# Modeling Perceived Lightness

Mattia Moro, Renqing Cuomao, Tessnim Bechrifa, Daniel Bucher

April 2025

## 1 Introduction/Background

We want to answer the following scientific question:

**Scientific Question**

What is the importance of background, hue and illuminance on the perceived lightness of reflecting surfaces?

To answer the scientific question we choose the following explanatory factors:

- **Background Reflectance**: Categorical factor with White, Gray and Black.
- **Hue**: Categorical factor with Red, Yellow, Green and Blue
- **Illuminance or Intensity**: Continuous factor measured in the foot-candle unit. four levels defined with 1.4, 4.5, 72 and 135 foot-candles.

In the following sections we explain how we measure their influence on a lighting score with the help of analysis of variance (ANOVA). The end goal is to have a linear model of the following form:

$$\begin{aligned}
\hat{y} = \beta_0 \\
&+ \beta_1 \cdot \mathbf{1}(\text{Background} = \text{White}) + \beta_2 \cdot \mathbf{1}(\text{Background} = \text{Gray}) + \beta_3 \cdot \mathbf{1}(\text{Background} = \text{Black}) \\
&+ \beta_4 \cdot \mathbf{1}(\text{Hue} = \text{Red}) + \beta_5 \cdot \mathbf{1}(\text{Hue} = \text{Yellow}) + \beta_6 \cdot \mathbf{1}(\text{Hue} = \text{Green}) + \beta_7 \cdot \mathbf{1}(\text{Hue} = \text{Blue}) \\
&+ \beta_8 \cdot \mathbf{1}(\text{Intensity} = 1.4) + \beta_9 \cdot \mathbf{1}(\text{Intensity} = 4.5) + \beta_{10} \cdot \mathbf{1}(\text{Intensity} = 72) \\
&+ \beta_{11} \cdot \mathbf{1}(\text{Intensity} = 135)
\end{aligned}$$

where

- $\hat{y}$: Predicted Lighting Score
- $\beta_0$: Intercept; model's predicted value (or mean response) when all predictors are set to their baseline (reference) levels.
- $\mathbf{1}(\cdot)$: Indicator function that equals 1 if the condition is true, 0 otherwise
- $\beta_1$ ... $\beta_{11}$ : Additional effect on Lighting Score when corresponding coefficient is 1

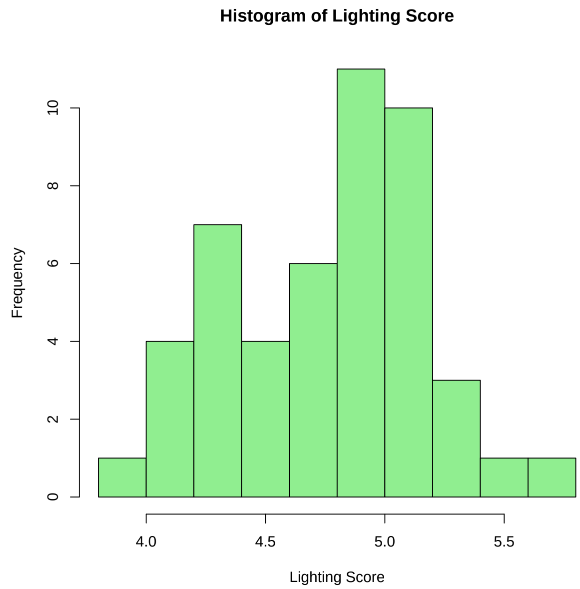



Figure 1: Histogram of observed lighting scores.

