GGA training session 3 – Trial Protocols and Best Data Practices

Adam H. Sparks

Curtin Biometry and Agricultural Data Analytics

September 5, 2024

GRDC Grains Research & Development Corporation | Curtin University | The University Of Queensland Australia | The University of Adelaide

AAGI Analytics for the Australian Grains Industry

On-Farm Research Success

1. Ask the Right Question (refer to Mark's session this AM)
2. **Ask what data will be collected**
3. Consider paddock history and variation
4. Pick the right trial design (refer to Point 3 above & previous session)
    1. Include a control treatment for a baseline!
5. Replicate and randomise (refer to previous session)
6. **Collect the data**
7. Share the raw data with AAGI

Ensuring You Have Quality Data

1. Ensure the trial design is valid and fit for purpose.
2. Ensure the trial design is recorded correctly.
3. Identify and control sources of noise.
4. Implement protocols to avoid mistakes in running the trial and recording of observations.
5. **Always** provide the raw data to the biometrician.
6. Record ancillary details.

Don’t Optimise for Convenience

Time of Sowing 1

		Wheat														Barley								Oats				
Replicate 1	Buffer	Magenta	Magenta	Mace	Mace	Manning	Manning	Forrest	Forrest	Wedgetail	Wedgetail	Currawong	Currawong	Buffer	Buffer	Hindmarsh	Hindmarsh	Urambie	Urambie	Oxford	Oxford	Buffer	Buffer	Bannister	Bannister	Eurabbie	Eurabbie	Buffer
Replicate 2	Buffer	Manning	Manning	Wedgetail	Wedgetail	Currawong	Currawong	Magenta	Magenta	Mace	Mace	Forrest	Forrest	Buffer	Buffer	Urambie	Urambie	Oxford	Oxford	Hindmarsh	Hindmarsh	Buffer	Buffer	Eurabbie	Eurabbie	Bannister	Bannister	Buffer
Replicate 3	Buffer	Currawong	Currawong	Magenta	Magenta	Forrest	Forrest	Wedgetail	Wedgetail	Manning	Manning	Mace	Mace	Buffer	Buffer	Oxford	Oxford	Hindmarsh	Hindmarsh	Urambie	Urambie	Buffer	Buffer	Bannister	Bannister	Eurabbie	Eurabbie	Buffer

56 m

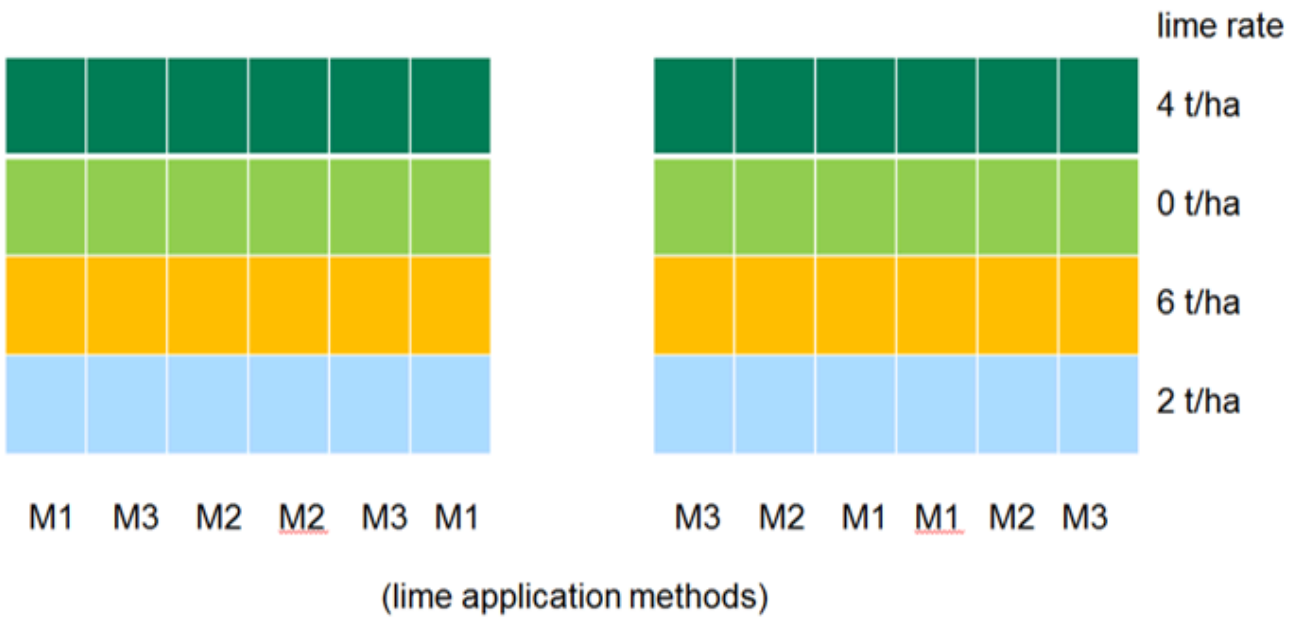
Time of Sowing 2

Replicate 1	Buffer	Currawong	Currawong	Mace	Mace	Wedgetail	Wedgetail	Forrest	Forrest	Magenta	Magenta	Manning	Manning	Buffer	Buffer	Urambie	Urambie	Oxford	Oxford	Hindmarsh	Hindmarsh	Buffer	Buffer	Bannister	Bannister	Eurabbie	Eurabbie	Buffer
Replicate 2	Buffer	Magenta	Magenta	Forrest	Forrest	Manning	Manning	Mace	Mace	Wedgetail	Wedgetail	Currawong	Currawong	Buffer	Buffer	Hindmarsh	Hindmarsh	Urambie	Urambie	Oxford	Oxford	Buffer	Buffer	Eurabbie	Eurabbie	Bannister	Bannister	Buffer
Replicate 3	Buffer	Mace	Mace	Manning	Manning	Forrest	Forrest	Wedgetail	Wedgetail	Currawong	Currawong	Magenta	Magenta	Buffer	Buffer	Oxford	Oxford	Hindmarsh	Hindmarsh	Urambie	Urambie	Buffer	Buffer	Bannister	Bannister	Eurabbie	Eurabbie	Buffer

