Analysis of Longevity in Male Fruit Flies in Relation to Reproductive Effort

Olachi Mbakwe, William Bevidas, Claudia Silverstein

2023-10-31

Introduction and Background

There have always been theories throughout history on the physiological cost of reproduction in terms of reduced lifespan[1]. Although it has not been determined that there is a significant impact on the longevity of human males on reproduction, two scientists, Linda Partidge and Marion Farquhar decided to test the impact on sexual reproduction of male fruit flies. This is because while some species do not need to have a partner to reproduce, fruit flies do, which is why there were chosen for this study.

In order to make any statistical discoveries within this study, we would need to see that if a male fruit fly reproduces, their lives are cut significantly shorter than those that do not reproduce within their life. Although we cannot determine anything for humans through this study, we can make some inference on if sexual reproduction impacts the longevity of a living object and this could potentially begin a study on how it affects a human life time. While there might be ethical dilemmas involved here and other things to figure out for that part of the research, fruit flies are a good way to begin to understand the affect of sexual activity on longevity.

Overall, this experiment is trying to help explore whether or not if the reproduction process for male fruit flies affects their longevity? If there turns out to be a statistically significant impact on the fruit flies longevity, our post hoc question would include the following.

- Does the amount of other flies in the area make a statistical difference?
- Does being in an area with virgin fruit flies have a statistical difference from being in an area with pregnant fruit flies?

Study Design and Methods

In order to address our comparative research question concerning male fruit fly longevity in relation to their reproductive conditions, we've designed a quasi-experimental study. This study retrospectively analyzes data collected from an experiment involving male fruit flies exposed to different reproductive scenarios. To collect our data, scientists randomly divided 125 male fruit flies into five groups of 25. The longevity was quantitatively collected by the number of days the fly was alive. The final data collection is publicly available. Our primary response is the longevity of male fruit flies. This has been operationalized as the total number of days they were alive.

[1] https://www.nature.com/articles/294580a0.pdf

Data Acquisition

The dataset was compiled from previous experimental trials where male fruit flies (Drosophila melanogaster) were exposed to different mating conditions. This data has been systematically recorded and categorized based on the reproductive condition each male was subjected to during the trial period.

Population and Data Collection

Our population of interest is composed of male fruit flies (Drosophila melanogaster) from a controlled laboratory experiment.The dataset was compiled from previous experimental trials where male fruit flies were exposed to various mating conditions. This data has been systematically recorded and categorized based on the reproductive condition each male was subjected to during the trial period. After importing the dataset, we transformed the 'condition' variable into an ordered factor to accurately reflect the range of mating conditions encountered. The conditions range from the absence of females ('Null') to exposure to multiple virgin females ('EightVirgin'), chosen to simulate different degrees of sexual competition and reproductive opportunities. We hypothesize that these variables may significantly influence the longevity of the flies.

Conditions Defined:

- 'Null': Males with no exposure to females, serving as a baseline control group.
- 'OnePreg': Males exposed to one pregnant female per day.
- 'OneVirgin': Males exposed to one virgin female per day.
- 'EightPreg': Males exposed to eight pregnant females per day.
- 'EightVirgin': Males exposed to eight virgin females per day.

Sampling Method

From the dataset, a stratified sampling approach was used to ensure an equal representation of flies from each reproductive condition. This balanced design helps in comparing the longevity across different reproductive scenarios and controlling for potential confounding variables.

Analytical Approach

The Hasse diagram (Figure 1) for the study, indicates a hierarchical structure of the factors, from a single factor of reproduction to the specific conditions and the individual male fruit flies within those conditions. This suggests a nested experimental design and implies that we can explore not only the main effects of reproductive conditions but also potential interactions.

