

Tugas Mata Kuliah Manajemen dan Analisis Data dengan R

Identifikasi Outlier, Uji Normalitas dan Pembuatan Scatterplot

Dibuat untuk memenuhi Tugas Mata Kuliah Manajemen dan Analisis Data dengan R

Oleh:

Nurul Hidayah Chairunnisa

NPM: 131520220001



Program Studi Magister Epidemiologi

Fakultas Kedokteran

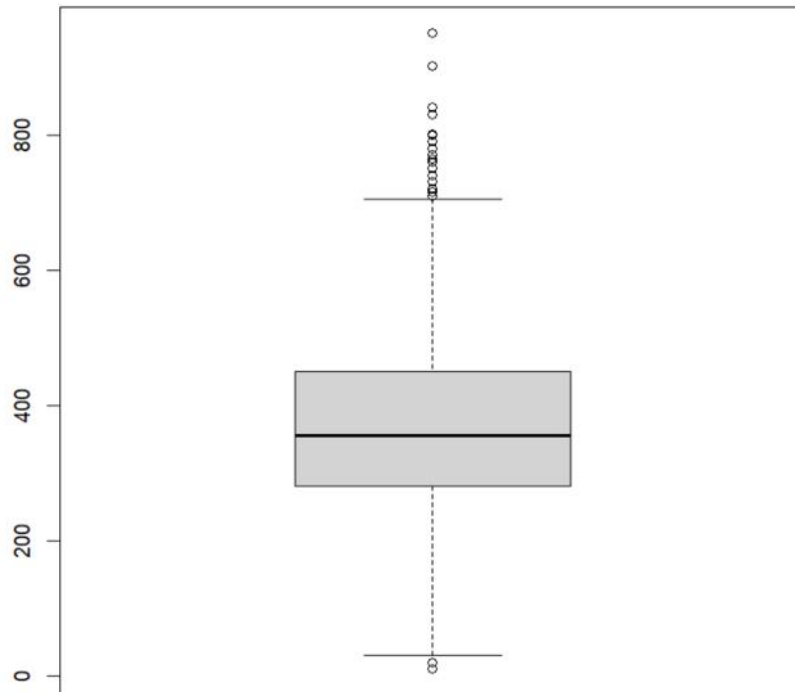
Universitas Padjadjaran

2023

1. Mengidentifikasi outlier berdasarkan kriteria dan visualisasi grafik boxplot.

```
boxplot(w5peflj_ashp$pef)
```

Hasil:



2. Menentukan nilai cutoff outlier (batas atas dan bawah) dari grafik boxplot tersebut.

```
boxplot.stats(w5peflj_ashp$pef)$out
```

```
quartiles <- quantile(w5peflj_ashp$pef, probs=c(.25, .75), na.rm = FALSE)
```

```
IQR <- IQR(w5peflj_ashp$pef)
```

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

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

Hasil:

```
> boxplot.stats(w5peflj_ashp$pef)$out
[1] 951 801 750 720 710 800 740 800 730 800 710 901 800 740 800 710 20 10 830 720 730 710 710 730 790
[26] 901 720 800 720 800 800 800 800 790 730 730 710 750 710 710 710 720 800 780 800 730 730 720 20
[51] 740 720 715 710 840 765 710 710 800 780 770 710 710 720 740 720 800 720 760 720 710 720 710 720 800
[76] 740 740 740 800 720 710
```

quartiles	Named num [1:2] 280 450
IQR	170
Lower	Named num 25
Upper	Named num 705

3. Membuat dataset yang tidak berisi outlier sesuai cutoff nomor 2

```
data_no <- subset(w5peflj_ashp$pef, w5peflj_ashp$pef > Lower & w5peflj_ashp$pef < Upper)
```

Hasil:

values										
data_no	num	[1:26232]	210	160	230	205	170	120	200	210 ...

