Tutorial letter 101/3/2017

Applied Statistics II
STA2601

Semesters 1 & 2

Department of Statistics

IMPORTANT INFORMATION:
This tutorial letter contains important information about your module and includes the assignment questions for both semesters.
1 INTRODUCTION

Dear Student

Welcome to this module. We trust your studies will be rewarding and successful!

The module is called APPLIED STATISTICS II. The module is the follow-up on the module STA1502 (Statistical Inference I). The name Applied Statistics was chosen because of its double meaning: Data analysis is in effect applied statistical theory and you will learn how to apply the statistical software packages SAS JMP and R. This means that you must have access to a suitable computer for a component of practical work.

This module will equip you with a proper basis in statistical knowledge, introduce you to a statistical package and highlight the value of thorough statistical know-how that the business and outside world require of students who major in Statistics! Knowledge of statistics will enable you to conduct quantitative research and statistical literacy will enable you to understand research reports you might encounter as a scientist in your everyday life or enable you to understand statistical reports you might encounter as a manager in your business.

We trust that you will work seriously and continuously. We hope that you will enjoy this module and wish you all the best!

1.1 Tutorial matter

Take note that every tutorial letter you will be receiving is important and you have to read them all immediately and carefully. Some information contained in these tutorial letters may be urgent, while others may, for example, contain examination information. So, it is wise to keep them all in a file!

Some of this tutorial matter may not be available when you register. Tutorial matter that is not available when you register will be posted to you as soon as possible, but is also available on myUnisa.

At the time of registration, you will receive an inventory letter that will tell you what you have received in your study package and also show items that are still outstanding. Also see the brochure entitled my Studies @ Unisa.

Check the study material that you have received against the inventory letter. You should have received all the items listed in the inventory, unless there is a statement like “out of stock” or “not available”. If any item is missing, follow the instructions on the back of the inventory letter without delay.

Shortly after registration The Department of Despatch should supply you with the following tutorial matter for this module:

- **Tutorial letter 101.** Read it and save it as it contain important information as well as your assignments for the semester.

- **A study guide** written by a lecturer to guide you through the relevant sections in the prescribed book. Use it together with the textbook as the guide indicates the relevant prescribed sections, explaining difficult concepts in more detail, giving additional examples and exercises, etc.
- Other tutorial letters to further assist you with your studies, will be dispatched to you throughout the semester.

If you have access to the Internet, you can view the study guide and tutorial letters for the modules for which you are registered on the University’s online campus, myUnisa, at http://my.unisa.ac.za.

**There are two types of tutorial letters:**

- The 100-series (e.g. Tutorial letter 101, 102, 103, etc.) containing general information, assignment questions, information about your lecturer or the examination, a trial paper, etc.
- The 200-series (e.g. Tutorial letter 201, 202, 203, etc.) containing the solutions to the assignments and the trial paper.

## 2 PURPOSE OF AND OUTCOMES FOR THE MODULE

### 2.1 Purpose

Students credited with this unit standard, will be able to identify the correct technique, manage the statistical softwares SAS JMP and R to do the computations and interpret the results for decisions regarding tests for normality, independence and hypothesis concerning means, proportions, variances and regression. Students should be able to solve applied statistics problems arising in government and industry.

### 2.2 Outcomes

Qualifying students will be able to:

- describe various probability distributions and illustrate their applications as probabilities associated with critical values from the tables.
- describe desirable properties of estimators for population parameters and derive these estimators through the methods of maximum likelihood and least squares.
- use statistical softwares SAS JMP and R.
- do statistical estimation and hypothesis testing for a single population.
- test for normality by employing various techniques.
- do statistical estimation and hypothesis testing involving two populations.
- do statistical estimation and hypothesis testing involving more than two populations.
- measure relationships between variables.
3 LECTURER(S) AND CONTACT DETAILS

3.1 Lecturer(s)

The lecturer responsible for this module is as follows:

Ms S. Muchengetwa  
GJ GERWEL (C-Block), **Floor 6, Office 6-05**  
Tel: (011) 670-9253  
Cell: 074 065 9020  
E-mail address: muches@unisa.ac.za

You might also want to write to us. Letters should be sent to:

Ms S. Muchengetwa  
Department of Statistics  
PO Box 392  
UNISA  
0003

All queries that are not of a purely administrative nature but are about the content of this module should be directed to me. Please have your study material with you when you contact me. E-mail address is included above.

**PLEASE NOTE:** Letters to lecturers may not be enclosed with or inserted into assignments.

3.2 Department

The departmental secretary can be contacted at 011 670-9255 for other queries.

3.3 University

If you need to contact the University about matters not related to the content of this module, please consult the publication *My Studies @ Unisa* that you received with your study material. This brochure contains information on how to contact the University (e.g. to whom you can write for different queries, important telephone and fax numbers, addresses and details of the times certain facilities are open).

Always have your student number at hand when you contact the University.

4 MODULE RELATED RESOURCES

4.1 Prescribed books

The prescribed book for this semester is


You have to buy this book. Please consult the list of official booksellers and their addresses listed in *My Studies @ Unisa*. Prescribed books can be obtained from the University's official booksellers. If
you have difficulty locating your book(s) at these booksellers, please contact the Prescribed Books Section at 012 429 4152 or e-mail vospresc@unisa.ac.za. If you cannot find the book you can buy the latest edition.

You need to purchase one other publication. The publication is a book of tables containing the normal, $t$-, chi-squared and $F$-tables.


Foreign students may have difficulty in obtaining this book. If you are unable to obtain this book you may use any other book of tables, but keep in mind that the tables used in the examination will be the ones from Stoker.

### 4.2 Recommended books

There are no recommended books for this module.

### 4.3 Electronic Reserves (e-Reserves)

There are no e-Reserves for this module.

### 4.4 Library services and resources information

For brief information go to: http://www.unisa.ac.za/contents/studies/docs/myStudies-at-Unisa2017-brochure.pdf
For more detailed information, go to the Unisa website: http://www.unisa.ac.za/, click on Library
For research support and services of Personal Librarians, go to: http://www.unisa.ac.za/Default.asp?Cmd=ViewContent&ContentID=7102
The Library has compiled numerous library guides:

- find recommended reading in the print collection and e-reserves -http://libguides.unisa.ac.za/request/undergrad
- request material - http://libguides.unisa.ac.za/request/request
- postgraduate information services - http://libguides.unisa.ac.za/request/postgrad
- finding, obtaining and using library resources and tools to assist in doing research - http://libguides.unisa.ac.za/Research_Skills
- how to contact the Library/find us on social media/frequently asked questions - http://libguides.unisa.ac.za/ask

### 5 STUDENT SUPPORT SERVICES FOR THE MODULE

For information on the various student support systems and services available at Unisa (e.g. student counseling, tutorial classes, language support), please consult the publication *my Studies @ Unisa* that you received with your study material.
5.1 Contact with Fellow Students

5.1.1 Study Groups

It is advisable to have contact with fellow students. One way to do this is to form study groups. Please consult the publication my Studies@Unisa to find out how to obtain the addresses of students in your region.

5.1.2 myUnisa

If you have access to a computer that is linked to the internet, you can quickly access resources and information at the University. The myUnisa learning management system is Unisa’s online campus that will help you to communicate with your lecturers, with other students and with the administrative departments of Unisa - all through the computer and the internet.

To go to the myUnisa website, start at the main Unisa website, http://www.unisa.ac.za, and then click on the “Login to myUnisa” link on the right-hand side of the screen. This should take you to the myUnisa website. You can also go there directly by typing in http://my.unisa.ac.za.

