Statistics 312/612, fall 2016
Homework # 8
Due: Wednesday 9 November

The ExptDesign handout described a simple example where a treatment, represented by a factor $\mathbb{F}$ (changed from $\mathbb{T}$) on t levels, is applied to each of $n = tk$ experimental units (plots), with each level being applied k times. With F_i denoting the $n \times 1$ dummy variable that indicated which plots received $\mathbb{F}$ at level i, the model assumes that, conditional on $\mathbb{F}$,

$$y = d + \sum\nolimits_{i=1}^{t} \tau_i F_i + \xi \qquad \text{with } \mathbb{E}\xi = 0 \text{ and } \mathrm{var}(\xi) = \sigma^2 I_n.$$

The column vector $d = [d_1, \dots, d_n]$ represents largely unknown contributions coming from the plots themselves, regardless of which level of treatment is applied.

The handout described two methods, CRD (sometimes called CAR by mistake) and RBD, for making $\mathbb{F}$ random (independently of ξ). For the estimator

$$\widehat{\Delta} = \langle y, F_1 - F_2 \rangle / k = \tau_1 - \tau_2 + \langle d + \xi, F_1 - F_2 \rangle / k,$$

the handout showed that $\mathbb{E}\widehat{\Delta} = \tau_1 - \tau_2$ for both CRD and RBD. It also showed that

$$k^2 \mathrm{var}(\widehat{\Delta}) = 2k\sigma^2 + 2\|d\|^2/t + \sum\nolimits_{\alpha=1}^{n} \sum\nolimits_{\beta=1}^{n} \mathbb{1}\{\alpha \neq \beta\} d_\alpha d_\beta g(\alpha, \beta)$$

where, for $\alpha \neq \beta$,

$$\begin{aligned} g(\alpha, \beta) &= \mathbb{E}\Big[(F_1[\alpha] - F_2[\alpha])(F_1[\beta] - F_2[\beta])\Big] \\ &= 2\mathbb{P}\{\mathbb{F}[\alpha] = 1, \mathbb{F}[\beta] = 1\} - 2\mathbb{P}\{\mathbb{F}[\alpha] = 1, \mathbb{F}[\beta] = 2\}. \end{aligned}$$

Write $\overline{d}$ for $n^{-1} \sum_\alpha d_\alpha$, the average of the d_α's. Also, for RBD, write $\mathcal{B}_i$ for the set of plots in the ith block and let $\overline{d}_i = t^{-1} \sum_{\alpha \in \mathcal{B}_i} d_\alpha$, the average of the d_α's within the ith block, for $i = 1, \dots, k$.

For this homework I want you to show that

<1> $$\mathrm{var}(\widehat{\Delta}) = \frac{2}{k}\Big(\sigma^2 + (n-1)^{-1} \sum\nolimits_\alpha (d_\alpha - \overline{d})^2\Big) \qquad \text{under CRD}$$

<2> $$\mathrm{var}(\widehat{\Delta}) = \frac{2}{k}\left(\sigma^2 + (n-k)^{-1} \sum\nolimits_{i=1}^{k} \sum\nolimits_{\alpha \in \mathcal{B}_i} (d_\alpha - \overline{d}_i)^2\right) \qquad \text{under RBD}$$

[1] Define $D_\alpha = d_\alpha - \overline{d}$.

(i) (10 points) Explain why, for both cases, $\mathrm{var}(\widehat{\Delta})$ is unchanged if replace d_α by D_α for each α. Of course you should not assume the validity of <1> or <2> to answer this question.

(ii) (10 points) Explain why $2 \sum_{\alpha < \beta} D_\alpha D_\beta = -\sum_a D_\alpha^2$.

[2] For CRD:

(i) (10 points) Calculate $g(\alpha, \beta)$. Explain your reasoning.

(ii) (10 points) Using the result from (i) derive <1>. Explain your reasoning.

[3] For RBD, write $\overline{D}_i$ for $= t^{-1} \sum_{\alpha \in \mathcal{B}_i} D_\alpha$.

(i) (10 points) Explain why

$$\sum_{\alpha \in \mathcal{B}_i} D_\alpha^2 - \frac{1}{t-1} \sum_{\alpha \in \mathcal{B}_i} \sum_{\beta \in \mathcal{B}_i} \mathbb{1}\{\alpha \neq \beta\} D_\alpha D_\beta = \frac{t}{t-1} \sum\nolimits_{\alpha \in \mathcal{B}_i} (D_\alpha - \overline{D}_i)^2.$$

Hint: Compare with Problem [1](ii).

(ii) (5 points) Calculate $g(\alpha, \beta)$ if plots α and β belong to the same block. Explain your reasoning.

(iii) (5 points) Calculate $g(\alpha, \beta)$ if plots α and β belong to different blocks. Explain your reasoning.

(iv) (10 points) Use the results from (i) through (iii) to derive <2>. Explain your reasoning.

[4] (10 points) For RBD, explain why

$$\sum\nolimits_{\alpha=1}^{n}(d_\alpha - \overline{d})^2 = \sum\nolimits_{i=1}^{k}\sum\nolimits_{\alpha\in\mathcal{B}_i}(d_\alpha - \overline{d}_i)^2 + t\sum\nolimits_{i=1}^{k}(\overline{d}_i - \overline{d})^2.$$

[5] (extra credit) Problem [3] involved a lot of algebra. Here is a neater method. Let $B_1, \dots, B_k$ denote the vectors of dummy variables for the blocks. Decompose d as a component in the subspace of $\mathbb{R}^n$ spanned by $B_1, \dots, B_k$ plus a component M that is orthogonal to that subspace. Show that

$$\mathrm{var}\langle d, F_1 - F_2\rangle = \mathrm{var}\langle M, F_1 - F_2\rangle \quad \text{AND} \quad \mathrm{var}(F_1 - F_2) = c_1 I_n - c_2 \sum\nolimits_{i\le k} B_i B_i^T.$$

for constants c_1 and c_2. Then explain why $\mathrm{var}\langle M, F_1 - F_2\rangle = c_1 \|M\|^2$.