

Nama : Sri Hartati

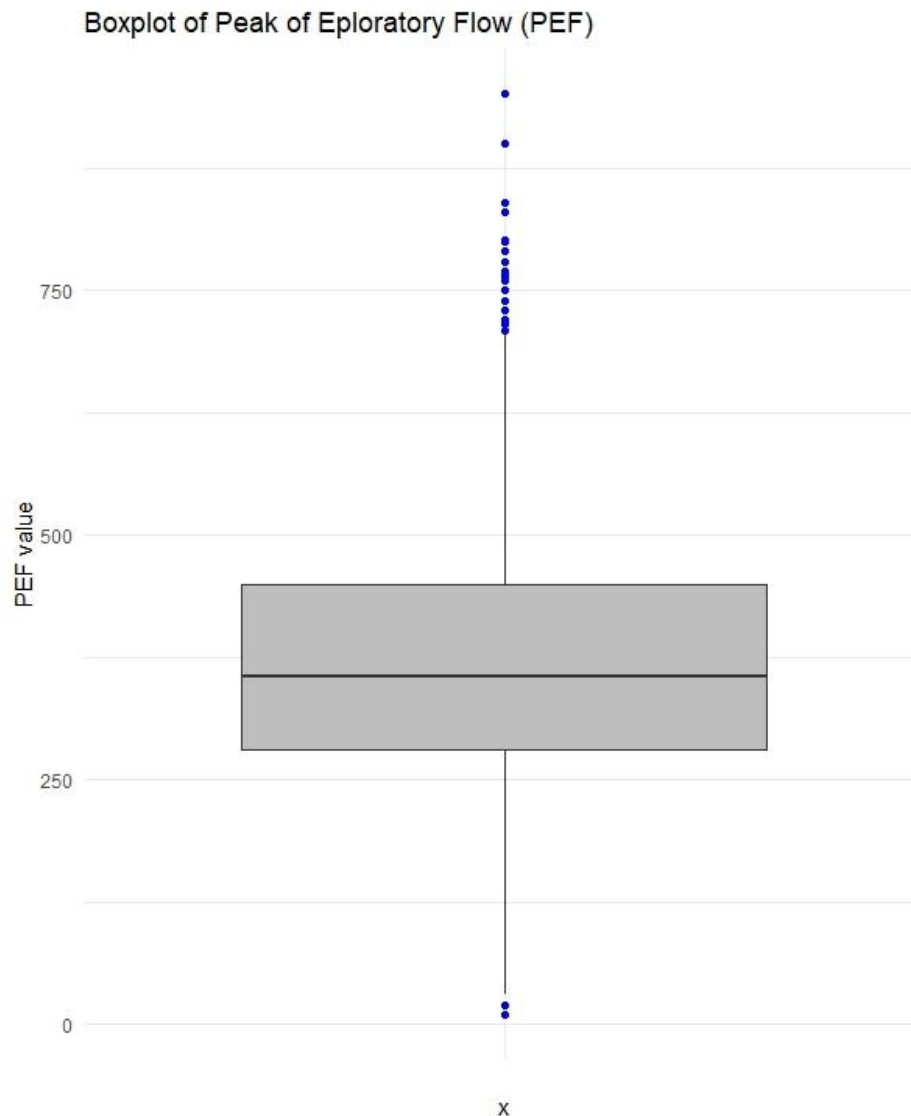
NPM : 131520220006

Mata Kuliah : Manajemen dan Analisis Data dengan R

Prodi : Magister Epidemiologi

Universitas Padjadjaran 2023

1. Mengidentifikasi outlier berdasarkan kriteria dan visualisasi boxplot.



Titik berwarna biru adalah outlier. Terdapat outlier di bawah nilai minimum dan di atas nilai maksimum.

