

Course Information

This is a sample study guide used in a previous semester.
It is not current and is only an example. 2004 1st Semester.

1. Background Assumed

It is assumed that students coming into the STATS 20x course have passed one of the STATS 10x courses at this university. In particular it is assumed that students have a reasonable knowledge about Hypotheses tests, *P-values*, Confidence Intervals, Tables of Counts, One-way Analysis of Variance and Linear Regression. **If you have entered STATS 20x without studying STATS 10x** at this university (e.g. *direct entry* from school) you will not have covered some of these topics. If this is the case, or that you feel your background in these topics is weak, we suggest that you read Chapters 3 and 7 to 12 of *Chance Encounters: A First Course in Data Analysis and Inference* by C.J. Wild and G.A.F. Seber (particularly Chapters 8, 9 and 10). Direct-entry students should also make themselves known to the Course Coordinator within the first few days.

It will be assumed that students coming into STATS 20x have no prior experience using the statistical computer package **R**. However, it will be assumed that students coming into STATS 20x have already been exposed to computers.

2. Aims of the Course

The process of statistical inquiry begins with questions about the real world. Data are collected and then analysed to see what it has to say about the answers to those original questions. This results in conclusions and ideas for further inquiry. STATS 20x concentrates largely (but not solely) on the analysis phase of this process.

We build on the ideas introduced in STATS 10x to extend the range of types of data, which you will be able to analyse. From the STATS 10x course, Chapters 2–3 (tools for exploring data) and Chapters 10–12 (analysing data on continuous variables, Tables of Counts and Regression) are particularly relevant to Stage 2. Confidence intervals (Chapter 8) and hypothesis tests and *P-values* (Chapter 9) are essential tools, which we will use frequently to extrapolate from sample data to what is going on in a wider population.

One example of the ways we extend the range of data, which you can analyse is provided by Regression. In STATS 10x, you learned to use the information in a single explanatory variable to explain or predict the behaviour of a response variable of interest. Clearly, the more information one has, the better one can do. This year, you will learn how to use the information in a whole collection of explanatory variables to explain or predict the behaviour of the response (Multiple Regression). You will also learn how to fit nonlinear trend curves.

All serious data analysis is now done by computer so this is very much a computer based course. The package we use is called **R**. Nevertheless, this is a course in data analysis, not a course in **R**. We teach you how to look at data, the types of analyses and summaries you should be looking at and how to interpret them, and we teach this in such a way that you can transfer your knowledge to working with other computer packages. **R** is principally the tool that you will use to do your assignment work.



Course Information

If you apply yourself to this course, you should take away with you:

- an awareness of a wider array of types of data as given in the “ Topics” section below;
- the ability to analyse reasonably straightforward examples of each type of data and interpret the results in the terms of the original question;
- the ability to recognise when examples are not straightforward and when expert assistance may be required;
- facility in practical data analysis with a computer package (**R**) which you can take away with you after the course, free of charge, for job or private use.

3. STATS 20x Web Survey – Assignment 1



STATS 20x is about collecting and analysing data. We wish to create a large data set relevant to you. We need your help to do this.

Log onto <http://www.stat.auckland.ac.nz/~u47520x/> to complete the STATS 20x web survey.

Completion of this web survey forms part of assignment 1 and the survey is to be **completed by 11:59pm, Sunday 7th March.**

You will have to enter your ID number to obtain **credit towards ASSIGNMENT 1.**

A screenshot of a web browser displaying the "STATS20x Web Survey" form. The browser window title is "STATS20x Web Survey - Microsoft Internet Explorer". The address bar shows "http://www.stat.auckland.ac.nz/~u47520x/". The form has a blue header with the title "STATS20x Web Survey". Below the header is a light blue box containing "Instructions":

- Please answer **all** questions
- Completion and submission of this survey by the end of Friday 7 March will gain you credit for Assignment 4
- Your ID is needed so you can be awarded the marks for Assignment 4
- Your ID will **not** be stored with your responses
- The information collected from this survey will be used **only** for data analysis for examples and exercises in this course

Below the instructions is a white form area with the following fields:

- Student ID:
- Course:
- Sex: Male Female
- Date of birth:
- 1. My eyes are
- 2. I have body piercings. (Count each individual piercing.)
- 3. I have tattoos.
- 4. I wear glasses (spectacles)/contact lenses: Yes No
- 5. On average, I sleep for hours per night.
- 6. On average, I study for hours per week.
- 7. On average, I watch television for hours per week.

