COMPLETELY RANDOMIZED DESIGNS (CRD)

- For now, t unstructured treatments (e.g. no factorial structure)
- "Completely randomized" means no restrictions on the randomization of units to treatments (except for sample sizes ordinarily fixed in advance), so with N available units:
  - Randomly select n_1 units for assignment to trt 1
  - Randomly select n_2 remaining units for assignment to trt 2
  - ...
  - Last n_t units are assigned to trt t
  - Or, any other procedure s.t. all $N!/(n_1!n_2!...n_t!)$ possible assignment patterns have equal probability
- Ordinarily reflects model and assumptions for 1-way ANOVA

"Cell Means" model:

- $y_{i,j} = \mu_i + \epsilon_{i,j}$
- $i = 1...t$, $j = 1...n_i$, $\sum_{i=1}^{t} n_i = N$
- μ_i represents $E[y_{i,j}]$, a response corresponding to treatment i,
  - in the laboratory in which the experiment was performed
  - on the day the experiment was run ...
- $E[\epsilon_{i,j}] = 0$, $Var[\epsilon_{i,j}] = \sigma^2$, independent, perhaps normal
- ϵ represents (in part) unit-to-unit variation; "indep" is justified (in part) by complete randomization
- Matrix representation:
  - $\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\epsilon}$
  - $[N \times 1] = [N \times t][t \times 1] + [N \times 1]$

$$\begin{pmatrix} y_{1,1} \\ ... \\ y_{1,n_1} \\ y_{2,1} \\ ... \\ y_{2,n_2} \\ ... \\ y_{t,1} \\ ... \\ y_{t,n_t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ ... \\ \mu_t \end{pmatrix} + \begin{pmatrix} \epsilon_{1,1} \\ ... \\ \epsilon_{1,n_1} \\ \epsilon_{2,1} \\ ... \\ \epsilon_{2,n_2} \\ ... \\ \epsilon_{t,1} \\ ... \\ \epsilon_{t,n_t} \end{pmatrix}$$

"Effects" model:

- $y_{i,j} = \alpha + \tau_i + \epsilon_{i,j}$
- This *says* the same thing mathematically ...
  - Let $\alpha =$ anything, then $\tau_i = \mu_i - \alpha$
- α now represents everything *in common* to all experimental runs, τ_i is the deviation from this due *only* to treatment i
- Matrix representation:
  - $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\theta} = (\alpha, \tau_1, \tau_2, ..., \tau_t)'$
  - $[N \times 1] = [N \times (t+1)][(t+1) \times 1] + [N \times 1]$
  - $\mathbf{X}$ is a different matrix now ... will often use "$\mathbf{X}$" generically
  - Less-than-full-rank

$$
\begin{pmatrix} y_{1,1} \\ \cdots \\ y_{1,n_1} \\ y_{2,1} \\ \cdots \\ y_{2,n_2} \\ \cdots \\ y_{t,1} \\ \cdots \\ y_{t,n_t} \end{pmatrix}
=
\begin{pmatrix}
1 & 1 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
1 & 1 & 0 & \cdots & 0 \\
1 & 0 & 1 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
1 & 0 & 1 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
1 & 0 & 0 & \cdots & 1 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
1 & 0 & 0 & \cdots & 1
\end{pmatrix}
\begin{pmatrix} \alpha \\ \tau_1 \\ \tau_2 \\ \cdots \\ \tau_t \end{pmatrix}
+
\begin{pmatrix} \epsilon_{1,1} \\ \cdots \\ \epsilon_{1,n_1} \\ \epsilon_{2,1} \\ \cdots \\ \epsilon_{2,n_2} \\ \cdots \\ \epsilon_{t,1} \\ \cdots \\ \epsilon_{t,n_t} \end{pmatrix}
$$

ESTIMATION: review

- $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$ ... (general $\mathbf{X}$, iid errors)

  (first piece is expectation of $\mathbf{y}$ ... the "noiseless model")

- MLE for normal errors, LS in any case, is any sol'n to:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\theta}} = \mathbf{X}'\mathbf{y}$$

  - "Normal equations", as in "norm", not "normal dist'n"
  - $\mathbf{X}$ of full column rank: $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
  - Otherwise: $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$
    - generalized inverse, non-unique
    - $\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}$

- Need to review statistical properties of $\hat{\boldsymbol{\theta}}$

EXPECTATIONS:

- Recall that for random $\mathbf{y}$ with $E[\mathbf{y}] = \mathbf{m}$, $\quad Var[\mathbf{y}] = \mathbf{V}$
  - $\mathbf{m}$ vector of means, $\mathbf{V}$ matrix of variances and covariances
  - $E[\mathbf{Ay}] = \mathbf{Am}$
  - $Var[\mathbf{Ay}] = \mathbf{AVA}'$

    (generalization of "constant-squared" in the scalar case)

- So,
  - $E[\hat{\boldsymbol{\theta}}] = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'(\mathbf{X}\boldsymbol{\theta})$ ... generally
  - $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$ ... full rank

- In the less-than-full-rank case, $\hat{\boldsymbol{\theta}}$ is not uniquely determined, but consider *estimable functions*:
  - $\mathbf{c}'\boldsymbol{\theta}$, where $\mathbf{c}' = \mathbf{l}'\mathbf{X}$ for some N-vector $\mathbf{l}$
  - $\widehat{\mathbf{c}'\boldsymbol{\theta}} = \mathbf{c}'\hat{\boldsymbol{\theta}} = (\mathbf{l}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$
- $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' = \mathbf{H}$ is invariant to choice of g-inverse
  - "projection matrix", into space spanned by columns of $\mathbf{X}$
  - symmetric $N \times N$, idempotent, $\mathbf{H}$ for "hat" because $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\theta}} = \mathbf{H}\mathbf{y}$
  - $\mathbf{c}'\hat{\boldsymbol{\theta}} = \mathbf{l}'\mathbf{X}\hat{\boldsymbol{\theta}} = \mathbf{l}'\mathbf{H}\mathbf{y}$
  - $E[\mathbf{c}'\hat{\boldsymbol{\theta}}] = \mathbf{l}'\mathbf{H}\mathbf{X}\boldsymbol{\theta} = \mathbf{l}'\mathbf{X}\boldsymbol{\theta} = \mathbf{c}'\boldsymbol{\theta}$
  - Likewise for $\mathbf{C} = \begin{pmatrix} \mathbf{c}_1' \\ \ldots \\ \mathbf{c}_m' \end{pmatrix}$, $E[\mathbf{C}\hat{\boldsymbol{\theta}}] = \mathbf{C}\boldsymbol{\theta}$ if $\mathbf{C} = \mathbf{L}\mathbf{X}$

- CRD Examples:
  - τ_i isn't estimable
  - $\mu_i = \alpha + \tau_i$ is estimable, but is a function of "experiment" as well as "treatment"
  - $\mu_i - \mu_j = \tau_i - \tau_j$ is estimable ... comparative

(see $\mathbf{X}$, slide 5)

VARIANCES:

- Recall for estimable $\mathbf{C}\boldsymbol{\theta}$,

$$\mathbf{C}\hat{\boldsymbol{\theta}} = (\mathbf{LX})(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} = \mathbf{LHy}$$

- $Var[\mathbf{C}\hat{\boldsymbol{\theta}}] = \mathbf{LH}(\sigma^2\mathbf{I})\mathbf{H}'\mathbf{L}' = \sigma^2\mathbf{LHL}'$

  $= \sigma^2(\mathbf{LX})(\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{L}') = \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'$ ... generally

- $Var[\mathbf{C}\hat{\boldsymbol{\theta}}] = \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$ ... full rank

- in particular, $Var[\hat{\boldsymbol{\theta}}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ ... full rank

ALLOCATION:

- Interest in estimating $\mathbf{C}\boldsymbol{\theta}$ well, i.e. with minimum variance
- Allocate N units to treatments to minimize the average contrast variance
- $min_{n_i, i=1 \ldots t}\ ave_j\ \sigma^2 \mathbf{c}_j'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_j$, subject to $\sum_{i=1}^{t} n_i = N$
  - or g-inv if $\mathbf{X}$ less than full rank and $\mathbf{C}\boldsymbol{\theta}$ estimable
  - $\mathbf{c}_j'$, the jth row in $\mathbf{C}$
  - σ^2 is a positive constant, doesn't matter
- $min_{n_i, i=1 \ldots t}$ trace $\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$, subject to $\sum_{i=1}^{t} n_i = N$
- $min_{n_i, i=1 \ldots t}$ trace $\mathbf{C}'\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}$, subject to $\sum_{i=1}^{t} n_i = N$
  - *linear optimality*, generalization of *A-optimality*
- If $\mathbf{X}$ is not of full rank but $\mathbf{C}\boldsymbol{\theta}$ is estimable, same argument applies to $(\mathbf{X}'\mathbf{X})^{-}$

