Statistics 641, Fall 2011
Homework #4
Answers

1. Suppose that we have a binomial response with expected failure probabilities $\pi_0 = 0.2$ and $\pi_1 = 0.1$ for treatments 0 (control) and 1 (experimental) respectively. Calculate the required sample sizes to achieve 90% power (type II error rate of 10%) at level $\alpha = .01$ assuming

   - Equal allocation to groups 0 and 1

     $$N = \frac{\bar{\pi}(1-\bar{\pi})(Z_{1-\alpha/2} + Z_{1-\beta})^2}{\xi_0\xi_1(\pi_0 - \pi_1)^2}$$

     and with equal allocation $\bar{\pi} = (\pi_0 + \pi_1)/2 = 0.15$.

     $$\frac{4 \times .15(1-.15)(2.58+1.28)^2}{(0.2-0.10)^2} = 759.9$$

     so $N = 760$.

   - 2:1 randomization to groups 0, 1 respectively

     $\bar{\pi} = (2\pi_0 + \pi_1)/3 = 0.167$ and

     $$\frac{9 \times .167(1-.167)(2.58+1.28)^2}{2(0.2-0.10)^2} = 931.2$$

     so $N = 932$.

   - 2:1 randomization to groups 1, 0 respectively

     $\bar{\pi} = (\pi_0 + 2\pi_1)/3 = 0.133$ and

     $$\frac{9 \times .133(1-.133)(2.58+1.28)^2}{2(0.2-0.10)^2} = 774.8$$

     so $N = 775$.

     Note that N is sensitive to $\bar{\pi}$, which in turn depends on ξ_j. When more subjects are allocated to the treatment with the lower rate, fewer subjects are required to detect a fixed difference in probabilities.

2. Heart patients have a greater risk of a second heart attack (MI) immediately following the first MI, then they do later on. Suppose (simplistically) that with standard therapy, the hazard rate λ is constant .12/year during the first 6 months following MI and constant .05/year thereafter. Suppose further that a new treatment is expected to reduce these rates by 25%. We wish to perform a study of patients enrolled immediately following an MI with (uniform) enrollment and followup of either:

(a) 1.5 year enrollment, 4 year followup

(b) 2 year enrollment, 3.5 year followup

Compute the required sample sizes for each of these two designs assuming equal numbers of patients in each treatment group. (*Hint: First find the required number of events, then find the probability of an event by writing the cumulative hazard function as piecewise linear function and calculating the survivor function.*)

First, assuming $\alpha = 0.05$ and $\beta = .1$, according to Schoenfeld's formula we have that the required number of events is

$$D = \frac{4(1.96 + 1.28)^2}{\log(.75)^2} = 508$$

The average hazard function is (t in years)

$$\bar{\lambda}(t) = \begin{cases} (0.12 + 0.12 \times .75) = 0.105 & \text{if } t \leq 0.5 \\ (0.05 + 0.05 \times .75) = 0.04375 & \text{if } t > 0.5 \end{cases}$$

When $t \leq 0.5$, the cumulative hazard function is:

$$\bar{\Lambda}(t) = \int_0^t \bar{\lambda}(t) ds = \int_0^t 0.105 ds = 0.105t.$$

For $t > .5$, the cumulative hazard function is:

$$\begin{aligned} \bar{\Lambda}(t) &= \int_0^t \bar{\lambda}(u) du = \int_0^{0.5} \bar{\lambda}(u) du + \int_{0.5}^t \bar{\lambda}(u) du \\ &= 0.105 \times 0.5 + 0.04375 \times (t - 0.5) = 0.030625 + 0.04375t \end{aligned}$$

Hence, the mean survivor function is

$$\bar{S}(t) = \begin{cases} e^{-0.105t} & \text{if } t \leq 0.5 \\ e^{-0.030625 - 0.04375t} & \text{if } t > 0.5 \end{cases}$$

We need to compute the probability of failure by integrating over the censoring distribution which is uniform on an interval $(F - R, F)$ where F is the maximum follow-up time, and R is the length of the enrollment period.