A black arrow points from the "Student ID" field in the form to the "Instructions" box.

Course Information

4. Topics Taught

Introduction

Introduction to the statistical computer package **R**.

A review of the concepts of Statistical Inference from STATS 10x.

Review of 10x

Analysis techniques from STATS 10x and Case Studies.

Models, Assumptions and Transformations

The linear model and the underlying assumptions of the model. An introduction to Analysis of Variance and Multiple Comparison procedures. Transforming data to satisfy the assumptions of the linear model.

Two-way Analysis of Variance

2 grouping factors, understanding interactions between factors.

Categorical Data

Chi-square test, odds and odds ratio.

Regression

Regression as trend + scatter + watch out for outliers (revision of STATS 10x), fitting parametric trend curves (lines, polynomials), models for Y and models for $\log(Y)$, smoothing, transforming to linearity, multiple regression, dummy variables, prediction, model diagnostics, logistic regression.

Time Series

Components, plots, time series regression analysis and forecasting.

5. Course Materials

5.1 Materials required

The following materials will be required:

- Lecture Workbook
- CD-ROM
- At least one 3½" floppy disk

The Lecture Workbook and CD-ROM may be purchased at a price of \$30.00 at the **Student Resource Centre** [Building 303 (Science Centre)].

The Lecture Workbook defines the examinable content of STATS 20x in 2004.

Calculators used in STATS 10x will suffice for Stage 2.

5.2 Recommended reading

We strongly recommend the STATS 10x textbook *Chance Encounters: A First Course in Data Analysis and Inference* by C.J. Wild and G.A.F. Seber.



Course Information

6. Assessment

Your Final Mark for your STATS 20x statistics course is calculated as:

Final Mark = Assignments (20%) + Term Test (20%) + Final Exam (60%)

or = Assignments (10%) + Term Test (10%) + Final Exam (80%)

WHICHEVER IS HIGHER!

- Note:**
- *A minimum of 45% in the exam* is necessary to pass the paper (in addition to a minimum overall mark of 50%).
 - *All five assignments* are used to calculate the 10% or 20% contribution to your final mark.

7. People

COURSE COORDINATOR:

	Building	Room	Ext	Email
<i>Mike Forster</i>	T721	309	88759	m.forster@auckland.ac.nz

COURSE ADMINISTRATORS:

	Building	Room	Ext	Email
<i>David Smith</i>	303	226	85390	dp.smith@auckland.ac.nz
<i>Christine Miller</i>	303	204	84275	c.miller@auckland.ac.nz
<i>Mike Forster</i>	T721	309	88759	m.forster@auckland.ac.nz

Course Administrators are responsible for all inquiries about changing courses, assignment extensions, problems with sitting tests, mark records and complaints about assignment marking.

LECTURERS

	Building	Room	Ext	Email
City:				
<i>Mike Forster</i>	T721	309	88759	m.forster@auckland.ac.nz
Tamaki:				
<i>Russell Millar</i>	T721	333	85289	millar@stat.auckland.ac.nz



Course Information

8. Services Provided

8.1 Lectures

LECTURE TIMES

City:	Monday – Friday	10:00am
	Monday – Friday	12:00 noon
Tamaki:	Monday – Friday	8:30am

Lecture Handouts may be provided very occasionally in class. The Lecture Workbook is not complete — the lecturers will often write additional material on a slide, or leave important points off the handout for you to fill in after class discussion. Since the lectures in STATS 20x define the examinable content of the course, if you cannot make a class, you should obtain the notes from a classmate or from your CD-ROM.

8.2 Statistics Computer Laboratory

In the City there is one computer laboratory available for STATS 20x students to use. This laboratory is located in the basement of Building 303 (Science Centre). Refer to §8.6 (page vii) for a map showing the location of the laboratory.

You should learn to use the laboratory, even if you have a home machine. There are *tutor-demonstrators* working in the laboratories for some of the hours that they are open. These people are there to help you with any problems you experience with the laboratory machines and their software.

Opening hours in the city will be available on Cecil.

At Tamaki the laboratories are open Monday – Saturday, with hours for each laboratory posted on doors to buildings containing the laboratories and on laboratory doors.

8.3 Internet Information

STATS 20x uses the University of Auckland’s computer supported learning platform, Cecil. It will be necessary for students to have access to the internet for connection to Cecil.