Source: Dr Karyn Reeves, SAGI-West

AAGI Analytics for the Australian Grains Industry

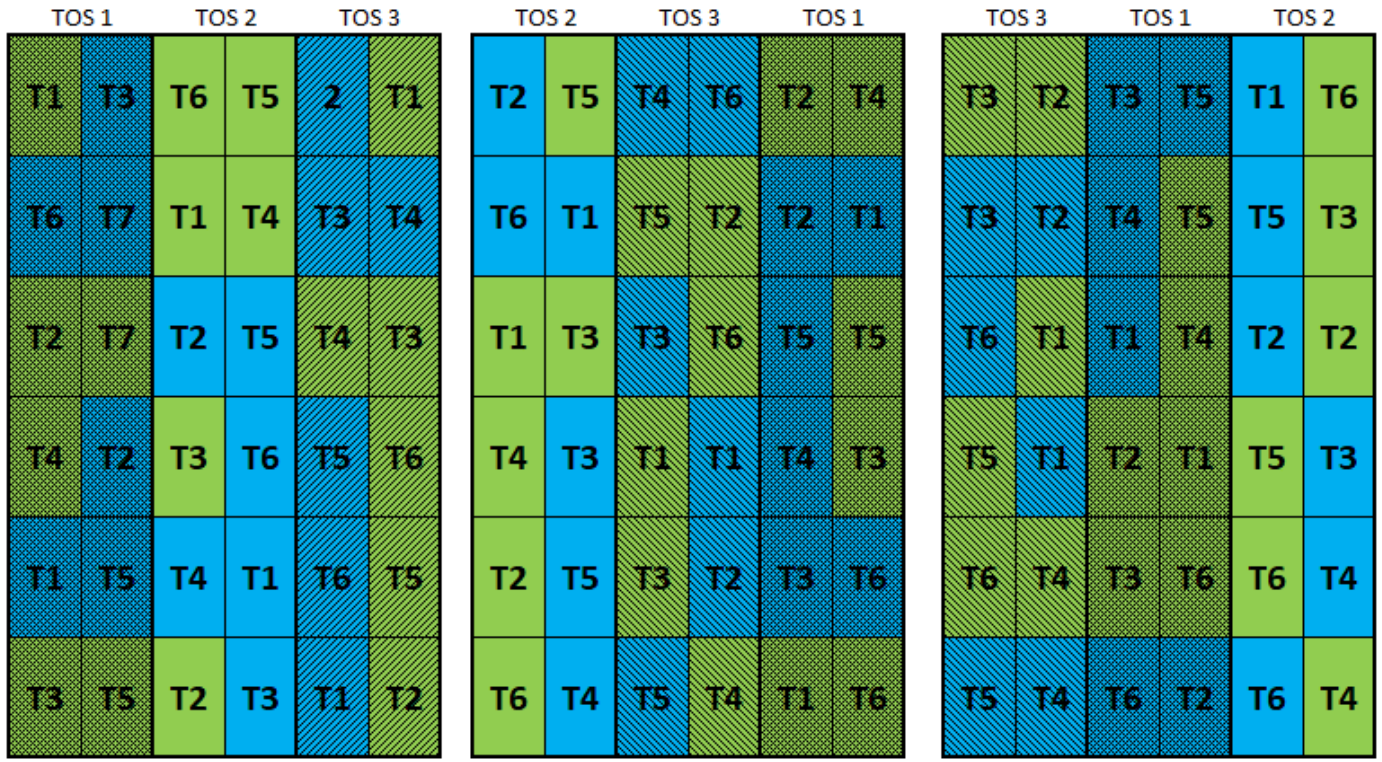
Don’t Optimise for Convenience



4 lime rates by 3 application methods (M1, M2, M3)

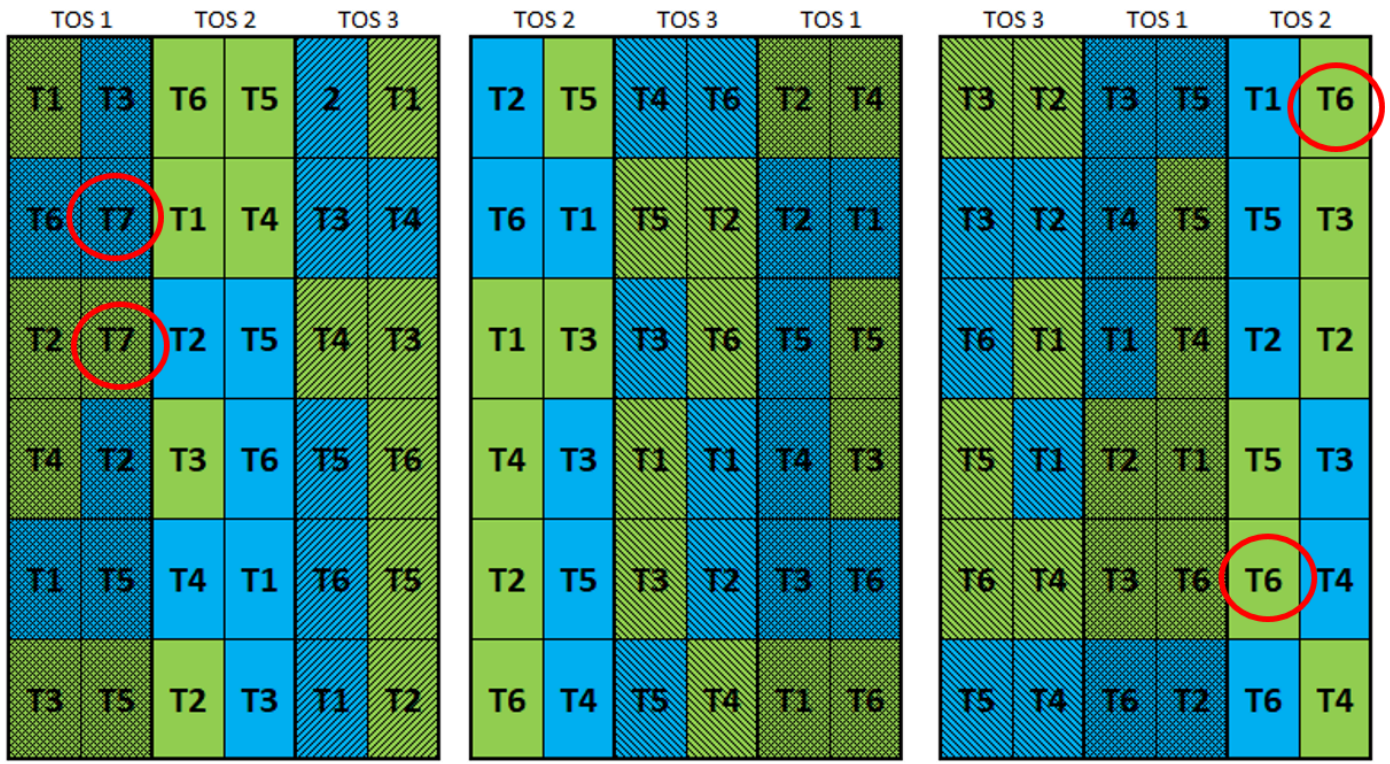
Source: Andrew Van Burgel, DPIRD

Ensure the Design is Recorded Correctly



Source: Dr Karyn Reeves, SAGI-West

Ensure the Design is Recorded Correctly



Source: Dr Karyn Reeves, SAGI-West

AAGI
Analytics for the Australian Grains Industry

Controlling Measurement Error and Uncertainty

Source:SVGRepo

Controlling Measurement Error and Noise

Source: Dr Karyn Reeves, SAGI-West

AAGI
Analytics for
the Australian
Grains Industry

Controlling Measurement Error and Noise

	T1 Rep1	T1 Rep2	T2 Rep1	T2 Rep2	T3 Rep1	T3 Rep2
14/08/2016 5:15	-0.91	-0.622	-1.189	-1.057	-0.677	-0.954
14/08/2016 5:30	-0.903	-0.648	-1.04	-1.023	-0.724	-0.932
14/08/2016 5:45	-0.768	-0.619	-0.702	-0.901	-0.577	-0.791
14/08/2016 6:00	-0.728	-0.595	-0.828	-0.898	-0.619	-0.772
14/08/2016 6:15	-0.765	-0.617	-0.821	-0.946	-0.648	-0.814
14/08/2016 6:30	-1.057	-0.886	-1.124	-1.204	-0.906	-1.057
14/08/2016 6:45	-0.922	-0.867	-0.978	-1.041	-0.78	-0.958
14/08/2016 7:00	-0.988	-0.833	-1.041	-1.101	-0.835	-0.953
14/08/2016 7:15	-0.142	-0.033	0.38	0.183	-0.003	-0.08
14/08/2016 7:30	1.602	1.691	2.628	2.041	1.795	1.619
14/08/2016 7:45	3.763	3.655	4.777	3.91	3.971	3.674
14/08/2016 8:00	6.018	5.671	6.484	5.762	6.094	5.589
14/08/2016 8:15	8.043	7.694	8.752	7.724	8.454	7.667
14/08/2016 8:30	8.897	8.458	9.098	8.228	9.463	8.288
14/08/2016 8:45	9.959	9.011	10.257	8.989	10.559	8.924
14/08/2016 9:00	11.361	10.219	12.139	10.963	11.685	10.014
14/08/2016 9:15	12.399	11.186	13.824	12.023	12.762	11.102
14/08/2016 9:30	13.526	12.411	15.373	13.554	14.122	12.137
14/08/2016 9:45	14.573	13.253	16.464	14.686	15.159	13.136
14/08/2016 10:00	15.969	15.468	17.658	15.993	16.441	14.275
14/08/2016 10:15	17.317	16.967	18.747	17.164	17.742	16.037
14/08/2016 10:30	18.051	17.813	19.977	18.194	18.41	17.385
14/08/2016 10:45	18.523	18.173	20.176	18.697	18.816	17.803
14/08/2016 11:00	19.155	19.057	20.697	19.749	19.288	18.593
14/08/2016 11:15	19.877	20.128	21.103	20.677	19.884	19.479
14/08/2016 11:30	20.035	20.155	21.14	20.643	19.858	19.523
14/08/2016 11:45	20.949	21.215	22.053	21.567	20.897	20.428
14/08/2016 12:00	22.175	22.148	23.417	22.763	22.056	21.488
14/08/2016 12:15	21.271	21.25	22.359	21.695	21.159	20.794
14/08/2016 12:30	21.335	21.119	21.975	21.564	21.004	20.637

Source: Dr Karyn Reeves, SAGI-West

Controlling Measurement Error and Noise

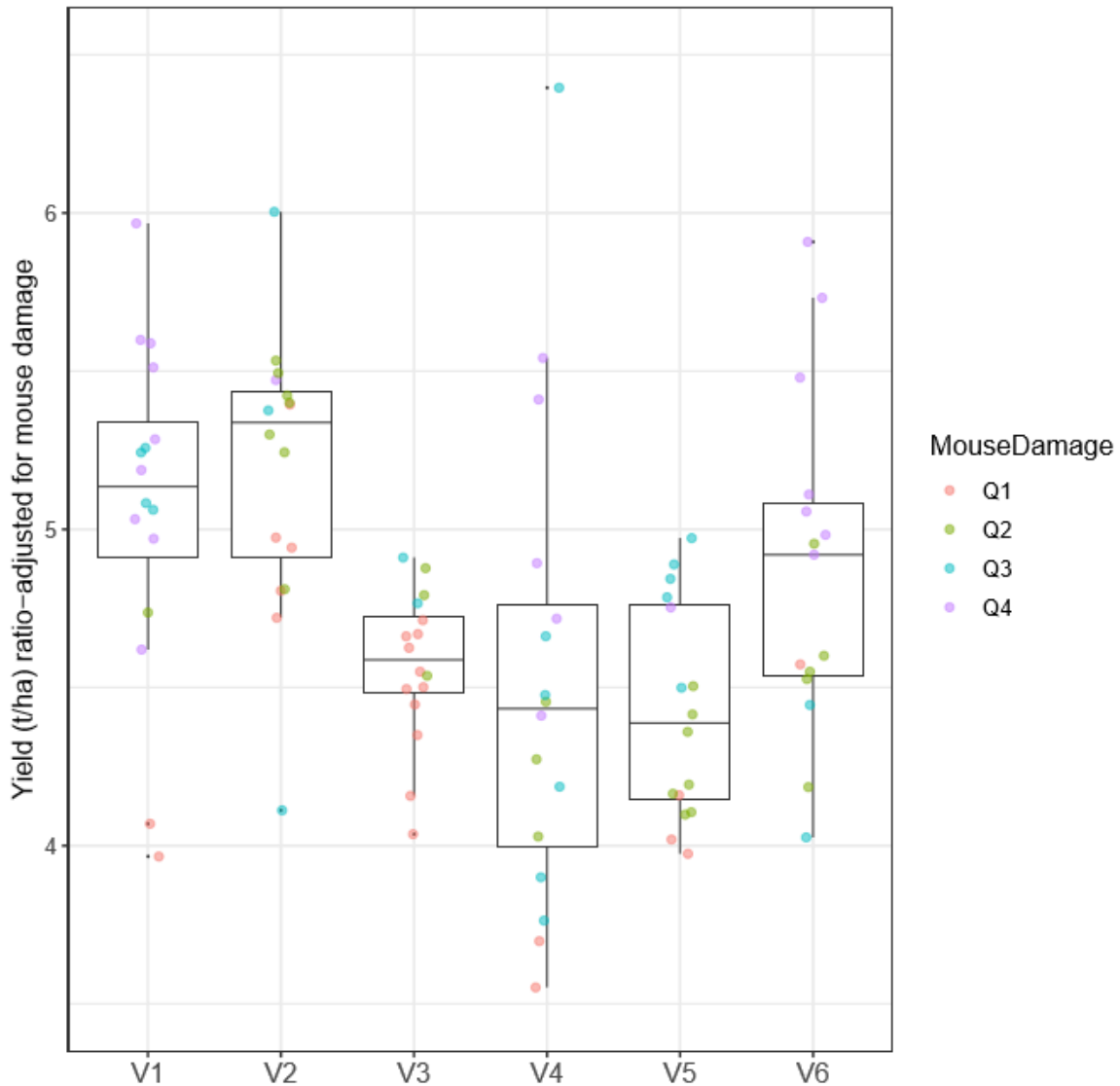
	T1 Rep1	T1 Rep2	T2 Rep1	T2 Rep2	T3 Rep1	T3 Rep2
25/08/2016 11:15	20.061	20.735	21.493	21.433	21.818	20.194
25/08/2016 11:30	21.464	21.388	23.952	23.816	23.621	21.4
25/08/2016 11:45	22.056	22.613	24.533	25.259	24.106	22.753
25/08/2016 12:00	22.727	23.228	26.035	25.974	24.888	23.02
25/08/2016 12:15	22.028	22.708	25.069	24.195	24.15	22.382
25/08/2016 12:30	18.777	19.275	20.387	20.341	19.914	30.568
25/08/2016 12:45	18.26	17.87	19.011	19.354	18.992	32.178
25/08/2016 13:00	18.698	17.695	20.29	19.551	18.919	33.81
25/08/2016 13:15	18.889	18.006	21.683	27.506	19.686	33.315
25/08/2016 13:30	17.755	17.336	20.45	27.42	19.429	27.86
25/08/2016 13:45	17.237	18.74	20.679	32.214	19.724	25.991
25/08/2016 14:00	17.123	19.529	20.683	32.839	20.348	24.585
25/08/2016 14:15	16.796	18.401	18.942	25.009	19.959	22.295
25/08/2016 14:30	16.563	17.679	17.714	22.471	19.159	20.523
25/08/2016 14:45	16.422	17.215	17.02	22.665	18.651	19.103
25/08/2016 15:00	16.124	16.784	16.464	19.859	17.956	18.725
25/08/2016 15:15	15.919	16.52	16.17	18.009	17.449	18.029
25/08/2016 15:30	15.371	21.446	15.57	22.594	16.698	21.086
25/08/2016 15:45	14.978	20.922	15.093	21.731	16.116	20.483
25/08/2016 16:00	14.592	20.489	14.726	21.513	15.641	20.183
25/08/2016 16:15	14.213	20.006	14.286	20.375	15.125	19.703
25/08/2016 16:30	13.79	18.895	13.806	19.133	14.576	18.487
25/08/2016 16:45	13.352	17.652	13.361	17.791	13.99	17.316
25/08/2016 17:00	12.838	16.424	12.776	16.275	13.355	16.129
25/08/2016 17:15	12.216	15.139	12.056	13.715	12.641	14.575
25/08/2016 17:30	11.348	10.565	11.048	9.706	11.565	10.433
25/08/2016 17:45	10.251	7.901	9.74	7.883	10.294	7.927
25/08/2016 18:00	9.511	7.052	9.107	7.405	9.491	7.184
25/08/2016 18:15	9.201	7.273	8.86	7.585	9.183	7.336
25/08/2016 18:30	9.025	7.192	8.539	7.421	9.046	7.228

