HW 7

No name needed

Due 3/14/2025

# HW 7 Instructions:

* This is an individual assignment, but you may discuss the work with other students, which must be documented in the last question. The answers should be your own.

Continuing with Fabusova et al. (2024) and the provided data set (“moths.xlsx”) to answer the following questions.

* Fabusova M, Gaston KJ, Troscianko J. 2024 Pulsed artificial light at night alters moth flight behaviour. Biol. Lett. 20: 20240403. <https://doi.org/10.1098/rsbl.2024.0403>

library(readxl)  
moth <- read\_excel("moths.xlsx",   
 sheet = "complete\_dataset\_MAIN FAMILIES.",   
 na = "NA")  
  
library(lubridate)  
  
moth <- moth %>% mutate(Date = factor(ymd(Date))) %>% dplyr::select(-Sunset\_time)  
  
moth2 <- moth %>% drop\_na(Date, ID, Location, Treatment, Treatment\_y\_n, Colour, Spectra, Moon\_Phase, Temp, Wind\_speed\_ms, Humidity, Common\_name, Latin\_name, Family, Sub\_family, Latency) #Slight change from prior HW to keep more observations!  
  
moth2 <- moth2 %>% mutate(Treatment = fct\_recode(factor(Treatment),  
 "Cold Phosphor" = "1",  
 "Warm Phosphor" = "2",  
 "Cold RGB" = "3",  
 "Warm RGB" = "4"), #5 recoding not needed - controls dropped  
 logLatency = log(Latency),  
 Colour = factor(Colour),  
 Spectra = factor(Spectra))

Note: Use moth2 for the remaining questions.

To avoid 2-WAY ANOVA interaction models prior to discussing them fully, we fit the “Treatment” variable as a four level. For now, we will focus just on the randomly assigned predictor part of the model only and compare using the two individual binary variables to using Treatment that we used previously.

1. The linear model with just Treatment is provided. Add the coefficients to the estimated model:

* where is 1 for a Warm Phosphorous observation and 0 otherwise, is 1 for a Cold RGB observation and 0 otherwise, and is 1 for a Warm RGB observation and 0 otherwise.

lmTreat <- lm(logLatency ~ Treatment, data = moth2)  
lmTreat %>% tbl\_regression(intercept = T)

| **Characteristic** | **Beta** | **95% CI***1* | **p-value** |
| --- | --- | --- | --- |
| (Intercept) | 1.8 | 1.5, 2.0 | <0.001 |
| Treatment |  |  |  |
| Cold Phosphor | — | — |  |
| Warm Phosphor | 0.65 | 0.25, 1.0 | 0.001 |
| Cold RGB | 0.64 | 0.24, 1.0 | 0.002 |
| Warm RGB | 0.81 | 0.42, 1.2 | <0.001 |
| *1*CI = Confidence Interval | | | |

1. Generate estimated means for both the Cold Phosphorous level and the Warm RGB level based on the estimated model you wrote out in question 2. Show your work.

**Estimated mean for Cold Phosphorous is 1.8 log seconds, as cold phosphor is the baseline so it is just the intercept. Estimated mean for warm RGB is 2.61 log seconds.**

coldP <- 1.8   
coldP

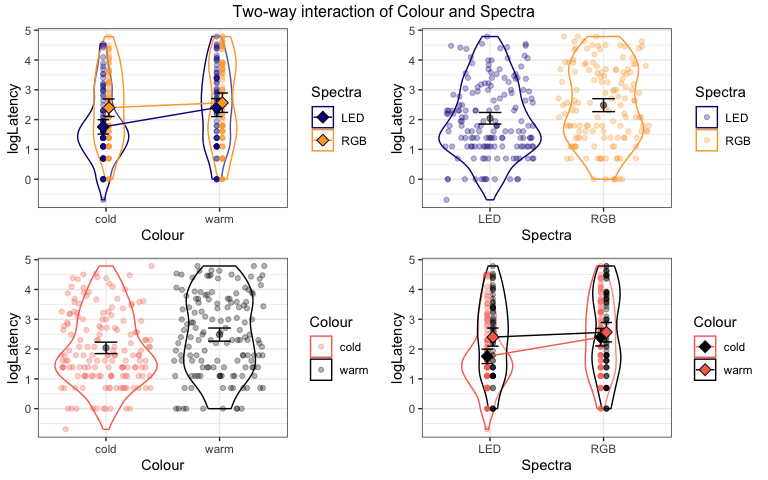
## [1] 1.8

warmRGB <- 1.8 + 0.81  
warmRGB

## [1] 2.61

1. Make an interaction plot based on the Colour and Spectra. No discussion.

moth2 %>% ggintplot(response = 'logLatency', groupvars = c('Colour','Spectra'))



1. Fit a linear model with Colour, Spectra, and the interaction of those two variables. Write out the estimated model, defining your indicator variables.

