

Manajemen Data Relasional (STK261)

Teknik Wrangling

Dr. Agus Mohamad Soleh

Departemen Statistika Fakultas Matematika dan IPA Ganjil 2021



Bentuk umum:

Peubah = ekspresi_numerik;

Ekspresi_numerik terdiri dari:

- 1. Konstanta **numerik**
- 2. Peubah **numerik**
- 3. Operator **numerik**
- 4. Functions: SQRT, ABS, MIN, MAX, SIN, dll

```
Contoh:
    a. x=(-b + SQRT(b^{**2} - 4*a*c))/2*a;
    b.IF jk=1 THEN jenis='pria';
      ELSE jenis='wanita';
    c. DATA metric;
      INFILE statlab;
      INPUT bobotpon 16-19;
      bobotkg = bobotpon/2.2;
```



$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

$$d = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

$$y = e^{e} \log 20x$$

- 1. Konstanta numerik: 4, 2, 20
- 2. Peubah numerik: x, b, a, c, d
- 3. Operator numerik: -, +, *, /, **
- 4. Functions: SQRT, LOG

urutan dari yang paling tinggi : pangkat; kali & bagi; plus & minus

Membuat var karakter menjadi numerik

```
DTA A;

INPUT x$;

CARDS;

1 2 3 4 5;

DATA B;

SET A;

X_Num = x*1;
```

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

$$x = (-b+SQRT(b**2-4*a*c))/(2*a);$$

$$d = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$$

$$d = ((1/SQRT(2*2,317))*(e**((-1/2)*z**2));$$

$$y = e^{e} \log 20x$$

$$y = LOG(20*x);$$



Fungsi-Fungsi Numerik:

```
INT(arg): Returns the integer portion of argument.
```

```
e.g. x=INT(4.32); gives x=4.
```

LOG(arg): Natural logarithm.

MAX(arg,arg,...): Largest non-missing value.

MIN(arg,arg,...): Smallest non-missing value.

MEAN(arg,arg,...): Arithmetic mean of non-missing values.

SUM(arg,arg,...): Sum of non-missing values.

ROUND(arg,round-off-unit): Rounds to nearest round-off unit.

Contoh: x=ROUND(12.65); returns x=13

y=ROUND(12.65, .1); returns y=12.7.



```
* Modify a data set with assignment statements;
DATA homegarden;
INFILE 'c:\MyRawData\Garden.txt';
INPUT Name $ Tomato Zucchini Peas Grapes;
    Zone=16;
    Type='home';
    Zucchini=Zucchini*10;
    Total=Tomato+Zucchini+Peas+Grapes;
    PerTom=(Tomato/Total)*100;
PROC PRINT DATA=homegarden;
    TITLE 'home gardening survey';
RUN;
```

Bentuk umum:

Peubah = ekspresi_string;

- Ekspresi_string terdiri dari:
 - 1. Konstanta karakter atau string
 - 2. Peubah string
 - 3. Operator **string**: ||
 - 4. Functions: TRIM, LENGTH, COMPRESS



Contoh:

DATA nama;

fungsi length membuat tidk perlu menggunakan input secara banyaknya kolom

* pemisahnya adalah spasi

```
LENGTH nama1 $ 20 nama2 $ 20 nama3 $ 40;
INPUT nama1 $ nama2 $;
                                 menggabungkan 2 var karakter
nama3 = TRIM(nama2)
                                           TRIM(nama1);
             fungsi untuk menghilangkan spasi
LINES;
             di ujung kiri dan di kanan
                                      jadi ini menggabungkan nama2 dengan koma dan dengan nama1
              * berbeda dengan COMPRESS yang
               menghapus seluruh spasi
Aji Wigena
Budi Santoso
Rudi Nurtanio
PROC PRINT DATA=nama;
TITLE 'Daftar Nama';
```



Character:

LEFT(arg): Left aligns a character expression.

e.g.
$$a='$$
 my cat'; $x=LEFT(a)$; $\rightarrow x='$ my cat'.

mengambil sebagian karakter bagian kiri

LENGTH(arg): Returns the length of an argument not counting trailing blanks (missing values have a length of 1).

