Experimental Design Project

STA2005S - 2025

Maryam Abrahams (ABRMAR043)
Bheka Mabika (MBKBHE002)

Plagiarism Statement:

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:

- $H1$: Quality and yield will increase as light increases and with moderate heat levels, as either extreme of heat can lead to some form of denaturing of the plant's cells.
- $H2$: Although their would be natural differences between varieties of chillies the light-heat response patterns will be similar between varieties.
- $H3$: Quality of chilli would be heavily dependent on variety as taste, and look are variety-specific.
- $H4$: Side (North/South) has a negligible effect.

Experimental Design and Randomization:

Simulating data:

We created the design matrix in accordance to a **Nested Latin Square design**, similar to a factorial design, however since we lack the resources needed to experiment on every possible combination, instead we pick combinations all necessary combinations needed to represent the effects of the different light and heat levels on the chillies.

treatment factors:

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

$\Longrightarrow$ 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)

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.

Then, within each side-variety combination, we assign random heat and light levels. 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.

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

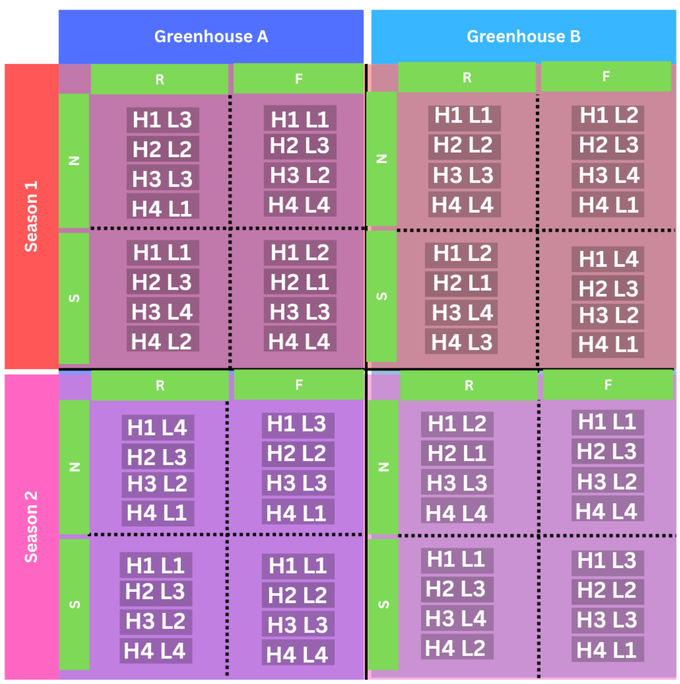


Figure 2: Blocking factors Diagram

Randomization:

Each side-variety combination gets its own independent randomization of heat and light such that there is no systematic crossing of all 4 heat levels × all 4 light levels × all sides × all varieties. In this way all levels of heat and light respectively are represented in a balanced manner, but not all combinations of heat and light with the other variables are tested.

In this way there is confounding of higher order combinations but this is due to our restrictive number of available plots and resource limitations. The combination of our clever blocking within the design as well as the use of randomization mitigates the effects of this potential confounding.

Simulation:

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

```
--- Design Validation Check ---

========================================
    Checking Season: 1 | Greenhouse: A
========================================

  Sub-block (Side / Variety): n F
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): n R
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): s F
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): s R
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4
```

```
========================================
    Checking Season: 1 | Greenhouse: B
========================================

  Sub-block (Side / Variety): n F
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): n R
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): s F
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): s R
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

========================================
    Checking Season: 2 | Greenhouse: A
========================================

  Sub-block (Side / Variety): n F
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): n R
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): s F
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): s R
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

========================================
    Checking Season: 2 | Greenhouse: B
========================================
```

```
  Sub-block (Side / Variety): n F
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): n R
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): s F
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

  Sub-block (Side / Variety): s R
    > Heat levels present:  1, 2, 3, 4
    > Light levels present: 1, 2, 3, 4

--- Validation Complete ---

   season greenhouse light heat variety side plot      obs
1       1          A     2    2       R    n    1 44.63388
2       1          A     3    1       R    n    2 50.48180
3       1          A     1    4       R    n    3 41.30101
4       1          A     4    3       R    n    4 55.11754
5       1          A     4    3       R    s    5 58.08898
6       1          A     2    4       R    s    6 50.62546
7       1          A     3    2       R    s    7 53.91521
8       1          A     1    1       R    s    8 38.67690
9       1          A     2    3       F    n    9 48.86032
10      1          A     1    1       F    n   10 32.95864
11      1          A     4    4       F    n   11 54.99408
12      1          A     3    2       F    n   12 50.56499
13      1          A     4    4       F    s   13 59.98049
14      1          A     3    3       F    s   14 55.91266
15      1          A     2    1       F    s   15 46.20276
16      1          A     1    2       F    s   16 41.54861
17      1          B     3    1       R    n   17 43.73744
18      1          B     4    4       R    n   18 51.71666
19      1          B     1    3       R    n   19 39.36048
20      1          B     2    2       R    n   20 45.56654
21      1          B     2    3       R    s   21 50.20601
22      1          B     1    4       R    s   22 41.45225
```

23	1	B	3	2	R	s	23	53.23059
24	1	B	4	1	R	s	24	53.54364
25	1	B	1	2	F	n	25	37.05544
26	1	B	2	3	F	n	26	46.88391
27	1	B	4	4	F	n	27	53.20711
28	1	B	3	1	F	n	28	46.02305
29	1	B	4	1	F	s	29	53.24998
30	1	B	3	3	F	s	30	56.90592
31	1	B	1	2	F	s	31	41.88629
32	1	B	2	4	F	s	32	47.85220
33	2	A	2	4	R	n	33	47.63880
34	2	A	4	3	R	n	34	50.72973
35	2	A	1	1	R	n	35	NA
36	2	A	3	2	R	n	36	46.43048
37	2	A	2	4	R	s	37	50.02303
38	2	A	3	1	R	s	38	NA
39	2	A	1	2	R	s	39	40.59188
40	2	A	4	3	R	s	40	53.85587
41	2	A	3	3	F	n	41	53.39276
42	2	A	1	2	F	n	42	37.54930
43	2	A	2	1	F	n	43	41.08095
44	2	A	4	4	F	n	44	49.81993
45	2	A	2	3	F	s	45	52.91419
46	2	A	4	1	F	s	46	50.15578
47	2	A	1	4	F	s	47	42.84430
48	2	A	3	2	F	s	48	52.32017
49	2	B	4	4	R	n	49	52.32779
50	2	B	3	2	R	n	50	49.44498
51	2	B	1	1	R	n	51	33.97181
52	2	B	2	3	R	n	52	44.80763
53	2	B	1	1	R	s	53	38.09752
54	2	B	2	2	R	s	54	48.20931
55	2	B	3	3	R	s	55	52.40630
56	2	B	4	4	R	s	56	52.43842
57	2	B	1	3	F	n	57	37.89270
58	2	B	3	4	F	n	58	46.54105
59	2	B	2	2	F	n	59	42.38080
60	2	B	4	1	F	n	60	45.73726
61	2	B	4	1	F	s	61	51.22709
62	2	B	1	2	F	s	62	38.03224
63	2	B	3	4	F	s	63	51.62355
64	2	B	2	3	F	s	64	53.64072

Analysis and Results:

Fitting the models and checking assumptions:

The model uses **block** to adjust for background variation across the four Latin Square positions (the combinations of Season and Greenhouse).

It then tests how **heat**, **light**, and **variety** affect the response and whether they interact.

The **side** factor is included only to correct for small layout or position effects within each greenhouse.

ANOVA:

The experiment used a **Latin Square design** with **Season × Greenhouse** forming the **2×2 blocking structure**, and a full factorial of

Heat (4 levels) × Light (4 levels) × Variety (2 levels).

