GLM Practical Session 1

alexaoh

9/28/2021

Linear Regression for Cholesterol

```
data <- read.csv2("COL.csv", header = T)
summary(data)
```

```
#>        A               H               W               C
#>  Min.   : 9.00   Min.   :103.0   Min.   :37.30   Min.   : 67.5
#>  1st Qu.:12.00   1st Qu.:130.5   1st Qu.:53.23   1st Qu.:166.5
#>  Median :15.00   Median :151.5   Median :66.60   Median :217.8
#>  Mean   :14.71   Mean   :147.4   Mean   :64.57   Mean   :218.2
#>  3rd Qu.:18.00   3rd Qu.:167.2   3rd Qu.:74.95   3rd Qu.:262.4
#>  Max.   :20.00   Max.   :187.0   Max.   :89.70   Max.   :438.5
```

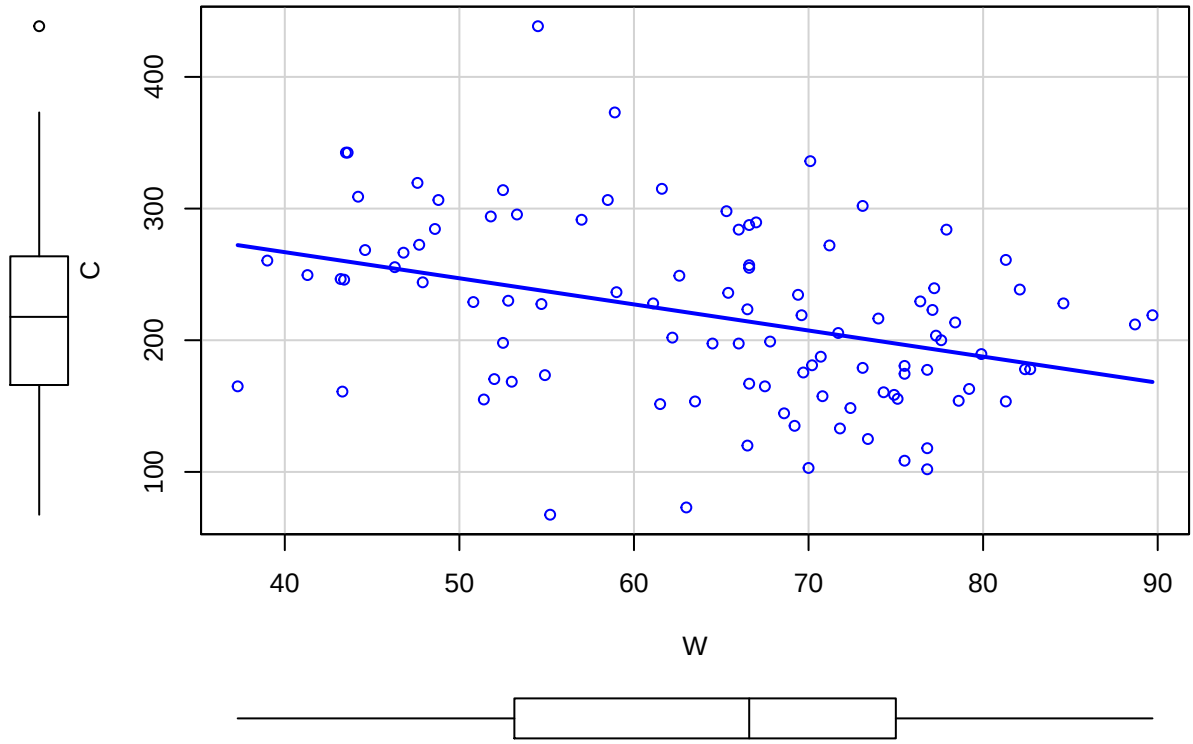
```
p <- 2
n <- dim(data)[1]

# Fit linear model.
lm.fit <- lm(C~W, data = data)
summary(lm.fit)
```