e.g. a=' my cat ';
$$x=LENGTH(a)$$
; $\rightarrow x=7$. spasi dihitung menghitung banyaknya karakter

SUBSTR(arg,position,n): Extracts a substring from an argument starting at 'position' for 'n' characters or until end if no 'n'.

| Mulai dari data kolom ke 2 | Sebanyak 3 karakter yang akan diambil

e.g.
$$x=SUBSTR('(814)863-5678',2,3); \rightarrow x='814'.$$



Pernyataan SET

Bentuk umum:

SET namagugusdataSAS;

- Tidak ada pernyataan INPUT bersamaan dengan pernyataan SET
- Untuk membuat subset data dari suatu gugusdataSAS

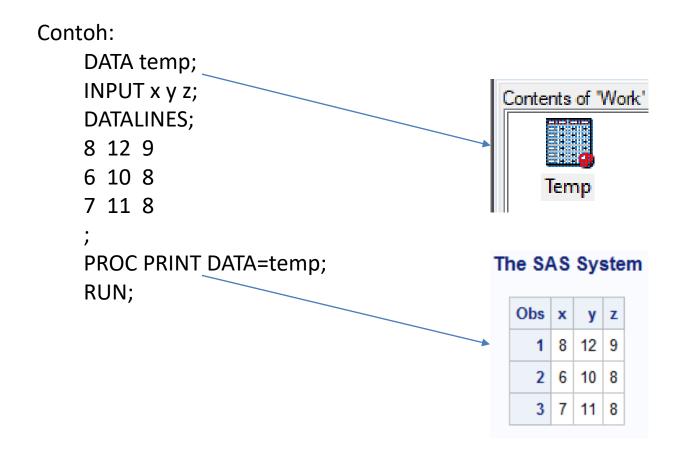
```
Contoh:

DATA temp;
INFILE 'D:/xyzdata';
INPUT x y z;
DATA jumlah;

SET temp;
jum = x+y+z;
PROC PRINT DATA=jumlah;
```



Pernyataan SET





Pernyataan SET

Contoh: DATA temp; INPUT x y z; Explorer DATALINES; Contents of 'Work' 8 12 9 6 10 8 7 11 8 Jumlah Temp PROC PRINT DATA=temp; The SAS System DATA jumlah; **SET** temp; Obs x z jum jum = x+y+z;12 9 PROC PRINT DATA=jumlah; 10 8 24 RUN;

3 7 11 8

26



Pernyataan DROP dan KEEP

Bentuk umum:

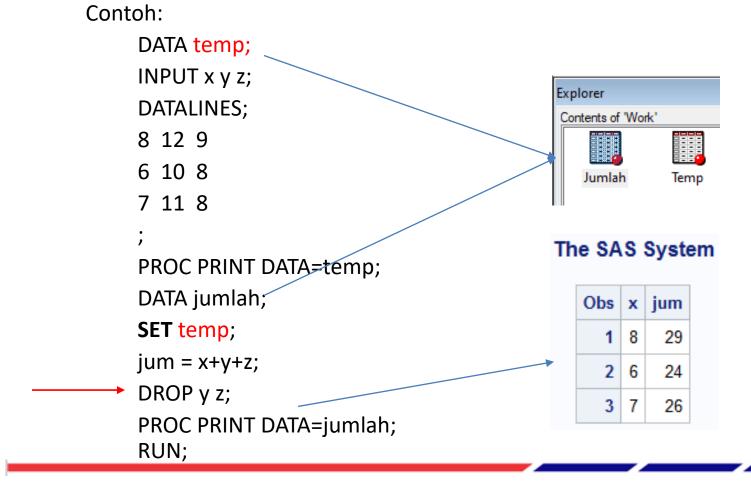
```
DROP peubah ...;
KEEP peubah ...;
```

```
DATA temp;
INFILE 'D:/xyzdata';
INPUT x y z;
DATA jumlah;
SET temp;
jum = x+y+z;
DROP x y z;
tidak boleh dilakukan secara bersama-sama

NEEP jum;
PROC PRINT DATA=jumlah;
```