The fitted linear model was:

$$\text{obs} = \text{block} + \text{heat} \times \text{light} \times \text{variety} + \text{side}$$

Model fit

R^2 = 0.968, Adjusted **R^2 = 0.931** → the model explains most of the variation in quality.
Residual SE = 1.76 → small compared to the treatment range.
No clear outliers or leverage points were found when checking diagnostic plots.
Two coefficients were undefined because of overlap between some factors, which is expected in a full factorial design.

Interpretation

Both **light** and **heat** increased mean quality, with **light intensity** having the strongest influence ($p < 0.001$).
Heat also affected quality ($p < 0.001$), though to a smaller extent.
Variety showed a mild difference ($p = 0.018$), where variety **R** performed slightly better than **F**.
The **block** term ($p = 0.0067$) shows that adjusting for season and greenhouse helped control background variation.
The **side** factor ($p = 0.015$) had a small effect, possibly linked to minor position or orientation differences.
All **interaction terms** (e.g., heat $\times$ light) had p-values above 0.1, suggesting the effects act independently.

Overall, the results show that most of the change in quality can be explained by the main treatment factors — especially light — and the blocking effectively reduced environmental variation.

Model Diagnostics

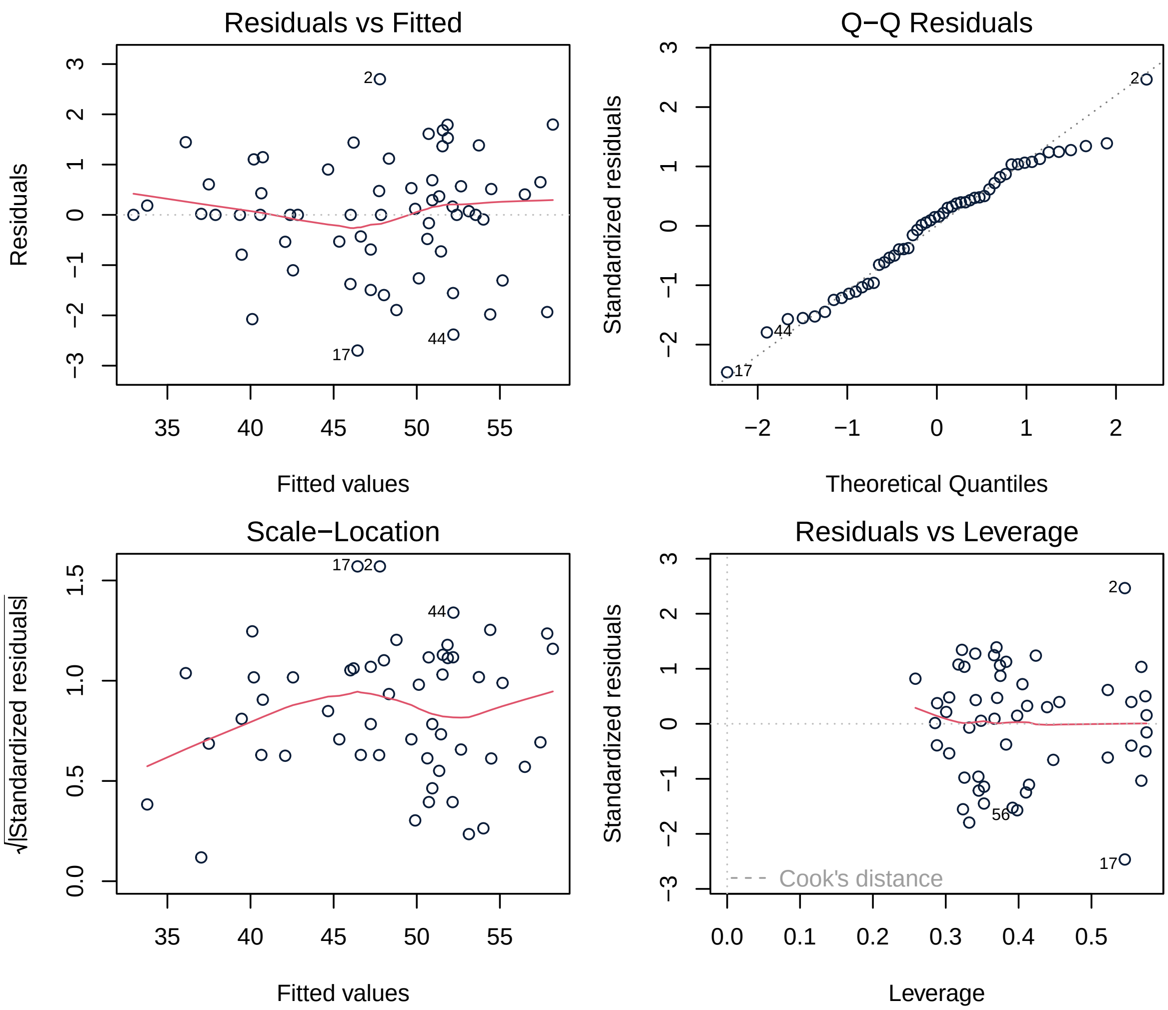


