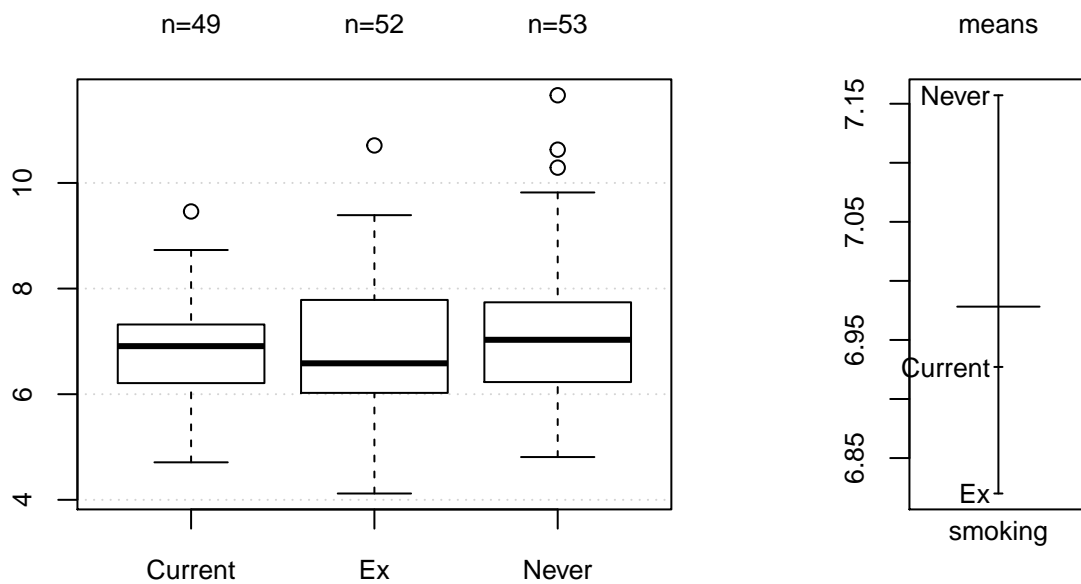


Applied Statistics using R

FACTSHEET

UNIT 3



Medical Statistics Team

University of Aberdeen

This booklet should not be reproduced without permission from the Medical Statistics Team

© Unless otherwise stated all content Copyright University of Aberdeen

Copyright Notice

Unless otherwise stated all content Copyright University of Aberdeen. The University of Aberdeen subscribes to the Copyright Licensing Agency's Higher Education Photocopying and Scanning Licence. You may access, download and print out a copy of any material included under the terms of this licence.

Any digital or print copy supplied to or made by you are for use in connection with this Course of Study. You may retain such copies after the end of the course, but strictly for your own personal use.

All copies (including electronic copies) shall include this Copyright Notice and shall be destroyed and/or deleted if and when required by the University of Aberdeen.

Except as provided by copyright law, no further copying, storage or distribution (including by e-mail) is permitted without the consent of the copyright holder.

The author has moral rights in the work. No distortion, mutilation, or other modifications of the work, or any other derogatory treatment prejudicial to the honour or reputation of the author is permitted.

Contents

1	Introduction to Hypothesis Testing	1
1.1	Overview of hypothesis testing	1
1.2	Null and alternative hypotheses	1
1.3	One and two tailed tests	1
1.4	Example	1
1.5	The p-value	2
1.6	Parametric versus non-parametric hypothesis tests	2
1.7	Hypothesis testing versus confidence intervals	2
1.8	Errors in hypothesis testing	2
1.9	Multiple testing	2
1.10	Practical (or clinical) importance versus statistical significance	3
2	One-sample t-test	4
2.1	When is it appropriate?	4
2.2	Example	4
2.3	Null hypothesis	4
2.4	Assumptions	4
2.5	Using R	4
2.6	Steps to interpretation	5
3	Independent samples t-test	6
3.1	When is it appropriate?	6
3.2	Example	6
3.3	Null hypothesis	6
3.4	Assumptions	6
3.5	Using R	7
3.6	Steps to interpretation	12

4	Paired samples t-test	13
4.1	When is it appropriate?	13
4.2	Example	13
4.3	Null hypothesis	13
4.4	Assumptions	13
4.5	Using R	14
4.6	Steps to interpretation	16
5	Analysis of Variance (ANOVA)	17
5.1	When is it appropriate?	17
5.2	Example:	17
5.3	Null hypothesis	17
5.4	Assumptions	17
5.5	Using R	18
5.6	Steps to interpretation	23
6	Chi-squared test with continuity correction (2×2 Tables)	25
6.1	Chi-squared test with continuity correction	25
6.2	Using R	26
6.3	Fisher's exact test	28
7	Chi-squared test for trend	30
7.1	When is it appropriate?	30
7.2	Chi-squared test for trend	30
7.3	Using R	30
8	Chi-squared test for association ($r \times c$)	33
8.1	When is it appropriate?	33
8.2	Chi-squared test for association	33
8.3	Using R	33
9	McNemar Test	37
9.1	When is it appropriate?	37
10	Mann-Whitney test	40
10.1	When is it appropriate?	40
10.2	Using R	40

11 Kruskal-Wallis test	46
11.1 When is it appropriate?	46
11.2 Using R	46
12 Wilcoxon Matched pairs test	50
12.1 When is it appropriate?	50
12.2 Using R	50

Chapter 1

Introduction to Hypothesis Testing

1.1 Overview of hypothesis testing

- Define the null and alternative hypotheses
- Collect data
- Calculate the probability (the p-value) of obtaining these results (or more extreme results) given the null hypothesis
- Use the p-value to decide whether to reject the null hypothesis

1.2 Null and alternative hypotheses

- The *null* hypothesis (H_0) is that there is *no difference* in the *population*
- The *alternative* hypothesis (H_1) is that there is a *difference* in the *population*

NB: the null and alternative hypotheses concern the whole population of interest, not the sample

1.3 One and two tailed tests

Usually a two-tailed (or two-sided) test is assumed. This means that the alternative hypothesis reflects that the results could go in either direction.

1.4 Example

- H_0 : In the population mean blood pressure is the same before and after taking drug A
- Two-tailed H_1 : In the population mean blood pressure is not the same before and after taking drug A
- One-tailed H_1 : In the population mean blood pressure is lower after taking drug A

1.5 The p-value

The p-value is the probability of obtaining our results, or more extreme results, if the null hypothesis were true.

- In general, if the p-value is less than 5% ($p < 0.05$) we reject the null hypothesis in favour of the alternative hypothesis. The result is referred to as being statistically significant.
- If the p-value is greater than or equal to 5% ($p \geq 0.05$) we do not have enough evidence to reject the null hypothesis. The result is not statistically significant and we don't reject the null hypothesis.

1.6 Parametric versus non-parametric hypothesis tests

Parametric tests are based on an assumption that the data follow a known probability distribution, often the Normal distribution. Non-parametric tests do not assume any specific distribution for the data.

1.7 Hypothesis testing versus confidence intervals

Confidence intervals and hypothesis tests are closely related. If the hypothesised value (e.g., zero for a difference in means) for an effect lies outside the 95% confidence interval then the hypothesised value is implausible and the p-value will be less than 0.05.

A confidence interval gives more information than a p-value – as well as indicating statistical significance it also gives the possible range of plausible values for an effect.

1.8 Errors in hypothesis testing

- Type I error: rejecting the null hypothesis when it is true
- Type II error: not rejecting the null hypothesis when it is false

1.9 Multiple testing

When the null hypothesis is true and several hypothesis tests are conducted, the chance of making at least one Type I error is increased. Strategies to combat this include:

- Limiting the number of hypothesis tests conducted
- Using a reduce level of significance (e.g., $p < 0.001$ instead of $p < 0.05$)
- Bonferroni approach: divide the significance level by the number of tests conducted (e.g., if there are ten tests use $p < 0.005$ instead of $p < 0.05$)

1.10 Practical (or clinical) importance versus statistical significance

A test can be statistically significant ($p < 0.05$) but the observed effect may not be practically or clinically important. More commonly, especially for small samples, a study may show an effect of a clinically important magnitude but which is not statistically significant ($p > 0.05$). Interpretation of a hypothesis test should consider both practical and statistical significance.

Chapter 2

One-sample t-test

2.1 When is it appropriate?

If you have one continuous variable and you wish to compare its mean to a pre-specified value.

2.2 Example

To test whether the mean diastolic blood pressure equals 75mmHg.

2.3 Null hypothesis

The mean diastolic blood pressure equals 75mmHg in the population.

2.4 Assumptions

- This is a parametric test so has an assumption which should be checked before proceeding
- Normal distribution of the continuous variable.
- Use a histogram to check this.

2.5 Using R

Read the data into R as `data.frame` (DF) if it is not available in the R environment.

2.5.1 One-sample t-test

- Note that the data is read as `data.frame` `DF` and available in the current R environment.
- The R function `t.test` is used to carry out one sample t-test.
- Place continuous variable diastolic blood pressure (`diastolic`) column as `x`.
- Type the hypothesised value of the mean diastolic blood pressure under `mu`
- Other arguments are shown but these are default arguments.
- Check the help file for `t.test` for further details.

```
t.test(x = DF$diastolic,
       alternative = 'two.sided',
       mu = 75, paired = FALSE,
       conf.level = 0.95)
```

One Sample t-test

```
data: DF$diastolic
t = 2.4311, df = 162, p-value = 0.01614
alternative hypothesis: true mean is not equal to 75
95 percent confidence interval:
 75.45380 79.38056
sample estimates:
mean of x
 77.41718
```

We can see that the mean is slightly higher than the hypothesised mean.

2.6 Steps to interpretation

1. Look at the result of the t-test: what is the p-value, what is the mean difference and what is the 95% Confidence interval for the mean difference?
 - (a) For the one-sample t-test, $p=0.02$.
 - (b) Mean difference is 2.42 indicating that the sample mean of 77.42mmHg is 2.42mmHg higher than the hypothesised population mean of 75.
 - (c) The 95% CI for this mean difference is (0.45, 4.38) indicating that the mean diastolic blood pressure in the population is between 0.45 and 4.38mmHg higher than the hypothesised mean diastolic blood pressure of 75mmHg.
2. A conclusion would include all the information in point one and the general statement that the mean diastolic blood pressure is significantly higher than 75mmHg in the population.

Chapter 3

Independent samples t-test

3.1 When is it appropriate?

If you have one continuous variable and a binary variable and you wish to compare the mean of the continuous variable between the two groups of the binary variable.

