Experimental Design Project

STA2005S - 2025

Maryam Abrahams (ABRMAR043)
Bheka Mabika (MBKBHE002)

Table of contents

Introduction:

The two chilli farmers, Ms Hopeful and Mr Growing, compete head to head for the hearts and minds of the South African chilli market. But despite their names Ms Hopeful seems to always get higher chilli yields than Mr Growing. In Mr Growing's eyes this cannot continue.

Aim:

This experiment aims to systematically evaluate the effect of light levels, heat levels and chilli variety on crop quality and yield. With our results we will be able to give evidence-based recommendations on optimal growing conditions.

Particularly we seek to answer:

1. Which light and heat settings produce the highest quality chillies?

2. Does the above depend on variety?

3. Does quality depend on variety?

4. How big are the differences?

Priori Hypothesis:

Prior to collecting the data, we hypothesize:

- H_1: Mean quality will increase with an increase in light levels

- H_2: Mean Quality will increase with intermediate heat levels (2-3), as extreme heat may damage plant cells while insufficient heat limits growth.

- H_3: There may (or may not) be a Heat $\times$ Light interaction.

- H_4: Both varieties will show similar response patterns to light and heat treatments. But mean quality will differ by variety.

- H_5: Plot side (North/South) will have negligible effect on quality after accounting for other factors.

Experimental Design and Randomization:

Simulating data:

We created the design matrix in accordance to a **Nested Latin Square design**, testing all 16 heat $\times$ light combinations in each block where each block represented a greenhouse-season combination and is created from a 2$\times$2 latin square design with greenhouse (A, B) as columns and season (1, 2) as rows.

treatment factors:

- Light: 4 levels (1, 2, 3, 4)

- Heat: 4 levels (1, 2, 3, 4)

$\implies$ 16 unique treatments

blocking factors:

- Season: (1, 2)

- Greenhouse: (A, B)

- Side of greenhouse: (N = north-facing, S = south-facing)

- Chilli variety: (R = red-hot, F = furious)

response variables: quality score, which combines taste, yield, and look

Side-Variety combination assignments:

	H1	H2	H3	H4
L1	RN	FN	RS	FS
L2	FN	RN	FS	RS
L3	RS	FS	RN	FN
L4	FS	RS	FN	RN

->|

Example Greenhouse setup:

	R		F	
N	H1	L1	H1	L2
	H2	L2	H2	L1
	H3	L3	H3	L4
	H4	L4	H4	L3
S	H1	L3	H1	L4
	H2	L4	H2	L3
	H3	L1	H3	L2
	H4	L2	H4	L1

Figure 1: Blocking example greenhouse

Design:

Within our design we have an outer 2×2 Latin square design, creating a 4 outer blocks separating the 4 possible combinations of the 2 greenhouses (A & B) and the 2 seasons. Each of these represents a separate experimental block.

- Block 1 = Season 1 $\times$ Greenhouse A

- Block 2 = Season 1 $\times$ Greenhouse B

- Block 3 = Season 2 $\times$ Greenhouse A

- Block 4 = Season 2 $\times$ Greenhouse B

Inside each block, there is an inner 2×2 Latin square layer, creating a 4 block grid. The variety (*Redhot* or *Furious*) and side (*North* or *South*) were assigned according to the plan shown below:

Side	Variety	Description
North	Redhot (RN)	4 plots
North	Furious (FN)	4 plots
South	Redhot (RS)	4 plots
South	Furious (FS)	4 plots

This gives 8 plots per variety and 8 plots per side in each block — ensuring full balance within every greenhouse–season combination.

The overarching structure is thus such that: Heat-Light is nested within Side-variety which are is nested Season-greenhouse combinations.
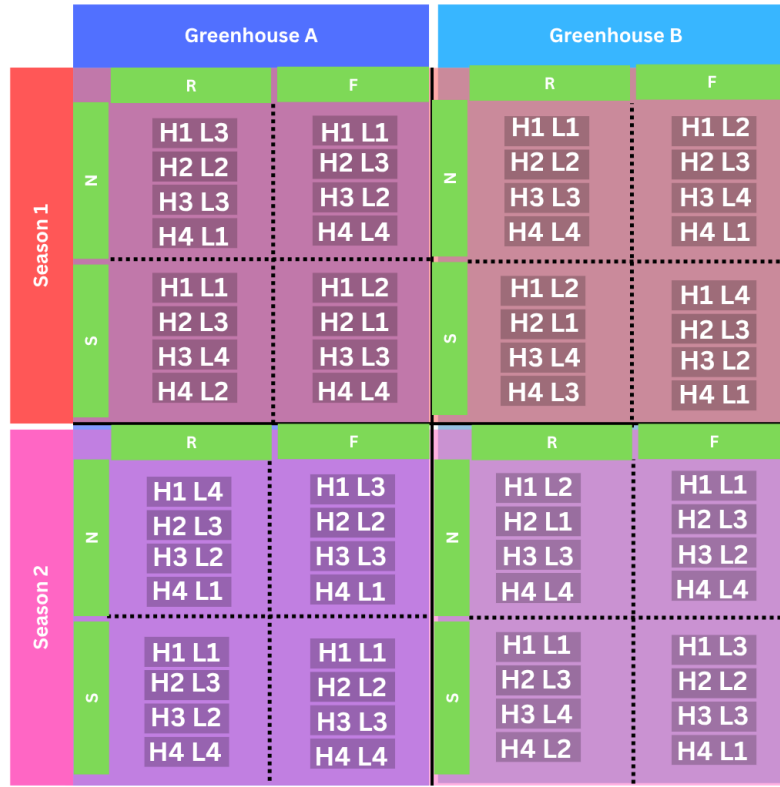

Figure 2: Blocking factors Diagram

Randomization:

Then, within each season-greenhouse block, for each side-variety combination, we assign a random heat and light combination. This means that Inside each block, all **16 combinations**

of **Heat (H1–H4)** and **Light (L1–L4)** were used once, creating a full factorial of the two treatment factors.

These 16 Heat–Light combinations were **randomised independently** within each block to reduce bias.

Constraints:

Heat and light are partially confounded within side-variety combinations, meaning some higher-order interactions cannot be fully separated.

We are limited to 2 greenhouses, 2 seasons, and 16 plots withni each of the greenhouses, 8 plots per side.

Simulation:

The simulated quality data was generated using the `get.observations()` function with a fixed random seed (`set.seed(22)`) to ensure reproducibility.

