Statistics 312/612, fall 2016 Homework # 5 Due: Wednesday 12 October

- [1] Suppose \mathbb{F} is a factor variable (of length n) with k levels and \mathbb{G} is another factor variable (of length n) with ℓ levels. Let $F = (F_1, \ldots, F_k)$ and $G = (G_1, \ldots, G_\ell)$ denote the corresponding matrices of dummy variables.
 - (i) (5 points) Suppose none of the F_i 's is a zero vector. Show that $\operatorname{rank}(\mathbb{1}_n, F) = \operatorname{rank}(F) = k$. Hint: If $\sum_i a_i F_i = 0$ why must all a_i 's be zero?
 - (ii) (5 points) Let X be the $n \times (k + \ell + 1)$ matrix $(\mathbb{1}_n, F, G)$. Explain why rank $(X) \le k + \ell + 1$.

From now on assume $F_i * G_j \neq 0$ for all i, j.

(iii) (10 points) Define a $(k + \ell + 1) \times 2$ matrix by

$$D = (d_1, d_2) = \begin{pmatrix} 0_1 & 0_1 \\ \mathbb{1}_k & 0_k \\ 0_\ell & \mathbb{1}_\ell \end{pmatrix} \quad \text{where } 0_m = \texttt{rep(0,m)}.$$

Let X_1 denote the $(n+2) \times (k+\ell+1)$ matrix obtained by appending the two rows of D^T to X. Show that the equations

$$X_1 \begin{pmatrix} m \\ a \\ b \end{pmatrix} = 0$$
 with $a \in \mathbb{R}^k$ and $b \in \mathbb{R}^\ell$

have a unique solution: $m = 0, a = 0_k, b = 0_\ell$. Explain why rank $(X_1) = k + \ell + 1$. Hint: Write out the equations as expressions involving m, a_i, b_j . Try summing over the subscripts.

- (iv) (10 points) Explain why $\operatorname{rank}(X_1) \leq \operatorname{rank}(X) + 2$.
- (v) (5 points) Explain why rank $(X) = k + \ell 1$.

P.T.O.

[2] (20 points) The following plot shows pairs of "good" coefficients for samples of size 48 take from the BC data with replacement. That is, for each sample of 48 rows I ran lm(rate ~ Ht + Hp, BC, subset= ???), calculated the estimates for the eight constrained coefficients, then plotted a blue circle for the coefficients shown on the axes. The red dot shows the corresponding point for lm(rate ~ Ht + Hp, BC).

source("HW5.2.R")
draw.plot("A","I",50) # my solution to the problem





Produce a similar plot yourself. Make sure you show your source code, clearly commented. Don't forget set.seed(0) if you want your plot to contain the same points as the plots produced by the TAs. Feel free to make your plot look prettier.