Male Fly Longevity Study

Kevin Mouck, Zachary Kemp, Nate Smith

March 31, 2023

Introduction and Background

The cost of reproduction on the longevity of organisms has been a long studied question. The reduced lifespan of females as a consequence of increased reproduction has been well established, but this same pattern is not as clearly seen for males who only have the sole responsibility of contributing a gamete to the offspring (Partridge & Farquhar, 1981).

Using Partridge and Farquhar's methodology, we want to explore if a males living condition with females flies (no females, one virgin female, eight virgin females, one pregnant female, eight pregnant females) affects the longevity of male flies. If the male's living condition does statistically impact their longevity, we can create post hoc questions.

These questions can include whether a the longevity of a male fly is statistically different from being with virgin or pregnant flies, or which male living condition is statistically different from others.

Study Design

An experiment was created using a population of flies collected in Dahomey. 125 of these male flies were randomly divided into five groups of 25 each. The first group acted as a control where the male flies were kept with no female flies. The next group of 25 male flies were kept with one virgin female fly per day, and the next 25 male flies were kept with eight female virgin flies per day. The last two groups of 25 were similar except that these two groups were either kept with one pregnant female fly per day or eight pregnant female flies per day.

The primary response in this experiment is the longevity of male flies measured in the number of days they remained alive. The factors of the study are the five different living conditions that male flies were placed in.

Figure 1 shows a Hasse diagram for the study where a One-way ANOVA model is applicable with sufficient degrees of freedom so we can estimate our effects and residuals.

Our null hypothesis is that there is no statistically significant affect of male flies living condition on their longevity.

Our alternative hypothesis is that the male flies living condition has a statistically significant affect on their longevity.

Our overall Type I risk will be 1% and we will control the Simultaneous Confidence Interval error rate using Tukey's HSD. Our unusualness threshold will also be set at 1%.

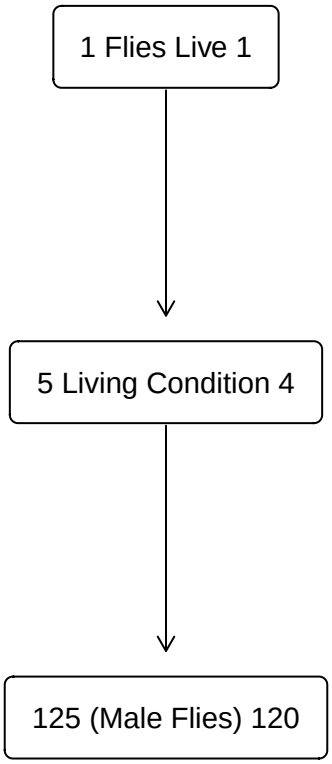

Figure 1: Hasse Diagram for the Fly Longevity Study

Data Exploration

Table 1: Summary Statistics for Male Fly Longevity

	n	Min	Q1	Median	Q3	Max	MAD	SAM	SASD	Sample Skew	Sample Ex. Kurtosis
Null	25	33	58	66	73	100	11.861	65.60	14.393	-0.010	0.042
OnePreg	25	21	49	60	76	106	20.756	60.16	20.155	0.220	-0.368
OneVirgin	25	35	51	59	68	104	13.343	61.24	16.179	0.833	0.370
EightPreg	25	39	47	58	69	88	16.309	59.24	14.354	0.408	-0.978
EightVirgin	25	18	33	44	53	72	16.309	42.56	13.866	0.073	-0.743

Table 1 shows values of the descriptive statistics grouped by each living condition of the male flies. Males flies living with eight virgin females flies have longevities shorter than the other groups. The second shortest lifespan seems to be the male flies with eight pregnant females flies, but these values fall closer to the other groups than to the eight virgin females group.