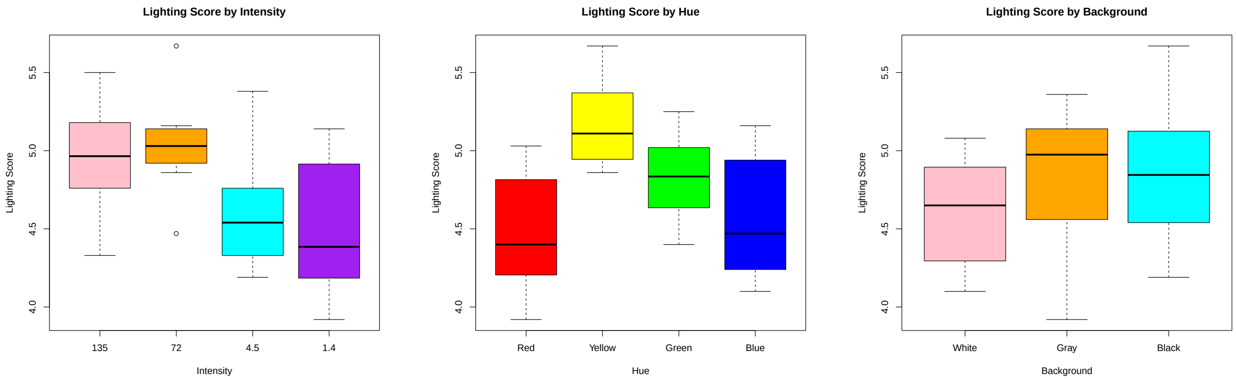
# 2 EDA

First, we start by exploring the structure of our response variable. To do this, we generate a histogram of the observed lighting scores. We observe that the data approximately follow a bimodal distribution, with the two modes occurring around 4.25 and 5.

We then proceed to generate a five-number summary of the observed intensity scores ( Figure 1). We obtain a minimum of 3.92, a median of 4.86, a mean of 4.76, and a maximum of 5.67. Since the mean and the median are quite similar, we can say that the distribution of our response variable is fairly symmetric.

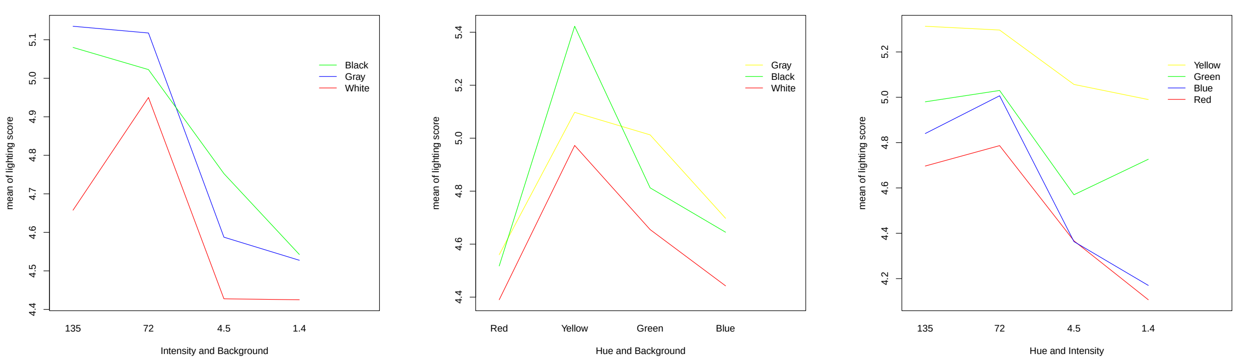
To inspect the effects of intensity, hue, and background reflectance on the lighting score, we generate different boxplots of the observations, grouping them each time by the explanatory variable. For intensity, we observe that higher levels are associated with higher lighting scores (Figure 2a). For hue, we observe that yellow and green hues are associated with higher lighting scores, while blue and red are associated with lower ones (Figure 2b). Finally, for background reflectance, we observe that the lighting score decreases as we go from gray to black to white (Figure 2c).

To assess whether there are interactions between the explanatory variables, we create interaction plots. Looking at the interaction plot between intensity and background (Figure 3a), we see that the lines are not perfectly parallel, suggesting a potential interaction effect. Repeating the procedure for hue and background (Figure 3b), we observe a similar interaction pattern. Finally, looking at the interaction plot between hue and intensity, we see a similar effect, which is more pronounced for lower intensity scores (Figure 3c). These observations suggest that interaction effects are worth considering, but further analyses are needed.




