

```

---
title: "Predicting Caterpillar Mass Project"
author: "Daniel Hrusovsky, Aydan Lacroix"
date: "2024-12-10"
output: html_document
---

```{r}
library(Stat2Data)
library(ggplot2)
library(leaps)
library(olsrr)
library(car)

#Retrieving Caterpillar data
data("Caterpillars")

#Cleaning variables to binary responses
Caterpillars$Fgp <- ifelse(Caterpillars$Fgp == "Y", 1, 0)
Caterpillars$Mgp <- ifelse(Caterpillars$Mgp == "Y", 1, 0)
Caterpillars$ActiveFeeding <- ifelse(Caterpillars$ActiveFeeding == "Y", 1, 0)

#Overview of Caterpillar data
summary(Caterpillars)
```

#Attempting a linear model to begin

```{r}
#Creating the full linear model containing every variable
full_model = lm(Mass ~ ActiveFeeding + Fgp + Mgp + Intake + WetFrass + DryFrass + Cassim +
Nfrass + Nassim, data = Caterpillars)

#Calculating the variation inflation factor for each predictor
vif(full_model)
```

```{r}
#Removing unnecessary predictors from the linear model

removed1_model = lm(Mass ~ ActiveFeeding + Fgp + Mgp + WetFrass + DryFrass + Nfrass +
Nassim, data = Caterpillars)
vif(removed1_model)

removed2_model = lm(Mass ~ ActiveFeeding + Fgp + Mgp + WetFrass + DryFrass + Nassim,
data = Caterpillars)
vif(removed2_model)

removed3_model = lm(Mass ~ ActiveFeeding + Fgp + Mgp + WetFrass , data = Caterpillars)
vif(removed3_model)
```

```{r}
#Visualizations determining if the linear model is an adequate representation of the data

plot(removed3_model)
qplot(WetFrass, Mass, data=Caterpillars)
qplot(LogWetFrass, LogMass, data=Caterpillars)
scatterplotMatrix(~ Mass + WetFrass , data=Caterpillars)
scatterplotMatrix(~ LogMass + LogWetFrass , data=Caterpillars)
```

```

```

```{r}
#Transitioning to a logarithmic model

log_model = lm(LogMass ~ ActiveFeeding + Fgp + Mgp + LogWetFrass, data = Caterpillars)
summary(log_model)

#Calculating the variation inflation factor for the logarithmic model
vif(log_model)
```

#Visualizations
```{r}
#Visualizations to determine adequate representation of the data

plot(log_model)
```

#ANOVA Table
```{r}

#Variance analysis of the logarithmic model
anova(log_model)
```

#Confidence intervals
```{r}

#Confidence interval of the logarithmic model
confint(log_model)
exp(confint(log_model))
```

#Prediction Intervals
```{r}

#Prediction interval of the model at a 95% level when the caterpillar is in active
feeding, a full growth period, a maximum growth period, and a logarithmic wet frass
consumption at -1.7 grams/day
exp(predict(log_model, interval = "prediction", level = .95, newdata =
data.frame(ActiveFeeding=1, Fgp=1, Mgp=1, LogWetFrass = -1.7)))

exp()
```

```