3.2 Example

To compare the mean total cholesterol between male and female patients (see the `cardiacdata.csv` dataset).

3.3 Null hypothesis

The mean total cholesterol of male patients is the same as the mean total cholesterol for female patients in the population.

3.4 Assumptions

This is a parametric test so has assumptions which should be checked before proceeding

- Independent groups e.g. male or female; treatment A or treatment B
- Normal distribution of continuous variable in each group – use a histogram to check this; note if this assumption is not met you need to use the Mann-Whitney test instead (see Factsheet Mann-Whitney test).
- Equal variance of continuous variable between groups – tested using Levene's test. This is done as you do the t-test (see below)

3.5 Using R

Read the data into R as `data.frame` (DF) if it is not available in the R environment.

3.5.1 Summary Statistic

- It is useful to use the summary statistics to check the summary statistics and plots.
- Use the `DescTools::Desc` function to obtain the summary statistics and plots.

```
# Boxplot
```

```
DescTools::Desc(tchol ~ sex, data = DF, plotit = TRUE)
```

```
-----  
tchol ~ sex
```

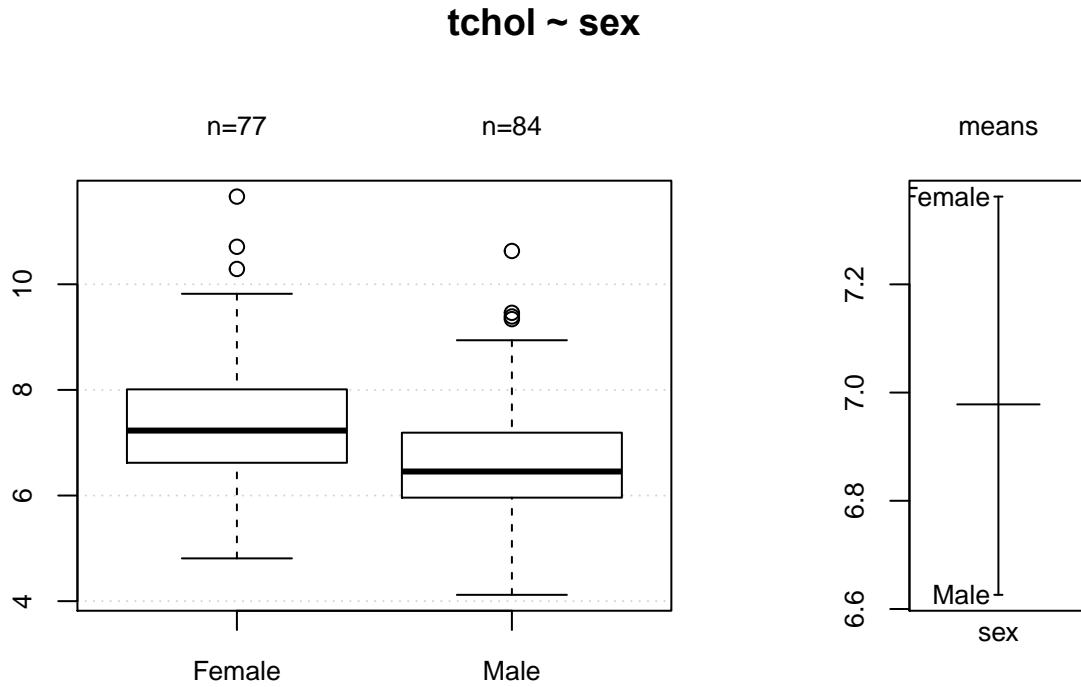
Summary:

n pairs: 163, valid: 161 (98.8%), missings: 2 (1.2%), groups: 2

	Female	Male
mean	7.362	6.626
median	7.230	6.455
sd	1.257	1.238
IQR	1.390	1.215
n	77	84
np	47.826%	52.174%
NAs	1	1
Os	0	0

Kruskal-Wallis rank sum test:

Kruskal-Wallis chi-squared = 15.934, df = 1, p-value = 6.561e-05



Histogram

```
DescTools::Desc(DF[DF$sex=='Female', 'tchol'], plotit = TRUE)
```

```
DF[DF$sex == "Female", "tchol"] (numeric)
```

length	n	NAs	unique	0s	mean	meanCI
78	77	1	75	0	7.362	7.077
	98.7%	1.3%		0.0%		7.648
.05	.10	.25	median	.75	.90	.95
5.564	5.748	6.620	7.230	8.010	8.880	9.236
range	sd	vcoef	mad	IQR	skew	kurt
6.850	1.257	0.171	1.067	1.390	0.715	1.096

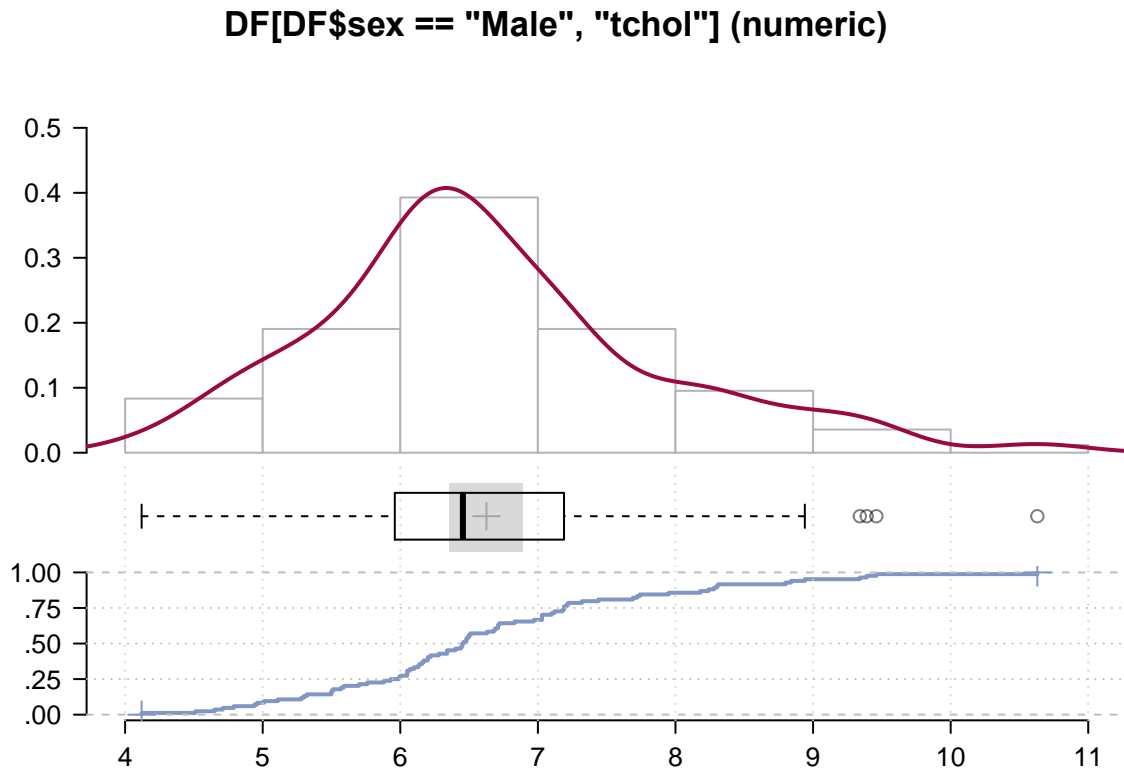
```
lowest : 4.81, 4.96, 5.27, 5.5, 5.58
```

```
highest: 9.09, 9.82, 10.29, 10.71, 11.66
```



```
DF[DF$sex == "Male", "tchol"] (numeric)
```

```
lowest : 4.12, 4.51, 4.65, 4.71, 4.79
highest: 8.94, 9.34, 9.39, 9.46, 10.63
```



3.5.2 Levene's test

- Note that the data is read as data.frame `DF` and available in the current R environment.
- First, we will conduct the Levene's test to check the equal variance of continuous variable between groups.
- The R function `LeveneTest` from the `library(DescTools)` or the function `leveneTest` from the `library(car)` conducts the Levene's test. Here we will be using `library(DescTools)` that you have used earlier for descriptive statistics.
- We will use the formula option of R (`y ~ x`) to run the test
- For this example, `y` is the continuous variable total cholesterol (`tchol`) column.
- The `x` is the factor variable sex (`sex`) column.

```
library(DescTools)
```

```
DescTools::LeveneTest(tchol ~ sex, data = DF)
```

```
Levene's Test for Homogeneity of Variance (center = median)
```

```
  Df F value Pr(>F)
group 1  0.0361 0.8495
159
```

3.5.3 Independent samples t-test

- The R function `t.test` is used to carry out independent samples t-test.
- If the data on two groups are available in two different columns, we can use argument `x` and `y` of each group.
- Our dataset has the variable and group as two columns, We will, therefore, use the formula option of R (`y ~ x`)
- For this example, `y` is the continuous variable total cholesterol (`tchol`) column.
- The `x` is the factor variable sex (`sex`) column.
- If the p-value is > 0.05 , you can assume equal variance and use `var.equal = TRUE` in the `t.test`.
- If the $p < 0.05$ then you cannot assume equal variances and use `var.equal = FALSE` in the `t.test`.
- In general, it is a standard practice in R to include `var.equal = FALSE` and let the function `t.test` to test the homogeneity of variance and use appropriate t-test. Hence you do not need to conduct the Levene test explicitly.
- We consider the default hypothesised value of the mean difference, i.e. $\mu = 0$
- Other arguments are shown but these are default arguments.
- Check the help file for `t.test` for further details.

```
t.test(tchol ~ sex, data = DF,
       alternative = 'two.sided',
       mu = 0, paired = FALSE,
       var.equal = TRUE, conf.level = 0.95)
```

Two Sample t-test

```
data:  tchol by sex
t = 3.7402, df = 159, p-value = 0.0002563
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3473626 1.1246721
sample estimates:
mean in group Female    mean in group Male
      7.362208           6.626190
```


3.6 Steps to interpretation

1. Check Levene's test to see if you can assume equal variances.
 - a. If the p-value is > 0.05 , you can assume equal variance and use `var.equal = TRUE` option in the `t.test` function.
 - b. If the $p < 0.05$ then you cannot assume equal variances and must use `var.equal = FALSE` option in the `t.test` function. Note this is the default option for the `t.test` function. The function automatically checks for the equality of the variance and use appropriate t-test.
 - c. In the example $p = 0.85$, so we can assume equal variances and use the top line of the results for the independent t-test.
2. Look at the result of the t-test: what is the p-value, what is the mean difference and what is the 95% Confidence interval for the mean difference?
 - a. For the independent t-test, $p < 0.001$.
 - b. Mean difference is 0.74 (7.36 - 6.62) indicating that on average total cholesterol is 0.74mmol/l more for females compared to males.
 - c. The 95% CI for this mean difference is (0.35, 1.12) indicating that the mean total cholesterol for female patients in the population is between 0.35 and 1.12mmol/l higher than that for male patients.
3. A conclusion would include all the information in point two and the general statement that female patients have significantly higher total cholesterol compared to male patients.

