**Issues in designing the experiments with 6 treatments, 3 biological replicates and 2 technical replicates assigning to 9 runs of 4-plex experiment**

This write-up describes an implication for finding an optimal design for one specific set of design parameters. The design parameters are

Phase 1 experiment - 6 treatments, 3 biological replicates, 2 technical replicates,

Phase 2 experiment – 9 runs and 4 tags.

For the Phase 1 experiment, the 6 treatments are denoted by “a”, “b”, “c”, “d”, “e” and “f”. Since 3 biological replicates are used, this means 3 animals are assigned to each treatment which gives a total of 15 animals. These 15 animals are denoted by upper case letters of “A” to “R”. The theoretical ANOVA of the Phase 1 experiment can be presented as follows,

$ANOVA

DF Ani

Between Ani

Trt 5 1

Residual 12 1

$EF

Trt eff.Trt

Between Ani

Trt 3 1

Note that all treatment information is in the between animals stratum.

The pattern (which I have been using) for assigning the animals to the runs and tags is to group pair of animals and allocating them in a quadrant of 2 runs and 2 tags. For this case, the total number of runs needed is 9; hence, the last pair of animals is assigned to the last run. The pair of animas can be Animals “A” and “B”, Animals “C” and “D” to Animals “Q” and “R”. The allocation of the animals to runs and tags can be shown as follows,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Run | Tag | | | |
| 114 | 115 | 116 | 117 |
| 1 | A | B | C | D |
| 2 | B | A | D | C |
| 3 | E | F | G | H |
| 4 | F | E | H | G |
| 5 | I | J | K | L |
| 6 | J | I | L | K |
| 7 | M | N | O | P |
| 8 | N | M | P | O |
| 9 | Q | Q | R | R |

The bold box in this design represents each pair of the animals. Note that the animal is confounded with both runs and tags. More specifically, the animal is confounded with a tag contrast of 114, 115 versus 116, 117. For the relationship between runs and animals, the runs can be separated into 5 groups according to the pairs of animals that are assigned. This means 4 DF associated with the animals are confounded with the runs, or we can also say that 4Df associated with the animals should be in the between runs stratum.

The treatment allocation to runs and tags is based on the assignments of treatments to animals of the Phase 1 experiments. The treatment design is shown as follows,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Run | Tag | | | |
| 114 | 115 | 116 | 117 |
| 1 | a | b | c | d |
| 2 | b | a | d | c |
| 3 | e | f | a | b |
| 4 | f | e | b | a |
| 5 | c | d | e | f |
| 6 | d | c | f | e |
| 7 | a | b | c | d |
| 8 | b | a | d | c |
| 9 | e | e | f | f |

The bold box in this design represents each pair of treatments from the pair of the animals. The treatment is also confounded with both runs and tags. The treatment is also confounded with the confounded with a tag contrast of 114, 115 versus 116, 117.

The theoretical ANOVA table for this design can be written as follows,

$ANOVA

DF e Ani Run

Between Run

Between Ani

Trt 2 1 2 4

Residual 2 1 2 4

Residual 4 1 0 4

Within

Between Ani

Tag 1 1 2 0

Trt 5 1 2 0

Residual 7 1 2 0

Residual

Tag 2 1 0 0

Residual 12 1 0 0

$EF

Tag Trt eff.Tag eff.Trt

Between Run

Between Ani

Trt 3/2 1/4

Residual

Within

Between Ani

Tag 9 2/3 1 1/9

Trt 3540/751 590/751

Residual

Tag 9 1

A valid test for the treatment differences can be conducted based on this theoretical ANOVA table, because the coefficients of the between animals variance components of the treatment and residual mean squares are identical in the between animals within runs stratum. However, from the fixed effects table, there is 1/4 of treatment information in the between runs and 1/9 of treatment information confounded with tags; hence, there is (590/751 =) 0.7856192 of pure treatment information remaining based on animals and treatments allocation to runs and tags described above.

We can then study the concurrence matrix of treatments and runs

Trt

Trt a b c d e f

a 6 6 4 4 2 2

b 6 6 4 4 2 2

c 4 4 6 6 2 2

d 4 4 6 6 2 2

e 2 2 2 2 8 8

f 2 2 2 2 8 8

Note that due to the grouping of the pairs of animals and treatments, it causes the treatment “e” and “f” to appear 8 times together in the same run. In addition, both treatment “e” and “f” only appear together with the other treatments twice in the same run. We may be able to increase the average efficiency factor of treatment by spreading the treatments “e” and “f” by pairing them to the other treatments while assigning the animals and treatments to runs and tags. It may also decrease the average efficiency factor of treatment in the between runs stratum.