Please consult the publication my Studies @ Unisa which you received with your study material for more information on myUnisa.

5.1.3 Discussion classes

There are no discussion classes offered in this module. Should the need for discussion classes arise in future, students will be informed in advance about actual dates and venues.

5.2 Free computer and internet access

Unisa has entered into partnerships with establishments (referred to as Telecentres) in various locations across South Africa to enable you (as a Unisa student) free access to computers and the Internet. This access enables you to conduct the following academic related activities: registration; online submission of assignments; engaging in e-tutoring activities and signature courses; etc. Please note that any other activity outside of these are for your own costing e.g. printing, photocopying, etc.

For more information on the Telecentre nearest to you, please visit www.unisa.ac.za/telecentres.
6 MODULE-SPECIFIC STUDY PLAN

<table>
<thead>
<tr>
<th>SEMESTER 1</th>
<th>Study units for preparing your assignments</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment 1</td>
<td>Chapter 1 to Chapter 2</td>
<td>Registration 11 March</td>
<td>10 March</td>
</tr>
<tr>
<td></td>
<td>Start writing your assignment</td>
<td>11 March</td>
<td>17 March</td>
</tr>
<tr>
<td>Assignment 2</td>
<td>Chapter 3 to Chapter 6</td>
<td>18 March</td>
<td>24 March</td>
</tr>
<tr>
<td></td>
<td>Start writing your assignment</td>
<td>25 March</td>
<td>31 March</td>
</tr>
<tr>
<td>Assignment 3</td>
<td>Chapter 6 to Chapter 8</td>
<td>1 April</td>
<td>14 April</td>
</tr>
<tr>
<td></td>
<td>Start writing your assignment</td>
<td>15 April</td>
<td>21 April</td>
</tr>
</tbody>
</table>

Note: For the text book, you will see the instructions in you workbook on which pages to read.

<table>
<thead>
<tr>
<th>SEMESTER 2</th>
<th>Study units for preparing your assignments</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment 1</td>
<td>Chapter 1 to Chapter 2</td>
<td>Registration 12 August</td>
<td>11 August</td>
</tr>
<tr>
<td></td>
<td>Start writing your assignment</td>
<td>12 August</td>
<td>18 August</td>
</tr>
<tr>
<td>Assignment 2</td>
<td>Chapter 3 to Chapter 6</td>
<td>19 August</td>
<td>25 August</td>
</tr>
<tr>
<td></td>
<td>Start writing your assignment</td>
<td>26 August</td>
<td>01 September</td>
</tr>
<tr>
<td>Assignment 3</td>
<td>Chapter 6 to Chapter 8</td>
<td>02 September</td>
<td>22 September</td>
</tr>
<tr>
<td></td>
<td>Start writing your assignment</td>
<td>23 September</td>
<td>29 September</td>
</tr>
</tbody>
</table>

Note: For the text book, you will see the instructions in you workbook on which pages to read.

7 MODULE PRACTICAL WORK AND WORK-INTEGRATED LEARNING

There are no practicals for this module.

8 ASSESSMENT

8.1 Assessment plan

The assessment in this module consists of three assignments and an examination.

Your final mark for the module is determined from your semester mark and your examination mark.

The semester mark forms 20% and the examination mark 80% of the final mark. The semester mark is composed of 30% of assignment 1, 35% of assignment 2 and 35% of assignment 3 of the marks you receive. An assignment submitted late or not at all will give you 0%. If you do well in
your assignments you have a good semester mark and that can make all the difference between a pass or fail or between a distinction or simply a pass!

The three assignments prescribed for this module must be seen as part of the learning process. The typical assignment question is a reflection of a typical examination question. There are fixed submission dates for the assignments and each assignment is based on specific chapters in the study guide. You have to adhere to these dates as assignments are only marked if they are received on or before the due dates.

You will only get examination admission if you submit the first assignment by its due date. You should complete all assignments as well as you can, since

- they are the sole contributors towards your semester mark,
- they form an integral part of the learning process and indicate the form and nature of the questions you can expect in the examination.

**Assignments and Learning**

Assignments are seen as part of the learning material for this module. As you do the assignment, study the reading texts, consult other resources, discuss the work with fellow students or tutors or do research, you are actively engaged in learning. Looking at the assessment criteria given for each assignment, and the feedback you receive in your marked assignment, will help you to understand what is required of you more clearly.

### 8.2 General assignment numbers

The three assignments are numbered 01, 02 and 03 for each semester.

#### 8.2.1 Unique assignment numbers

Please note that each assignment has its unique six-digit assignment number which has to be written on the cover of your assignment upon submission. The unique numbers are given later on in this tutorial letter; you will find them in the heading of each set of assignment questions.

#### 8.2.2 Due dates for assignments

The closing dates for the submission of the assignments are:

<table>
<thead>
<tr>
<th>Assignment for SEMESTER 1</th>
<th>Sections from the following Chapters are covered</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chapters 1 and 2 of Study Guide and Workbook</td>
<td>17 March 2017</td>
</tr>
<tr>
<td>2</td>
<td>Chapters 3, 4, 5 and 6 of Study Guide and Workbook</td>
<td>31 March 2017</td>
</tr>
<tr>
<td>3</td>
<td>Chapters 6, 7, and 8 of Study Guide and Workbook</td>
<td>21 April 2017</td>
</tr>
<tr>
<td>Assignment for SEMESTER 2</td>
<td>Sections from the following Chapters are covered</td>
<td>Due Date</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>1</td>
<td>Chapters 1 and 2 of Study Guide and Workbook</td>
<td>18 August 2017</td>
</tr>
<tr>
<td>2</td>
<td>Chapters 3, 4, 5 and 6 of Study Guide and Workbook</td>
<td>01 September 2017</td>
</tr>
<tr>
<td>3</td>
<td>Chapters 6, 7, and 8 of Study Guide and Workbook</td>
<td>29 September 2017</td>
</tr>
</tbody>
</table>

### 8.3 Submission of assignments

For detailed information on assignments, please refer to the *my Studies @ Unisa* brochure, which you received with your study package.

To submit an assignment via myUnisa:

- Go to myUnisa.
- Log in with your student number and password.
- Select the module.
- Click on assignments in the menu on the left-hand side of the screen.
- Click on the assignment number you wish to submit.
- Follow the instructions.

For general information and requirements as far as assignments are concerned, see the brochure *my Studies @ Unisa* which you received with your study material.

### 8.4 Assignments

This tutorial letter 101 contains the assignments for both semesters, so select the semester you are enrolled for and do the set of assignments for that semester only. The assignments for Semester 1 are in Appendix A, pages 14–25. The assignments for Semester 2 are in Appendix B, pages 26–38. Solutions to the assignments will be posted to ALL students registered for this module a while after the closing date of the relevant assignment. Solutions will also be available on myUnisa.

### 9 OTHER ASSESSMENT METHODS

There are no other assessment methods for this module.
10 Examination

10.1 Examination Admission

You need to have a final mark of 50% to pass this module and 75% to obtain a distinction.

In this module a maximum of 20 marks is added to your examination mark (out of 80) to form your final mark. This 20% contribution comes from the marks you obtained for the three assignments and is called your semester mark. If you do well in your assignments you have a good semester mark and that can make all the difference between a pass or fail or between a distinction or simply a pass!