Note: If the normality assumption is not met, use *Mann-Whitney test* (check the relevant chapter in Factsheet).

Chapter 4

Paired samples t-test

4.1 When is it appropriate?

If you have two measurements of the same continuous variable at different time points and you wish to test for a change in the mean of the variable.

4.2 Example

To compare the mean systolic blood pressure at baseline and at follow-up (see the `cardiacdata.csv` dataset).

4.3 Null hypothesis

The mean systolic blood pressure at follow-up is the same as the mean systolic blood pressure at baseline in the population.

4.4 Assumptions

This is a parametric test so has assumptions which should be checked before proceeding.

- Normal distribution of the *difference* in systolic blood pressure at follow-up and at baseline.
- Calculate a new variable which is the difference and check its distribution using a histogram.
- If the assumption is not met you need to use the Wilcoxon test (see Factsheet chapter on Wilcoxon test).

4.5 Using R

Read the data into R as `data.frame` (DF) if it is not available in the R environment.

4.5.1 Summary Statistic

- It is useful to use the summary statistics to check the summary statistics and plots.
- Use the `DescTools::Desc` function to obtain the summary statistics and plots.

```
DF$diff_systolic <- as.numeric(DF$systolic - DF$systolic10)
DescTools::Desc(DF$diff_systolic, plotit = TRUE)
```

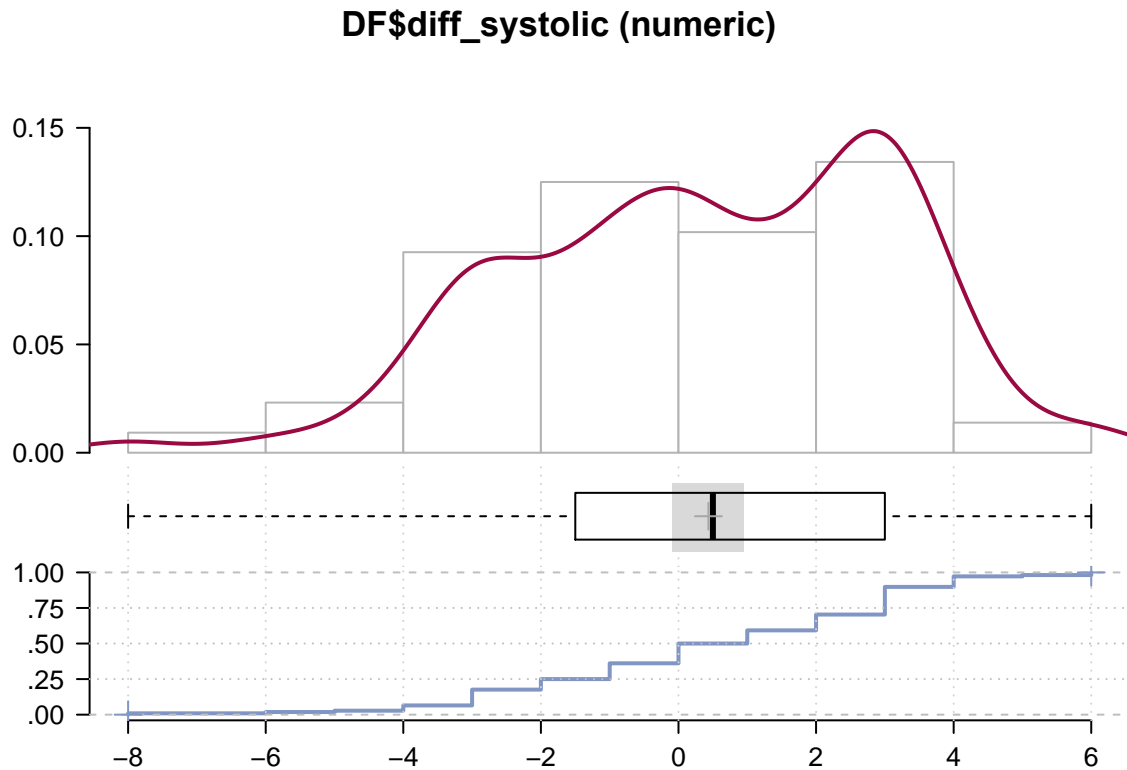
```
DF$diff_systolic (numeric)
```

length	n	NAs	unique	Os	mean	meanCI
163	108	55	14	15	0.44	-0.08
	66.3%	33.7%		9.2%		0.95
.05	.10	.25	median	.75	.90	.95
-4.00	-3.00	-1.25	0.50	3.00	3.30	4.00
range	sd	vcoef	mad	IQR	skew	kurt
14.00	2.72	6.25	3.71	4.25	-0.37	-0.33

```
lowest : -8.0, -6.0, -5.0, -4.0 (4), -3.0 (12)
```

```
highest: 2.0 (12), 3.0 (21), 4.0 (8), 5.0, 6.0 (2)
```

```
heap(?): remarkable frequency (19.4%) for the mode(s) (= 3)
```



4.5.2 Paired samples t-test

- The R function `t.test` is used to carry out paired sample t-test.
- Our dataset has baseline and follow-up systolic blood pressure values in two different columns; we can use argument `x` and `y` of each group.
- For this example, `x` is the baseline systolic blood pressure (`systolic`) column.
- The argument `y` is the follow-up systolic blood pressure (`systolic10`) column.
- We consider the default hypothesised value of the mean difference, i.e. $\mu = 0$
- Change the argument `paired = TRUE` to indicate paired samples t-test
- Other arguments are shown but these are default arguments.
- Check the help file for `t.test` for further details.

```
t.test(x = DF$systolic, y = DF$systolic10,
       alternative = 'two.sided',
       mu = 0, paired = TRUE,
       conf.level = 0.95)
```

Paired t-test

```

data: DF$systolic and DF$systolic10
t = 1.6617, df = 107, p-value = 0.09949
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.08396881  0.95433918
sample estimates:
mean of the differences
      0.4351852

```

You should also generate the table with the descriptive information (mean and standard deviation) for systolic blood pressure at baseline and at follow-up. We can see that there is a small decrease in systolic blood pressure. You can also calculate the correlation between baseline and follow-up systolic blood pressure. We can see they are very highly correlated as one would expect, with a $p\text{-value} < 0.001$ (check the Factsheet on *Correlation*).

4.6 Steps to interpretation

1. Look at the result of the t-test: what is the p-value, what is the mean difference and what is the 95% Confidence interval for the mean difference?
 - a. For the paired t-test, $p = 0.1$.
 - b. Mean difference is 0.435 indicating that on average systolic blood pressure lowers by 0.435mmHg from baseline to follow-up.
 - c. The 95% CI for this mean difference is (-0.08, 0.95) indicating that the mean systolic blood pressure in the population is between 0.08mmHg higher and 0.95mmHg lower at follow-up.
2. A conclusion would include all the information in point one and the general statement that there is no significant change in systolic blood pressure from baseline to follow-up.

Note: If the normality assumption is not met, use *Wilcoxon Matched Pairs test* (check the relevant chapter in Factsheet).

Chapter 5

Analysis of Variance (ANOVA)

5.1 When is it appropriate?

If you have one continuous variable and a categorical variable with more than 2 groups and you wish to compare the mean of the continuous variable between the groups of the categorical variable.

5.2 Example:

To compare the mean total cholesterol between 3 categories of smoking group (Current, Ex, Never)

5.3 Null hypothesis

The mean total cholesterol is the same in all 3 smoking groups in the population.

5.4 Assumptions

This is a parametric test so has assumptions which should be checked before proceeding.

- Independent groups e.g. Current, Ex, Never
- Normal distribution of continuous variable in each of the 3 groups – use a histogram to check this; note if this assumption is not met you need to use the Kruskal-Wallis test instead (see Factsheet on Kruskal-Wallis).
- Equal variance of continuous variable between groups – tested using Levene's test. This is an option that can be selected as you do the ANOVA test (see below)

5.5 Using R

5.5.1 Summary Statistic

- It is useful to use the summary statistics to check the summary statistics and plots.
- Use the `DescTools::Desc` function to obtain the summary statistics and plots.

```
DescTools::Desc(tchol ~ smoking, data = DF, plotit = TRUE)
```

```
tchol ~ smoking
```

Summary:

n pairs: 163, valid: 154 (94.5%), missings: 9 (5.5%), groups: 3

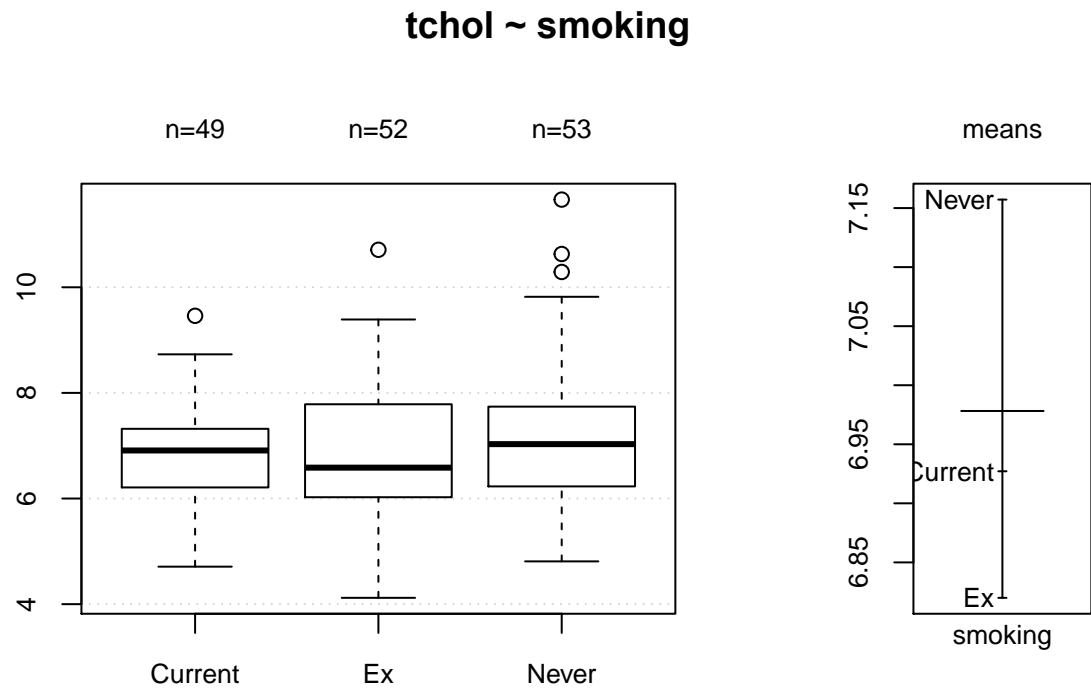
	Current	Ex	Never
mean	6.927	6.820	7.157
median	6.910	6.585	7.030
sd	0.991	1.376	1.450
IQR	1.110	1.635	1.510
n	49	52	53
np	31.818%	33.766%	34.416%
NAs	1	0	1
Os	0	0	0

