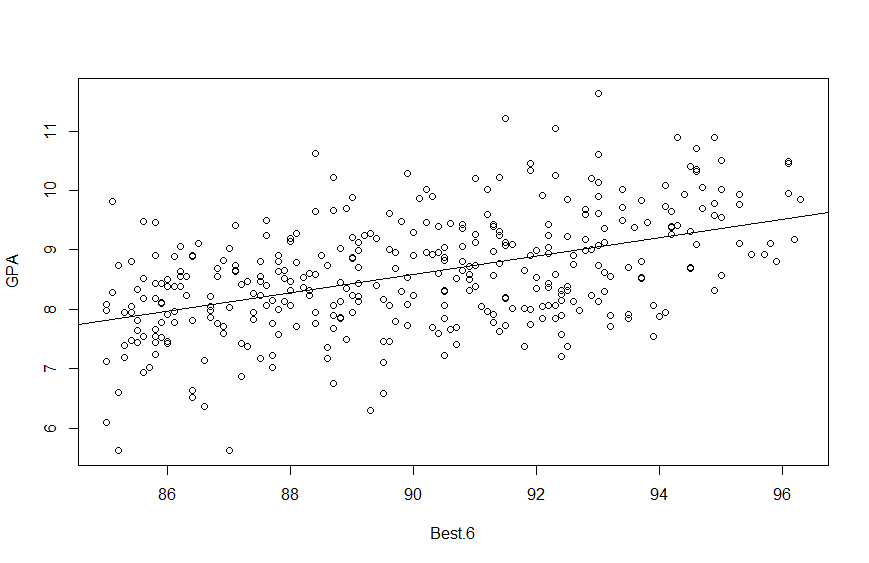
1.

(a) Is there a relationship between university grades and high school average using the best six OACs?



Residuals:

Min 1Q Median 3Q Max

-2.49671 -0.56908 0.02352 0.53000 2.58352

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -5.35498 1.31289 -4.079 5.57e-05 \*\*\*

Best.6 0.15496 0.01458 10.632 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

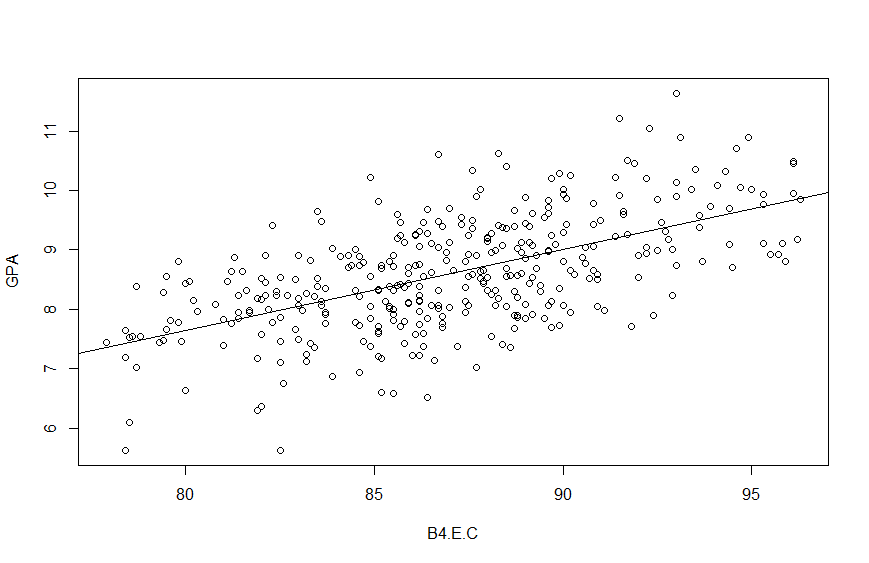
Residual standard error: 0.8295 on 361 degrees of freedom

Multiple R-squared: 0.2385, Adjusted R-squared: 0.2363

F-statistic: 113 on 1 and 361 DF, p-value: < 2.2e-16

藉由繪製散布圖以及Simple Linear Regression可以看得出來GPA與Best.6有線性關係。

(b)Is there a relationship between university grades and high school average using the best four OACs?



Residuals:

Min 1Q Median 3Q Max

-2.35057 -0.52668 0.02093 0.52157 2.22092

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.321975 0.854028 -3.89 0.000119 \*\*\*

B4.E.C 0.137001 0.009807 13.97 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7658 on 361 degrees of freedom

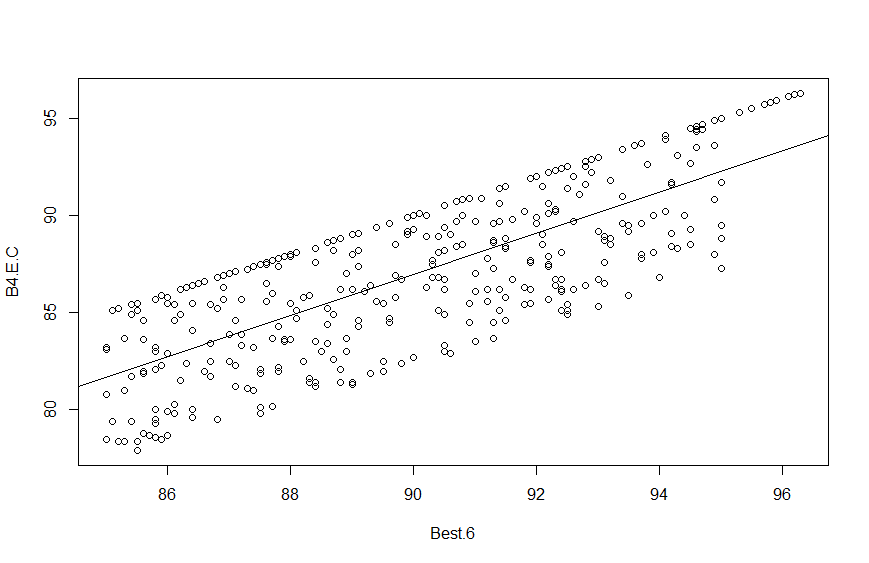
Multiple R-squared: 0.3509, Adjusted R-squared: 0.3491

F-statistic: 195.2 on 1 and 361 DF, p-value: < 2.2e-16

藉由繪製散布圖以及Simple Linear Regression可以看得出來GPA與Best.E.C有線性關係。

(c) Write a comment to the university’s academic vice president describing your statistical analysis and your recommendations.

由上述可以發現B6與B4.E.C與GPA的關係十分相似，也可以從下圖發現B6與B4.E.C存在著線性關係。單以B6作為dependent variable，以B4.E.C，R squared可以高達0.5948。因此，即使採取最高分的六個成績B6之中不包含English跟Calculus，還是可以多少代表著B4.E.C，藉由下面的模型預測出B4.E.C。



B4.E.C = -8.27950 + 1.05824\* B6

Call:

lm(formula = B4.E.C ~ Best.6, data = OACs)

Residuals:

Min 1Q Median 3Q Max

-4.9532 -2.4154 0.2506 2.7235 3.3233

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -8.27950 4.14097 -1.999 0.0463 \*

Best.6 1.05824 0.04597 23.020 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

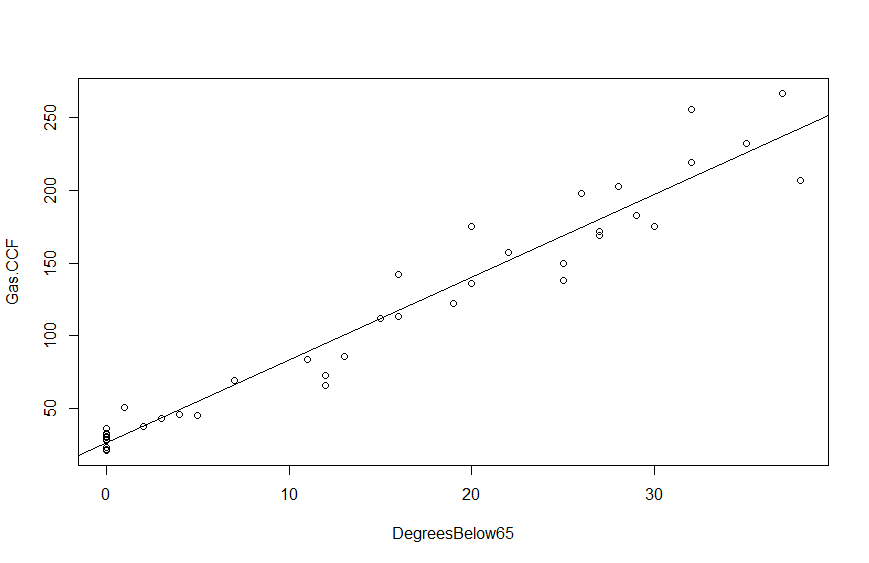
Residual standard error: 2.616 on 361 degrees of freedom

Multiple R-squared: 0.5948, Adjusted R-squared: 0.5937

F-statistic: 529.9 on 1 and 361 DF, p-value: < 2.2e-16

2.

(a) Use R to fit a simple linear model with the date “gas\_consumption.txt”. Do the analysis, make the plot, and summarize the results.



Residuals:

Min 1Q Median 3Q Max

-36.056 -6.404 0.580 5.460 47.102

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 26.7274 3.2199 8.301 1.07e-10 \*\*\*

DegreesBelow65 5.6928 0.1818 31.316 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.09 on 46 degrees of freedom

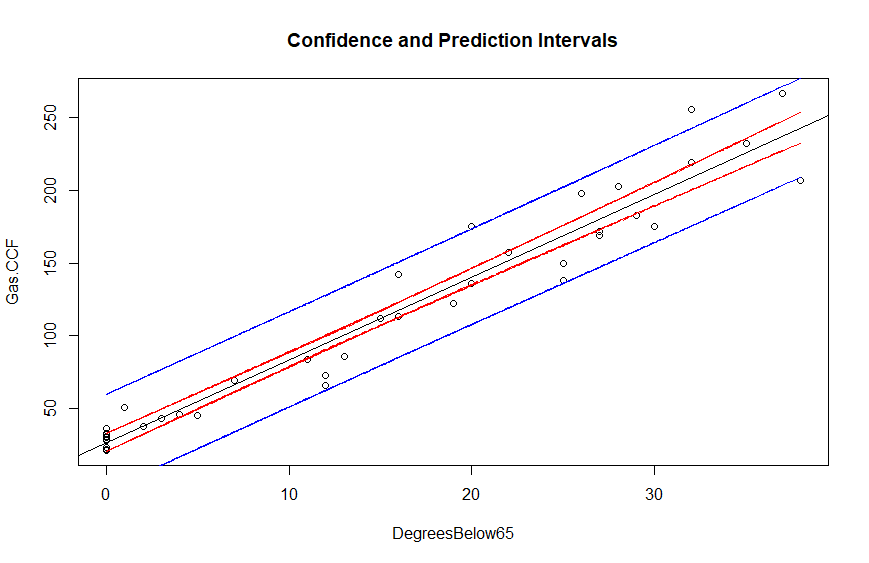
Multiple R-squared: 0.9552, Adjusted R-squared: 0.9542

F-statistic: 980.7 on 1 and 46 DF, p-value: < 2.2e-16

(Intercept)與DegreeBelow65的p-value都十分小，代表此截距與係數可信。而R-squared高達0.9552，幾乎所有dependent variable的變異都被解釋到了。

在DegreeBelow=0時，預測Gas.CCF=26.72，而DegreeBelow65每上升1單位，Gas.CCF就會上升5.69

(b) Modify the script provided for this lecture to create the “Confidence and Prediction Intervals” plot shown on lecture note p. 2-32 and re-printed below.



Code:

xy <- data.frame(X=DegreesBelow65)

yhat <- predict(fit, newdata=xy, interval="confidence")

ci <- data.frame(lower=yhat[,"lwr"], upper=yhat[,"upr"])

yhat.pi <- predict(fit, newdata=xy, interval="prediction")

pi <- data.frame(lower=yhat.pi[,"lwr"], upper=yhat.pi[,"upr"])

plot(DegreesBelow65,Gas.CCF, main ="Confidence and Prediction Intervals",

ylab = "Gas.CCF",

xlab = "DegreesBelow65")

abline(fit)

lines(xy$X, ci$lower, lty=2, col="red")

lines(xy$X, ci$upper, lty=2, col="red")

lines(xy$X, pi$lower, lty=2, col="blue")

lines(xy$X, pi$upper, lty=2, col="blue")

3. Find a linear model that appropriately reveals the relationship among these variables. Clearly state your model building process and interpret the results of your final model.

我先將Cars跟Speed分別對Accidents做SLD後，發現Cars這項的p-value與R-squared表現比較好，繼續加入Speed後，發現兩個independent variables的p-value都下降，且R-squared與Adjusted R-squared都上升，因此選定這項為final model。

其中Cars與Speed這兩項變數的p-value都>0.05，因此不能否定H0: Coefficient of Cars = 0跟H0: Coefficient of Speed = 0，其中可以利用這兩個變數解釋的變異只有約5.55%，而在Global model test中的F-statistic: 1.674, p-value: 0.1965，代表我們不行拒絕H0: Coefficient of Cars = Coefficient of Speed = 0

Call:

lm(formula = Accidents ~ Cars + Speed)

Residuals:

Min 1Q Median 3Q Max

-7.8859 -0.8428 -0.0385 1.4427 4.2786

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -12.8719 13.7901 -0.933 0.355

Cars 0.3733 0.2587 1.443 0.155

Speed 0.2699 0.2232 1.209 0.232

Residual standard error: 2.408 on 57 degrees of freedom

Multiple R-squared: 0.05548, Adjusted R-squared: 0.02234

F-statistic: 1.674 on 2 and 57 DF, p-value: 0.1965