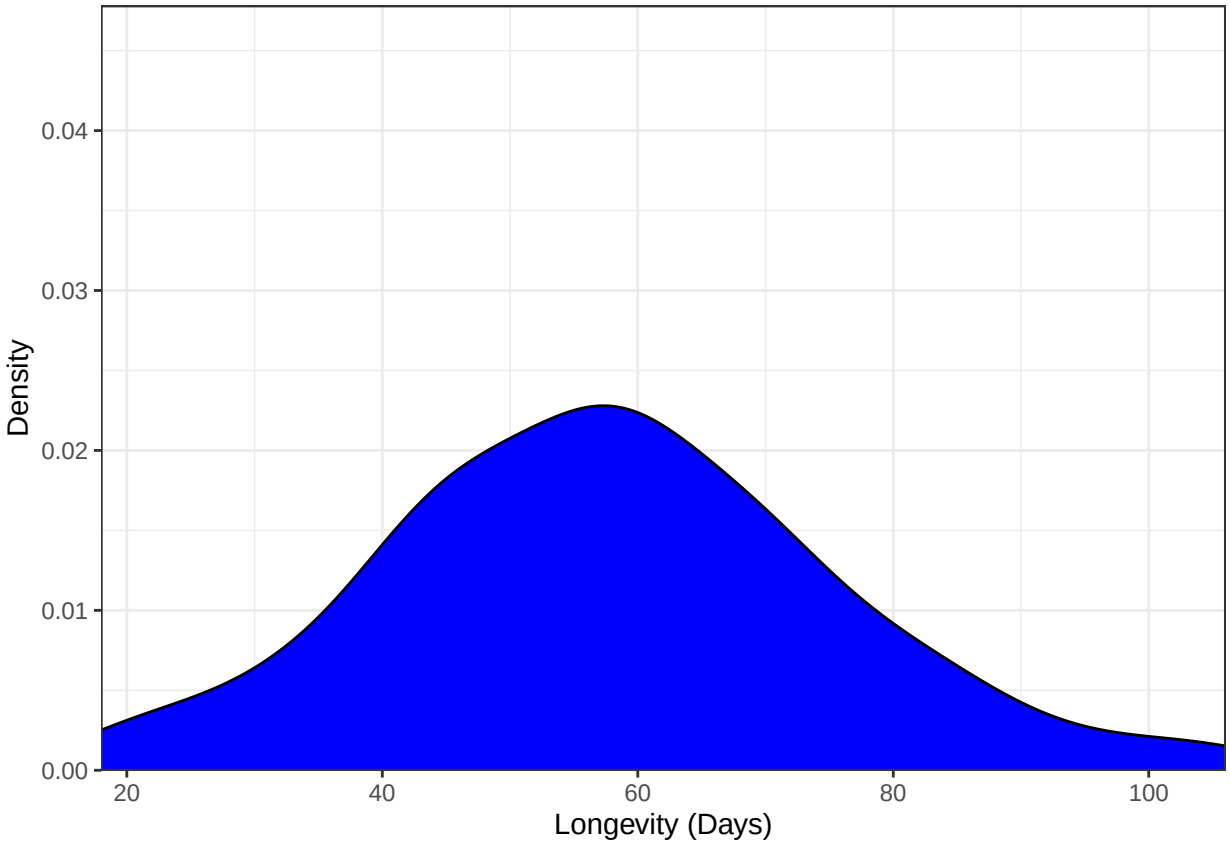

Figure 2: Density Plot for Male Fly Longevity

Figure 2 shows a density curve for male fly longevity. The curve appears to only have one peak and is thus unimodal and most of the values are located between 55 and 60 days.

Figure 3 shows box plots for the various conditions. From Figure 1 we can see that the sample arithmetic mean for the eight virgin female flies group is lower in value than the rest of the groups. It is also evident from the boxplots that the one pregnant female flies group has the largest sample arithmetic standard deviation. The control group has two outliers to it and the one virgin female fly group has one outlier.