Kruskal-Wallis rank sum test:

Kruskal-Wallis chi-squared = 1.7511, df = 2, p-value = 0.4166

Warning:

Grouping variable contains 7 NAs (4.29%).



```
# Histogram
DescTools::Desc(DF[DF$smoking=='Current', 'tchol'], plotit = TRUE)
```

DF[DF\$smoking == "Current", "tchol"] (numeric)

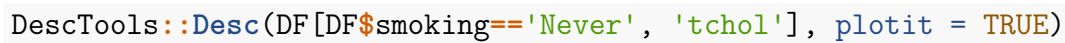
length	n	NAs	unique	Os	mean	meanCI
57	49	8	48	0	6.927	6.642
	86.0%	14.0%		0.0%		7.212

.05	.10	.25	median	.75	.90	.95
5.504	5.748	6.210	6.910	7.320	8.318	8.664

range	sd	vcoef	mad	IQR	skew	kurt
4.750	0.991	0.143	0.771	1.110	0.248	-0.089

lowest : 4.71, 4.96, 5.5, 5.51, 5.7

highest: 8.39, 8.64, 8.68, 8.73, 9.46

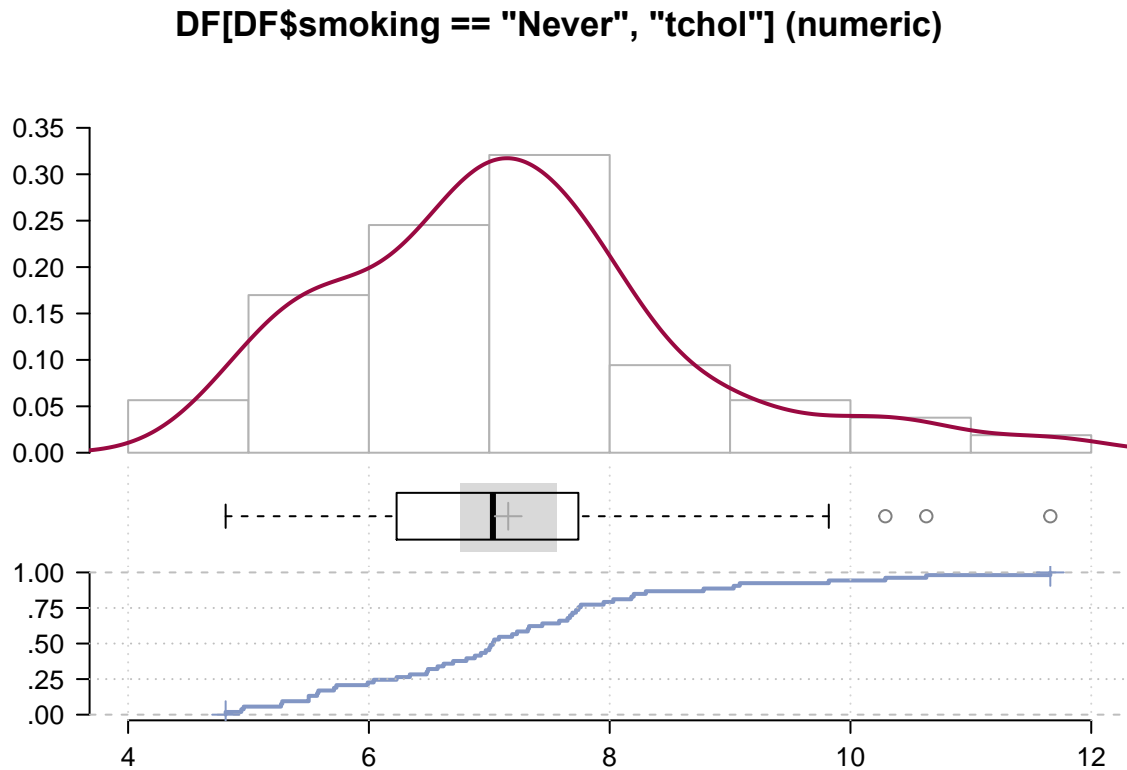


```
DF[DF$smoking == "Never", "tchol"] (numeric)
```

length	n	NAs	unique	0s	mean	meanCI
61	53	8	52	0	7.157	6.758
	86.9%	13.1%		0.0%		7.557
.05	.10	.25	median	.75	.90	.95
5.146	5.500	6.230	7.030	7.740	8.980	10.008
range	sd	vcoef	mad	IQR	skew	kurt
6.850	1.450	0.203	1.082	1.510	0.809	0.786

lowest : 4.81, 4.94, 4.96, 5.27, 5.28

highest: 9.08, 9.82, 10.29, 10.63, 11.66



5.5.2 Levene's test

- Note that the data is read as data.frame DF and available in the current R environment.
- First, we will conduct the Levene's test to check the equal variance of continuous variable between groups.
- The R function `LeveneTest` from the `library(DescTools)` or the function `leveneTest` from the `library(car)` conducts the Levene's test. Here we will be using `library(DescTools)` that you have used earlier for descriptive statistics.
- We will use the formula option of R ($y \sim x$) to run the test
- For this example, y is the continuous variable total cholesterol (`tchol`) column.
- The x is the factor variable smoking group (`smoking`) column.
- If the p-value is > 0.05 , you can assume equal variance of y and use ANOVA shown later
- If the $p < 0.05$ then you cannot assume equal variances of y and use alternative to ANOVA (for example, Kruskal-Wallis test instead; see Factsheet Kruskal-Wallis).

```
library(DescTools)
```

```
DescTools::LeveneTest(tchol ~ smoking, data = DF)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	2	2.0088	0.1377
	151		

5.5.3 ANOVA

- The p-value is greater than 0.05 ($p > 0.05$); therefore we can assume equal variance of `tchol`.
- The R function `aov` (stands for *analysis of variance*) is used to carry out ANOVA.
- We will use the formula option of R (`y ~ x`) to fit the model
- For this example, `y` is the continuous variable total cholesterol (`tchol`) column.
- The `x` is the factor variable smoking group (`smoking`) column.
- Use the function `anova(fm)` to print the analysis of variance table of the fitted model
- Use the function `summary(fm)` to print the summary information of the fitted table

```
fm <- aov(tchol ~ smoking, data = DF)
```

```
anova(fm)
```

Analysis of Variance Table

Response: `tchol`

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
smoking	2	3.117	1.5584	0.9305	0.3966
Residuals	151	252.900	1.6748		

5.6 Steps to interpretation

1. Check Levene's test to see if you can assume equal variances (the first table).
 - a. If the p-value is greater than 0.05 ($p > 0.05$) you can assume equal variance.
 - b. If the $p < 0.05$ then you cannot assume equal variances.
 - c. In the example $p = 0.138$ so we can assume equal variances.
2. Look at the result of the ANOVA (second) table: what is the p-value?
 - a. The p value here is 0.397 (which is $p > 0.05$) so we can conclude that there is no evidence that mean total cholesterol is different between smoking groups in the population.
3. How do we know which groups differ from others? We know this by looking at the table for the Post Hoc test (Scheffe) that we have produced. **Note that it is only appropriate to look at the post hoc results if the overall ANOVA is significant!**

Note: If the normality assumption or equality of variance is not met, use *Kruskal-Wallis test* (check the relevant chapter in Factsheet).

5.6.1 Post-hoc comparison of means

- The overall p-value for `smoking` from the ANOVA table is greater than 0.05 ($p > 0.05$); therefore, there is no evidence that mean total cholesterol is different between smoking groups in the population.
- Hence we do NOT need to conduct any post-hoc comparison of means.
- If the p-value is less than 0.05 ($p > 0.05$), we may conduct post-hoc comparison of means.
- The following code shows how to conduct post-hoc comparison of means
- *Remember this is for demonstration ONLY; we will not conduct the post-hoc test in real scenario for this data as the overall $p > 0.05$*
- The R function `PostHocTest` from the `library(DescTools)` can conduct several post-hoc tests.
- Here we conduct Scheffe test (this is a post-hoc test that shows which groups differ)

```
DescTools::PostHocTest(fm, method = 'scheffe')
```

```
Posthoc multiple comparisons of means : Scheffe Test
  95% family-wise confidence level
```

```
$smoking
```

	diff	lwr.ci	upr.ci	pval
Ex-Current	-0.1071429	-0.7441375	0.5298518	0.9172
Never-Current	0.2300270	-0.4040455	0.8640994	0.6696
Never-Ex	0.3371698	-0.2873264	0.9616661	0.4125

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note: The Scheffe test produces the *adjusted* p-values. If the adjusted p-value for a mean difference is less than 0.05 ($p < 0.05$), the mean difference is considered as statistically significant.

Chapter 6

Chi-squared test with continuity correction (2×2 Tables)

A common situation is when you want to examine a relationship between two binary variables. This leads to a 2×2 contingency table containing exactly four cells. To perform a hypothesis test on these data, one of three tests may be appropriate depending on the situation:

- If the two variables are independent the **Chi-squared test with continuity correction** should be used, but only if the assumptions for the chi-squared test are met: i.e.
 - No more than 20% of cells have expected counts less than 5
 - No cell has an expected count less than 1
- See separate chapter on Pearson's chi-squared test for more details.
- For the 2×2 table situation, the assumptions of the chi-squared test will only be met if all four cells have expected counts above 5.
- If the two variables are independent and the assumptions of the chi-squared test are not met (i.e. any one cell has an expected count less than 5) it is not appropriate to use the continuity-corrected chi-squared test. In this situation an exact test (**Fisher's exact test**) may be used instead. Note: for larger tables, it may be possible to combine categories if the assumptions are not met but this is not possible for the 2×2 table situation.
- If the two variables are paired (not independent) the **McNemar test** should be used (check the relevant chapter in Factsheet).

6.1 Chi-squared test with continuity correction

6.1.1 Example

We wish to determine whether there is an association between smoking status (`smokegr`) and social class (`soccl2`) in the dataset `cardiacdata.csv`. NB: we will use the versions of these variables that have been categorised into two groups (`smokegr` and `soccl2`).

6.1.2 Null hypothesis

In the population equal proportions of people are smokers in these two groups: “high” and “low” social class.

6.1.3 Assumptions

- The two groups are independent
- All four cells of the 2×2 table have expected values greater than 5

6.2 Using R

Read the data into R as `data.frame` (DF) if it is not available in the R environment.

6.2.1 Summary Statistic

- It is also useful to use the summary statistics to check the assumptions of the chi-squared test interpret the results.
- We can use the function `gmodels::CrossTable` from the `gmodels` package to obtain the summary statistics.
- Load the `library(gmodels)` if it is not already loaded
- Note that the `gmodels` package also provides the chi-squared test outputs.
- Note we have put values as `FALSE` for arguments that we do not need.
- We, however, will discuss the chi-squared test with continuity correction in the next section.
- Alternatively, you can use the function `DescTools::Desc` function to obtain the summary statistics.