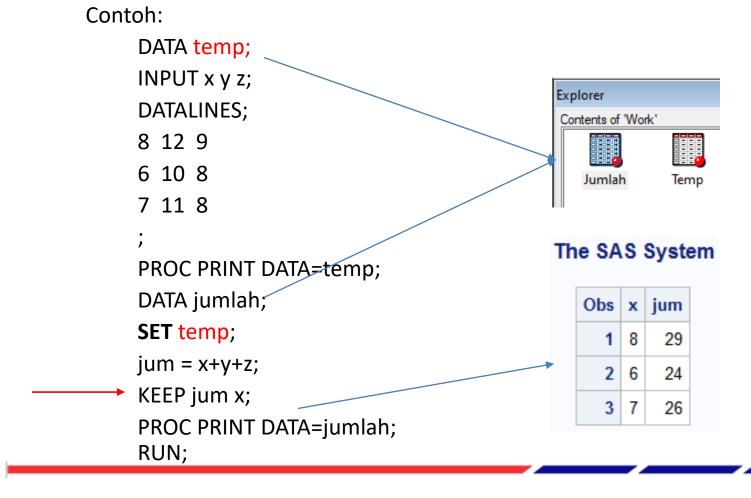


Pernyataan DROP dan KEEP





Pernyataan DROP dan KEEP



Subsetting Variabel

1. Pernyataan sendiri

```
DATA B;
SET A
DROP x y;
```

2. Opsi dari pernyataan SET

ini bisa diproses lebih cepat, karena tidak membaca seluruh data dari A terlebih dahulu

```
DATA B; SET A (DROP = x y);
```

3. Opsi dari tahapan data

```
DATA B (DROP = x y);
SET A;
```



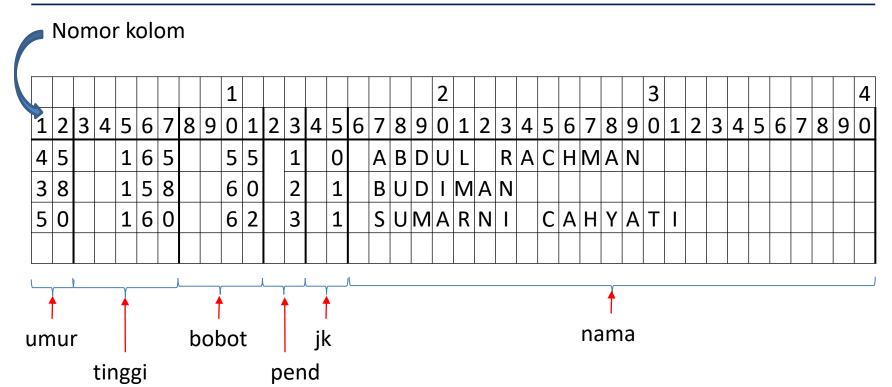
```
IF expression; mengambil nilai yang true saja
IF expression THEN DELETE; menghapus nilai yang false
```

```
IF Sex='f';
IF Sex='m' THEN DELETE;

* Choose only comedies;
DATA comedy;
    INFILE 'c:\MyRawData\Shakespeare.dat';
    INPUT Title $ 1-26 Year Type $;
    IF Type='comedy';
PROC PRINT DATA=comedy;
RUN;
```