EXAMPLE

- Full-rank cell means model, $\boldsymbol{\theta} = \boldsymbol{\mu}$

$$(\mathbf{X}'\mathbf{X})^{-1} = \mathsf{diag}(1/n_1, 1/n_2, ...1/n_t)$$

- Interest in estimating differences between the "control" (treatment 1) and each of the other treatments

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & ... & 0 \\ 1 & 0 & -1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 1 & 0 & 0 & ... & -1 \end{pmatrix}$$

$- \quad t-1 \quad \times \quad t$

$$\mathbf{C'C} = \begin{pmatrix} t-1 & -1 & -1 & ... & -1 \\ -1 & 1 & 0 & ... & 0 \\ -1 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ -1 & 0 & 0 & ... & 1 \end{pmatrix}$$

– $t \quad \times \quad t$

- trace $\mathbf{C'C}(\mathbf{X'X})^{-1} = (t-1)/n_1 + \sum_{i=2}^{t} 1/n_i$
- want to:
  - minimize, w.r.t. $n_1, n_2, ...n_t$,

$$(t-1)/n_1 + \sum_{i=2}^{t} 1/n_i$$

  - subject to:

$$\sum_{i=1}^{t} n_i = N$$

- Constrained Minimization by *Method of Lagrangian Multipliers*:

$$\phi = [(t-1)/n_1 + \textstyle\sum_{i=2}^{t} 1/n_i] + \lambda(\textstyle\sum_{i=1}^{t} n_i - N)$$

(first term is what you want to minimize, second is what is constrained to be zero)

$$\partial\phi/\partial n_1 = -(t-1)/n_1^2 + \lambda$$
$$\partial\phi/\partial n_i = -1/n_i^2 + \lambda \quad i = 2...t$$
$$\partial\phi/\partial\lambda = \textstyle\sum_{i=1}^{t} n_i - N$$

  – Set each to zero and solve:

$$n_1 = \sqrt{(t-1)/\lambda}$$
$$n_i = \sqrt{1/\lambda} \quad i = 2...t$$

$$n_1 = \frac{\sqrt{t-1}N}{\sqrt{t-1}+(t-1)} \quad n_i = \frac{N}{\sqrt{t-1}+(t-1)}$$

- Note: This is really a continuous optimization technique applied to a discrete problem. In practice, need to round solutions to integer values, or use as starting point in a discrete search
- Note also that $Cov[\widehat{\mu_1 - \mu_i}, \widehat{\mu_1 - \mu_j}] = \sigma^2/n_1$
  - relatively larger n_1 reduces these ...
- Postscript:
  - intuition might say you need *less* info about control, since you have more experience with it ...
  - but this means you intend to USE that information ... e.g. via a Bayesian prior

EXAMPLE: Less-Than-Full-Rank version

- "Effects" model, $\boldsymbol{\theta} = (\alpha \ \tau_1 \ ... \ \tau_t)'$

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & n_1 & n_2 & ... & n_t \\ n_1 & n_1 & 0 & ... & 0 \\ n_2 & 0 & n_2 & ... & 0 \\ ... & ... & ... & ... & ... \\ n_t & 0 & 0 & ... & n_t \end{pmatrix}$$

$$(\mathbf{X}'\mathbf{X})^- = \begin{pmatrix} 0 & 0 & 0 & ... & 0 \\ 0 & 1/n_1 & 0 & ... & 0 \\ 0 & 0 & 1/n_2 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1/n_t \end{pmatrix}$$

- General algorithm for g-inverse of symmetric matrices (e.g. "Linear Models", Searle, Wiley)
  - find a full-rank, principal minor of maximum dimension
  - invert it
  - fill in zeros everywhere else

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & -1 & 0 & ... & 0 \\ 0 & 1 & 0 & -1 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 1 & 0 & 0 & ... & -1 \end{pmatrix}$$

- $\mathbf{C}'\mathbf{C}$ and this $(\mathbf{X}'\mathbf{X})^-$ are as before, but with leading row/column of zero ... so problem and solution are the same

HYPOTHESIS TESTING: review

- Alternative (full) and Null (restricted) hypotheses:
  - $\text{Hyp}_A : \mathbf{y} = \mathbf{X}_A \boldsymbol{\theta}_A + \boldsymbol{\epsilon}$
  - $\text{Hyp}_0 : \mathbf{y} = \mathbf{X}_0 \boldsymbol{\theta}_0 + \boldsymbol{\epsilon} = \mathbf{X}_A \mathbf{P} \boldsymbol{\theta}_0 + \boldsymbol{\epsilon}$
- Columns of $\mathbf{X}_0$ are all linear combinations of columns of $\mathbf{X}_A$
- $\mathbf{X}_A$ is N-by-p_A, $\mathbf{P}$ is p_A-by-p_0, $p_0 < p_A$