Using the simulated annealing algorithm, another design with a higher average efficiency factors of treatment in the between animals within runs stratum was found. The allocation of animals to the runs and tags is as follows,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Run | Tag | | | |
| 114 | 115 | 116 | 117 |
| 1 | A | B | C | D |
| 2 | B | A | D | C |
| 3 | E | F | G | H |
| 4 | F | E | H | G |
| 5 | J | K | N | O |
| 6 | K | J | O | N |
| 7 | M | I | Q | L |
| 8 | I | M | L | Q |
| 9 | P | P | R | R |

The pattern of assignment is still the same, which is grouping pair of animals and allocate them in the 2-by-2 quadrants. The last pair of animals is then assigned to the last run. Comparing to the previous design, the first four runs are identical. The remaining runs indicate different pairing of the animals of “J” and “K”, “O” and “N”, “M” and “I”, “Q” and “L”, and “P” and “R”. The treatment allocation can then be shown as follows,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Run | Tag | | | |
| 114 | 115 | 116 | 117 |
| 1 | a | b | c | d |
| 2 | b | a | d | c |
| 3 | e | f | a | b |
| 4 | f | e | b | a |
| 5 | d | e | b | c |
| 6 | e | d | c | b |
| 7 | a | c | e | f |
| 8 | c | a | f | e |
| 9 | d | d | f | f |

From previous design, the pairs of treatments were “a” and “b”, “c” and “d”, and “e” and “f”. For the new design, the treatment pairs are not identical as appeared in last 5 runs. The different treatment pairs of last five runs are “d” and “e”, “b” and “c”, “a” and “c”, and “d” and “f”. Another important point apart from the treatment pairing is the treatments assigned in the same runs. For the previous design, treatment “e” always appears in the same runs as the treatment “f” which is also noted from the concurrence matrix.

We will start looking at the concurrence matrix of treatments and runs, which is follows

Trt

Trt a b c d e f

a 6 4 4 2 4 4

b 4 6 4 4 4 2

c 4 4 6 4 4 2

d 2 4 4 8 2 4

e 4 4 4 2 6 4

f 4 2 2 4 4 8

Note the treatments “d” and “f” are the treatments assigned in the last runs which is equivalent to the treatment “e” and “f” of the previous design. For the new design, these two treatments only appear 4 times together in the same run, unlike the previous design, the treatments “e” and “f” were appearing 8 times together in the same run. In addition, for the previous design, both treatment “e” and “f” only appear together with the other treatments twice in the same run. For the new design, the treatment “d” appears twice in the same run as treatment “a” and ”e” and four times in the same runs as treatment “b” and “c”. As for treatment “f” of the new design, it appears twice in the same run as treatment “b” and ”c” and four times in the same runs as treatment “a” and “e”. Therefore, the pair of treatment in the last run has now been spread to pair with the other treatments for the new design. Now let’s look at the theoretical ANOVA table of the new design which can be shown as follows,

$ANOVA

DF e Ani Run

Between Run

Between Ani

Trt 4 1 2 4

Residual 4 1 0 4

Within

Between Ani

Tag 1 1 2 0

Trt 5 1 2 0

Residual 7 1 2 0

Residual

Tag 2 1 0 0

Residual 12 1 0 0

$EF

Tag Trt eff.Tag eff.Trt

Between Run

Between Ani

Trt 3/4 1/8

Residual

Within

Between Ani

Tag 9 2/3 1 1/9

Trt 7920/1577 1320/1577

Residual

Tag 9 1

The random effects table of the new design is identical to the random effects table of the previous design. Therefore, a valid test for the treatment differences can still be conducted. From the fixed effects table of the new design, the amount of treatment information in the between runs stratum become 1/8 which is lower than before. There is still 1/9 of treatment information confounded with the tag. This means there is (1320/1577 =) 0.8370 of pure treatment information remaining in the between animals within runs stratum, that is 0.051 more than the previous design.

Therefore, this suggests that the new design is better than the previous design. In addition, this also suggests that for experiment with 6 treatments with 3 (and may be also for odd number) biological replicates, I cannot use the same pair of animals and treatments when assigning them to the all runs and tags. The treatment pairs have to be different to minimise the confounding of treatment with the runs, which will also maximise the treatment information in the between animals within runs stratum.

