

## MTH5120 Statistical Modelling I

## Practical 1

This practical introduces you to fitting regression models using MINITAB. *You do not need to hand in anything this week.* However, next week's course-work will require you to do very similar analyses on a new set of data so the better you understand what we are doing now the easier the course-work will be for you next time.

When you log on you should select the Mathematics Department and the environment

STM Statistical Modelling.

You should put all your files for this course in a suitable directory (for example in your home directory usually mapped to the `g:/` drive).

Now begin MINITAB.

### Example 1

To calibrate an instrument which measures the lactic acid concentration in blood, an investigator used twenty samples of known concentration ( $x$ ) and then obtained the concentration as measured on the instrument ( $y$ ). The results were

$x$	1	1	1	1	3	3	3	3
$y$	1.1	0.7	1.8	0.4	3.0	2.4	4.9	4.4
$x$	5	5	5	5	10	10	10	10
$y$	4.5	7.3	8.2	6.2	12.0	13.1	12.6	13.2
$x$	15	15	15	15				
$y$	18.7	19.7	17.4	17.1				

Enter the  $x$  values into Column C1 of the data window and the  $y$  values into Column C2. Label the columns as  $x$  and  $y$ .

1. Plot  $y$  against  $x$ . Add a title. Get into the habit of giving your plots titles.
2. Does the relationship between  $y$  and  $x$  seem to be linear? Write down the linear model and the assumptions usually made about the random error.

3. Fit the model to the data by **Stat**→**Regression**→**Regression...**. Select  $y$  as the response and  $x$  as the predictor. See the numerical output in the Session Window. To see the fitted line on the plot use **Stat**→**Regression**→**Fitted Line Plot...**
4. What would you expect “to observe” at  $x = 0$ ? Is the intercept statistically significant?
5. Refit the model without intercept (go to **Options** in **Stat**→**Regression**→**Regression...**). What has changed?
6. Check the model assumptions. Does the normal plot give a straight line?
7. Does the plot of residuals versus fitted values suggest a random spread of points? Note that because the  $x$  values are repeated this plot will have a number of vertical lines. We are looking to see if they are all of a similar spread, which will confirm that the errors have a constant variance, and are all centered at zero, which will confirm that a model linear in  $x$  is appropriate.
8. The plot of residuals against order isn’t very useful here as we don’t know that the data were collected in the order given in the table.
9. Interpret the regression equation (model fit) in terms of the practical meaning of  $x$  and  $y$ . Suppose a new sample has concentration  $x = 7$ . What would we estimate to be the value of  $y$ ? Obtain this estimate.
10. Test the null hypothesis that the slope is zero, versus the alternative hypothesis that it is nonzero. Comment briefly, in non-statistical language, about the result of the test.

### Example 2

A fire insurance company wants to relate the amount of fire damage in major residential fires to the distance between the residence and the nearest fire station. The study was conducted in a large suburb of a major city; a sample of fifteen recent fires in the suburb was selected. The amount of damage,  $y$  (£000), and the distance,  $x$  (km), between the fire and the nearest fire station are given in the following table.

$x$	3.4	1.8	4.6	2.3	3.1	5.5	0.7	3.0
$y$	26.2	17.8	31.3	23.1	27.5	36.0	14.1	22.3
$x$	2.6	4.3	2.1	1.1	6.1	4.8	3.8	
$y$	19.6	31.3	24.0	17.3	43.2	36.4	26.1	

1. Plot the data.
2. Write down the regression model the data suggest. Also, write the model assumptions.
3. Fit the simple linear regression model to the data, plot the fitted line and write down its equation.
4. Look at the residual plots, do they give any reason to doubt our model?
5. Test the null hypothesis that the regression is not significant (the slope is zero) and interpret the result in non-technical language.
6. Estimate the cost of the damage for a residence 3.5 km from the nearest fire station.
7. Interpret the meaning of the slope and of the intercept parameters.

## Writing a Minitab Project Report

When writing a report remember:

- Include only the information which is relevant to the question.
- Always write down the model you work on and the assumptions.
- When testing a hypothesis, write down the null hypothesis and the alternative, the test function and its distribution, the p-value and the result of the test.
- Results should be commented on in technical (statistical) language and then interpreted in a non-technical (practical) way.
- Comments should always be brief and to the point.

You may create a report of your work using the ReportPad folder in the Project Manager (ctrl I). From the Graphs directory in the Project Manager copy the pictures into your report (right-clicking on the graph's name will show a set of options, one of them is Append to Report). Make the graphs small, so that two graphs are in one line (saving paper). Usually, if there are important results in the Session window you should also append them to the report.

To save the Report in your special for this course directory (create it now and call it SMI) as an \*.rtf file choose

File ► Save Report As

You may then edit it further in Microsoft Word and save as a \*.doc file (it takes less space than its \*.rtf counterpart).

You can also edit the Session Window Output if necessary. By default, the Session window is read-only, so that you cannot accidentally delete results. To be able to edit text you have to make the Session window editable:

- Press [Ctrl]+[M] to make the Session window active.
- Pull down the Editor menu.
- If there is no check mark next to Output Editable, then select it to enable Session window editing.
- If there is already a check mark next to Output Editable, then Session window editing is already enabled.