Constructing Two-Level Trend Robust Designs

Robert W. Mee
University of Tennessee

rmee@utk.edu

Outline

- Choice of run order – Does it matter?
- Quote from 1966
- Simple example with drift in the errors
- Daniel and Wilcoxon's systematic run order
- Blocking versus run order covariates
- Sorting on two or three contrasts
- Examples – 2^{8-3}, 2^{13-9}, 12-run Plackett-Burman
- Summary comments

2^k Designs with Sequential Runs

- Consider experiments with runs performed in sequence, e.g., one run per day for 16 days
- No identifiable basis for grouping sets of runs – other than closeness in time
- How do we determine the run order?
  - Unrestricted randomization?
  - Blocking, with randomization within blocks?
  - A systematically determined run order?

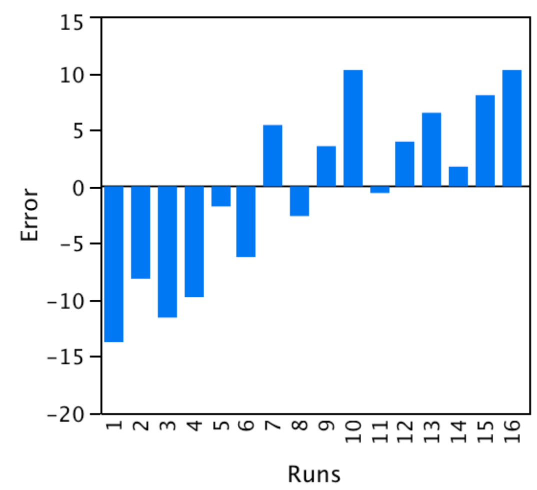
Does it matter which?

Insight from Daniel and Wilcoxon

Only a small fraction of all experimental work in the physical and engineering sciences meets the orthodox design statistician's requirements for objective randomization. How does it happen then that a considerable part of all this work produces useful, even valid, results? It happens because randomization, while generally sufficient, is not always necessary... almost as soon as (randomization) is urged, sometimes dogmatically, the demand is modified and for obvious reasons. Full randomization would in many situations guarantee results which, while entirely valid, would not detect any effects.... Randomization is used, then, after we have exhausted our knowledge of the behavior of the system under study and have taken serious steps to control what can be controlled.

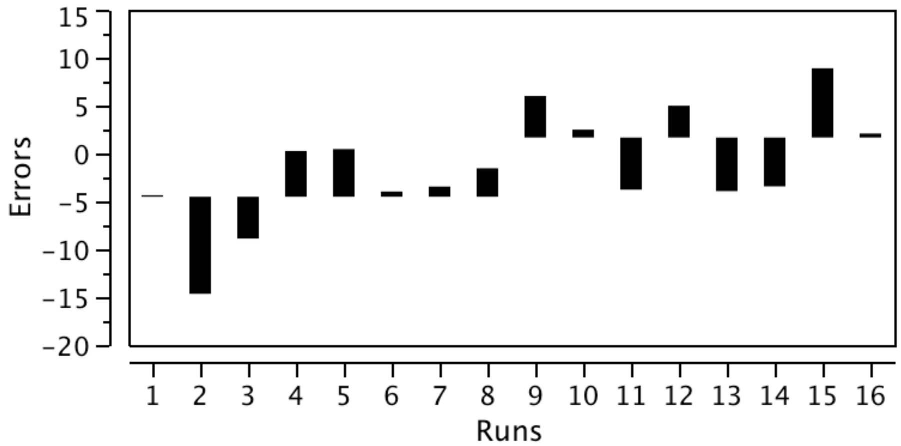
Suppose there is a drift in the errors

- To illustrate, suppose we have a 16-run design such as an unreplicated 2^4.
- Let the error contain two components:
  - IID deviations with mean zero and variance 25.
  - Drift = (Run number – 8.5)
- What is the effect of complete randomization?
  - An error variance of 46, nearly doubled by the drift!



Blocking Reduces the Variance Due to Drift

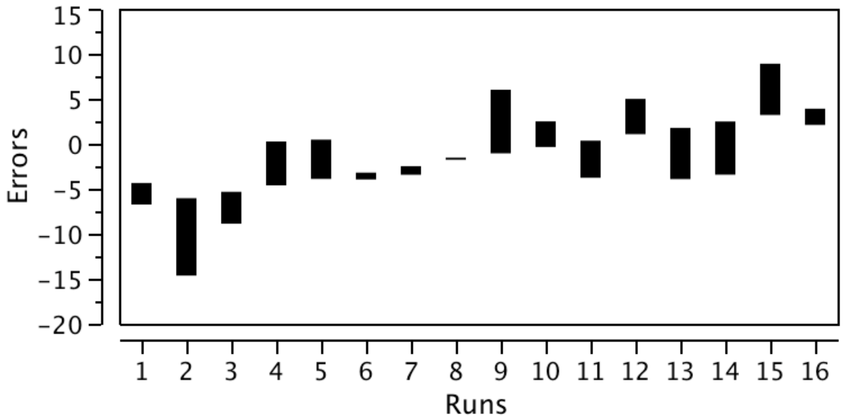
- Two blocks of size 8
  - Errors are more consistent within blocks
  - The residual variance is reduced to about 30



  - This gain is achieved by sacrificing info for “ABCD”

Using “Run” as a Covariate

- Including “Run” as a covariate would reduced the unexplained error variation to 25, which is better than with blocks



- But this covariate will be correlated with many factorial effect estimates.
- What run order will avoid such correlations??