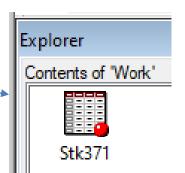
Contoh DATA





Contoh: tanpa pernyataan IF

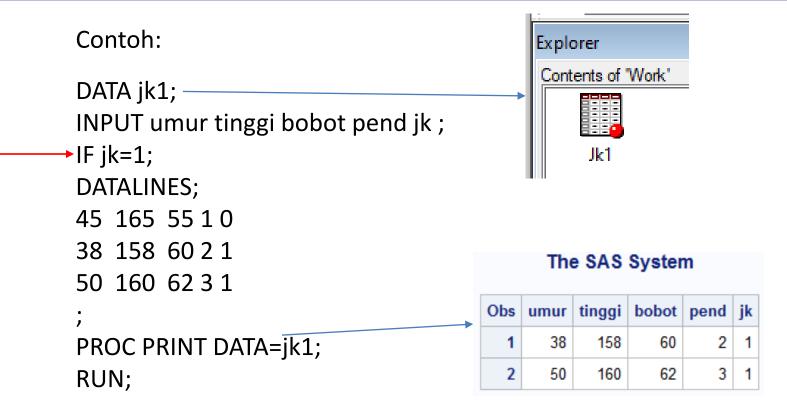
```
DATA stk371;
INPUT umur tinggi bobot pend jk;
DATALINES;
45 165 55 1 0
38 158 60 2 1
50 160 62 3 1
;
PROC PRINT DATA=stk371;
RUN;
```



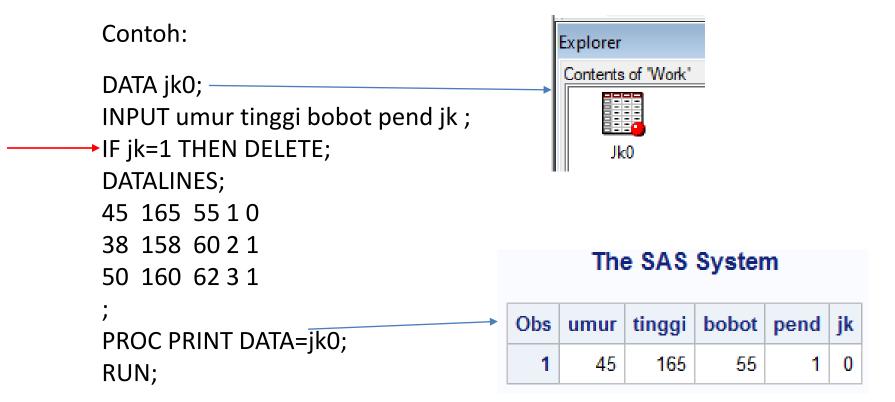
	•							
Obs	umur	tinggi	bobot	pend	jk			
1	45	165	55	1	0			
2	38	158	60	2	1			
3	50	160	62	3	1			

The SAS System











Bentuk umum:

```
DO var1 = var2 TO var3 [BY pertambahan];
:
:
END;
```

```
DATA contoh1;
INPUT pon @@;

kilo = pon/2.2;

LINES;

O.1 0.2 ... 19.9 20.0

DATA contoh2;

DO pon = 0.1 TO 20.0 BY 0.1;

kilo = pon / 2.2;

OUTPUT;

dengan OUTPUT, data yang diprint akan menjadi 1 baris

END;
```

```
DATA A;

INPUT X;

DO i=1 TO 5;

x=x+i;

END;

CARDS;

2

3

4
```



```
Contoh:

DATA normal1;

DO z = -4 TO 4.05 BY 0.2;

density = (1.0/SQRT(2*3.1415926536))*EXP(-z**2/2);

OUTPUT;

END;

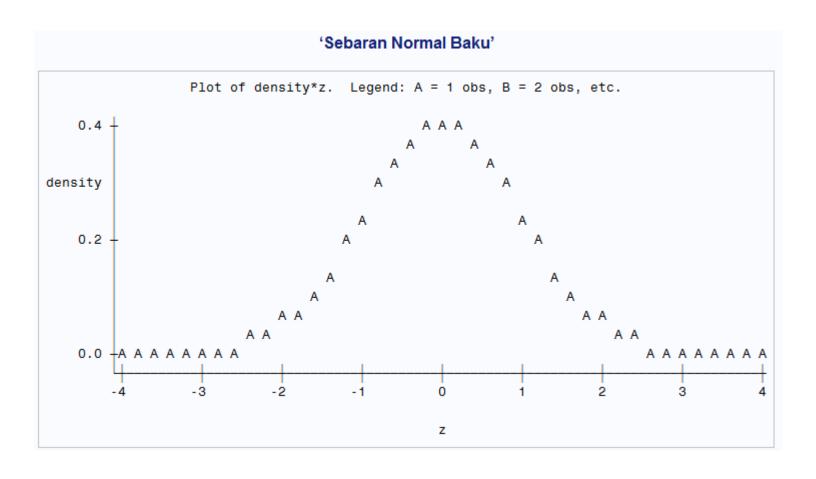
PROC PLOT DATA=normal1;

PLOT density*z / HAXIS=-4 TO 4 BY 1;

TITLE 'Sebaran Normal Baku';

RUN;
```





```
DATA normal2;
DO i = 1 TO 40;
    x = 50 + 10*NORMAL(0);
    OUTPUT;
END;
PROC CHART DATA=normal2;
VBAR x / MIDPOINTS=20 TO 80 BY 5;
TITLE 'Bangkitan data normal dengan rataan 50 dan simpangan 10';
RUN;
```

```
DATA seragam;
DO i = 1 TO 100;
y = UNIFORM(0);
OUTPUT;
END;
PROC CHART DATA=seragam;
VBAR y / MIDPOINTS=0.05 TO 0.95 BY 0.1;
TITLE 'Bangkitan data seragam (uniform) dengan selang nilai (0,1)';
RUN;
```



```
DATA normal3;
n = 40;
DO i = 1 \text{ TO } 500;
   DO j = 1 TO n;
         x = 50 + 10*NORMAL(0);
         jum = jum + x;
   END;
   Rataan = jum / n;
   OUTPUT;
END;
PROC PRINT DATA=normal3;
TITLE 'Bangkitan data 500 kali masing2 dengan n=40';
PROC CHART DATA=normal3;
VBAR rataan / MIDPOINTS=40 TO 60 BY 2;
RUN;
```