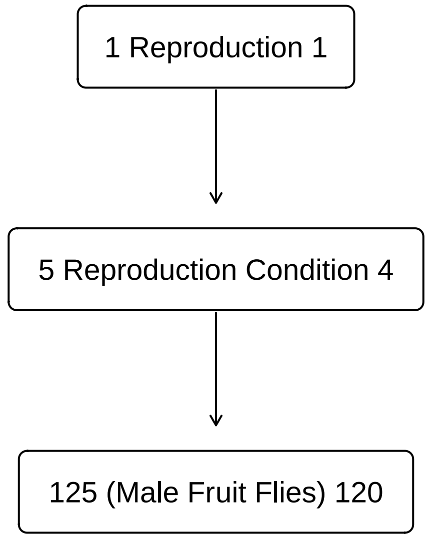


Figure 1: Hasse Diagram for Fruit Fly Longevity

Hypotheses

The null hypothesis for our study is that there is no statistically significant impact of reproductive condition on male fruit fly longevity. The alternative hypothesis is that there is a statistically significant impact. We express these hypotheses as:

$$H_0 : y_{ij} = \mu + \varepsilon_{ij}$$

$$H_A : y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

where y_{ij} represents a fly's longevity, μ the overall mean longevity, α_i the effect of being in reproductive condition i, and ε_{ij} the residual effects.

Type I Error Control

We have decided to set our overall Type I error rate at 6%. For multiple comparisons, we will maintain the Simultaneous Confidence Interval error rate at the same level using Tukey's HSD. Our threshold for declaring results as statistically unusual will be set at 4%.

Exploration of the Data

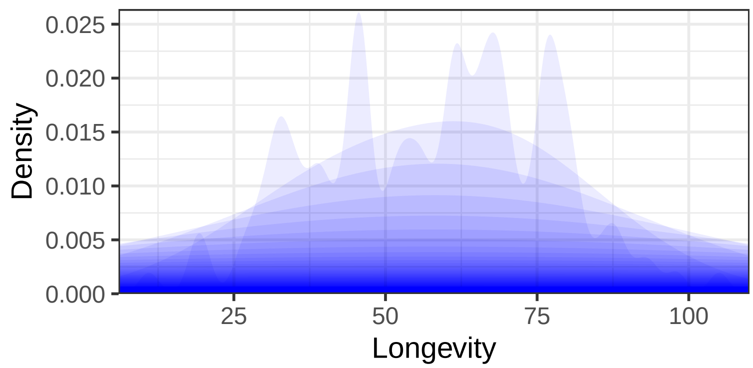


Figure 2: Shadowgram of Fruit Flies Longevity

Figure 2 provides the shadowgram for our 125 male fruit flies longevity . In examining the shadowgram, we can that there is a dominant modal clump in the center (from 25 to 85 points) with separation in the background. While we know that we have five groups based upon condition of the fruit fly, Figure 2 suggests that there are significant differences in terms of performance between them.

Table 1: Summary Statistics for Fruit Flies Longevity

	n	Min	Q1	Median	Q3	Max	MAD	SAM	SASD	Sample Skew	Sample Ex. Kurtosis
Null	25	36	46	61	76	92	22.239	61.96	16.599	0.111	-1.237
OnePreg	25	29	63	70	79	105	13.343	68.60	18.055	-0.433	-0.196
OneVirgin	25	32	58	68	76	98	13.343	65.20	15.901	-0.439	-0.266
EightPreg	25	27	50	60	65	88	11.861	56.92	14.714	-0.175	-0.449
EightVirgin	25	11	31	35	46	63	13.343	36.80	12.285	-0.054	-0.468

Table 1 displays the values of many of the descriptive statistics for each of the treatments used described above. When looking at the five number summary, it is easy to see that having eight virgin fruit flies in the same room results in a much smaller median, *Sample Arithmetic Mean (SAM)*, and generally a smaller range of values. It also looks like the treatment of having one pregnant fruit fly in the room has the highest longevity value according to the *SAM* as well as the highest median. The ranges are quite close together between the last three treatments. According to the box plots in Figure 3, the null treatment seems to have the highest variability and the two treatments with pregnant fruit flies both have outliers while the others do not. There also seems to be a decrease in longevity as the female fruit fly numbers increase and whether or not they were pregnant.