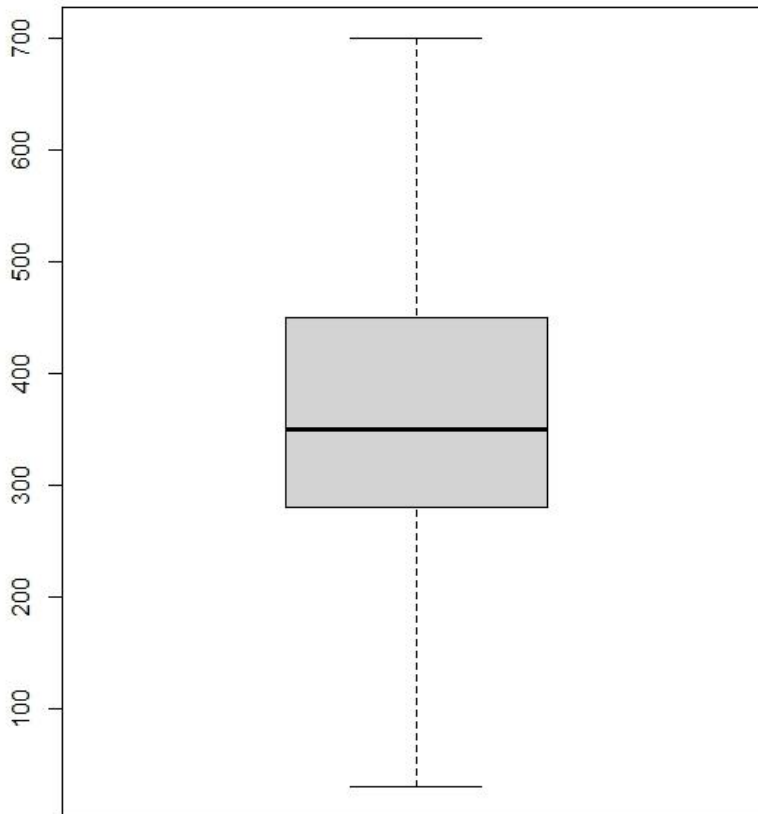
2. Menentukan cutoff dari outlier (batas atas dan batas bawah).

Dari perhitungan di program RStudio, didapatkan nilai sebagai berikut:

$Q1 = 280$ $Q3 = 450$ $IQR = 170$

- Kriteria batas atas outlier jika nilai $PEF > Q3 + (1.5 * IQR)$, yaitu
nilai $PEF > 450 + (1.5 * 170)$
nilai $PEF > 705$
- Kriteria batas bawah outlier jika nilai $PEF < Q1 - (1.5 * IQR)$, yaitu
nilai $PEF < 280 - (1.5 * 170)$
nilai $PEF < 25$

3. Membuat dataset yang tidak berisi outlier sesuai cutoff no 2.
Boxplot dari dataset yang tidak berisi outlier.



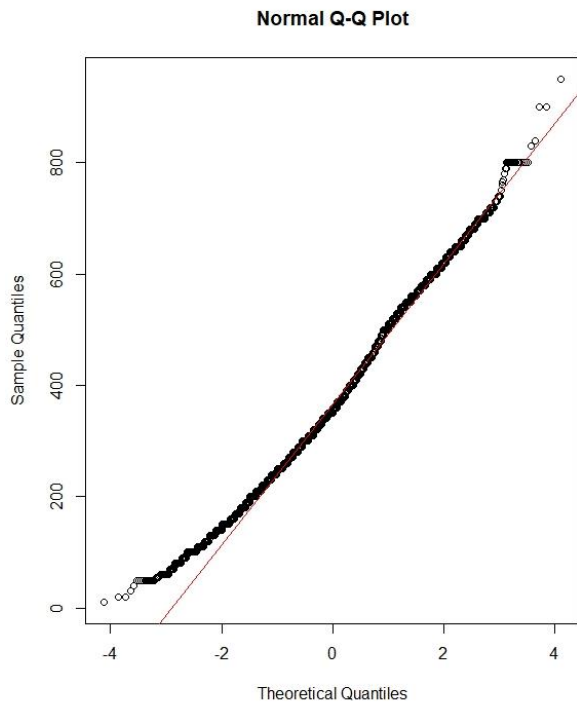
4. Melakukan tes normalitas pada dataset dengan outlier dan tanpa outlier

Karena jumlah sampel yang besar maka digunakan fungsi `lillie.test` untuk menguji normalitas, dengan *null hypothesis* nya adalah data terdistribusi normal.