```
library(gmodels)

gmodels::CrossTable(x = DF$smokegr, y = DF$soccl2,
                    expected = FALSE, prop.r = TRUE, prop.c = TRUE, prop.t = FALSE,
                    prop.chisq = FALSE, chisq = FALSE, format = 'SPSS')
```

Cell Contents

Count
Row Percent
Column Percent

```
|-----|
```

Total Observations in Table: 156

DF\$smokegr	DF\$soccl2		Row Total
	High	Low	
Yes	16	34	50
	32.000%	68.000%	32.051%
	26.230%	35.789%	
No	45	61	106
	42.453%	57.547%	67.949%
	73.770%	64.211%	
Column Total	61	95	156
	39.103%	60.897%	

6.2.2 Using R to conduct Chi-squared test with continuity correction

- The R function `chisq.test` is used to carry out Chi-squared test with continuity correction.
- The R function `chisq.test` considers both table object as well as the column vector of factor variable.
- For this example, `x` is the smoking group (`smokegr`) column.
- The argument `y` is the social class (`soccl2`) column.
- Change the argument `correct = TRUE` to suggest continuity correction
- Check the help file for `chisq.test` for further details.

```
tab <- table(DF$smokegr, DF$soccl2)
```

```
tab
```

```
      High Low
Yes    16  34
No     45  61
```



```
chisq.test(tab, correct = TRUE)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data:  tab
X-squared = 1.1508, df = 1, p-value = 0.2834
```

```
# Alternatively, use the vector for each variable
# chisq.test(x = DF$smokegr, y = DF$soccl2, correct = TRUE)
```

6.2.3 Steps to interpretation

The results indicate that fewer people in the *high* social class group smoke (26.2% versus 35.8% in the *low* social class group). The lowest expected count is 19.6. As this is greater than 5 the assumptions are met and the appropriate test to use is the chi-squared test with continuity correction. The relevant p-value is 0.283; this is not statistically significant (as it is $p > 0.05$).

Conclusion:

There is no evidence that smoking rates differ between those of *high* and *low* social class in the population.

6.3 Fisher's exact test

- In order to carry out Fisher's exact test you would perform exactly the steps as above.
- In the above example if any of the four expected counts had been less than 5 the chi-squared test would not have been valid.
- R will show warning that Chi-squared approximation may be incorrect.
- Check also the expected cell frequency from the `gmodels::CrossTable` function.
- In such situation, use Fisher's exact test
- It is recommended to use the 2-sided p-value for the Fisher's exact test.

6.3.1 Using R to conduct Fisher's exact test

- The R function `fisher.test` is used to carry out Fisher's exact test.
- The R function `fisher.test` considers both table object as well as the column vector of factor variable.
- For this example, `x` is the smoking group (`smokegr`) column.

- The argument `y` is the social class (`soccl2`) column.
- Note that we consider `alternative = 'two.sided'`
- Check the help file for `fisher.test` for further details.

```
fisher.test(x = DF$smokegr, y = DF$soccl2, alternative = 'two.sided')
```

Fisher's Exact Test for Count Data

```
data:  DF$smokegr and DF$soccl2
p-value = 0.2245
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.2919432 1.3624845
sample estimates:
odds ratio
 0.6397302
```

Note: The *Fisher's exact test* shown above is for demonstration ONLY. Since the data meet the assumptions of the Chi-squared test with continuity correction, we will *not* conduct the Fisher's exact test for this particular scenario.__

Chapter 7

Chi-squared test for trend

7.1 When is it appropriate?

When you have two independent categorical (ordinal) variables or one binary and one ordinal variables, that form an $r \times c$ contingency table, where r = number of rows and c = number of columns, and the assumptions of the chi-squared test are met: each observation is independent of all the others (i.e., one observation per subject) and no more than 20% of the expected counts are less than 5 and all individual expected counts are 1 or greater are met.

7.2 Chi-squared test for trend

7.2.1 Example

To compare the proportion of males/females (**sex**) across their leisure activity levels (**leisact**) in the dataset `cardiacdata.csv`. There are four leisure activity categories: No activity, At most light, At most moderate, At most strenuous.

7.2.2 Null hypothesis

In the population, there are equal proportions in each leisure activity category for males and females.

7.2.3 Assumptions

Individual expected cell count not < 1 , at least 80% of the expected counts should be > 5

7.3 Using R

Read the data into R as `data.frame` (DF) if it is not available in the R environment.

7.3.1 Summary Statistic

- It is also useful to use the summary statistics to check the assumptions of the chi-squared test interpret the results.
- We will use the function `gmodels::CrossTable` from the `gmodels` package to obtain the summary statistics.
- Load the `library(gmodels)` if it is not already loaded
- Note that the `gmodels` package also provides the chi-squared test outputs.
- Note we have put values as `FALSE` for arguments that we do not need.
- We, however, will discuss the chi-squared test for trend in the next section.
- Alternatively, you can use the function `DescTools::Desc` function to obtain the summary statistics.

```
library(gmodels)

gmodels::CrossTable(x = DF$sex, y = DF$leisact,
                    expected = FALSE, prop.r = TRUE, prop.c = TRUE, prop.t = FALSE,
                    prop.chisq = FALSE, chisq = FALSE, format = 'SPSS')
```

Cell Contents

```
|-----|
|              Count |
|          Row Percent |
|      Column Percent |
|-----|
```

Total Observations in Table: 163

	DF\$leisact				
DF\$sex	No	Light	Moderate	Strenuous	Row Total
Female	11	33	31	3	78
	14.103%	42.308%	39.744%	3.846%	47.853%
	78.571%	53.226%	44.286%	17.647%	
Male	3	29	39	14	85
	3.529%	34.118%	45.882%	16.471%	52.147%
	21.429%	46.774%	55.714%	82.353%	
Column Total	14	62	70	17	163
	8.589%	38.037%	42.945%	10.429%	

7.3.2 Using R to conduct Chi-squared test for trend

- The function `DescTools::MHChisqTest` is used to carry out Chi-squared test for trend.
- This is also called as Mantel-Haenszel Chi-Square Test.
- The function considers only the table object.
- For this example, `x` is the sex (`sex`) column.
- The argument `y` is the leisure activity (`leisact`) column.
- Check the help file for `MHChisqTest` for further details.

```
tab <- as.array(table(DF$sex, DF$leisact))
```

```
tab
```

	No	Light	Moderate	Strenuous
Female	11	33	31	3
Male	3	29	39	14

```
DescTools::MHChisqTest(tab)
```

Mantel-Haenszel Chi-Square

```
data:  tab
X-squared = 11.34, df = 1, p-value = 0.0007584
```

7.3.3 Steps to interpretation

From the output it is evident that the proportion of females doing at most strenuous activity (3.8%) is lower than the proportion of males (16.5%). Also, the proportion of female doing no activity (14.1%) is higher than the proportion of male (3.5%) doing no activity at all.

The p-value for the Chi-squared test for trend is $p = 0.001$ ($p < 0.05$), obtained from Linear-by-Linear association, indicating we reject the null hypothesis and conclude that there seems to be a linear trend between gender and leisure activity level at baseline in the population.

Conclusion:

In other words, there is an evidence of a trend whereby males tend to do more strenuous activities than females in the population.

Chapter 8

Chi-squared test for association ($r \times c$)

8.1 When is it appropriate?

When you have two independent categorical (nominal) variables that form an $r \times c$ contingency table, where r = number of rows and c = number of columns, and the assumptions of the chi-squared test are met: each observation is independent of all the others (i.e., one observation per subject) and no more than 20% of the expected counts are less than 5 and all individual expected counts are 1 or greater are met.

8.2 Chi-squared test for association

8.2.1 Example

To compare the proportion of people who smoke (`smokegr`) by marital status (`maritalstatus`) in the dataset `cardiacdata.csv`. There are three marital status categories: Never married, Married and Divorced.

8.2.2 Null hypothesis

There is no association between smoking status and marital status in the population.

8.2.3 Assumptions

No expected cell counts < 1 , at least 80% of the expected counts should be > 5 .

8.3 Using R

Read the data into R as `data.frame` (DF) if it is not available in the R environment.

8.3.1 Summary Statistic

- It is also useful to use the summary statistics to check the assumptions of the chi-squared test interpret the results.
- Use the function `gmodels::CrossTable` from the `gmodels` package to obtain the summary statistics.
- Load the `library(gmodels)` if it is not already loaded
- Note that the `gmodels` package also provides the chi-squared test outputs.
- Note we have put values as `FALSE` for arguments that we do not need.
- We, however, will discuss the chi-squared test for trend in the next section.
- Alternatively, you can use the function `DescTools::Desc` function to obtain the summary statistics.

