Homework 4

Sara Gonzalez

May 22, 2023

# Set-up

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.1 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(here)

## here() starts at /Users/saragonzalez/Desktop/github/ENVS-193DS\_homework-04\_Gonzalez-Sara

library(lterdatasampler)  
library(naniar)  
library(flextable)

##   
## Attaching package: 'flextable'  
##   
## The following object is masked from 'package:purrr':  
##   
## compose

library(car)

## Loading required package: carData  
##   
## Attaching package: 'car'  
##   
## The following object is masked from 'package:dplyr':  
##   
## recode  
##   
## The following object is masked from 'package:purrr':  
##   
## some

library(performance)  
library(broom)  
library(ggeffects)

# organize using here package  
fishes <- read\_csv(here("data", "ntl6\_v12.csv"))

# filter data based on what columns we need  
perch\_data <- fishes %>%  
 filter(spname=="YELLOWPERCH") %>%   
 subset(select = c(year4, spname, length, weight))

# Problem 1

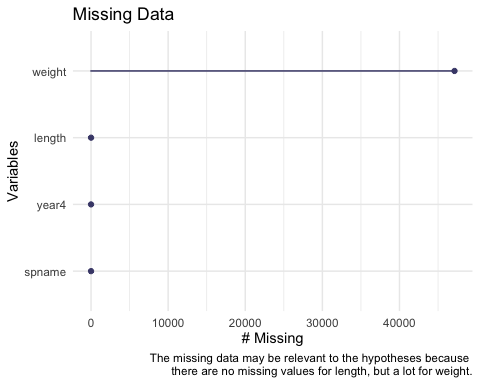
## Hypotheses

1. Mathematical Hypotheses
   1. Null: The predictor variable does not predict the response variable.
   2. Alternative: The predictor variable does predict the response variable.

* Biological Hypotheses
  1. Null: Fish length does not significantly predict fish weight for trout perch.
  2. Alternative: Fish length significantly predicts fish weight for trout perch.

## Missing Data

gg\_miss\_var(perch\_data)+  
 labs(title = "Missing Data", caption = "The missing data may be relevant to the hypotheses because   
 there are no missing values for length, but a lot for weight.") # added title and caption to plot



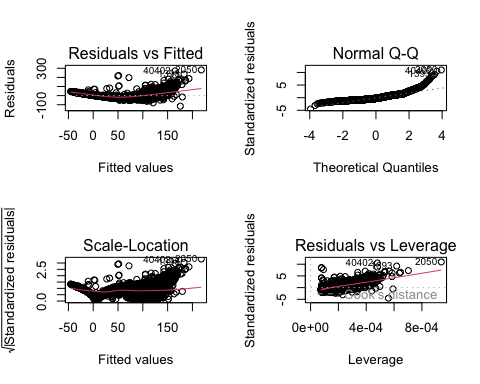
## Test

perch\_model <- lm(weight ~ length, data = perch\_data)  
perch\_model # y intercept and slope for fish length

##   
## Call:  
## lm(formula = weight ~ length, data = perch\_data)  
##   
## Coefficients:  
## (Intercept) length   
## -65.5171 0.8287

## Check Assumptions

# to display all four at once:  
par(mfrow = c(2,2)) # set up 2 x 2 grid  
plot(perch\_model)



1. The residual vs. fitted plot showcases the variance and tests for homoscedasticity. Since the residuals are relatively even and randomly distributed, I would say this meets the homoscedastic assumption. The normal qq plot tests for normal distribution. Since it is for the most part on the dotted line, I think the data meets the normal distribution assumption. The scale-location plot essentially says the same thing as Residuals vs Fitted, except using the square root of the standardized residuals. Again, the data looks randomly distributed, so I would say it is another way to show that the homoscedastic assumption is met. The residuals vs leverage plot tells you if there are any influential outliers. There are no outliers outside the dotted line on the left side, I will say none of the outliers are influential towards the model estimates.

## Results

1. summarize

summary(perch\_model)

##   
## Call:  
## lm(formula = weight ~ length, data = perch\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -115.628 -18.496 -6.575 15.037 282.109   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -65.517059 0.569359 -115.1 <2e-16 \*\*\*  
## length 0.828678 0.003731 222.1 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 25.69 on 12771 degrees of freedom  
## (47139 observations deleted due to missingness)  
## Multiple R-squared: 0.7943, Adjusted R-squared: 0.7943   
## F-statistic: 4.933e+04 on 1 and 12771 DF, p-value: < 2.2e-16

1. ANOVA table

model\_squares <- anova(perch\_model) # use anova to make table  
model\_squares

## Analysis of Variance Table  
##   
## Response: weight  
## Df Sum Sq Mean Sq F value Pr(>F)   
## length 1 32560510 32560510 49328 < 2.2e-16 \*\*\*  
## Residuals 12771 8429850 660   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

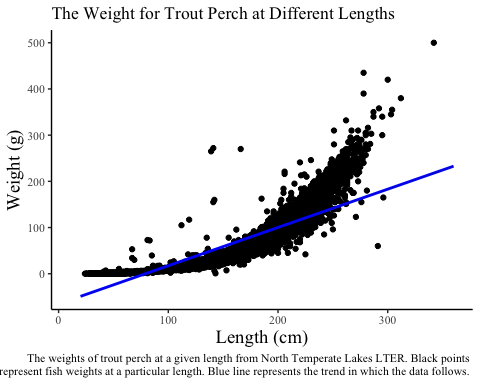
model\_squares\_table <- tidy(model\_squares) %>%   
 mutate(p.value = case\_when(  
 p.value < 0.001 ~ "< 0.001")) %>%   
 flextable() %>%   
 set\_header\_labels(df = "Degrees of Freedom", sumsq = "Sum of Squares", meansq = "Mean of Squares") # adding more informative titles  
model\_squares\_table

| term | Degrees of Freedom | Sum of Squares | Mean of Squares | statistic | p.value |
| --- | --- | --- | --- | --- | --- |
| length | 1 | 32,560,510 | 32,560,509.9039 | 49,328.31 | < 0.001 |
| Residuals | 12,771 | 8,429,850 | 660.0775 |  |  |

1. The ANOVA table relates to the information from the summary object in that it extract information from the summary table (p-value and degrees of freedom) and tells you relevant information about degrees of freedom, test statistic, p-value, R^2 as well as where the latter three came from.
2. Results Summary We found that our model including length significantly predicted the weight of the perch (F(1, 12771) = 49,330, p < 0.001, α = 0.05). With a 95% confidence interval, we found that for every unit increase in weight, there is a 0.83 ± 0.004 cm increase in length.
3. Visualization

predictions <- ggpredict(perch\_model, terms = "length")  
View(predictions) # terms corresponds to whatever predictor was in model

plot\_predictions <- ggplot(data = perch\_data, aes(x = length, y = weight))+  
 geom\_point() + # first plot underlying data  
 # plotting model predictions from predictions object from ggeffects  
 geom\_line(data = predictions, aes(x = x, y = predicted), color = "blue", linewidth = 1) +  
 geom\_ribbon(data = predictions, aes(x = x, y = predicted, ymin = conf.low, ymax = conf.high), alpha = 0.2) +  
 labs(x = "Length (cm)", y = "Weight (g)", title = "The Weight for Trout Perch at Different Lengths", caption = "The weights of trout perch at a given length from North Temperate Lakes LTER. Black points   
 represent fish weights at a particular length. Blue line represents the trend in which the data follows. ") +  
 theme\_classic()+  
 theme(text = element\_text(family = "Times New Roman"),  
axis.title = element\_text(size = 14),  
 )  
plot\_predictions



link to repo: <https://github.com/gnlsara/ENVS-193DS_homework-04_Gonzalez-Sara>