

# Komputasi Statistika

## Statistika Deskriptif

---



# TOPIK



**Ukuran Pemusatan**



**Ukuran Penyebaran**



**Korelasi**

# Ukuran Pemusatan

## Mean

- Mean merangkum semua informasi dalam data.
- Mean adalah satu titik yang dapat dilihat sebagai titik pusat massa data.

**Sampel :** 
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

**Populasi:** 
$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

$n$  adalah ukuran sampel,  
 $N$  adalah ukuran populasi

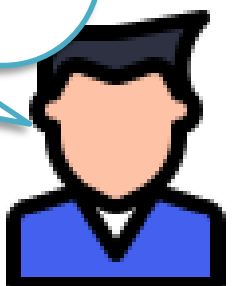
## Median

- Median adalah **nilai tengah** data dalam artian separuh data berada di bawahnya dan separuh di atasnya. Median dapat dinotasikan  $Q_2$ .
- Menghitung persentil sampel:
  1. Data diurutkan dari nilai terkecil hingga terbesar
  2. Menghitung perkalian (*ukuran sampel*)  $\times$  (*proporsi*) =  $np$ 
    - Jika  $np$  bukan bilangan integer, maka bulatkan ke atas dan temukan nilai pada urutan tersebut.
    - Jika  $np$  integer, disebut  $k$ , hitung rata-rata data ke- $k$  dan  $(k+1)$

## Modus

- Modus kumpulan data adalah nilai yang paling **sering muncul**.
- Nilai yang paling sering muncul terlihat dari frekuensi data.

When did you choose to use one of them?  
What are the characteristics of each of them?



# Ukuran Penyebaran

## Variansi

- Deviasi kuadrat rata-rata dari titik data dari *mean-nya*.

**Populasi:** 
$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

**Sampel :** 
$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

## Deviasi Standar

- Akar kuadrat dari variansi

**Populasi:** 
$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

**Sampel :** 
$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

## Range

$$R = \text{Maksimum} - \text{Minimum}$$

## Interquartile Range

$$IQR = Q_3 - Q_1$$

- $Q_3$  (persentil ke-75) adalah **75% pengamatan sampel**  $\leq$  nilai kuartil ketiga
- $Q_1$  (persentil ke-25) adalah **25% pengamatan sampel**  $\leq$  nilai kuartil kesatu

# Dataset

```
data("mtcars") #memanggil data di R
?mtcars #mengetahui tentang data mtcars
head(mtcars) #mengecek 6 observasi awal
tail(mtcars) #mengecek 6 observasi terakhir
str(mtcars) #mengetahui struktur data
```

## Motor Trend Car Road Tests

### Description

The data was extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

### Usage

`mtcars`

### Format

A data frame with 32 observations on 11 (numeric) variables.

- [, 1] mpg Miles/(US) gallon
- [, 2] cyl Number of cylinders
- [, 3] disp Displacement (cu.in.)
- [, 4] hp Gross horsepower
- [, 5] drat Rear axle ratio
- [, 6] wt Weight (1000 lbs)
- [, 7] qsec 1/4 mile time
- [, 8] vs Engine (0 = V-shaped, 1 = straight)
- [, 9] am Transmission (0 = automatic, 1 = manual)
- [, 10] gear Number of forward gears
- [, 11] carb Number of carburetors

# Menghitung Mean

```
attach(mtcars) #untuk selalu menyertakan seluruh dataset pada tiap command
```

```
sum(wt)/nrow(mtcars) #menghitung manual nilai mean  
mean(wt) #menghitung mean dengan fungsi
```

```
#jika tanpa fungsi attach  
sum(mtcars$wt)/nrow(mtcars) #menghitung manual nilai mean  
mean(mtcars$wt) #menghitung mean dengan fungsi
```

```
getmean <- function(x) {  
  n <- length(x)  
  xbar <- 0  
  for (i in c(1:n))  
  {  
    xbar <- xbar + ((x[i]/n))  
  }  
  xbar  
}  
getmean(wt)
```