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

a. Dataset dengan outliers

```
lillie.test(w5peflj_ashp$pef)
```

Hasil:

```
> lillie.test(w5peflj_ashp$pef)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  w5peflj_ashp$pef
D = 0.055561, p-value < 2.2e-16
```

b. Dataset tanpa outliers

```
lillie.test(data_no)
```

Hasil:

```
> lillie.test(data_no)

      Lilliefors (Kolmogorov-Smirnov) normality test

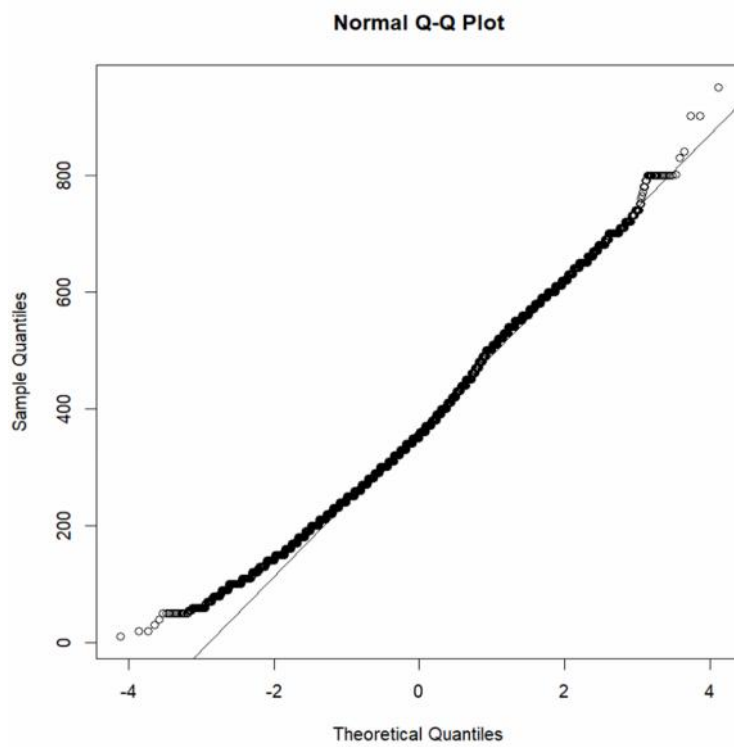
data:  data_no
D = 0.053763, p-value < 2.2e-16
```

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

a. Dataset dengan outlier

```
qqnorm(w5peflj_ashp$pef); qqline(w5peflj_ashp$pef)
```

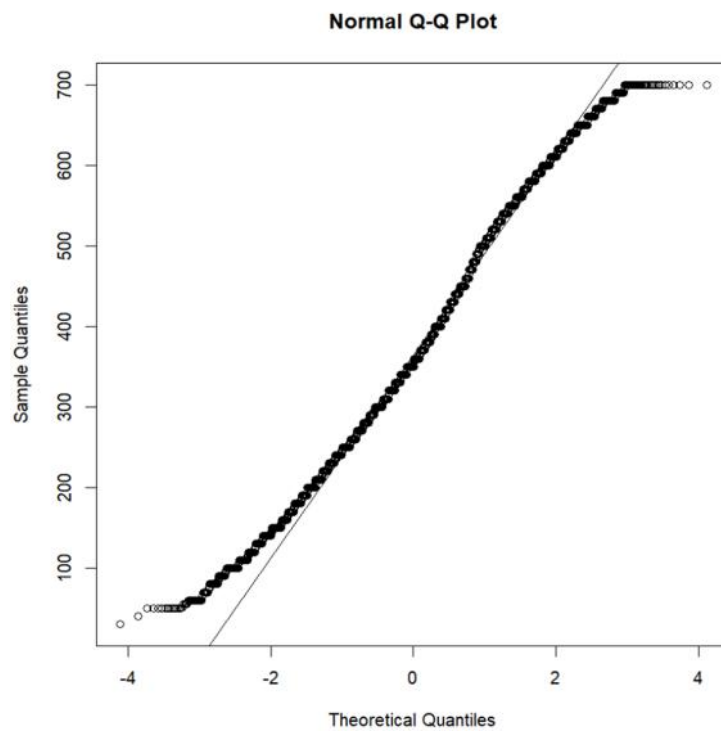
Hasil:



b. Dataset tanpa outlier

`qqnorm(data_no); qqline(data_no)`

Hasil:



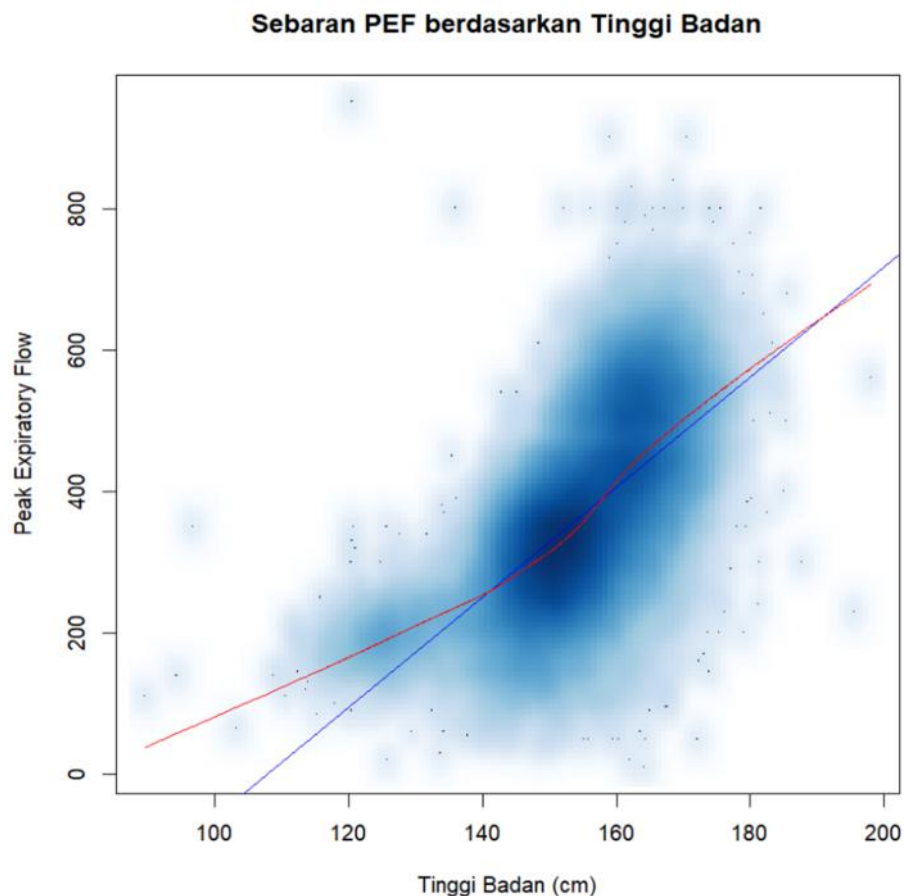
6. Membuat scatterplot yang memperlihatkan hubungan antara pef dengan tinggi badan, dengan penambahan garis linear/regresi dan smoothed dengan loes (*local regression smoothing*)

```
smoothScatter(w5peflj_ashp$pef~w5peflj_ashp$height, xlab = "Tinggi Badan (cm)",  
              ylab = "Peak Expiratory Flow",  
              main = "Sebaran PEF berdasarkan Tinggi Badan")
```

```
abline(lm(w5peflj_ashp$pef~w5peflj_ashp$height, data = w5peflj_ashp), col = "blue")
```

```
lines(lowess(w5peflj_ashp$height, w5peflj_ashp$pef), col = "red")
```

Hasil:



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

```
smoothScatter(w5peflj_ashp$pef~w5peflj_ashp$age, xlab = "Usia (Tahun)",  
              ylab = "Peak Expiratory Flow",  
              main = "Sebaran PEF berdasarkan Usia")
```

```
abline(lm(w5peflj_ashp$pef~w5peflj_ashp$age, data = w5peflj_ashp), col = "blue")
```

```
lines(lowess(w5peflj_ashp$age, w5peflj_ashp$pef), col = "red")
```

Hasil:

