Due: November 14

1. A three arm trial is about to be conducted where two new treatments, say, treatment 2 and treatment 1, are to be compared to the standard treatment 0. The response is a continuous measurement which we denote by Y and, absent any additional information, we will assume that the treatment-specific variances are all equal for design purposes. The null hypothesis of treatment equality

$$H_0 : \mu_0 = \mu_1 = \mu_2,$$

where μ_0, μ_1 and μ_2 denote the treatment-specific mean responses for treatments 0, 1, 2 respectively is to be tested against the alternative that some treatment difference exists. The investigators of the trial come to you for advise on sample size considerations. They are mostly interested in comparing the new treatments to the standard treatment and not particularly interested in comparing the two new treatments to each other. After quizzing

the investigators about important differences they tell you that the average response on the standard treatment is about 100 with a standard deviation of 40. They also tell you that if a new treatment can increase the mean response to 115 or more, then this would be an important treatment difference to detect. Their belief is that both new treatments will yield clinically important differences and they insist that the study be powered with that alternative in mind. That is, $\mu_{0A} = 100$, $\mu_{1A} = 115$ and $\mu_{2A} = 115$.

a. What test would you use to test H_0?

b. If you were to randomize with equal probability to all three arms, then how large a sample size is necessary so that a test at the .05 level of significance would have 90% power to detect the clinically important difference given by the investigators?

c. In contrast, what is the sample size necessary so that a test at the .05 level of significance would be guaranteed to have 90% power if any two treatments had a mean difference of 15 or more?

Regardless of your answers above, a clinical trial was conducted and the results are summarized in the table below

treatment	sample average	sample standard deviation	sample size
0	130.2	62.3	210
1	125.8	40.6	202
2	149.5	43.7	195

If we denote by Y_{ij} the response for the i-th individual in treatment j, where $j = 0, 1, 2$ and $i = 1, \ldots, n_j$, then the treatment-specific sample average is

$$\bar{Y}_j = \frac{\sum_{i=1}^{n_j} Y_{ij}}{n_j}, j = 0, 1, 2$$

and the sample standard deviation is

$$s_j = \sqrt{\frac{\sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2}{n_j - 1}}, j = 0, 1, 2.$$

Looking at the data, it seems that the assumption that the treatment-specific variances are all the same may not be so reasonable. Keep this in mind when constructing the test statistics for parts d. and e.

d. What is the p-value for the data above? (To get credit you must show what test statistic you used) What are your conclusions regarding the null hypothesis?

e. It was always intended to make pairwise comparisons between treatment 1 versus treatment 0 and treatment 2 versus treatment 0. Accounting for multiple comparisons what conclusions would you draw regarding these comparisons? (Show all your calculations including relevant p-values and discuss how you reached your conclusion).

2. There are several treatments that have been proven to be effective in increasing CD4 counts for patients with HIV disease. However, it is well known that such treatments over time become resistant to the HIV virus and lose their effectiveness. Consequently, there is still room for new treatments to be considered that are equivalent in effectiveness to the current treatments. A pharmaceutical company has developed a new drug that they believe is similar in effectiveness to the drugs currently used and decide to conduct an equivalency (non-inferiority) trial. Because the distribution of CD4 counts is right skewed it is more convenient to work with the logarithm (natural) of CD4 counts as this will have a distribution that is closer to a normal distribution. Patients with advanced disease will be recruited for this clinical trial. Based on previous data the median CD4 counts is about 200 for such patients which translates to median log CD4 counts of 5.3. Also from past experience the effective treatment can increase CD4 counts by 50% over a period of one year or an increase in log CD4 counts of .41. This would correspond to increasing the median CD4 counts from 200 to 300. Based on these considerations the primary outcome Y is the difference in log CD4 count for each patient one year after treatment is initiated.

You are asked to help design this non-inferiority trial. A randomized clinical trial will be conducted where patients will be randomized to either a current effective treatment (positive control) versus the new treatment. As stated above, the primary endpoint will be the difference in the log CD4 count one year after treatment. For planning purposes we assume that the mean log CD4 count difference for patients on positive control will be .41 with a standard deviation of .25. The FDA has decided that equivalence can be claimed if the company can prove that the response in CD4 counts is within 10% of the current effective drug. On the log scale this would correspond to a mean log CD4 count difference for patients on the new drug

of .315 or greater or a tolerable difference of .095 on the log scale. We also assume that the standard deviation of response (log CD4 count difference) will be the same as the other treatment; namely, .25. Let μ_2 denote the mean response for patients if they were treated with positive control and μ_1 the mean response for patients if they were assigned the new drug.

a. In terms of μ_1 and μ_2 state the null and alternative hypotheses that will be tested in this equivalency trial.

