Biost/Stat 578A
Emerson, Spr 13
April 22, 2013

Homework #2: Written problems to be submitted electronically by 11:59 pm on Sunday, April 28.

The following notation applies to all problems.

Suppose $Y_1, \ldots, Y_n$ are independent random variables measuring treatment outcome in a randomized, placebo controlled random trial. Let X_i be an indicator that the ith patient received the experimental treatment, and let W_i be an indicator that the ith patient has advanced disease. Further, let π represent the proportion of potentially eligible patients with advanced disease.

We presume the following conditional distributions:

- Without loss of generality, we presume the conditional distribution $Y_i|X_i = 0, W_i = 0 \sim (0, 1)$. That is, subjects without advanced disease who receive placebo have treatment outcomes following a distribution that has mean 0 and variance 1, but is otherwise unspecified.
- Subjects with advanced disease who are treated with placebo have treatment outcomes distributed according to $Y_i|X_i = 0, W_i = 1 \sim (\omega, \tau^2)$.
- Subjects without advanced disease who receive the experimental treatment have treatment outcomes distributed according to $Y_i|X_i = 1, W_i = 0 \sim (\delta, \psi^2)$.
- Subjects with advanced disease who receive the experimental treatment have treatment outcomes distributed according to $Y_i|X_i = 1, W_i = 1 \sim (\nu, \xi^2)$.

In all problems, we are interested in contrasting the mean treatment outcome across groups who receive the experimental treatment or placebo, and we will estimate the treatment effect using sample means.

In each problem, provide explicit formulae whenever possible. When evaluation of the results depends too heavily on the exact choice of parameters, you can illustrate that dependence in particular settings. Settings of greatest interest might include:

- The simplifying assumption of $\tau^2 = \psi^2 = \xi^2 = 1$.
- The case of a subgroup for which there is no treatment effect, so $\omega = \nu$. You might want to then consider how the choice of π and ω affects your results.
- A parameterization of the subgroup effect in which $\omega = 0$ and you consider a range of $\nu/\delta \in (0, 1)$ for a range of π.
- Cases in which $min\,(n\pi, n(1-\pi)) = 20$, 50, or 100.

1. Consider first the setting of a completely randomized study in which W_i is an unmeasured latent variable: We accrue n subjects to the study, and we independently randomize each patient with probability 0.5 to receive the experimental treatment.
   a. Find expressions for $\mu_x = E[Y_i|X_i = x]$ and $\sigma_x^2 = Var(Y_i|X_i = x)$ for $x = 0, 1$.
   b. Let $\theta = \mu_1 - \mu_0$ and $\hat{\theta}$ be the distribution-free estimator of θ. Find an approximate distribution for $\hat{\theta}$ as a function of ω, δ, ν, τ, ψ, ξ, π, and n.
2. Consider now the setting of a block randomized study in which W_i is an unmeasured latent variable. We accrue n subjects to the study, and we ensure that exactly $n/2$ subjects are randomly chosen to receive the experimental treatment.
   a. Find expressions for $\mu_x = E[Y_i|X_i = x]$ and $\sigma_x^2 = Var(Y_i|X_i = x)$ for $x = 0, 1$.
   b. Let $\theta = \mu_1 - \mu_0$ and $\hat{\theta}$ be the distribution-free estimator of θ. Find an approximate distribution for $\hat{\theta}$ as a function of ω, δ, ν, τ, ψ, ξ, π, and n.
   c. How does efficiency of the study in this problem compare to that in problem 1? In particular,

consider the case where $\omega = 0$, $\delta = \nu$, $\tau^2 = 1$, and $\psi^2 = \xi^2$. (This latter comparison isolates the efficiency associated with block randomization as opposed to complete randomization.)

3. Consider now the setting of a block randomized study in which W_i is a measured variable used in an adjusted analysis (but we do not stratify randomization using W_i). We accrue n subjects to the study, and we ensure that exactly $n/2$ subjects are randomly chosen to receive the experimental treatment.
   a. Find expressions for $\mu_{xw} = E[Y_i|X_i = x, W_i = w]$ and $\sigma^2_{xw} = Var(Y_i|X_i = x, W_i = w)$ for $x = 0, 1$ and $w = 0, 1$.
   b. Let $\theta = (1 - \pi)(\mu_{10} - \mu_{00}) + \pi(\mu_{11} - \mu_{01})$ and $\hat{\theta}$ be the distribution-free estimator of θ. Find an approximate distribution for $\hat{\theta}$ as a function of ω, δ, ν, τ, ψ, ξ, π, and n.
   c. How does efficiency of the study in this problem compare to that in problem 2? In particular, consider the case where $\nu = \delta + \omega$, $\tau^2 = 1$, and $\psi^2 = \xi^2$.
4. Consider now the setting of a block randomized study in which W_i is a measured variable used in a stratified randomization (but we do not perform an adjusted analysis using W_i). We accrue n subjects to the study, and we ensure that exactly $n/2$ subjects are randomly chosen to receive the experimental treatment, and on each of the treatment arms we have the same number of subjects with advanced disease.
   a. Find expressions for $\mu_x = E[Y_i|X_i = x]$ and $\sigma^2_x = Var(Y_i|X_i = x)$ for $x = 0, 1$.
   b. Let $\theta = \mu_1 - \mu_0$ and $\hat{\theta}$ be the distribution-free estimator of θ. Find an approximate distribution for $\hat{\theta}$ as a function of ω, δ, ν, τ, ψ, ξ, π, and n.
   c. How does efficiency of the study in this problem compare to that in problem 3?
5. Consider now the setting of a block randomized study in which W_i is a measured variable used in a stratified randomization and we do perform an adjusted analysis using W_i. We accrue n subjects to the study, and we ensure that exactly $n/2$ subjects are randomly chosen to receive the experimental treatment, and on each of the treatment arms we have the same number of subjects with advanced disease.
   a. Find expressions for $\mu_{xw} = E[Y_i|X_i = x, W_i = w]$ and $\sigma^2_{xw} = Var(Y_i|X_i = x, W_i = w)$ for $x = 0, 1$ and $w = 0, 1$.
   b. Let $\theta = (1 - \pi)(\mu_{10} - \mu_{00}) + \pi(\mu_{11} - \mu_{01})$ and $\hat{\theta}$ be the distribution-free estimator of θ. Find an approximate distribution for $\hat{\theta}$ as a function of ω, δ, ν, τ, ψ, ξ, π, and n.
   c. How does efficiency of the study in this problem compare to that in problems 3 and 4?
6. Supposing that we can accrue r patients per month in the combined advanced and early stage disease groups, under what conditions would it require less calendar time to complete a study?