The following information can be obtained from Cecil.

- Course information
- Current assignments, including data sets
- Coursework marks
- Computer tutorials, including data sets
- Frequently asked questions
- Information about **R** (including **R**-help)
- Notice board providing up-to-date announcements
- A course web board for communication



Course Information

8.4 Obtaining R

The package used in this course is available in versions for Windows 95/98/NT, 2000, ME, XP. A CD-ROM is provided with your Lecture Workbook. This CD-ROM contains R, plus data sets for the course.

To install R at home refer to page 1 of **Computer Tutorials** in the R-Package or “Getting Started” on your Course CD-ROM.

8.5 Other Forms of Assistance

Statistics Assistance Room (City)

Statistics Assistance Room tutors will be able to assist with conceptual problems with the course and its assignments. You may also get, but should not expect, assistance with problems with R.

The Assistance Room is located in B09 – refer to §7.6 Basement Map, below.

Opening hours will be posted on Cecil.

Common Room For Maori And Pasifika Students

Some help with STATS.20x for Maori and Pasifika students will be available in rooms 115, 117, 120 and 122 (first floor of Building 303 (Science Centre)). Room 120 has computers and can be used to relax, meet other students, and make a cup of coffee. Room 122 is for quiet study and tutorials. Rooms 115 and 117 are used for one-to-one tutoring and small group tutorials. For more information contact Viliami Latu (Room 121, ext 83063), Susan Wingfield (Room 119, ext 84934) or Garry Nathan (Room 118, ext 84931).

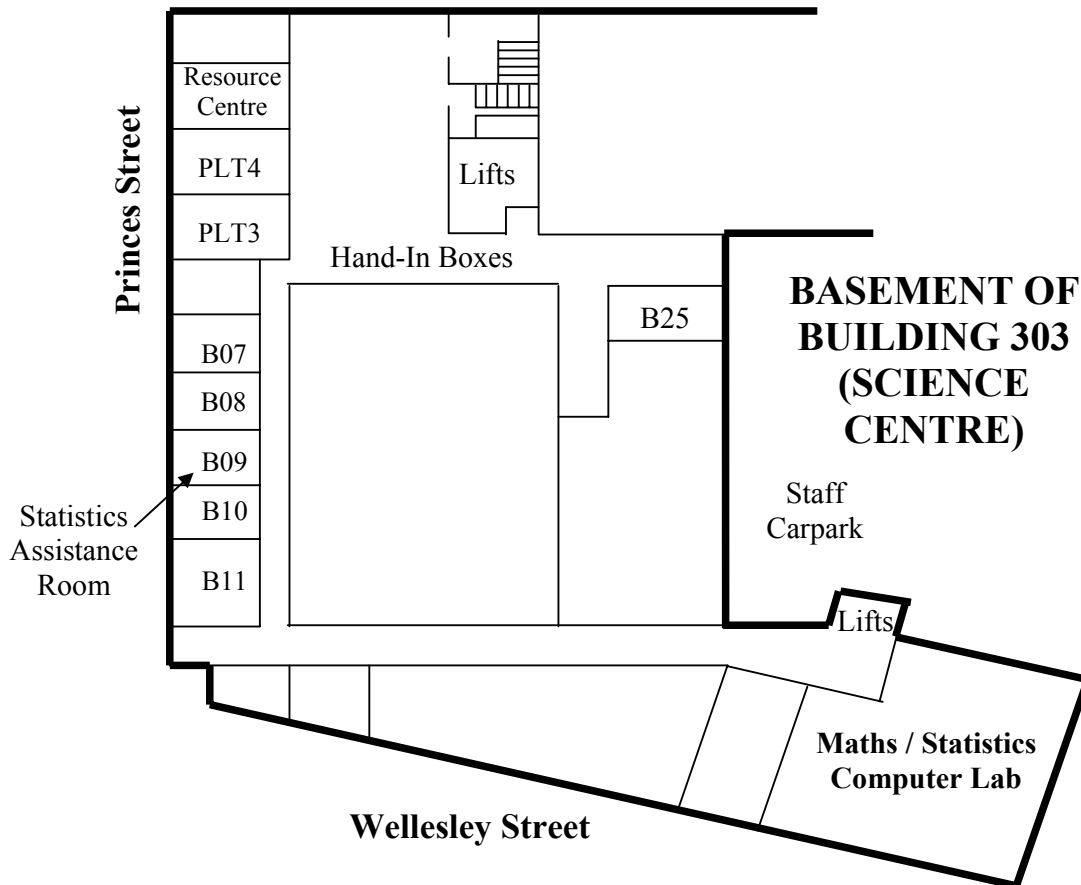
Lecturers

Your Lecturers are happy to help you after class and at posted office hours. They may also be available when you can find them outside those times.