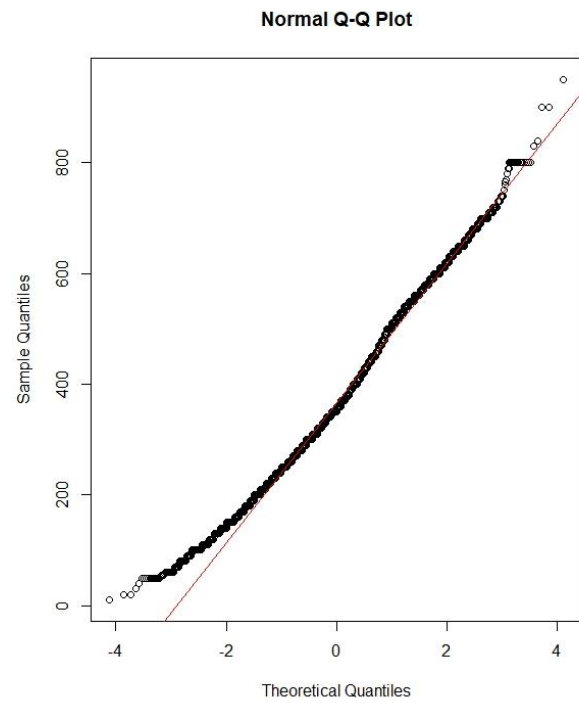
Hasil yang didapat dari tes normalitas menunjukkan bahwa:

- Pada dataset dengan outlier didapatkan p-value $< 2.2e-16$ berarti terdapat cukup bukti untuk menolak *null hypothesis*. Maka data tidak terdistribusi normal.
- Pada dataset tanpa outlier didapatkan p-value $< 2.2e-16$ berarti terdapat cukup bukti untuk menolak *null hypothesis*. Maka data tidak terdistribusi normal.