Source: Dr Karyn Reeves, SAGI-West

	T1 Rep1	T1 Rep2	T2 Rep1	T2 Rep2	T3 Rep1	T3 Rep2
30/08/2016 0:00	4.952	0.851	6.286	1.632	4.906	0.683
30/08/2016 0:15	4.872	0.781	5.896	1.393	4.813	0.716
30/08/2016 0:30	4.506	0.529	5.903	1.118	4.58	0.505
30/08/2016 0:45	4.27	0.086	6.024	0.71	4.347	0.021
30/08/2016 1:00	4.098	-0.287	5.813	0.334	4.029	-0.422
30/08/2016 1:15	3.83	-0.397	5.05	0.509	3.717	-0.544
30/08/2016 1:30	3.744	-0.439	5.26	0.442	3.576	-0.416
30/08/2016 1:45	3.492	-0.299	5.341	0.344	3.522	-0.548
30/08/2016 2:00	3.592	0.091	5.086	0.544	3.573	-0.132
30/08/2016 2:15	3.586	-0.259	4.863	0.193	3.617	-0.45
30/08/2016 2:30	3.436	-0.472	4.664	0.217	3.336	-0.751
30/08/2016 2:45	3.293	-0.653	4.736	-0.114	3.197	-0.988
30/08/2016 3:00	3.032	-0.994	4.409	-0.254	2.835	-0.99
30/08/2016 3:15	2.713	-1.207	4.068	-0.331	2.737	-1.317
30/08/2016 3:30	2.85	-1.225	4.094	-0.583	2.497	-1.426
30/08/2016 3:45	2.713	-1.077	3.972	-0.716	2.584	-1.266
30/08/2016 4:00	2.575	-1.331	4.284	-0.51	2.74	-1.481
30/08/2016 4:15	2.576	-1.768	4.095	-0.912	2.517	-1.868
30/08/2016 4:30	2.125	-1.763	3.493	-0.985	2.479	-1.907
30/08/2016 4:45	2.219	-2.205	3.46	-1.232	2.089	-2.406
30/08/2016 5:00	2.059	-2.236	3.809	-1.195	1.811	-2.294
30/08/2016 5:15	1.932	-0.721	3.614	-1.641	1.749	-2.096
30/08/2016 5:30	2.076	-1.391	3.57	-1.445	1.876	-2.308
30/08/2016 5:45	1.907	-1.324	3.2	-1.319	1.966	-1.886
30/08/2016 6:00	2.041	-1.666	3.179	-1.735	1.803	-1.945
30/08/2016 6:15	1.951	-1.46	3.584	-1.264	1.668	-1.524
30/08/2016 6:30	2.504	-0.045	3.717	0.49	2.008	0.369
30/08/2016 6:45	3.253	0.658	4.275	1.74	2.938	1.769
30/08/2016 7:00	3.947	2.253	4.725	2.408	3.657	2.512
30/08/2016 7:15	4.405	2.663	5.094	2.792	4.018	2.709