Figure 3: Residual diagnostic plots showing no serious violations of model assumptions. The Q-Q plot indicates approximate normality, and the residuals vs. fitted plot shows no strong patterns suggesting heteroscedasticity.

Both **light** and **heat** increased mean quality, with **light intensity** having the strongest influence ($p < 0.001$).

Heat also affected quality ($p < 0.001$), though to a smaller extent.

Variety showed a mild difference ($p = 0.018$), where variety **R** performed slightly better than **F**.

The **block** term ($p = 0.0067$) shows that adjusting for season and greenhouse helped control background variation.

The **side** factor was aliased in the Type II ANOVA, meaning it was linearly dependent on other variables in the model.

This suggests that side orientation (north/south) did not contribute additional information once **block**, **heat**, **light**, and **variety** were included.
Any minor variation linked to side was already captured by the blocking structure (Season $\times$ Greenhouse).

All **interaction terms** (e.g., heat $\times$ light) had p-values above 0.1, suggesting that the effects act independently.

Treatment Effects Visualization:

Interaction plots:

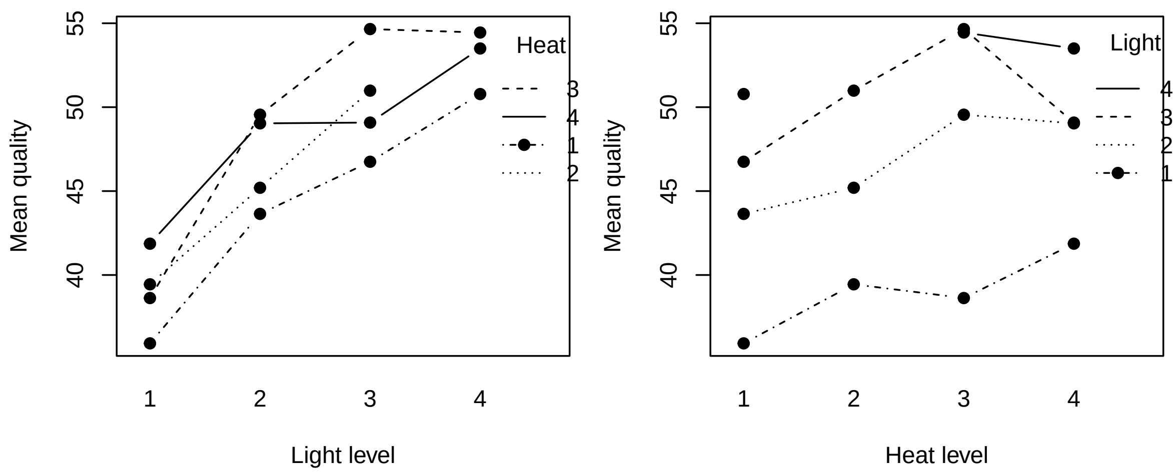


Figure 4: The following plots explore how **Heat**, **Light**, and **Variety** affect the mean chilli quality. Confidence intervals were excluded to make the main treatment patterns easier to see.