I now want to construct the

Let slightly different design

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Run | Tag | | | |
| 114 | 115 | 116 | 117 |
| 1 | A | B | C | F |
| 2 | B | A | F | C |
| 3 | E | F | G | H |
| 4 | F | E | H | G |
| 5 | I | J | K | L |
| 6 | J | I | L | K |
| 7 | M | N | O | P |
| 8 | N | M | P | O |
| 9 | Q | Q | R | R |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Run | Tag | | | |
| 114 | 115 | 116 | 117 |
| 1 | a | b | c | f |
| 2 | b | a | f | c |
| 3 | e | d | a | b |
| 4 | d | e | b | a |
| 5 | c | d | e | b |
| 6 | d | c | b | e |
| 7 | a | f | c | d |
| 8 | f | a | d | c |
| 9 | e | e | f | f |

$ANOVA

DF e Ani Run

Between Run

Between Ani

Trt 4 1 2 4

Residual 4 1 0 4

Within

Between Ani

Tag 1 1 2 0

Trt 5 1 2 0

Residual 7 1 2 0

Residual

Tag 2 1 0 0

Residual 12 1 0 0

$EF

Tag Trt eff.Tag eff.Trt

Between Run

Between Ani

Trt 3/4 1/8

Residual

Within

Between Ani

Tag 9 2/3 1 1/9

Trt 66513/13333 22171/26666

Residual

Tag 9 1

(22171/26666 =) 0.8314

slightly different design

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Run | Tag | | | |
| 114 | 115 | 116 | 117 |
| 1 | A | B | C | F |
| 2 | B | A | F | C |
| 3 | E | F | G | H |
| 4 | F | E | H | G |
| 5 | I | J | K | L |
| 6 | J | I | L | K |
| 7 | M | N | O | P |
| 8 | N | M | P | O |
| 9 | Q | Q | R | R |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Run | Tag | | | |
| 114 | 115 | 116 | 117 |
| 1 | a | b | c | f |
| 2 | b | a | f | c |
| 3 | e | d | a | b |
| 4 | d | e | b | a |
| 5 | f | d | c | b |
| 6 | d | f | b | c |
| 7 | a | c | e | d |
| 8 | c | a | d | e |
| 9 | f | f | e | e |

$ANOVA

DF e Ani Run

Between Run

Between Ani

Trt 4 1 2 4

Residual 4 1 0 4

Within

Between Ani

Tag 1 1 2 0

Trt 5 1 2 0

Residual 7 1 2 0

Residual

Tag 2 1 0 0

Residual 12 1 0 0

$EF

Tag Trt eff.Tag eff.Trt

Between Run

Between Ani

Trt 3/4 1/8

Residual

Within

Between Ani

Tag 9 2/3 1 1/9

Trt 7920/1577 1320/1577

Residual

Tag 9 1

(1320/1577 =) 0.8370

The similarity between the last design and the second design has been founded by representing the grouping of the treatment pairs using the graphs. Each vertex represents the treatment. The grouping of treatment pairs are represented by the edges. This can be shown as follows

The above graph can also be used for different treatment pairing.

I have found another graph with for generating optimal designs with same design parameters.

[,1] [,2] [,3] [,4]

[1,] "AA" "AB" "AD" "AF"

[2,] "AB" "AA" "AF" "AD"

[3,] "AJ" "AL" "AG" "AI"

[4,] "AL" "AJ" "AI" "AG"

[5,] "AC" "AP" "AH" "AE"

[6,] "AP" "AC" "AE" "AH"

[7,] "AM" "AK" "AN" "AO"

[8,] "AK" "AM" "AO" "AN"

[9,] "AQ" "AQ" "AR" "AR"

> matrix(design.df$Trt, nrow = nBlk, ncol = nPlot, byrow = TRUE)

[,1] [,2] [,3] [,4]

[1,] "a" "b" "d" "f"

[2,] "b" "a" "f" "d"

[3,] "d" "f" "a" "c"

[4,] "f" "d" "c" "a"

[5,] "c" "d" "b" "e"

[6,] "d" "c" "e" "b"

[7,] "a" "e" "b" "c"

[8,] "e" "a" "c" "b"

[9,] "e" "e" "f" "f"

The next experiment with new set of design parameters is as follows

Phase 1 experiment - 6 treatments, 5 biological replicates, 2 technical replicates,

Phase 2 experiment – 15 runs and 4 tags.

> summary.aov.twoPhase(design.df, blk.str2 = "Run", blk.str1 = "Ani",

+ trt.str = "Tag + Trt")

$ANOVA

DF e Ani Run

Between Run

Between Ani

Trt 5 1 2 4

Residual 2 1 2 4

Residual 7 1 0 4

Within

Between Ani

Tag 1 1 2 0

Trt 5 1 2 0

Residual 16 1 2 0

Residual

Tag 2 1 0 0

Residual 21 1 0 0

$EF

Tag Trt eff.Tag eff.Trt

Between Run

Between Ani

Trt 10/13 1/13

Residual

Within

Between Ani

Tag 15 2/5 1 1/25

Trt 184075/21193 11743/13520

Residual

Tag 15 1

> matrix(design.df$Ani, nrow = nBlk, ncol = nPlot, byrow = TRUE)

[,1] [,2] [,3] [,4]

[1,] "AP" "AI" "AM" "AK"

[2,] "AI" "AP" "AK" "AM"

[3,] "AG" "AD" "AT" "AU"