AAGI Analytics for the Australian Grains Industry

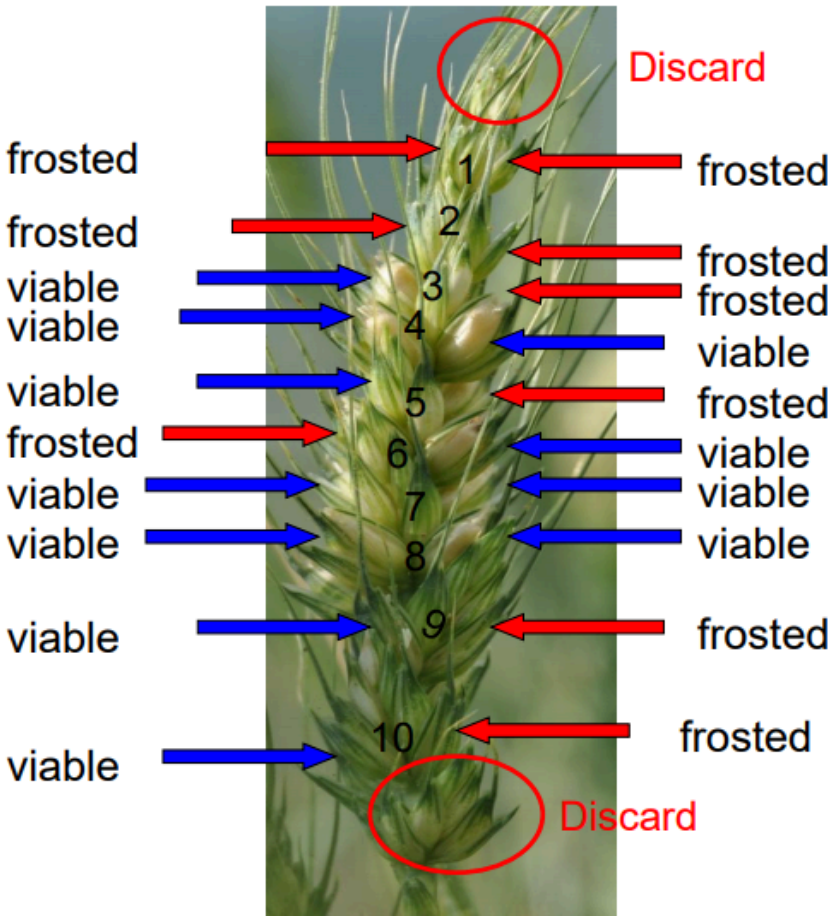
Always Provide the Full Set of Raw Data



Source: Dr Karyn Reeves, SAGI-West

Record Ancillary Details

Frost Induced Sterility (FIS) estimation



On this side of the head

10 pairs of spikelets,= 20 florets in total,

9 florets frosted

11 florets viable

Count the other side of the head as well

FIS = 18/40x100 = 45%

Source: Dr Ben Biddulph, DPIRD

AAGI Analytics for the Australian Grains Industry

Check Your Data

Supplied data			
B1	W2	W2	B1
B3	W1	W2	B2
B4	W1	W1	B5
B2	W3	W4	B4
B5	W4	W4	B3
W3	B1	B3	W4
W1	B5	B4	W3
W2	B4	B2	W3
W1	B1	B3	W5
W5	B2	B5	W5
B4	W1	W4	B4
B1	W3	W5	B1
B5	W2	W5	B4
B3	W1	W1	B5
B2	W3	W2	B3
W2	B3	B1	W2
W4	B2	B1	W2
W5	B2	B2	W5
W3	B3	B4	W4
W3	B5	B5	W1
B5	W4	W2	B3
B2	W5	W3	B4
B2	W5	W4	B1
B5	W5	W1	B3

Designed trial			
B1	B4	B5	B3
B3	B1	B2	B4
B4	B5	B2	B1
B2	B3	B5	B1
B5	B2	B3	B4
W3	W2	W4	W5
W1	W4	W3	W5
W2	W5	W4	W1
W1	W3	W2	W4
W5	W3	W1	W2
W2	W1	W4	W3
W1	W3	W5	W2
W1	W2	W5	W4
W3	W1	W5	W4
W4	W3	W2	W5
B1	B3	B4	B2
B5	B2	B1	B3
B4	B2	B3	B5
B1	B3	B4	B5
B2	B5	B1	B4
W2	W4	W1	W3
W2	W5	W4	W3
W1	W5	W3	W5
W4	W1	W2	W5

Correct Response			
1.10	0.27	5.68	2.17
0.00	0.00	2.08	1.92
0.34	0.95	9.68	3.06
0.00	5.70	2.43	5.46
0.00	4.41	7.39	2.99
0.00	3.15	9.93	1.48
0.54	3.83	9.08	4.26
0.18	2.48	8.78	2.22
0.00	1.72	4.14	4.04
0.39	4.75	19.54	8.33
1.24	3.79	12.99	1.39
4.54	1.53	15.92	1.50
1.68	1.17	7.88	0.64
2.31	1.69	8.77	1.95
0.41	10.37	8.54	1.34
0.28	8.56	7.97	1.09
0.59	8.21	12.45	6.11
0.98	12.16	17.09	6.40
0.67	11.78	16.52	10.03
0.60	16.20	30.93	12.75
3.05	5.86	12.49	8.18
2.14	3.12	23.97	14.62
1.25	9.83	12.56	4.88
2.92	4.78	14.16	2.40

Supplied Response			
1.10	0.27	7.54	2.17
0.00	0.00	5.68	1.92
0.34	0.95	2.08	3.06
0.00	5.70	9.68	5.46
0.00	4.41	2.43	2.99
0.00	3.15	7.39	1.48
0.54	3.83	9.93	4.26
0.18	2.48	9.08	2.22
0.00	1.72	8.78	4.04
0.39	4.75	4.14	8.33
1.24	3.79	19.54	1.39
4.54	1.53	12.99	1.50
1.68	1.17	15.92	0.64
2.31	1.69	7.88	1.95
0.41	10.37	8.77	1.34
0.28	8.56	8.54	1.09
0.59	8.21	7.97	6.11
0.98	12.16	12.45	6.40
0.67	11.78	17.09	10.03
0.60	16.20	16.52	12.75
3.05	5.86	30.93	8.18
2.14	3.12	12.49	14.62
1.25	23.97	12.56	4.88
2.92	9.83	14.16	2.40