```
library(gmodels)
```

```
gmodels::CrossTable(x = DF$maritalstatus, y = DF$smokegr,
                    expected = FALSE, prop.r = TRUE, prop.c = TRUE, prop.t = FALSE,
                    prop.chisq = FALSE, chisq = FALSE, format = 'SPSS')
```

Cell Contents

```
|-----|
|              Count |
|              Row Percent |
|              Column Percent |
|-----|
```

Total Observations in Table: 156

DF\$maritalstatus	DF\$smokegr		Row Total
	Yes	No	
Never_married	22	29	51
	43.137%	56.863%	32.692%
	44.000%	27.358%	
Married	12	46	58
	20.690%	79.310%	37.179%
	24.000%	43.396%	
Divorced	16	31	47
	34.043%	65.957%	30.128%
	32.000%	29.245%	
Column Total	50	106	156
	32.051%	67.949%	

8.3.2 Chi-squared test for association

- The R function `chisq.test` is used to carry out Chi-squared test for association.
- The R function `chisq.test` considers both table object as well as the column vector of factor variable.
- For this example, `x` is the marital status (`maritalstatus`) column.
- The argument `y` is the smoking group (`smokegr`) column.
- Change the argument `correct = FALSE` to exclude continuity correction
- Check the help file for `chisq.test` for further details.

```
tab <- table(DF$maritalstatus, DF$smokegr)
```

```
tab
```

	Yes	No
Never_married	22	29
Married	12	46
Divorced	16	31

```
chisq.test(tab, correct = FALSE)
```

Pearson's Chi-squared test

data: tab

X-squared = 6.4014, df = 2, p-value = 0.04073

8.3.3 Steps to interpretation

The p-value for Pearson's chi-squared is $p = 0.041$ ($p < 0.05$) indicating we reject the null hypothesis and conclude there is an association between smoking status and marital status in the population. A higher proportion of people who have never married smoke (44%) compared with those who are divorced (32%) or married (24%).

Conclusion:

We conclude there is an association between smoking status and marital status in the population.

Note: If the data do not meet the assumptions of the Chi-squared test, you can apply the *Fisher's exact test* using the function `fisher.test` for the $r \times c$ scenario as discussed earlier (SPSS can only apply *Fisher's exact test* for 2×2 scenario).

Chapter 9

McNemar Test

9.1 When is it appropriate?

When you have two paired (i.e. non-independent) binary groups that form a 2x2 contingency table and you wish to determine whether this an association between these two groups.

9.1.1 Example

To compare whether the proportion of people who smoke is the same between baseline (`smokegr`) and ten-year follow-up (`smokegr10`) in the dataset `cardiacdata.csv`.

9.1.2 Null hypothesis

In the population the proportions who smoke are the same at baseline and at ten-year follow-up.

9.1.3 Assumptions

- The two groups are related or dependent.
- In this case we have paired data as we have data for each individual for both time points.

9.1.4 Summary Statistic

- It is also useful to use the summary statistics to check the assumptions of the chi-squared test interpret the results.
- Use the function `gmodels::CrossTable` from the `gmodels` package to obtain the summary statistics.
- Load the `library(gmodels)` if it is not already loaded
- Note that the `gmodels` package also provides the chi-squared test outputs.

- Note we have put values as `FALSE` for arguments that we do not need.
- We, however, will discuss the McNemar test in the next section.
- Alternatively, you can use the function `DescTools::Desc` function to obtain the summary statistics.

```
library(gmodels)

gmodels::CrossTable(x = DF$smokegr, y = DF$smokegr10,
                    expected = FALSE, prop.r = TRUE, prop.c = TRUE, prop.t = FALSE,
                    prop.chisq = FALSE, chisq = FALSE, format = 'SPSS')
```

Cell Contents

Count
Row Percent
Column Percent

Total Observations in Table: 104

DF\$smokegr	DF\$smokegr10		Row Total
	Yes	No	
Yes	23	4	27
	85.185%	14.815%	25.962%
	100.000%	4.938%	
No	0	77	77
	0.000%	100.000%	74.038%
	0.000%	95.062%	
Column Total	23	81	104
	22.115%	77.885%	

9.1.5 Using R to conduct McNemar test

- The R function `mcnemar.test` is used to carry out McNemar's chi-squared test.
- The R function `mcnemar.test` considers both table object as well as the column vector of factor variable.

- For this example, `x` is the smoking group (`smokegr`) column.
- The argument `y` is the social class (`soccl2`) column.
- Change the argument `correct = TRUE` to suggest continuity correction
- Check the help file for `mcnemar.test` for further details.

```
tab <- table(DF$smokegr, DF$smokegr10)
```

```
tab
```

	Yes	No
Yes	23	4
No	0	77

```
mcnemar.test(tab, correct = TRUE)
```

McNemar's Chi-squared test with continuity correction

```
data: tab
```

```
McNemar's chi-squared = 2.25, df = 1, p-value = 0.1336
```

9.1.6 Steps to interpretation

We observe that 4 (14.8%) of the 27 people smoking at baseline no longer smoked at ten-year follow-up. All 77 (100%) non-smokers at baseline were still non-smokers at follow-up. The p-value resulting from the McNemar test ($p = 0.134$), however, is not statistically significant as it is greater than 0.05.

Conclusion:

There is therefore no evidence of a difference between rates of smoking at baseline and ten-year follow-up in this population.

Chapter 10

Mann-Whitney test

10.1 When is it appropriate?

If you have one continuous variable and a categorical variable with two categories (binary variable) and you wish to compare the distribution of the continuous variable between the groups and

- the assumptions of the independent samples t-test are not met
- the continuous data are ranked or
- small sample size

10.1.1 Example

To compare the distribution of HDL cholesterol (`hdlchol`) among patients who smoke and do not smoke (`smokegr`) in the dataset `cardiacdata.csv`.

10.1.2 Null hypothesis

The distribution of HDL cholesterol is the same for smokers and non-smokers in the population.

10.1.3 Assumptions

- Data are skewed (checked using a histogram and/or boxplot of `hdlchol` in each group) and the groups are independent

10.2 Using R

Read the data into R as `data.frame` (DF) if it is not available in the R environment.

10.2.1 Summary Statistic

- It is useful to use the summary statistics to check the summary statistics and plots.
- Use the `DescTools::Desc` function to obtain the summary statistics and plots.

