Statistics in Clinical Trials

Exercises

Dr Nick Fieller

Probability & Statistics, SoMaS

University of Sheffield

visiting

2012

(this page left blank for notes)

1. Read the article referred to in §1.8, this can be accessed from the web address given there or from the link given in the course web pages. Use the facility on the BMJ web pages to find related articles both earlier and later. Key reason for doing this is that it is a common question at interviews to discuss current issues in statistics (e.g. Civil Service, biomedical and financial industries). In biomedical industries in particular you will be expected to know about ethical issues ***and to have opinions on them***.
2. Revision of t-tests and non-parametric tests. The data set HoursSleep which can be accessed from the course website gives the results from a cross-over trial comparing two treatments for insomnia. Group 1 had treatment A in period 1 whilst group 2 had B (and then the other treatment in period 2). Use a t-test to assess the differences between the mean numbers of hours sleep on the two treatments in period 1. Compare the p-values obtained using separate and pooled variance options. Next assess the difference in medians of the two groups using a non-parametric Mann-Whitney test. Compare the p-value obtained from this test with those from the two versions of the t-test. The functions required are **`t.test()`** and **`wilcox.test()`**. Use the **`help()`** system to find out how to use them.
3. Using your general knowledge compare the following two theories against the Bradford-Hill Criteria:

   (i) Smoking causes lung cancer

   (ii) The MMR (mumps, measles and rubella) vaccine given to young babies causes autism in later childhood.

©NRJF, University of Sheffield, 2012

4. For each of the proposed trials listed below, select the most appropriate study design, allocating onne design to onne trial. (Onne≡'one and only one'!)

Trial

A Comparison of surgery and 3 months radiotherapy in treating lung cancer.

B Comparison of new and standard drugs for relief from chronic arthritis

C Use of diet control and drug therapy for cure of hypertension

D Comparison of absorption speed of new and standard anaesthetics.

Design

a Crossover

b Parallel Group

c Sequential

d Factorial

©NRJF, University of Sheffield, 2012

5. In a recent radio programme an experiment was proposed to investigate whether common garden snails have a homing instinct and return to their 'home territory' if they are moved to some distance away. The proposal is that you should collect a number of snails, mark them with a distinctly coloured nail varnish, and place all of them in your neighbour's garden. Your neighbour should do likewise (using a different colour) and place their snails in your garden. You and your neighbour should each observe how many snails returned to their own garden and how many stayed in their neighbour's. (See http://downloads.bbc.co.uk/radio4/so-you-want-to-be-a-scientist/Snail-Swapping-Experiment-Instructions.pdf for full details)

   (a) What flaws does the design of this experiment have?

   (b) How could the design of the experiment be improved?

©NRJF, University of Sheffield, 2012

The next three questions are designed to give practise at locating the available (if any) published evidence to support claims of medical connections. The most reliable evidence is provided by published (in peer-reviews journals) [double blind] randomized controlled clinical trials of adequate size.

6. What evidence is there that grapefruit juice should never be taken by people receiving treatment for high cholestrol with statins?
7. What evidence is there that taking fish oil helps schoolchildren concentrate?
8. On a recent BBC Radio programme (Front Row, Friday 03/10/08, http://www.bbc.co.uk/radio4/arts/frontrow/) there was an interview with Bettany Hughes, (http://www.bettanyhughes.co.uk/), a historian, who was talking about gold (in relation to an exhibition of a gold statue of Kate Moss in the British Museum). She made the surprising statement

   "....*ingesting gold can cure some forms of cancer*."

   I would only regard this as true if there has been a randomized controlled clinical trial where one of the treatments was gold taken by mouth and where the measured outcome was cure of a type of cancer.

   The task is to find a record of such a clinical trial or else find a plausible source that might explain this historian's rash statement.

©NRJF, University of Sheffield, 2012

Questions on Constructing Randomization Lists

Look at Martin Bland's guide to randomization software

http://www-users.york.ac.uk/~mb55/guide/randsery.htm

and see what is available and what descriptions there are available. He also has guides to many other sources for software of use in Medical Statistics.

9. Patients are to be allocated randomly to 3 treatments. Construct a randomization list

i) for a simple, unrestricted random allocation of 24 patients

ii) for a restricted allocation stratified on the following factors with 4 patients available in each factor combination:

Sex: M or F Age: <30; 30≤&<50; ≥50.

10. Patients are to be randomly assigned to active and placebo treatments in the ratio 2:1. To ensure 'balance' a block size of 6 is to be used. Construct a randomisation list for a total sample size of 24.

11. Patients are to be randomly assigned to active and placebo treatments in the ratio 3:2. To ensure 'balance' a block size of 5 is to be used. Construct a randomisation list for a total sample size of 30.