Course Information

8.6 Basement Map



9. Tests and Examinations

9.1 Format

There is one Mid Semester Test contributing 10% or 20% to the final grade and a Final Examination contributing 60% or 80% of the final grade.

- **The test** will be
 - a 1 hour test
 - entirely multiple choice
 - “closed book” (though a formula sheet will be provided)
- **The exam** will be
 - a *3 hour exam*
 - approximately 30% multiple choice and 70% written.
 - “closed book”
 - a formula sheet will be provided
 - format will be the same as Semester 2, 2003



Course Information

9.2 Test and Exam Dates

The test will be held on *Monday 26th April at 6:30pm – 7:30pm.*

The exam will be held on *Wednesday 23rd June at 9:15am – 12:30pm.*

Note: You must bring your student ID card to all tests and examinations.

9.3 Inability to Sit

If you know that you will not be able to sit the test at the appointed time, you should see a Course Administrator:

	Building	Room	Ext	Email
<i>David Smith</i>	303	226	85390	dp.smith@auckland.ac.nz
<i>Christine Miller</i>	303	204	84275	c.miller@auckland.ac.nz
<i>Mike Forster</i>	T721	309	88759	m.forster@auckland.ac.nz

If you have a legitimate reason (e.g., sickness or bereavement) alternative provisions will be made.

[*Outside job commitments and going overseas on holiday will not be treated as legitimate* — you have sufficient notice to schedule most other commitments so that they do not conflict with this date.]

If you are unable to sit the test and have not made prior arrangements, you will have to apply for special consideration using a form available from Registry.

9.4 Aegrotat and Compassionate Passes

If you should be sick or injured prior to or at the time of the final exam and are unable to sit the exam or think that your performance was seriously affected, you may be eligible for an Aegrotat pass.

If through bereavement or some circumstance beyond your control (other than your own illness or injury) you are unable to sit the exam or think that your performance was seriously affected, you may be eligible for a Compassionate pass.

To be eligible for either of these, your lecturers must be able to certify that:

- (a) work done by you during the course was well above average, or
- (b) if work done by you during the course was not well above average, after taking into consideration your semester's work *and your performance in the exams*, including an exam in another subject, you are clearly worthy of a pass.

Application forms can be obtained from Registry, and must be submitted within seven days of the exam along with a medical certificate or other evidence. Please see the University Calendar for more detailed information.



Course Information

10. Assignments

Assignments count towards your final grade. All assignments are counted and all contribute equally.

10.1 Assignment Due Dates

Assignment Number	Due date
1	Monday 15 th March
2	Monday 29 th March
3	Monday 26 th April
4	Monday 10 th May
5	Monday 24 th May

10.2 Handing In

Assignments must be inserted into the correct box in the **City** and at **Tamaki** by **4:00pm** on the due date. Late assignments will **not** be accepted. Before you hand in your assignment it is *strongly recommended* that you make a photocopy or retain an electronic copy for your records. ***Student Resource Centre staff will not accept late assignments.***

City Campus: Hand-in boxes are located in the foyer of the basement of Building 303 (Science Centre). Refer to the map in Section 8.6 (page vii).

Tamaki Campus: Hand-in boxes are located in the foyer of the Tamaki Student Resource Centre, Building 710.1.

10.3 Handing In & Getting Back

Marked assignments are returned in return boxes placed in the same general area as was used for STATS 10x last year.

City Campus: Assignment-return pigeonholes are located in the foyer of the basement of the Building 303 (Science Centre). Refer to the map in Section 8.6 (page vii).

Tamaki Campus: Assignments will be returned to the pigeonholes in the Tamaki Student Resource Centre, Building 710.1.

Model answers will be posted on Cecil.



Course Information

10.4 Extensions

If you need an extension for handing in an assignment because of sickness or bereavement you should see a Course Administrator:

City:	David Smith	Room 226	Phone: ext 85390
	Christine Miller	Room 204	Phone: ext 84275
Tamaki:	Mike Forster	Room T721.309	Phone: ext 88759

Pressure from other university work will not be accepted as justification for an extension.