(a) Lighting by intensity (b) Lighting by hue (c) Lighting by background

Figure 2: Lighting Score by the explanatory factors




(a) Lighting score by intensity and background (b) Lighting score by hue and background (c) Lighting score by intensity and hue

Figure 3: Interaction plots showing the combined effect of the explanatory factors

# 3 Model Fitting

We analyze the `lightsat.txt` dataset, which contains subjective ratings of lighting conditions based on combinations of visual factors. The response variable is `Lighting_Score`, a continuous measure of perceived lighting quality. The predictors are three categorical factors:

- **Background Reflectance**: with levels White, Gray, and Black.
- **Hue**: with levels Red, Yellow, Green, and Blue.
- **Intensity (Illuminance)**: with levels 135, 72, 4.5, and 1.4 foot-candles.

We treat all three predictors as categorical variables and define the following reference levels:

- `Background = White`
- `Hue = Red`
- `Intensity = 135`

We begin by fitting a main-effects ANOVA model using R:

```
lightsat$Background <- relevel(lightsat$Background, "White")
lightsat$Hue <- relevel(lightsat$Hue, "Red")
lightsat$Intensity <- relevel(lightsat$Intensity, "135")
mod_main <- aov(Lighting_Score ~ Background + Hue + Intensity, data = lightsat)
```

The ANOVA table for the main-effects model is:

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Background | 2 | 0.5678 | 0.2839 | 7.38 | 0.00191 ** |
| Hue | 3 | 3.2168 | 1.0723 | 27.89 | 8.38e-10 *** |
| Intensity | 3 | 2.5110 | 0.8370 | 21.77 | 1.90e-08 *** |
| Residuals | 39 | 1.4995 | 0.0384 | | |

All three main effects are statistically significant at the 5% level.

To explore potential interactions, we fit a full factorial model:

```
mod_full <- aov(Lighting_Score ~ Background * Hue * Intensity, data = lightsat)
```

The three-way interaction term was found to be non-significant and was removed. We then fit a simplified interaction model including only pairwise interactions:

```
mod_pairwise <- aov(Lighting_Score ~ Background + Hue + Intensity
                                  + Background:Hue
                                  + Background:Intensity
                                  + Hue:Intensity, data = lightsat)
```

ANOVA table for this model is:

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Background | 2 | 0.5678 | 0.2839 | 10.96 | 0.00077 *** |
| Hue | 3 | 3.2168 | 1.0723 | 41.40 | 2.77e-08 *** |
| Intensity | 3 | 2.5110 | 0.8370 | 32.32 | 1.85e-07 *** |
| Background:Hue | 6 | 0.3284 | 0.0547 | 2.11 | 0.1022 |
| Background:Intensity | 6 | 0.2787 | 0.0464 | 1.79 | 0.1572 |
| Hue:Intensity | 9 | 0.4262 | 0.0474 | 1.83 | 0.1319 |
| Residuals | 18 | 0.4662 | 0.0259 | | |

We then formally compare the two models using a partial $F$-test:

| | Res.Df | RSS | Df | Sum of Sq | F | Pr(>F) |
|---|---|---|---|---|---|---|
| M1 | 39 | 1.4995 | | | | |
| M2 | 18 | 0.4662 | 21 | 1.0333 | 1.900 | 0.0869 |

Since the p-value ($p = 0.0869$) is above the 0.05 threshold, the addition of the interaction terms does not significantly improve the model. We therefore retain the main-effects model for further interpretation and diagnostics.

# Model Assessment

To ensure the validity of our ANOVA model, we carefully evaluated the standard assumptions. These are:

1. **Errors have mean zero:** The average of the residuals should be approximately zero. This can be visually assessed in residual plots, where residuals should be symmetrically distributed around the horizontal axis at zero.

2. **Errors are normally distributed:** Residuals should follow a normal distribution. This is checked using a Q-Q plot and a boxplot of residuals.

3. **Errors have constant variance (homoscedasticity):** The spread of the residuals should be roughly constant across all fitted values or groups. We test this with Levene's test and residual plots.