b. Let n_1, $\bar{Y}_1$, and s_1 denote the sample size, sample average and sample standard deviation computed from the patients in the trial that would be randomized to the new treatment and similarly n_2, $\bar{Y}_2$, and s_2 for patients in the trial that would be randomized to positive control. In terms of these what would you use to compute the test statistic? At the .05 level of significance when would you reject the null hypothesis?

c. In designing your trial the investigators want to do some additional testing on resistance for the new drug. Therefore in order to get more experience they decide to randomize twice as many patients on the new drug versus positive control; that is, n_1 is twice n_2. If you wanted 90% or greater power to be able to declare equivalency if, in truth, your new treatment had a population mean response equal to or better than that of the positive control then how large a sample size will be needed?

d. Regardless of your answer to part (c.) a clinical trial was conducted with 150 patients; 100 randomized to the new treatment and 50 to the positive control. The sample average response and sample standard deviation on the new treatment were .40 and .22 respectively and the sample average response and sample standard deviation on the positive control were .39 and .26 respectively. Compute the value of your test statistic and whether or not you will be able to conclude that your new treatment is equivalent to control.

3. Let T denote a positive continuous random variable with hazard rate

$$\lambda(t) = \gamma t, \ \gamma > 0,$$

where time t is in units of years. In terms of γ

(a) What is the survival distribution $S(t) = P(T \geq t)$?

(b) What is the median survival time?

4. In the table below are the results of a small clinical trial comparing two treatments. The endpoint of interest is time to death (in years) which may be right censored.

Treatment A		Treatment B	
time to failure or censoring	failure indicator	time to failure or censoring	failure indicator
1.43	1	2.30	1
0.95	0	1.98	1
3.89	0	4.03	1
2.25	1	3.30	0
0.73	1	4.20	0
3.24	1	1.58	1

Using only a calculator and showing all your work (i.e. risk set calculations)

(a) Combining the two treatments together compute and plot the Kaplan-Meier estimator of the survival curve.

(b) Test the null hypothesis of equality of the treatment specific survival distributions using a two-sided logrank test. Compute the standardized test statistic and the p-value.

5. In designing a clinical trial the investigators tell you that the standard treatment used for the treatment of multiple myeoloma results in a median survival of 5.0 years and that the survival distribution is well approximated

by an exponential distribution. Two new promising treatments are to be tested against the standard treatment in a three-arm randomized clinical trial with equal allocation to all three arms. After quizzing the investigators they decide that they want at least 90% power to detect any difference between the standard treatment and either new treatment if either new treatment can increase the median survival to 7.0 years or greater using a three sample logrank test at the .05 level of significance. Again after quizzing them further they tell you that they believe that they can accrue 200 patients per year into the study.

- Assume that all three survival distributions are exponentially distributed and that you compute your power calculations under the least favorable configuration; namely, where the median survival for the worst treatment is 5.0 years, the median survival for the best treatment is 7.0 years and for the least favorable configuration the median survival of the third treatment is the geometric mean of 5 and 7 years; i.e., $\sqrt{(5 \times 7)}$ or 5.92 years. You can use the formula given at the end of page 137 of Chapter 8 of the notes to determine the requisite number of deaths for the least favorable configuration. Assuming no loss due to patient drop-out or competing risks, compute and plot all the possible combinations of accrual period and total length of study that would meet the goals of the study.

6. In class we discussed CALGB 8541 (the three arm randomized clinical trial of women with stage II node positive breast cancer) that we've been using for illustration. I have given you access to a subset of these data on the class web-site. The dataset is in a file "calrisk.dat". I have only included patients that were randomized to either treatment 2 (low dose CAF) or treatment 3 (standard dose CAF). There are four variables in this file. Variable 1 denotes days on study, variable 2 denotes failure indicator (1=death, 0=censored), variable 3 is treatment indicator (1 if they received treatment 3 and 0 if they received treatment 2), and finally variable 4 is a risk indicator (coded as 0 or 1).

I want you to analyze the data looking at the relationship of treatment and the risk indicator. I want you to consider two groups, women with risk indicator equal to 1, and women with risk indicator equal to 0. Using SAS or R

(a) First I want you to compare the survival distribution of these two groups (risk indicator 1 versus risk indicator 0), by plotting the survival curves and computing the two-sample logrank test. What are your conclusions?

(b) I want you to analyze the treatment comparisons separately within the strata for women with risk indicator 1 and for women with risk indicator 0. That is, for each stratum separately, plot the treatment specific survival curves and compute the logrank test. What are your conclusions?

Note: The dataset 'calrisk.dat' has a header. Therefore if you are using R you can read in the data using read.table('calrisk.dat',header=TRUE). For example, if you use
x=read.table('calrisk.dat',header=TRUE),
then the first three variables are x$V1, x$V2, x$V3, and the fourth variable is x$risk.