* where is 1 for a Warm observation and 0 otherwise, and is 1 for a RGB observation and 0 otherwise

int\_model <- lm(logLatency ~ Colour \* Spectra, data = moth2)  
int\_model %>% tbl\_regression(intercept = T)

| **Characteristic** | **Beta** | **95% CI***1* | **p-value** |
| --- | --- | --- | --- |
| (Intercept) | 1.8 | 1.5, 2.0 | <0.001 |
| Colour |  |  |  |
| cold | — | — |  |
| warm | 0.65 | 0.25, 1.0 | 0.001 |
| Spectra |  |  |  |
| LED | — | — |  |
| RGB | 0.64 | 0.24, 1.0 | 0.002 |
| Colour \* Spectra |  |  |  |
| warm \* RGB | -0.48 | -1.1, 0.09 | 0.10 |
| *1*CI = Confidence Interval | | | |

1. Use the interaction model to generate estimated means for both a Cold, Phosphorous level observation and a Warm, RGB observation based on the estimated model you wrote out in question 5. Show your work.

* Note: These should match prior answers, but the work should be different.

**Estimated mean for Cold Phosphorous is 1.8 log seconds, as cold phosphor is the baseline so it is just the intercept. Estimated mean for warm RGB is 2.61 log seconds, with beta coefficients for both Warm and RGB being added and the interaction term subtracted. (See work below)**

coldP <- 1.8  
coldP

## [1] 1.8

warmRGB <- 1.8 + 0.65 + 0.64 - 0.48  
warmRGB

## [1] 2.61

1. Generate an F-test for the interaction of Colour and Spectra from the previous interaction model and report an evidence sentence for the interaction.

**There is moderate [weak] evidence against the null hypothesis of no interaction of colour and spectra on the log-latency time (F(1,299) = 2.7, p-value = 0.1) so we conclude there is [not] an interaction and keep it [drop it] from the model.**

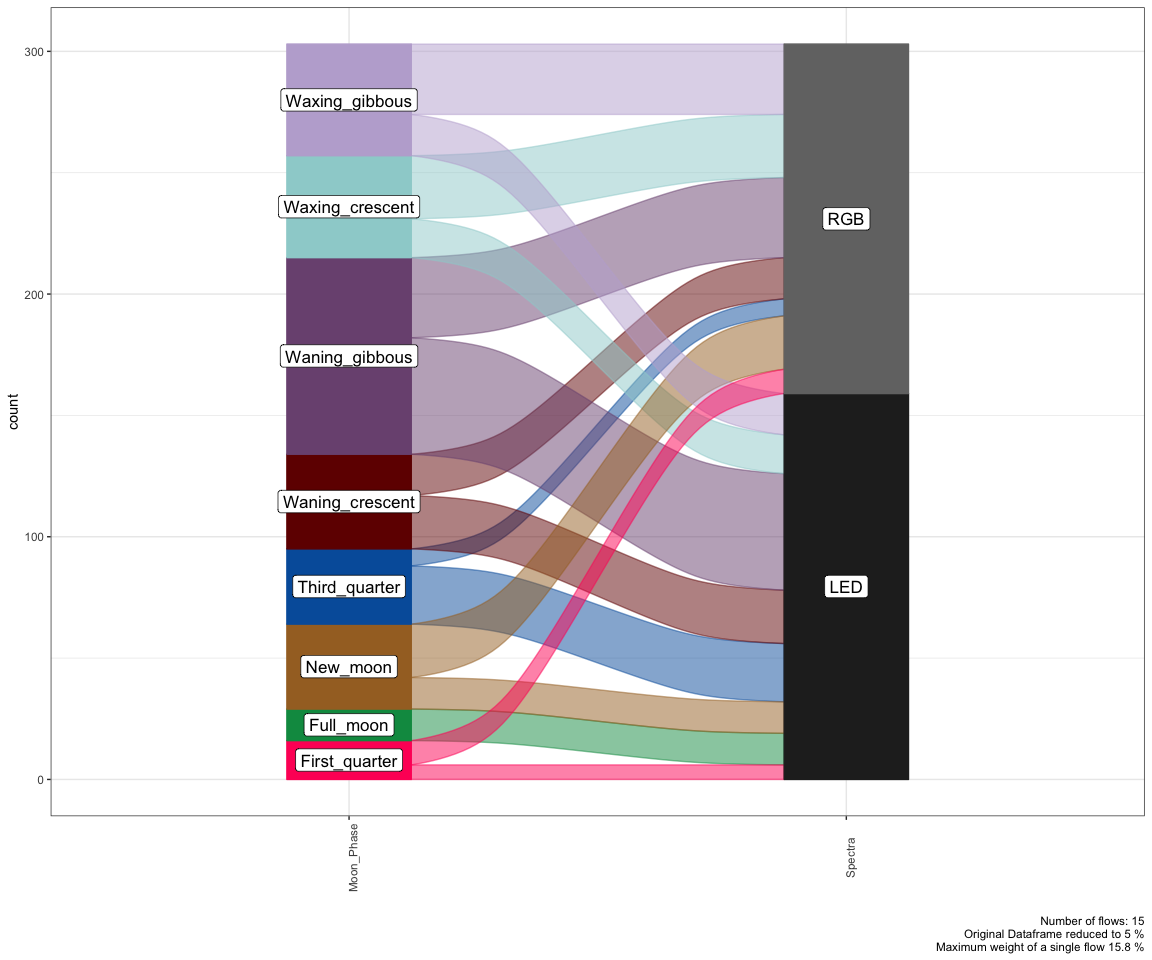
Anova(int\_model)