The generated design matrix was also validated with a loop which checks for correct randomization and that each heat-light combination is used once per side-variety combination.

Analysis and Results:

Fitting the models and checking assumptions:

Table 2: Model Coefficients Summary

	Estimate	Std. Error	t value	Pr($>$|t|)
(Intercept)	33.051	1.359	24.315	0.000
block2.A	-2.104	0.627	-3.354	0.002
block1.B	-1.458	0.656	-2.224	0.033
block2.B	-2.309	0.688	-3.357	0.002
sides	3.739	0.472	7.927	0.000
heat2	5.302	1.472	3.601	0.001
heat3	8.407	1.768	4.755	0.000
heat4	7.226	1.748	4.133	0.000
light2	9.773	1.788	5.467	0.000

	Estimate	Std. Error	t value	Pr(>|t|)
light3	13.797	1.716	8.040	0.000
light4	16.417	1.483	11.072	0.000
varietyR	2.880	1.362	2.114	0.042
heat2:light2	-3.394	2.393	-1.418	0.165
heat3:light2	-1.437	2.201	-0.653	0.518
heat4:light2	-3.006	2.278	-1.319	0.195
heat2:light3	-1.338	1.798	-0.744	0.461
heat3:light3	-1.070	2.083	-0.514	0.611
heat4:light3	-4.553	2.566	-1.774	0.084
heat3:light4	-0.724	2.407	-0.301	0.765
heat4:light4	-2.373	1.841	-1.289	0.206
heat2:varietyR	-1.860	1.564	-1.189	0.242
heat3:varietyR	-4.775	1.692	-2.822	0.008
heat4:varietyR	-2.455	1.552	-1.582	0.122
light2:varietyR	0.379	1.460	0.260	0.796
light3:varietyR	-1.572	1.481	-1.062	0.295
light4:varietyR	-1.626	1.585	-1.026	0.312

Model Interpretation:

We fitted a linear model to analyze the experimental design, using **Season × Greenhouse** (`block`) to control for background variation. The model tested the main effects of **heat**, **light**, **variety**, and **side**, as well as all two-way interactions involving the primary treatment factors (`heat:light`, `heat:variety`, `light:variety`).

Initially, a more complex model was considered, which included `side` nested within `block` (`side:block`) and a full three-way interaction (`heat:light:variety`). However, this version resulted in over-parameterization and non-estimable coefficients, likely due to the two missing data points. Therefore, this simplified and practical model was adopted to ensure all key treatment effects could be robustly estimated, in line with our project guidelines.

Because two observations were missing, the design became slightly unbalanced. To address this, we used **Type II ANOVA** (`car::Anova()`) for valid, order-independent F-tests of significance. The `summary(fit)` output was used only to interpret **coefficient direction and relative magnitude**, as its p-values are not reliable in unbalanced designs.

From the coefficients, the **intercept (~33.05)** represents the baseline quality under reference conditions. Positive coefficients for **Heat 2–4** (especially Heat 3: +8.41 and Heat 4: +7.23) indicate that increasing heat generally improved quality. The effect of **light** was even more pronounced, with Light 2–4 showing strong positive effects (up to +16.42 for Light 4). The **side effect** (`sides`) was also significant, suggesting that the greenhouse side had a measurable

impact on outcomes. Overall, the model fit was excellent, with an **adjusted R² of 0.93** and a **residual standard error of ~1.61**.

ANOVA

The experiment followed a **Latin Square design**, with **Season × Greenhouse** (`block`) as the blocking structure to account for background environmental variability. To ensure model stability and interpretability, we fitted a simplified fixed-effects model including all main effects (`block`, `side`, `heat`, `light`, `variety`) and two-way treatment interactions (`heat:light`, `heat:variety`, `light:variety`).

The fitted model was: `obs ~ block + side + heat*light + variety + heat:variety + light:variety`

$$obs = block + side + heat + light + variety$$
$$+ (heat : light) + (heat : variety) + (light : variety)$$

Table 3: Type II ANOVA Results

	Sum Sq	Df	F value	Pr(>F)
block	43.9712658	3	5.6515190	0.0028015
side	162.9796258	1	62.8421158	0.0000000
heat	215.2706776	3	27.6682127	0.0000000
light	1455.9992346	3	187.1360137	0.0000000
variety	0.0332905	1	0.0128362	0.9104245
heat:light	11.4327538	8	0.5510339	0.8099978
heat:variety	22.0876775	3	2.8388751	0.0515297
light:variety	7.3069467	3	0.9391440	0.4318596
Residuals	93.3651972	36	NA	NA

The ANOVA results showed highly significant main effects of heat, light, and side, but insignificant interactions between heat and variety as well as between light and variety ($p > 0.05$). These effects indicate that increasing temperature and light intensity both improved outcomes. The blocking factors (Season × Greenhouse) also explained a significant portion of variation ($p = 0.003 < 0.05$), confirming that the blocking was effective in controlling background differences.

Overall, the model captured the major treatment effects and their interactions well, explaining approximately 96% of the total variation ($R^2 = 0.96$) in observed responses.

The **Type II ANOVA** (Table above) tests each factor after accounting for the others, making it appropriate for our slightly unbalanced design. The note *"model has aliased coefficients"* confirms that some higher-order interactions were not estimable due to the two missing plots; however, the sums of squares were still correctly computed by model comparison.

Significant effects were found for **Block** ($p = 0.0028$), **Side** ($p < 0.001$), **Heat** ($p < 0.001$), and **Light** ($p < 0.001$). This indicates that chilli quality varied across greenhouses/seasons, was higher on one side of the greenhouse, and increased significantly with both higher heat and light levels. The **Variety** factor and all interactions were not significant ($p > 0.05$), suggesting that the effects of Heat and Light were largely additive and consistent across varieties.

Overall, the model captures strong main effects of **Heat**, **Light**, and **Side** on chilli quality, with minimal evidence of interaction or variety-dependent responses.