Daniel and Wilcoxon (1966 *Techno.*)

- Sort the runs of a 2^n factorial – [or $2^{(n+p)-p}$] – by n factorial effect contrasts $c_1, c_2, \ldots, c_n$.
  - Only these n effects are correlated with a linear covariate, Run.
  - Only n(n-1)/2 effects are correlated with an (orthogonal) quadratic component $(\text{Run} - \text{mean})^2$
  - The remaining effect contrasts are orthogonal to Run and Run^2
- For example, n = 5: 5+10 effects correlated with linear or quadratic trend; 16 are orthogonal

D&W (1966) Examples

- 2^4 and 2^5
  - The linear covariate Run is most correlated with c_1
  - The quadratic covariate (Run – mean)^2 is most correlated with c_1c_2
- 2^{6-2} and 2^{14-9}
  - These two resolution IV designs have two degrees of freedom for 3-factor interactions
  - Choose the c_j's so that c_1 and c_1c_2 match these 3-factor interactions

Nominal Block Effect or 2 Covariates

- Four blocks are created by sorting on c_1 and c_2:
  - One may include in the fitted model a term for "Blocks" (with 3 d.f.), which will explain:
    - > 93.75% of the variance due to any linear trend,
    - > 70.3% of the variance due to any quadratic component in the errors.
  - The other factorial effects may be slightly biased (for any particular randomization) due to within-block trends

Block Effect or Two Covariates

- If one creates a run order by sorting on c_1, c_2, c_3, …
  - One may include Run and $(\text{Run} - \text{mean})^2$ in the model as covariates
  - Only c_1 and c_1c_2 need be dropped from the model to allow for these covariates.
  - The covariates effectively remove linear and quadratic trend in the errors, but cause some minor loss of efficiency (VIFs ≤ 1.25) primarily in two other effect estimates.

Covariates and the Choice of c_3

- Sort on c_1, then c_2, then c_3
- Replace c_1 with Run, c_1c_2 with (Run – mean)^2
- The two largest VIFs (1.25) in a saturated model are for c_2 and c_1c_3
- What about the other 3 contrasts?
  - c_3 has a very weak correlation with Run;
  - c_2c_3 has a very weak correlation with (Run – mean)^2
  - $c_1c_2c_3$ is orthogonal to both
- Choose (c_1, c_2, c_3) with this in mind.

Example 1: Minimum Aberration 2^{8-3}

- This is resolution IV (but not even)
  - Basic factors A, B, C, D, E
  - Generated factors F = CDE, G = ABDE, H = ABCE
- There are 3 degrees of freedom not devoted to main effects and two-factor interactions
  - ACD, BCD, ABCD
- Systematic design run orders:

  Option 1: c_1 = ACD, c_2 = AB, c_3 = ABCD

  Option 2: c_1 = ABCD, c_2 = A, c_3 = B

2^{8-3} Fraction with First Run Order

- Option 1: Sort on c_1 = ACD, c_2 = AB, c_3 = ABCD
- Omit ACD (c_1) and BCD (c_1c_2)
- Include Run and $(\text{Run}-16.5)^2$ as covariates
- This design
  - Has minimum aberration
  - Provides 28 degrees of freedom for main effects and two-factor interactions
  - Has slightly lower precision for AB (c_2) and B (c_1c_3) due to correlations with covariates

2^{8-3} Fraction with Second Run Order

- Option 1: Sort on : c_1 = ABCD, c_2 = A, c_3 = B
- Omit ABCD (c_1) and BCD (c_1c_2)
- Include Run and $(\text{Run-16.5})^2$ as covariates
- This design
  - Has minimum aberration
  - Provides 28 degrees of freedom for main effects and two-factor interactions
  - Has slightly lower precision for A (c_2) and ACD (c_1c_3) due to correlations with covariates

Consider the Alternative Blocked 2^{8-3}

- This is resolution IV (but not even)
  - Basic factors A -- E
  - Generated factors F = CDE, G = ABDE, H = ABCE
- Confound ACD, BCD, and AB with blocks
- If we arrange blocks by sorting on ABD and then AB:
  - This design is akin to the first systematic order
  - But we consider AB to be estimable

Example 2: 2^{13-9}

- Consider the design for which AB and AC are not confounded with any main effects.
- With only 2 d.f. after including main effects, we cannot construct a design with 4 blocks
- We can sort on c_1 = AB and c_2 = BC = N, so that AB and AC are most correlated with Run and $(\text{Run} - 8.5)^2$
- Fit a model with 13 main effects and these two covariates. If no evidence for trend appears, fit a model with 15 orthogonal effects.

Example 3: OA(12, 2^{10}, 2)

- Sort the rows of this Plackett-Burman design by the idle column X_{11} and then by any two-factor interaction column
- Fit a saturated model with 10 main effects and a covariate for Run.
- The largest VIF for a factor main effect is 1.15, and the rest are smaller than 1.05.

Summary Remarks

- For smaller designs, requiring robustness to linear trend should suffice.
  - For larger designs, choose a run order such that Run and $(\text{Run-mean})^2$ have no strong correlations with effects of interest.
- These systematic run orders are often similar to designs constructed with four blocks, although systematic run order designs are more flexible.
- Systematic run order designs should be utilized for applications with a likely trend in the errors.