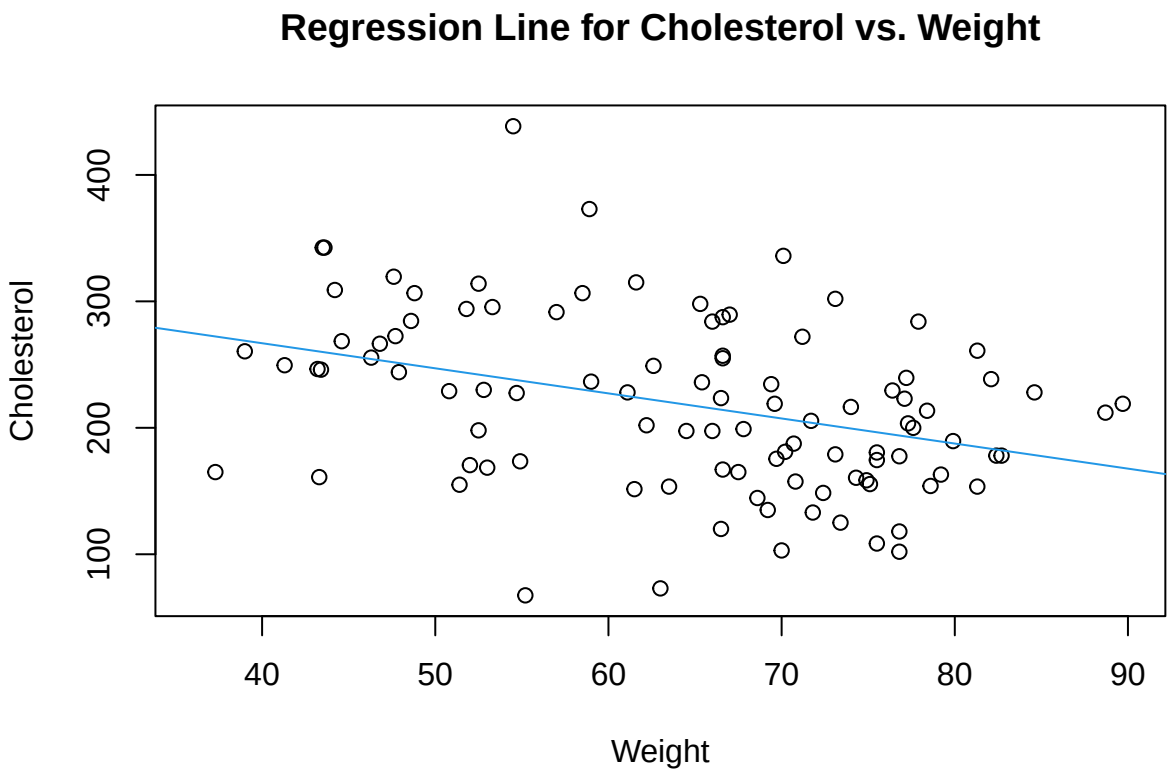
```
#>
#> Call:
#> lm(formula = C ~ W, data = data)
#>
#> Residuals:
#>     Min      1Q  Median      3Q     Max
#> -169.24  -39.81   -4.49   47.19  200.37
#>
#> Coefficients:
#>             Estimate Std. Error t value Pr(>|t|)
#> (Intercept) 346.2251    33.1983   10.43  < 2e-16 ***
#> W            -1.9835     0.5046   -3.93 0.000158 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 63.55 on 98 degrees of freedom
#> Multiple R-squared:  0.1362, Adjusted R-squared:  0.1274
#> F-statistic: 15.45 on 1 and 98 DF,  p-value: 0.0001581
```

Scatterplot of Points and Regression Line.

```
# Can be done manually and with a function.
scatterplot(C~W, smooth = F, data = data)
```