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

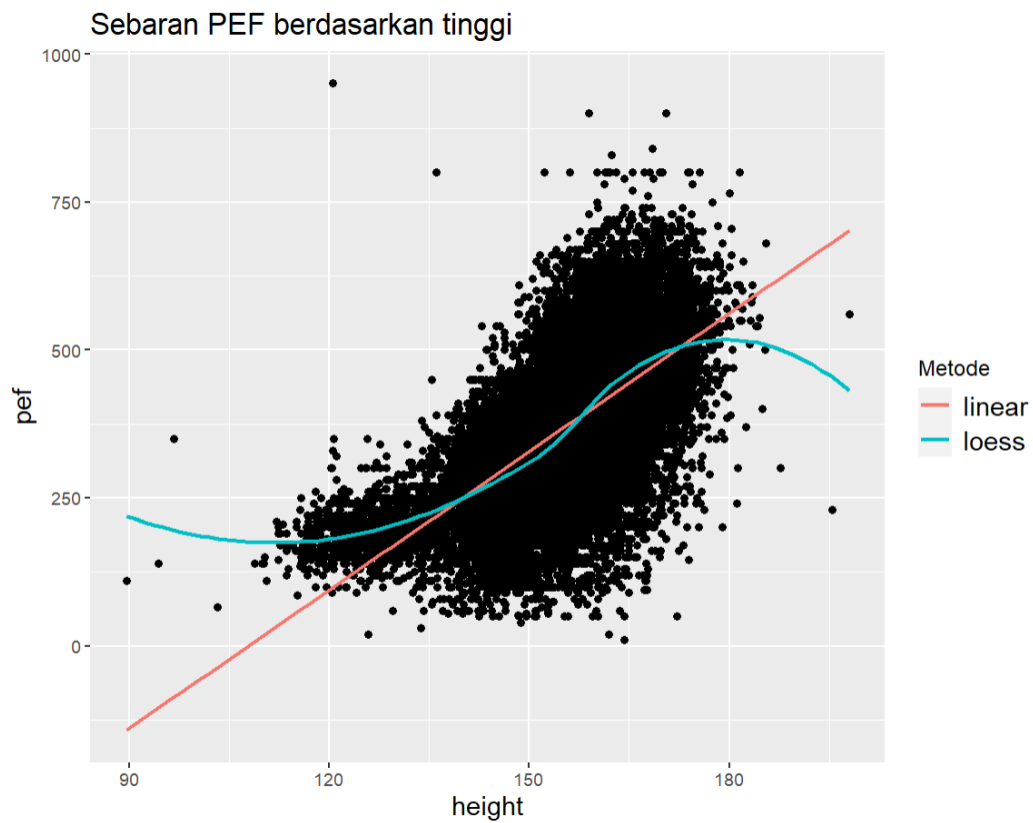


QQ plot pada dataset dengan outlier

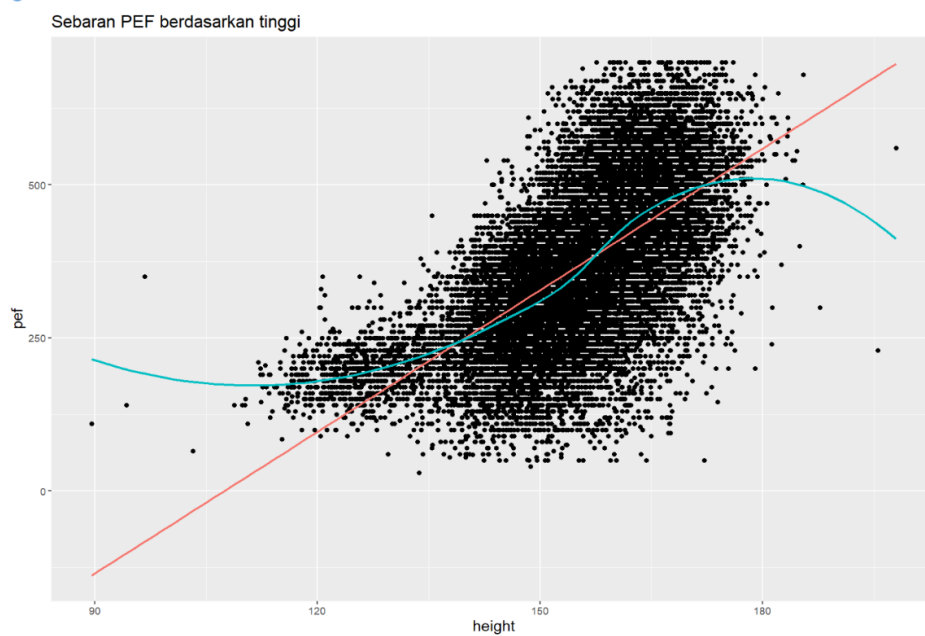


QQ plot pada dataset tanpa outlier

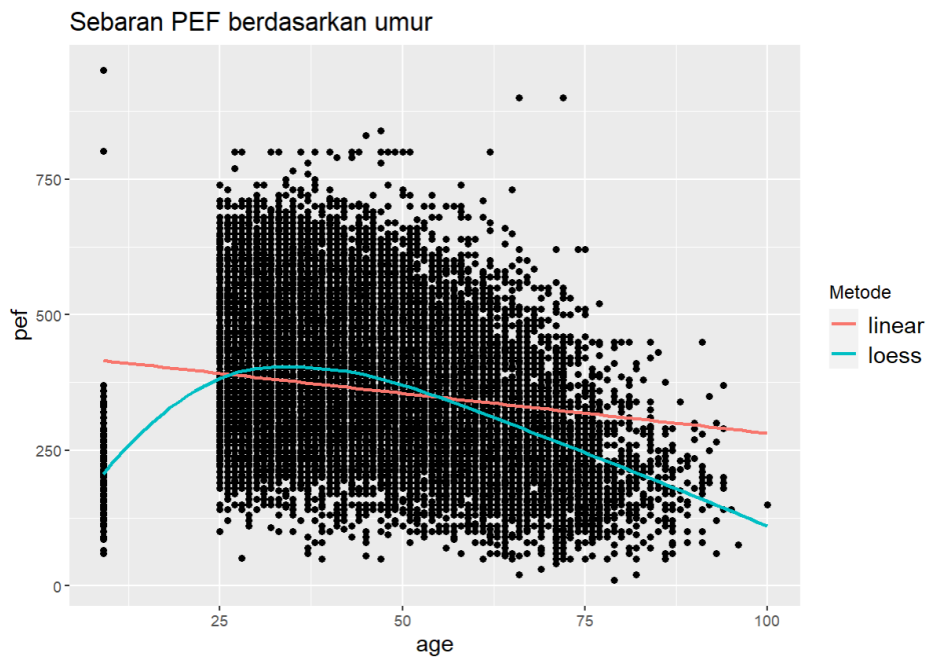
6. Membuat scatterplot yang memperlihatkan hubungan antara PEF dengan height, dengan penambahan garis linear/regresi dan smoothed dengan loes (local regression smoothing).
- a. Pada dataset dengan outlier