$$\Pr\{\text{failure}\} = \frac{1}{R} \int_{F-R}^F 1 - \bar{S}(u) du$$

Since for both scenarios above, $F - R > 0.5$, the integrand only needs to be evaluated for $u > 0.5$ so we don't need to break it into pieces.

$$\begin{aligned} \Pr\{\text{failure}\} &= \frac{1}{R} \int_{F-R}^F 1 - e^{-0.030625 - 0.04375u} du \\ &= 1 + \frac{e^{-0.030625}}{0.04375R} e^{-0.04375u} \Big|_{F-R}^F \\ &= 1 + \frac{e^{-0.030625}}{0.04375R} \left(e^{-0.04375F} - e^{-0.04375(F-R)} \right) \end{aligned}$$

Plugging in $F = 4$, $R = 1.5$, we have

$$\Pr\{\text{failure}\} = 0.15855$$

so the required number of subjects is

$$\frac{507}{0.15855} = 3198$$

Plugging in $F = 3.5$, $R = 2$, we have

$$\Pr\{\text{failure}\} = 0.13036$$

so the required number of subjects is

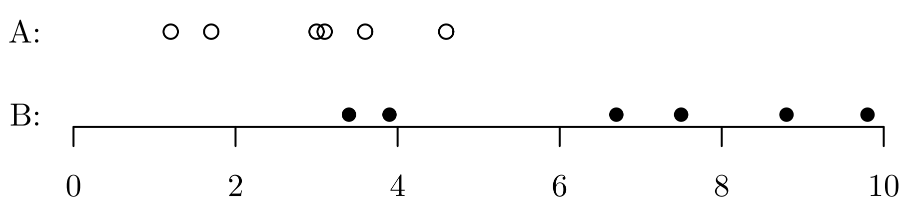
$$\frac{507}{0.13036} = 3889$$

3. Suppose that we observe the following continuous responses for groups A and B.

A: 3.0, 1.7, 4.6, 3.6, 1.2, 3.1
B: 3.9, 6.7, 7.5, 3.4, 8.8, 9.8

Let $\hat{\mu}$ be then difference in means, $\bar{X}_B - \bar{X}_A$. Compute a one-sided p-value under the randomization distribution generated by the random allocation rule ($N_A = N_B$). (Note that you don't need to generate the entire randomization distribution for $\hat{\mu}$.)

First, we have $\hat{\mu} = 6.683 - 2.867 = 3.817$. A simple plot of the data shows that most of the observations in group B are well to the right of the observations in group A:



There are only small number of treatment allocations that produce a more extreme mean difference than the one observed, specifically, we need only consider allocations such that the mean in group B is at least as large as the observed mean, 6.683. These are enumerated below:

Allocated to B	mean
3.9, 4.6, 6.7, 7.5, 8.8, 9.8	6.883
3.6, 4.6, 6.7, 7.5, 8.8, 9.8	6.833
3.4, 4.6, 6.7, 7.5, 8.8, 9.8	6.800
3.1, 4.6, 6.7, 7.5, 8.8, 9.8	6.750
3.0, 4.6, 6.7, 7.5, 8.8, 9.8	6.733
3.6, 3.9, 6.7, 7.5, 8.8, 9.8	6.717
3.4, 3.9, 6.7, 7.5, 8.8, 9.8	6.683

Therefore, there are 7 allocations yielding a mean difference at least as extreme as the one observed. There are a total of

$$\binom{12}{6} = 924$$

possible allocations, so the one-sided p-value is $7/924 = .0076$.

4. Suppose that we have two treatments with a 1:1 permuted block randomization with blocks of size 6 (i.e., within each block of 6 we randomly allocate 3 to each treatment). We enroll 12 subjects and in the two blocks we observe the following summary tables:

Group	D	A	Total
1	3	0	3
2	0	3	3
	3	3	6

Group	D	A	Total
1	2	1	3
2	0	3	3
	2	4	3

(a) Calculate the size of the reference set (all possible allocations of treatments to subjects).