Menambahkan secara kolom:

MERGE A B;

SET A; SET B: menambah secara baris (ke bawah) :

SET A B;

Bentuk umum:

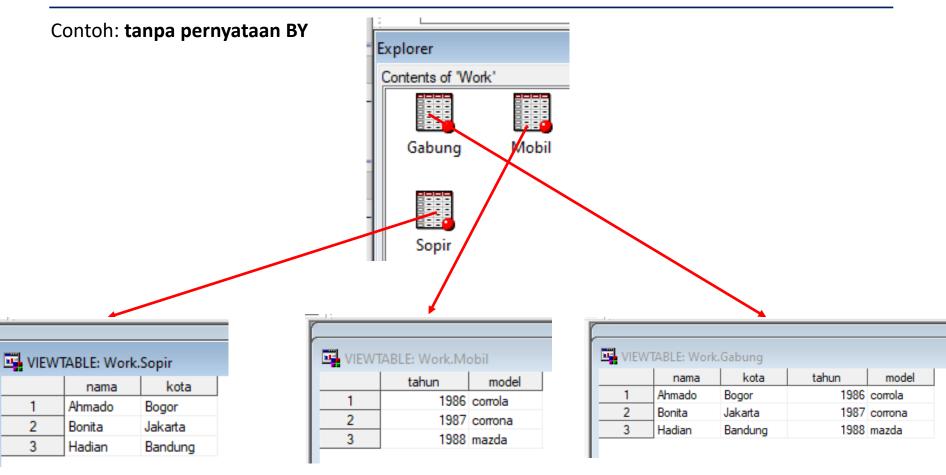
MERGE namadataSAS1 namadataSAS2;

 Menggabungkan dua atau lebih gugus data SAS berdasarkan setiap pengamatan yang sesuai atau pernyataan BY



```
Contoh: tanpa pernyataan BY
      DATA sopir;
      INPUT nama $ kota $;
      CARDS:
      Ahmado Bogor
      Bonita Jakarta
                                                 Explorer
      Hadian Bandung
                                                 Contents of 'Work'
      DATA mobil;_
      INPUT tahun model $;
                                                    Gabung
                                                               Mobil
      CARDS;
      1986 corrola
      1987 corrona
      1988 mazda
                                                     Sopir
      PROC PRINT DATA=sopir;
      PROC PRINT DATA mobil;
      DATA gabung:
      MERGE sopir mobil;
      PROC PRINT;
```

