Lab 1 - Clinical trials

Introduction

A doctor knocks on your office door one day and asks you to be involved in a clinical trial she is running. The trial is investigating the effect of giving vitamin D supplements to pregnant women, and in particular how that will affect their newborn babies' serum calcium levels 1 week after birth. The trial will be run with two groups, where the pregnant women volunteers will be randomly allocated to receive a Vitamin D supplement or a placebo. The only covariate she plans to obtain is whether each baby was breast fed or bottle fed.

Step 1 - Sample size estimation

The doctor is not sure what the treatment effect will be exactly, nor what the standard deviation will be. Therefore she asks you to compute the number of patients required under the following assumptions. Carry out the calculations in R. Don't forget to interpret your results.
Hint: You might benefit from defining a function.

1. The difference is 0.25, the standard deviation is 0.99, the significance level is 0.05 and the power is 0.9.
2. As in 1. above but the power is reduced to 0.8.
3. As in 1. above but the standard deviation increases to 1.5.
4. As in 1. above but the difference reduces to 0.2.
5. The doctor then wonders whether rather than a continuous response a binary response would be better, because her colleagues use the following binary scale:
   - hypocalcaemia - serum calcium $< 7.5mg/100ml$.
   - non-hypocalcaemia - serum calcium $\geq 7.5mg/100ml$. Modify the function above to compute the sample size if the probability of hypocalcaemia is 0.15 on the vitamin D treatment and 0.25 on the control. Assume that the power is 0.95 and the significance level is 0.05.

Step 2 - Randomisation

Following your sample size calculations she decides to go with your first estimate in (a), and try and recruit 330 patients per group, giving 660 in total. Each patient must be randomised into the two treatment groups, which here we call treatments (A, B).

1. Initially, the doctor wishes to use a simple randomisation scheme. Use the `sample()` function to randomly draw 660 times from the pair $\{A, B\}$.
2. However, the doctor is now worried about this approach. What is the main problem with this simple randomisation scheme? How can you check if the problem has occurred, and if it has how extreme it is?

3. The doctor wants to know, under this simple randomisation scheme, what the probability is that the 2 treatments differ in the number of patients by more than 10.

4. The doctor decides that the problem illustrated in 1. to 3. is too much of a risk, and needs to be prevented. This can be done by constructing a random permuted block design, with 6 blocks of size 110. This guarantees that equal numbers of patients are allocated to each treatment after 110, 220, 330, 440, 550 and 660. patients have entered the trial. Check that you indeed have 330 patients enrolled on treatment A and 330 on treatment B.

5. All results you get will be random and change slightly every time you run your code. How can you achieve reproducible results that remain constant when using R functions with a random component?

Step 3 - Analysis of the data

The doctor conducts the trial and returns a year later with the data for you to analyse. Unfortunately due to patient drop out there are only 620 patients with complete data to analyse, as 40 patients have withdrawn from the study. The following results were obtained from the babies one week after their birth, and have been split by the single binary covariate, which is feeding type (artificial (bottle) or breast feeding). Note, in the end the treatment and control are not given to the same numbers of patients because medically it was not appropriate for some patients to receive the vitamin D supplement.

Table 1: Results of vitamin D study.

	Artificially fed		**Breast fed**	
	Vitamin D	**Control**	**Vitamin D**	**Control**
No. of infants	169.00	285.00	64.00	102.00
Mean calcium	9.20	8.78	9.79	9.64
SD	1.10	1.28	1.17	1.26

The data from this trial are available in `vitaminD.csv` and contain the following data on 620 babies:

- `calcium` - serum calcium level in mg per 100ml.
- `feeding` - artificially (bottle) fed (denoted `Art`) or breast fed (denoted `Br`).
- `treatment` - Vitamin D supplement (denoted VitD) or placebo (denoted Control).

The data can be read in using the following commands.

```
dat <- read.csv("vitaminD.csv")
head(dat)
```

```
##   calcium feeding treatment
## 1    9.12     Art      VitD
## 2    9.92     Art      VitD
## 3    9.99     Art      VitD
## 4    7.52     Art      VitD
## 5   10.54     Art      VitD
## 6   11.30     Art      VitD
```

1. Carry out separate analyses for the two subgroups (artificially fed and breast fed) to determine if there is a significant difference in serum calcium levels between vitamin D (treatment) and placebo groups within each subgroup. What do you conclude?

2. Use the `interaction.plot()`, `lm()` and `confint()` functions to establish whether there is any evidence of a difference in the size of the vitamin D effect on serum calcium levels in the two subgroups. Provide an appropriate interval estimate of the size of the vitamin D effect. Don't forget to interpret your results!