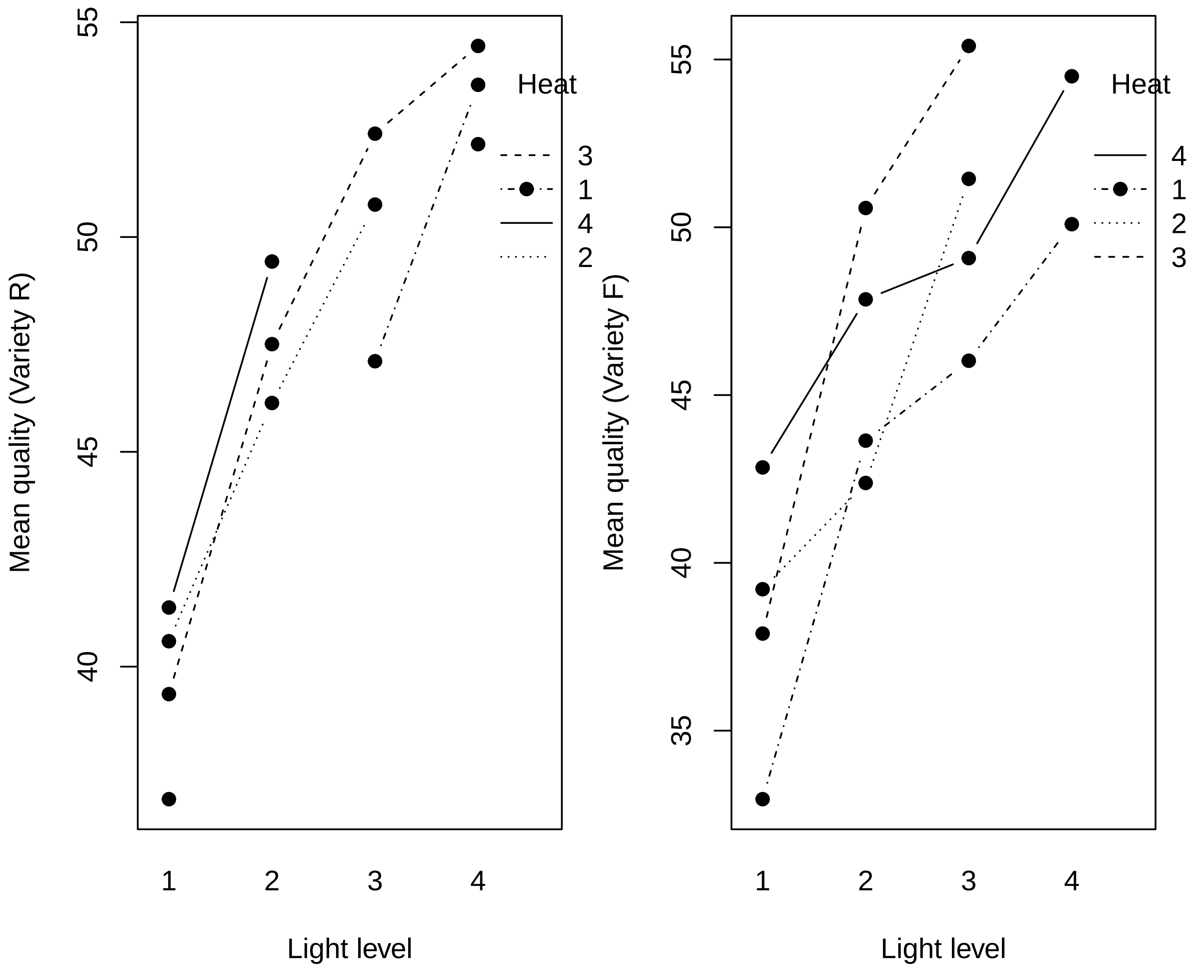
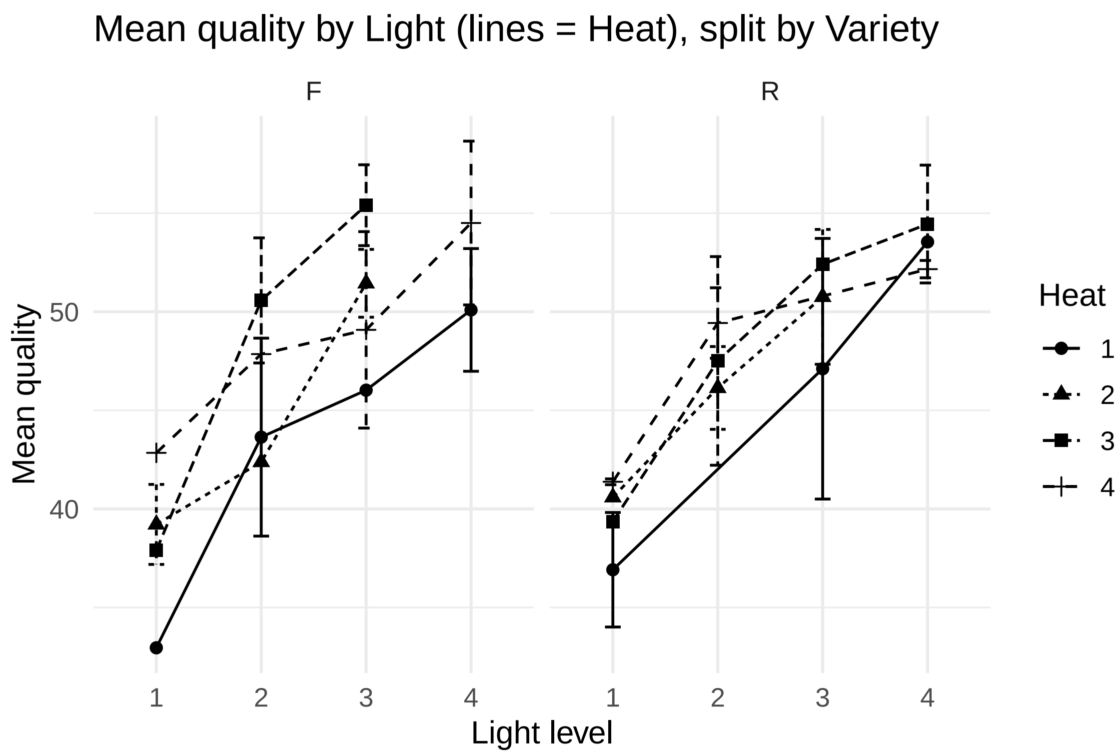
