

**TUGAS 3\_ *OUTLIER***

**MATA KULIAH MANAJEMEN DAN ANALISIS DATA DENGAN R**



**Isni Nurul Aini**

**NPM. 131520220003**

**PROGRAM STUDI MAGISTER EPIDEMIOLOGI**

**FAKULTAS KEDOKTERAN**

**UNIVERSITAS PADJADJARAN**

**2023**

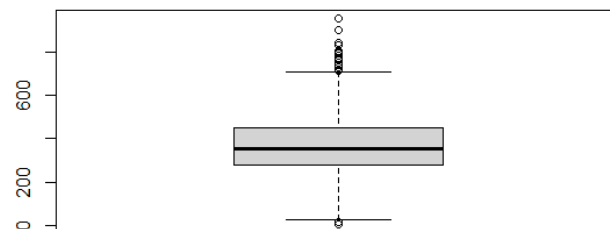
```
#reading data
```

```
pef_na_pef
```

## #1. Mengidentifikasi outlier berdasarkan kriteria dan visualisasi grafik boxplot

```
summary(pef_na_pef$pef)
```

```
boxplot(pef_na_pef$pef,col="aquamarine",main="PEF Score")
```



## #2. Menentukan cut off outlier (batas atas dan bawah) dari grafik boxplot tersebut

```
min(boxplot(pef_na_pef$pef, plot = FALSE)$out)
```

```
max(boxplot(pef_na_pef$pef, plot = FALSE)$out)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10.0	280.0	355.0	367.4	450.0	951.0

```
quartiles <- quantile(pef_na_pef$pef, probs=c(.25, .75), na.rm=FALSE)
```

```
quartiles
```

```
IQR <- IQR(pef_na_pef$pef)
```

```
IQR
```

```
Lower <- quartiles[1] - 1.5*IQR
```

```
Lower
```

```
Upper <- quartiles[2] + 1.5*IQR
```

```
Upper
```

### #3. Membuat dataset yang tidak berisi outlier sesuai cut off no 2

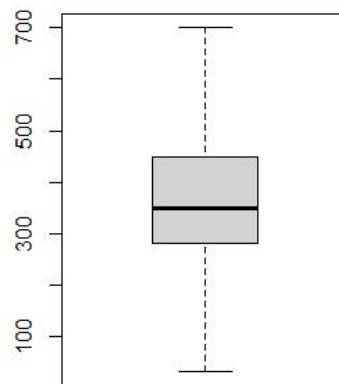
```
pef_na_pef_no <- subset(pef_na_pef, pef_na_pef$pef > Lower
```

```
    pef_na_pef $pef < Upper)
```

```
summary(pef_na_pef_no$pef)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
30.0	280.0	350.0	366.2	450.0	700.0

```
boxplot(pef_na_pef_no$pef)
```



### #4. Melakukan tes normalitas pada dataset dengan outlier dan tanpa outlier

#For Big Sample (Kolmogorov-Smirnov)

```
library(nortest)
```

```
lillie.test(pef_na_pef$pef)
```

```
Lilliefors (kolmogorov-smirnov) normality test  
data:  pef_final$pef  
D = 0.055561, p-value < 2.2e-16
```

```
lillie.test(pef_na_pef_no$pef)
```

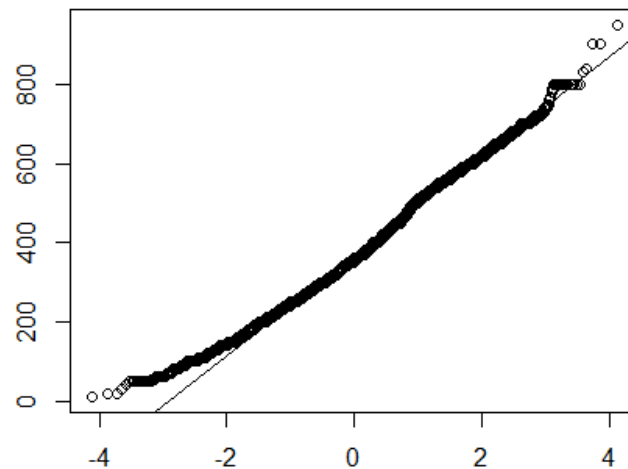
```
Lilliefors (kolmogorov-smirnov) normality test  
data:  pef_final_no$pef  
D = 0.053763, p-value < 2.2e-16
```

