Experimental Designs

James Dann

September 3, 2021

Context

- Weigh 8 different items with unreliable pan scales

Context

- Weigh 8 different items with unreliable pan scales
- As few instances of measuring as possible
- As high an accuracy as possible

Unreliable?

- Each trial gives us a (independent) random variable w_i

Unreliable?

- Each trial gives us a (independent) random variable w_i
- Let the true mass of object i be θ_i

Unreliable?

- Each trial gives us a (independent) random variable w_i
- Let the true mass of object i be θ_i
- Let the estimate of the mass of object i be $\hat{\theta}_i$

Naive Scheme

Left	Right	Result
1	-	w_1
2	-	w_2
3	-	w_3
4	-	w_4
5	-	w_5
6	-	w_6
7	-	w_7
8	-	w_8

Naive Scheme

Left	Right	Result
1	-	w_1
2	-	w_2
3	-	w_3
4	-	w_4
5	-	w_5
6	-	w_6
7	-	w_7
8	-	w_8

- 8 trials total
- $\hat{\theta}_i = w_i$
- $\mu(\hat{\theta}_i) = \theta_i$
- $\text{Var}(\hat{\theta}_i) = \sigma^2$

Still Naive Scheme

Left	Right	Result
1	-	w_1
2	-	w_2
3	-	w_3
4	-	w_4
5	-	w_5
6	-	w_6
7	-	w_7
8	-	w_8
1	-	v_1
2	-	v_2
3	-	v_3
4	-	v_4
5	-	v_5
6	-	v_6
7	-	v_7
8	-	v_8

- 16 trials total
- $\hat{\theta}_i = \frac{w_i + v_i}{2}$
- $\mu(\hat{\theta}_i) = \theta_i$
- $\mathrm{Var}(\hat{\theta}_i) = \frac{\sigma^2}{2}$

Some Magical Numbers

$$S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

1248	3567
2358	1467
3468	1257
4578	1236
5618	2347
6728	1345
7138	2456

Some Magical Numbers

$$S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

1248	3567
2358	1467
3468	1257
4578	1236
5618	2347
6728	1345
7138	2456

- It can be verified that each triple occurs in exactly one cell

Some Magical Numbers

$$S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

1248	3567
2358	1467
3468	1257
4578	1236
5618	2347
6728	1345
7138	2456

- It can be verified that each triple occurs in exactly one cell

Some Magical Numbers

$$S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

1248	3567
2358	1467
3468	1257
4578	1236
5618	2347
6728	1345
7138	2456

- It can be verified that each triple occurs in exactly one cell
- Each pair occurs in exactly three cells

Some Magical Numbers

$$S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

1248	3567
2358	1467
3468	1257
4578	1236
5618	2347
6728	1345
7138	2456

- It can be verified that each triple occurs in exactly one cell
- Each pair occurs in exactly three cells
- It is clear that each row contains each number once

Some Magical Numbers

$$S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

1248	3567
2358	1467
3468	1257
4578	1236
5618	2347
6728	1345
7138	2456

- It can be verified that each triple occurs in exactly one cell
- Each pair occurs in exactly three cells
- It is clear that each row contains each number once
- Each pair occurs in the same cell in exactly 3 rows

Some Magical Numbers

$$S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

1248	3567
2358	1467
3468	1257
4578	1236
5618	2347
6728	1345
7138	2456

- It can be verified that each triple occurs in exactly one cell
- Each pair occurs in exactly three cells
- It is clear that each row contains each number once
- Each pair occurs in the same cell in exactly 3 rows
- Each pair occurs in different cells in exactly 4 rows

Improved? Scheme

Left	Right	Result
12345678	-	w_0
1248	3567	w_1
2358	1467	w_2
3468	1257	w_3
4578	1236	w_4
5618	2347	w_5
6728	1345	w_6
7138	2456	w_7

- 8 trials total
- $\hat{\theta}_i =$???
- $\mathrm{Var}(\hat{\theta}_i) =$???