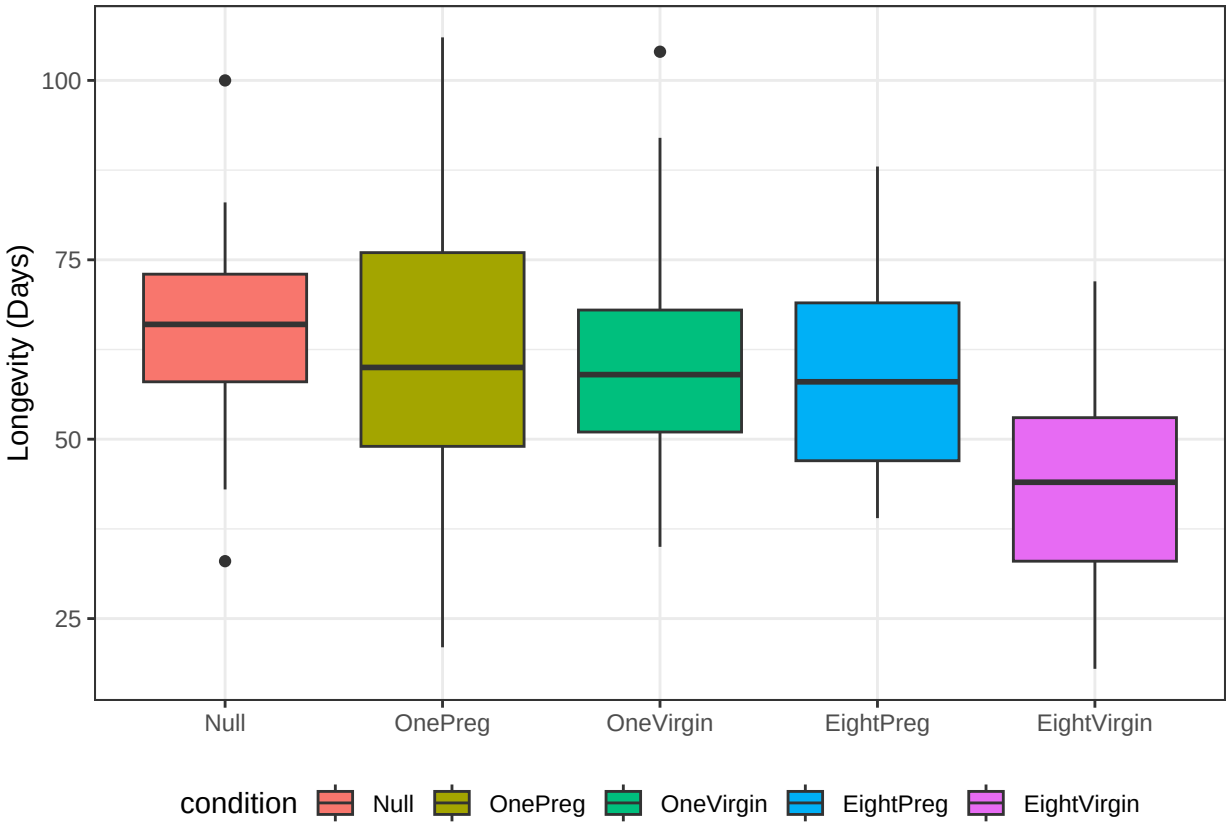

Figure 3: Side-by-side Box PLots for Male Fly Longevity

Results

We must test for three assumptions in order to use the ANOVA F test. These assumptions include the residuals following a Normal distribution, homoscedasticity, and the independence of observations.