Currently admission to the examination is only based on the proof that you are actively involved in your studies. This proof is based on the submission of your first assignment before a fixed given date. Admission therefore does not rest with the department and if you do not submit that particular assignment in time, we can do nothing to give you admission. Although you are most probably a part time student with many other responsibilities, work circumstances will not be taken into consideration for exemption from assignments or the eventual admission to the examination.

No concession will be made to students who do not qualify for the examination.

10.2 Examination Period

This module is offered in a semester period of fifteen weeks. This means that

- if you are registered for the first semester, you will write the examination in May/June 2017 and should you fail and qualify for a supplementary examination, that supplementary examination will be written in October/November 2017.

- if you are registered for the second semester, you will write the examination in October/November 2016 and should you fail and qualify for a supplementary examination, that supplementary examination will be written in May/June 2018.

The examination section will provide you with information regarding the examination in general, examination venues, examination dates and examination times. Eventually, your results will also be processed by them and sent to you.

10.3 Examination Paper

Your examination will be a 2 hour examination. The questions will be similar to the assignment questions, but there will also be questions on theory. Should you have a final mark of less than 50%, it implies that you failed the module STA2601. However, should your results be within a specified percentage (usually from 40% to 49%), you will be given a second chance in the form of a supplementary examination on the dates as specified in 10.2. If you fail the examination with less than 40%, the semester mark will not count to help you pass.

10.4 Previous Examination Papers

Previous examination papers are available to students on myUnisa. In addition, you will receive a trial paper towards the end of the semester that you can use as an indication of typical
examination questions. Solutions to this trial paper is also sent out in a follow-up tutorial letter. Remember that the examples, exercises, activities in the guide as well as your assignment questions are also indicators of typical examination questions.

10.5 Tutorial Letter with Information on the Examination

As mentioned before, you will receive a tutorial letter containing a trial paper. Should the lecturer want to discuss any matter about the examination, it will be included in this tutorial letter. In the study guide you are given clear indications of the sections in the textbook that you have to know and can be tested on in the examination. Remember that you have to work continuously and do not treat statistics as any other subject, where it may be possible to study only selected sections of the work. All the topics are interlinked and you will definitely run into trouble if you skip sections!

You are automatically admitted to the exam on the submission of Assignment 01 by a specific date – see Section 8.1. Please note that lecturers are not responsible for exam admission, and ALL enquiries about exam admission should be directed by e-mail to exams@unisa.ac.za.

11 FREQUENTLY ASKED QUESTIONS

The my Studies @ Unisa brochure contains an A-Z guide of the most relevant study information. Please refer to this brochure for any other questions.

12 SOURCES CONSULTED

Several books were consulted in preparing this tutorial letter.

13 CONCLUSION

Remember that there are no "short cuts" to studying and understanding statistics. You need to be dedicated, work consistently and practice, practice and practice some more! We hope that you will enjoy studying this module and we wish you success in your studies.

Your lecturer
ADDENDUM A: FIRST SEMESTER ASSIGNMENTS

A.1 Assignment 01

ONLY FOR SEMESTER 1 STUDENTS
ASSIGNMENT 01
Unique Nr.: 752438
Fixed closing date: 17 March 2017

QUESTION 1

(a) The discrete random variable $X$ has the probability function:

$$P(X = x) = \begin{cases} 
  kx & x = 2, 4, 6 \\
  k(x-2) & x = 8 \\
  0 & \text{otherwise}
\end{cases}$$

where $k$ is a constant

(i) Show that $k = \frac{1}{18}$. \hfill (2)
(ii) Construct the probability distribution of $X$. \hfill (2)
(iii) Calculate the mean value of $X$. \hfill (3)
(iv) Calculate the variance of $X$. \hfill (4)
(v) Would you say that the distribution is symmetrical? \hfill (6)

(b) Let $X_1, X_2, X_3$ and $X_4$ be a random sample of size $n = 4$ drawn from a $n \, (\mu; \sigma^2)$ distribution.

Consider the following two estimators of the population mean $\mu$.

$$T_1 = \frac{X_1 + X_2 + X_3 + X_4}{4} \quad \text{(The sample mean)}$$

$$T_2 = \frac{X_1 + 2X_2 + 2X_3 + X_4}{6} \quad \text{(A weighted mean)}$$
(i) Show that both $T_1$ and $T_2$ are unbiased estimators of $\mu$. 

(ii) Which estimator do you prefer and why?

**QUESTION 2**

Suppose that $X_1; X_2; \ldots; X_{10}$ is a random sample from a $\mathcal{N}(65; 16)$ distribution.

We also define the following:

$$
V_1 = \sum_{i=1}^{5} \left[ \frac{(X_i - \mu)}{\sigma} \right]^2
$$

$$
V_2 = \sum_{i=7}^{10} \left[ \frac{(X_i - \mu)}{\sigma} \right]^2
$$

$$
W = \sum_{i=1}^{10} \left[ \frac{(X_i - \mu)}{\sigma} \right]^2
$$

(a) Find an expression for $f_{X_4X_{10}}(x_4; x_{10})$. 

(b) Find $P(X_1 > 73)$. 

(c) Find $P(61 < X_1 < 69)$. 

(d) Is $W \sim \chi^2_9$? 

(e) Is $Z = \frac{X_5 - 65}{4} \sim \mathcal{N}(0; 1)$? 

(f) Suppose that $\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i$ and $Y = \sum_{i=1}^{10} \left[ \frac{X_i - \bar{X}}{\sigma} \right]^2$.

Do you think that $f_{\bar{X},Y}(u; v) = f_{\bar{X}}(u) f_Y(v)$? 

(g) What is $\text{Var}(Y)$ where $Y = \sum_{i=1}^{10} \left[ \frac{X_i - \bar{X}}{\sigma} \right]^2$? 

(h) What is $E(W)$?
(i) What is the distribution of $V_1$?  

(j) What is the distribution of $V_2$?

QUESTION 3

Complete the following statements in your answer book (i.e. give the missing words and do not waste time rewriting everything):

(a) The statistic $T$ is called an unbiased estimator for the parameter $\theta$ if ....................

(b) Let $X_1; X_2; \ldots; X_n$ be a random sample from a population with unknown expected value $\mu$ and unknown variance $\sigma^2$. An unbiased estimator for the population variance is $\hat{\sigma}^2 =$ ....................

(c) The efficiency of two estimators of the same parameter is a function of their ....................

(d) When multiple measurements or observations are made on each of the individuals or units in a sample the assumption of ....................... is violated.

(e) If random variable $Y$ has a $n \sim (0; 1)$ distribution, $Y^2$ has a ......................... distribution with ......................... degree(s) of freedom.

(f) The significance level of a statistical test is defined as $\alpha =$ .........................

(g) The hypothesis “The blood platelet count of male cancer patients is lower than the blood platelet count of healthy men”, requires a .......................-sided test of significance.

(h) An appropriate null hypothesis for the above test (in g) could be ................................. .

(i) If the observed difference is “statistically highly significant” it means that the P-value will be very ................................. .
**QUESTION 4**

Let $X_1, X_2, ..., X_n$ be a random sample of size $n$ from a discrete distribution with probability function

$$P(X = r) = \frac{\lambda^r e^{-\lambda}}{(1 - e^{-\lambda}) r!} \quad \text{for } r = 1; 2; ...$$

[Please note:

In a real life situation a random sample will result in, for example, $X_1 = 2; X_2 = 3; X_3 = 3; X_4 = 10; X_5 = 2$ etc ... . Do not fall into the trap to argue that $X_1 = 1; X_2 = 2; ...; X_n = n$ because this is only one very specific outcome out of the millions of other possibilities. Denote the sample outcome by $X_1 = r_1; X_2 = r_2; ...; X_n = r_n$.]