4. **Errors are independent:** Residuals should not be correlated with each other. Independence is primarily ensured by the experimental design (randomization), and can be further inspected via residual patterns.

## Normality and Mean Zero of Residuals

- **Q-Q plot of residuals (Figure 4a):** The residuals closely follow the reference line, which supports the assumption that they are approximately normally distributed.

- **Boxplot of residuals (Figure 4b):** The distribution appears symmetric enough and centered around zero. This supports both normality and the assumption that errors have mean zero.

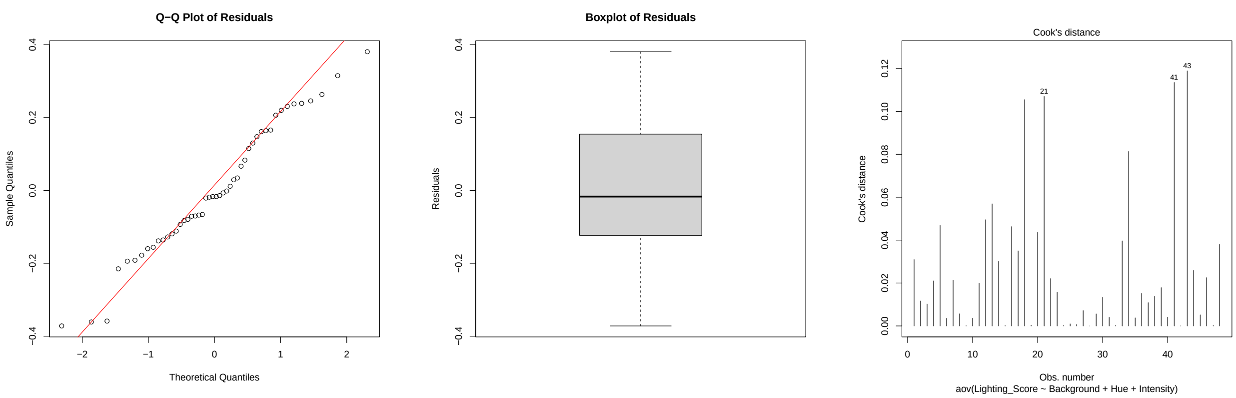
- **Cook's distance (Figure 4c):** A few data points (notably observations 21, 41, and 43) have slightly higher influence, but none are extreme enough to drastically affect the model. This confirms that no single observation has an too much impact.

### Homoscedasticity and Independence

- **Residuals vs. fitted values (Figure 5a):** The residuals are mostly centered around 0, but still show a shallow curved shape. This suggests the presence of a trend, but still with limited 'damage'.

- **Levene's test:** Applied across Background, Hue, and Intensity, the test returned non-significant $p$-values ($p > 0.05$), which supports the assumption of homoscedasticity.

- **Residuals vs. group means (Figure 5b):** The spread of residuals appears consistent across fitted values, with no systematic trend, suggesting independence and equal variance.