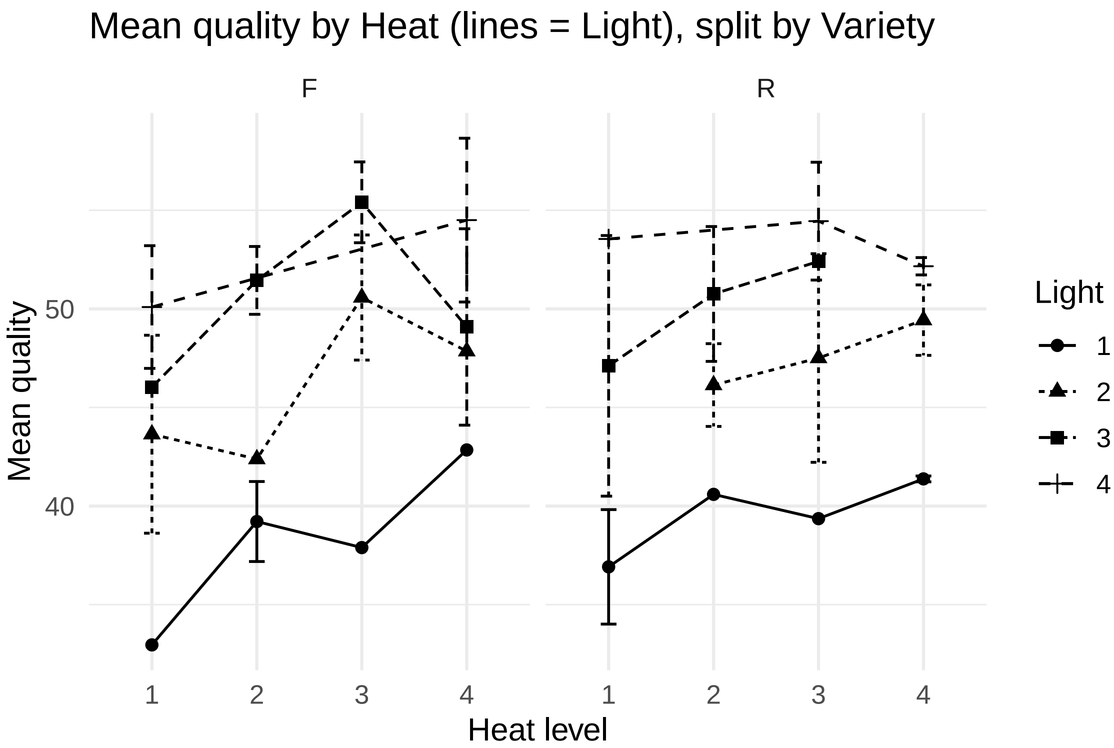