Model Diagnostics:
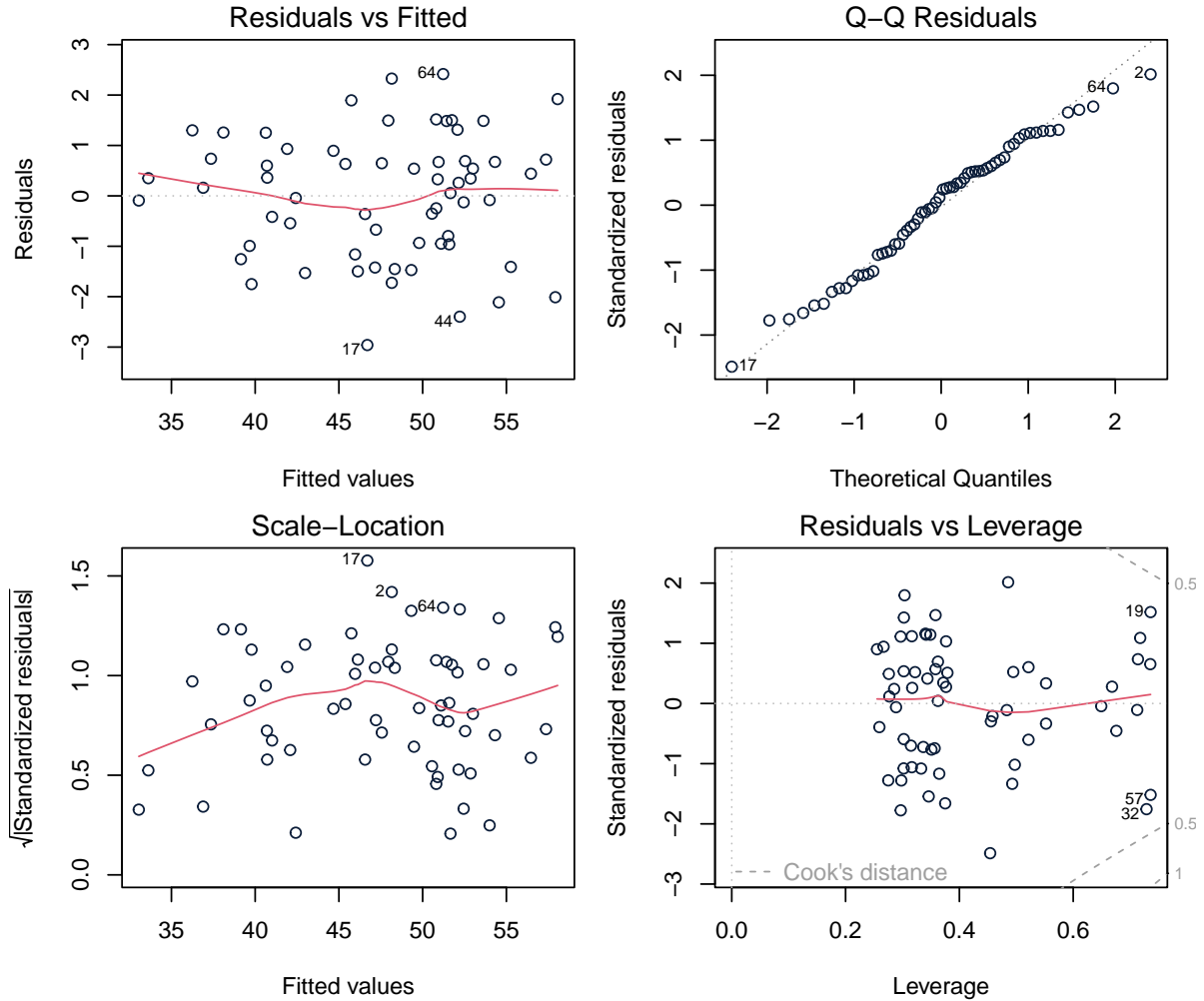
Figure 3: Residual diagnostic plots showing no serious violations of model assumptions. The Q-Q plot indicates approximate normality, and the residuals vs. fitted plot shows no strong patterns suggesting heteroscedasticity.

The model assumptions were checked using standard residual plots. The Residuals vs Fitted plot shows a random scatter of points around zero, indicating that the assumption of linearity and homoscedasticity (constant variance) is satisfied. There are no visible patterns or funnel shapes.

The Q-Q plot shows that the residuals lie close to the straight line, confirming that the normality assumption is reasonable. The Scale-Location plot also shows roughly constant spread

across fitted values, supporting equal variance.

Finally, the residuals vs leverage plot shows that no points have unusually high leverage or cooks distance, indicating that no single observation unduly influenced the fitted model.

Overall, these diagnostic plots confirm that the model fits the data well, and the results from the type 2 anova can be considered reliable and valid for interpretation.

Interaction plots:

The following plots explore how **Heat, Light, and Variety** affect mean chilli quality. Confidence intervals were excluded to highlight the main treatment patterns clearly.

Figure 4: Interaction between Heat and Light levels on chilli quality. Lines represent different Light levels.

Figure 5: Interaction between Heat levels and Variety on chilli quality. Lines show how Redhot and Furious varieties respond differently to temperature.
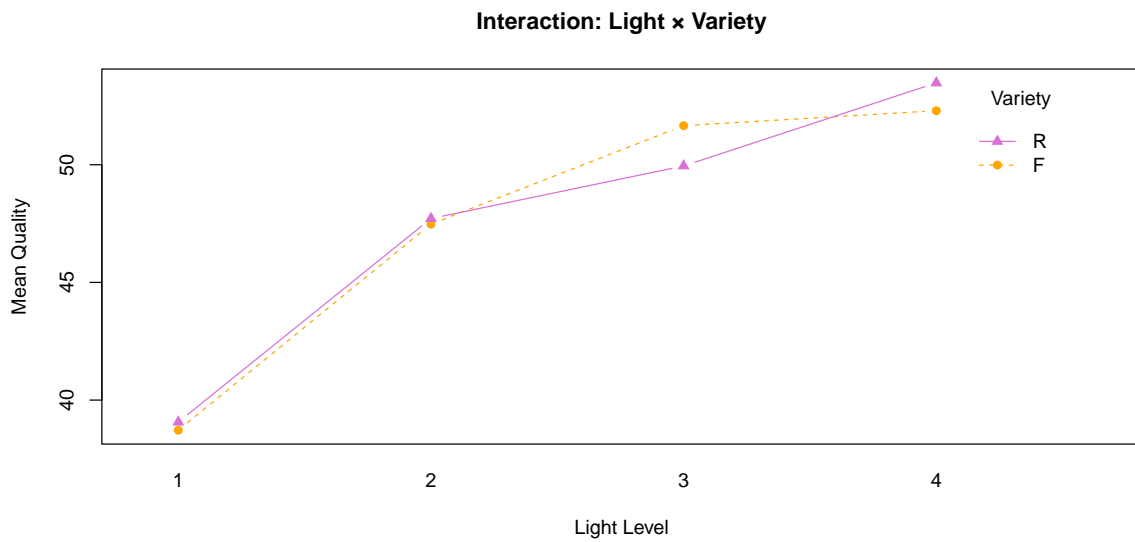
Figure 6: Interaction between Heat levels and Variety on chilli quality. Lines show how Redhot and Furious varieties respond differently to temperature.