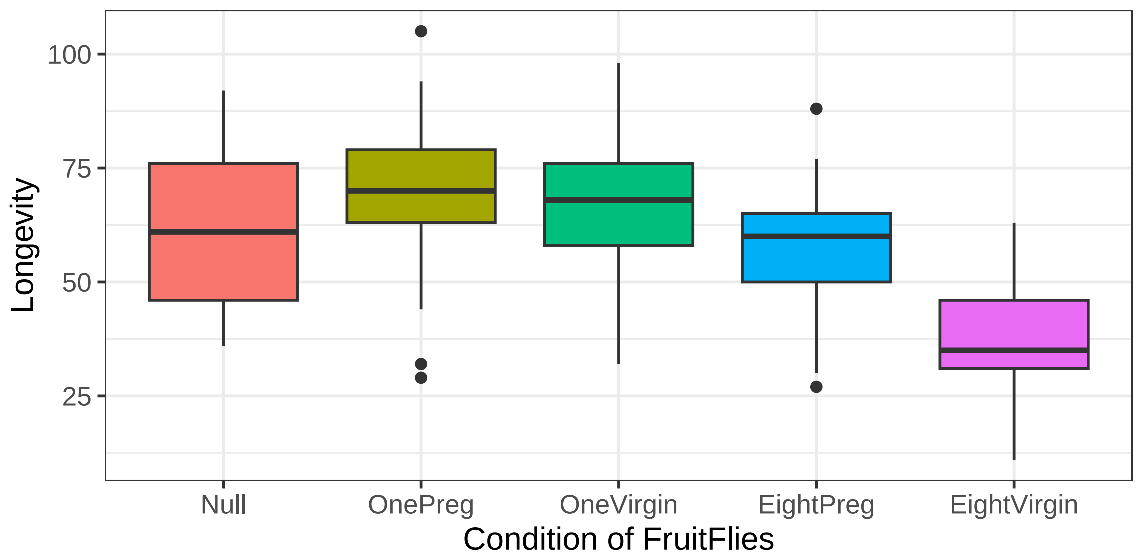


Figure 3: Side-by-side Box Plots of Longevity by Condition

The density plots in Figure 4 help to show how the distribution of each of the treatment values looks. The null condition looks to be spread across the entire range of values from all measurements. This reflects the box plots where the null treatment had the most variability. Having 1 fruit fly that is either pregnant or not shows to be quite close together. Having 8 pregnant fruit flies seems to be pretty close to the values of having 1 fly, but having 8 virgin fruit flies looks to have much smaller values than the rest.

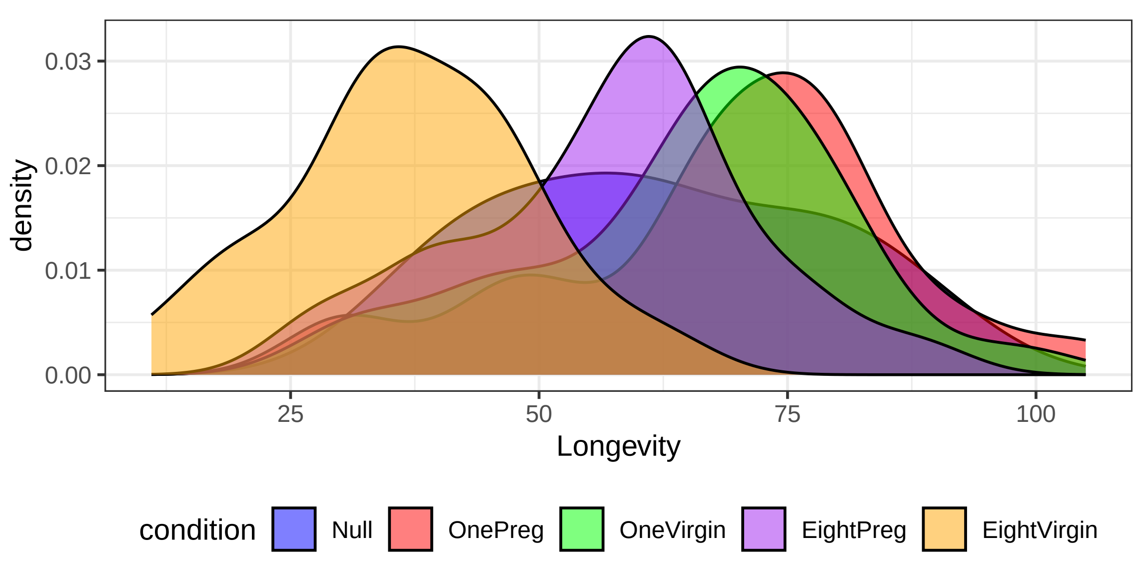


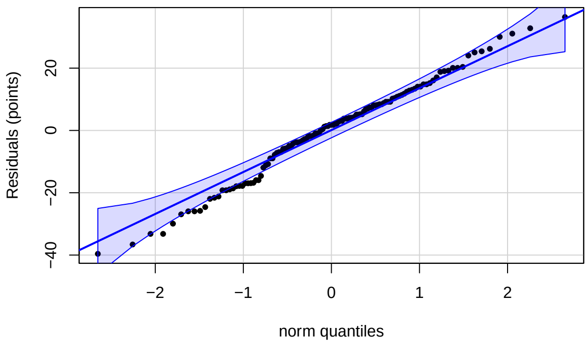
Figure 4: Densities Plots of Longevity by Condition

Results

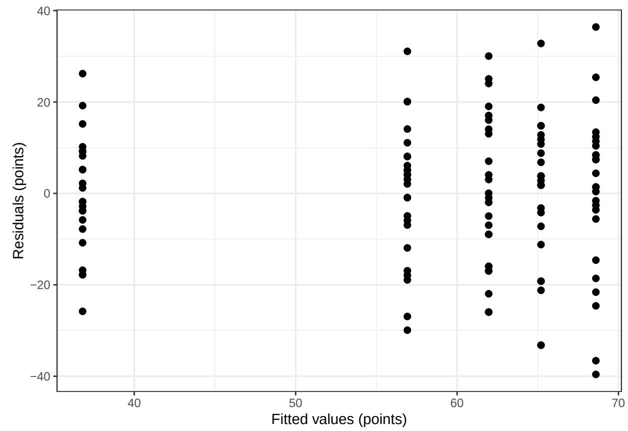
To answer our main research question, we will seek to use the parametric shortcut known as the ANOVA F test. There are three assumptions that our data must satisfy to use this approach: residuals follow a

Gaussian distribution, homoscedasticity, and independence of observations.

Assumptions



(a) QQ Plot



(b) Strip Chart

Figure 5: Assessing Assumptions for Song Knowledge Study

Firstly, looking at the Gaussian assumption of Figure 5a, which shows the QQ plot for our residuals within a 90% confidence envelope, we can see that a small number of observations fall outside this envelope. However, given that we have 125 observations, these few outliers that fall outside the 90% confidence envelope are not a significant concern. Additionally, from Table **??**, all conditions except for the null condition exhibit some slight negative skewness, while the null condition appears to exhibit zero skewness. Further, all five groups have similar levels of excess kurtosis, suggesting fewer potential outliers in the data than expected for a Gaussian distribution. The consistent skewness and kurtosis issues across groups reinforce the Gaussian assumption.

In Figure 5b, the strip chart for assessing the homoscedasticity assumption presents five strips, each with 25 points. Looking at the strips we see that the length increases from left to right along the horizontal axis which suggests a very slight megaphone shape. However, since no group's length is more than twice that of another, concerns about homoscedasticity are minimal. Referring back to Table **??**, the EightPreg group had the smallest values for variation, as measured by *SASD* and *MAD*, though the differences are not substantial. The visual and statistical evidence allows us to proceed without concern for homoscedasticity.

For the issue of independence of observations, the study design by Partridge and Farquhar guarantees this as each male fruit fly is isolated in its own vial, preventing interaction, and each is randomly assigned a specific reproductive condition, effectively controlling for and isolating variables that affect longevity. Thus, the design and methodology of the study ensure the independence of observations.

Given that we have a balanced design, we will cautiously proceed with the parametric ANOVA F test and our planned post hoc analysis (as needed).

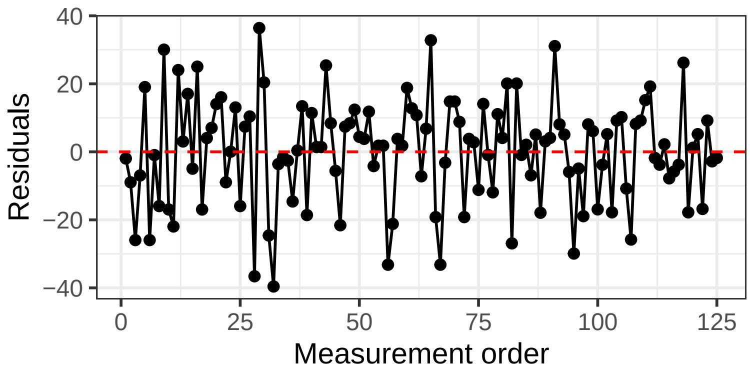


Figure 6: Index plot of Fruit Flies Residual

```
## [1] 2.161944
```

The index plot for the Fruit Flies data does not exhibit any indication of a patterns. Additionally, the Durbin-Watson statistic of 2.16 falls within the acceptable range, indicating no evidence of autocorrelation among observations and supporting the assumption of independent observations for our data.

Omnibus Results

Table 2: Modern ANOVA Table for Resin Lifetimes Study

Source	SS	df	MS	F	p-value	Eta Sq.	Omega Sq.	Epsilon Sq.
condition	15760.85	4	3940.2120	16.1248	< 0.0001	0.3496	0.3261	0.3279
Residuals	29322.80	120	244.3567					

Note. Computer rounding has made the p-value look like zero.

As we can see from Table 2,the condition a fruit fly is put under accounts for approximately 16.1248 times as much variation as the residuals. Under the null model that the reproduction condition does not impact the longevity of fruit flies, we would anticipate observing a value for the F ratio at least as large as we did ~ 0.01% of the time we repeat the study. Since our p-value is less than our unusualness threshold (0.0001 $<$ 0.04), we will reject the null hypothesis and decide to act as if the reproduction condition for fruit flies does impact the longevity of fruit flies. When looking at the values for η^2, ω^2, and ϵ^2, we can see that the effect sizes are significantly large since Cohen and Field's suggested rules of thumb consider a large effect size to be above 0.26 and 0.14 for their respective effect sizes. This compliments the results of the hypothesis test because having a large effect size means that there would have been a larger effect, leading to potential differences throuhgout the model.

Table 3: Point Estimates from the Fruit Flies Study

	Estimate
Grand Mean	57.90
Null	4.06
OnePreg	10.70
OneVirgin	7.30
EightPreg	-0.98
EightVirgin	-21.10

In general, a group of fruit flies will live, on average, approximately r round(dummy.coef(fruitflies_Model)[[1]], 2) days; this is our estimate for baseline longevity ($\widehat{\mu_{\bullet\bullet}}$).. We can also see the factor level (treatment) effects ($\widehat{\alpha_i}$) estimates. For males exposed to eight virgin females per day exhibited a reduction in longevity by NAdays, whereas males exposed to one virgin female per day showed a decrease of 7.3 days. Males exposed to one pregnant female per day experienced a reduction of -0.98 days. In contrast, males with no female exposure, serving as our baseline control, had a point estimate effect of 10.7days, and males exposed to eight pregnant females per day had a reduced longevity of -21.1 days. This analysis indicates that males in the Eight Virgin and Eight Preg conditions experienced a greater reduction in longevity compared to the baseline ($GSAM$).

Post Hoc

Given our decision to reject the null hypothesis, it is appropriate to proceed with post hoc analyses.

For post hoc analyses, we will examine pairwise comparisons among the five conditions experienced by the fruit flies: null, one pregnant female (onepreg), one virgin female (onevirgin), eight pregnant females (eightpreg), and eight virgin females (eightvirgin).

Table 4: Post Hoc Tukey HSD Comparisons

	Difference	Lower Bound	Upper Bound	Adj. p-Value
OnePreg-Null	6.64	-5.606	18.886	0.563
OneVirgin-Null	3.24	-9.006	15.486	0.948
EightPreg-Null	-5.04	-17.286	7.206	0.785
EightVirgin-Null	-25.16	-37.406	-12.914	0.000
OneVirgin-OnePreg	-3.40	-15.646	8.846	0.939
EightPreg-OnePreg	-11.68	-23.926	0.566	0.069
EightVirgin-OnePreg	-31.80	-44.046	-19.554	0.000
EightPreg-OneVirgin	-8.28	-20.526	3.966	0.338
EightVirgin-OneVirgin	-28.40	-40.646	-16.154	0.000
EightVirgin-EightPreg	-20.12	-32.366	-7.874	0.000

From Table 4, the Post Hoc Tukey HSD Comparisons show that the differences in longevity between OnePreg and Null, OneVirgin and Null, and EightPreg and Null fruit flies occur with adjusted p-values of 56.3%, 94.8%, and 78.5% respectively under the null hypothesis, suggesting that these differences are not statistically significant and can be considered typical variations when no real effect is present. In contrast, the comparison between EightVirgin and Null is highly significant with an adjusted p-value of 0.000, indicating a statistically significant difference in longevity.

Further comparisons show that the differences between OneVirgin and OnePreg, EightPreg and OnePreg, and EightPreg and OneVirgin do not reach statistical significance, bearing adjusted p-values of 93.9%, 6.9%, and 33.8%, respectively, all exceeding the alpha threshold conventionally set at 0.05. However, the comparisons involving the EightVirgin group—specifically when compared against OnePreg, OneVirgin, and EightPreg—each demonstrate statistically significant discrepancies with an adjusted p-value of 0.000. This consistency in significance suggests a substantial effect of the EightVirgin condition on fruit fly longevity. This consistency in significance suggests a substantial effect of the EightVirgin condition on fruit fly longevity. The final comparison drawn between the EightVirgin and EightPreg groups solidifies this finding, with a significant divergence and an adjusted p-value of 0.000.

These results indicate that while certain group differences could be due to chance, the consistently significant outcomes associated with the EightVirgin condition across several comparisons point to a meaningful influence of this particular treatment on the lifespan of the fruit flies.

Table 5: Post Hoc Comparison Effect Sizes

Pairwise Comparison	Cohen's d	Hedge's g	Prob. Superiority
Null vs. OnePreg	-0.383	-0.377	0.393
Null vs. OneVirgin	-0.199	-0.196	0.444
Null vs. EightPreg	0.321	0.316	0.590
Null vs. EightVirgin	1.723	1.696	0.888
OnePreg vs. OneVirgin	0.200	0.197	0.556
OnePreg vs. EightPreg	0.709	0.698	0.692
OnePreg vs. EightVirgin	2.059	2.027	0.927
OneVirgin vs. EightPreg	0.541	0.532	0.649
OneVirgin vs. EightVirgin	1.999	1.967	0.921
EightPreg vs. EightVirgin	1.484	1.461	0.853

Discussions and Limitations

In examining the impact of reproductive condition on male fruit fly longevity, our study has provided valuable insights. From our data, we found that exposure to eight virgin females per day has a marked negative impact on male longevity. However, there are several limitations within our study that must be acknowledged. Firstly, there is the specificity of our study because we examined a single species of fruit fly in a controlled laboratory setting, originally collected in Dahomey in 1970. Secondly, our sample size may not have been large enough to detect subtler effects or to ensure that our findings are representative of the broader population of fruit flies. A larger sample would enhance the statistical power of the study and could potentially reveal effects that our current research was unable to observe. Additionally, the incorporation of more varied attributes could provide us with further insight.

References and Materials Consulted

Partridge, L., & Farquhar, M. (1981, December 10). Sexual activity reduces lifespan of male fruitflies. Nature. Retrieved from https://www.nature.com/articles/294580a0.pdf.

Author Contributions

The authors of this report would like to acknowledge their individual contributions to the report. Both authors contributed to ongoing discussions about study design and analysis.

- Olachi Mbakwe contributed to the Study Design and Methods, Exploration of the Data, analysis of data(Assumptions), coding, and writing of the report.
- William Bevidas contributed to the Introduction and Background, Exploration of the Data, analysis of data(Omnibus Results), coding, and writing of the report.
- Claudia Silverstein contributed to the Discussions and Limitations, analysis of data (Post Hoc), coding, and writing of the report.

Code Appendix

```
knitr::opts_chunk$set(
    echo = FALSE,
    fig.align = "center",
    message = FALSE,
    warning = FALSE,
    dpi = 300 # helps create higher quality graphics in Word
)


# Add additional packages by name to the following list ----
packages <- c(
  "tidyverse", "knitr", "kableExtra", "hasseDiagram",
  "psych", "car", "parameters"
  )
lapply(X = packages, FUN = library, character.only = TRUE, quietly = TRUE)

# Loading Helper Files and Setting Global Options ----
options(knitr.kable.NA = "")
options("contrasts" = c("contr.sum", "contr.poly"))
source("https://raw.github.com/neilhatfield/STAT461/master/rScripts/ANOVATools.R")

source("https://raw.github.com/neilhatfield/STAT461/master/rScripts/shadowgram.R")


fruitflies <- read_csv("fruitflies.csv")

### Set year to an ordered factor

fruitflies$condition <- factor(
x = fruitflies$condition,
levels = c("Null", "OnePreg", "OneVirgin","EightPreg","EightVirgin")
)

modelLabels <- c("1 Reproduction 1", "5 Reproduction Condition 4", "125 (Male Fruit Flies) 120")

modelMatrix <- matrix(
  data = c(FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, TRUE, TRUE, FALSE),
  nrow = 3,
  ncol = 3,
  byrow = FALSE
)

hasseDiagram::hasse(
 data = modelMatrix,
 labels = modelLabels
)
# Creating a shadowgram of scores ----
# Note: you do not have to use shadowgrams.
# You can use a histogram or any other kind of data visualization.
shadowgram(
  dataVec = fruitflies$longevity,
```

```
  label = "Longevity",
  layers = 50,
  color = "blue",
  aStep = 4
)


# Descriptive statistics on longevity by condition ----
fruitfliesStats <- psych::describeBy(
x = fruitflies$longevity,
group = fruitflies$condition,
na.rm = TRUE,
skew = TRUE,
ranges = TRUE,
quant = c(0.25, 0.75),
IQR = FALSE,
mat = TRUE
)

fruitfliesStats %>%
tibble::remove_rownames() %>%
tibble::column_to_rownames(
var = "group1"
) %>%
dplyr::select(
n, min, Q0.25, median, Q0.75, max, mad, mean, sd, skew, kurtosis
) %>%
knitr::kable(
caption = "Summary Statistics for Fruit Flies Longevity",
digits = 3,
format.args = list(big.mark = ","),
align = rep('c', 11),
col.names = c("n", "Min", "Q1", "Median", "Q3", "Max", "MAD", "SAM", "SASD",
"Sample Skew", "Sample Ex. Kurtosis"),
booktabs = TRUE
) %>%
kableExtra::kable_classic_2(
font_size = 12,
latex_options = c("scale_down", "HOLD_position")
)
ggplot(
  data = fruitflies,
  mapping = aes(x = condition, y = longevity, fill = condition)
) +
  geom_boxplot() +
  theme_bw() +
  xlab("Condition of FruitFlies") +
  ylab("Longevity") +
  theme(
    legend.position = "none",
    text = element_text(size = 12)
  )
```

```r
fruitflies %>%
  drop_na(condition) %>%
  ggplot(mapping = aes(x = longevity, fill = condition)) +
  geom_density(na.rm = TRUE, alpha = 0.5) +
  theme_bw() +
  scale_fill_manual(values = c("blue", "red", "green", "purple", "orange")) + # Make sure to adjust the
  xlab("Longevity") +
  theme(legend.position = "bottom")
# Fit the model and parametric shortcut ----
fruitflies_Model <- aov(
formula = longevity ~ condition,
data = fruitflies,
na.action = "na.omit"
)

# Assumption Assessment Visualizations ----
## Gaussian Residuals Assumption
car::qqPlot(
  x = fruitflies_Model$residuals,
  distribution = "norm",
  envelope = 0.90,
  id = FALSE,
  pch = 20,
  ylab = "Residuals (points)"
)

## Strip Chart for Homoscedasticity ----
ggplot(
  data = data.frame(
    residuals = fruitflies_Model$residuals,
    fitted = fruitflies_Model$fitted.values
  ),
  mapping = aes(x = fitted, y = residuals)
) +
  geom_point(size = 2) +
  theme_bw() +
  xlab("Fitted values (points)") +
  ylab("Residuals (points)")

ggplot(
data = data.frame(
residuals = fruitflies_Model$residuals,
index = 1:length(fruitflies_Model$residuals)
),
mapping = aes(x = index, y = residuals)
) +
geom_point(size = 1.5) +
geom_line() +
theme_bw() +
geom_hline(
yintercept = 0,
linetype = "dashed",
color = "red"
```

```
) +
xlab("Measurement order") +
ylab("Residuals")


car::durbinWatsonTest(fruitflies_Model)$dw


#| fig.cap = "Modern ANOVA Table for Fruit Flies Longevity Studyl",
#| #| fig.height = 2,
#| fig.width = 4,
#| fig.pos = "H"

# Modern ANOVA Table ----


parameters::model_parameters(
model = fruitflies_Model,
effectsize_type = c("eta", "omega", "epsilon")
) %>%
  dplyr::mutate(
p = ifelse(
test = is.na(p),
yes = NA,
no = pvalRound(p, digits = 4)
)
) %>%
knitr::kable(
digits = 4,
col.names = c(
"Source", "SS", "df", "MS", "F", "p-value",
"Eta Sq.", "Omega Sq.", "Epsilon Sq."),
caption = "Modern ANOVA Table for Resin Lifetimes Study",
booktabs = TRUE,
align = c("l", rep("c", 8))
) %>%
kableExtra::kable_classic(
font_size = 10,
latex_options = c("HOLD_position")) %>%
  kableExtra::footnote(
general = "Computer rounding has made the p-value look like zero.",
general_title = "Note. ",
footnote_as_chunk = TRUE
)

# Point Estimates for Parametric Shortcut ----

pointEst <- dummy.coef(fruitflies_Model)
pointEst <- unlist(pointEst)
names(pointEst) <- c("Grand Mean","Null", "OnePreg", "OneVirgin","EightPreg","EightVirgin")
data.frame("Estimate" = pointEst) %>%
knitr::kable(
digits = 2,
```

```
caption = "Point Estimates from the Fruit Flies Study",
format = "latex",
booktabs = TRUE,
align = "c"
) %>%
kableExtra::kable_classic(
font_size = 12,
latex_options = c("HOLD_position")
)


# Post Hoc via Tukey HSD ----
hsdfruitflies <- TukeyHSD(
x = fruitflies_Model, # Your aov/lm object
conf.level = 0.95 # 1 -- Your overall Type I Error level
)


## Kable Code for Tukey HSD
knitr::kable(
x = hsdfruitflies$condition, # Notice the factor's name
digits = 3,
caption = "Post Hoc Tukey HSD Comparisons",
col.names = c("Difference", "Lower Bound",
"Upper Bound", "Adj. p-Value"),
align = 'cccc',
booktabs = TRUE,
) %>%
kableExtra::kable_classic_2("striped",font_size = 12,full_width = FALSE,latex_options = "HOLD_position")

# Post Hoc Effect Sizes ----
anova.PostHoc(fruitflies_Model) %>%
  knitr::kable(
    digits = 3,
    caption = "Post Hoc Comparison Effect Sizes",
    col.names = c("Pairwise Comparison","Cohen's d", "Hedge's g",
                  "Prob. Superiority"),
    align = 'lccc',
    booktabs = TRUE
  ) %>%
kableExtra::kable_classic_2(
  "striped",
  font_size = 12,
  full_width = FALSE,
  latex_options = "HOLD_position"
  )
```