Assumptions

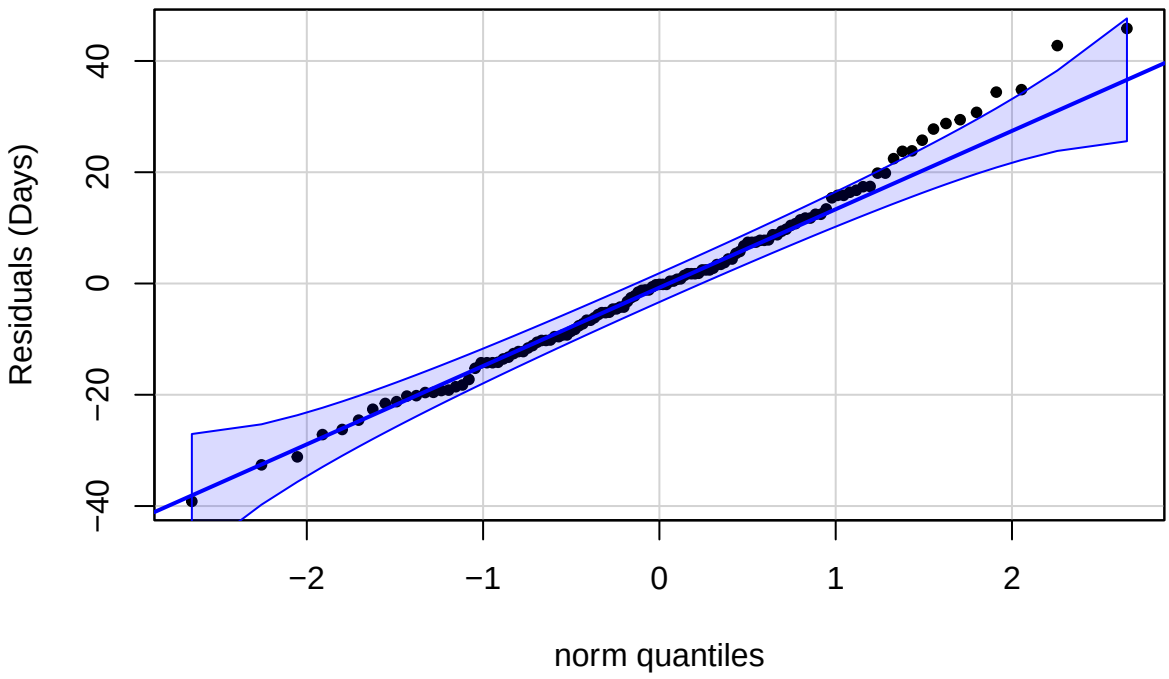

Figure 4: QQ Plot of the Residuals

Figure 4 shows a QQ plot of the residuals with a 90% confidence envelope. Most of the residuals fall into the confidence envelope, with only around 10 of the 125 residuals falling outside of the envelope. Since this is a 90% confidence envelope, we want no more than 10% of the residuals on the outside. 10 residuals outside of the envelope corresponds to less than 10% of the total residuals so we can say that our residuals satisfy the assumption of coming from a Normal distribution.

Figure 5 is a strip chart which allows us to assess the assumption of homoscedasticity. The third group has the most vertical height which appears to be less than double the height of the first or second group. There is no apparent pattern in the groups. Because we have a balanced design that leads to more robustness and these two features of the strip chart are not in extreme violation, we can assume homoscedasticity.

To assess the independence of observations, we can look at the study design. The males were kept independently from each other for the duration of the study and were put with females only one male at a time. Because the males were kept separate from one another, we can assume an independence of observations.