Improved? Scheme

Left	Right	Result
12345678	-	w_0
1248	3567	w_1
2358	1467	w_2
3468	1257	w_3
4578	1236	w_4
5618	2347	w_5
6728	1345	w_6
7138	2456	w_7

Let $\hat{\theta}_3 = \frac{1}{8}(w_0 - w_1 + w_2 + w_3 - w_4 - w_5 - w_6 + w_7)$

Improved? Scheme

Left	Right	Result
12345678	-	w_0
1248	3567	w_1
2358	1467	w_2
3468	1257	w_3
4578	1236	w_4
5618	2347	w_5
6728	1345	w_6
7138	2456	w_7

Let $\hat{\theta}_3 = \frac{1}{8}(w_0 - w_1 + w_2 + w_3 - w_4 - w_5 - w_6 + w_7)$

Improved? Scheme

Left	Right	Result
12345678	-	w_0
1248	3567	w_1
2358	1467	w_2
3468	1257	w_3
4578	1236	w_4
5618	2347	w_5
6728	1345	w_6
7138	2456	w_7

Let $\hat{\theta}_3 = \frac{1}{8}(w_0 - w_1 + w_2 + w_3 - w_4 - w_5 - w_6 + w_7)$

$$\begin{aligned}
\mu(\hat{\theta}_3) &= \frac{1}{8}(\mu(w_0) - \mu(w_1) + \dots) \\
&= \frac{1}{8}(\theta_3 + \dots - (-\theta_3 + \dots) + \dots) \\
&= \frac{8\theta_3 + \dots}{8} = \theta_3
\end{aligned}$$

Improved? Scheme

Left	Right	Result
12345678	-	w_0
1248	3567	w_1
2358	1467	w_2
3468	1257	w_3
4578	1236	w_4
5618	2347	w_5
6728	1345	w_6
7138	2456	w_7

Let $\hat{\theta}_3 = \frac{1}{8}(w_0 - w_1 + w_2 + w_3 - w_4 - w_5 - w_6 + w_7)$

$$\begin{aligned}\mathrm{Var}(\hat{\theta}_2) &= \mathrm{Var}((w_0 - w_1 + w_2 + w_3 - w_4 - w_5 - w_6 + w_7)/8) \\ &= \frac{1}{64}\sum_{i=0}^{7}\mathrm{Var}(w_i) \\ &= \frac{\sigma^2}{8}\end{aligned}$$

Improved Scheme

Left	Right	Result
12345678	-	w_0
1248	3567	w_1
2358	1467	w_2
3468	1257	w_3
4578	1236	w_4
5618	2347	w_5
6728	1345	w_6
7138	2456	w_7

- 8 trials total
- $\hat{\theta}_i = \frac{1}{8} \sum_j \text{left}(i,j) w_j$
- $\mu(\hat{\theta}_i) = \theta_i$
- $\text{Var}(\hat{\theta}_i) = \sigma^2/8$

So what, I have linear algebra

- Matrix inverses exist and I don't like combinatorics

So what, I have linear algebra

- Matrix inverses exist and I don't like combinatorics
- You are just solving $A\theta + \varepsilon = \omega$, which can be solved with $\hat{\theta} = A^{-1}\omega$
- We can randomly fill the scales and take an inverse

Why combinatorics

- Combinatorics provides a unique optimal solution in general

Why combinatorics

- Combinatorics provides a unique optimal solution in general
- If $A^{-1} = (b_{ij})$, then $\mathrm{Var}(\hat{\theta}_i) = \sigma^2 \sum_{j=0}^{7} b_{ij}^2$

Why combinatorics

- Combinatorics provides a unique optimal solution in general
- If $A^{-1} = (b_{ij})$, then $\mathrm{Var}(\hat{\theta}_i) = \sigma^2 \sum_{j=0}^{7} b_{ij}^2$
- Since A has entries in $\{-1, 1\}$, we must have $\sum_{j=0}^{7} |b_{ij}| \geq 1$

Why combinatorics

- Combinatorics provides a unique optimal solution in general
- If $A^{-1} = (b_{ij})$, then $\text{Var}(\hat{\theta}_i) = \sigma^2 \sum_{j=0}^{7} b_{ij}^2$
- Since A has entries in $\{-1, 1\}$, we must have $\sum_{j=0}^{7} |b_{ij}| \geq 1$
- The variance is minimised when all $b_{ij} = \pm\frac{1}{8}$

Why combinatorics

- Combinatorics provides a unique optimal solution in general
- If $A^{-1} = (b_{ij})$, then $\mathrm{Var}(\hat{\theta}_i) = \sigma^2 \sum_{j=0}^{7} b_{ij}^2$
- Since A has entries in $\{-1, 1\}$, we must have $\sum_{j=0}^{7} |b_{ij}| \geq 1$
- The variance is minimised when all $b_{ij} = \pm\frac{1}{8}$
- This is exactly the combinatorial solution

