Homework 3, due: October 1

1. In a small clinical trial, permuted block randomization was used to assign one of two antiviral treatments (A or B) to eight patients with HIV disease. Four blocks of size 2 were chosen; each with one A and one B chosen at random. The endpoint of interest was change from baseline in log CD4 counts after eight weeks on treatment. The results of this trial are summarized as follows:

	Block 1	
treatment	A	B
response	.3	.9
	Block 2	
treatment	B	A
response	1.0	.2
	Block 3	
treatment	B	A
response	.5	.7
	Block 4	
treatment	A	B
response	.1	.2

The test statistic used to evaluate whether treatment B has better response than treatment A is based on $\bar{Y}_B - \bar{Y}_A$, where $\bar{Y}_B$ is the average response among patients assigned to treatment B and $\bar{Y}_A$ is the average response for treatment A. We assume the sharp null hypothesis and reject in favor of treatment B if the test statistic is sufficiently large (i.e., one-sided test).

(a) Find the distribution of the test statistic that is induced by the permuted block randomization under the sharp null hypothesis (i.e., design based inference).

(b) What is the p-value for the clinical trial described above?

(c) What are the advantages and disadvantages of this kind of design as compared to using simple randomization where each treatment is assigned with probability one half independently of each other.

2. Let the random variable $X \sim b(n, \pi)$. The sample proportion is denoted by $p = X/n$.

(a) Prove that
$$\left(\frac{n}{n-1}\right) p(1-p)$$
is an unbiased estimator for $\pi(1 - \pi)$, $n \geq 2$.

Consider data from N studies using the same treatment regimen. We observe $(X_i, n_i), i = 1, \ldots, N$, where n_i denotes the number of patients in the i-th study and X_i denotes the number of these patients that respond to treatment. It will be assumed that these data are from a hierarchical model where $(X_i, n_i, \pi_i), i = 1, \ldots, N$ are assumed to be iid random vectors. The π_i are assumed to be study-specific response rates which themselves represent an iid sample from some distribution. Conditional on π_i and n_i, X_i is assumed to follow a binomial distribution. Specifically,

$$X_i | n_i, \pi_i \sim b(n_i, \pi_i).$$

(b) Find an unbiased estimator for the expectation
$$E\{\pi_i(1 - \pi_i)\}$$
and prove that it is unbiased.

3. Consider the following hierarchical model which may be appropriate in conceptualizing results from different studies using the same regimen but where the response of interest is a continuous measurement. The data from N such studies are defined as

$$\{(Y_{ij}, j = 1, \ldots, n_i), i = 1, \ldots, N\},$$

where Y_{ij} denotes the response from the j-th individual (among n_i individuals) in the i-th study. The underlying hierarchical model assumes that there exists N iid random vectors

$$\{(Y_{ij}, j = 1, \ldots, n_i), \mu_i, \sigma_i^2\},$$

2

for $i = 1, \ldots, N$. The random vectors $(\mu_i, \sigma_i^2), i = 1, \ldots, N$ are themselves assumed to be iid unobserved random effects corresponding to the study-specific mean and variance of response which are allowed to vary according to some distribution. In particular, we will denote the expected value of the study-specific means by

$$\mu^* = E(\mu_i)$$

and the variance of the study-specific means by

$$\sigma_{*\mu}^2 = var(\mu_i).$$

<u>Conditional</u> on (μ_i, σ_i^2, n_i), we will assume that the data from the i-th study represent n_i iid observations with mean μ_i and variance σ_i^2. Specifically, we assume that for the i-th study

$$E(Y_{ij}|\mu_i, \sigma_i^2, n_i) = \mu_i, \ var(Y_{ij}|\mu_i, \sigma_i^2, n_i) = \sigma_i^2, j = 1, \ldots, n_i.$$

The major reason for this conceptualization is to determine whether study to study variation in the mean response is substantial. Specifically, we want to estimate $\sigma_{*\mu}^2 = var(\mu_i)$. In the table below, we summarize results from ten studies. For each study $i = 1, \ldots, 10$, we give the sample size n_i, the sample mean response

$$\bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i},$$

and the sample variance

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{n_i - 1}.$$

study number	sample mean	sample variance	sample size
1	59	2475	52
2	23	1971	48
3	41	2009	47
4	28	1771	163
5	40	2484	135
6	31	2139	150
7	34	2275	37
8	20	1411	111
9	30	2379	62
10	25	1622	100

3

(a) In terms of $(\bar{Y}_i, s_i^2, n_i), i = 1, \ldots, N$ derive an unbiased estimator for $\sigma_{*\mu}^2$.

(b) Derive this estimate for the data provided in the table above.

4. Suppose the goal of a clinical trial is to estimate the difference in mean response between two treatments assumed to have the same variance. It costs 150 dollars to treat a patient on regimen A and 100 dollars to treat on regimen B. An investigator has 30,000 dollars to spend on this trial.

(a) What is the optimal allocation of patients to treatments A and B subject to the budgetary constraints of the investigator? That is, what allocation of patients would result in the smallest variance of the estimator of treatment difference.

(b) If the investigator insisted on equal allocation to both treatments, then how much more money would need to be spent to get the same degree of precision (i.e. variance of estimator) as the allocation in part (a).