Source: Dr Karyn Reeves, SAGI-West

AAGI
Analytics for the Australian Grains Industry

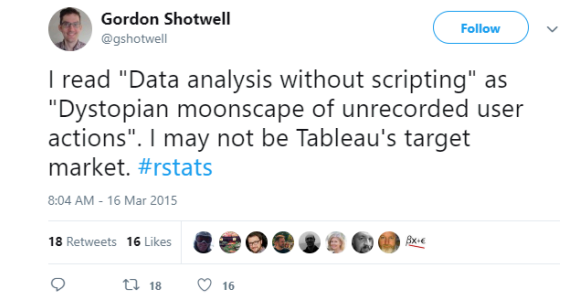
Effective Use of Spreadsheets for Sharing Data With AAGI

GRDC Grains Research & Development Corporation | Curtin University | The University Of Queensland Australia | The University of Adelaide

AAGI Analytics for the Australian Grains Industry

Best Practices to Follow

- All columns supplied with the trial design should be retained
- Every plot needs a unique identifier (*e.g.*, plot number)
- Avoid copy-paste (easy to introduce errors)
- Ensure embedded equations reference the correct columns
  - Or better yet, don't do calculations in the spreadsheet at all



Source: Dr Karyn Reeves, SAGI-West

Basic Rules for Data in Spreadsheets

- Be Consistent
  - Use the same layout if you have multiple files
  - Use the same variable names (also if you have multiple files)
  - Only use one variable name, *e.g.*, *“S10”* is different than *“S 10”*

Choose Good Names for Things

Good Name	Good Alternative	Avoid
max_temp_C	MaxTemp	Maximum Temp (˚C)
precipitation_mm	precip	precmm
mean_year_growth	MeanYearGrowth	Mean growth/year
yield_kg_ha	yield	yield kg/ha
observation_o1	first_observation	1st Obs.

Write Dates as YYYY-MM-DD (ISO 8601)

	A	B	C
1	Date	Assay date	Weight
2		12/9/05	54.9
3		12/9/05	45.3
4	12/6/2005	e	47
5		e	45.7
6		e	52.9
7		1/11/2006	46.1
8		1/11/2006	38.6

A spreadsheet with inconsistent date formats. This spreadsheet does not adhere to recommendations for consistency of date format. From Broman and Woo (2018). Also see https://xkcd.com/1179.

No Empty Cells

A

	A	B	C
1	id	date	glucose
2	101	2015–06–14	149.3
3	102		95.3
4	103	2015–06–18	97.5
5	104		117.0
6	105		108.0
7	106	2015–06–20	149.0
8	107		169.4

B

	A	B	C	D	E	F	G	H	I
1		1 min				5 min			
2	strain	normal		mutant		normal		mutant	
3	A	147	139	166	179	334	354	451	474
4	B	246	240	178	172	514	611	412	447

Examples of spreadsheets that violate the *"no empty cells"* recommendation. (a) A spreadsheet where only the first of several repeated values was included. (b) A spreadsheet with a complicated layout and some implicit column headers, from Broman and Woo (2018).

Put Just One Thing in a Cell

- e.*g.*, Don't use *"rep-plot"* for a header and use *"1-1"*, *"1-2"*…*"2-1"*, in the column, etc.
- Do use *"rep"* and *"plot"* as headers and have the values in separated columns

Make It a Rectangle

A

	A	B	C	D	E	F
1						
2		101	102	103	104	105
3	sex	Male	Female	Male	Male	Male
4						
5		101	102	103	104	105
6	glucose	134.1	120.0	124.8	83.1	105.2
7						
8		101	102	103	104	105
9	insulin	0.60	1.18	1.23	1.16	0.73

B

	A	B	C	D	E	F	G
1	1MIN						
2			Normal			Mutant	
3	B6	146.6	138.6	155.6	166	179.3	186.9
4	BTBR	245.7	240	243.1	177.8	171.6	188.1
5							
6	5MIN						
7			Normal			Mutant	
8	B6	333.6	353.6	408.8	450.6	474.4	423.8
9	BTBR	514.4	610.6	597.9	412.1	447.4	446.5

C

	A	B	C	D	E	F	G
1							
2	Date	11/3/14					
3	Days on diet	126					
4	Mouse #	43					
5	sex	f					
6	experiment		values			mean	SD
7	control		0.186	0.191	1.081	0.49	0.52
8	treatment A		7.414	1.468	2.254	3.71	3.23
9	treatment B		9.811	9.259	11.296	10.12	1.05
10							
11	fold change		values			mean	SD
12	treatment A		15.26	3.02	4.64	7.64	6.65
13	treatment B		20.19	19.05	23.24	20.83	2.17

D

	A	B	C	D	E	F
1		GTT date	GTT weight	time	glucose mg/dl	insulin ng/ml
2	321	2/9/15	24.5	0	99.2	lo off curve
3				5	349.3	0.205
4				15	286.1	0.129
5				30	312	0.175
6				60	99.9	0.122
7				120	217.9	lo off curve
8	322	2/9/15	18.9	0	185.8	0.251
9				5	297.4	2.228
10				15	439	2.078
11				30	362.3	0.775
12				60	232.7	0.5
13				120	260.7	0.523
14	323	2/9/15	24.7	0	198.5	0.151
15				5	530.6	off curve lo

Examples of spreadsheets with nonrectangular layouts. These layouts are likely to cause problems in analysis, from Broman and Woo (2018).