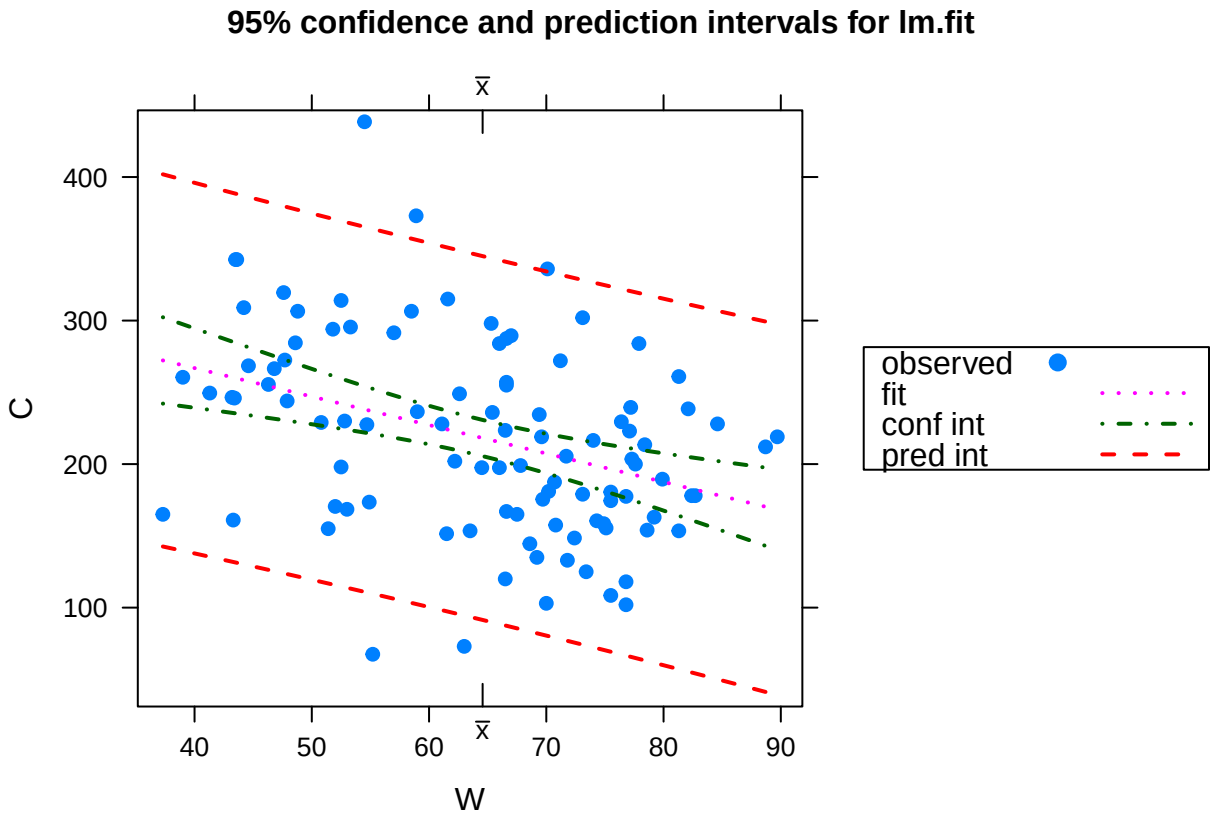

```
plot(data[, "W"], data[, "C"], main = "Regression Line for Cholesterol vs. Weight",
     xlim = c(36, 90), ylim = c(66, 440), xlab = "Weight", ylab = "Cholesterol")
abline(lm.fit, col = 4)
```




```
# Could do the plot from above with the scatterplot function above (comes from 'car' package).
```