```
# Boxplot
```

```
DescTools::Desc(hdlchol ~ smokegr, data = DF, plotit = TRUE)
```

```
-----  
hdlchol ~ smokegr
```

Summary:

n pairs: 163, valid: 154 (94.5%), missings: 9 (5.5%), groups: 2

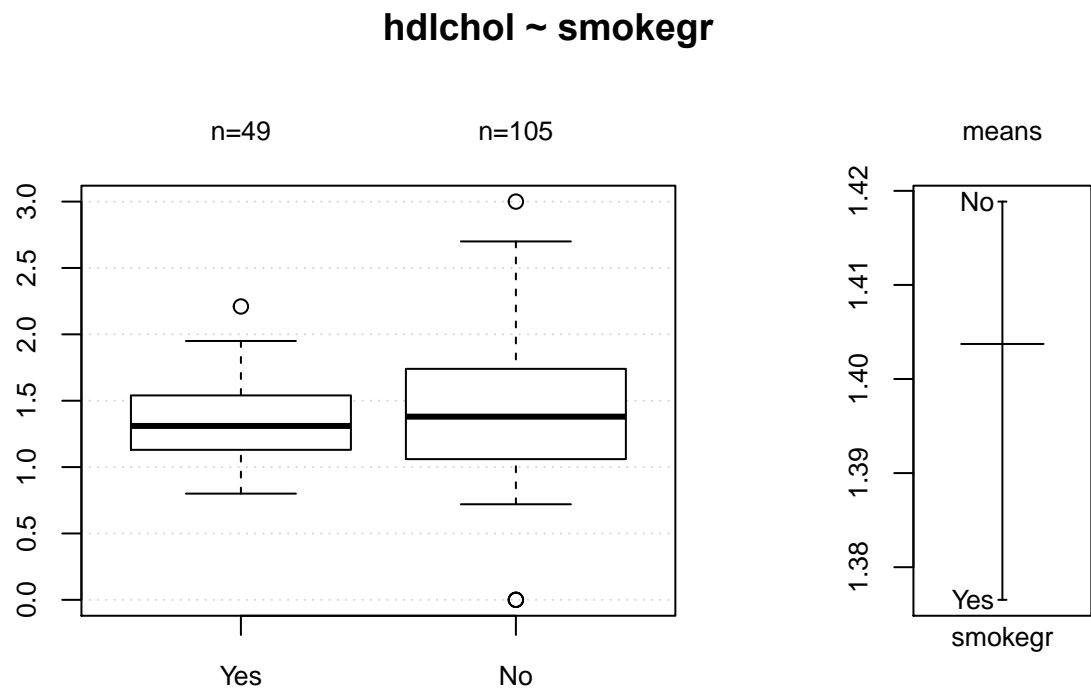
	Yes	No
mean	1.377	1.419
median	1.310	1.380
sd	0.318	0.510
IQR	0.410	0.680
n	49	105
np	31.818%	68.182%
NAs	1	1
Os	0	2

Kruskal-Wallis rank sum test:

Kruskal-Wallis chi-squared = 0.02592, df = 1, p-value = 0.8721

Warning:

Grouping variable contains 7 NAs (4.29%).



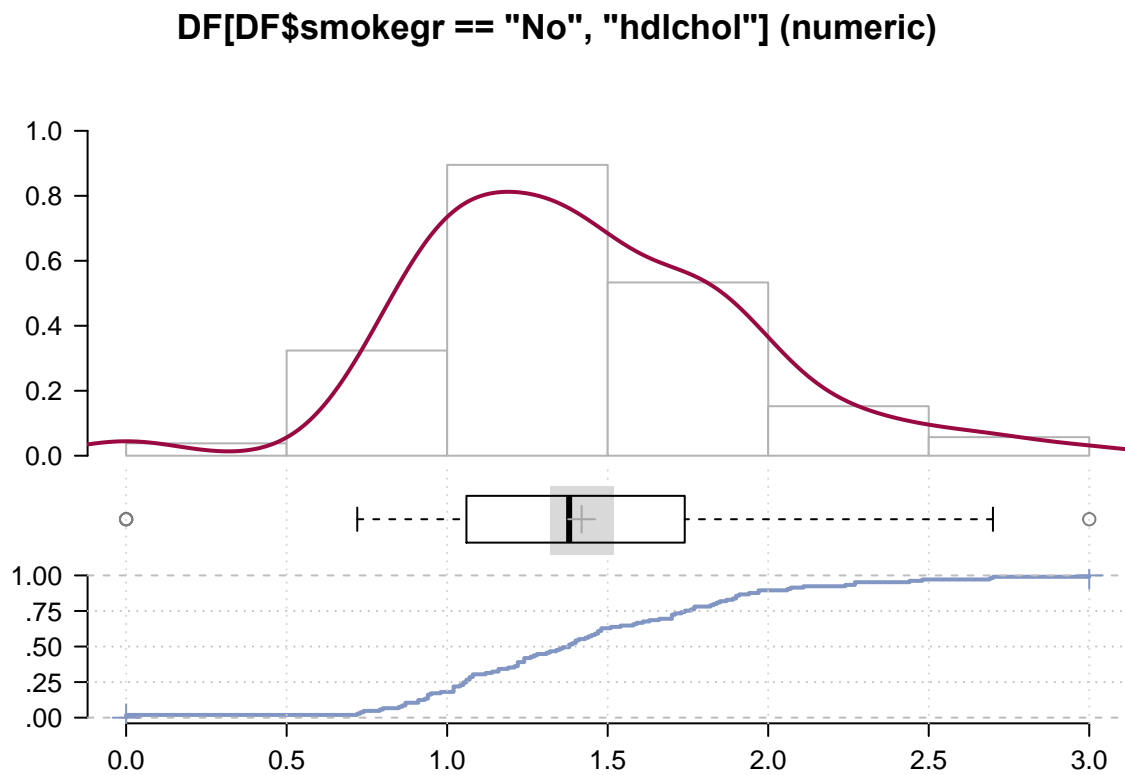
```
# Histogram
DescTools::Desc(DF[DF$smokegr=='No', 'hdlchol'], plotit = TRUE)
```

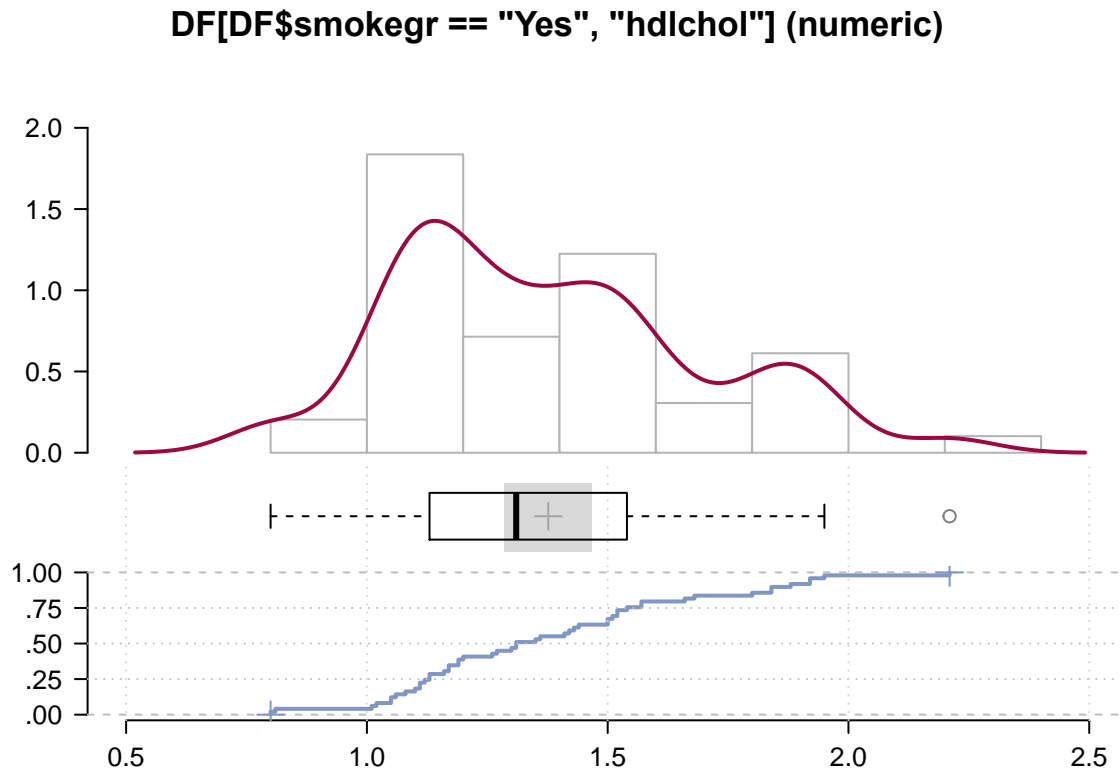
DF[DF\$smokegr == "No", "hdlchol"] (numeric)

length	n	NAs	unique	0s	mean	meanCI
113	105	8	78	2	1.419	1.320
	92.9%	7.1%		1.8%		1.517
.05	.10	.25	median	.75	.90	.95
0.792	0.886	1.060	1.380	1.740	2.024	2.270
range	sd	vcoef	mad	IQR	skew	kurt
3.000	0.510	0.359	0.489	0.680	0.364	0.801

lowest : 0.0 (2), 0.72, 0.73, 0.74, 0.79

highest: 2.44, 2.48, 2.69, 2.7, 3.0





10.2.2 Mann-Whitney test

- The R function `wilcox.test` performs one- and two-sample Wilcoxon tests on vectors of data
- The two-sample Wilcoxon tests is also known as ‘Mann-Whitney’ test
- If the data on two groups are available in two different columns, we can use argument `x` and `y` of each group.
- Our dataset has the variable and group as two columns, We will, therefore, use the formula option of R (`y ~ x`)
- For this example, `y` is the continuous variable HDL cholesterol (`hdlchol`) column.
- The `x` is the factor variable smoking group (`smokegr`) column.
- We consider the default hypothesised value of the mean difference, i.e. $\mu = 0$
- Some other arguments are shown but we can accept all default values for those arguments.
- Check the help file for `wilcox.test` for further details.

```
wilcox.test(hdlchol ~ smokegr, data = DF, na.action = na.exclude)
```

Wilcoxon rank sum test with continuity correction

data: hdlchol by smokegr

W = 2531, p-value = 0.8736

alternative hypothesis: true location shift is not equal to 0

10.2.3 Steps to interpretation

The p-value for the Mann-Whitney test is $p = 0.723$ indicating we do not reject the null hypothesis and conclude there is no difference in distribution of HDL cholesterol between the smokers and non-smokers. To fully present our results we need the median and inter quartile range for HDL cholesterol in each of the groups.

Conclusion:

The median (IQR) HDL cholesterol for smokers 1.31 (1.12, 1.56) compared to non-smokers 1.38 (1.06, 1.76) and the difference in these distributions is not significant, $p = 0.723$.

Chapter 11

Kruskal-Wallis test

11.1 When is it appropriate?

If you have one continuous variable and a categorical variable (more than two categories) and

- you wish to compare the distribution of the continuous variable between the groups and
- the assumptions of ANOVA are not met or
- the continuous data are ranked or
- small sample size

11.1.1 Example

To compare the distribution of triglyceride level (`triglyceride`) among patients of different diabetes status (`diabetic`) in the dataset `cardiacdata.csv`. There are three diabetic groups: normal, impaired glucose tolerant, diabetic

11.1.2 Null hypothesis

The distribution of amount triglyceride level is the same for the three diabetic groups in the population.

11.1.3 Assumptions

- Data are skewed (checked using a histogram of triglyceride in each group) and the groups are independent

11.2 Using R

Read the data into R as `data.frame` (DF) if it is not available in the R environment.

11.2.1 Summary Statistic

- It is useful to use the summary statistics to check the summary statistics and plots.
- Use the `DescTools::Desc` function to obtain the summary statistics and plots.
- Note that the `DescTools::Desc` function produces the Kruskal-Wallis rank sum test by default, but we will be using the appropriate function in the section.

```
DescTools::Desc(triglyceride ~ diabetic, data = DF, plotit = TRUE)
```

```
-----
triglyceride ~ diabetic
```

Summary:

n pairs: 163, valid: 155 (95.1%), missings: 8 (4.9%), groups: 3

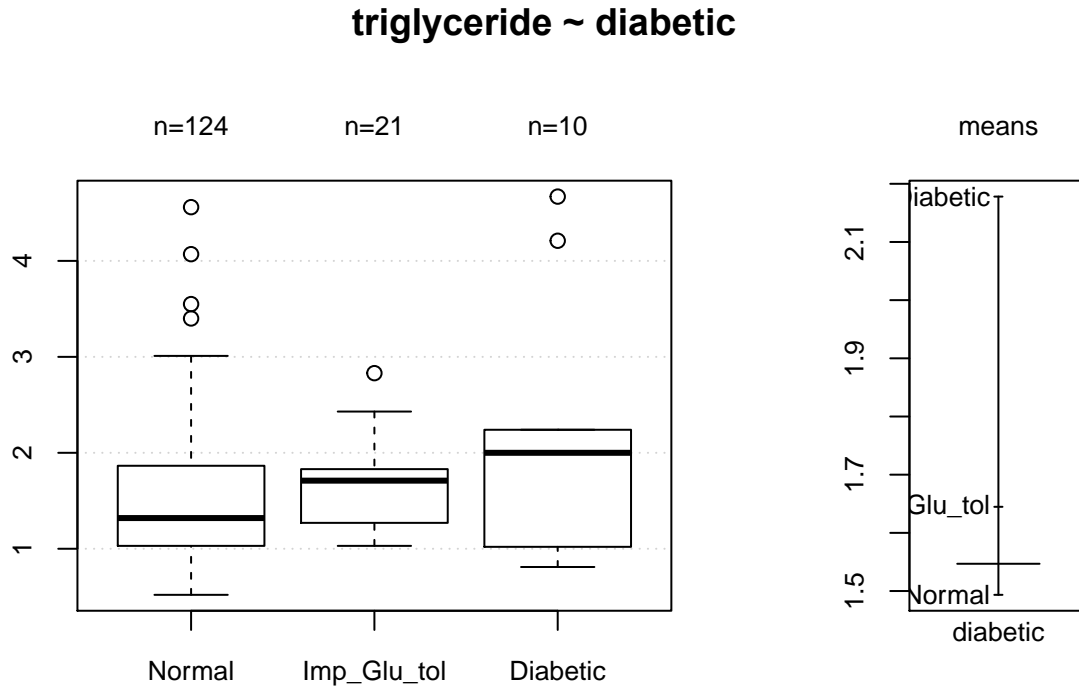
	Normal	Imp_Glu_tol	Diabetic
mean	1.493	1.645	2.178
median	1.320	1.710	2.000
sd	0.699	0.472	1.311
IQR	0.827	0.560	1.018
n	124	21	10
np	80.000%	13.548%	6.452%
NAs	0	0	1
Os	0	0	0

Kruskal-Wallis rank sum test:

Kruskal-Wallis chi-squared = 5.5397, df = 2, p-value = 0.06267

Warning:

Grouping variable contains 7 NAs (4.29%).



11.2.2 Kruskal-Wallis test

- The R function `kruskal.test` performs Kruskal-Wallis rank sum test.
- You can input the numeric vector and group as two arguments `x` and `g` respectively.
- Alternatively, you can use formula interface
- Here, we will use the formula option of R (`y ~ x`)
- For this example, `y` is the continuous variable Triglycerides (`triglyceride`) column.
- The `x` is the factor variable diabetic condition (`diabetic`) column.
- We can accept other arguments with its default values.
- Check the help file for `kruskal.test` for further details.