- Test for treatment effects: cell means model

$$\mathbf{X}_A = \begin{pmatrix} \mathbf{1} & & & \\ & \mathbf{1} & & \\ & & \mathbf{1} & \\ & & & \mathbf{1} \end{pmatrix}, \mathbf{P} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{X}_0 = \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \\ \mathbf{1} \\ \mathbf{1} \end{pmatrix}$$

- Test for treatment effects: effects model

$$\mathbf{X}_A = \begin{pmatrix} \mathbf{1} & \mathbf{1} & & & \\ \mathbf{1} & & \mathbf{1} & & \\ \mathbf{1} & & & \mathbf{1} & \\ \mathbf{1} & & & & \mathbf{1} \end{pmatrix}, \mathbf{P} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{X}_0 = \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \\ \mathbf{1} \\ \mathbf{1} \end{pmatrix}$$

- Test for "additional terms": partitioned model
  - $\text{Hyp}_0 : \mathbf{y} = \mathbf{X}_1\boldsymbol{\theta}_1 + \boldsymbol{\epsilon}$
  - $\text{Hyp}_A : \mathbf{y} = \mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2 + \boldsymbol{\epsilon}$
- This is a special case of the general form:

$$\mathbf{X}_A = (\mathbf{X}_1, \mathbf{X}_2),\ \mathbf{P} = \begin{pmatrix} \mathbf{I} \\ \mathbf{0} \end{pmatrix},\ \mathbf{X}_0 = \mathbf{X}_A\mathbf{P} = \mathbf{X}_1$$

- But $\mathbf{X}_0$ doesn't *have* to be a subset of columns from $\mathbf{X}_A$

- For either model:

$$
\begin{aligned}
\text{SSE} &= \textstyle\sum_i \sum_j (y_{i,j} - \hat{y}_{i,j})^2 \text{ ... sometimes "resid"} \\
&= \textstyle\sum_i \sum_j (y_{i,j} - \mathbf{x}_i'\hat{\boldsymbol{\theta}})^2, \text{ where } \mathbf{x}_i' \text{ is row from } \mathbf{X} \\
&= \textstyle\sum_i \sum_j (y_{i,j} - \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y})^2 \\
&= \text{squared length of } \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y} \\
&= \mathbf{y}'(\mathbf{I}-\mathbf{H})(\mathbf{I}-\mathbf{H})\mathbf{y} = \mathbf{y}'(\mathbf{I}-\mathbf{H}-\mathbf{H}+\mathbf{H}^2)\mathbf{y} \\
&= \mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y} \text{ (key result for what's coming up)} \\
&= \mathbf{y}'\mathbf{y} - (\mathbf{H}\mathbf{y})'(\mathbf{H}\mathbf{y}) \\
&= \text{length}^2 \text{ of } \mathbf{y} \text{ - length}^2 \text{ of } \mathbf{H}\mathbf{y}
\end{aligned}
$$

- Given truth of Hyp_A (which means Hyp_0 MAY also be true), the validity of Hyp_0 is judged by how much worse the data fit the corresponding model:

$$\text{SST} = \text{SSE}(\text{Hyp}_0) - \text{SSE}(\text{Hyp}_A)$$

$$= \mathbf{y}'(\mathbf{I} - \mathbf{H}_0)\mathbf{y} - \mathbf{y}'(\mathbf{I} - \mathbf{H}_A)\mathbf{y}$$

$$= \mathbf{y}'\mathbf{H}_A\mathbf{y} - \mathbf{y}'\mathbf{H}_0\mathbf{y} = \mathbf{y}'(\mathbf{H}_A - \mathbf{H}_0)\mathbf{y}$$

- Notation: Here and later, the subscript on $\mathbf{H}$ is the subscript on the corresponding $\mathbf{X}$. Here it is 0 or A, later also 1 or 2.