Generalisations

- The Hadamard conjecture implies this can be done for any $4n$

Generalisations

- The Hadamard conjecture implies this can be done for any $4n$
- Can construct a scheme for any $4n$ such that $4n - 1$ is a prime power
- Can construct a scheme for any product of integers that work

Context

- Efficiently testing samples for a hypothetical virus

Context

- Efficiently testing samples for a hypothetical virus
- AIDS, Zika, SARS, spanish flu, bubonic plague

Context

- Efficiently testing samples for a hypothetical virus
- AIDS, Zika, SARS, spanish flu, bubonic plague
- Negative test implies all samples are uninfected

Context

- Efficiently testing samples for a hypothetical virus
- AIDS, Zika, SARS, spanish flu, bubonic plague
- Negative test implies all samples are uninfected
- Assume that there are at most d infected samples

Context

- Efficiently testing samples for a hypothetical virus
- AIDS, Zika, SARS, spanish flu, bubonic plague
- Negative test implies all samples are uninfected
- Assume that there are at most d infected samples
- Testing scheme doesn't change depending on results

More Magic Numbers

$$S = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$$

0 1 2
3 4 5
6 7 8
0 4 8
1 5 6
2 3 7
0 7 5
1 8 3
2 6 4
0 3 6
1 4 7
2 5 8

- Easy to check that each pair occurs in exactly one cell
- The intersection of any two cells has at most one element

Group Testing Example

Sample	Tests
a	0 1 2
b	3 4 5
c	6 7 8
d	0 4 8
e	1 5 6
f	2 3 7
g	0 7 5
h	1 8 3
i	2 6 4
j	0 3 6
k	1 4 7
l	2 5 8

- 9 tests in total
- 12 samples tested
- Assume there are 2 or less positive samples

Group Testing Example

Sample	Tests
a	0 1 2
b	3 4 5
c	6 7 8
d	0 4 8
e	1 5 6
f	2 3 7
g	0 7 5
h	1 8 3
i	2 6 4
j	0 3 6
k	1 4 7
l	2 5 8

- 9 tests in total
- 12 samples tested
- Assume there are 2 or less positive samples

Group Testing Example

Sample	Tests
a	0 1 2
b	3 4 5
c	6 7 8
d	0 4 8
e	1 5 6
f	2 3 7
g	0 7 5
h	1 8 3
i	2 6 4
j	0 3 6
k	1 4 7
l	2 5 8

- 9 tests in total
- 12 samples tested
- Assume there are 2 or less positive samples

Group Testing Example

Sample	Tests
a	0 1 2
b	3 4 5
c	6 7 8
d	0 4 8
e	1 5 6
f	2 3 7
g	0 7 5
h	1 8 3
i	2 6 4
j	0 3 6
k	1 4 7
l	2 5 8

- 9 tests in total
- 12 samples tested
- Assume there are 2 or less positive samples

Group Testing Example

Sample	Tests
a	0 1 2
b	3 4 5
c	6 7 8
d	0 4 8
e	1 5 6
f	2 3 7
g	0 7 5
h	1 8 3
i	2 6 4
j	0 3 6
k	1 4 7
l	2 5 8

- 9 tests in total
- 12 samples tested
- Assume there are 2 or less positive samples
- Can detect if 3 or more samples are positive

Group Testing Counterexample

Sample	Tests
a	0 1 2
b	3 4 5
c	6 7 8
d	0 4 8
e	1 5 6
f	2 3 7
g	0 7 5
h	1 8 3
i	2 6 4
j	0 3 6
k	1 4 7
l	2 5 8

- 9 tests in total
- 12 samples tested
- Assume there are 2 or less positive samples
- Can detect if 3 or more samples are positive

Generalisations

- Can test $q^2 + q$ samples with q^2 tests, assuming a max of $q - 1$ positive samples.

Generalisations

- Can test $q^2 + q$ samples with q^2 tests, assuming a max of $q - 1$ positive samples.
- Can test $n(1 + 6n)$ or $\frac{(6n+3)(6n+2)}{6}$ samples with $6n + 1$ or $6n + 3$ tests, assuming a max of 2 positive samples.

Generalisations

- Can test $q^2 + q$ samples with q^2 tests, assuming a max of $q - 1$ positive samples.
- Can test $n(1 + 6n)$ or $\frac{(6n+3)(6n+2)}{6}$ samples with $6n + 1$ or $6n + 3$ tests, assuming a max of 2 positive samples.
- In general, any $(v, k, 1)$ design implies the existence of a group test of $\frac{v(v-1)}{k(k-1)}$ samples over v tests if there are at most $k - 1$ positive samples.