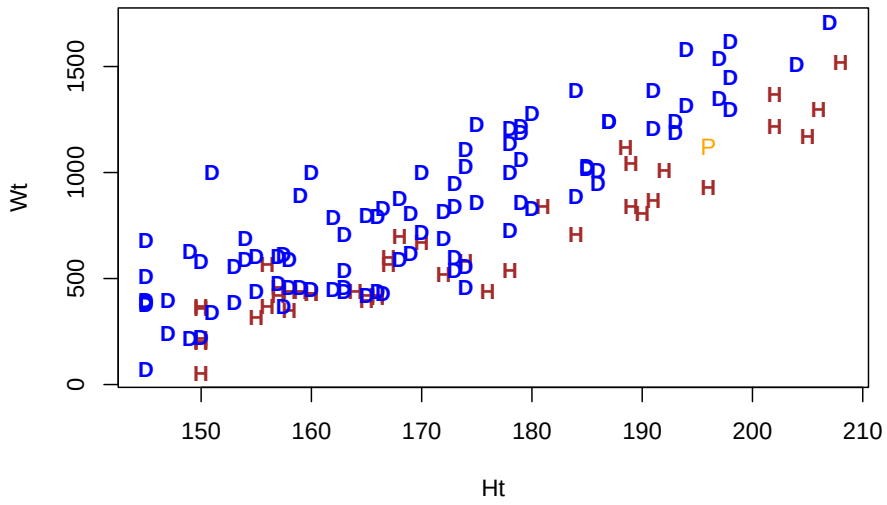
Assignment 1 STATS 330

Anish Hota

2025-03-26

Plot

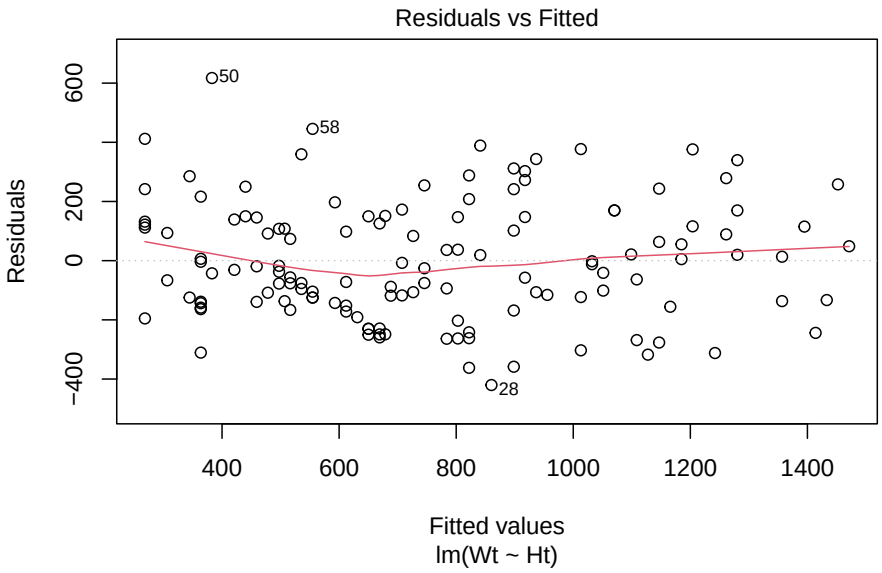
```
PPW.df <- read.csv("PPW.csv")
plot(Wt~Ht, type="n",data=PPW.df)
text(Wt~Ht, labels=substr(How,1,1),
     col=ifelse(How=="Drown","blue", "brown"),
     data=PPW.df, cex=0.9, font=2)
points(196,1125,pch = "P", col = "orange")
```

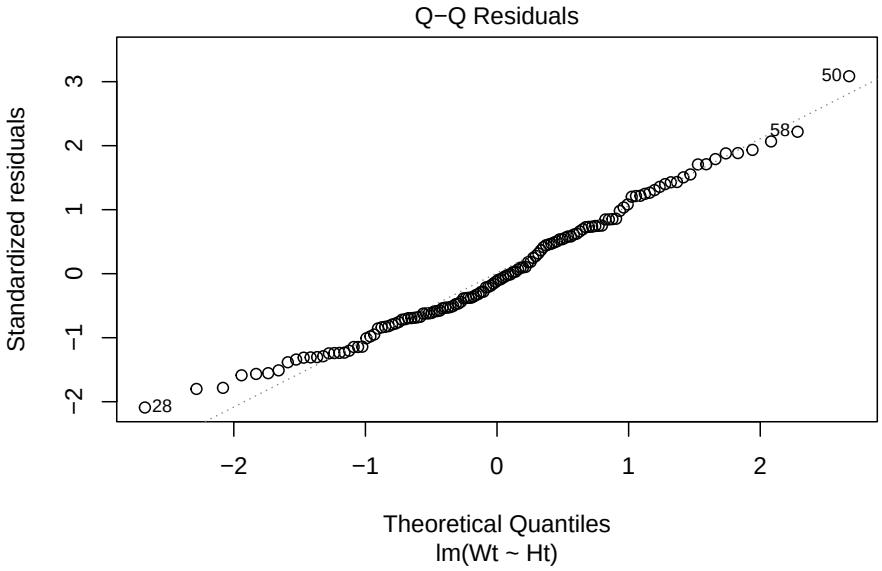


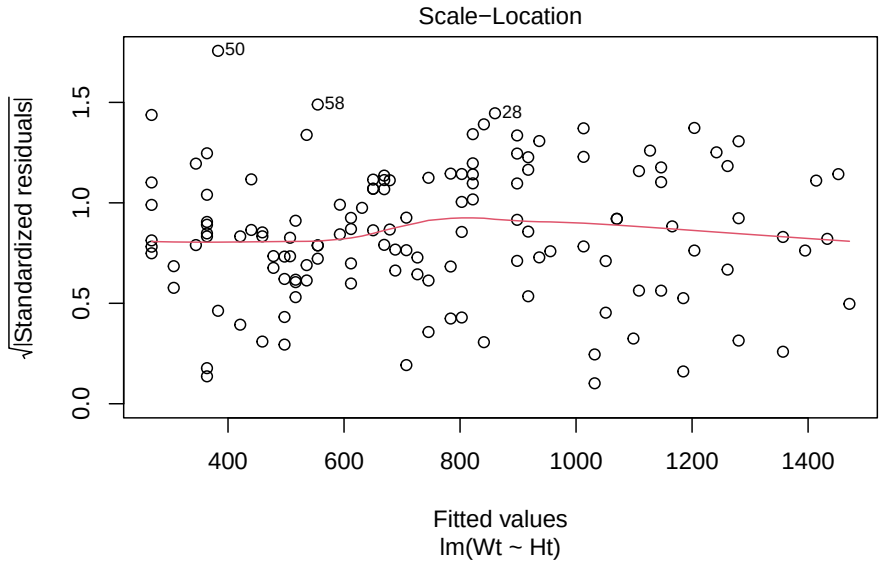
The point P is the height and weight of lungs of the person who got killed in this case, this is why it is essential to look at this point when inspecting the data.

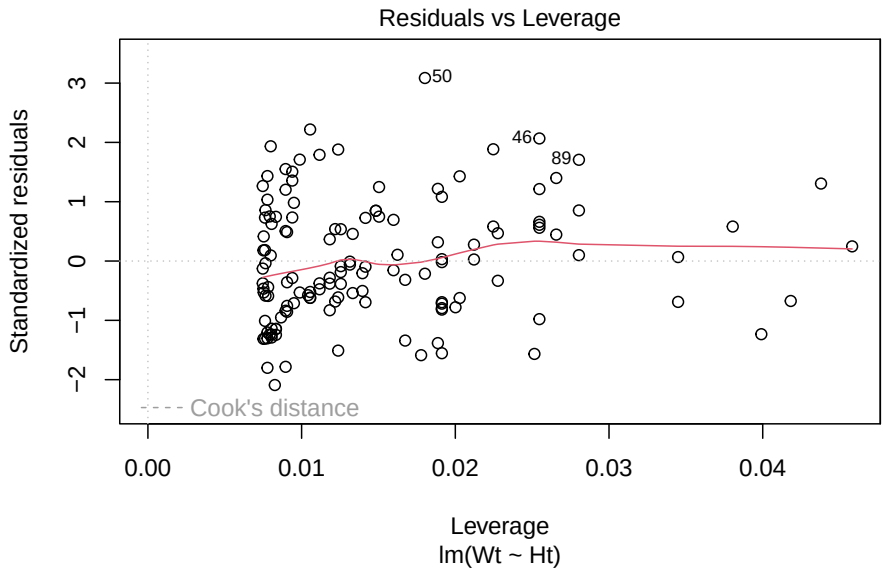
Model Building and Check Assumptions

```
PPW.fit <- lm (Wt ~Ht, data = PPW.df)
plot(PPW.fit)
```

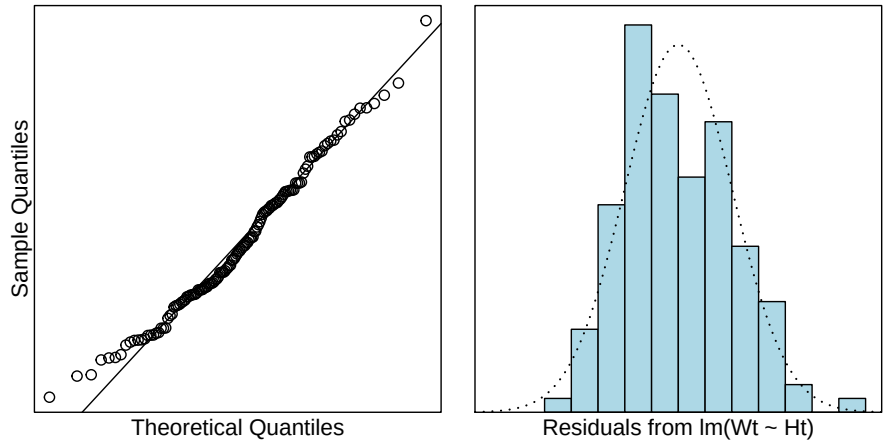

Residuals vs Fitted
Residuals
Fitted values
lm(Wt ~ Ht)
50
58
28


Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(Wt ~ Ht)
50
58
28

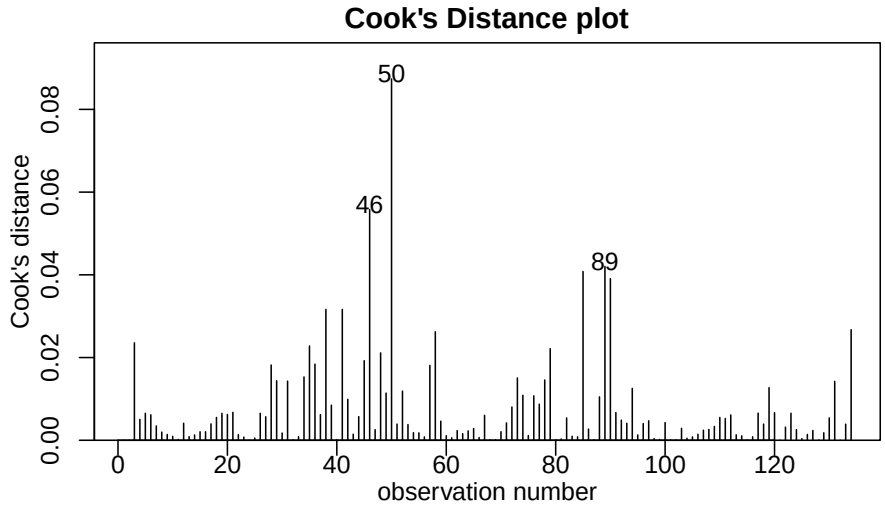

Scale−Location
√|Standardized residuals|
Fitted values
lm(Wt ~ Ht)
50
58
28




```
normcheck(PPW.fit)
```




```
cooks20x(PPW.fit)
```



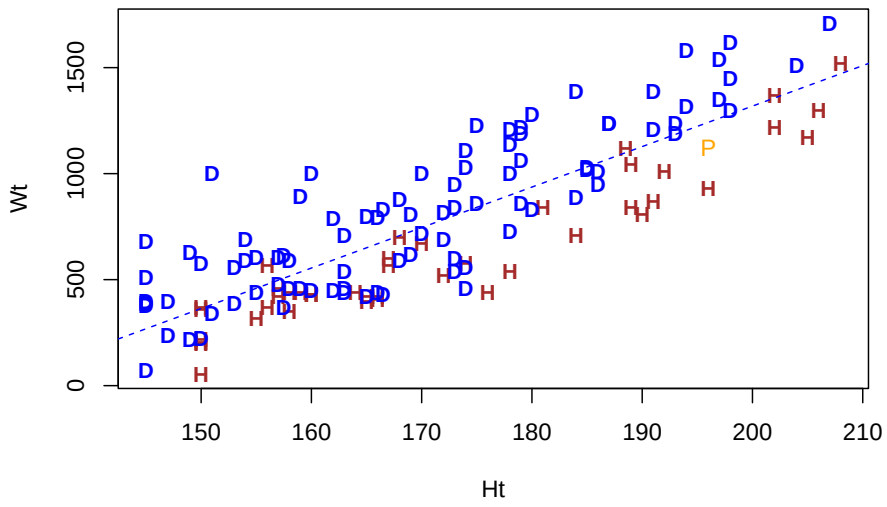

```
summary(PPW.fit)
```

```
##
## Call:
## lm(formula = Wt ~ Ht, data = PPW.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -420.21 -139.08  -22.42  144.10  617.18
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2500.651    181.264  -13.80   <2e-16 ***
## Ht             19.096      1.057   18.06   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 201.9 on 132 degrees of freedom
## Multiple R-squared:  0.7118, Adjusted R-squared:  0.7097
## F-statistic: 326.1 on 1 and 132 DF,  p-value: < 2.2e-16
```

The data seems to be normalized and there are no outstanding points in data. So all assumptions are satisfied. The data explains 71.2% of the data.

Fitted Model

```
plot(Wt~Ht, type="n",data=PPW.df)
text(Wt~Ht, labels=substr(How,1,1),
     col=ifelse(How=="Drown","blue", "brown"),
     data=PPW.df, cex=0.9, font=2)
points(196,1125,pch = "P", col = "orange")
abline(PPW.fit, col ="blue", lty = 2)
```

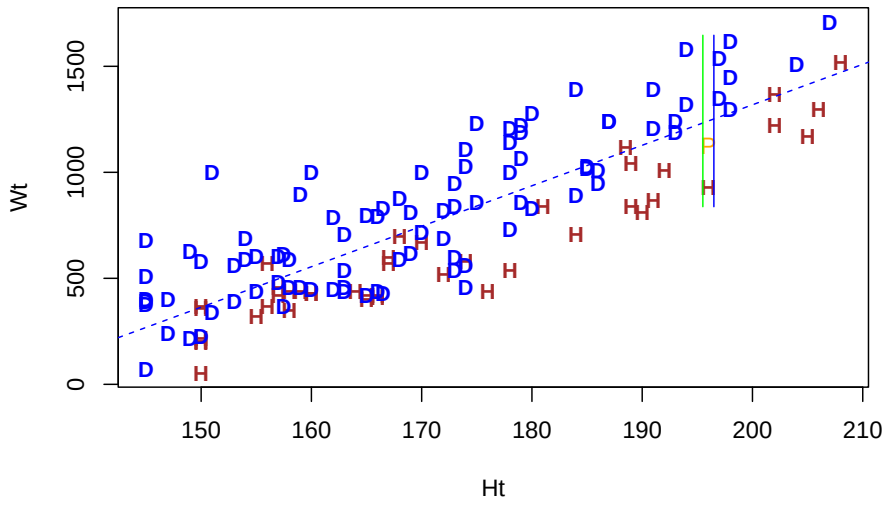


This data shows that there is a positive increase in weight compared to height. This relates to our case as we are seeing if the height of the man got killed could have an effect of how he died, as he was a pretty tall man.

Prediction model

```
plot(Wt~Ht, type="n",data=PPW.df)
text(Wt~Ht, labels=substr(How,1,1),
```

```
     col=ifelse(How=="Drown","blue", "brown"),
     data=PPW.df, cex=0.9, font=2)
points(196,1125,pch = "P", col = "orange")
abline(PPW.fit, col ="blue", lty = 2)
preddata <- data.frame(Ht = c(196,196))
preds <- predict(PPW.fit, preddata, interval = "prediction")
segments(196-.5,preds[1,2], 196-.5, preds[1,3], col="green")
segments(196+.5,preds[2,2], 196+.5, preds[2,3], col="blue")
```



This prediction interval shows the person in question and compares this person weight and height to other people in this data and checks to see if there are any similar traits.

GLM

```
glm.model <- glm(Wt ~Ht, data = PPW.df)
summary(glm.model)
```

```
##
## Call:
## glm(formula = Wt ~ Ht, data = PPW.df)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2500.651    181.264  -13.80   <2e-16 ***
## Ht             19.096      1.057   18.06   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 40747.97)
##
##     Null deviance: 18666034  on 133  degrees of freedom
## Residual deviance:  5378732  on 132  degrees of freedom
## AIC: 1806.7
##
## Number of Fisher Scoring iterations: 2
```

```
sqrt(deviance(glm.model)) / df.residual(glm.model)
```

```
## [1] 17.56977
```

```
1 - sum(resid(glm.model)^2) / sum((PPW.df$Wt - mean(PPW.df$Wt))^2)
```

```
## [1] 0.7118439
```

Binomial Logit model

```
PPW.df$Hung <- ifelse(PPW.df$How == "Hung",1,0)
logitmodel <- glm(Hung ~ Ht + Wt, data = PPW.df, family = "binomial")
summary(logitmodel)
```

```
##
## Call:
## glm(formula = Hung ~ Ht + Wt, family = "binomial", data = PPW.df)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -25.465397   5.227893  -4.871 1.11e-06 ***
## Ht            0.181604   0.036738   4.943 7.68e-07 ***
## Wt           -0.008898   0.001713  -5.195 2.05e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 168.18  on 133  degrees of freedom
## Residual deviance: 120.69  on 131  degrees of freedom
## AIC: 126.69
##
## Number of Fisher Scoring iterations: 5
```

Prediction for PPW

```
ppwcase <- data.frame(Ht = 196, Wt = 1125)

pred <- predict(logitmodel, newdata = ppwcase, type = "link", se.fit = TRUE)

lower <- pred$fit - 1.96 * pred$se.fit
upper <- pred$fit + 1.96 * pred$se.fit
round(c(plogis(lower), plogis(pred$fit), plogis(upper)), 3)
```

```
##     1     1     1
## 0.325 0.530 0.725
```

This shows the probability that PPW was hung rather than drowned. It proves that the probability that PPW was hung is between 32.5% and 72.5%.

Final Statement

The probability suggest that there is a high chance that he was hung meaning that means that we can not favour the defense and therefore they can be proven guitly based on these statistics.