```
kruskal.test(triglyceride ~ diabetic, data = DF, na.action = na.exclude)
```

Kruskal-Wallis rank sum test

data: triglyceride by diabetic

Kruskal-Wallis chi-squared = 5.5397, df = 2, p-value = 0.06267

11.2.3 Steps to interpretation

The p-value for Kruskal-Wallis is $p = 0.031$ indicating we reject the null hypothesis and conclude there is a difference in distribution of triglyceride between the three levels of diabetic status. To fully present our results we need the median and inter quartile range for triglyceride in each of the groups.

Conclusion:

The median (IQR) triglyceride level for normal patients was 1.32 (1.025, 1.868), impaired glucose tolerant 1.71 (1.22, 1.85) and for diabetic patients 2.05 (1.02, 3.24) and the difference in these distributions is significant $p = 0.031$.

Chapter 12

Wilcoxon Matched pairs test

12.1 When is it appropriate?

If you have a continuous variable in paired groups and the differences are skewed e.g. measuring body mass index at two time points on the same people, perhaps before and after an intervention. Wilcoxon test is used when the assumption of normality for the paired sample t-test is not met or when you have a small sample size.

12.1.1 Example

To compare the distribution of BMI between baseline and follow-up in `cardiacdata.sav`.

12.1.2 Null hypothesis

There is no difference/ no change in the distribution of BMI in the population between baseline and follow up.

This could also be written as:

The distribution of difference in BMI in the population between baseline and follow up is around zero.

12.1.3 Assumptions

- No distributional assumptions are made but Wilcoxon test is used when differences are skewed (checked using a histogram) after first computing the variable equal to the difference between baseline and follow-up BMI.

12.2 Using R

Read the data into R as `data.frame` (DF) if it is not available in the R environment.

12.2.1 Summary Statistic

- It is useful to use the summary statistics to check the summary statistics and plots.
- Use the `DescTools::Desc` function to obtain the summary statistics and plots.

```
DescTools::Desc(DF$bmi, plotit = TRUE)
```

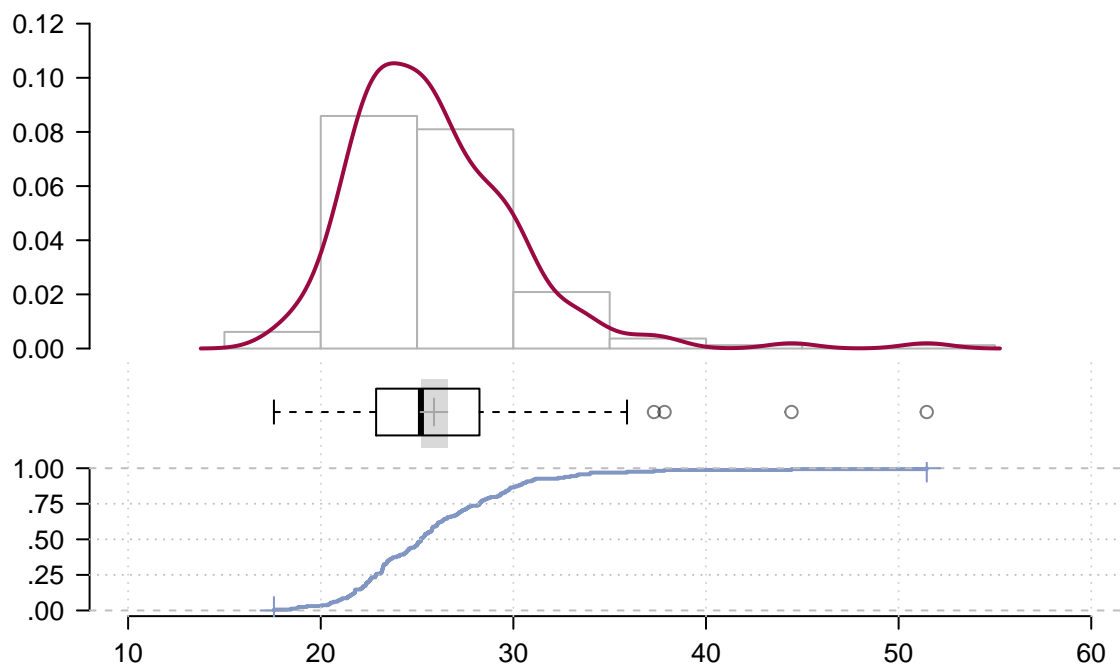
DF\$bmi (numeric)

length	n	NAs	unique	0s	mean	meanCI
163	163	0	152	0	25.879	25.183
	100.0%	0.0%		0.0%		26.574
.05	.10	.25	median	.75	.90	.95
20.525	21.520	22.875	25.200	28.240	30.602	33.238
range	sd	vcoef	mad	IQR	skew	kurt
33.890	4.498	0.174	3.810	5.365	1.813	6.766

lowest : 17.57, 18.46, 18.77, 18.84, 19.27

highest: 35.9, 37.3, 37.85, 44.44, 51.46

DF\$bmi (numeric)



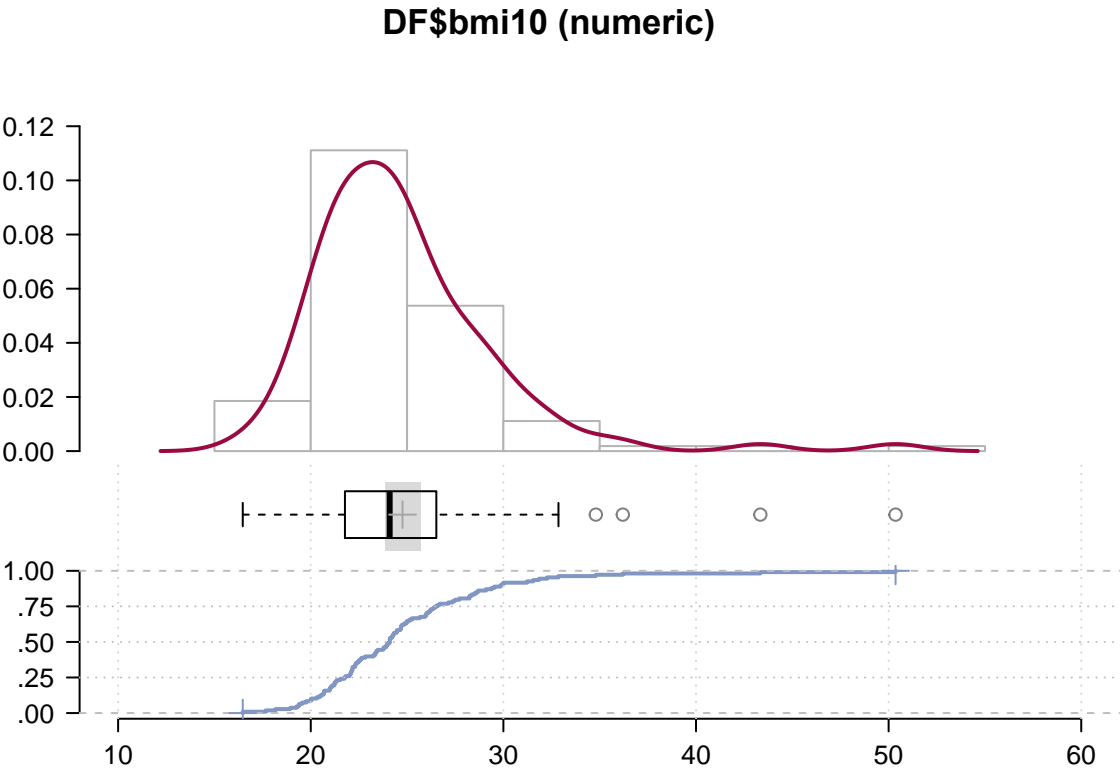

```
DescTools::Desc(DF$bmi10, plotit = TRUE)
```

DF\$bmi10 (numeric)

length	n	NAs	unique	0s	mean	meanCI
163	108	55	104	0	24.7641	23.8426
	66.3%	33.7%		0.0%		25.6855
.05	.10	.25	median	.75	.90	.95
19.3840	20.2230	21.7875	24.0950	26.4900	29.8770	32.1230
range	sd	vcoef	mad	IQR	skew	kurt
33.8900	4.8306	0.1951	3.4915	4.7025	2.0673	7.4349

lowest : 16.47, 17.67, 18.17, 18.94, 19.27

highest: 32.86, 34.8, 36.2, 43.34, 50.36



12.2.2 Wilcoxon Matched pairs test

- The R function `wilcox.test` performs Wilcoxon Matched pairs on vectors of paired data
- If the data on two groups are available in two different columns, we can use argument `x` and `y` of each group.
- For this example, we have `x` is the continuous variable BMI (`bmi`) at the baseline
- The `y` is the is the continuous variable BMI at the follow-up (`bmi10`)
- Include `paired = TRUE` to indicate Wilcoxon Matched pairs test
- We consider the default hypothesised value of the mean difference, i.e. `mu = 0`
- Other arguments are considered at its default values.
- Check the help file for `wilcox.test` for further details.

```
wilcox.test(x = DF$bmi, y = DF$bmi10, paired = TRUE, na.action = na.exclude)
```

Wilcoxon signed rank test with continuity correction

data: DF\$bmi and DF\$bmi10

V = 5886, p-value < 2.2e-16

alternative hypothesis: true location shift is not equal to 0

12.2.3 Steps to interpretation

The output provides the Wilcoxon test p-value which is found to be $p < 0.001$ indicating we can reject the null hypothesis of no change in BMI.

The direction of this change (i.e. at which point the BMI is higher/lower) can be ascertained by examining the medians and IQR for baseline and follow up periods. This can be obtained using the Explore menu.

Body mass index at baseline (kg/m^2); median (IQR): 25.2 (22.9, 28.2) Follow up body mass index (kg/m^2); median (IQR): 24.1 (21.8, 26.5)

Above estimates imply that there were more occasions where follow up BMI was smaller than baseline BMI (negative ranks) than occasions where follow up BMI was larger (actually zero).

Conclusion:

We can conclude that there was a significant reduction in BMI at the follow up period ($p < 0.001$).