EXAMPLE

- Full-rank (cell means) model
- $\text{Hyp}_0 : \mu_1 = \mu_2 = ... = \mu_t$, or $y_{i,j} = \mu + \epsilon_{i,j}$
- Hyp_A : at least two differ, or $y_{i,j} = \mu_i + \epsilon_{i,j}$

$$\mathbf{X}_0 = \begin{pmatrix} \mathbf{1}_{n_1} \\ \mathbf{1}_{n_2} \\ ... \\ \mathbf{1}_{n_t} \end{pmatrix} \quad , \quad \mathbf{H}_0 = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' = \frac{1}{N}\mathbf{J}_{N\times N}$$

$$\mathbf{X}_A = \begin{pmatrix} \mathbf{1}_{n_1} & ... & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & ... & \mathbf{0}_{n_2} \\ ... & ... & ... \\ \mathbf{0}_{n_t} & ... & \mathbf{1}_{n_t} \end{pmatrix}, \; \mathbf{H}_A = \begin{pmatrix} \frac{1}{n_1}\mathbf{J}_{n_1\times n_1} & ... & \mathbf{0}_{n_1\times n_t} \\ \mathbf{0}_{n_2\times n_1} & ... & \mathbf{0}_{n_2\times n_t} \\ ... & ... & ... \\ \mathbf{0}_{n_t\times n_1} & ... & \frac{1}{n_t}\mathbf{J}_{n_t\times n_t} \end{pmatrix}$$

Letting $\mathbf{y}_i$ represent the n_i-element vector of observations associated with treatment i,

$$\begin{aligned} \mathsf{SST} = \mathbf{y}'\mathbf{H}_A\mathbf{y} - \mathbf{y}'\mathbf{H}_0\mathbf{y} &= \textstyle\sum_i \mathbf{y}_i'(\frac{1}{n_i}\mathbf{J})\mathbf{y}_i - \mathbf{y}'(\frac{1}{N}\mathbf{J})\mathbf{y} \\ &= \textstyle\sum_i \frac{1}{n_i} y_{i.}^2 - \frac{1}{N} y_{..}^2 \\ &= \textstyle\sum_i n_i(\bar{y}_{i.})^2 - N(\bar{y}_{..})^2 \\ &= \textstyle\sum_i n_i(\bar{y}_{i.} - \bar{y}_{..})^2 \end{aligned}$$

EXAMPLE

- Reduced-rank (effects) model, same problem
- Due to the invariance of the hat matrices, $\mathbf{H}_0$ and $\mathbf{H}_A$ are the same as before
- So, the resultant sums of squares are also the same

EXPECTATIONS

- Recall that for random $\mathbf{y}$ with $E[\mathbf{y}] = \mathbf{m}, \;\; Var[\mathbf{y}] = \mathbf{V}$
  - $E[\mathbf{y}'\mathbf{A}\mathbf{y}] = \mathbf{m}'\mathbf{A}\mathbf{m} + \text{trace}(\mathbf{AV})$
- Suppose Hyp$_A$ really is true (perhaps also Hyp$_0$), so:
  - $E[\mathbf{y}] = \mathbf{X}_A\boldsymbol{\theta}_A$
  - $E[\text{SSE(Hyp}_A)]$

$$= \boldsymbol{\theta}'_A\mathbf{X}'_A(\mathbf{I} - \mathbf{H}_A)\mathbf{X}_A\boldsymbol{\theta}_A \;+\; \text{trace}((\mathbf{I} - \mathbf{H}_A)\sigma^2\mathbf{I}))$$

$$= \boldsymbol{\theta}'_A\mathbf{X}'_A(\mathbf{X}_A - \mathbf{X}_A)\boldsymbol{\theta}_A \;+\; \sigma^2(\text{trace}(\mathbf{I}) \;-\; \text{trace}(\mathbf{H}_A))$$

$$= 0 \;+\; \sigma^2(N \;-\; \text{rank}(\mathbf{X}_A))$$

(because for idempotent matrices, rank=trace,

and $\text{rank}(\mathbf{H}) = \text{rank}(\mathbf{X})$)

- $E[\text{ SSE(Hyp}_0\text{) }]$

$$= \boldsymbol{\theta}_A'\mathbf{X}_A'(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_A\boldsymbol{\theta}_A \; + \; \text{trace}((\mathbf{I}-\mathbf{H}_0)\sigma^2\mathbf{I}))$$
$$= \boldsymbol{\theta}_A'\mathbf{X}_A'(\mathbf{I}-\mathbf{H}_0)\mathbf{X}_A\boldsymbol{\theta}_A \; + \; \sigma^2(N - \text{rank}(\mathbf{X}_0))$$
$$= Q_{A|0}(\boldsymbol{\theta}_A) \; + \; \sigma^2(N - \text{rank}(\mathbf{X}_0))$$

