Biost/Stat 533
Emerson, Spr 14
Homework #1
March 31, 2014

Written problems to be handed in Monday, April 7.

Under no circumstances may you refer to a homework key from this or other classes. While you may work with other students to derive a solution, when you write up your final solution, you may not refer to any other source. You must be able to develop your answer as if it were being done in a closed book, closed notes examination. You must provide a signed pledge to that effect:

On my honor I have neither given nor received unauthorized aid on the completion of this homework.

1. The vast majority of statistics used for estimation, inference, and prediction involve a sum of some (possibly transformed) observations, with most such cases ultimately concerned with the differences in the means across groups. Suppose we have two groups of interest. Further suppose we can observe n (possibly transformed) independent random variables under a sampling scheme such that $Y_{0j} \sim (\mu_0, \sigma_0^2 < \infty)$ for $j = 1, \ldots, n_0$ and $Y_{1j} \sim (\mu_1, \sigma_1^2 < \infty)$ for $j = 1, \ldots, n_1 = n - n_0$. Let $\overline{Y}_{i\cdot} = \sum_{j=1}^{n_i} Y_{ij}/n_i$.

   a. Under the restriction of a fixed value for n and the probability model described above, what is the optimal choice of n_0 to have maximal precision in an unbiased, distribution-free estimate $\hat{\theta}$ of $\theta = \mu_1 - \mu_0$?

   b. Suppose we are able to randomly assign each subject to one of the two groups. Further suppose that we choose to randomize n subjects in an optimal ratio of r subjects in group 1 to 1 subject in group 0. We consider two possible randomization strategies:

      – "complete randomization" in which a biased coin in tossed for each subject to determine whether or not the subject is in group 1, and

      – "blocked randomization" subjects are sequentially assigned to treatment groups according to a random permutation of n_0 0's and n_1 1's, where $n_1/n_0 = r$ and $n_1 + n_0 = n$.

      Show that the unconditonal relative efficiency of $\hat{\theta}$ under blocked randomization compared to complete randomization is infinite.

   c. After placing reasonable restrictions on the complete randomization strategy to avoid the infinite relative efficiency, explore the gains in unconditional efficiency of blocking for selected choices of n (say n = 20, 50, 100, 200, 500, and 1000) and r (say r = 1, 2, 3, 5, 10).

2. Suppose Y_i, $i = 1, 2, \ldots$ are independent, identically distributed random variables having mean μ, variance σ^2, and finite third and fourth central moments ω and ζ, respectively. Find the asymptotic joint distribution of the sample mean and sample variance.

3. Consider again the probability model of problem 1 with n_0 and n_1 satisfying $n = n_0 + n_1$ and $r = n_1/n_0$ (not necessarily optimal).

   a. Find the asymptotic distribution (as $n \to \infty$) of the t statistic that presumes equal variances.

   b. Suppose we are interested in testing the strong null hypothesis $H_0 : Y_{1j} \sim Y_{0j}$. Under what conditions will is it asymptotically valid to make inference when assuming that the statistic in part (a) has a t distribution? When would such a procedure be asymptotically conservative? When would it be asymptotically anti-conservative?

   c. Suppose we are interested in testing the weak null hypothesis $H_0 : \mu_1 = \mu_0$. Under what conditions will is it asymptotically valid to make inference when assuming that the statistic in part (a) has a t distribution? When would such a procedure be asymptotically conservative? When would it be asymptotically anti-conservative?

d. Find the asymptotic distribution (as $n \to \infty$) of the t statistic that allows unequal variances.

e. Suppose we are interested in testing the strong null hypothesis $H_0 : Y_{1j} \sim Y_{0j}$. Under what conditions will is it asymptotically valid to make inference when assuming that the statistic in part (d) has a t distribution? When would such a procedure be asymptotically conservative? When would it be asymptotically anti-conservative?

f. Suppose we are interested in testing the weak null hypothesis $H_0 : \mu_1 = \mu_0$. Under what conditions will is it asymptotically valid to make inference when assuming that the statistic in part (d) has a t distribution? When would such a procedure be asymptotically conservative? When would it be asymptotically anti-conservative?

4. Suppose that for $k = 1, \ldots, K$ we have totally independent estimators $\hat{\theta}_k \sim (\theta, V_k)$ (that is, $\hat{\theta}_k$ is an unbiased estimator of θ, and that the standard error of $\hat{\theta}_k$ is $\sqrt{V_k}$.

   a. For $K = 2$, we consider weighted linear combinations of the form $\hat{\theta} = w_1\hat{\theta}_1 + w_2\hat{\theta}_2$. Without appealing to the Gauss-Markov theorem, find expressions for w_k such that $\hat{\theta}$ is unbiased for θ and has the smallest variance of all such linear combinations.

   b. Generalize your result to an arbitrary value of K of $\hat{\theta} = \sum_{k=1}^{K} w_k\hat{\theta}_k$. (It is okay to solve the general case first, it you want.)