12. In the comparison of a new drug A with a standard drug B it is required that patients are assigned to drugs A and B in the proportions 3:1 respectively. Illustrate how this may be achieved for a group of 32 patients, and provide an appropriate randomization list. Comment on the rationale for selecting a greater proportion of patients for drug A.

13. The table below gives the age (≤55/>55), gender (M/F), disease stage (I/II/III) of subjects entering a randomized controlled clinical trial at various intervals and who are to be allocated to treatment or placebo in approximately equal proportions immediately on entry.

order of entry	Age	Gender	Stage
1	≤55	F	III
2	≤55	M	III
3	≤55	M	I
4	≤55	F	I
5	>55	F	II
6	≤55	F	III
7	>55	F	I
8	>55	M	III
9	≤55	M	III
10	>55	F	III
11	≤55	F	III
12	≤55	M	I
13	>55	F	I

i) Use a minimization method designed to achieve an overall balance between the factors to allocate these subjects in the order given to the two treatments and provide the resulting list of allocations.

ii) Cross-tabulate the treatment received with each [separate] factor.

iii) Construct a list to allocate the subjects to treatment completely randomly without taking any account of any prognostic factor and compare the balance between treatment groups achieved on each of the factors.

©NRJF, University of Sheffield, 2012

Sample size questions

(in all cases take the significance level as 0.05)

The commands in **R** for calculation of power, sample size etc are `power.t.test()` and `power.prop.test()`. Note that typing the ↑ recalls the last **R** command and use of Backspace and the ← key allows you to edit the command and run a new version

14. How many subjects are needed to achieve a power of 80% when the standard deviation is 1.5 to detect a difference in two populations means of 0.8 using a two sample t-test? (Note that **R** gives the number needed in each group, i.e. total is twice number given)

15. How many subjects are needed to achieve a power of 80% when the standard deviation is 1.5 to detect a difference in one population mean from a specified value of 0.8 using a one sample t–test?

16. Do you have an explanation for why the total numbers in Q14 and Q15 are so different?

17. How many subjects are needed to detect a change of 20% from a standard incidence rate of 50% using a two sample test of proportions with a power of 90%?

18. How many subjects are needed to detect a change from 30% to 10% using a two sample test of proportions with a power of 90%?

19. How many subjects are needed to detect a change from 60% to 80% using a two sample test of proportions with a power of 90%?

©NRJF, University of Sheffield, 2012

20. How many subjects are needed to detect a change from 50% to 30% using a two sample test of proportions with a power of 90%?

21. How many subjects are needed to detect a change from 75% to 55% using a two sample test of proportions with a power of 90%?

22. How many subjects are needed to detect a change from 40% to 60% using a two sample test of proportions with a power of 90%?

23. Questions 17–21 all involve changes of 20% and a power of 90%. Why are the answers not all identical?

24. Without doing any calculations (neither by hand nor in **R**) write down the number of subjects needed to detect a change from 45% to 25% using a two sample test of proportions with a power of 90

25. A trial for the relief of pain in patients with osteoarthritis of the knee is being planned on the basis of a pilot survey which gave a 25% placebo response rate against a 45% active treatment response rate.

a) How many patients will be needed to be recruited to a trial which in a two-sided 5% level test will detect a difference of this order of magnitude with 90% power? (Calculate this first 'by hand' and then using a computer package and compare the answers).

b) With equal numbers in placebo and active groups, what active rates would be detected with power in the range 50% to 95% and group sizes 60 to 140? (Calculate for power in steps of 15% and group sizes in steps of 20).

26. Woollard & Cooper (1983) Clinical Trials Journal, 20, 89-97, report a clinical trial comparing Moducren and Propranolol as initial therapies in essential hypertension. These authors propose to compare the change in initial blood pressure under the two drugs.

    a. Given that they can recruit only 100 patients in total to the study, calculate the approximate power of the two-sided 5% level t-test which will detect a difference in mean values of 0.5σ, where σ is the common standard deviation.

    b. How big a sample would be needed in each group if they required a power of 95%? (Calculate this first 'by hand' and then using a computer package and compare the answers).

27. In a clinical trial of the use of a drug in twin pregnancies an obstetrician wishes to show a significant prolongation of pregnancy by use of the drug when compared to placebo. She assesses that the standard deviation of pregnancy length is 1.5 weeks, and considers a clinically significant increase in pregnancy length of 1 week to be appropriate.

    a. How many pregnancies should be observed to detect such a difference in a test with a 5% significance level and with 80% power?

    b. It is thought that between 40 and 60 pregnancies will be observed to term during the course of the study. What range of increases in length of pregnancy will the study have a reasonable chance (i.e. between 70% and 90%) of detecting?