**#5. Membuat grafik QQ line untuk membandingkan visualisasi nilai pef pada dataset dengan outlier dan tanpa outlier**

```
qqnorm(pef_na_pef$pef); qqline(pef_na_pef$pef)
```

**(DENGAN OUTLIERS)**

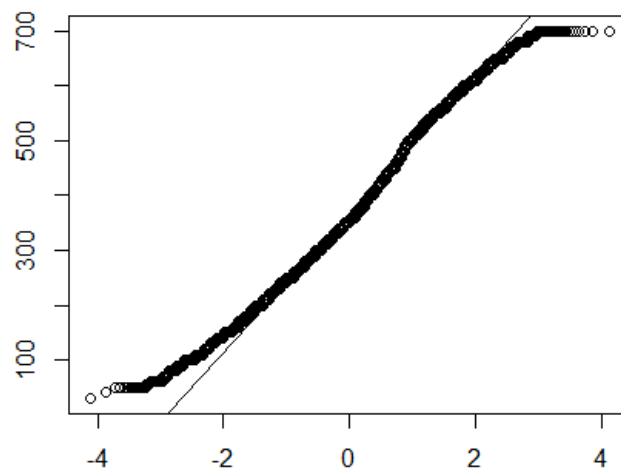
**Normal Q-Q Plot**



```
qqnorm(pef_na_pef_no$pef); qqline(pef_na_pef_no$pef)
```

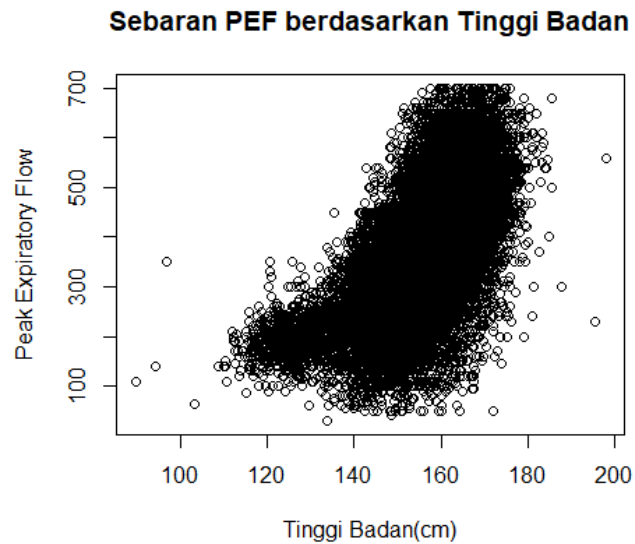
**(TANPA OUTLIERS)**

**Normal Q-Q Plot**

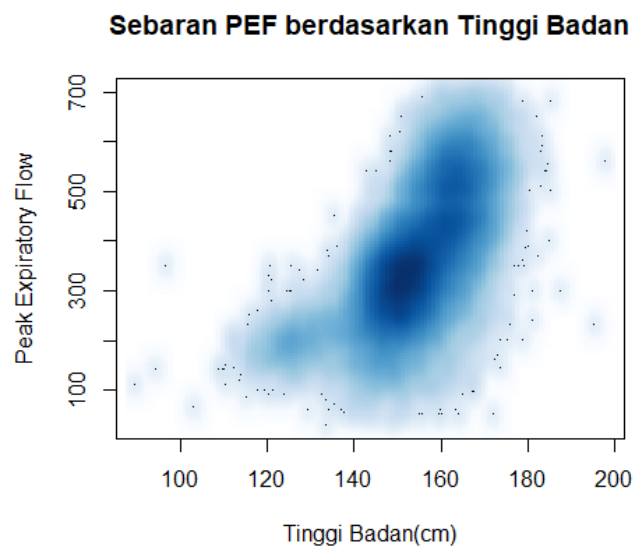


**#6. Membuat scatterplot yang memperlihatkan hubungan antara pef dengan height, dengan penambahan garis linear/regresi dan smoothed dengan loes (local regression smoothing)**

```
plot(pef_na_pef_no$pef~ pef_na_pef_no$height, xlab = "Tinggi Badan(cm)",  
     ylab="Peak Expiratory Flow", main="Sebaran PEF berdasarkan Tinggi Badan")
```