**Membuat Fungsi  
Sendiri**

# Menghitung Median

```
df <- sort(mtcars$wt) #mengurutkan data 1 variabel secara parsial
head(df) #melihat 6 observasi awal
n <- length(df) #menghitung banyaknya observasi
p <- 0.5 #persentil 50%
np <- n*p

#karena hasil np integer, sehingga median ada di observasi ke-16 dan 17
x16 <- df[16] #mengambil data ke-16
x17 <- df[17] #mengambil data ke-17
median <- (x16+x17)/2 #menghitung median dengan manual
median

median <- median(wt) #menghitung median dengan fungsi median()
median

median <- quantile(wt,0.5) #menghitung median dengan fungsi quantile()
median
```

# Menghitung Modus

```
table(cyl) #tabel frekuensi setiap level variabel cyl  
l=length(levels(factor(wt))) #mencari banyak level  
modus <- names(sort(table(cyl))[1]) #mengurutkan frekuensi, mencari level terakhir  
modus
```

```
getmode <- function(x) {  
  uniqx <- unique(x)  
  uniqx[which.max(tabulate(match(x, uniqx)))]  
}  
mode <- getmode(cyl)  
print(mode)
```

**Membuat Fungsi  
Sendiri**



# Menghitung Ukuran Penyebaran

```
variansi <- var(wt)
variansi

dev_standar <- sqrt(variansi)
dev_standar

range <- max(wt)-min(wt)
range
range <- range(wt)[2]-range(wt)[1]
range

iqr <- IQR(wt)
iqr
```



# Fungsi summary() di R

```
attach(mtcars)

allstdes <- summary(mtcars)
wtcyl_stdes <- allstdes[,c(2,6)] #mengambil cyl di kolom 2 dan wt di kolom 6
wtcyl_stdes
```

```
> wtcyl_stdes
```

cyl	wt
Min. :4.000	Min. :1.513
1st Qu.:4.000	1st Qu.:2.581
Median :6.000	Median :3.325
Mean :6.188	Mean :3.217
3rd Qu.:8.000	3rd Qu.:3.610
Max. :8.000	Max. :5.424

OUTPUT

# Fungsi stat.desc() dari Paket pastecs di R

```
install.packages("pastecs") #install package dulu jika belum ter-install di R
library(pastecs) #memanggil library pastecs
stat.desc(mtcars)
stat.desc(mtcars[,c(2,6)]) #mengambil cyl di kolom 2 dan wt di kolom 6
```

```
> stat.desc(mtcars[,c(2,6)])
```

	cyl	wt
nbr.val	32.0000000	32.0000000
nbr.null	0.0000000	0.0000000
nbr.na	0.0000000	0.0000000
min	4.0000000	1.5130000
max	8.0000000	5.4240000
range	4.0000000	3.9110000
sum	198.0000000	102.9520000
median	6.0000000	3.3250000
mean	6.1875000	3.2172500
SE.mean	0.3157093	0.1729685
CI.mean.0.95	0.6438934	0.3527715
var	3.1895161	0.9573790
std.dev	1.7859216	0.9784574
coef.var	0.2886338	0.3041285