[4,] "AD" "AG" "AU" "AT"

[5,] "AW" "AN" "AC" "AS"

[6,] "AN" "AW" "AS" "AC"

[7,] "AB" "AY" "AX" "BC"

[8,] "AY" "AB" "BC" "AX"

[9,] "BD" "AE" "AV" "AZ"

[10,] "AE" "BD" "AZ" "AV"

[11,] "AQ" "AR" "BB" "AO"

[12,] "AR" "AQ" "AO" "BB"

[13,] "AJ" "BA" "AF" "AH"

[14,] "BA" "AJ" "AH" "AF"

[15,] "AA" "AA" "AL" "AL"

> matrix(design.df$Trt, nrow = nBlk, ncol = nPlot, byrow = TRUE)

[,1] [,2] [,3] [,4]

[1,] "d" "c" "a" "e"

[2,] "c" "d" "e" "a"

[3,] "a" "d" "b" "c"

[4,] "d" "a" "c" "b"

[5,] "e" "b" "c" "a"

[6,] "b" "e" "a" "c"

[7,] "b" "a" "f" "e"

[8,] "a" "b" "e" "f"

[9,] "f" "e" "d" "b"

[10,] "e" "f" "b" "d"

[11,] "e" "f" "d" "c"

[12,] "f" "e" "c" "d"

[13,] "d" "c" "f" "b"

[14,] "c" "d" "b" "f"

[15,] "a" "a" "f" "f"

Phase 1 experiment - 6 treatments, 7 biological replicates, 2 technical replicates,

Phase 2 experiment – 21 runs and 4 tags.

$ANOVA

DF e Ani Run

Between Run

Between Ani

Trt 5 1 2 4

Residual 5 1 2 4

Residual 10 1 0 4

Within

Between Ani

Tag 1 1 2 0

Trt 5 1 2 0

Residual 25 1 2 0

Residual

Tag 2 1 0 0

Residual 30 1 0 0

$EF

Tag Trt eff.Tag eff.Trt

Between Run

Between Ani

Trt 10/7 5/49

Residual

Within

Between Ani

Tag 21 2/7 1 1/49

Trt 62439/5066 1641/1864

Residual

Tag 21 1

> matrix(design.df$Ani, nrow = nBlk, ncol = nPlot, byrow = TRUE)

[,1] [,2] [,3] [,4]

[1,] "BD" "AS" "BI" "BA"

[2,] "AS" "BD" "BA" "BI"

[3,] "BF" "AX" "AP" "BO"

[4,] "AX" "BF" "BO" "AP"

[5,] "AI" "AK" "AB" "AA"

[6,] "AK" "AI" "AA" "AB"

[7,] "BN" "AE" "BM" "AM"

[8,] "AE" "BN" "AM" "BM"

[9,] "AW" "AZ" "AR" "BE"

[10,] "AZ" "AW" "BE" "AR"

[11,] "BL" "BG" "BH" "BC"

[12,] "BG" "BL" "BC" "BH"

[13,] "BK" "AN" "AD" "AF"

[14,] "AN" "BK" "AF" "AD"

[15,] "AU" "BB" "BP" "AH"

[16,] "BB" "AU" "AH" "BP"

[17,] "AC" "AY" "AT" "BJ"

[18,] "AY" "AC" "BJ" "AT"

[19,] "AV" "AL" "AQ" "AO"

[20,] "AL" "AV" "AO" "AQ"

[21,] "AG" "AG" "AJ" "AJ"

> matrix(design.df$Trt, nrow = nBlk, ncol = nPlot, byrow = TRUE)

[,1] [,2] [,3] [,4]

[1,] "f" "a" "e" "c"

[2,] "a" "f" "c" "e"

[3,] "b" "f" "d" "e"

[4,] "f" "b" "e" "d"

[5,] "c" "e" "b" "a"

[6,] "e" "c" "a" "b"

[7,] "d" "e" "c" "a"

[8,] "e" "d" "a" "c"

[9,] "e" "b" "f" "a"

[10,] "b" "e" "a" "f"

[11,] "b" "c" "d" "e"

[12,] "c" "b" "e" "d"

[13,] "a" "b" "d" "f"

[14,] "b" "a" "f" "d"

[15,] "c" "d" "f" "b"

[16,] "d" "c" "b" "f"

[17,] "c" "a" "b" "f"

[18,] "a" "c" "f" "b"

[19,] "d" "f" "e" "c"

[20,] "f" "d" "c" "e"

[21,] "a" "a" "d" "d"

Phase 1 experiment - 6 treatments, 9 biological replicates, 2 technical replicates,

Phase 2 experiment – 27 runs and 4 tags.

I have been able to speed up the simulated annealing algorithm by adjusting the swapping methods.

The new swapping method now swaps the pair of treatments, this avoid destroying the patterns.