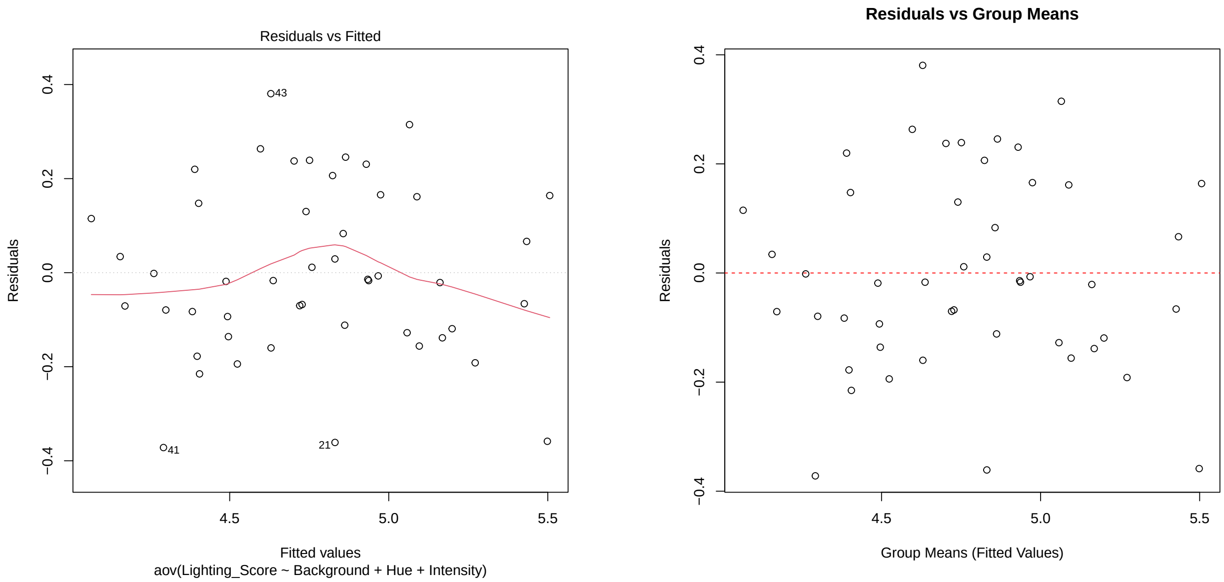



(a) Q-Q plot of residuals (b) Boxplot of residuals (c) Cook's distance

Figure 4: The plots confirm that residuals are approximately normally distributed and centered around zero. Although a few points show higher influence, especially visible from cook's distance plot, they remain within acceptable limits

## Final Estimated Model (main effects only)

The interaction terms were dropped because none were statistically significant at the 5 % level (see Model Fitting section). Using **White** background, **Red** hue, and **135 fc** intensity as the reference levels, the fitted main–effects model is




(a) residuals vs fitted values (b) residuals vs group means

Figure 5: Residuals show no clear pattern or heteroscedasticity and remain roughly centered around zero, supporting constant variance and independence assumptions, despite the shallow trend in the residual vs fitted value

$$\begin{aligned}\widehat{\text{LightingScore}} &= 4.5 \\ &\quad + 0.22\,\mathbf{1}\big(B = \text{Gray}\big) + 0.23\,\mathbf{1}\big(B = \text{Black}\big) \\ &\quad + 0.67\,\mathbf{1}\big(H = \text{Yellow}\big) + 0.33\,\mathbf{1}\big(H = \text{Green}\big) + 0.10\,\mathbf{1}\big(H = \text{Blue}\big) \\ &\quad - 0.45\,\mathbf{1}\big(I = 1.4 \text{ fc}\big) - 0.36\,\mathbf{1}\big(I = 4.5 \text{ fc}\big) + 0.22\,\mathbf{1}\big(I = 72 \text{ fc}\big)\end{aligned}$$

where $\mathbf{1}(\cdot)$ is the indicator function (1 if the condition is true, 0 otherwise). All coefficients are shown with a maximum of two significant digits.

# 4 Conclusion

**Recap.** Exploratory plots hinted that *background*, *hue* and *intensity* all matter, but formal ANOVA showed only the three main effects are significant ($p < 0.002$) while every interaction is not ($p > 0.10$). We therefore kept a main-effects model with White / Red / 135 fc as reference levels.

**Interpretation (numbers = fitted mean change).**

- *Background.* Darker surrounds make the sample look lighter: Gray +0.22, Black +0.23.
- *Hue.* Perceived lightness rises from Red (0) → Blue +0.10 → Green +0.33 → Yellow +0.67.

- *Intensity.* Relative to 135 fc, mid–high 72 fc is slightly brighter (+0.22) while low levels are dimmer (1.4 fc −0.45, 4.5 fc −0.36).

**Takeaways.** Place the target against a dark background, use yellow or green hues, and light it at a mid-high (not maximal) intensity to maximise perceived lightness. The model explains 76 % of the variance; changes smaller than $\approx 0.2$ units are likely below perceptual threshold. Future work should sample more intensity levels and a larger observer pool to test subtler interactions.