Interpretation of Interaction Plots

Interpretation of Interaction Plots

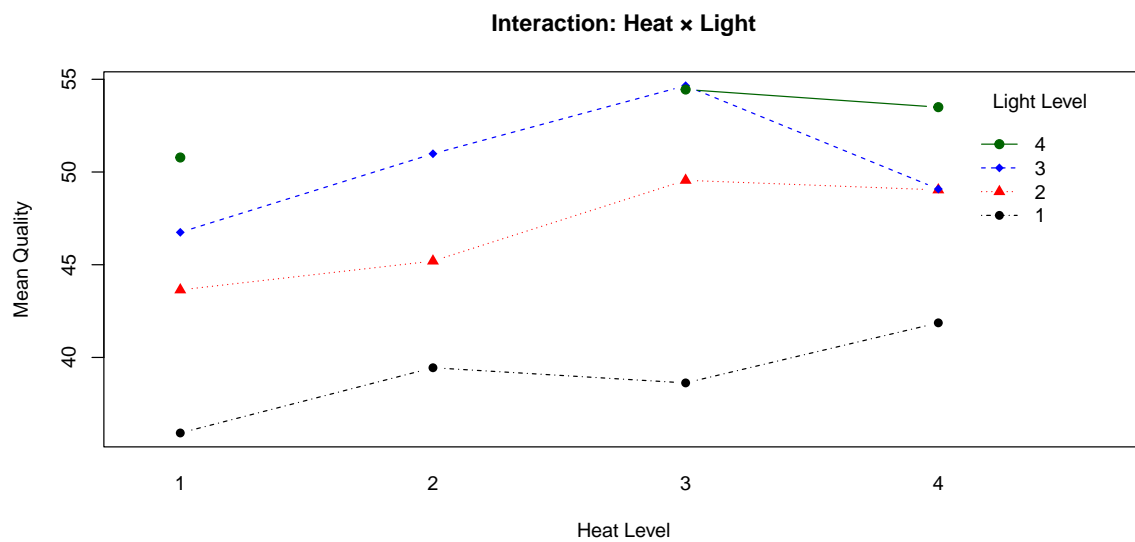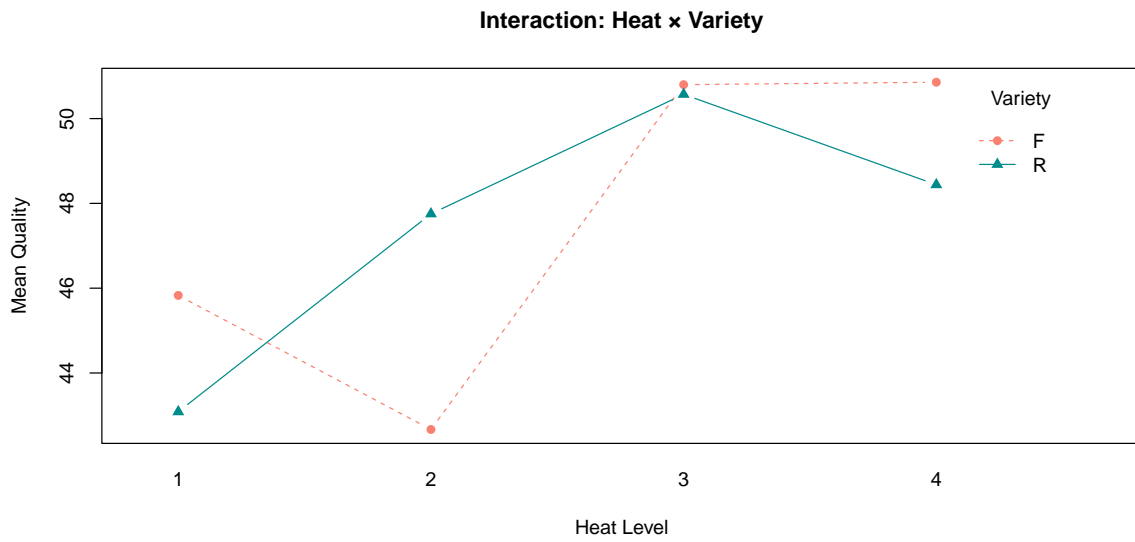The interaction plots reveal the following patterns:

1) **Heat × Light:** Quality increases with Heat from level 1 to 3 for all Light levels, with a slight plateau or dip at Heat 4. The lines are roughly parallel, indicating minimal interaction. This shows that Heat and Light act mostly additively, with each factor independently improving chilli quality. Differences between Light levels remain consistent across Heat levels, highlighting strong main effects.

2) **Heat × Variety:** Both varieties (Redhot and Furious) improve with increasing Heat. The lines are mostly parallel, with only a small divergence at Heat 4, where Redhot slightly outperforms Furious. This suggests that the Heat effect is consistent across varieties, with minimal Heat × Variety interaction.

3) **Light × Variety:** Quality rises sharply with Light from 1 to 3 and flattens slightly at Light 4. Lines for both varieties nearly overlap, indicating that Light affects both varieties similarly. This implies minimal Light × Variety interaction, confirming Light as a strong main effect rather than an interactive one.

Overall, the plots show that Heat and Light are the dominant factors, with both varieties responding similarly. Optimal quality is generally achieved around Heat 3–4 combined with Light 3–4.

Conclusions:

This section fits the simplified model

$$\text{obs} = \text{block} + \text{side} + \text{heat} + \text{light} + \text{variety}$$
$$+ (\text{heat} : \text{light}) + (\text{heat} : \text{variety}) + (\text{light} : \text{variety})$$

, runs Type-II ANOVA, extracts estimated marginal means with the emmeans package, answers the four farmer questions, and builds a cost-aware decision table.

Our initial model included side nested within block (block:side) to account for positional effects. However, this caused over-parametrisation and non-estimable marginal means due to missing values and limited replication.