We have two blocks of size 6. Within each block we allocate 3 to group 1 and the remaining 3 to group 2. There are $\binom{6}{3} = 20$ ways of doing this within each block. Thus the reference set has $20 \times 20 = 400$ allocations.

(b) If x_j is the number of deaths in group 1 for block j, find the sample space for $U(0) = \sum_j x_j - E[x_j]$ and corresponding sampling probabilities under the randomization distribution. (*Hint: x_j has a hypergeometric distribution*).

In x_j be the entry in the upper left corner of block j, $j = 1, 2$.

x_1 takes 4 possible values: 0,1,2,3. $\quad E[x_1] = 3 \times 3/6 = 1.5$
x_2 takes 3 possible values: 0,1,2. $\quad E[x_1] = 2 \times 3/6 = 1$

$U(0)$ takes values according to the following table:

$x_2 \backslash x_1$:	0	1	2	3
0	-2.5	-1.5	-0.5	0.5
1	-1.5	-0.5	0.5	1.5
2	-0.5	0.5	1.5	2.5

Each x_j has a hypergeometric distribution, so if m_j is the total number of deaths in block j (3 or 2), then the x_j has probability

$$\Pr\{x_j = x\} = \frac{\binom{3}{x}\binom{3}{m_j - x}}{\binom{6}{m_j}}$$

$$\Pr\{x_1 = 0\} = \Pr\{x_1 = 3\} = \frac{\binom{3}{0}\binom{3}{3}}{\binom{6}{3}} = \tfrac{1}{20},$$

$$\Pr\{x_1 = 1\} = \Pr\{x_1 = 2\} = \frac{\binom{3}{1}\binom{3}{2}}{\binom{6}{3}} = \tfrac{9}{20}$$

Similarly,

$$\Pr\{x_2 = 0\} = \Pr\{x_2 = 2\} = \frac{\binom{3}{0}\binom{3}{2}}{\binom{6}{2}} = \tfrac{1}{5}, \qquad \Pr\{x_1 = 1\} = \frac{\binom{3}{1}\binom{3}{1}}{\binom{6}{2}} = \tfrac{3}{5}$$

Thus,

$$\begin{aligned}
\Pr\{U(0) = -2.5\} &= \Pr\{x_1 = x_2 = 0\} = \frac{1}{20} \times \frac{1}{5} = \frac{1}{100} \\
\Pr\{U(0) = -1.5\} &= \Pr\{x_1 = 1, x_2 = 0\} + \Pr\{x_1 = 0, x_2 = 1\} = \frac{9}{20} \times \frac{1}{5} + \frac{1}{20} \times \frac{3}{5} = \frac{12}{100} \\
\Pr\{U(0) = -0.5\} &= \Pr\{x_1 = 2, x_2 = 0\} + \Pr\{x_1 = 1, x_2 = 1\} + \Pr\{x_1 = 2, x_2 = 0\} \\
&= \frac{9}{20} \times \frac{1}{5} + \frac{9}{20} \times \frac{3}{5} + \frac{1}{20} \times \frac{1}{5} = \frac{37}{100}
\end{aligned}$$

By symmetry,

$$\Pr\{U(0) = 2.5\} = \frac{1}{100} \qquad \Pr\{U(0) = 1.5\} = \frac{12}{100} \qquad \Pr\{U(0) = 0.5\} = \frac{37}{100}$$

Hence $U(0)$ takes values $\{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$ with probabilities $\{1/100, 12/100, 37/100, 37/100, 12/100, 1/100\}$.

(c) Calculate the one-sided randomization p-value for the observed data.

The observed value of $U(0)$ is $(3 - 1.5) + (2 - 1) = 2.5$. The one-sided randomization p-value is $\Pr\{U(0) \geq 2.5\} = 0.01$ from the distribution in part (b).

(Note that the stratified chi-square statistic (Mantel-Haenszel) is $2.5^2/.708 = 8.824$ *which corresponds to a large-sample p-value of 0.003.)*