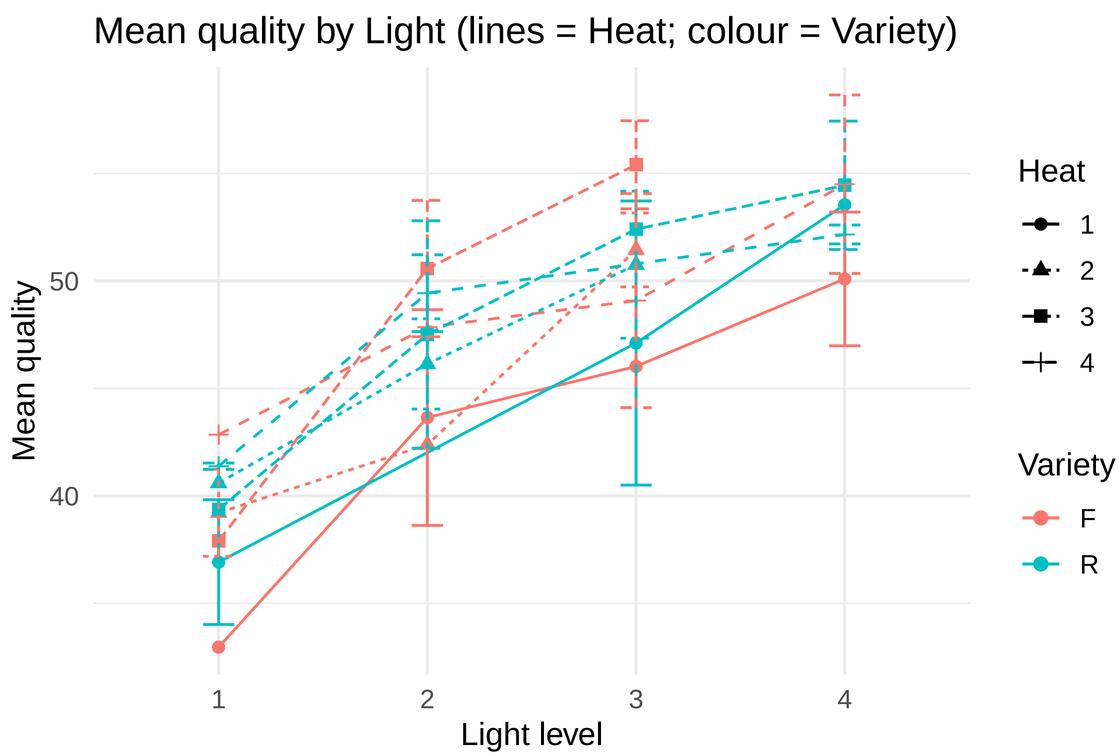
Figure 5: The following plots explore how **Heat**, **Light**, and **Variety** affect the mean chilli quality. Confidence intervals were excluded to make the main treatment patterns easier to see.