Cross-Over Trials

28. Senn and Auclair (Statistics in Medicine, 1990, **9**) report on the results of a clinical trial to compare the effects of single inhaled doses of 200μg salbutamol (a well established bronchodilator) and 12μg formoterol (a more recently developed bronchodilator) for children with moderate or severe asthma. A two-treatment, two-period crossover design was used with 13 children entering the trial, and the observations of the peak expiratory flow, a measure of lung function where large values are associated with good responses, were taken. The following summary of the data is provided.

Group 1: formoterol → salbutamol ($n_1 = 7$)				
	Period 1	Period 2	Sum (1 + 2)	Difference(1 - 2)
mean	337.1	306.4	643.6	30.7
s.d.	53.8	64.7	114.3	33.0
Group 2: salbutamol → formoterol ($n_2 = 6$)				
	Period 1	Period 2	Sum (1 + 2)	Difference(1 - 2)
mean	283.3	345.8	629.2	-62.6
s.d.	105.4	70.9	174.0	44.7

a) Specify a model for peak expiratory flow which incorporates treatment, period and carryover effects.

b) Assess the carryover effect, and, if appropriate, investigate treatment differences. In each case specify the hypotheses of interest and illustrate the appropriateness of the test.

29. A and B are two hypnosis treatments given to insomniacs one week apart. The order of receiving the treatment is randomized between patients. The measured response is the number of hours sleep during the night. Data are given in the following table.

patient	**period 1**		**period 2**	
1	A	9	B	0
2	B	11	A	14
3	B	7	A	3
4	B	12	A	8
5	A	8	B	8
6	A	11	B	1
7	A	4	B	4
8	B	3	A	4
9	A	13	B	2
10	B	7	A	3
11	A	1	B	2
12	A	13	B	1
13	A	6	B	3
14	B	5	A	6
15	B	6	A	8
16	B	3	A	7

a) Calculate the mean for each treatment in each period and display the results graphically.

b) Assess the carryover effect.

c) If appropriate, assess the treatment and period effects.

(NB These data are available in on the course web pages)

30. Two ointments A and B have been widely used for the treatment of athlete's foot. In a recent report the following results were noted, where response indicated temporary relief from the outbreak.

	Response	No Response
Ointment A	174	96
Ointment B	149	121

a) Based on these results the report concluded that ointment A was more effective than ointment B. Use the Mantel-Haenszel test to verify this conclusion.

b) Further investigation into the source of the data revealed that the data had been pooled from two clinics. The results from individual clinics were:

	Ointment A		Ointment B	
Clinic	Response	No response	Response	No response
1	129	71	113	87
2	45	25	36	34

Reassess the evidence in the light of these additional facts.

©NRJF, University of Sheffield, 2012

31. (Artificial data from Ben Goldacre, 06/08/11).

Imagine a study was conducted to examine relationship between heavy drinking of alcohol and developing ling cancer, obtaing the following results:

	Cancer	No cancer
Drinker	366	2300
Non-Drinker	98	1856

a) Calculate the ration of the odds of developing cancer for drinkers to non-drinkers. What conclusions do you draw from this odds ratio?

b) It transpires that 330 of the drinkers developing cancer were smokers and 1100 of the drinkers who smoked did not, with corresponding figures for the non-drinkers of 47 and 156. Calculate the odds ratios separately for smokers and non-smokers. What conclusions do you draw?.

32. In a clinical trial of the use of a drug in twin pregnancies an obstetrician wishes to show a significant prolongation of pregnancy by use of the drug when compared to placebo. She assesses that the standard deviation of pregnancy length is 1.5 weeks, and considers a clinically significant increase in pregnancy length of 1 week to be appropriate.

iv) How many pregnancies should be observed to detect such a difference in a test with a 5% significance level and with 80% power?

v) It is thought that between 40 and 60 pregnancies will be observed to term during the course of the study. What range of

increases in length of pregnancy will the study have a reasonable chance (i.e. between 70% and 90%) of detecting?

33. Given below is an edited extract from an SPSS session analysing the results of a two period crossover trial to investigate the effects of two treatments A (standard) and B (new) for cirrhosis of the liver. The figures represent the maximal rate of urea synthesis over a short period and high values are desirable. Patients were randomly allocated to two groups: the 8 subjects in group 1 received treatment A in period 1 and B in period 2. Group 2 (13 subjects) received the treatments in the opposite order.

i) Specify a suitable model for these data which incorporates treatment, period and carryover effects.

ii) Assess the evidence that there is a carryover effect from period 1 to period 2.

iii) Do the data provide evidence that there is a difference in average response between periods 1 and 2?

iv) Assess whether the treatments differ in effect, taking into account the results of your assessments of carryover and period effects.

v) Repeat the statistical analysis in R

Extract from SPSS Analysis of Crossover Trial on Liver Treatment

Summarize

Case Summaries(a)

Patnum	Group	Period1	Period2	Sum1+2	PeriodDiff	TreatDiff
1.00	1.00	48.00	51.00	99.00	-3.00	-3.00
2.00	1.00	43.00	47.00	90.00	-4.00	-4.00
3.00	1.00	60.00	66.00	126.00	-6.00	-6.00
4.00	1.00	35.00	40.00	75.00	-5.00	-5.00
5.00	1.00	36.00	39.00	75.00	-3.00	-3.00
6.00	1.00	43.00	46.00	89.00	-3.00	-3.00
7.00	1.00	46.00	52.00	98.00	-6.00	-6.00
8.00	1.00	54.00	42.00	96.00	12.00	12.00
9.00	2.00	31.00	34.00	65.00	-3.00	3.00
10.00	2.00	51.00	40.00	91.00	11.00	-11.00
11.00	2.00	31.00	34.00	65.00	-3.00	3.00
12.00	2.00	43.00	36.00	79.00	7.00	-7.00
13.00	2.00	47.00	38.00	85.00	9.00	-9.00
14.00	2.00	29.00	32.00	61.00	-3.00	3.00
15.00	2.00	35.00	44.00	79.00	-9.00	9.00
16.00	2.00	58.00	50.00	108.00	8.00	-8.00
17.00	2.00	60.00	60.00	120.00	.00	.00
18.00	2.00	82.00	63.00	145.00	19.00	-19.00
19.00	2.00	51.00	50.00	101.00	1.00	-1.00
20.00	2.00	49.00	42.00	91.00	7.00	-7.00
21.00	2.00	47.00	43.00	90.00	4.00	-4.00

T-Test

Independent Samples Test

	Mean Difference	Std. Error Difference	t	Df	Sig. (2-tailed)
Sum1+2	2.7308	8.7046	.314	18.683	.757
PeriodDiff	-5.9423	2.9429	-2.019	17.646	.059
TreatDiff	1.4423	2.9429	.490	17.646	.630

Summarize

Case Summaries(a)

				Summ1+2	PeriodDiff	TreatDiff
GROUP	1.00	1		99.00	-3.00	-3.00
		2		90.00	-4.00	-4.00
		3		126.00	-6.00	-6.00
		4		75.00	-5.00	-5.00
		5		75.00	-3.00	-3.00
		6		89.00	-3.00	-3.00
		7		98.00	-6.00	-6.00
		8		96.00	12.00	12.00
		Total	N	8	8	8
			Mean	93.5000	-2.2500	-2.2500
			Std. Deviation	16.1688	5.8979	5.8979
	2.00	1		65.00	-3.00	3.00
		2		91.00	11.00	-11.00
		3		65.00	-3.00	3.00
		4		79.00	7.00	-7.00
		5		85.00	9.00	-9.00
		6		61.00	-3.00	3.00
		7		79.00	-9.00	9.00
		8		108.00	8.00	-8.00
		9		120.00	.00	.00
		10		145.00	19.00	-19.00
		11		101.00	1.00	-1.00
		12		91.00	7.00	-7.00
		13		90.00	4.00	-4.00
		Total	N	13	13	13
			Mean	90.7692	3.6923	-3.6923
			Std. Deviation	23.6684	7.4876	7.4876
	Total	N		21	21	21
		Mean		91.8095	1.4286	-3.1429
		Std. Deviation		20.7235	7.3863	6.8065

34. Several studies have considered the relationship between elevated blood glucose levels and occurrence of heart problems. The results of two similar studies are summarized below.

	Study 1			**Study 2**		
	heart problems			heart problems		
glucose level	yes	no		yes	no	
elevated	61	1284	1345	32	996	1028
not elevated	82	1930	2012	25	633	658
	143	3214	3357	57	1629	1686

ii) What can be concluded from these data regarding the influence of glucose on heart problems?

iii) Do you have any doubts on the validity of the form of analysis you have used?

35. A randomized, parallel group, placebo controlled trial was undertaken to assess the effect on children of a cream in reducing the pain associated with venepuncture at the induction of anaesthesia. A binary response of Y=0 for 'did not hurt' and Y=1 for 'hurt' was recorded for each of the 40 children who entered the trial, together with the treatment given (x1) and two covariates, sex (x2) and age (x3), which were thought might affect pain levels. A logistic model was fitted and the following details are available.

Factor	Reg. Coeff.	Standard Error of Coefficient
Intercept	2.058	1.917
x1: treatment (0 = placebo, 1 = cream)	-1.543	0.665
x2: sex (0 = boy, 1 = girl)	0.609	0.872
x3: age (years)	-0.461	0.214

i) Interpret and assess the treatment effect and also the effects of sex and age.

ii) Estimate the relative risk of hurting with the cream compared to the placebo.