(a) Find the likelihood function for the sample. (5)

(b) Show that

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = -n - \frac{ne^{-\lambda}}{1 - e^{-\lambda}} + \frac{\sum_{i=1}^{n} r_i}{\lambda}$$

(5)

(c) Find the maximum likelihood estimator for $\lambda$. [Hint: Your answer will be an implicit solution for $\lambda$. In other words, $\lambda$ is solved in terms of $\lambda$ and however hard you may try it is impossible to solve explicitly for $\lambda$. This means that you cannot “take $\lambda$ to the left” and equate it to a known expression.] Do not feel discouraged by “all the calculus” in this problem. One cannot be a good statistician without a sound knowledge of mathematics! (5)

[15]

**QUESTION 5**

(a) Write down, in general terms, the method of obtaining a least squares estimator. (4)

(b) Let $X_1, X_2, ..., X_n$ be independent random variables from a distribution with expected value $2\theta$. Find the least squares estimator for $\theta$. (6)

[10]

[100]
QUESTION 1
A psychology student conducted an experiment dealing with reaction times and recorded the following times to the nearest hundred of a second.

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.48</td>
<td>1</td>
</tr>
<tr>
<td>0.96</td>
<td>1</td>
</tr>
<tr>
<td>0.52</td>
<td>1</td>
</tr>
<tr>
<td>0.36</td>
<td>1</td>
</tr>
<tr>
<td>0.49</td>
<td>1</td>
</tr>
<tr>
<td>0.58</td>
<td>1</td>
</tr>
<tr>
<td>0.64</td>
<td>1</td>
</tr>
<tr>
<td>0.64</td>
<td>1</td>
</tr>
<tr>
<td>0.44</td>
<td>1</td>
</tr>
<tr>
<td>0.39</td>
<td>1</td>
</tr>
<tr>
<td>0.59</td>
<td>1</td>
</tr>
<tr>
<td>0.37</td>
<td>2</td>
</tr>
<tr>
<td>0.66</td>
<td>2</td>
</tr>
<tr>
<td>0.57</td>
<td>2</td>
</tr>
<tr>
<td>0.68</td>
<td>2</td>
</tr>
<tr>
<td>0.64</td>
<td>2</td>
</tr>
<tr>
<td>0.81</td>
<td>2</td>
</tr>
<tr>
<td>1.02</td>
<td>1</td>
</tr>
<tr>
<td>0.54</td>
<td>1</td>
</tr>
<tr>
<td>0.38</td>
<td>1</td>
</tr>
<tr>
<td>0.67</td>
<td>1</td>
</tr>
<tr>
<td>0.74</td>
<td>1</td>
</tr>
<tr>
<td>0.81</td>
<td>1</td>
</tr>
<tr>
<td>0.91</td>
<td>1</td>
</tr>
<tr>
<td>0.69</td>
<td>2</td>
</tr>
<tr>
<td>0.54</td>
<td>2</td>
</tr>
<tr>
<td>0.55</td>
<td>2</td>
</tr>
<tr>
<td>0.52</td>
<td>2</td>
</tr>
<tr>
<td>0.94</td>
<td>2</td>
</tr>
<tr>
<td>0.39</td>
<td>2</td>
</tr>
<tr>
<td>0.57</td>
<td>2</td>
</tr>
<tr>
<td>0.66</td>
<td>2</td>
</tr>
<tr>
<td>0.44</td>
<td>2</td>
</tr>
<tr>
<td>0.68</td>
<td>2</td>
</tr>
<tr>
<td>0.74</td>
<td>2</td>
</tr>
<tr>
<td>0.91</td>
<td>2</td>
</tr>
<tr>
<td>0.66</td>
<td>2</td>
</tr>
<tr>
<td>0.54</td>
<td>2</td>
</tr>
<tr>
<td>0.68</td>
<td>2</td>
</tr>
<tr>
<td>0.83</td>
<td>2</td>
</tr>
<tr>
<td>0.52</td>
<td>2</td>
</tr>
<tr>
<td>0.37</td>
<td>2</td>
</tr>
<tr>
<td>0.56</td>
<td>2</td>
</tr>
<tr>
<td>0.90</td>
<td>2</td>
</tr>
<tr>
<td>0.47</td>
<td>2</td>
</tr>
<tr>
<td>0.74</td>
<td>1</td>
</tr>
<tr>
<td>0.71</td>
<td>1</td>
</tr>
<tr>
<td>0.42</td>
<td>1</td>
</tr>
<tr>
<td>0.60</td>
<td>1</td>
</tr>
<tr>
<td>0.58</td>
<td>1</td>
</tr>
<tr>
<td>0.74</td>
<td>1</td>
</tr>
<tr>
<td>0.55</td>
<td>1</td>
</tr>
<tr>
<td>0.49</td>
<td>1</td>
</tr>
<tr>
<td>0.58</td>
<td>1</td>
</tr>
<tr>
<td>0.64</td>
<td>1</td>
</tr>
</tbody>
</table>

(a) Use Excel but do not include the output to complete this table. (9)

\[
\begin{align*}
N &= 100 \\
\sum X_i &= 36.6 \\
\overline{X} &= \ldots \\
\sum (X_i - \overline{X})^2 &= 1.4572 \\
\sum (X_i - \overline{X})^3 &= \ldots \\
\sum (X_i - \overline{X})^4 &= \ldots
\end{align*}
\]

(b) (i) Test whether this distribution is symmetric. (Use \( \alpha = 0.10 \).) (7)

(ii) Test whether this distribution has a kurtosis of a normal distribution. (Use \( \alpha = 0.10 \).) (7)

(iii) Does this distribution originate from a normal population? (1)

(c) The 60 observations were classified into six classes with equal probability for each class interval and test the null hypothesis that the observations come from a \( n (0.6; 0.16^2) \) distribution. Let \( \alpha = 0.10 \).

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>( O_i )</th>
<th>( \hat{e}_i = n \overline{X}_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X &lt; 0.445 )</td>
<td>10</td>
<td>\ldots</td>
</tr>
<tr>
<td>( 0.455 \leq X &lt; 0.531 )</td>
<td>8</td>
<td>\ldots</td>
</tr>
<tr>
<td>( 0.531 \leq X &lt; 0.600 )</td>
<td>14</td>
<td>\ldots</td>
</tr>
<tr>
<td>( 0.600 \leq X &lt; 0.669 )</td>
<td>7</td>
<td>\ldots</td>
</tr>
<tr>
<td>( 0.669 \leq X &lt; 0.755 )</td>
<td>12</td>
<td>\ldots</td>
</tr>
<tr>
<td>( X \geq 0.755 )</td>
<td>9</td>
<td>\ldots</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>60</strong></td>
<td><strong>60</strong></td>
</tr>
</tbody>
</table>
(i) Show how intervals 1 and 2 are derived. (10)

(ii) Calculate the missing expected frequencies in the above table. (1)

(iii) At the 0.05 level, use a chi-square goodness-of-fit test to test if the 60 observations in the sample come from a normal distribution with mean 0.6 and standard deviation 0.16. (10)

QUESTION 2

In an apartment complex, a sociologist asks a sample of workers if they are satisfied with their work. He then classifies type of work into three categories. Using the following results, test the hypothesis that the job satisfaction is not related to type of work.