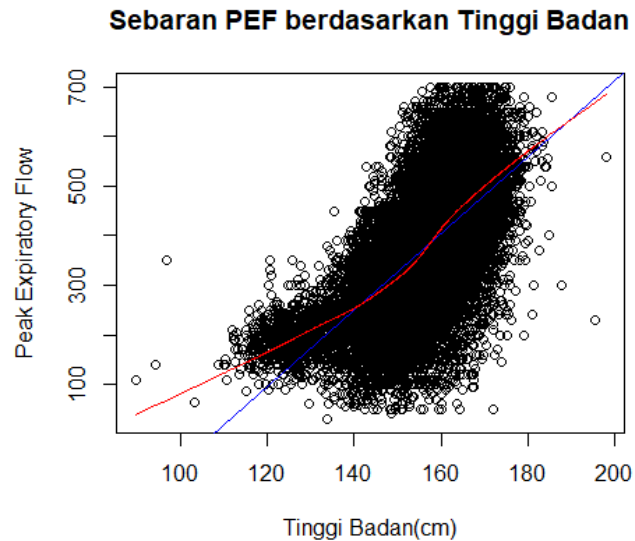


```
smoothScatter(pef_na_pef_no$pef~ pef_na_pef_no$height, xlab = "Tinggi Badan(cm)",  
              ylab="Peak Expiratory Flow", main="Sebaran PEF berdasarkan Tinggi Badan")
```



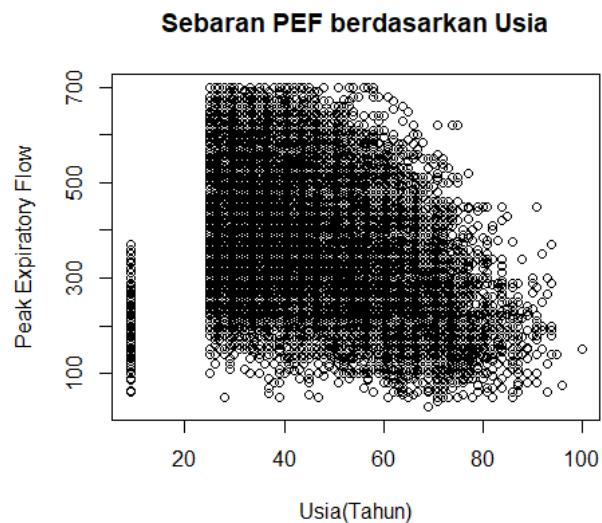
```
abline(lm(pef_na_pef_no$pef~ pef_na_pef_no$height, data = pef_na_pef_no), col = "blue")
```

```
lines(lowess(pef_na_pef_no$height, pef_na_pef_no$pef), col = "red")
```

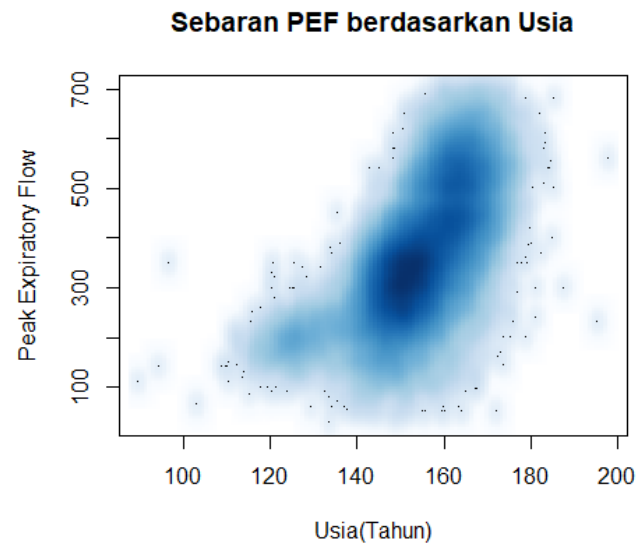


**#7. Membuat scatterplot yang memperlihatkan hubungan antara pef dengan umur, dengan penambahan garis linear/regresi dan *smoothed dengan loes (local regression smoothing)***

```
plot(pef_na_pef_no$pef~ pef_na_pef_no$age, xlab = "Usia(Tahun)",  
      ylab="Peak Expiratory Flow", main="Sebaran PEF berdasarkan Usia")
```



```
smoothScatter(pef_na_pef_no$pef~ pef_na_pef_no$height, xlab = "Usia(Tahun)",
  ylab="Peak Expiratory Flow", main="Sebaran PEF berdasarkan Usia")
```



```
abline(lm(pef_na_pef_no$pef~ pef_na_pef_no$age, data = pef_na_pef_no), col = "blue")
lines(lowess(pef_na_pef_no$age, pef_na_pef_no$pef), col = "red")
```