(subscript $A|0$ comes from space($\mathbf{X}_A$) -space($\mathbf{X}_0$))

- For the partitioned model:

$$= (\boldsymbol{\theta}_1'\mathbf{X}_1' + \boldsymbol{\theta}_2'\mathbf{X}_2')(\mathbf{I}-\mathbf{H}_1)(\mathbf{X}_1\boldsymbol{\theta}_1 + \mathbf{X}_2\boldsymbol{\theta}_2)$$
$$+ \sigma^2(N - \text{rank}(\mathbf{X}_1))$$
$$= \boldsymbol{\theta}_2'\mathbf{X}_2'(\mathbf{I}-\mathbf{H}_1)\mathbf{X}_2\boldsymbol{\theta}_2 \; + \; \sigma^2(N - \text{rank}(\mathbf{X}_1))$$

($\mathbf{X}_1$ terms vanish ... why?)

$$= Q_{2|1}(\boldsymbol{\theta}_2) \; + \; \sigma^2(N - \text{rank}(\mathbf{X}_1))$$

(subscript $2|1$ comes from space($\mathbf{X}_2$) -space($\mathbf{X}_1$))

- $E[\text{ SST }] = Q_{2|1}(\boldsymbol{\theta}_2) \; + \; (\text{rank}(\mathbf{X}_A) - \text{rank}(\mathbf{X}_0))\sigma^2$

EXAMPLE: All-Treatments-Equal Hypothesis

- In either parameterization:
  - $\text{rank}(\mathbf{X}_A) - \text{rank}(\mathbf{X}_0) = t - 1$
  - $\mathbf{H}_0 = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' = \frac{1}{N}\mathbf{J}$
- Full rank:
  - $\mathbf{X}_A'(\mathbf{I} - \mathbf{H}_0)\mathbf{X}_A = \mathbf{X}_A'\mathbf{X}_A - \frac{1}{N}\mathbf{X}_A'\mathbf{J}\mathbf{X}_A$
    $= \text{diag}(\mathbf{n}) - \frac{1}{N}\mathbf{n}\mathbf{n}'$, where $\mathbf{n} = (n_1, n_2, ...n_t)'$

$$Q_{A|0}(\boldsymbol{\mu}) = \sum_{i=1}^t n_i\mu_i^2 - \frac{1}{N}\left(\sum n_i\mu_i\right)^2$$
$$= \sum_{i=1}^t n_i(\mu_i - \bar{\mu})^2, \quad \bar{\mu} = \frac{1}{N}\sum_{i=1}^t n_i\mu_i$$

- Reduced rank:
  - $\mathbf{X}_2'(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2 = \mathbf{X}_A'(\mathbf{I} - \mathbf{H}_0)\mathbf{X}_A$ in full-rank case
    $Q_{2|1}(\boldsymbol{\tau}) = \sum_{i=1}^t n_i(\tau_i - \bar{\tau})^2, \quad \bar{\tau} = \frac{1}{N}\sum n_i\tau_i$

DISTRIBUTION THEORY: review

1. $\mathbf{y} \sim MVN(\mathbf{m}, \sigma^2\mathbf{I})$,
   $\mathbf{A}$ and $\mathbf{B}$ p.s.d. symmetric,
   Any two of:
   - $\mathbf{A}$ and $\mathbf{B}$ idempotent
   - $\mathbf{A}+\mathbf{B}$ idempotent
   - $\mathbf{AB}=\mathbf{0}$

   Implies all of:
   - $\frac{1}{\sigma^2}\mathbf{y}'\mathbf{A}\mathbf{y} \sim \chi^{2'}(\text{rank}(\mathbf{A}), \frac{1}{\sigma^2}\mathbf{m}'\mathbf{A}\mathbf{m})$
   - $\frac{1}{\sigma^2}\mathbf{y}'\mathbf{B}\mathbf{y} \sim \chi^{2'}(\text{rank}(\mathbf{B}), \frac{1}{\sigma^2}\mathbf{m}'\mathbf{B}\mathbf{m})$
   - the two quadratic forms are independent
2. For idempotent $\mathbf{A}$, $\text{rank}(\mathbf{A}) = \text{trace}(\mathbf{A})$

