##REGRESI LINEAR

* Fungsi untuk import data

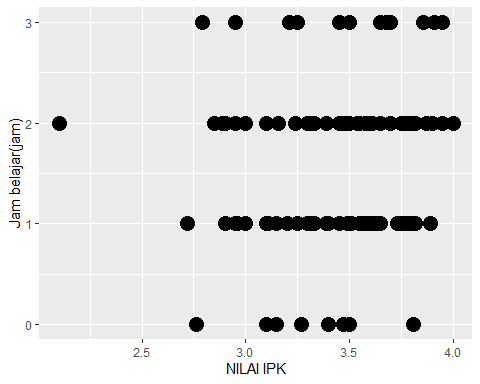
library(readxl)  
data <- read\_excel("C:/Users/Lenovo/Downloads/data.xlsx")  
data$IPK <- as.numeric(data$IPK)

* Fungsi untuk membuat scatterplot hubungan IPK dengan jam belajar

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.1.2

ggplot(data, aes(x=IPK, y=`Jam belajar`)) + geom\_point(size=5)+  
 labs(x="NILAI IPK", y="Jam belajar(jam)")



* Fungsi untuk mendeskripsikan variable baru antar data kategori

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.1.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

data = mutate(data, Jam\_belajar= case\_when(  
 data$`Jam belajar` <= 1 ~ "cepat ",  
 data$`Jam belajar` <= 2 ~ "sedang ",  
 data$`Jam belajar` >= 2 ~ "lama "  
),  
nilai = case\_when(  
 data$IPK <= 3 ~ "grade C ",  
 data$IPK <= 3.5 ~ "grade B ",  
 data$IPK >= 3.5 ~ "grade A "))

* Fungsi untuk melihat frekuensi data

table(data$Jam\_belajar)

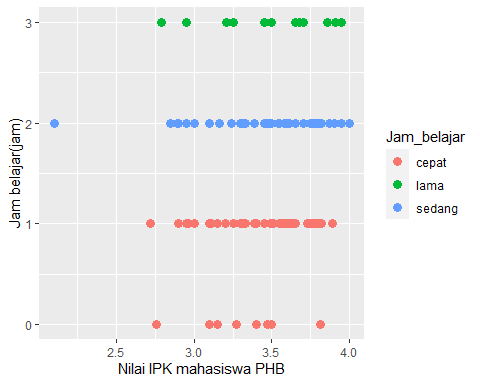
##   
## cepat lama sedang   
## 63 15 49

table(data$nilai)

##   
## grade A grade B grade C   
## 62 49 16

* Fungsi untuk membuat grafik plot(scatterplot)

ggplot(data, aes(x=IPK, y=`Jam belajar`, color = Jam\_belajar)) +  
 geom\_point(size=3)+  
 labs(x="Nilai IPK mahasiswa PHB", y="Jam belajar(jam)")



* Fungsi untuk membuat data yang baru dan menghapus data apabila data outlier

databaru = filter(data, `Jam belajar` < 3)  
databaru

## # A tibble: 112 x 13  
## Nama `Jenis kelamin` Usia Prodi `Kesesuaian prodi` Pembiayaan `Kota asal`  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Fabil 0 20 2 1 0 0  
## 2 Amel 1 19 2 0 1 0  
## 3 Widia 1 21 3 1 0 1  
## 4 Sinta 1 20 7 0 0 1  
## 5 Ami 1 20 2 1 0 0  
## 6 Taufik 0 19 1 0 0 1  
## 7 Ziky 0 22 1 1 0 0  
## 8 Silvia 1 20 1 1 2 1  
## 9 Gilang 0 20 4 1 2 1  
## 10 Putri 0 22 3 0 0 0  
## # ... with 102 more rows, and 6 more variables: Kos <dbl>, Jam belajar <dbl>,  
## # Kendaraan <dbl>, IPK <dbl>, Jam\_belajar <chr>, nilai <chr>

* Fungsi untuk mengetahui nilai korelasi antar variable jam belajar dan IPK

cor(databaru$IPK, databaru$`Jam belajar`)

## [1] 0.1212045

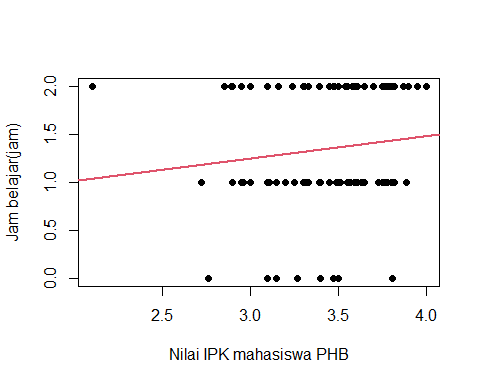
* Fungsi untuk mencari persamaan regresi

lm(`Jam belajar`~IPK, databaru)

##   
## Call:  
## lm(formula = `Jam belajar` ~ IPK, data = databaru)  
##   
## Coefficients:  
## (Intercept) IPK   
## 0.5493 0.2339

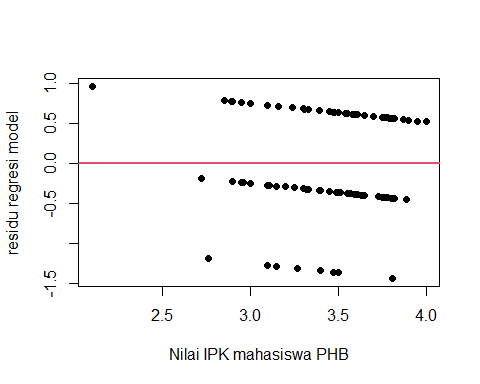
* Fungsi untuk membuat scatterplot dengan garis regresi dan korelasi

plot(databaru$IPK, databaru$`Jam belajar`, pch=16,  
 xlab = "Nilai IPK mahasiswa PHB",  
 ylab = "Jam belajar(jam)")  
abline(lm(databaru$`Jam belajar` ~ databaru$IPK), col=2, lwd=2)  
text(3.4, 40, paste("y = 0.5493+0.2339x"))



* Fungsi untuk membuat plot residu(tidak ada outlier) dengan garis linear

model = lm(`Jam belajar` ~ IPK, databaru)  
residu = resid(model)  
plot(databaru$IPK, residu, pch=16,  
 xlab = "Nilai IPK mahasiswa PHB",  
 ylab = "residu regresi model")  
abline(h=0, col=2, lwd=2)



* Fungsi untuk melihat nilai plot residu

summary(model)

##   
## Call:  
## lm(formula = `Jam belajar` ~ IPK, data = databaru)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.4406 -0.3921 -0.2816 0.6062 0.9594   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.5493 0.6336 0.867 0.388  
## IPK 0.2339 0.1827 1.281 0.203  
##   
## Residual standard error: 0.6258 on 110 degrees of freedom  
## Multiple R-squared: 0.01469, Adjusted R-squared: 0.005733   
## F-statistic: 1.64 on 1 and 110 DF, p-value: 0.203

* Fungsi untuk memprediksi data(co: hubungan jam belajar dengan IPK)

nilaibaru = data.frame(IPK=3.85)  
predict(model, nilaibaru)

## 1   
## 1.449991