OUTPUT

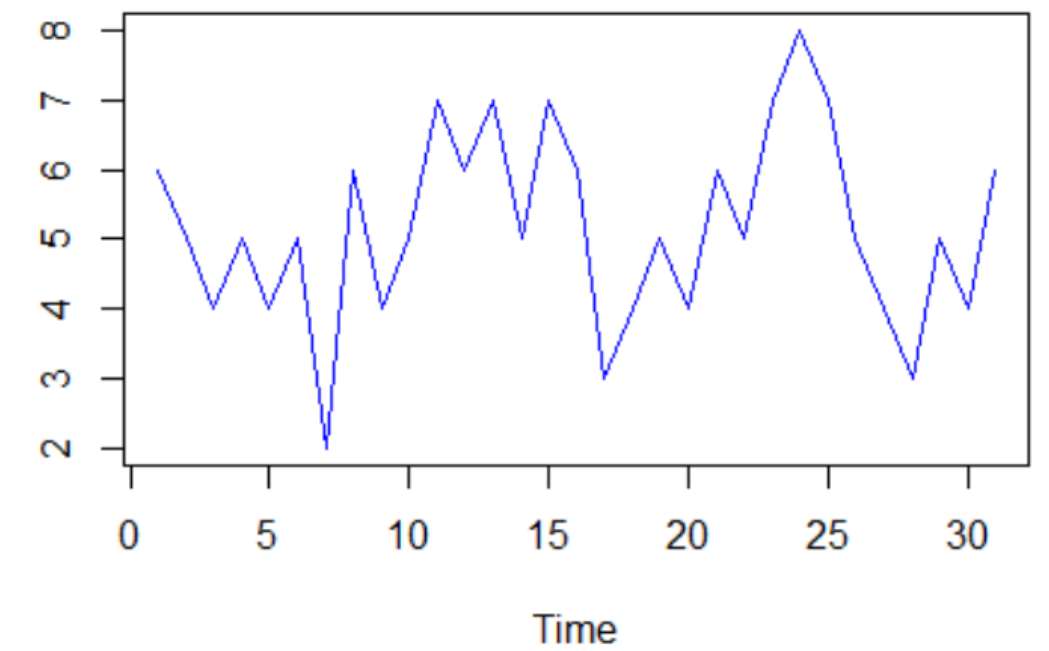
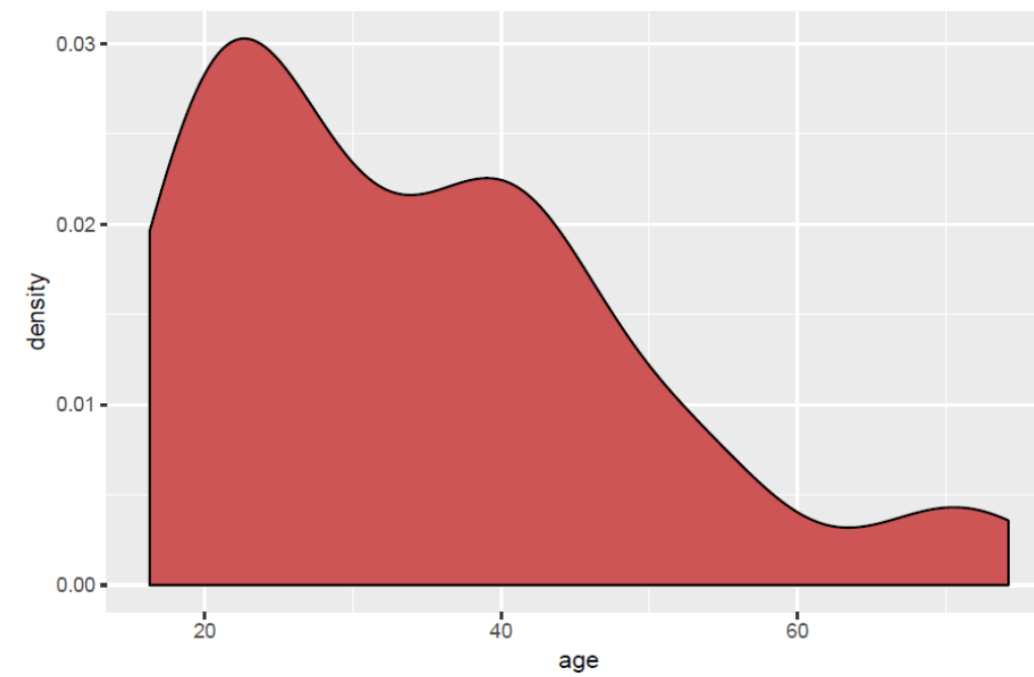
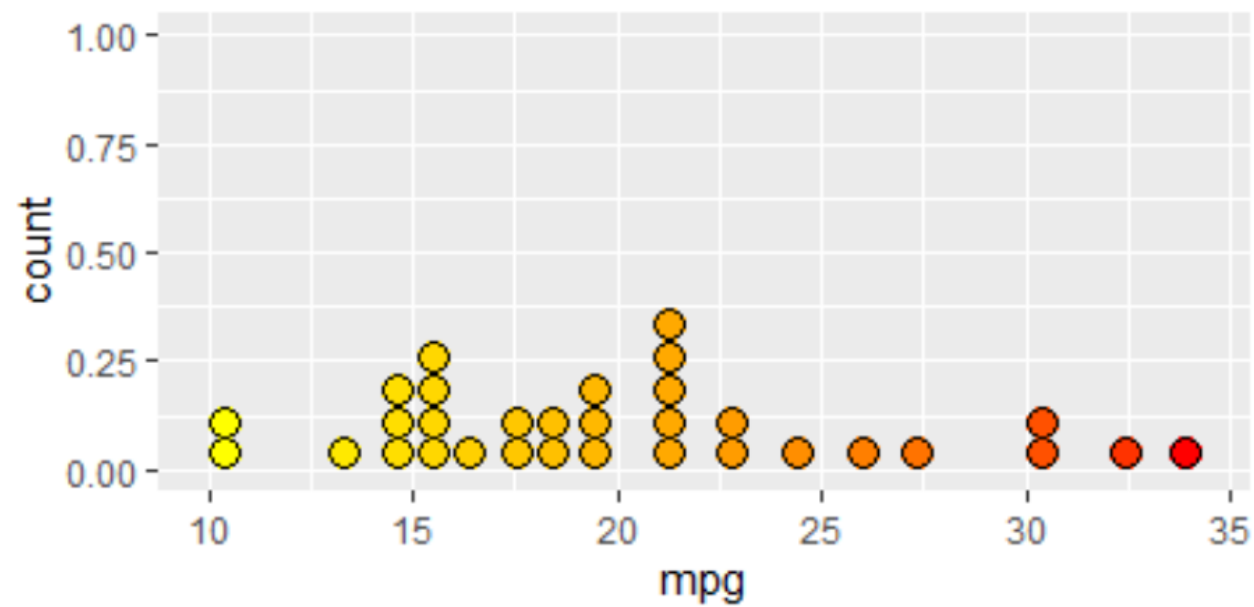
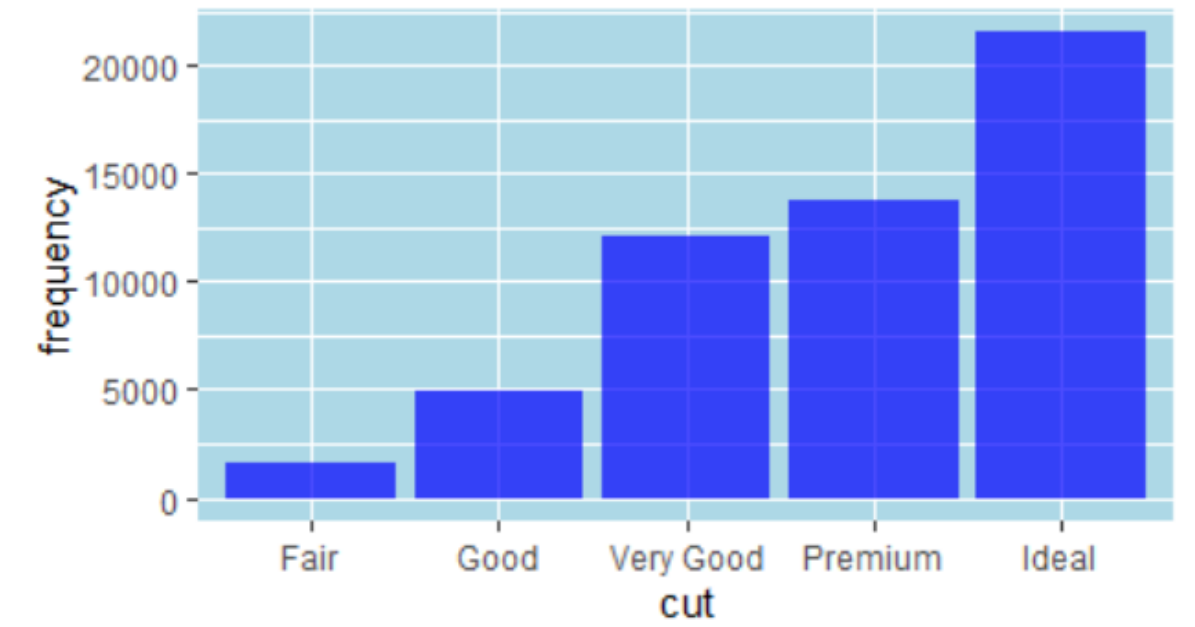
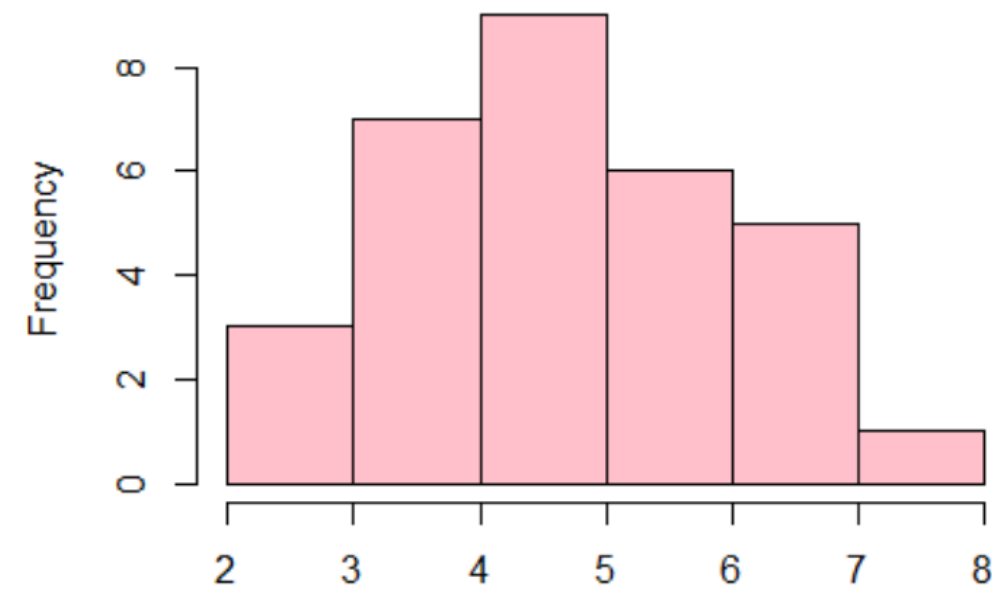