10.5 Marks

Class lists of coursework marks will be posted regularly on Cecil. Please check your marks. If a mark is incorrectly recorded or if you think mistakes have been made in the marking, please see one of the Course Administrators.

10.6 Aims and Content

Assignments are one of the most important learning experiences in this course. They enable you to apply the theory and techniques presented in lectures to actual data sets. They also enable you to develop the communication skills that are important if you wish to communicate the results of your analysis clearly and concisely. We do not want the following comment to apply to your reports:

"Like other occult techniques of divination the statistical method has a private jargon deliberately contrived to obscure its method from non-practitioners." Ashley-Perry
Statistical Axioms

Instead we want you to be able to present a report which is intelligible and interesting to both practitioners and non-practitioners of statistics.

10.7 Presentation of Assignments

A high standard of presentation is expected. Reports must be prepared using a word processor or typed or neatly and legibly hand-written in clear, coherent English. Annotated computer printouts from **R** will not be accepted. Any graphs, tables or other computer output that you wish to include in your report should be cut out from the computer output and pasted into your report.

- Use standard A4 sized paper.
- Number each page and write your name in the top right hand corner of each page with surname (family name) underlined.
- Include page numbers in the bottom right hand corner of each page.
- Fill out **both sides** of a Statistics Department assignment cover sheet.
- Place the cover sheet at the front of your assignment.
- Staple the cover sheet and your assignment together.
- Fold the cover sheet and assignment in half **lengthwise** with the cover sheet facing outwards.



Course Information

10.8 Collaboration

Working together is an important part of the learning process. We encourage you to discuss assignment problems with one another, with tutors and lecturing staff. However, **you must not copy** another person's assignment. You can work together to decide what should be done, but ultimately, **you must produce your own computer output and write your own report.**

You must not collaborate during tests and examinations (this includes sharing a calculator).

We view cheating on assignment work as seriously as cheating in an examination. Cheating is:

- copying all or part of another student's assignment;
- allowing another student to copy all or part of your assignment. **This is treated as seriously as copying another student's assignment;**
- taking a copy of another student's work without their knowledge. This is treated as theft and will be referred to the University Discipline Committee.

The Penalties for copying another person's work include:

- **loss of some or all marks for the assignment/test;**
- **loss of additional coursework marks;**
- **requesting that the students involved withdraw from the paper;**

11. Tutorials

These tutorials are mainly designed to assist students who have difficulty with statistics or with using **R**, although all 20x students are welcome to attend. They will be held weekly, beginning in Week 1. Times will be announced in lectures and on Cecil.

The focus is on analysing data using **R**, interpreting the output we get from **R** and writing up Technical Notes on the analysis and an Executive Summary of the main findings from the analysis. (The above make up about 85% of the assessment in STATS 20x.)

In each tutorial (except for Tutorial 1), the tutor will review the work covered during the week, and will work through two or three data sets. In the Tutorial section of your Lecture Workbook, you will find the data sets and the names of the variables in those data sets for each tutorial. You can access the data files from Cecil or from your course CD. They are all in `.txt` format and all have header lines of variable names.

You will get the most benefit from these tutorials if you use the **R** commands that are given to analyse the data, and look carefully at the output you have generated. If you wish to, you may cut and paste your **R** output and plots into a Word processing document and bring it to the tutorial with you. If you have time, it would be a good idea to jot down some brief comments in pen about the analysis, much as we do in the Case Studies done in lectures.



Course Information

NOTE: At the end of the tutorial, you **may** hand in your **R** output. If you are borderline pass/fail for the course, we will take your tutorial work into account when making our decision. The tutorial sheets will be returned through the assignment return boxes. Make sure you put your Student ID Number on the tutorial sheet before handing it in.

After the tutorial, you should attempt to write up the Technical Notes and Executive Summaries.

If you do as we suggest in preparing for, and following through on these tutorials, you will find doing your assignments considerably easier, and faster.

The complete **R** output, the Technical Notes and Executive Summaries will be posted on Cecil after the tutorials.

11.1 Tutorial Room and Times

Tutorial Rooms in the City and at Tamaki will be announced in class and posted on Cecil.

Tutorial times will also be announced in class and posted on Cecil.