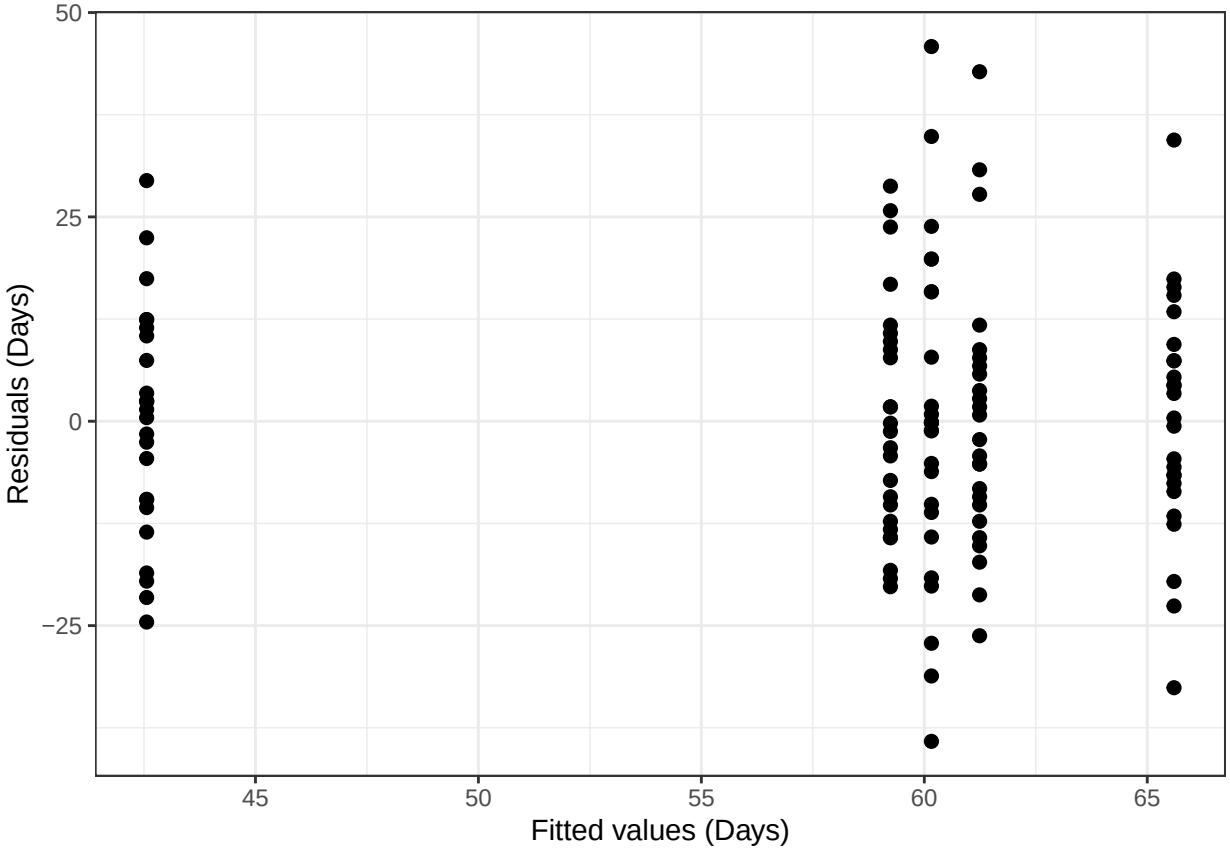

Figure 5: Strip Chart

ANOVA One-way Omnibus Results

Table 2: Modern ANOVA Table for Male Fly Longevity Study

Source	SS	df	MS	F	p-value	Eta Sq.	Omega Sq.	Epsilon Sq.
condition	7814.16	4	1953.5400	7.6703	< 0.0001	0.2036	0.1759	0.1771
Residuals	30562.64	120	254.6887					

Table 2 shows that a male fly's living condition accounts for 7.67 times as much variation as the residuals. Under the null model that the living condition of male flies does not affect their longevity, we would only observe this F value less than 0.01% of the time; out p-value is smaller than our unusualness threshold of 0.01 meaning that we reject the hull hypothesis and act as if a the living condition of male flies affects their longevity. This is supported by our values of eta, omega, and epsilon squared where between 17% and 20% of the variation in longevity could be explained by a male fly's living condition.

Table 3: Point Estimates from the Male Fly Longevity Study

	Estimate
Grand Mean	57.76
Null	7.84
OnePreg	2.40
OneVirgin	3.48
EightPreg	1.48
EightVirgin	-15.20

Table 3 shows that ignoring the living conditions, the male flies total days lived is 57.76 times as large as the total number of flies. For the group of flies in the control, they had an additional performance of 7.84 days per fly while the male flies with eight virgin female flies had a boost of -15.2 days per fly. The group of eight virgin female flies performed worse than the baseline.

