Regresion Lineal

Mario H. Palomino Ludena

25/6/2021

### Importar los datos  
df = read.csv("DuffyIndustries\_RawData.csv")  
### Mostrar la cabecera de los datos  
head(df)

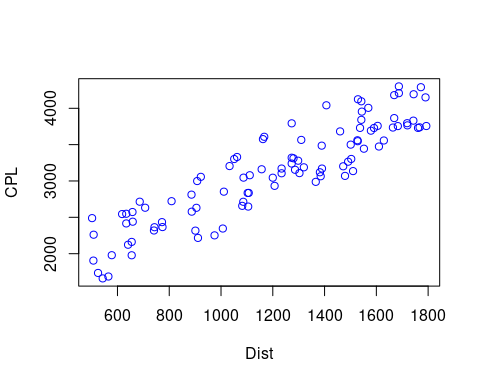
## ID CPL Dist LdTime TrlLng Wgt Equpt  
## 1 1 3692 1579 1 53 20559 DRY  
## 2 2 3279 1298 12 48 17025 REF  
## 3 3 3120 1382 11 48 26735 DRY  
## 4 4 3205 1033 1 53 26175 DRY  
## 5 5 3188 1320 3 53 17994 DRY  
## 6 6 2835 1103 9 53 32206 DRY

tail(df)

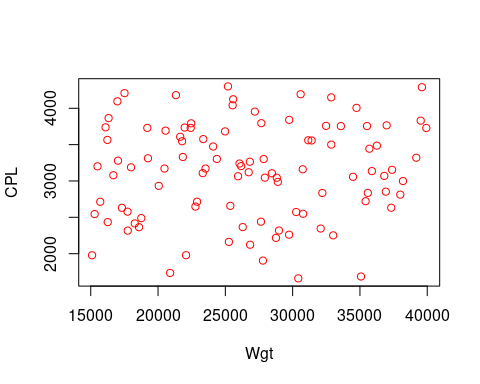
## ID CPL Dist LdTime TrlLng Wgt Equpt  
## 95 95 3608 1167 2 53 21646 REF  
## 96 96 3330 1062 0 53 21843 REF  
## 97 97 2489 502 1 48 18756 REF  
## 98 98 2722 809 2 48 35421 REF  
## 99 99 1978 655 3 48 15100 DRY  
## 100 100 4208 1687 3 53 17510 REF

# Análisis de Correlación

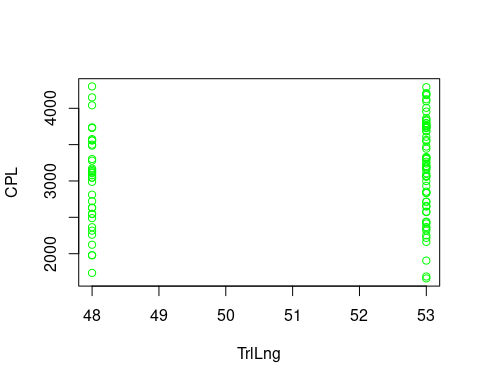
# Cómo se relacionan las variables?  
attach(df)  
plot(Dist,CPL, col= "blue")



plot(Wgt,CPL, col= "red")



plot(TrlLng,CPL, col = "green")



# Generar una matriz de correlación  
# No considerar las variables no numéricas  
cor(df[,c(-1,-7)])

## CPL Dist LdTime TrlLng Wgt  
## CPL 1.00000000 0.904213069 -0.093313675 0.141566808 0.075854669  
## Dist 0.90421307 1.000000000 0.004662807 0.174592785 0.119108525  
## LdTime -0.09331368 0.004662807 1.000000000 -0.040163096 0.066091539  
## TrlLng 0.14156681 0.174592785 -0.040163096 1.000000000 0.006411846  
## Wgt 0.07585467 0.119108525 0.066091539 0.006411846 1.000000000

# Especificación del modelo

fit = lm(CPL ~ Dist)

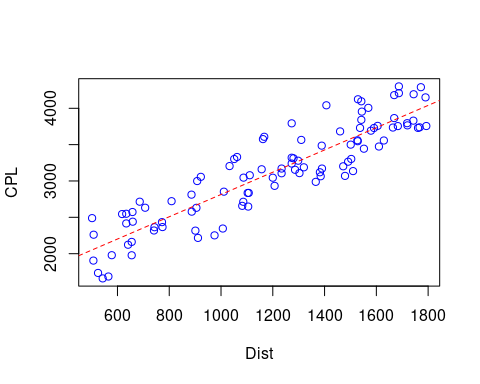
# Resultados del Modelo

summary(fit)

##   
## Call:  
## lm(formula = CPL ~ Dist)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -523.94 -217.05 -47.13 210.10 603.31   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.282e+03 9.261e+01 13.85 <2e-16 \*\*\*  
## Dist 1.532e+00 7.309e-02 20.96 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 281.3 on 98 degrees of freedom  
## Multiple R-squared: 0.8176, Adjusted R-squared: 0.8157   
## F-statistic: 439.3 on 1 and 98 DF, p-value: < 2.2e-16

# Representación gráfica del Modelo

plot(Dist,CPL, col = "blue")  
abline(fit, col = "red",lty = 2)



library(HH)

## Loading required package: lattice

## Loading required package: grid

## Loading required package: latticeExtra

## Loading required package: multcomp

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

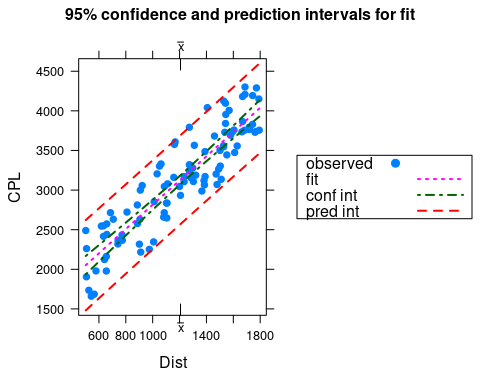
## Loading required package: MASS

##   
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':  
##   
## geyser

## Loading required package: gridExtra

ci.plot(fit)



# Comunicar el valor del intervalo de confianza, pero además el valor de predicción. El primero construye un intervalo sobre cuál sería el valor medio, el segundo sobre cuál sería el valor de la predicción.