We simplified the model by treating side as a main effect. This avoids the over-parameterization of prior models and yields stable, interpretable estimates for heat, light, and variety. This approach is justified as it appropriately controls for side as a known nuisance variable.

Q1 — Which Heat × Light is best?

Answer:

The top four treatment combinations were:

Heat	Light	Mean Quality	95% CI
3	4	55.79	[53.09, 58.49]
4	4	54.12	[52.85, 55.40]
3	3	52.85	[51.05, 54.66]
2	3	50.93	[49.53, 52.34]

The combination of Heat 3 and Light 4 produced the highest estimated mean quality (55.79).

Quality increases consistently and substantially with higher Light levels, which appears to be the most dominant environmental factor. The effect of Heat is also generally positive but less linear. The top-performing setting (H3, L4) was significantly better than most other combinations, though the second-best (H4, L4) was statistically comparable (a non-significant difference of 1.67 units).

Q2 — Do best settings depend on Variety?

Answer: Yes, there is evidence for a specific interaction. While the overall Anova table shows the light:variety interaction is not significant (p = 0.432) and the general heat:variety interaction is only borderline (p = 0.0515), this masks a more important detail.

The model's coefficient table (summary(fit)) reveals a highly significant heat3:varietyR interaction (p = 0.0077). This specific interaction indicates that the effect of Heat Level 3 is not consistent for both varieties. The negative coefficient suggests that Variety 'R' (Furious) responds significantly worse to Heat Level 3 compared to the baseline 'F' (Redhot) variety.

Therefore, our top recommendation (Heat 3) is primarily driven by the strong performance of 'Redhot' under that condition and may not be the optimal heat level for 'Furious'.

Q3 — Does quality depend on Variety?

Answer: No. When averaged across all treatment conditions, there is no evidence of an overall difference in quality between the two varieties.

The Anova main effect for variety was not statistically significant (F = 0.013, p = 0.910). This is strongly supported by the emmeans pairwise contrast, which found a trivial and non-significant difference of only 0.10 units between the two varieties (p = 0.850).

This means that after accounting for all other factors (like blocking, side, and the significant interaction identified in Q2), neither variety provides a statistically reliable advantage overall. The important difference is not the main effect, but rather the interaction—how the varieties respond differently to specific heat levels.

Question 4: How big are the differences?

The environmental factors exerted a much larger influence on quality than the genetic variety. The main effects of light (F = 187.1) and heat (F = 27.7) were both highly significant (p < 0.0001) and were the most powerful drivers of chilli quality. Increasing light from Level 1 to Level 4, for example, could boost quality by over 15 units.

Crucially, the side main effect was also highly significant (F = 62.8, p < 0.0001). This indicates a strong, systematic difference in quality between the 'North' and 'South' positions within the greenhouses, validating its use as a key factor in the model.

In contrast, the main effect of variety (p = 0.91) and the heat:light interaction (p = 0.81) were both non-significant. Overall, light intensity was the dominant factor, followed by heat and the greenhouse side.

Compact Letter Display:

Table 5: CLD for Heat × Light Interactions

	heat	light	emmean	SE	df	lower.CL	upper.CL	.group
1	1	1	34.89	0.91	36	32.03	37.75	a
2	2	1	39.27	0.82	36	36.70	41.83	ab
4	4	1	40.89	0.98	36	37.82	43.97	bc
3	3	1	40.91	1.20	36	37.16	44.66	bc
5	1	2	44.86	1.46	36	40.28	49.43	bcd
6	2	2	45.83	0.93	36	42.93	48.74	cd
8	4	2	47.85	0.93	36	44.94	50.75	de
9	1	3	47.90	1.03	36	44.68	51.13	def
12	4	3	49.35	1.58	36	44.38	54.32	defg

	heat	light	emmean	SE	df	lower.CL	upper.CL	.group
7	3	2	49.44	0.69	36	47.26	51.61	def
13	1	4	50.50	0.90	36	47.68	53.31	defg
10	2	3	50.94	0.69	36	48.76	53.11	efg
11	3	3	52.85	0.89	36	50.07	55.64	fg
16	4	4	54.12	0.63	36	52.15	56.10	g
15	3	4	55.79	1.33	36	51.62	59.97	g
14	2	4	NA	NA	NA	NA	NA	
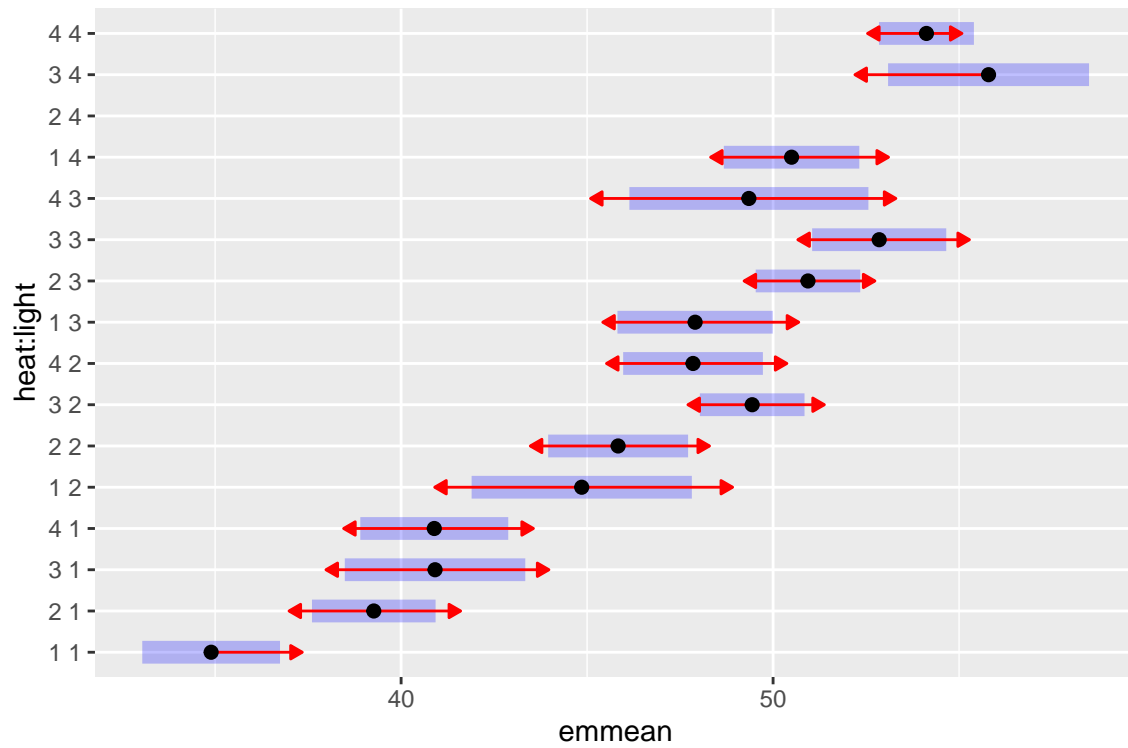
This display clearly shows that the top two combinations, (Heat 3, Light 4) and (Heat 4, Light 4), both share the letter 'g'. This confirms our earlier finding that while (H3, L4) has the highest mean, it is not statistically superior to (H4, L4).