Post Hoc

Table 4: Post Hoc Tukey HSD Comparisons

	Difference	Lower Bound	Upper Bound	Adj. p-Value
OnePreg-Null	-5.44	-20.469	9.589	0.748
OneVirgin-Null	-4.36	-19.389	10.669	0.870
EightPreg-Null	-6.36	-21.389	8.669	0.623
EightVirgin-Null	-23.04	-38.069	-8.011	0.000
OneVirgin-OnePreg	1.08	-13.949	16.109	0.999
EightPreg-OnePreg	-0.92	-15.949	14.109	1.000
EightVirgin-OnePreg	-17.60	-32.629	-2.571	0.001
EightPreg-OneVirgin	-2.00	-17.029	13.029	0.992
EightVirgin-OneVirgin	-18.68	-33.709	-3.651	0.001
EightVirgin-EightPreg	-16.68	-31.709	-1.651	0.003

Table 4 shows us all the possible pairwise comparisons. Using the unusualness threshold of 0.01, we can only reject the null hypothesis for the comparisons involving the eight virgin female flies. When this group was paired against all other groups, it was the only one that showed a statistically significant difference in the longevity of male flies.

Table 5: Post Hoc Comparison Effect Sizes

Pairwise Comparison	Cohen's d	Hedge's g	Prob. Superiority
Null vs. OnePreg	0.311	0.306	0.587
Null vs. OneVirgin	0.285	0.280	0.580
Null vs. EightPreg	0.442	0.436	0.623
Null vs. EightVirgin	1.630	1.605	0.876
OnePreg vs. OneVirgin	-0.059	-0.058	0.483
OnePreg vs. EightPreg	0.053	0.052	0.515
OnePreg vs. EightVirgin	1.017	1.001	0.764
OneVirgin vs. EightPreg	0.131	0.129	0.537
OneVirgin vs. EightVirgin	1.240	1.220	0.810
EightPreg vs. EightVirgin	1.182	1.163	0.798

Cohen's d and Hedge's g both are a standardized effect size that measures the difference between sample arithmetic mean for both groups. The largest values for these again come from the comparisons between the eight virgin female fly group compared with the other groups. This is also true for the probability of superiority where the greatest magnitude away from 0.5 is in the comparisons with the eight virgin female fly groups. The probability of superiority tells us the probability the observation from the first group will be larger than the observation from the seconf group given each observation is sampled randomly for each group. The largest probability of superiority is 0.81 between the one virgin and eight virgin group, meaning that we would find a male fly with a larger value of longevity from the one virgin group 81% of the time compared to the eight virgin group if we randomly sampled an observation from each group.

Discussion

We found that the living conditions of males flies appears to affect the longevity of male flies. From the data, it seems as if males flies living with eight virgin female flies had significantly shorter lifespans than the male flies in the rest of the living conditions. There did not seem to be a significant difference in longevity based on if the flies were virgin, pregnant, one female, or eight female, but rather the combination of eight virgin females led to a reduction in lifespan.

This study could be expanded to flies in different parts of the world to see if the pattern remains the same. We could also expand the attributes we track to become more accurate in our study such as the size of each fly or the age in days of the fly when the study began.

References and Materials Consulted

Hatfield, N. J. (2023). Unit 2: Study Design. In STAT 461: Analysis of Variance (ANOVA) Course Materials. Pennsylvania State University.

Hatfield, N. J. (2023). Unit 3: Oneway ANOVA. In STAT 461: Analysis of Variance (ANOVA) Course Materials. Pennsylvania State University.