<table>
<thead>
<tr>
<th>Job satisfaction</th>
<th>Type of Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>White collar</td>
<td>Blue collar</td>
</tr>
<tr>
<td>Satisfied</td>
<td>81</td>
</tr>
<tr>
<td>Dissatisfied</td>
<td>49</td>
</tr>
</tbody>
</table>

Use SAS JMP or R to determine whether job satisfaction is not related to type of work at $\alpha = 0.05$.

(a) Produce the Mosaic Plot and interpret it. (11)

(b) State the appropriate null and alternative hypothesis for this test. (2)

(c) What test statistic is used to test these hypotheses and what is the value of the test statistic? (2)

(d) Looking at the row percentages in your output, can you draw any conclusions? (3)

(e) What is your final conclusion? (2)
QUESTION 3
A random sample of twelve psychologists and psychiatrists were questioned on their opinions regarding “behavioural modification” as a useful technique. The following opinions were found:

<table>
<thead>
<tr>
<th>Is behavioural modification useful?</th>
<th>Yes</th>
<th>No</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychologists</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Psychiatrists</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Column total</td>
<td>6</td>
<td>6</td>
<td>12</td>
</tr>
</tbody>
</table>

Test the hypothesis that psychiatrists feel less favourable towards the usefulness of behavioural modification. Use $\alpha = 0.05$.

[10]

QUESTION 4

(a) Let $X_1; X_2; \ldots; X_n$ be a random sample from a normal distribution with known mean $\mu$. Use the distribution of $U = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$ to show that a $100(1 - \alpha)$% two-sided confidence interval for $\sigma^2$ is given by

$$\left[ \frac{\sum (X_i - \mu)^2}{\chi^2_{\frac{\alpha}{2}; n}} ; \frac{\sum (X_i - \mu)^2}{\chi^2_{1 - \frac{\alpha}{2}; n}} \right]$$

(7)

(b) To study telephone calls at a business office, the office manager times incoming calls and outgoing calls for 1 day. She finds that 17 incoming calls last an average of 5.16 minutes with a standard deviation of 1.12 minutes and 12 outgoing calls last an average of 4.13 minutes with a standard deviation of 2.36 minutes.

(i) Test the hypothesis that the variances of the populations are equal at the 5% level of significance.

(ii) Determine a 95% confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$ and interpret the results.

(Clearly show how you compute the critical values.)

[25]

[100]
A.3 Assignment 03

QUESTION 1
A random sample of soil specimens was obtained, and the amount of organic matter (%) in the soil was determined for each specimen, resulting in the accompanying data (from "Engineering Properties of Soil," Soil Science, 1998; 93 - 102).

\[
\begin{array}{cccccccccccccccc}
1.10 & 5.09 & 0.97 & 1.59 & 4.60 & 0.32 & 0.55 & 1.45 & 0.14 & 4.47 \\
1.20 & 3.50 & 5.02 & 4.67 & 5.20 & 2.69 & 3.98 & 3.15 & 3.03 & 2.21 \\
0.69 & 4.47 & 3.31 & 1.17 & 0.76 & 1.17 & 1.57 & 2.62 & 1.66 & 2.05 \\
\end{array}
\]

\[ \sum X_i = 74.4 \quad \sum X_i^2 = 260.0742 \]

(a) Produce a SAS JMP or R output to answer the following questions.

(i) What assumption(s) is/are necessary in order to conduct the statistical test specified in (b) below? Are they met? Give a brief discussion (8).

(ii) Do this data suggest that the true average percentage of organic matter in such soil is something other than 3%? Use a 10% level of significance and test two-sided. (10).

(iii) Would your conclusion be different if \( \alpha = 0.05 \) had be used. Justify? (2)

(iv) Is there any reason to reject the null hypothesis \( H_0: \sigma = 1.5 \)? Use a 10% level of significance and test two-sided. (10)

(b) Suppose that another sample of 20 specimens is taken and the following statistics are obtained:

\[ \bar{Y} = 2.43 \quad S_Y^2 = 2.4749 \]

Would you say that the mean percentage of organic matter of population B is less than the mean percentage of organic matter of population A? Use \( \alpha = 0.05 \).

[Hint: assume that both variances are unknown but equal. Assume that \( S_X^2 = 2.6056 \) for the sample with \( n_1 = 30 \) from population A.] (10)
QUESTION 2

A pharmaceutical manufacturer has been researching new medications formulas to provide quicker relief of arthritis pain. Their laboratories have produced three different medications and they want to determine if the different medications produce different responses. Fifteen people who complained of arthritis pains were recruited for an experiment; five were randomly assigned to each medication. Each person was asked to take the medicine and report the length of time until some relief was felt (minutes). The results are shown below.

<table>
<thead>
<tr>
<th>Time in minutes until relief is felt</th>
<th>Medication 1</th>
<th>Medication 2</th>
<th>Medication 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

DO NOT USE SAS JMP or R. DO THIS MANUALLY:
(Regard the data as random samples from normal populations.)

(a) What are the values of $S_1^2$, $S_2^2$, and $S_3^2$.

(b) (i) Compute the “ordinary” average of the three variances computed in (a).

(ii) Compute the MSE according to the definition in the study guide. What do you notice?

(c) Do you think it is reasonable to assume that the assumption of independence is met?

(d) Test at the 5% level of significance whether the population means of the four different detergents differ.

(i) State the null and alternative hypotheses.

(ii) State the rejection region and conclusion.

(e) Perform multiple comparisons on all pairs of means. Discuss your results.
QUESTION 3

A behaviorist has performed the following experiment. For each of 10 sets of identical twins who were born 30 years ago, he recorded their annual incomes, according to which twin was born first. The results (in R10 000s) are shown below:

<table>
<thead>
<tr>
<th>Twin set</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>First born</td>
<td>32</td>
<td>36</td>
<td>21</td>
<td>30</td>
<td>49</td>
<td>27</td>
<td>39</td>
<td>38</td>
<td>56</td>
<td>44</td>
</tr>
<tr>
<td>Second born</td>
<td>44</td>
<td>43</td>
<td>28</td>
<td>39</td>
<td>51</td>
<td>25</td>
<td>32</td>
<td>42</td>
<td>64</td>
<td>44</td>
</tr>
</tbody>
</table>

(a) Can he infer at 5% significance level that there is a difference in income between the twins? Clearly state the hypothesis implied by the question and how it can be tested. Give the rejection region and the conclusions. (13)

(b) Produce a SAS JMP or R output to support your calculations. (5)

(c) What type of data is this? (2)

QUESTION 4

Five companies submit samples of paint to a firm that is considering the purchase of a large quantity. Six samples of paint are tested to find the paint with the shortest drying time. The drying times are:

<table>
<thead>
<tr>
<th>Company</th>
<th>Drying times (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>34 36 29 38 35 32</td>
</tr>
<tr>
<td>II</td>
<td>30 34 32 31 28 30</td>
</tr>
<tr>
<td>III</td>
<td>27 32 31 30 34 30</td>
</tr>
<tr>
<td>IV</td>
<td>28 35 29 37 37 33</td>
</tr>
<tr>
<td>V</td>
<td>34 31 36 38 40 37</td>
</tr>
</tbody>
</table>

Use SAS JMP or R to answer the following questions.

(a) Use Levene’s test to determine if the five groups have equal population variances? Use $\alpha = 0.05$ level of significance. (State your hypothesis and justify your answer.) (10)

(b) Are there significant differences in the mean drying times for the paints at the 5% level of significance?

Justify your answer by giving attention to the following detail:

(i) State the appropriate null and alternative hypothesis for this test.
(ii) What test statistic is used to test these hypotheses?

(iii) What is the value of the test statistic?