Future Reccomendations:

Primary Recommendation: Based on this experiment, the setting of Heat 3 × Light 4 produced the highest mean quality (55.79).

- **Caveat:** This recommendation is strongest for the 'Redhot' (F) variety. Due to the significant heat3:varietyR interaction (p = 0.008), the 'Furious' (R) variety responded less well to Heat 3.

Cost-Benefit Considerations: For a more robust or cost-effective solution, the decision_tbl highlights practical alternatives:

1. Heat 4 × Light 4: This is the second-best option, achieving a mean quality (54.12) only 1.67 units below the maximum. This setting may be a safer general recommendation as it did not show a negative interaction with either variety.

2. Heat 3 × Light 3: This setting is a larger step down, with a mean quality (52.85) that is 2.94 units below the maximum.

Variety Selection: There is no evidence of an overall quality difference between Redhot and Furious (p = 0.85). The choice should instead be based on the intended heat setting. Redhot performs exceptionally well at Heat 3. If a farmer must use Heat 4 (perhaps for other reasons), the varieties may perform more similarly.

Positional Effects: The side effect was highly significant (p < 0.0001). This suggests a consistent environmental difference (e.g., sun exposure, airflow) between 'North' and 'South' sides that strongly impacts quality. Future planting and harvesting should account for this systematic variation.

Study Limitations

1. **Sample Size**: With a small sample size from resource limitations there are missing observation(s) thus our power to detect small effects is limited. Confidence intervals for some comparisons are wide.

2. **Missing Data:** The nonEst (non-estimable) result for the heat2:light4 combination in the emmeans output is due to missing data. This creates a "hole" in our understanding of the treatment landscape, and we cannot make recommendations for that specific setting.

3. **Complex Interactions:** We found one specific heat:variety interaction. It is possible that other, more complex three-way interactions (e.g., heat:light:variety) exist, but our model did not include this term.

4. **Quality Metric:** The composite score combines multiple factors. Future studies should examine yield, taste, and appearance separately.

5. **Greenhouse Effects:** Results are specific to this greenhouse setup and may not generalize to open-field conditions.

Future Directions

- Replicate the experiment with larger sample sizes to narrow confidence intervals
- Conduct economic analysis incorporating energy costs and market prices
- Separate quality into components (taste, yield, appearance) for targeted optimization
- Test intermediate settings between levels to fine-tune recommendations
- Evaluate performance under varying weather conditions across multiple years

———————————————————

References:

- STA2005S Experimental Design Project (2025) handout. Department of Statistical Sciences, UCT.
- R Core Team (2025). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna.
- Fox, J. & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). (used for `car::Anova`)
- Lenth, R. V. (2024). *emmeans: Estimated Marginal Means.* R package.
- Wickham, H., François, R., Henry, L., & Müller, K. (2023). *dplyr: A Grammar of Data Manipulation.* R package.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer. (used for plots)

Data and Design:

Table 6: Experimental Design Genrated Matrix

season	greenhouse	light	heat	variety	side	plot
1	A	2	2	R	n	1
1	A	3	1	R	n	2
1	A	1	4	R	n	3
1	A	4	3	R	n	4
1	A	4	3	R	s	5
1	A	2	4	R	s	6
1	A	3	2	R	s	7
1	A	1	1	R	s	8
1	A	2	3	F	n	9
1	A	1	1	F	n	10
1	A	4	4	F	n	11
1	A	3	2	F	n	12
1	A	4	4	F	s	13
1	A	3	3	F	s	14
1	A	2	1	F	s	15
1	A	1	2	F	s	16
1	B	3	1	R	n	17
1	B	4	4	R	n	18
1	B	1	3	R	n	19
1	B	2	2	R	n	20
1	B	2	3	R	s	21
1	B	1	4	R	s	22
1	B	3	2	R	s	23
1	B	4	1	R	s	24
1	B	1	2	F	n	25
1	B	2	3	F	n	26
1	B	4	4	F	n	27
1	B	3	1	F	n	28
1	B	4	1	F	s	29
1	B	3	3	F	s	30
1	B	1	2	F	s	31
1	B	2	4	F	s	32
2	A	2	4	R	n	33
2	A	4	3	R	n	34
2	A	1	1	R	n	35
2	A	3	2	R	n	36

season	greenhouse	light	heat	variety	side	plot
2	A	2	4	R	s	37
2	A	3	1	R	s	38
2	A	1	2	R	s	39
2	A	4	3	R	s	40
2	A	3	3	F	n	41
2	A	1	2	F	n	42
2	A	2	1	F	n	43
2	A	4	4	F	n	44
2	A	2	3	F	s	45
2	A	4	1	F	s	46
2	A	1	4	F	s	47
2	A	3	2	F	s	48
2	B	4	4	R	n	49
2	B	3	2	R	n	50
2	B	1	1	R	n	51
2	B	2	3	R	n	52
2	B	1	1	R	s	53
2	B	2	2	R	s	54
2	B	3	3	R	s	55
2	B	4	4	R	s	56
2	B	1	3	F	n	57
2	B	3	4	F	n	58
2	B	2	2	F	n	59
2	B	4	1	F	n	60
2	B	4	1	F	s	61
2	B	1	2	F	s	62
2	B	3	4	F	s	63
2	B	2	3	F	s	64

