

Regression Assignment 1

Shinjon Ghosh

2025-01-17

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

data(iris)
names(iris)

## [1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"

head(iris)

##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1         5.1         3.5          1.4          0.2  setosa
## 2         4.9         3.0          1.4          0.2  setosa
## 3         4.7         3.2          1.3          0.2  setosa
## 4         4.6         3.1          1.5          0.2  setosa
## 5         5.0         3.6          1.4          0.2  setosa
## 6         5.4         3.9          1.7          0.4  setosa

ir1 <- lm(Petal.Length ~ Petal.Width, data = iris)
summary(ir1)

##
## Call:
## lm(formula = Petal.Length ~ Petal.Width, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33542 -0.30347 -0.02955  0.25776  1.39453
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.08356    0.07297   14.85  <2e-16 ***
## Petal.Width  2.22994    0.05140   43.39  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##  
## Residual standard error: 0.4782 on 148 degrees of freedom  
## Multiple R-squared:  0.9271, Adjusted R-squared:  0.9266  
## F-statistic: 1882 on 1 and 148 DF,  p-value: < 2.2e-16
```

1. $\text{Petal.Length} = 1.08356 + 2.22994 \cdot \text{Petal.Width}$

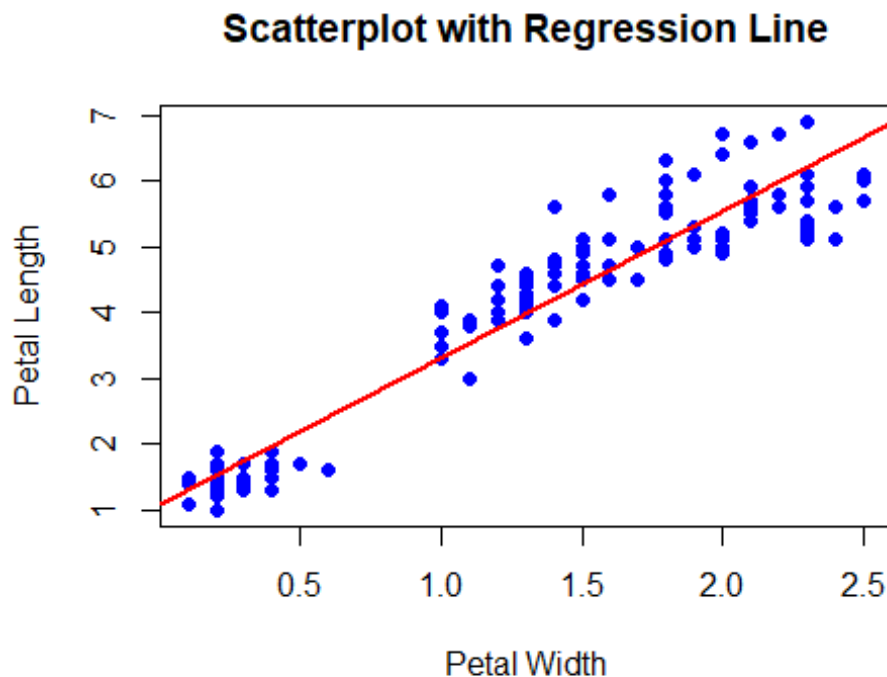
Where β_0 is Intercept and β_1 is Slope and the value is $\beta_0 = 1.08356$ & $\beta_1 = 2.22994$.

2. Null Hypothesis $H_0: \beta_1 = 0$;

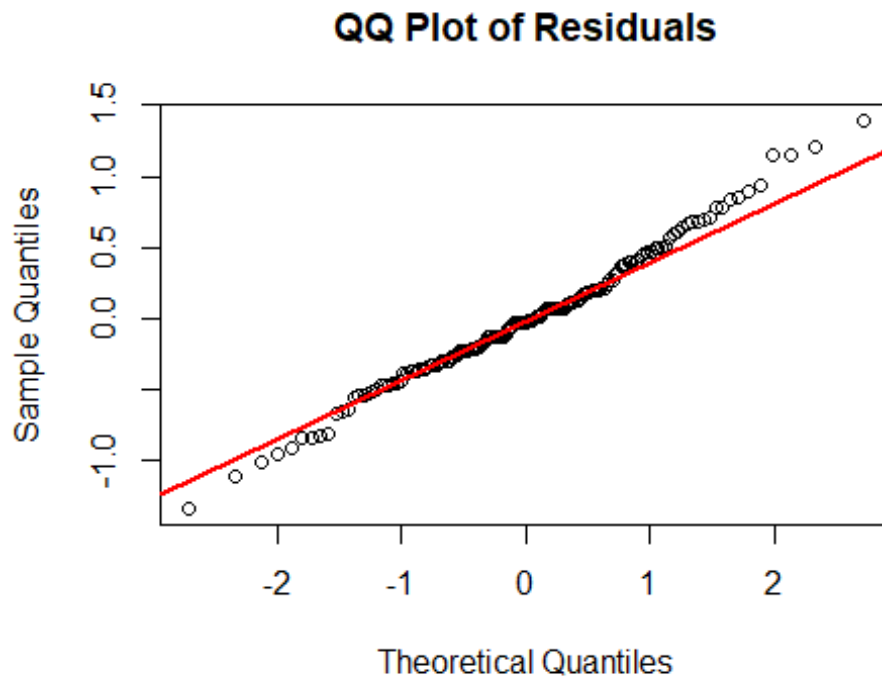
Alternative Hypothesis $H_a: \beta_1 \neq 0$

From the R-code, we get P-value is $< 2e-16$ which is less than 0.05. We reject null hypothesis. So, there is a significant linear relationship between Petal.Length and Petal.Width.

```
plot(iris$Petal.Width, iris$Petal.Length,  
     xlab = "Petal Width", ylab = "Petal Length",  
     main = "Scatterplot with Regression Line",  
     pch = 19, col = "blue")  
abline(ir1, col = "red", lwd = 2)
```



```
qqnorm(resid(ir1), main = "QQ Plot of Residuals")  
qqline(resid(ir1), col = "red", lwd = 2)
```



4) Most residual points fall close to the red line, indicating that the residuals are approximately normally distributed. There are slight deviations at the tails (both left and right extremes). This suggests that there might be some slight non-normality in the residuals, particularly in the extreme values.