## Anova Table (Type II tests)  
##   
## Response: logLatency  
## Sum Sq Df F value Pr(>F)  
## Colour 13.14 1 8.2350 0.004402  
## Spectra 12.77 1 8.0049 0.004981  
## Colour:Spectra 4.31 1 2.7012 0.101323  
## Residuals 477.03 299

1. The researchers might have been interested in the interaction between moon phase and the spectra. Explain why it is not possible to estimate and test for an interaction of these two variables using the provided alluvial diagram and possibly other information you can find from the data set.

**Looking at the alluvial diagram, the full moon phase is only associated with the LED spectra level. This restricts moon phase to only the LED level of the spectra. For an interaction you need variation in both variables across ALL levels, and we do not have that.**

moth2 %>% dplyr::select(Moon\_Phase, Spectra) %>% alluvial\_wide(fill\_by = "first\_variable")



tally(Moon\_Phase ~ Spectra, data = moth2)

## Spectra  
## Moon\_Phase LED RGB  
## First\_quarter 6 10  
## Full\_moon 13 0  
## New\_moon 13 22  
## Third\_quarter 24 7  
## Waning\_crescent 22 17  
## Waning\_gibbous 48 33  
## Waxing\_crescent 16 26  
## Waxing\_gibbous 17 29

1. Suppose that the researchers had posited that the effects of the spectral type on logLatency might differ based on the family of the moths. Select *ALL* of the following that are correct based on the ggintplot results.

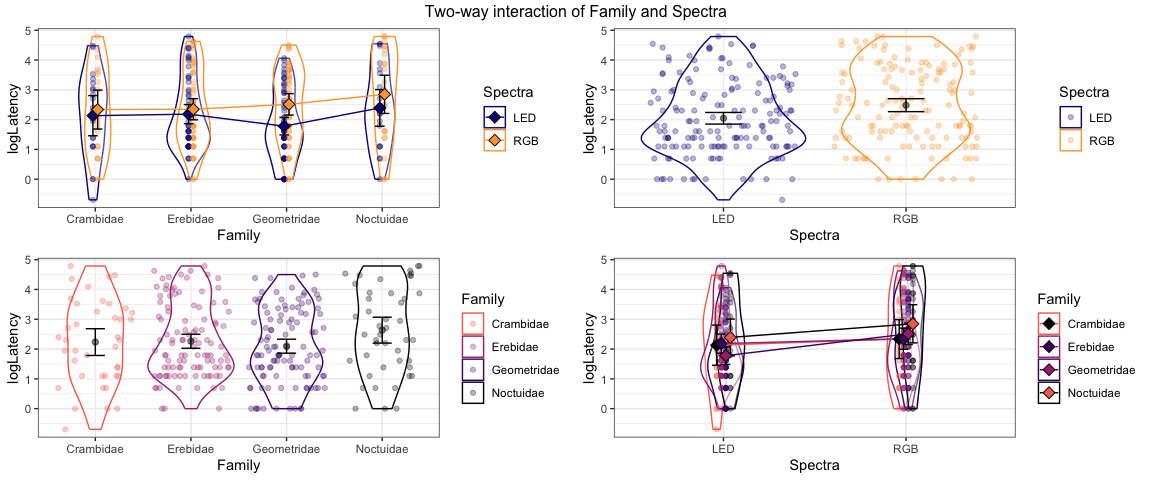
**a) For *Crambidae* and *Erebidae* families, the estimated mean log-latency is similar, but the difference is a little larger for *Noctuidae* and the difference is the largest for the *Geometridae* families.**

**b) The estimated mean log-latency is higher for RGB than for LED in the upper right panel of the plot, but this result should be viewed with caution as there could be an interaction between family and spectra that could make this more complicated.**

**c) The estimated mean log-latency is highest for *Noctuidae* moths that were exposed to RGB spectra.**

1. Because there is a negative number in the strip for the *Crambidae* and LED, there must be a calculation error as latency is a time and so can’t be negative.

ggintplot(response = "logLatency", groupvars = c("Family","Spectra"), data = moth2)



1. Document any collaborations or pertinent discussions with other students and the resources that I am providing and the *Sleuth* that you used to complete this assignment *or report that you did not have any*. If you used generative AI, report which question(s) and how/what you asked for.

**NONE**