

Sparrow Dataset Analysis

Francisco Solis

R Coding with the Sparrow Dataset

This data is a stratified random sample of 116 Savannah sparrows at Kent Island. The **weight (in grams)** and **wing length (in mm)** were obtained for birds from nests that were reduced, controlled, or enlarged. The data are available in the file `Sparrows` from the R package `Stat2Data`.

Objectives a,b,c,d,e

- a. Load in Sparrows dataset and display the first 5 rows of data.
- b. Create scatterplot of weight versus wing length with a fitted regression line.
- c. Fit a simple linear regression and display the summary output.
- d. `par(mfrow=c(1,2))` to display the following two diagnostic plots side-by-side:
 - the residual vs fitted
 - Normal Q-Q
- e. `par(mfrow=c(1,2))` to display the following two diagnostic plots side-by-side:
 - Cook's distance
 - Residual vs Leverage

```
install.packages('Stat2Data',repos='https://mirror.csclub.uwaterloo.ca/CRAN/')
## Installing package into 'C:/Users/francisco solis jr/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)
```

```
## package 'Stat2Data' successfully unpacked and MD5 sums checked
```

```
##
```

```
## The downloaded binary packages are in
```

```
## C:\Users\francisco solis jr\AppData\Local\Temp\RtmpKaLPP9\downloaded_packages
```

```
library(Stat2Data)
```

```
# (a) Data structure and the first 5 rows
```

```
data("Sparrows")
```

```
str(Sparrows)
```

```
## 'data.frame': 116 obs. of 3 variables:
```

```
## $ Treatment : Factor w/ 3 levels "control","enlarged",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
## $ Weight : num 14.9 15 14.3 17 16 16.2 12.4 14.6 12.2 13 ...
```

```
## $ WingLength: num 29 31 25 29 30 31.5 23.5 26 25.5 27 ...
```

```
head(Sparrows,5)
```

```
##   Treatment Weight WingLength
## 1   control   14.9         29
## 2   control   15.0         31
## 3   control   14.3         25
## 4   control   17.0         29
## 5   control   16.0         30
```

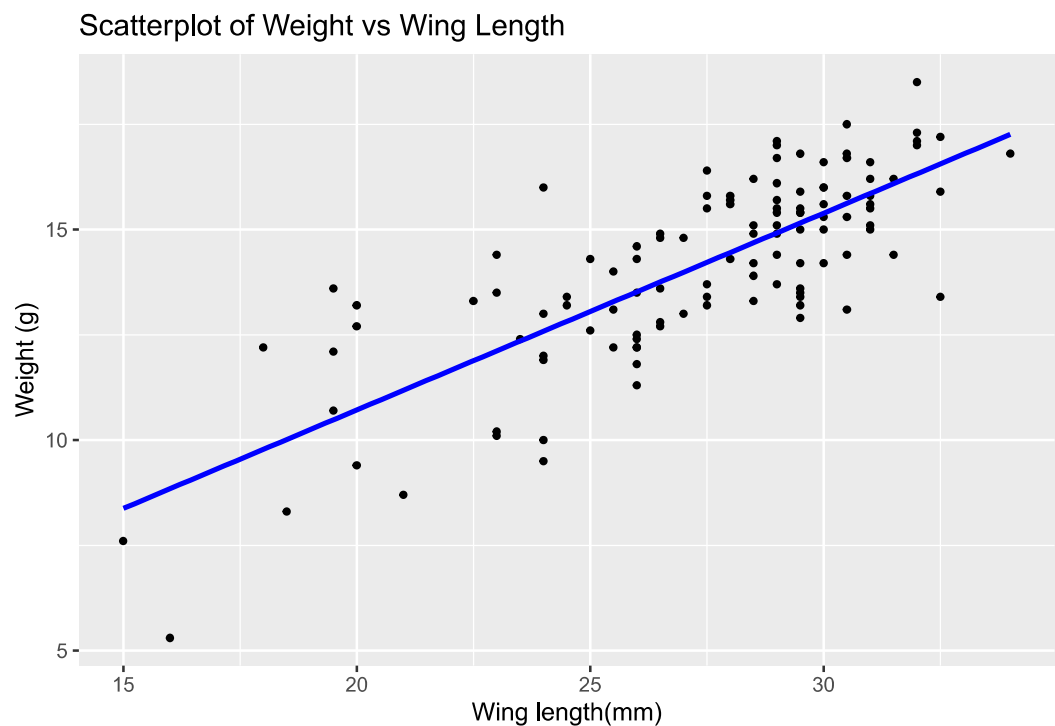
```
# (b) ggplot to form the scatterplot
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

```
ggplot(Sparrows, aes(x=WingLength, y=Weight))+  
  geom_point(color="black", size=1)+  
  geom_smooth(method="lm",se=FALSE,color="blue", linetype="solid")+  
  labs(title="Scatterplot of Weight vs Wing Length",  
        x="Wing length(mm)",  
        y="Weight (g)")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



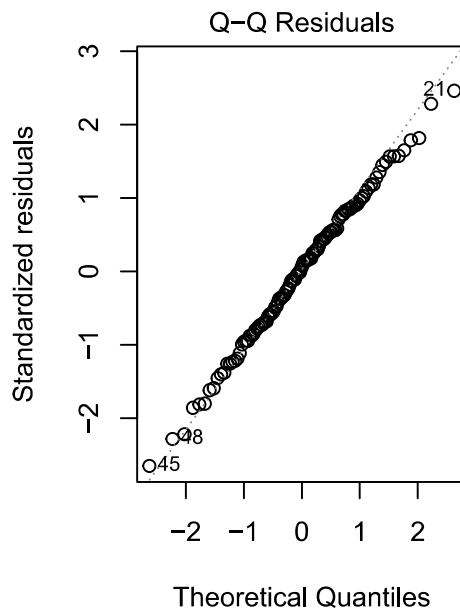
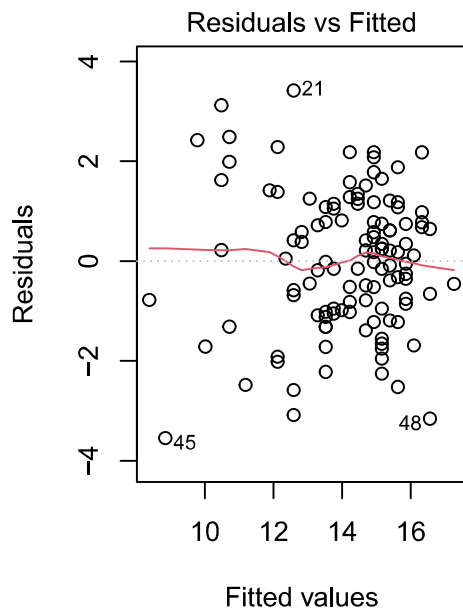
```
# (c) Fiting a regression model and outputting the summary
```

```
SLR<-lm(Weight~WingLength, data=Sparrows)
summary(SLR)
```

```
##
## Call:
## lm(formula = Weight ~ WingLength, data = Sparrows)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5440 -0.9935  0.0809  1.0559  3.4168
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.36549    0.95731   1.426   0.156
## WingLength   0.46740    0.03472  13.463 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.4 on 114 degrees of freedom
## Multiple R-squared:  0.6139, Adjusted R-squared:  0.6105
## F-statistic: 181.3 on 1 and 114 DF,  p-value: < 2.2e-16
```

```
# (d) Diagnostic plots (Residuals vs Fitted plot and Normal QQ plot)
```

```
par(mfrow=c(1,2))
plot(SLR,which=1:2)
```

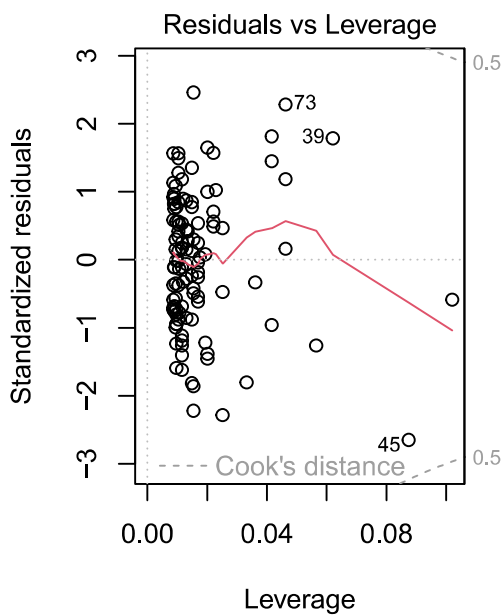
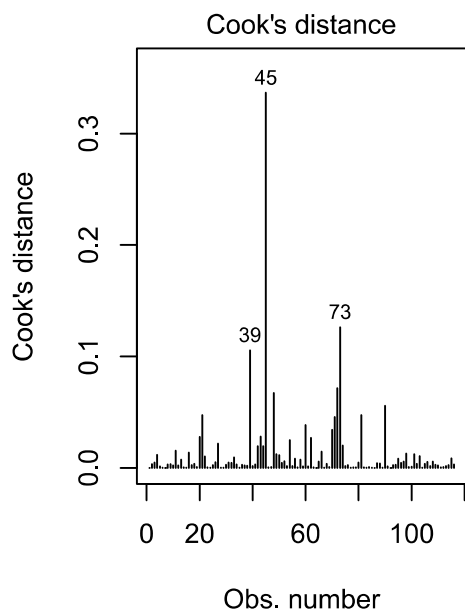


```
par(mfrow=c(1,1))
```

```
# (e) Diagnostic plots (Cook's distance and Leverage plots)
```

```
par(mfrow=c(1,2))
```

```
plot(SLR,which=4:5)
```



```
par(mfrow=c(1,1))
```