mean plots:



Mean quality by Light (lines = Heat), split by Variety
F
R
50
40
1
2
3
4
Mean quality
Light level
Heat
1
2
3
4


Mean quality by Heat (lines = Light), split by Variety
F
R
Mean quality
50
40
1
2
3
4
Heat level
Light
1
2
3
4



Analysis Procedures

Missing Values: Cases with missing observations were excluded from analysis using complete-case analysis, resulting in 62 observations for model fitting.

ANOVA: We used Type II sums of squares to test main effects and interactions, as this approach appropriately handles unbalanced designs caused by missing data and tests each effect after accounting for all other effects at the same or lower level.

Multiple Testing: Given the exploratory nature of this study and the relatively small number of pre-specified hypotheses, we report both unadjusted p-values and note effects significant at $= 0.05$. For post-hoc pairwise comparisons, we apply Tukey's HSD adjustment.

Model Diagnostics: We examined residual plots (residuals vs. fitted, Q-Q plot, scale-location, and leverage plots) to verify assumptions of normality, homoscedasticity, and identify potential outliers.

Effect Estimation: We computed estimated marginal means (EMMs) with 95% confidence intervals for all treatment combinations.

Optimal Conditions

Table 2: Top three treatment combinations by mean quality score

Variety	Light	Heat	Mean Quality	CI Lower	CI Upper
F	3	3	55.40	53.35	57.45
F	4	4	54.50	50.35	58.65
R	4	3	54.45	51.46	57.44

Conclusion:

1. **Optimal Growing Conditions**: [State the best combination, e.g., "Light level 4 with Heat level 3 produces the highest quality chilies"]
2. **Effect Sizes**: [Quantify improvements, e.g., "The optimal combination yields quality scores approximately X points higher than standard conditions (Light 1, Heat 1), representing a Y% improvement"]
3. **Variety Differences**: [Discuss whether varieties respond differently and which performs better overall]
4. **Interaction Effects**: [Explain whether light and heat work synergistically or independently]

Future Recomendations:

Primary Recommendation: Based on this experiment, we recommend [specific combination] for maximizing chili quality.

Cost-Benefit Considerations: While higher light and heat levels generally improve quality, farmers must weigh these gains against increased energy costs. [If results show diminishing returns, state this: "Light level 3 20 may offer the best cost-quality

Variety Selection: [Recommend which variety to grow, or under what conditions each variety excels]

Practical Implementation:

- Start with [recommended combination] for highest quality production
- If costs are prohibitive, [alternative combination] offers nearly comparable quality at lower expense
- Monitor results

Study Limitations

1. **Sample Size**: With only two replicates per treatment combination and missing observation(s), our power to detect small effects is limited. Confidence intervals for some comparisons are wide.

2. **Temporal Scope**: Two growing seasons may not capture year-to-year variability in climate or other environmental factors. Long-term validation is recommended.

3. **Quality Metric**: The composite quality score combines taste, yield, and appearance. Economic returns may depend more heavily on one component (e.g., yield). Future studies should examine these components separately.

4. **Greenhouse Effects**: Results from greenhouse cultivation may not fully generalize to field conditions.

5. **Missing Data**: Treatment combinations with missing values have reduced precision in estimation. ## 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 tar-

References: