

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 lifespan¹. 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 they 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.

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.

¹<https://www.nature.com/articles/294580a0.pdf>

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.

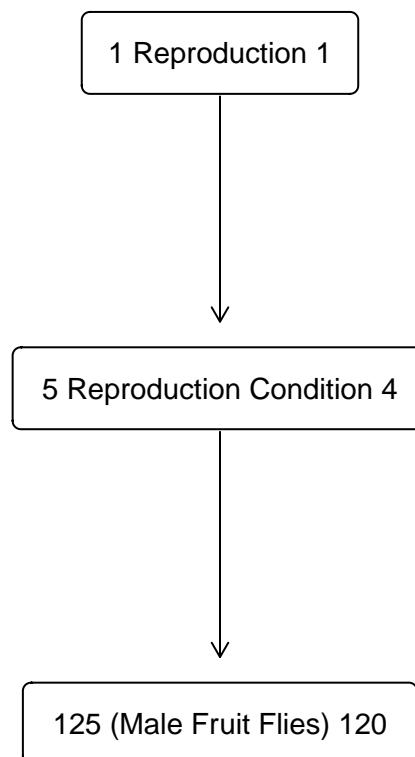
Conditions Defined: - 'Null': Males with no exposure to females, serving as a baseline control group. - 'OnePreg': Males exposed to one pregnant female. - 'OneVirgin': Males exposed to one virgin female. - 'EightPreg': Males exposed to eight pregnant females. - 'EightVirgin': Males exposed to eight virgin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

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 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.



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

y_{ij} represents a fly's longevity, μ the overall mean

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

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 boxplots in Figure 1, the null treatment seems to have the most variance 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.

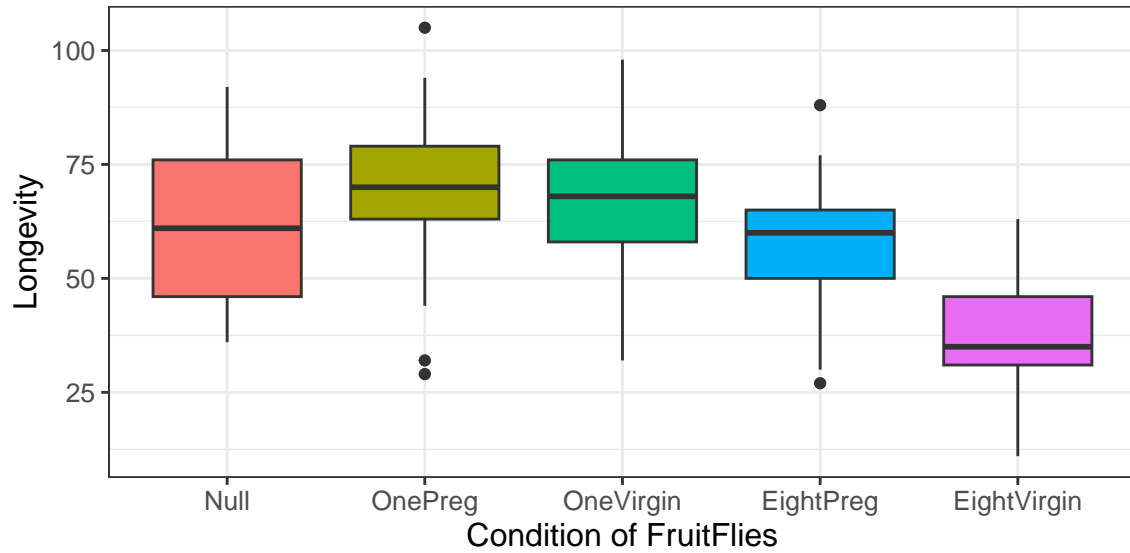


Figure 1: Side-by-side Box Plots of Longevity by Condition

The density plots in Figure 2 helps 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. 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.

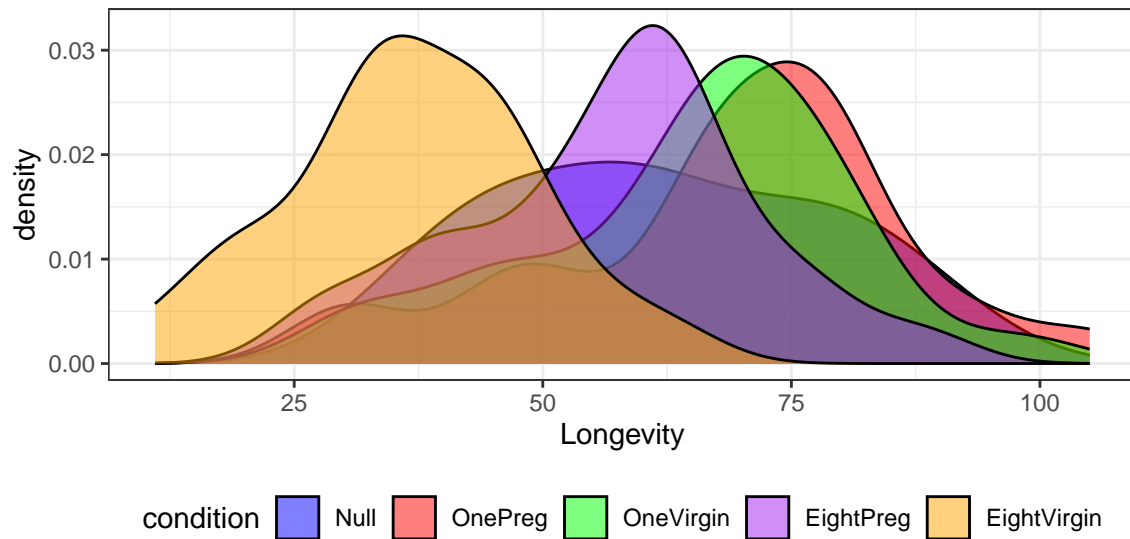


Figure 2: Densities Plots of Longevity by Condition

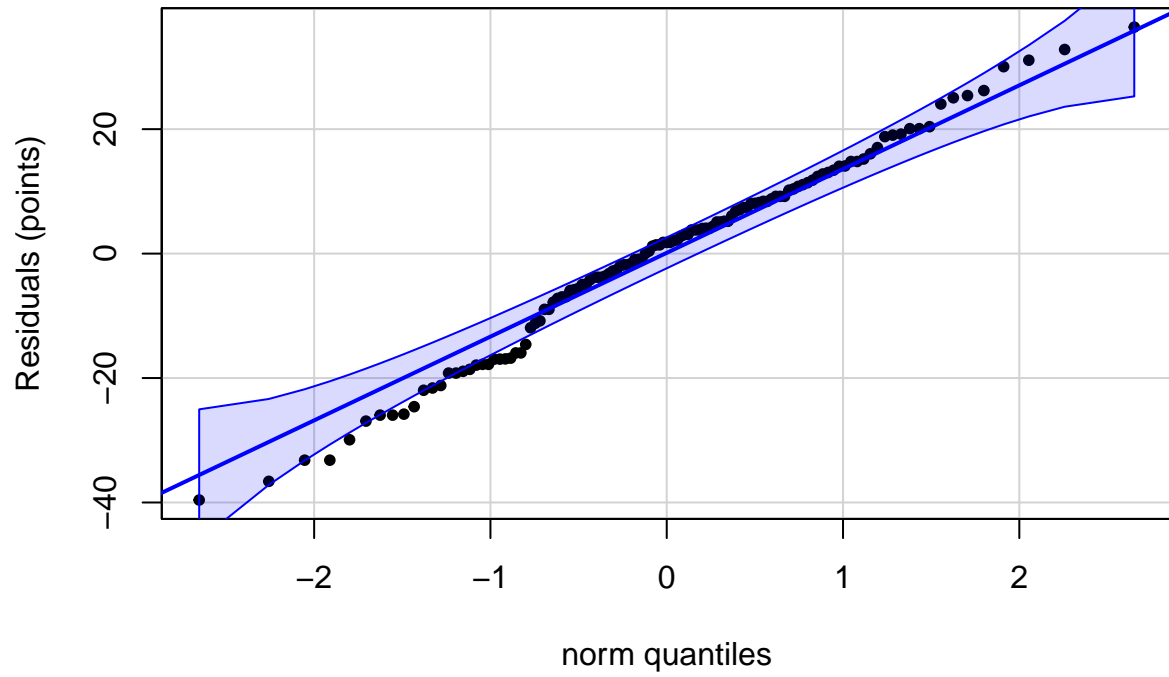
Results

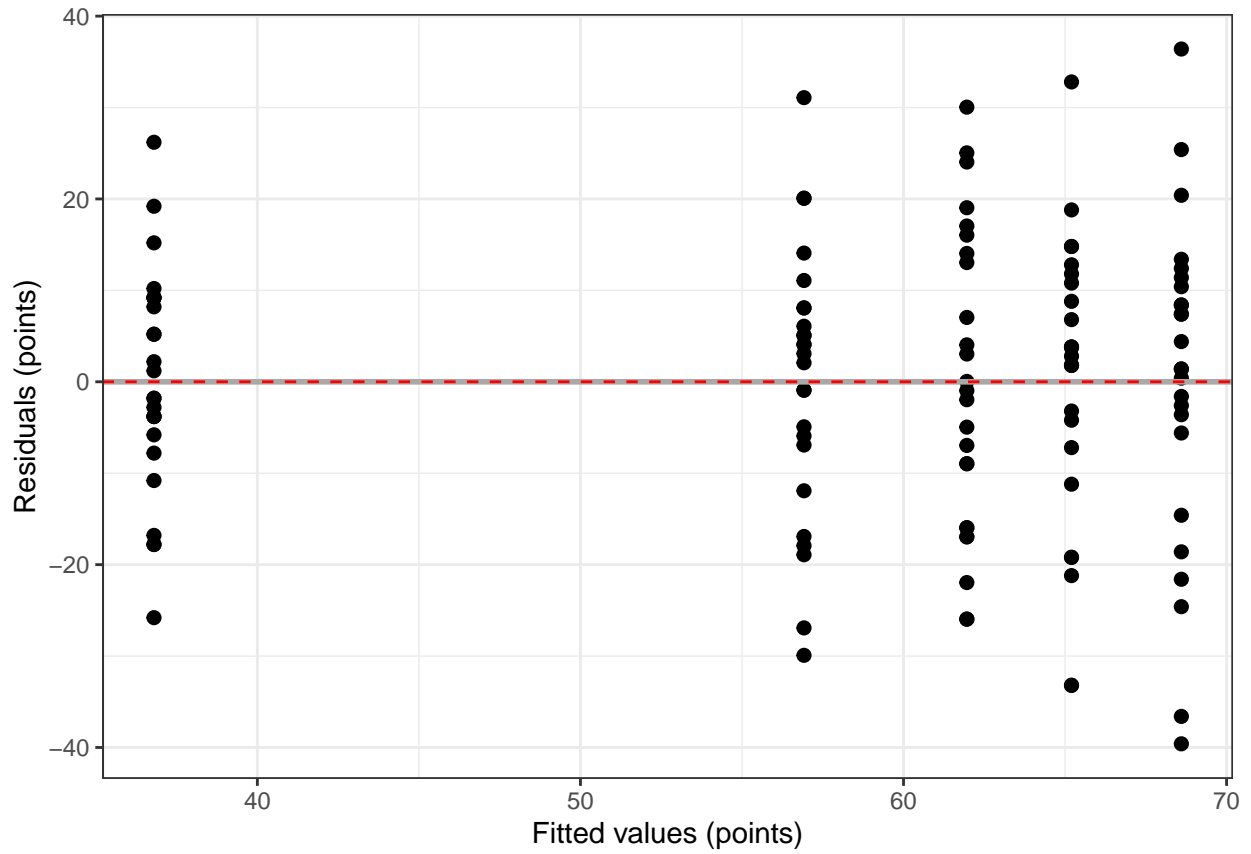
To answer our main research question, we will seek to use the parametric shortcut known as the ANOVA F test. There are three assumptions that our data must satisfy to use this approach: residuals follow a Gaussian distribution, homoscedasticity, and independence of observations.

Assumptions

model

Gaussian Residuals





Post Hoc for the Parametric Shortcut

Table 2: ANOVA Table for Fruit Flies Study

Source	SS	df	MS	F	p-value	Eta Sq.	Omega Sq.	Epsilon Sq.
condition	15760.85	4	3940.2120	16.1248	0	0.3496	0.3261	0.3279
Residuals	29322.80	120	244.3567					

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

The emmeans Approach

Table 4: Pairwise Post Hoc Comparison via Tukey HSD

Pair	Difference	SE	DF	t	p-value
Null - OnePreg	-6.64	4.421	120	-1.502	0.563
Null - OneVirgin	-3.24	4.421	120	-0.733	0.948
Null - EightPreg	5.04	4.421	120	1.140	0.785
Null - EightVirgin	25.16	4.421	120	5.691	0.000
OnePreg - OneVirgin	3.40	4.421	120	0.769	0.939
OnePreg - EightPreg	11.68	4.421	120	2.642	0.069
OnePreg - EightVirgin	31.80	4.421	120	7.192	0.000
OneVirgin - EightPreg	8.28	4.421	120	1.873	0.338
OneVirgin - EightVirgin	28.40	4.421	120	6.423	0.000
EightPreg - EightVirgin	20.12	4.421	120	4.551	0.000

Tukey's [& Kramer's] Honest Significant Difference

Table 5: 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

DescTools Method

Table 6: Post Hoc Newman-Keuls Comparisons

	Difference	Lower Bound	Upper Bound	Adj. p-Value
OnePreg-Null	6.64	-2.522	15.802	0.294
OneVirgin-Null	3.24	-4.089	10.569	0.465
EightPreg-Null	-5.04	-12.369	2.289	0.257
EightVirgin-Null	-25.16	-34.322	-15.998	0.000
OneVirgin-OnePreg	-3.40	-10.729	3.929	0.443
EightPreg-OnePreg	-11.68	-21.922	-1.438	0.046
EightVirgin-OnePreg	-31.80	-42.804	-20.796	0.000
EightPreg-OneVirgin	-8.28	-17.442	0.882	0.151
EightVirgin-OneVirgin	-28.40	-38.642	-18.158	0.000
EightVirgin-EightPreg	-20.12	-27.449	-12.791	0.000

Special Comparisons (Dunnett's)

Table 7: Post Hoc Comparisons–Dunnett's Test

	Difference	Lower Bound	Upper Bound	Adj. p-Value
OnePreg-Null	6.64	-3.014	16.294	0.371
OneVirgin-Null	3.24	-6.414	12.894	0.876
EightPreg-Null	-5.04	-14.694	4.614	0.612
EightVirgin-Null	-25.16	-34.814	-15.506	0.000

Post Hoc Effect Sizes

Table 8: 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

Connecting Letter Report

```
## $condition
##      OnePreg  OneVirgin      Null  EightPreg EightVirgin
##          "a"      "ab"      "ab"      "b"      "c"
```

Author Contributions

Table 9: Author Contributions and Responsibilities

Name	Responsibility
Olachi Mbakwe	Introduction
William Bevidas	Data Exploration
Claudia Silverstein	Results

Code Appendix

```
knitr::opts_chunk$set(  
  echo = FALSE,  
  fig.align = "center",  
  message = FALSE,  
  warning = FALSE,  
  dpi = 300 # helps create higher quality graphics in Word  
)  
  
# Load packages  
packages <- c("tidyverse", "knitr", "dplyr", "kableExtra", "hasseDiagram", "psych", "car", "parameters",  
  
lapply(  
  X = packages,  
  FUN = library,  
  character.only = TRUE  
)  
  
# Set options  
options(knitr.kable.NA = "")  
options(contrasts = c("contr.sum", "contr.poly"))  
  
# Load additional tools  
source("https://raw.githubusercontent.com/neilhatfield/STAT461/master/rScripts/ANOVATools.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
```

```

)
# Descriptive statistics on score by year in school ----
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)
  )

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) +
  geom_smooth(
    formula = y ~ x,
    method = stats::loess,
    method.args = list(degree = 1),
    se = FALSE,
    linewidth = 0.5
  ) +
  geom_hline(
    yintercept = mean(residuals(fruitflies_Model)),
    color = "darkgrey",
    linetype = "solid",
    linewidth = 1
  ) +
  geom_hline(
    yintercept = 0,
    color = "red",
    linetype = "dashed"
  ) +
  theme_bw() +
  labs(
    x = "Fitted values (points)",

```

```

    y = "Residuals (points)"
  )

# Modern ANOVA Table ----

parameters::model_parameters(
  model = fruitflies_Model,
  effectsize_type = c("eta", "omega", "epsilon")
) %>%
knitr::kable(
  digits = 4,
  col.names = c(
    "Source", "SS", "df", "MS", "F", "p-value",
    "Eta Sq.", "Omega Sq.", "Epsilon Sq."),
  caption = "ANOVA Table for Fruit Flies Study",
  booktabs = TRUE,
  align = c("l", rep("c", 8))
) %>%
kableExtra::kable_classic(
  font_size = 10,
  latex_options = c("HOLD_position")
)

# 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")
)

# Demo Code for emmeans Post Hoc ----
fruitfliesHPairs <- emmeans::emmeans(
  object = fruitflies_Model, # Your aov/lm object
  specs = pairwise ~ condition, # Creates all pairs of the levels of the factor listed
  adjust = "tukey", # How you want to control the error rate
  level = 0.9 # 1 - Your overall Type I Error Rate
)

```

```

## Make a professional looking table
knitr::kable(
x = fruitfliesHPairs$contrasts, # Grab the appropriate sub-object
digits = 3,
caption = "Pairwise Post Hoc Comparison via Tukey HSD",
col.names = c("Pair", "Difference", "SE", "DF", "t", "p-value"),
align = "lcccc",
booktabs = TRUE
) %>%
kableExtra::kable_classic_2("striped",font_size = 12,full_width = FALSE,latex_options = "HOLD_position")

### same code as the on above different ordering
# 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")

# Demo Code for DescTools Pairwise Post Hoc ----
dtPH <- DescTools::PostHocTest(
x = fruitflies_Model, # Your aov/lm object
method = "newmankeuls", # Your chosen method
conf.level = 0.9 # 1 -- Your Overall Type I Error Rate
)

## Kable Code for DescTools
knitr::kable(
x = dtPH$condition, # Notice the use of the factor name
digits = 3,
caption = paste( # Creates a nice title; copy at will
"Post Hoc",
attr(dtPH, "method"),
"Comparisons"
),
col.names = c("Difference", "Lower Bound",
"Upper Bound", "Adj. p-Value"),
align = 'lcccc',
booktabs = TRUE,
) %>%

```

```

kableExtra::kable_classic_2("striped",font_size = 12,full_width = FALSE,latex_options = "HOLD_position")

## Dunnett's Test via DescTools
dunnett <- DescTools::DunnettTest(
formula = longevity ~ condition,
data = fruitflies,
control = "Null", # Enter the level that you want to compare to
conf.level = 0.9 # 1 -- Your Overall Type I Error Rate
)
## Kable Code for Dunnett's Test
knitr::kable(
x = dunnett$`Null`, # Note the use of special treatment level
digits = 3,
caption = paste("Post Hoc Comparisons--Dunnett's Test"),
col.names = c("Difference", "Lower Bound",
"Upper Bound", "Adj. p-Value"),
align = 'lcccc',
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")

multcompView::multcompLetters4(
object = fruitflies_Model, # Your aov/lm object
comp = hsdfruitflies, # Your stored comparisons
threshold = 0.1 # Your Overall Type I Error Rate
)

# Create a data frame with the partner names and years
Contributions <- data.frame(
  Name = c(
    "Olachi Mbakwe",
    " William Bevidas",
    "Claudia Silverstein"
  ),
  Responsibilites = c(

```

```

    "Introduction",
    "Data Exploration",
    "Results"
  )
)

Contributions <- kable(Contributions,
  col.names = c("Name", "Responsibility"),
  caption = "Author Contributions and Responsibilities" %>%
  kable_classic_2("striped",font_size = 12,full_width = FALSE,latex_options = "HOLD_position")
Contributions

```