IMPLICATIONS for CRD

- $\frac{1}{\sigma^2}$ SSE(Hyp$_A$) $= \frac{1}{\sigma^2}\mathbf{y}'(\mathbf{I} - \mathbf{H}_A)\mathbf{y} \;\sim\; \chi^{2'}(N-t, 0)$
  i.e. "central"
- $\frac{1}{\sigma^2}$ SST$= \frac{1}{\sigma^2}\mathbf{y}'(\mathbf{H}_A - \mathbf{H}_0)\mathbf{y} \;\sim\; \chi^{2'}(t-1, Q(\boldsymbol{\tau})/\sigma^2)$

$$\frac{\text{SST}}{t-1} \Big/ \frac{\text{SSE (Hyp}_A)}{N-t} \sim F'(t-1, N-t, Q(\boldsymbol{\tau})/\sigma^2)$$

- or, replace $\boldsymbol{\tau}$ with $\boldsymbol{\mu}$ in Q ... same thing
- Other things being equal, larger $Q \rightarrow$ greater power

Example

- CRD $\quad N = 20 \quad t = 3 \quad n_1 = n_2 = 8,\ n_3 = 4$
- numerator df $= 2$ denominator df $= 17$
- Suppose $\mu_1 = 1\ \ \mu_2 = 2\ \ \mu_3 = 3\ \ \sigma = 1.5$
- Then
  - $\bar{\mu} = 1.8$
  - $Q(\ \boldsymbol{\mu}\ ) = 8(1 - 1.8)^2 + 8(2 - 1.8)^2 + 4(3 - 1.8)^2 = 11.2$
  - $Q(\ \boldsymbol{\mu}\ )/\sigma^2 = 4.9\overline{777}$

- R:
  - qf(.95, 2, 17) $\rightarrow$ 3.5915

    95th quantile of central F, i.e. critical value
  - 1-pf(3.5915, 2, 17, 4.9777) $\rightarrow$ 0.4308

    probability of value in critical region for non-central F
- SAS:
  - proc iml;
  - finv(.95, 2, 17); $\rightarrow$ 3.5915
  - 1-probf(3.5915, 2, 17, 4.9777); $\rightarrow$ 0.4308

PRACTICAL POINT

- May not know enough about problem to speculate about the (complete) true value of $\boldsymbol{\mu}$, but willing to guess:

$$D = \mu_{max} - \mu_{min} = \tau_{max} - \tau_{min}$$

- Bounds for Q (assume equal n_i)
  - Most favorable situation (Q greatest)
    - greatest "variance" of μ's
    - $t/2$ μ's $= \mu_{max}$ $t/2$ μ's $= \mu_{min}$
    - $Q = n_i \times t \times (\frac{1}{2}D)^2 = N \times D^2/4$
  - Least favorable situation (Q smallest)
    - smallest "variance" of μ's
    - 1 $\mu = \mu_{max}$ 1 $\mu = \mu_{min}$
    - others $= (\mu_{max} + \mu_{min})/2$
    - $Q = 2 \times n_i \times (\frac{1}{2}D)^2 = n_i \times D^2/2$

REDUCED NORMAL EQUATIONS FOR $\boldsymbol{\tau}$

- $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon} \ = \ \mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2\boldsymbol{\tau} + \boldsymbol{\epsilon}$
  - $\boldsymbol{\beta}$ are nuisance parameters
  - for our CRD effects model, $\mathbf{X}_1\boldsymbol{\beta} = \mathbf{1}\alpha$
- "Full" normal equations:

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\theta}} = \mathbf{X}'\mathbf{y}$$

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\tau}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{pmatrix}$$

$$\mathbf{X}_1'\mathbf{X}_1\hat{\boldsymbol{\beta}} + \mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\tau}} = \mathbf{X}_1'\mathbf{y}$$
$$\mathbf{X}_2'\mathbf{X}_1\hat{\boldsymbol{\beta}} + \mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\tau}} = \mathbf{X}_2'\mathbf{y}$$

$$\begin{aligned}\mathbf{X}_1'\mathbf{X}_1\hat{\boldsymbol{\beta}} + \mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\tau}} &= \mathbf{X}_1'\mathbf{y}\\ \mathbf{X}_2'\mathbf{X}_1\hat{\boldsymbol{\beta}} + \mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\tau}} &= \mathbf{X}_2'\mathbf{y}\end{aligned}$$

- Pre-multiply first equation by $\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^-$, to match leading 1st term of second equation
  - will need g-inverse if $\mathbf{X}_1$ is of less than full rank

$$\mathbf{X}_2'\mathbf{X}_1\hat{\boldsymbol{\beta}} + \mathbf{X}_2'\mathbf{H}_1\mathbf{X}_2\hat{\boldsymbol{\tau}} = \mathbf{X}_2'\mathbf{H}_1\mathbf{y}$$