(c) Perform all pairwise comparison using the Tukey-Kramer HSD method of multiple comparisons at the 5% level of significance. [Hint: See last year’s tutorial 101 on myunisa under announcement for the type of outputs you should generate.]

(d) The firm wants to buy the paint with the shortest drying time. Can they determine, from this analysis, which paint to buy, or should they select paints for further testing? If so, which ones?

[40]

QUESTION 5

The controller of a large department store chain would like to predict the account balance (in millions) at the end of a billing period based upon the number of transactions made during the billing period. A random sample of 12 accounts was selected with the results given below:

<table>
<thead>
<tr>
<th>Account</th>
<th>Number of transactions</th>
<th>Account balance (in millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>36</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>175</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>69</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>198</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>120</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>84</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>150</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>78</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>100</td>
</tr>
</tbody>
</table>

(a) Determine the independent and dependent variables.  

(b) Plot a scatter diagram and comment on it.
(c) Use the least squares method to find the regression equation. (6)

(d) Give an interpretation of the numerical value of the regression coefficient $\hat{\beta}_1$. (2)

(e) Test for the significance of the slope, $\beta_1$ at the 5% level of significance.

**Justify your answer by giving attention to the following detail**

(i) State the appropriate null and alternative hypothesis for this test.

(ii) What test statistic is used to test these hypotheses?

(iii) What is the value of the test statistic? (10)

(f) Predict the account balance for an account which has had five transactions in the last billing period. (2)

(g) What is the standard error of your estimate in question (f)? (5)

(h) Derive a 95% confidence interval for the slope $\beta_1$. (5)

(i) Calculate the correlation coefficient, $r$ and comment on it. (5)

(j) Test $H_0 : \rho = 0.8$ (two-sided) at the 5% level of significance. (10)

(k) Calculate $R^2$ and interpret it. (2)

(l) Submit a SAS JMP or R output to support your calculations you have done manually. (10)
ADDENDUM B: SECOND SEMESTER ASSIGNMENTS

B.1 Assignment 01

ONLY FOR SEMESTER 2 STUDENTS
ASSIGNMENT 01
Unique Nr.: 666015
Fixed closing date: 18 August 2017

QUESTION 1

(a) The discrete random variable \( X \) can take only the values 1, 2 and 3. For these values the cumulative distribution function is defined by:

\[
F(x) = \frac{x^3 + k}{40} \quad x = 1, 2, 3
\]

where \( k \) is a constant

(i) Show that \( k = 13 \). (3)

(ii) Find the probability distribution of \( X \). (4)

(iii) Calculate the mean value of \( X \). (3)

(iv) Calculate the variance of \( X \). (4)

(v) Would you say that the distribution is symmetrical? (6)

(b) Let \( X_1 \) and \( X_2 \) be independent normal variables with mean \( \mu \) and variance \( \sigma^2 \). Suppose that

\[
T_1 = \frac{4}{7}X_1 + \frac{3}{7}X_2; \\
T_2 = \frac{2}{3}X_1 + \frac{3}{3}X_2 \quad \text{and} \\
T_3 = \frac{1}{4}X_1 + \frac{3}{4}X_2
\]

(i) Show that they are all unbiased estimators of \( \mu \). (7)

(ii) Which estimator is the most efficient? (8)
QUESTION 2

Comment on the following statements.

(a) The variance of any random variable can never be a negative value.

(b) The best method of estimating an unknown parameter is the method of least squares.

(c) A type I error is made if $H_0$ is not rejected when $H_1$ is true.

(d) One would expect a negative correlation between the length (in cm) and the mass (in kg) of a child.

(e) $r = 0$ indicates a lack of relationship between $X$ and $Y$.

(f) During the 1950s there was a high correlation between the consumption of soft drinks and the number of polio cases. This indicates that the consumption of soft drinks caused polio.

(g) The only important assumption underlying one-way analysis of variance is that the samples must be independent.

(h) The statistic $T$ is called an unbiased estimator for the parameter $\theta$ if $E(T) = \theta$.

(i) If you reject the null hypothesis $\sigma_1^2 = \sigma_2^2$ on the basis of sample data, when in fact no difference exists between the variances of two populations, you have made a type II error.

(j) If the sample size $n$ increases, then the power of the test decreases.

[20]
QUESTION 3

Suppose that \( X_1, X_2, \ldots, X_9 \) is a random sample from a \( n(15; 9) \) distribution and that

\[
\bar{X} = \frac{1}{9} \sum_{i=1}^{9} X_i \quad \text{and} \quad Y = \sum_{i=1}^{9} \left[ \frac{X_i - \bar{X}}{\sigma} \right]^2.
\]

Suppose that we also define

\[
V_1 = \sum_{i=2}^{6} \left[ \frac{(X_i - 15)}{3} \right]^2
\]

\[
V_2 = \sum_{i=3}^{9} \left[ \frac{(X_i - 15)}{3} \right]^2
\]

\[
V_3 = \sum_{i=1}^{7} \frac{[X_i - 15]^2}{9}
\]

(a) Is \( f_{X_1}(x_1) = f_{X_3}(x_3) \)?

(b) Find \( P(11 < X_2) \) or \( P(X_2 > 19) \).

(c) What is the distribution of \( \frac{(X_3 - 15)^2}{3} \)?

(d) What is \( E(V_3) \)?

(e) Let \( Y = \sum_{i=1}^{9} \left[ \frac{X_i - \bar{X}}{\sigma} \right]^2 \). Determine \( Var(Y) \).

(f) What is the distribution of \( U = \frac{V_1/5}{V_2/7} \) and hence \( \frac{1}{U} \)?

(g) Find a value \( a \) such that \( P(U < a) = 0.95 \).
QUESTION 4

(a) Let $X_1, X_2, \ldots, X_n$ be independent random variables such that

$$E(X_i) = \theta_1 + C_i \theta_2 \quad \text{for} \quad i = 1, 2, \ldots, n;$$

$$\text{Var}(X_i) = \sigma^2 \quad \forall \ i$$

with $\sigma^2$, $\theta_1$ and $\theta_2$ unknown parameters and $C_1, C_2, \ldots, C_n$ known constants. Find the least squares estimators of $\theta_1$ and $\theta_2$.

(b) Let $X_1; \ldots; X_n$ be independent random variables from a Bernoulli distribution with probability density function.

$$f_X(x) = \pi^x (1 - \pi)^{1-x} \quad \text{for} \quad x = 0 \text{ or } x = 1.$$ 

(In other words the outcome of each random experiment is either a success ($X = 1$) or a failure ($X = 0$) with $\pi = \text{probability of a success}$).

Find the maximum likelihood estimator of $\pi$. 

(15) 

(10) 

[25] 

[100]
QUESTION 1
A student of ancient Greek texts is doing a study on the authors of the epistles (letters) in the New Testament. (The New Testament was originally written in Greek.) The distribution of the sentence lengths in an epistle can be an indication of authorship. The student counted the number of words in each sentence in the epistle to the Romans and constructed the following frequency table:

<table>
<thead>
<tr>
<th>Number of words per sentence</th>
<th>Frequency (number of sentences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>67</td>
</tr>
<tr>
<td>6-10</td>
<td>144</td>
</tr>
<tr>
<td>11-15</td>
<td>87</td>
</tr>
<tr>
<td>16-20</td>
<td>42</td>
</tr>
<tr>
<td>21-25</td>
<td>43</td>
</tr>
<tr>
<td>26-30</td>
<td>14</td>
</tr>
<tr>
<td>31-35</td>
<td>12</td>
</tr>
<tr>
<td>36-40</td>
<td>6</td>
</tr>
<tr>
<td>41-45</td>
<td>7</td>
</tr>
<tr>
<td>46-50</td>
<td>9</td>
</tr>
<tr>
<td>&gt;50</td>
<td>5</td>
</tr>
</tbody>
</table>

She wants to test the hypothesis that the observed distribution originates from a normal distribution.

