Midterm2

Melissa Pulley

2024-04-21

1) Make a publication level figure that shows the weight of both infected and healthy tuna. Add a figure legend (caption) that accurately describes the figure. Be smart about the type of figure you use!

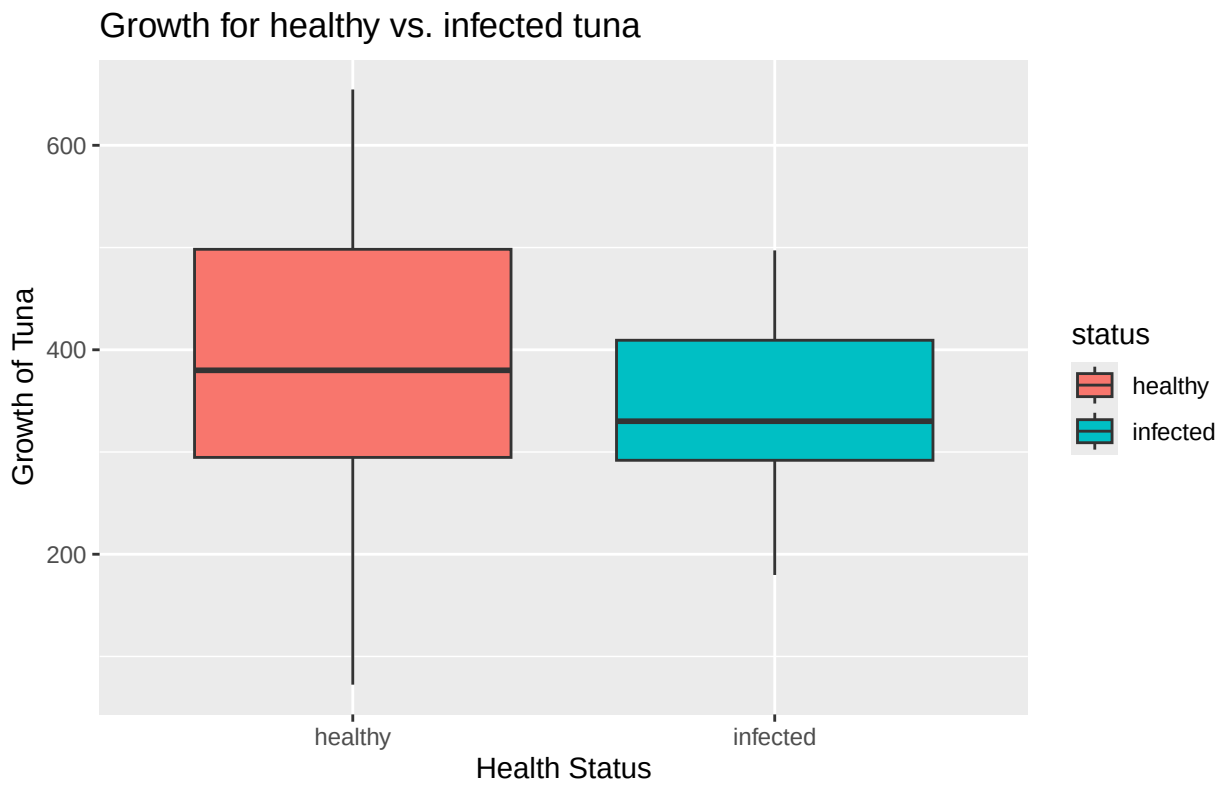
```
tuna = read.csv("tunadata.csv")
tuna$status = as.factor(tuna$status)
summary(tuna)
```

```
##      growth           status
##  Min.   : 72.4   healthy :25
##  1st Qu.:292.6   infected:25
##  Median :368.9
##  Mean   :364.3
##  3rd Qu.:439.4
##  Max.   :654.6
```

```
## Figure
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
ggplot(data=tuna, aes(x=status,y=growth, fill=status)) +
  geom_boxplot(outlier.colour="black", outlier.shape=15,
    outlier.size=2, notch=FALSE) +
    theme(plot.caption = element_text(hjust = 0.5)) +
    labs(title = "Growth for healthy vs. infected tuna",
         caption="This figure describe the size distribution of Yellowfish tuna based on their health s
         infected with Lacistorhynchidae tapeworm.") +
        ylab("Growth of Tuna") + xlab("Health Status")
```



This figure describe the size distribution of Yellowfish tuna based on their health status, healthy vs. infected with Lacistorhynchidae tapeworm.

2) Obtain the mean, median and IQR for both groups.

```
##
healthy_tuna = tuna$growth[which(tuna$status=="healthy")]
infected_tuna = tuna$growth[which(tuna$status=="infected")]

mean(healthy_tuna)
```

```
## [1] 390.156
```

```
median(healthy_tuna)
```

```
## [1] 379.9
```

```
quarts = quantile(healthy_tuna)
IQR = as.numeric(quarts[4] - quarts[2])
IQR
```

```
## [1] 203.6
```

```
mean(infected_tuna)
```

```
## [1] 338.4
```

```
median(infected_tuna)
```

```
## [1] 330.1
```

```
quarts = quantile(infected_tuna)
IQR = as.numeric(quarts[4] - quarts[2])
```

```
IQR
```

```
## [1] 117.4
```

For the healthy tuna, the mean growth is 390.156, the median is 379.9, and the IQR is 203.6. For the infected tuna, the mean growth is 338.4, median is 330.1, and the IQR is 117.4.

3) Based on the figure, mean, and IQR, do you thin there are differences in the data? Think biologically, what may cause the differences.

Yes, I think there are differences between the healthy and infected tuna sizes. Biologically, infected tuna are unable to fully use nutrients they consume for growth because the parasite consumes some of the nutrients from the tuna, making infected tuna smaller than healthy tuna.

4) Run an appropriate null hypothesis test (NHST) to test whether there are differences. Follow the 5 steps of the NHST process. Present the code and the output you got.

    1. The null hypothesis is that the mean of both population are equal. The alternative hypothesis is that at least one is different.
    2. We will use the tuna data.
    3. We will perform an ANOVA test.
    4. Because the p-value is above 0.05, we fail to reject the null hypothesis
    5. We cannot conclude that the mean of health vs. infected tuna are different.

```
#Consider the linear model
m1 = lm(growth~status, data = tuna)
summary(m1)
```

```
##
## Call:
## lm(formula = growth ~ status, data = tuna)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -317.76  -83.56   -9.86   91.37  264.44
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      390.16      23.55  16.569   <2e-16 ***
## statusinfected   -51.76      33.30  -1.554    0.127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 117.7 on 48 degrees of freedom
## Multiple R-squared:  0.04791,    Adjusted R-squared:  0.02808
## F-statistic: 2.416 on 1 and 48 DF,  p-value: 0.1267
```

```
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: growth
##           Df Sum Sq Mean Sq F value Pr(>F)
## status     1  33484   33484  2.4156 0.1267
## Residuals 48 665344   13861
```

5) Do you think the results accurately represent reality? If not, why? What biological processes might have affected your results?

I think the results are inaccurate and there are likely is a difference in mean growth of healthy vs. infected tuna. Logically, it feels that a fish with a parasite would impact the ability of the fish to thrive. Biologically, perhaps there is an issue with fairly sampling infected tuna as some may die from the disease.