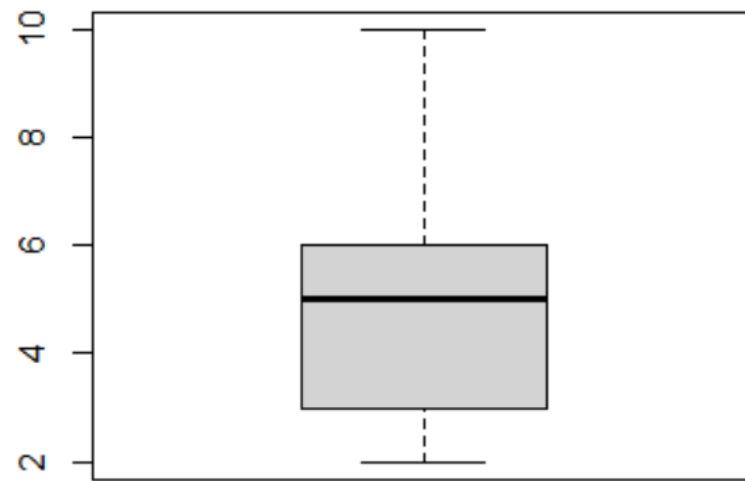
# Paket tidyverse di R

```
library(tidyverse)
mtcars %>%
  group_by(cyl) %>%
  summarise(Mean=mean(cyl), Median=median(cyl), St.Dev=sd(cyl))
```

```
> mtcars %>%
+   group_by(cyl) %>%
+   summarise(Mean=mean(cyl),
Median=median(cyl), St.Dev=sd(cyl))
# A tibble: 3 x 4
   cyl  Mean Median St.Dev
  <dbl> <dbl>   <dbl> <dbl>
1     4     4     4     0
2     6     6     6     0
3     8     8     8     0
```