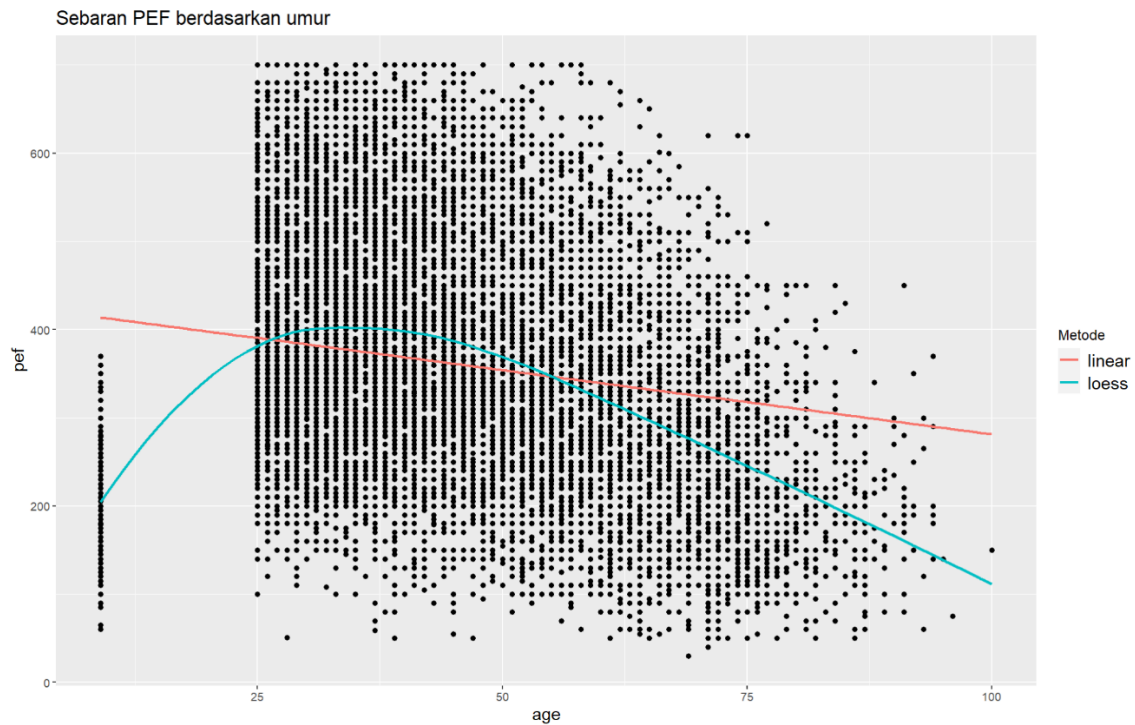
- b. Pada dataset tanpa outlier



7. Membuat scatterplot yang memperlihatkan hubungan antara PEF dengan umur, dengan penambahan garis linear/regresi dan smoothed dengan loes (local regression smoothing).
- a. Pada dataset dengan outlier



- b. Pada dataset tanpa outlier



R Script:

```
library(readxl)
library(readr)
library(ggplot2)
library(dplyr)
tugas_reglin <- w5_pef_lj_age_sex_height_pef
```

#mengetahui summary dari variabel pef

```
View(tugas_reglin)
summary(tugas_reglin$pef)
```

#nomor 1

#mengidentifikasi outlier variabel PEF berdasarkan visualisasi grafik boxplot

#website rujukan <https://statsandr.com/blog/outliers-detection-in-r/>

```
boxplot(tugas_reglin$pef)
ggplot(tugas_reglin) +
  aes(x= "", y = pef) +
  geom_boxplot(fill = "grey", outlier.color="blue") +
  labs(title="Boxplot of Peak of Exploratory Flow (PEF)", y="PEF value") +
  theme_minimal()
```