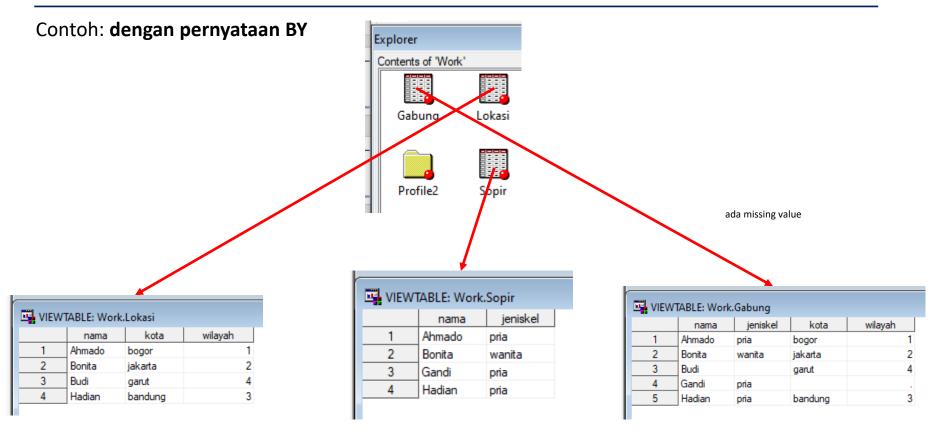






```
Contoh: dengan pernyataan BY
                                                   Explorer
      DATA sopir;
                                                    Contents of 'Work'
      INPUT nama $ jeniskel $;
      CARDS:
                                                               Lokasi
      Ahmado pria
      Bonita wanita
      Hadian pria
      Gandi pria
             harus di sortir
      PROC SORT DATA-sopir; BY nama;
      DATA lokasi;
      INPUT nama $ kota $ wilayah;
      CARDS;
      Budi garut 4
                                               PROC SORT DATA=lokasi; BY nama;
      Ahmado bogor 1
                                               DATA gabung;
      Bonita jakarta 2
                                               MERGE sopir lokasi; BY nama;
      Hadian bandung 3
                                               PROC PRINT;
```







Pernyataan PROC SORT

OUT digunakan apabila tidak ingin mengotak-atik gugus data awal * jadi gugus data tetap murni (tidak berubah)

```
PROC SORT DATA=gugusdataSAS [OUT=gugusdataSAS];
BY peubah-peubah;
BY [DESCENDING] peubah-peubah;
```

```
PROC SORT DATA=demo OUT=outsort; BY jk;
```



Pernyataan PROC SORT

```
proc format;
 value fjk 1='Laki-Laki' 2='Perempuan';
 value $ftklas 't1'='tinggi' 't2'='rendah';
libname stk 'D:\education\statcomp\KomStat\STK371\Ganjil201516\';
data stk.TBMHS;
infile 'D:\education\statcomp\KomStat\STK371\Ganjil201516\TBDATA2.txt';
input JK TINGGI;
if TINGGI<=165 then TKLAS='t1';
else TKLAS='t2';
proc sort data=stk.TBMHS out=stk.TBMHSORT;
by JK;
proc print;
var JK TKLAS;
format JK fjk. TKLAS $ftklas.;
run;
```



Pernyataan PROC SORT

pelum SORT (stk.TBMHS)

Obs	JK	TKLAS
1	Laki-Laki	rendah
2	Perempuan	tinggi
3	Perempuan	tinggi
4	Laki-Laki	rendah
5	Laki-Laki	tinggi

Setelah SORT (stk.TBMHSORT)

Obs	JK	TKLAS
1	Laki-Laki	rendah
2	Laki-Laki	rendah
3	Laki-Laki	tinggi
4	Perempuan	tinggi
5	Perempuan	tinggi

Format data

Long			Wide		
Nam	na Ujiar	n Nilai	Nam	na Nilai1	Nilai2
A	1	60	A	50	60
A	2	70	B	90	70



Terima Kasih



Inspiring Innovation with Integrity in Agriculture, Ocean and Biosciences for a Sustainable World