Statistics in Practice

Case Studies in Medicine

Teacher's Guide

Royal Statistical Society Centre for Statistical Education
Department of Mathematics, Statistics and Operational Research
Nottingham Trent University
Nottingham NG1 4BU

Phone: 0115 848 4476
Fax: 0115 848 4771
http://science.ntu.ac.uk/rsscse/

Relevant Web Sites

These include:

- http://www.nhlbi.nih.gov/guidelines/obesity/ob_home.htm
  (American) National Institute of Health Clinical Guidelines on Overweight and Obesity
- http://www.drrossfox.com/abotobs.html
  Medical Services Organization: About Obesity
- http://www.iotf.org/
  International Obesity TaskForce
- http://www.iaso.org/
  International Association for the Study of Obesity
- http://www.naaso.org/
  North American Association for the Study of Obesity
- http://www1.umn.edu/mnoc/
  Minnesota Obesity Center
- http://www.loop.com/~bkrentzman/
  Science of Obesity and Weight Control
- http://www.obesity.chair.ulaval.ca/Genes.html
  Human Obesity Gene Map
- http://www.paiweb.com/obesityrisk.html
  Obesity Risk Assessment Software
- http://www.idf.org/
  International Diabetes Federation
- http://www.learneducation.com/
  LEARN Education Center

Electronic versions of the resource pack may be downloaded from
http://www.cse.dmu.ac.uk/Courses/MScIDM/StatPack/
and
http://science.ntu.ac.uk/rsscse/schools/StatPack/

Bibliography

1. Altman, D.G., *Practical Statistics for Medical Research*, Chapman and Hall, 1991.

   A useful text providing background material to clinical trials and epidemiology which includes many worked examples.

Other useful introductory texts are:

2. Campbell, M.J. and Machin, D., *Medical Statistics: A Commonsense Approach*, Wiley, 2nd edn, 1993.

3. Daniel, W.W., *Biostatistics: A Foundation for the Health Sciences*, Wiley, 6th edn, 1995.

4. Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J. and Ostrowski, E., *A Handbook of Small Data Sets*, Chapman and Hall, 1994.

   As the title suggests, a useful source of data for classroom exercises.

5. Martin, P. and Pierce, R., *Practical Statistics for the Health Sciences*, Nelson, 1994.

6. Rees, D.G., *Essential Statistics for Medical Practice: A Case Study Approach*, Chapman and Hall, 1994.

   A useful source of additional case-study material.

7. Selvin, S., *Practical Biostatistical Methods*, Chapman and Hall, 1995.

Contents

Case Study One

Obesity - Can a New Drug Help?

This case study is an example of a *clinical trial* to assess the effectiveness of a new drug as part of a combination therapy (diet, exercise and drug) to treat obesity. Data is provided on 37 patients who received either the new combination therapy or a 'dummy' treatment (*placebo*).

A useful text giving further background to clinical trials is the book by Altman, D.G., *Practical Statistics for Medical Research*, Chapman and Hall, 1991.

The Student Card provides some simple background to clinical trials and the full data on the 37 patients taking part in this obesity study. Three questions of interest are identified:

i. What are the characteristics of the patients involved?
ii. Are the patients in the two groups similar at the start of the study?
iii. Is there any difference in terms of outcome for patients in the two treatment groups?

Note that, in Q(iii), the comparison of new treatment to placebo can be made by comparing the mean change in weight for patients in the two groups and/or by comparing the % of patients in the two groups who achieve a >5% reduction in weight.

The data can be used to illustrate:

i. data summary and exploration - students could be asked to provide summaries and appropriate diagrams for the whole dataset and also to summarise the characteristics of the patients in each of the two treatment groups prior to the start of treatment (are the two treatment groups similar at the start of the study?). They can go on to look at simple Box Plots and 1-variable statistics for the placebo versus new drug groups and compare them.

ii. a hypothesis test for the difference of two means (change in weight in the two treatment groups) with variance known (assuming a standard deviation (s.d.) of 2.57 kg as given on the Student Card).

Ideas for Student Projects

The area of obesity is a good source of project material which could include:

1. measuring BMI (body mass index), WHR (waist to hip ratio), simple body weight, etc. on a sample of people comparing, for instance, males and females;

2. conducting a questionnaire survey to examine people's knowledge of the causes and risks associated with obesity;

3. surveying eating habits and lifestyle (amount of regular exercise taken, etc.);

4. designing a study to collect the data and evaluate the risk of heart disease amongst a sample of adults. Use the obesity risk assessment model on the Internet (at http://www.paiweb.com/obesityrisk.html) to calculate the risks and life expectancies.
   (You will need to make up values of blood pressure.)

- patients with high blood pressure (hypertensive) have 1.8 times the risk of a major coronary event compared to patients with normal blood pressure (normotensive).

It can therefore be concluded that the results of this study are consistent with those of previous studies in relation to smoking and hypertension.

The risks associated with diabetes and MI in family history

Again by reference to Table 1 it can be seen that

- patients with diabetes have 1.8 times the risk of a major coronary event compared to non-diabetics;
- patients with MI in the family history have 1.4 times the risk of a major coronary event compared to those with no family history.

Therefore it can be concluded that the risks associated with diabetes and MI in family history are less than that associated with smoking.

Risks associated with obesity and triglycerides

- Body Mass Index (BMI) can be used as a measure of obesity. Table 1 shows that the risk of a major coronary event increases as BMI increases, with patients of BMI > 27.3 kg/m^2 having 1.6 times the risk of those with BMI < 25.0. This is less than the increased risk for smokers.
- The risk of a major coronary event associated with triglycerides increases with triglyceride level, a level > 166 mg/dl leading to 2.6 times the risk compared to a level < 105 mg/dl. This increased risk is therefore comparable to that associated with smoking.

iii. a χ^2 test of independence to examine whether or not there is any association between 'percentage of patients achieving a 5% reduction in weight' and 'treatment group'.

Students studying more advanced modules might also be asked to perform the following additional analyses.

iv. A t-test to compare the mean change in weight in the two treatment groups.

v. A non-parametric comparison of change in weight in the two groups. **Note** however that the data is real and includes tied observations in the change in weight.

Sample solutions for all of the above aspects now follow.

Data Summary - Obesity Study, All Patients

Note: The data provided is real and typical of many real datasets - it includes a missing value - for patient 134 we do not know if there was a family history of obesity.

Variable	n	Mean	Median	StDev	Min	Max
Age (yrs)	37	45.38	44.0	8.37	27.0	62.0
Height (cm)	37	167.97	169.0	8.29	152.0	183.0
Wt at Entry (kg)	37	98.24	101.5	13.97	67.7	127.5

Gender	n	%
Female	25	67.6%
Male	12	32.4%

Previous Weight Loss Attempts							
Weight Loss Attempts	1	2	3	4	5	6	7
Number of Patients	12	7	9	2	3	1	3

Data Summary: **Are the patients in the two groups similar at the start of the study?**

Variable	Treatment Group	n	Mean	Median	StDev	Min	Max
Age (yrs)	Placebo	19	46.37	45.0	8.91	27.0	62.0
	New Drug	18	44.33	42.5	7.87	31.0	59.0
Height (cm)	Placebo	19	166.74	168.0	8.88	152.0	180.0
	New Drug	18	169.28	169.5	7.65	156.0	183.0
Wt at Entry (kg)	Placebo	19	96.4	98.4	13.99	67.7	117.9
	New Drug	18	100.18	103.2	14.08	81.5	127.5

Number (%) of Patients for each Treatment by:

Gender

	Female	Male
Placebo	13 (68.4%)	6 (31.6%)
New Drug	12 (66.7%)	6 (33.3%)

Previous Weight Loss Attempts

Attempts	1	2	3	4	5	6	7
Placebo	5	4	4	2	2	1	1
New Drug	7	3	5	0	1	0	2

Case Study Three

An Epidemiological Study - The Risk of Heart Disease

This case study is an example of an *epidemiological study* in which the incidence of disease and the risks associated with factors such as lifestyle, the environment, etc. are investigated.

The PROCAM study examined the risk of heart disease amongst people at work and involved employees at 52 companies and authorities in Germany. The participation rate varied from company to company, ranging from 40% to 80% and averaging 60%.

The Student Card provides some background information relating to epidemiology and heart disease in general, together with a more detailed description of the PROCAM study. The Student Card includes summary tables of results from the data analysis. These may be used to investigate the following questions of interest, as identified on the card.

i. Other medical research suggests that smoking and hypertension (high blood pressure) lead to increased risk of heart disease. Does this study agree with previous findings?
ii. What can be concluded about the risks associated with Diabetes and Previous Myocardial Infarction (MI) compared to the risk associated with smoking?
iii. What are the risks associated with Obesity and Triglyceride Levels as suggested by the data from this study?

Sample outline solutions are now provided.

Does this study agree with previous findings?

Table 1 on the Student Card provides the relative risk of a major coronary event for a number of factors including smoking and hypertension.

It can be seen that:

- smokers have 2.3 times the risk of a major coronary event compared to non-smokers;

Hence the reduction in triglyceride level predicted for a patient with a weight loss of 5 kg would be about 27.9 mg/dl.

This shows the effect that a single point can have on a relatively small database. A closer consideration of the patient giving the outlying point confirmed that the data were correct. The company also knew of other larger studies confirming the essential accuracy of the results calculated from all 35 points.

Prediction for a weight loss of 10 kg - Dangers of extrapolation

The practice of extrapolation, where a prediction of the dependent variable *y* (change in triglyceride) is made for a value of the independent variable *x* (change in weight) outside the range of *x* values observed in the study and which have been used to calculate the regression line, should be used with great caution. The true relationship between the variables, although linear over the interval observed in the study, may change drastically outside the range that has been investigated. A weight loss of 10 kg is outside the range observed and extrapolation is therefore not recommended.

Family history of obesity?

	Yes	No
Placebo	12 (66.7%)	6 (33.3%)
New Drug	13 (72.2%)	5 (27.8%)

Motivation rating

Rating	1	2	3
Placebo	7	10	2
New Drug	7	10	1

(1 missing value in placebo group)

Clearly the characteristics of the patients in the two groups are very similar.

Comparison of mean change in weight in the two treatment groups - Variance known

Note that students will need to generate a new column of data:
(Weight at week 8 – Weight at week 0)

Data summary:

		Change in weight
Placebo group	$n_1 = 19$	mean $\bar{x}_1 = -1.363$
New treatment group	$n_2 = 18$	mean $\bar{x}_2 = -5.011$

s.d. $\sigma = 2.57$ kg

$$Z_{\text{calc}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{|-1.363-(-5.011)|}{2.57\sqrt{\frac{1}{19} + \frac{1}{18}}} = 4.3155$$

$Z_{\text{calc}} > 2.58$ indicates significance at the 1% level.

Therefore we conclude that there is strong evidence of a difference in the mean weight loss brought about by the two treatment regimens. The data indicates that the new drug treatment results in a greater mean weight loss than placebo.

χ^2 test of independence for '% of patients achieving 5% reduction in weight' and 'treatment group'

Note: Students will need to generate a new column of data by comparing the Change in Weight to the Weight at Week 0 to see, for each individual patient, whether or not he/she achieves a 5% reduction in weight. This new column should consist of code = 0 (did not achieve 5% reduction in weight) and code = 1 (did achieve 5% reduction in weight). The column can then be summarised in terms of the number of patients in each group who did/did not achieve the 5% reduction, resulting in the following 2 × 2 contingency table.

Treatment Group	Achieved 5% reduction in weight?		Total
	Yes	No	
Placebo	1	18	19
New Treatment	10	8	18
Total	11	26	37

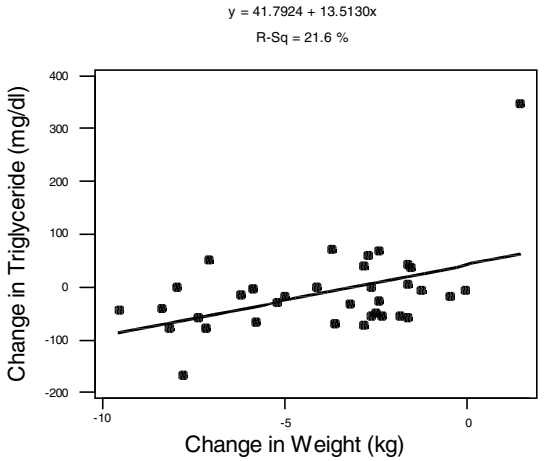
$$\sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} = 11.191 \qquad \chi^2_1(5\%) = 3.84 \quad \chi^2_1(1\%) = 5.99$$

$\chi^2_{calc} = 11.191 > 5.99$ indicates significance at the 1% level.

Therefore we conclude that there is strong evidence of a difference in the proportion of patients achieving a >5% reduction in weight in the two treatment groups.

The data indicates that the proportion achieving such a weight loss is higher for patients in the new drug group compared to the placebo (55.6% as compared to 5.3%).

Scatterplot showing the Fitted Regression Line



Prediction of change in triglyceride for a weight loss of 5 kg

$$\begin{aligned} \textit{Change in Triglyceride} &= 41.8 + 13.5(\textit{Change in Weight}) \\ &= 41.8 + 13.5(-5) \\ &= -25.8 \end{aligned}$$

Therefore, on average, a 25.8 mg/dl drop in triglyceride level would be predicted for a patient with a weight loss of 5 kg.

Discussion

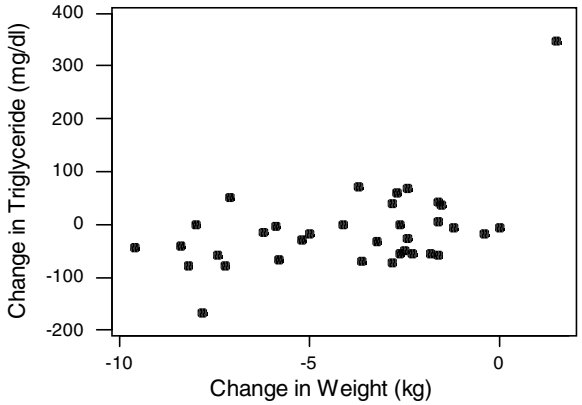
Before the above calculations it was stated that there was an apparent outlier in the top right hand corner of the graph. It is interesting to recalculate the correlation coefficient with this point removed.

These new calculations give a correlation coefficient of 0.336 (instead of 0.464) which is just less than the tabulated value for $p = 0.05$, so the evidence would be much weaker for rejecting the hypothesis that the true correlation is zero.

Similarly the regression line becomes:

$$\textit{Change in Triglyceride} = 4.63 + 6.5(\textit{Change in Weight})$$

Scatterplot of Change in Triglyceride v. Change in Weight



The scatterplot indicates that a linear relationship may be appropriate with one apparent outlier in the top right hand corner.

Calculation of correlation coefficient and linear regression line

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}} = 0.464$$

0.464 > 0.418 (the tabulated value at $p = 0.01$ for 33 degrees of freedom), indicating that there is strong evidence to reject a hypothesis that the true correlation is zero.

For the linear regression line $y = a + bx$,

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \text{ and } a = \bar{y} - b\bar{x}.$$

Therefore

Change in Triglyceride = 41.8 + 13.5(*Change in Weight*)

Comparison of mean change in weight in the two treatment groups - t-test (pooled estimate of variance)

Pooled estimate of the variance $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 - 1 + n_2 - 1} = 5.9536$

$$t_{calc} = \frac{|\bar{x}_1 - \bar{x}_2|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{|(-1.36) - (-5.01)|}{2.44\sqrt{\frac{1}{19} + \frac{1}{18}}} = 4.54$$

$t_{35}(5\%) = 2.032$
$t_{35}(1\%) = 2.727$

$t_{calc} = 4.54 > t_{35}(1\%)$ and therefore we conclude that there is strong evidence of a difference in the mean reduction in weight in the two groups. Examination of the data indicates that the mean weight reduction for the new treatment group is greater than for the placebo group.

Non-parametric comparison of change in weight in the two groups

Note: The data included in this case study arises from a real medical study and therefore includes tied values in change in weight.

Wilcoxon Rank Sum Test:

Let S = sum of ranks of the smaller sample (the new drug group) where tied values are given the mean of the ranks that would have been assigned if no ties existed.

Test statistic $T = S - \frac{n(n+1)}{2}$ so that $T_{calc} = 222 - \frac{18(19)}{2} = 51$

As both sample sizes clearly exceed 10, students may apply the Normal approximation such that

$$T \sim \text{N}\left[\frac{n_1(n_1 + n_2 + 1)}{2}, \frac{n_1 n_2(n_1 + n_2 + 1)}{12}\right].$$

In this example therefore $T \sim$ N(342, 1083) and $z = \frac{|222 - 342|}{32.91} = 3.646$

As 3.646 > 2.58, we conclude that there is strong evidence of a difference in the weight loss observed under the two treatments.

Case Study Two

Investigating the Relationship between Changes in Weight Loss and Triglyceride Level

This study is designed to investigate the relationship between two variables, weight loss and triglyceride level. Data is provided for 35 overweight patients who followed a treatment regimen of diet, exercise and drug for an eight week period.

The Student Study Card provides some background information describing the debate relating to the increased risk of heart disease brought about by high triglyceride levels, together with the full patient data and five questions of interest:

i. What are the characteristics of the patients involved?
ii. Does there appear to be a relationship between change in triglyceride level and change in weight?
iii. Is it possible to obtain a simple formula for predicting change in triglyceride level from change in weight?
iv. What change in triglyceride level should be predicted for a patient who loses 5 kg over an eight week period?
v. Is it sensible to use a straight line to make a prediction of change in triglyceride level for a patient who loses 10 kg in weight?

Note that students will need to generate new columns of data for weight loss (weight at week 8 – weight at baseline) and change in triglyceride level (triglyceride at week 8 – triglyceride at baseline).

The data can be used to illustrate:

i. data summary and exploration - students could be asked to provide appropriate summaries to characterise the 35 patients both at the start and end of the study, together with the changes in weight and triglyceride level;

ii. the use of a scatterplot to investigate the relationship between weight loss (x) and triglyceride (y);

iii. calculation of the product moment correlation coefficient and a test of zero correlation together with determination of the linear regression line which students could be asked to plot on their graph;

iv. use of both the fitted regression equation and the plotted line for predicting the change in triglyceride level brought about by a specified reduction in weight;

v. discussion of the dangers of extrapolation.

Sample solutions to the above now follow.

Data summary - Weight Loss and Triglyceride Study

Variable	n	Mean	Median	StDev	Min	Max
Weight at Baseline	35	98.52	99.2	14.29	69.4	133.1
Triglyceride at Baseline	35	168.5	154.0	77.3	57.0	446.0
Weight at Week 8	35	94.58	95.5	14.37	69.0	130.8
Triglyceride at Week 8	35	157.1	126.0	126.6	54.0	795.0
Change in Weight	35	-3.934	-2.8	2.761	-9.6	1.5
Change in Triglyceride	35	-11.4	-18.0	80.4	-169.0	349.0