- Subtract this from the second equation to get "reduced normal equations":

$$\mathbf{X}_2'(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2\hat{\boldsymbol{\tau}} = \mathbf{X}_2'(\mathbf{I} - \mathbf{H}_1)\mathbf{y}$$

- Pattern of "reduced" normal equations:
  - Same form found in weighted regression ... there $\mathbf{I} - \mathbf{H}_1$ is replaced with a "weight" matrix, any matrix proportional to $Var(\mathbf{y})$
  - $\mathbf{X}_2$-model projected into null space of $\mathbf{X}_1$
  - Same as regression of resid($\mathbf{y} : \mathbf{X}_1$) on resid($\mathbf{X}_2 : \mathbf{X}_1$)
  - Note reliance of form on $\mathbf{H}_1\mathbf{X}_2$ and $(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2$...

- In fact, suppose we transform the independent variables in this problem, pretending the model isn't really partitioned, and that the single model matrix is:

$$\mathbf{X}_{2|1} = (\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2$$

  - then the *usual* normal equations are:

$$\mathbf{X}'_{2|1}\mathbf{X}_{2|1}\hat{\boldsymbol{\tau}} = \mathbf{X}'_{2|1}\mathbf{y}$$

  - i.e. the correct result ... we've "corrected for $\boldsymbol{\beta}$" by "projecting $\mathbf{X}_2$ *out* of" the space spanned by $\mathbf{X}_1$
  - so, for example,
    * $\mathbf{c}'\boldsymbol{\tau}$ is estimable iff $\mathbf{c}' = \mathbf{l}'\mathbf{X}_{2|1}$
    * $Var(\widehat{\mathbf{c}'\boldsymbol{\tau}}) = \sigma^2\mathbf{c}'(\mathbf{X}'_{2|1}\mathbf{X}_{2|1})^{-}\mathbf{c}$

$\mathbf{X}'_{\mathbf{2|1}}\mathbf{X}_{\mathbf{2|1}}$ is called the "design information matrix" in the book, and sometimes denoted $\mathcal{I}_{2|1}$. For CRD's:

- $\mathbf{H}_1 = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' = \frac{1}{n}\mathbf{J}_{n\times n}$
- $\mathbf{X}_{2|1} = (\mathbf{I} - \mathbf{H}_1)\mathbf{X}_2 =$

$$\begin{pmatrix} (1-\frac{n_1}{N})\mathbf{1}_{n_1} & -\frac{n_2}{N}\mathbf{1}_{n_1} & \dots & -\frac{n_t}{N}\mathbf{1}_{n_1} \\ -\frac{n_1}{N}\mathbf{1}_{n_2} & (1-\frac{n_2}{N})\mathbf{1}_{n_2} & \dots & -\frac{n_t}{N}\mathbf{1}_{n_2} \\ \dots & \dots & \dots & \dots \\ -\frac{n_1}{N}\mathbf{1}_{n_t} & -\frac{n_2}{N}\mathbf{1}_{n_t} & \dots & (1-\frac{n_t}{N})\mathbf{1}_{n_t} \end{pmatrix}$$

- $\mathcal{I}_{2|1} = \mathbf{X}'_{2|1}\mathbf{X}_{2|1} = \text{diag}(\mathbf{n}) - \frac{1}{N}\mathbf{n}\mathbf{n}'$, where $\mathbf{n}' = (n_1, n_2, ..., n_t)$
- Note that $\mathbf{X}_{2|1}$ and $\mathcal{I}_{2|1}$ are each of rank $t-1$, reflecting the fact that $\boldsymbol{\tau}$ isn't estimable

So, for CRD's, given interest in $\boldsymbol{\tau}$ and accounting for α,

- the precision of $\widehat{\boldsymbol{c}'\boldsymbol{\tau}}$ is determined by $\sigma^2\mathbf{c}'\mathcal{I}_{2|1}^{-}\mathbf{c}$
  - one g-inverse of $\mathcal{I}_{2|1} = \text{diag}(\mathbf{n}) - \frac{1}{N}\mathbf{n}\mathbf{n}'$ is $\mathcal{I}_{2|1}^{-} = \text{diag}^{-1}(\mathbf{n})$ ... check it
- the power of the test for equality of τ's is determined by $\boldsymbol{\tau}'\mathcal{I}_{2|1}\boldsymbol{\tau}/\sigma^2$