Table 7: Simulated observations

season	greenhouse	light	heat	variety	side	plot	obs
1	A	2	2	R	n	1	44.63388
1	A	3	1	R	n	2	50.48180
1	A	1	4	R	n	3	41.30101
1	A	4	3	R	n	4	55.11754
1	A	4	3	R	s	5	58.08898
1	A	2	4	R	s	6	50.62546
1	A	3	2	R	s	7	53.91521
1	A	1	1	R	s	8	38.67690
1	A	2	3	F	n	9	48.86032
1	A	1	1	F	n	10	32.95864
1	A	4	4	F	n	11	54.99408
1	A	3	2	F	n	12	50.56499
1	A	4	4	F	s	13	59.98049
1	A	3	3	F	s	14	55.91266
1	A	2	1	F	s	15	46.20276
1	A	1	2	F	s	16	41.54861
1	B	3	1	R	n	17	43.73744
1	B	4	4	R	n	18	51.71666
1	B	1	3	R	n	19	39.36048
1	B	2	2	R	n	20	45.56654
1	B	2	3	R	s	21	50.20601
1	B	1	4	R	s	22	41.45225
1	B	3	2	R	s	23	53.23059
1	B	4	1	R	s	24	53.54364
1	B	1	2	F	n	25	37.05544
1	B	2	3	F	n	26	46.88391
1	B	4	4	F	n	27	53.20711
1	B	3	1	F	n	28	46.02305
1	B	4	1	F	s	29	53.24998
1	B	3	3	F	s	30	56.90592
1	B	1	2	F	s	31	41.88629
1	B	2	4	F	s	32	47.85220
2	A	2	4	R	n	33	47.63880
2	A	4	3	R	n	34	50.72973
2	A	1	1	R	n	35	NA
2	A	3	2	R	n	36	46.43048
2	A	2	4	R	s	37	50.02303
2	A	3	1	R	s	38	NA
2	A	1	2	R	s	39	40.59188

season	greenhouse	light	heat	variety	side	plot	obs
2	A	4	3	R	s	40	53.85587
2	A	3	3	F	n	41	53.39276
2	A	1	2	F	n	42	37.54930
2	A	2	1	F	n	43	41.08095
2	A	4	4	F	n	44	49.81993
2	A	2	3	F	s	45	52.91419
2	A	4	1	F	s	46	50.15578
2	A	1	4	F	s	47	42.84430
2	A	3	2	F	s	48	52.32017
2	B	4	4	R	n	49	52.32779
2	B	3	2	R	n	50	49.44498
2	B	1	1	R	n	51	33.97181
2	B	2	3	R	n	52	44.80763
2	B	1	1	R	s	53	38.09752
2	B	2	2	R	s	54	48.20931
2	B	3	3	R	s	55	52.40630
2	B	4	4	R	s	56	52.43842
2	B	1	3	F	n	57	37.89270
2	B	3	4	F	n	58	46.54105
2	B	2	2	F	n	59	42.38080
2	B	4	1	F	n	60	45.73726
2	B	4	1	F	s	61	51.22709
2	B	1	2	F	s	62	38.03224
2	B	3	4	F	s	63	51.62355
2	B	2	3	F	s	64	53.64072

