Lab 7

Kyle Salitrik
997543474

Part 1

The linear model obtained is:

$$\widehat{\text{hardness}} = 81.4793 + 1.8687*\text{Aluminum}$$

Which shows a positive relationship between the amount of aluminum present and the hardness of the cement. To investigate the accuracy of the model, we will examine the confidence and prediction intervals for points at the 50th and 90th quantiles. Before starting the analysis, I would assume that the 0.50 quantile to be a more trustworthy predictor because when reaching the 90th percentile, there are less data points to predict the behavior in that region.

The following table summarizes the information obtained:

	50th	90th
Quantile Value	7	11
Point Estimates	94.56058	102.03557
CI Lower Bound	87.99073	94.30998
CI Upper Bound	101.1304	109.7612
CI Width	13.13970	15.45117
PI Lower Bound	70.05417	77.19436
PI Upper Bound	119.0670	126.8768
PI Width	49.01281	49.68242

As stated before, the reason that the 90th percentile's width is larger is due to the fact that there are more points to inform the prediction for the 50th percentile than the 90th so the fitted value is more likely to have higher accuracy for the 50th percentile.

The prediction intervals are wider than the confidence intervals because the CI is simply concerned with the variance of the mean response based on the data obtained while the PI attempts to find an interval that a new response will fall in given an unobserved data point. This interval is larger because we are uncertain about the actual mean response, so to predict a point that would fall outside of this mean creates a larger interval.

CODE

```
> cement=read.table('cement.txt',header=TRUE,sep='\t')
> lm_aluminum = lm(Hardness_cement~Aluminum, data=cement)
> quantiles = quantile(cement$Aluminum, probs=c(0.5,0.9))
> estimates = predict(lm_aluminum, data.frame(Aluminum=quantiles))
> confidence = predict(lm_aluminum, data.frame(Aluminum=quantiles), interval=
"confidence", level=.95)
> prediction = predict(lm_aluminum, data.frame(Aluminum=quantiles), interval=
"prediction", level=.95)
> confwidth = cbind(confidence[,3]-confidence[,2])
> colnames(confwidth) <- c("CI Width")
> predwidth = cbind(prediction[,3]-prediction[,2])
> colnames(predwidth) <- c("PI Width")
> summary(lm_aluminum)

Call:
lm(formula = Hardness_cement ~ Aluminum, data = cement)

Residuals:
    Min      1Q  Median      3Q     Max
-16.061  -9.048   1.339   7.883  15.614

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  81.4793     4.9273   16.54 4.07e-09 ***
Aluminum      1.8687     0.5264    3.55  0.00455 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.73 on 11 degrees of freedom
Multiple R-squared:  0.5339,   Adjusted R-squared:  0.4916
F-statistic:  12.6 on 1 and 11 DF,  p-value: 0.004552

> quantiles
50% 90%
  7  11
> estimates
      50%       90%
 94.56058 102.03557
> confidence
          fit      lwr      upr
50%  94.56058 87.99073 101.1304
90% 102.03557 94.30998 109.7612
> prediction
          fit      lwr      upr
50%  94.56058 70.05417 119.0670
90% 102.03557 77.19436 126.8768
> confwidth
    CI Width
50% 13.13970
90% 15.45117
> predwidth
    PI Width
50% 49.01281
90% 49.68242
```