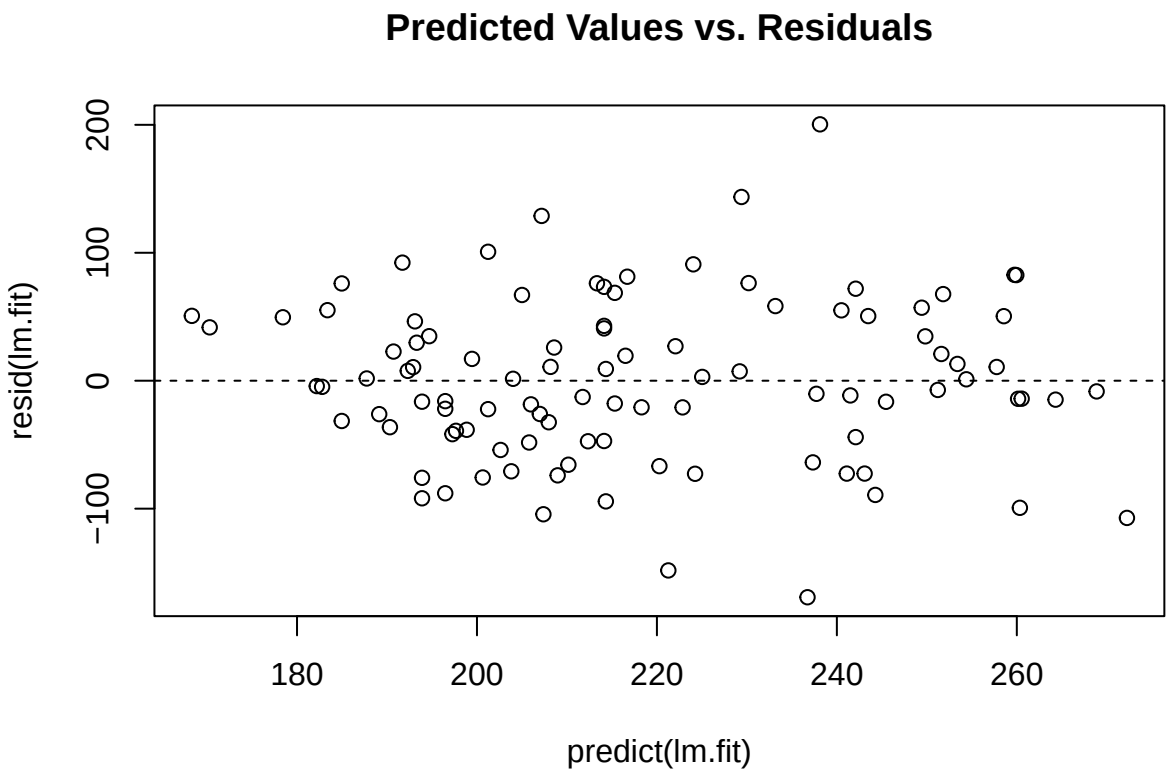
Plot Regression Line with Conf. and Pred. Intervals

```
# Plot confidence and prediction intervals with regression line (From package 'HH').
ci.plot(lm.fit)
```



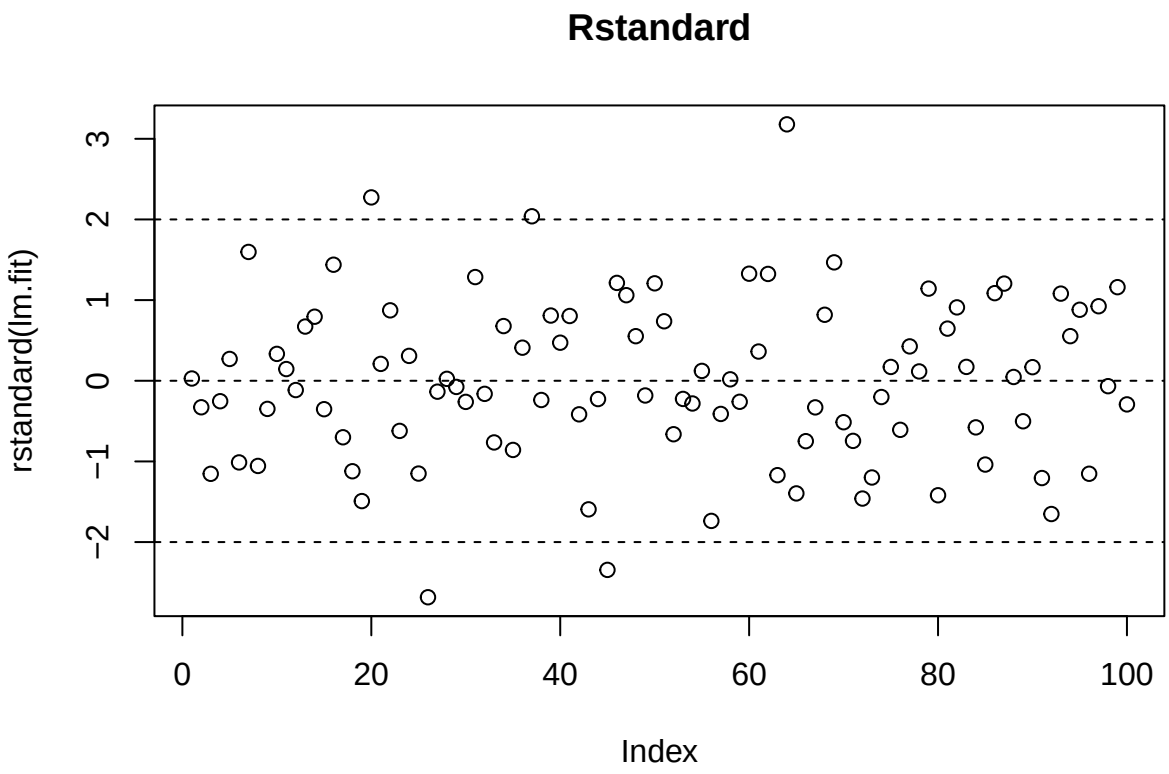
Plot Predicted Values vs. Residuals

```
# Plot the predicted values vs. residuals.
plot(predict(lm.fit), resid(lm.fit), main = "Predicted Values vs. Residuals")
abline(h=0, lty = 2)
```

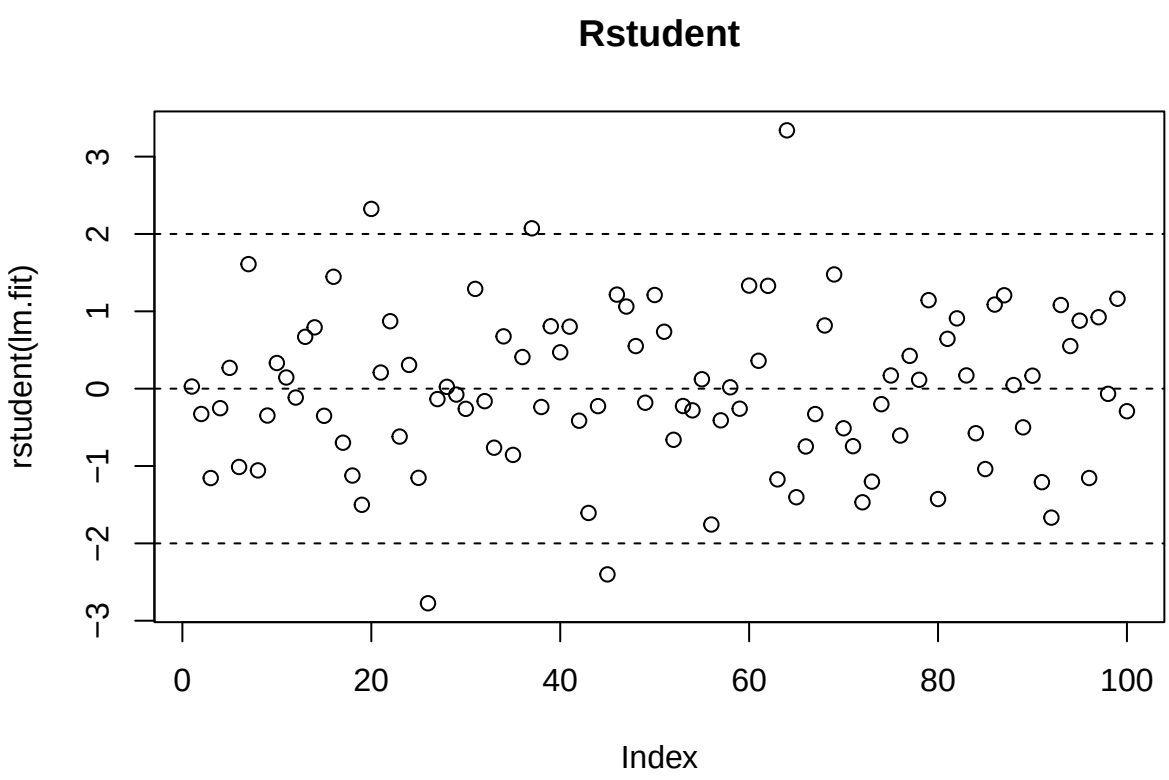


Plot Standardized/Studentized Residuals

```
plot(rstandard(lm.fit), main = "Rstandard")
abline(h=c(-2, 0, 2), lty = 2)
```




```
plot(rstudent(lm.fit), main = "Rstudent")
abline(h=c(-2, 0, 2), lty = 2)
```


Rstudent
rstudent(lm.fit)
3
2
1
0
−1
−2
−3
0
20
40
60
80
100
Index