Worksheet 1 MAS367
Sample Sizes, Allocation, Bias and Randomization

1.
The Beck depression inventory score is often used as the outcome measure in trials that assess treatments for depression. It has standard deviation 8 (no units). Suppose that in a trial comparing two treatments there is interest in detecting a difference of 3.5 between the mean Beck depression inventory scores in the two groups.

a) If the groups are of equal size and the power is 90% at the 5% level of significance, how many patients need to be recruited in total?

b) How would this number change if the power were reduced to 80%?

c) I can recruit 100 patients: what is the largest power I can have to detect a mean difference of 3.5?

d) If the value used for the standard deviation for the Beck depression inventory score should have been 10% larger, what would have been the effect on the numbers in a) and b)?

2.
Patients who contract bronchiolitis, a respiratory disease of infants, often suffer from wheezing for some time after they have been ostensibly cured. In a trial to see if an inhaled steroid, budesonide, can alleviate this after-effect, patients were randomized to receive budesonide or a placebo. The outcome was the presence of wheeze one year after the diagnosis of bronchiolitis.

a) It is thought that the proportion of children with wheeze at one year post-diagnosis is 0.65. There is interest in reducing this to 0.25. How many infants would have to be randomized to each of two equal-sized groups in order to have 80% power to detect this difference at the 5% significance level?
b) Suppose that the initial estimate of 0.65 is too high, and it should be 0.55. How many patients would be needed to detect a change to 0.15 (i.e. the same change in proportion) with 80% power?
c) If the initial estimate were too low and 0.75 was a more reasonable choice, how many patients would be needed to detect a change to 0.35 (again the same change, 0.4, in proportions) with 80% power? Comment on your answer.

3.
The method for determining the sample size for a trial with binary outcome uses the approximation:

$$\arcsin(\sqrt{p}) \sim N(\arcsin(\sqrt{p}), 1/(4n)),$$

where r successes have been observed in n trials and $p=r/n$: the population probability of success is p. Although no longer a standard method, this result can be used as a basis for the analysis of binary data. If two groups are to be compared, with observed proportions of successes from n_1 and n_2 trials of, respectively, p_1 and p_2, then show that a statistic testing the null hypothesis that the two population proportions are equal is:

$$T = \frac{\arcsin(\sqrt{p_1}) - \arcsin(\sqrt{p_2})}{\frac{1}{2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

To which distribution would you compare this statistic in order to perform the test? Using the result

$$\arcsin(\sqrt{p}) \approx \arcsin(\sqrt{p}) + (p - p)\frac{d}{du}\arcsin(\sqrt{u})\Big|_p ,$$

show that, approximately,

$$T = \frac{p_1 - p_2}{\sqrt{p(1-p)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

and relate this to one of the more common ways of testing this hypothesis.

4. Ten patients are randomly allocated to treatments A and B in a clinical trial. Unequal allocation is used, with Pr(Allocate to A) = 0.6. What is the probability that two or fewer patients are allocated to treatment B?

5. Suppose you were designing the trial outlined in question 1 and had decided to opt for 80% power. However it was thought appropriate that one treatment group should be twice the size of the other. What would be the sizes of the two groups? Suppose that allocation was to be made using random permuted blocks of length 6. How many blocks are there? What is the disadvantage of using a single block size? How many blocks of length 9 are there?