R code:

```
# Imports, libraries, and setup

#install.packages("EDproject_2.0.zip", repos = NULL, type = "win.binary")

library(EDproject)
library(ggplot2)
library(dplyr)
library(tidyr)
```

```r
library(EDproject)
library(car)
library(knitr)
library(emmeans)
library(multcomp)

# Set global format options
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE,
                      fig.width = 6, fig.height = 4, fig.align = "center")

# Making a color Palette for my document:
col_palette = c (
  "#0B1D39",
  "#333333",
  "#666666",
  "#D3D3D3",
  "#f8f8f8",
  "#F5F5F5",
  "#007ACC",
  "#1DB954",
  "#FF6F61",
  "#f9c74f"

)


# Outer Latin Square: Determines outer block position
outer_latin_square <- data.frame(
            season = c(1, 1, 2, 2),
            greenhouse = c("A", "B", "A", "B"),
            outer_block = c(1, 2, 3, 4)
)

# Assigning heat and light levels balanced
create_nested_ls <- function(block_id) {

  # 4 side-variety combos
  side_var_combos <- expand.grid(
    side = c("n", "s"),
    variety = c("R", "F")
  )
```

```r
  design <- data.frame()

  for (i in 1:nrow(side_var_combos)) {
    heat_order <- sample(1:4)
    light_order <- sample(1:4)
    combo_design <- data.frame(
      side = rep(side_var_combos$side[i], 4),
      variety = rep(side_var_combos$variety[i], 4),
      heat = heat_order,
      light = light_order
    )

    design <- rbind(design, combo_design)
  }

  return(design)
}

# Generating the design for all 4 outer blocks
all_designs <- list()

for (i in 1:nrow(outer_latin_square)){
  block_design <- create_nested_ls()
  block_design$season <- outer_latin_square$season[i]
  block_design$greenhouse <- outer_latin_square$greenhouse[i]
  block_design$outer_block <- outer_latin_square$outer_block[i]
  all_designs[[i]] <- block_design
}

# Combining blocks
final_design <- do.call(rbind, all_designs)
final_design$plot <- 1:nrow(final_design)

design_csv <- final_design[c("season",
                             "greenhouse", "light",
                             "heat", "variety", "side", "plot")]


cat("--- Design Validation Check ---\n\n")

# unique combinations of s and gh
all_main_blocks <- unique(design_csv[, c("season", "greenhouse")])
```

```r
for (i in 1:nrow(all_main_blocks)) {

  s <- all_main_blocks$season[i]
  gh <- all_main_blocks$greenhouse[i]

  cat(paste("=====================================\n"))
  cat(paste("    Checking Season:", s, "| Greenhouse:", gh, " \n"))
  cat(paste("=====================================\n"))


  current_block_data <- subset(design_csv, season == s & greenhouse == gh)

  side_variety_combos <- sort(unique(
    paste(current_block_data$side,
          current_block_data$variety)))


  for (sv in side_variety_combos) {
    cat("\n  Sub-block (Side / Variety):", sv, "\n")
    subset_data <- current_block_data[
      paste(current_block_data$side,
      current_block_data$variety) == sv, ]
    cat("    > Heat levels present: ", paste(
      sort(unique(subset_data$heat)), collapse=", "), "\n")
    cat("    > Light levels present:", paste(
      sort(unique(subset_data$light)), collapse=", "), "\n")
  }
  cat("\n")
}

cat("--- Validation Complete ---\n")

# Write file
write.csv(design_csv, "design.csv", row.names = FALSE, quote = FALSE)


# Obtaining observations:
design <- read.csv("design.csv")
mydata <- get.observations(design)

mydata
```

```r
# Convert variables to factors -----------------------------------------
dat <- na.omit(mydata)
dat$season      <- factor(dat$season)
dat$greenhouse  <- factor(dat$greenhouse)
dat$heat        <- factor(dat$heat)
dat$light       <- factor(dat$light)
dat$variety     <- factor(dat$variety)
dat$side        <- factor(dat$side)
dat$block       <- interaction(dat$season, dat$greenhouse, drop = TRUE)


# Fit model ------------------------------------------------------------
fit <- lm(obs ~ block + side + heat*light +
            variety + heat:variety + light:variety, data = dat)


# Model summary --------------------------------------------------------
summary(fit)


coef_table <- summary(fit)$coefficients
kable(coef_table, caption = "Model Coefficients Summary", digits = 3)


# Type II ANOVA
anova_results <- Anova(fit, type = 2)
kable(anova_results, caption = "Type II ANOVA Results")



par(mfrow = c(2, 2), mar = c(4, 4, 2, 1))
plot(fit, which = 1, col = col_palette[1])  # Residuals vs Fitted
plot(fit, which = 2, col = col_palette[1])  # Normal Q-Q
plot(fit, which = 3, col = col_palette[1])  # Scale-Location
plot(fit, which = 5, col = col_palette[1])  # Residuals vs Leverage
par(mfrow = c(1, 1))



# ===================================================================
#  Mean plots WITHOUT CIs
# ===================================================================


# Ensure factors are correctly coded

dat <- dat %>%
  mutate(
    heat    = factor(heat),
```

```r
    light   = factor(light),
    variety = factor(variety)
  )

# Summarise means

sumdat <- dat %>%
  group_by(light, heat, variety) %>%
  summarise(
    mean = mean(obs),
    .groups = "drop"
  )


# Heat × Light interaction plot

with(dat, interaction.plot(

  x.factor = heat,              # X-axis

  trace.factor = light,         # Lines

  response = obs,

  fun = mean,

  type = "b",

  col = c("black","red","blue","darkgreen"),

  pch = 16:19,

  xlab = "Heat Level",

  ylab = "Mean Quality",

  trace.label = "Light Level",

  main = "Interaction: Heat × Light"

))
```

```r
# Heat × Variety interaction plot

with(dat, interaction.plot(

  x.factor = heat,

  trace.factor = variety,

  response = obs,

  fun = mean,

  type = "b",

  col = c("salmon","darkcyan"),

  pch = 16:17,

  xlab = "Heat Level",

  ylab = "Mean Quality",

  trace.label = "Variety",

  main = "Interaction: Heat × Variety"

))

# Light × Variety interaction plot

with(dat, interaction.plot(

  x.factor = light,

  trace.factor = variety,

  response = obs,

  fun = mean,

  type = "b",
```

```r
  col = c("orange","orchid"),

  pch = 16:17,

  xlab = "Light Level",

  ylab = "Mean Quality",

  trace.label = "Variety",

  main = "Interaction: Light × Variety"

))


#Answer the farmers questions explicitly

# ======================
# Section 6 + 7: Code
# ======================
install.packages("multcomp")
# Drop missing rows (optional but useful for model fitting)
dat_complete <- na.omit(mydata)

# Create 'block' as interaction of season × greenhouse
dat_complete$block <- interaction(dat_complete$season,
                                  dat_complete$greenhouse, sep = ".")

# Convert all relevant variables to factors
dat_complete$season     <- factor(dat_complete$season)
dat_complete$greenhouse <- factor(dat_complete$greenhouse)
dat_complete$block      <- factor(dat_complete$block)
dat_complete$heat       <- factor(dat_complete$heat)
dat_complete$light      <- factor(dat_complete$light)
dat_complete$variety    <- factor(dat_complete$variety)
dat_complete$side       <- factor(dat_complete$side)

# =======================================================
# Step 2: Fit model (block as Season × Greenhouse)
# =======================================================

fit <- lm(obs ~ block + side + heat*light +
```

```r
              variety + heat:variety + light:variety, data = dat_complete)

# 1) Which heat & light give the best quality?
# ------------------------------------------
emm_hl <- emmeans(fit, ~ heat * light)    # EMMs for Heat x Light
emm_hl_sum <- summary(emm_hl)             # table with mean, SE, 95% CI
emm_hl_pairs <- pairs(emm_hl, adjust = "tukey")
# Tukey-adjusted pairwise tests

emm_hl_sum
emm_hl_pairs

# 2) Do settings depend on variety? (check interactions with variety)
# ------------------------------------------------------------------
#Anova(fit, type = 2)
# look at heat:variety, light:variety, and heat:light:variety rows

# 3) Does quality depend on variety? (main effect)
# -----------------------------------------------
emm_var <- emmeans(fit, ~ variety)
emm_var_sum <- summary(emm_var)
emm_var_sum

# 4) How big are the differences? (effect size with CI)
# ----------------------------------------------------
var_contr <- contrast(emm_var, "pairwise", adjust = "tukey")
var_contr

# 5) Build a small decision table for cost-aware recommendation (Section 7)
# ------------------------------------------------------------------------
# Rank Heat x Light means, compute difference from best (Δ)

# Create clean data frame
decision_df <- as.data.frame(emm_hl_sum)

# Remove problematic rows manually
decision_df <- decision_df[!is.na(decision_df$emmean)
                           & !is.na(decision_df$SE), ]

decision_tbl <- decision_df |>
  mutate(
    heat = as.numeric(as.character(heat)),
```

```r
    light = as.numeric(as.character(light)),
    mean = emmean,
    lower = lower.CL,
    upper = upper.CL
  ) |>
  arrange(desc(mean)) |>
  mutate(
    best_mean = first(mean),
    delta_vs_best = mean - best_mean
  ) |>
  dplyr::select(heat, light, mean, lower, upper, delta_vs_best)

decision_tbl

# Find cost-effective alternatives
near_optimal <- decision_tbl |>
  filter(delta_vs_best >= -3.0) |>
  mutate(energy_cost = heat + light) |>
  arrange(energy_cost)

near_optimal

library(multcompView)

cld_hl <- cld(emm_hl, adjust = "tukey", Letters = letters)
kable(cld_hl,
      caption = "CLD for Heat × Light Interactions",
      digits = 2)
plot(emm_hl, comparisons = TRUE)
```