OUTPUT

# Grafik



# Korelasi

- Identifikasi korelasi antar variabel dapat dilihat dari **scatter plot** dan **koefisien korelasi** ( $r$ ).
- **Scatter plot** dapat digunakan untuk mengetahui korelasi antar variabel secara visual dengan melihat persebaran titik-titik data.
- **Nilai koefisien korelasi** ( $r$ ) mengindikasikan kekuatan dan arah dari hubungan yang linear antar dua variabel, yaitu antara -1 dan +1.

## Koefisien Korelasi Pearson

**Populasi:**  $\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$

**Sampel:**  $r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$

atau

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n - 1)s_x s_y}$$

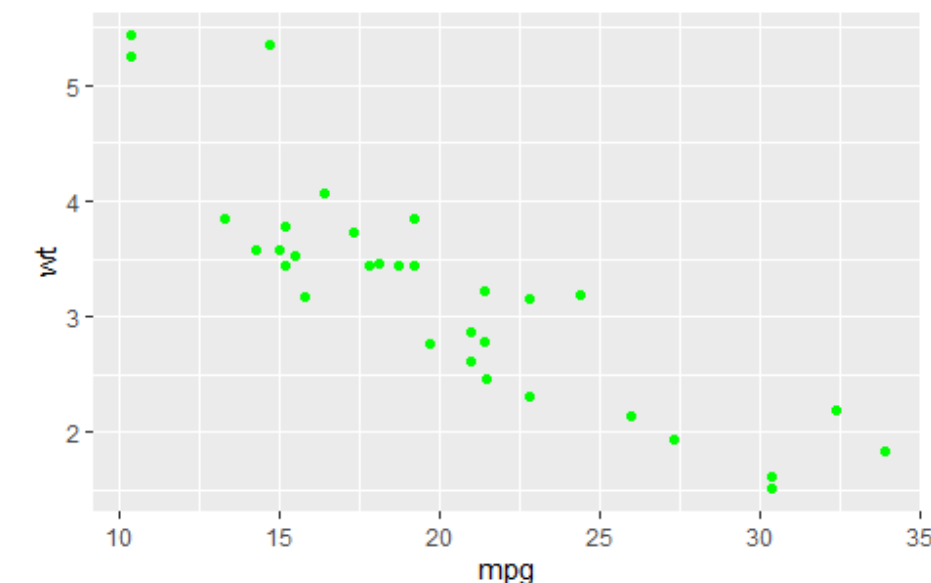
### Keterangan:

$cov(X,Y)$  kovariansi dari X dan Y,  $\sigma_X$  deviasi standar populasi dari X,  $\sigma_Y$  deviasi standar populasi dari Y,  $s_X$  deviasi standar sampel dari X,  $s_Y$  deviasi standar sampel dari Y

```
korelasi<-cor(mtcars$wt,mtcars$mpg)
```

```
> cor(mtcars$wt,mtcars$mpg)  
[1] -0.8676594
```

```
qplot(mpg, wt, data = mtcars, colour = I("green"))
```



# Tugas di Kelas

Buatlah fungsi pengurutan (sort)