.
- The mean and standard deviation of the case-wise differences are used to determine the test statistic.
- Under the null hypothesis, the expected value of the case-wise differences is zero (i.e $H_0 : \mu_d = 0$).

The paired t-test with `R`

- In the following procedure (next slide), there are two sets of values: the `Before` values and the `After` values.
- The `R` command is `t.test()`, with the additional specification "`paired=`".
- The alternative hypothesis is specified in the output. (Another way of expressing it: True mean of case-wise differences is not zero)
- Also included in the output is a 95% confidence interval for the sample mean of case-wise differences.

Test for Equality of Variance

- The $p-$value is 0.3764 (top right), above the threshold level of 0.0250.
- We fail to reject the null hypothesis.
- We can assume that there is no significant difference in sample variances. Therefore we can assume that both populations have equal variance.
- Additionally the 95% confidence interval (0.1237, 2.2125) contains the expected value under the assumption of equal variance i.e. 1.

Shapiro-Wilk Test for Normality

- We will often be required to determine whether or not a data set is normally distributed.
- Again, this assumption underpins many statistical models.
- The null hypothesis is that the data set is normally distributed.
- The alternative hypothesis is that the data set is not normally distributed.
- One procedure for testing these hypotheses is the Shapiro-Wilk test, implemented in `R` using the command `shapiro.test()`.

Variance Test (e)

- The p-value is 0.3764 (top right), above the threshold level of 0.0250.
- We fail to reject the null hypothesis.
- We can assume that there is no significant difference in sample variances. Therefore we can assume that both populations have equal variance.
- Additionally the 95% confidence interval (0.1237, 2.2125) contains the null values i.e. 1.

Shapiro-Wilk Test(a)

- We will often be required to determine whether or not a data set is normally distributed.
- Again, this assumption underpins many statistical models.
- The null hypothesis is that the data set is normally distributed.
- The alternative hypothesis is that the data set is not normally distributed.
- One procedure for testing these hypotheses is the Shapiro-Wilk test, implemented in `R` using the command `shapiro.test()`.
- (Remark: You will not be required to compute the test statistic for this test.)

Graphical Procedures for assessing Normality

- The normal probability (Q-Q) plot is a very useful tool for determining whether or not a data set is normally distributed.
- Interpretation is simple. If the points follow the trendline (provided by the second line of `R` code `qqline`).
- One should expect minor deviations. Numerous major deviations would lead the analyst to conclude that the data set is not normally distributed.
- The Q-Q plot is best used in conjunction with a formal procedure such as the Shapiro-Wilk test.

```
>qqnorm(CWdiff)
>qqline(CWdiff)
```