(a) The raw data $X_i$, for the above problem consists of 436 values, i.e. there are 436 sentences in this epistle and $X_i$ is the number of words in the i-th sentence. The following were calculated for your convenience:

\[
\begin{align*}
N &= 436 \\
\sum X_i &= 6238 \\
\bar{X} &= 14.31 \\
\frac{1}{n} \sum (X_i - \bar{X})^2 &= 124.942 \\
\frac{1}{n} \sum (X_i - \bar{X})^3 &= 2648.266 \\
\frac{1}{n} \sum (X_i - \bar{X})^4 &= 119812.018
\end{align*}
\]

(i) Test whether this distribution is symmetric. (Use $\alpha = 0.10$.)

(ii) Test whether this distribution is leptokurtic. (Use $\alpha = 0.05$.)
(iii) Does this distribution originate from a normal population? (1)

(b) There is a theory that sentence length distributions follow a so-called Sichel distribution with probability density function

\[ \phi(r) = \frac{\sqrt{1 - \theta}}{K_{\gamma} \sqrt{\theta}} \frac{(a_{\theta})^r}{r!} K_{r+\gamma}(a) \]

where \( \alpha, \theta, \) and \( \gamma \) are known constants and \( r = 1, 2, 3, \ldots \) the number of words per sentence.

Using the above (rather horrible!) probability density function, the expected frequencies for the same class intervals given in the problem were calculated and found to be:

<table>
<thead>
<tr>
<th>Number of words per sentence</th>
<th>Expected frequency (number of sentences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>78</td>
</tr>
<tr>
<td>6-10</td>
<td>132</td>
</tr>
<tr>
<td>11-15</td>
<td>90</td>
</tr>
<tr>
<td>16-20</td>
<td>50</td>
</tr>
<tr>
<td>21-25</td>
<td>34</td>
</tr>
<tr>
<td>26-30</td>
<td>12</td>
</tr>
<tr>
<td>31-35</td>
<td>13</td>
</tr>
<tr>
<td>36-40</td>
<td>10</td>
</tr>
<tr>
<td>41-45</td>
<td>5</td>
</tr>
<tr>
<td>46-50</td>
<td>6</td>
</tr>
<tr>
<td>&gt;50</td>
<td>6</td>
</tr>
</tbody>
</table>

Perform a goodness-of-fit test to see whether the sentence length distribution of the epistle to the Romans follows a Sichel distribution. Use \( \alpha = 0.05 \). (15)
QUESTION 2
Breeding experiments were carried out on Rhode Island Red fowl in a study of colour mutation. The following data show the number of chicks with white down (i.e. the first covering of a young bird) which resulted from three different types of parents.

<table>
<thead>
<tr>
<th>Type of parent</th>
<th>Colour of the down</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of chicks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>with coloured down</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of chicks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>with white down</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>210</td>
<td>50</td>
</tr>
<tr>
<td>B</td>
<td>146</td>
<td>54</td>
</tr>
<tr>
<td>C</td>
<td>34</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>390</td>
<td>110</td>
</tr>
</tbody>
</table>

Use SAS JMP or R to determine whether an association exists between the type of parent and the colour of the down. (Use $\alpha = 0.05$.)

(a) Produce the Mosaic Plot and interpret it. (11)

(b) State the appropriate null and alternative hypothesis for this test. (2)

(c) What test statistic is used to test these hypotheses and what is the value of the test statistic? (2)

(d) Looking at the row percentages in your output, can you draw any conclusions? (3)

(e) What is your final conclusion? (2)

QUESTION 3

(a) Twelve postgraduate students were classified according to gender and their final examination result. The results are given below.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Pass</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Do an exact 2x2 test at the 5% level of significance to see whether females performed better than males in this examination. (10)
(b) A sample of 35 pairs of observations is taken and a sample correlation coefficient is computed to be $R = 0.48$. Accept that the sample comes from a bivariate normal distribution with population correlation coefficient $\rho$.

(i) Do the data provide sufficient evidence to reject the null hypothesis that $\rho = 0.3$? Use a 1% significance level.  

(ii) Construct a 95% (two-sided) confidence interval for $\rho$.  

(c) Two independent random samples from two bivariate normal distributions yielded the following correlation coefficients:

Sample 1: $r_1 = 0.5; \ n_1 = 103$

Sample 2: $r_2 = 0.8; \ n_2 = 53$.

Is the correlation coefficient for population 1 significantly smaller than that for population 2?

QUESTION 4

A random sample of nine packets of potato chips (all marked 125 grams) were weighed and the following weights (in grams) were obtained:

125 ; 127 ; 122 ; 126 ; 121 ; 128 ; 125 ; 126 ; 125.

(a) Derive a 95% confidence interval for the unknown variance $\sigma^2$ if you assume that $\mu$ is unknown.  

(b) Derive a 95% confidence interval for the unknown variance $\sigma^2$ if you assume that $\mu = 125$. 

[35]
QUESTION 1
A manufacturer of rubber gloves (for household purposes) selected a random sample of 50 females and measured the palm width of the right hand of each female.
The table below shows the 50 observations.

<table>
<thead>
<tr>
<th>7.6</th>
<th>8.5</th>
<th>7.1</th>
<th>7.7</th>
<th>7.8</th>
<th>7.6</th>
<th>8.9</th>
<th>7.5</th>
<th>8</th>
<th>9.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.3</td>
<td>9.2</td>
<td>8</td>
<td>8.8</td>
<td>7.6</td>
<td>8.2</td>
<td>9.1</td>
<td>6.3</td>
<td>7.3</td>
<td>6.9</td>
</tr>
<tr>
<td>8.4</td>
<td>9</td>
<td>7</td>
<td>6.3</td>
<td>9</td>
<td>8.9</td>
<td>9.1</td>
<td>7.4</td>
<td>7.6</td>
<td>7.8</td>
</tr>
<tr>
<td>9.7</td>
<td>8.2</td>
<td>7.3</td>
<td>7.2</td>
<td>7.6</td>
<td>7.9</td>
<td>8.6</td>
<td>9.2</td>
<td>8.2</td>
<td>8</td>
</tr>
<tr>
<td>7.6</td>
<td>7.4</td>
<td>7.5</td>
<td>7.3</td>
<td>8</td>
<td>6.4</td>
<td>7.4</td>
<td>8.5</td>
<td>9.6</td>
<td>6.6</td>
</tr>
</tbody>
</table>

(a) Produce a SAS JMP or R output to answer the following questions.

(i) What can you conclude about normality from your SAS JMP output. (7).

(ii) Is there any reason to reject the null hypothesis $H_0: \mu = 8.5$? Test two-sided at the 5% level of significance. Show how you use and interpret the SAS JMP. (10).