Example A, what not to do.

Make It a Rectangle

	A	B	C	D	E	F	G	H	I	J	K
1			week 4			week 6			week 8		
2	Mouse ID	SEX	date	weight	glucose	date	weight	glucose	date	weight	glucose
3	3005	M	3/30/2007	19.3	635	4/11/2007	31	460.7	4/27/2007	39.6	530.2
4	3017	M	10/6/2006	25.9	202.4	10/19/2006	45.1	384.7	11/3/2006	57.2	458.7
5	3434	F	11/22/2006	26.6	238.9	12/6/2006	45.9	378	12/22/2006	56.2	409.8
6	3449	M	1/5/2007	27.5	121	1/19/2007	42.9	191.3	2/2/2007	56.7	182.5
7	3499	F	1/5/2007	19.8	220.2	1/19/2007	36.6	556.9	2/2/2007	43.6	446

A spreadsheet with two header rows. It is better to have a single header row, from Broman and Woo (2018).

Example B, what not to do.

Make It a Rectangle

	A	B	C	D	E	F
1	mouse_id	sex	week	date	glucose	weight
2	3005	M	4	3/30/2007	19.3	635
3	3005	M	6	4/11/2007	31	460.7
4	3005	M	8	4/27/2007	39.6	530.2
5	3017	M	4	10/6/2006	25.9	202.4
6	3017	M	6	10/19/2006	45.1	384.7
7	3017	M	8	11/3/2006	57.2	458.7
8	3434	F	4	11/22/2006	26.6	238.9
9	3434	F	6	12/6/2006	45.9	378
10	3434	F	8	12/22/2006	56.2	409.8
11	3449	M	4	1/5/2007	27.5	121
12	3449	M	6	1/19/2007	42.9	191.3
13	3449	M	8	2/2/2007	56.7	182.5
14	3499	F	4	1/5/2007	19.8	220.2
15	3499	F	6	1/19/2007	36.6	556.9
16	3499	F	8	2/2/2007	43.6	446

An example spreadsheet of the previous example's data in a rectangular layout, from Broman and Woo (2018).

Yes, do this!

Create a Data Dictionary

	A	B	C	D
1	name	plot_name	group	description
2	mouse	Mouse	demographic	Animal identifier
3	sex	Sex	demographic	Male (M) or Female (F)
4	sac_date	Date of sac	demographic	Date mouse was sacrificed
5	partial_inflation	Partial inflation	clinical	Indicates if mouse showed partial pancreatic inflation
6	coat_color	Coat color	demographic	Coat color, by visual inspection
7	crumblers	Crumblers	clinical	Indicates if mouse stored food in their bedding
8	diet_days	Days on diet	clinical	Number of days on high–fat diet

An example data dictionary, from Broman and Woo (2018).

Do Not Use Colour as Data

A

	A	B	C
1	id	date	glucose
2	101	2015–06–14	149.3
3	102	2015–06–14	95.3
4	103	2015–06–18	97.5
5	104	2015–06–18	1.1
6	105	2015–06–18	108.0
7	106	2015–06–20	149.0
8	107	2015–06–20	169.4

B

	A	B	C	D
1	id	date	glucose	outlier
2	101	2015–06–14	149.3	FALSE
3	102	2015–06–14	95.3	FALSE
4	103	2015–06–18	97.5	FALSE
5	104	2015–06–18	1.1	TRUE
6	105	2015–06–18	108.0	FALSE
7	106	2015–06–20	149.0	FALSE
8	107	2015–06–20	169.4	FALSE

Highlighting in spreadsheets.(a) A potential outlier indicated by highlighting the cell.(b) The preferred method for indicating outliers, via an additional column, from Broman and Woo (2018).

Save Data in Plain Text Files (*e.g.*, CSV)

A

	A	B	C	D	E
1	id	sex	glucose	insulin	triglyc
2	101	Male	134.1	0.60	273.4
3	102	Female	120.0	1.18	243.6
4	103	Male	124.8	1.23	297.6
5	104	Male	83.1	1.16	142.4
6	105	Male	105.2	0.73	215.7

B

```
id,sex,glucose,insulin,triglyc
101,Male,134.1,0.60,273.4
102,Female,120.0,1.18,243.6
103,Male,124.8,1.23,297.6
104,Male,83.1,1.16,142.4
105,Male,105.2,0.73,215.7
```

(a) An example spreadsheet. (b) The same data as a plain text file in CSV format, from Broman and Woo (2018).

Lastly

- No calculations in the raw data file
- Make backups
- Use data validation to avoid errors

Exercise

Exercise (20 min)

Working with your partner on the experiment you designed this morning.

1. Create a trial protocol that details the:
    1. Treatments and
    2. Data to be collected.
2. Draw or list the spreadsheet headings for the data you will be collecting to send to AAGI.
    1. Describe the date format you will use,
    2. Describe what you will use to describe missing data and
    3. Describe what descriptions need to be included in the data dictionary/protocol for AAGI to refer to.

Thank You

GRDC Grains Research & Development Corporation | Curtin University | The University Of Queensland Australia | The University of Adelaide

AAGI Analytics for the Australian Grains Industry

References

Broman, Karl W., and Kara H. Woo. 2018. “Data Organization in Spreadsheets.” *The American Statistician* 72 (1): 2–10. https://doi.org/10.1080/00031305.2017.1375989.