#mengeksraksi nilai outlier berdasarkan kriteria iqr

```
boxplot.stats(tugas_reglin$pef)$out
```

#mengetahui baris mana yang mengandung nilai outlier

```
out <- boxplot.stats(tugas_reglin$pef)$out
out_ind <- which(tugas_reglin$pef %in% c(out))
out_ind
```

#nomor 2

#menentukan batas atas dan batas bawah outlier

```
quartiles <- quantile(tugas_reglin$pef, probs=c(.25, .75), na.rm = FALSE)
IQR <- IQR(tugas_reglin$pef)
Lower <- quartiles[1] - 1.5*IQR
Upper <- quartiles[2] + 1.5*IQR
#kriteria batas atas outlier jika nilai pef > q3 + 1.5 iqr yaitu >705
#kriteria batas bawah outlier jika nilai pef < q1 - 1.5 iqr yaitu <25
```

#nomor 3

#membuat dataset yang tidak berisi outlier

```
data_nooutlier <- subset(tugas_reglin, tugas_reglin$pef > Lower & tugas_reglin$pef < Upper)
dim(data_nooutlier)
```

#memeriksa apakah di dataset yang baru terdapat outlier atau tidak

```
#menggunakan boxplot
boxplot(data_nooutlier$pef)
#menggunakan ggplot
ggplot(data_nooutlier) +
  aes(x= "", y = pef) +
  geom_boxplot(fill = "grey", outlier.color="blue") +
  labs(title="Boxplot of Peak of Eploratory Flow (PEF)", y="PEF value") +
  theme_minimal()
```

#nomor 4

```
#melakukan tes normalitas pada dataset dengan outlier dan tanpa outlier
#karena jumlah sampel nya besar jadi digunakan lillie.test
library(nortest)
lillie.test(tugas_reglin$pef)
lillie.test(data_nooutlier$pef)
```

#nomor 5

```
#membuat grafik QQ line untuk membandingkan dataset dengan outlier dan tanpa outlier
#dataset dengan outlier
qqnorm(tugas_reglin$pef)
qqline(tugas_reglin$pef, col="red")
```

#dataset tanpa outlier

```
qqnorm(data_nooutlier$pef)
qqline(data_nooutlier$pef, col="red")
```

#nomor 6

```
#membuat scatterplot antara pef dengan height
#dengan penambahan garis regresi linear dan smoothed dengan loess
plot(data_nooutlier$pef~data_nooutlier$height, xlab = "Tinggi (cm)", ylab = "Peak Expiratory Flow",
  main = "Sebaran PEF berdasarkan tinggi")
ggplot(data_nooutlier, aes(height, pef)) +
  geom_point() +
  geom_smooth(aes(color= "linear"), method = "lm", se = FALSE) +
  geom_smooth(aes(color= "loess"), method = "loess", se = FALSE) +
  ggtitle("Sebaran PEF berdasarkan tinggi") +
  labs(color= "Metode") +
  theme(plot.title = element_text(size = 15)) +
  theme(legend.text = element_text(size = 13)) +
  theme(axis.title.y = element_text(size = 13)) +
  theme(axis.title.x = element_text(size = 13))
```

#nomor 7


```
#membuat scatterplot antara pef dengan umur
#dengan penambahan garis regresi linear dan smoothed dengan loes
plot(data_nooutlier$pef~data_nooutlier$age, xlab = "Umur (tahun)", ylab = "Peak Expiratory Flow",
      main = "Sebaran PEF berdasarkan umur")
ggplot(data_nooutlier, aes(age, pef)) +
  geom_point() +
  geom_smooth(aes(color= "linear"), method = "lm", se = FALSE) +
  geom_smooth(aes(color= "loess"), method = "loess", se = FALSE) +
  ggtitle("Sebaran PEF berdasarkan umur") +
  labs(color= "Metode") +
  theme(plot.title = element_text(size = 15)) +
  theme(legend.text = element_text(size = 13)) +
  theme(axis.title.y = element_text(size = 13)) +
  theme(axis.title.x = element_text(size = 13))
```