(iii) Using the SAS JMP or R output, give a 95% confidence interval for the unknown mean $\mu$. Does this interval confirm your conclusion in question (ii)? (7)

(iv) Is there any reason to reject the null hypothesis $H_0: \sigma = 1.0$? Use a 5% level of significance and test two-sided. Clearly indicate what assumptions you make, and how you apply them. (10)

(b) Individuals from two independent random samples of 20 men and 25 women respectively were asked to do a social skills test. Assume the following statistics:

<table>
<thead>
<tr>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_1 = 57.4$</td>
<td>$\bar{X}_2 = 63.4$</td>
</tr>
<tr>
<td>$S_1 = 8.124$</td>
<td>$S_2 = 7.874$</td>
</tr>
</tbody>
</table>

(i) Test at the 5% level of significance the necessary assumptions. (11)
(ii) Does this confirm at the 5% level of significance that women are on average socially more skillful than men? Clearly state the hypothesis implied by the question and how it can be tested. Also give the rejection region and the conclusions. (10)

QUESTION 2

Four independent random samples of size \( n = 5 \) from assumed \( n \left( \mu_j; \sigma^2_j \right) \) distributions yield the following:

Sample 1 16 16 15 14 14
Sample 2 20 17 17 16 15
Sample 3 20 20 19 18 18
Sample 4 22 22 21 21 19

You may make use of the following computations:

\[
\begin{align*}
\sum y_{1i} &= 75 & \sum (y_{1i})^2 &= 1129 \\
\sum y_{2i} &= 85 & \sum (y_{2i})^2 &= 1459 \\
\sum y_{3i} &= 95 & \sum (y_{3i})^2 &= 1809 \\
\sum y_{4i} &= 105 & \sum (y_{4i})^2 &= 2211 \\
\end{align*}
\]

DO NOT USE SAS JMP or R. DO THIS MANUALLY:

(a) Test at the 5% level of significance whether the population variances differ significantly from one another. (15)

(b) Construct an ANOVA table and test at the 5% level whether the means differ significantly. What assumptions do you make? (20)

(c) Can you conclude that \( \mu_1 = \mu_2 \neq \mu_4 \)? (10)
QUESTION 3

The general manager of a chain of fast food chicken restaurants wants to determine how effective their promotional campaigns are. In these campaigns "20% off" coupons are widely distributed. These coupons are only valid for one week. To examine their effectiveness, the executive records the daily gross sales (in R10 000s) in one restaurant during the campaign and during the week after the campaign ends. The data is shown below.

<table>
<thead>
<tr>
<th>Day</th>
<th>Sales during campaign</th>
<th>Sales after campaign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunday</td>
<td>18.1</td>
<td>16.6</td>
</tr>
<tr>
<td>Monday</td>
<td>10.0</td>
<td>8.8</td>
</tr>
<tr>
<td>Tuesday</td>
<td>9.1</td>
<td>8.6</td>
</tr>
<tr>
<td>Wednesday</td>
<td>8.4</td>
<td>8.3</td>
</tr>
<tr>
<td>Thursday</td>
<td>10.8</td>
<td>10.1</td>
</tr>
<tr>
<td>Friday</td>
<td>13.1</td>
<td>12.3</td>
</tr>
<tr>
<td>Saturday</td>
<td>20.8</td>
<td>18.9</td>
</tr>
</tbody>
</table>

(a) Can they infer at the 5% significance level that sales increase during the campaign? Clearly state the hypothesis implied by the question and how it can be tested. Give the rejection region and the conclusions.

(b) Estimate with 95% confidence the mean difference and interpret.

(c) Produce a SAS JMP or R output to support your calculations.
 QUESTION 4

Five companies submit samples of paint to a firm that is considering the purchase of a large quantity. Six samples of paint are tested to find the paint with the shortest drying time. The drying times are:

<table>
<thead>
<tr>
<th>Amount of drug injected (CC)</th>
<th>Stress level at which adrenalin is released</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>14 21 16 18 23 15 19 22 26 27</td>
</tr>
<tr>
<td>10</td>
<td>13 17 19 18 21 16 25 18 22 23</td>
</tr>
<tr>
<td>20</td>
<td>16 22 18 23 22 25 19 17 21 25</td>
</tr>
<tr>
<td>30</td>
<td>21 19 24 22 28 23 22 28 24 20</td>
</tr>
<tr>
<td>40</td>
<td>26 22 23 25 27 29 22 24 27 26</td>
</tr>
</tbody>
</table>

Use SAS JMP or R to answer the following questions.

(a) There is considerable variation among the variances within each sample. Using Bartlett’s test, can you conclude using \( \alpha = 0.05 \) level of significance, that the variances are homogeneous? (State your hypothesis and justify your answer.) (10)

(b) Assuming equal variances, determine whether there are significant differences among the group at the 5% level of significance?

Justify your answer by giving attention to the following detail:

(i) State the appropriate null and alternative hypothesis for this test.

(ii) What test statistic is used to test these hypotheses?

(iii) What is the value of the test statistic?

(10)

(c) Perform all pairwise comparison using Each Pair, Student’s t method of multiple comparisons at the 5% level of significance. [Hint: See last year’s tutorial 101 on myunisa under announcement for the type of outputs you should generate. (14)

(d) Which method is better between Each Pair, Student’s t and All Pairs, Tukey HSD and why? (2)

(e) Obtain the means for each level of drug and plot them against cubic centimeters of drug injected. Do you notice an upward trend? (4)
QUESTION 5

The data in the table below represent the height ($X$) (in cm) and the mass ($Y$) (in kg) of several adult men. The heights were selected in advance and then the masses of a random group of men having the selected heights were obtained.

<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>Mass (kg)</th>
<th>$(X_i - \overline{X})$</th>
<th>$Y_i (X_i - \overline{X})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>160</td>
<td>62</td>
<td>-10</td>
<td>-620</td>
</tr>
<tr>
<td>160</td>
<td>68</td>
<td>-10</td>
<td>-680</td>
</tr>
<tr>
<td>160</td>
<td>68</td>
<td>-10</td>
<td>-680</td>
</tr>
<tr>
<td>165</td>
<td>70</td>
<td>-5</td>
<td>-350</td>
</tr>
<tr>
<td>165</td>
<td>68</td>
<td>-5</td>
<td>-340</td>
</tr>
<tr>
<td>165</td>
<td>74</td>
<td>-5</td>
<td>-370</td>
</tr>
<tr>
<td>170</td>
<td>70</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>170</td>
<td>82</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>170</td>
<td>78</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>175</td>
<td>77</td>
<td>5</td>
<td>385</td>
</tr>
<tr>
<td>175</td>
<td>83</td>
<td>5</td>
<td>415</td>
</tr>
<tr>
<td>175</td>
<td>82</td>
<td>5</td>
<td>410</td>
</tr>
<tr>
<td>180</td>
<td>84</td>
<td>10</td>
<td>840</td>
</tr>
<tr>
<td>180</td>
<td>86</td>
<td>10</td>
<td>860</td>
</tr>
<tr>
<td>180</td>
<td>88</td>
<td>10</td>
<td>880</td>
</tr>
</tbody>
</table>

| Total      | 2550      | 1140                   | 0                         |
| Total      | 0         | 750                    |

Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ where the $\epsilon_i$’s are independent $n(0; \sigma^2)$ variables.

(a) Compute unbiased estimates of $\beta_0$, $\beta_1$ and $\sigma^2$. (13)

(b) Find a 90% confidence interval for the slope ($\beta_1$) of the regression line computed in (a). (5)

(c) What is the expected mass of a man who is 178 cm tall? (1)

(d) Find a 90% confidence interval for the expected mass of a man who is 178 cm tall. [Hint: 178 is not an observed value!] (4)

(e) Would it be valid to use the estimated regression line in order to predict the mass of a boy who is 120 cm tall? Comment. (2)

(l) Submit a SAS JMP or R output to support your calculations you have done manually (10)