Partridge, L., Farquhar, M. Sexual activity reduces lifespan of male fruitflies. *Nature* **294**, 580–582 (1981). https://doi.org/10.1038/294580a0

Code Appendix

```
knitr::opts_chunk$set(echo = TRUE)
packages <- c("tidyverse", "knitr", "kableExtra", "sass",
"parameters", "emmeans", "DescTools", "multcompView", 'hasseDiagram')
lapply(packages, library, character.only = TRUE)
options(knitr.kable.NA = "")
options(contrasts = c("contr.sum", "contr.poly"))
source("https://raw.github.com/neilhatfield/STAT461/master/rScripts/ANOVATools.R")
modelLabels <- c("1 Flies Live 1", "5 Living Condition 4", "125 (Male Flies) 120")
modelMatrix <- matrix(
  data = c(FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, TRUE, TRUE, FALSE),
  nrow = 3,
  ncol = 3,
  byrow = FALSE
)
hasseDiagram::hasse(
 data = modelMatrix,
 labels = modelLabels
)
flyData <- read.csv('fruitflies.csv', header = TRUE, sep = ',')
flyData$condition <- factor(
  x = flyData$condition,
  levels = c('Null','OnePreg','OneVirgin','EightPreg','EightVirgin')
)
flyModel <- aov(
  formula = longevity ~ condition,
  data = flyData
)
scoreStats <- psych::describeBy(
  x = flyData$longevity,
  group = flyData$condition,
  na.rm = TRUE,
  skew = TRUE,
  ranges = TRUE,
  quant = c(0.25, 0.75),
  IQR = FALSE,
  mat = TRUE,
  digits = 4
)

scoreStats %>%
  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 Male Fly 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_styling(
    font_size = 12,
    latex_options = c("scale_down", "HOLD_position")
  )

ggplot2::ggplot(
  data = flyData,
  mapping = aes(x = longevity)
) +
ggplot2::geom_density(
  na.rm = TRUE,
  color = "black",
  fill = "blue"
) +
ggplot2::theme_bw() +
xlab("Longevity (Days)") +
ylab("Density") +
ggplot2::scale_x_continuous(
  expand = expansion(mult = 0, add = 0),
) +
ggplot2::scale_y_continuous(
  expand = expansion(mult = 0, add = c(0, 0.025))
)
ggplot2::ggplot(
data = flyData,
mapping = aes(x = condition, y = longevity, fill = condition)
) +
ggplot2::geom_boxplot(
na.rm = TRUE
) +
ggplot2::theme_bw() +
xlab(NULL) +
ylab("Longevity (Days)") +
theme(
legend.position = "bottom"
)
car::qqPlot(
x = flyModel$residuals,
distribution = "norm",
envelope = 0.90,
id = FALSE,
pch = 20,
ylab = "Residuals (Days)"
)
ggplot(
data = data.frame(
residuals = flyModel$residuals,
fitted = flyModel$fitted.values
```

```
),
mapping = aes(x = fitted, y = residuals)
) +
geom_point(size = 2) +
theme_bw() +
xlab("Fitted values (Days)") +
ylab("Residuals (Days)")
parameters::model_parameters(
  model = flyModel,
  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 Male Fly Longevity Study",
  booktabs = TRUE,
  align = c("l", rep("c", 8))
) %>%
kableExtra::kable_styling(
  font_size = 10,
  latex_options = c("HOLD_position")
)
pointEst <- dummy.coef(flyModel)
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 Male Fly Longevity Study",
booktabs = TRUE,
align = "c"
) %>%
kableExtra::kable_styling(
font_size = 12,
latex_options = c("HOLD_position")
)
flyPH <- TukeyHSD(
x = flyModel,
conf.level = 0.99
)
knitr::kable(
x = flyPH$condition,
digits = 3,
caption = "Post Hoc Tukey HSD Comparisons",
```

```
col.names = c("Difference", "Lower Bound",
"Upper Bound", "Adj. p-Value"),
align = 'lcccc',
booktabs = TRUE,
) %>%
kableExtra::kable_styling(
bootstrap_options = c("condensed", "boardered"),
font_size = 12,
latex_options = "HOLD_position"
)
anova.PostHoc(flyModel) %>%
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_styling(
bootstrap_options = c("condensed", "boardered"),
font_size = 12,
latex_options = "HOLD_position"
)
```