IPB University

# Manajemen Data Relasional

## (STA261)

## Bagian UAS: Penggunaan Software Statistics

Dr. Agus Mohamad Soleh

Departemen Statistika
Fakultas Matematika dan IPA
Ganjil 2021

# Pokok Bahasan

Membahas konsep dan perangkat lunak untuk pemasukan (entering) data, modifikasi data, manajemen data, dan transfer data SQL menggunakan paket program SAS dan R

SAS untuk 1 komputer harganya 100jt/tahun. IPB berkerjasama dengan SAS Indonesia

Data relasional -> formatnya tabel.
Big data -> formatnya 3 dimensi.

# Materi

## A. Program SAS

1. Pengantar *SAS System*
2. Pemasukan Data (DATA, INFILE, INPUT, DATALINES, LIBNAME, PROC CONTENTS)
3. Modifikasi Data (PROC FORMAT, IF-THEN, Functions, Pernyataan *assignment*, Struktur program, operasi data)
4. Pengelolaan Data (LIBNAME, Set & Merge, Combine, Subsetting, Recode)
5. Import Data (IMPORT dan EXPORT, SQL)

# Materi

B. Program R

1. Pengantar R
2. Pemasukan Data (vector, matrix, array, factor, list, data.frame)
3. Pengelolaan Data (Merge, Combine, Subsetting, Recode)
4. Transfer Data (DBI, R<dbms>)

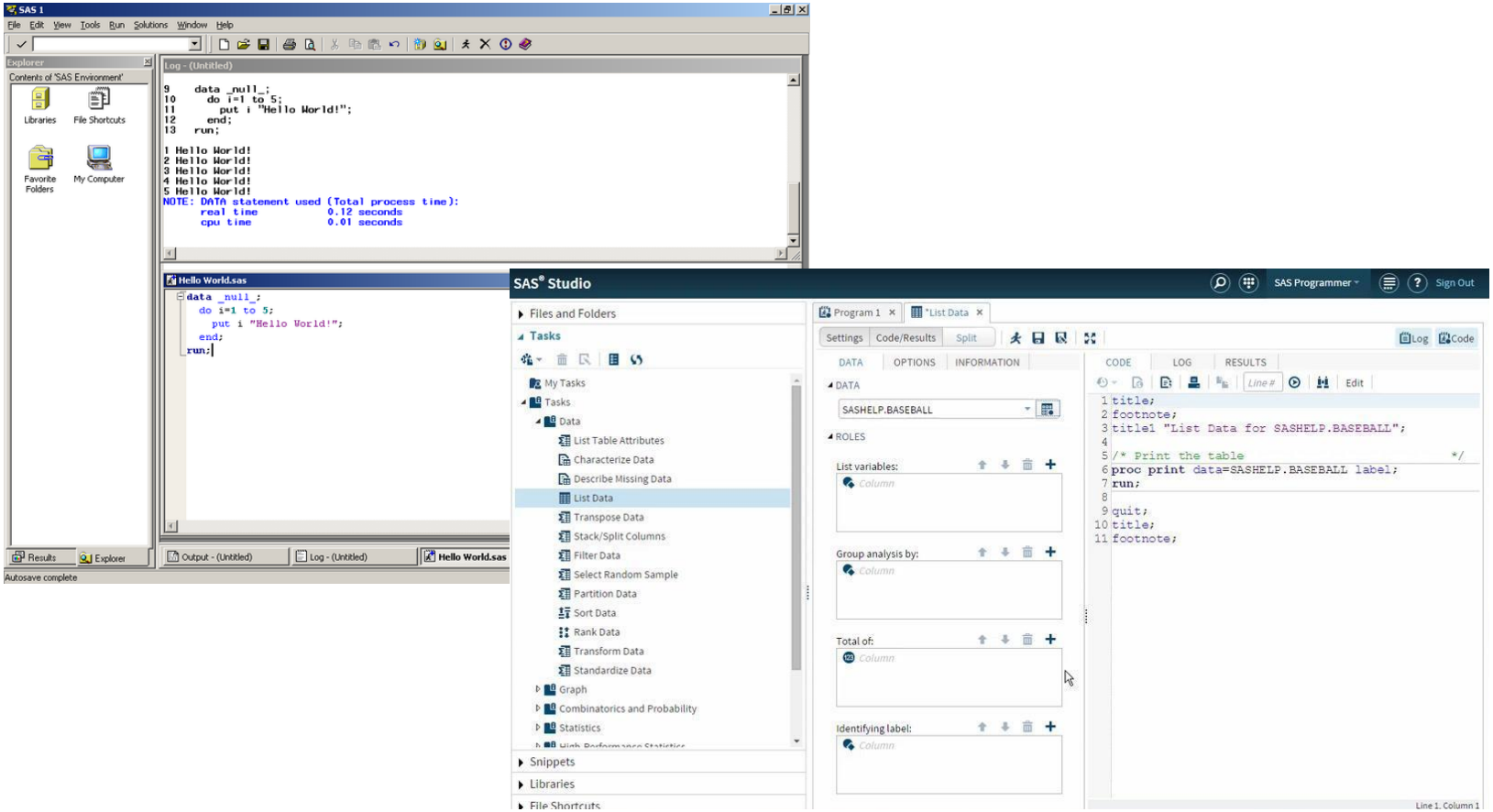
# PENGANTAR SAS SYSTEM

# Sistem SAS

**Sistem SAS** (*SAS System*) adalah paket pemrograman dan analisis data yang dikembangkan oleh SAS Institute, Cary NC

(James Goodnight)

# LINGKUNGAN SAS SYSTEM



# Modul SAS

- Modul inti
  - **SAS/BASE**
- Modul Analisis Statistika
  - SAS/TTEST, SAS/STAT, SAS/ETS, SAS/QC
- Modul Pengolahan Matriks
  - SAS/IML
- Modul Pemetaan
  - SAS/GRAPH
- Modul Sistem Informasi Eksekutif
  - SAS/EIS

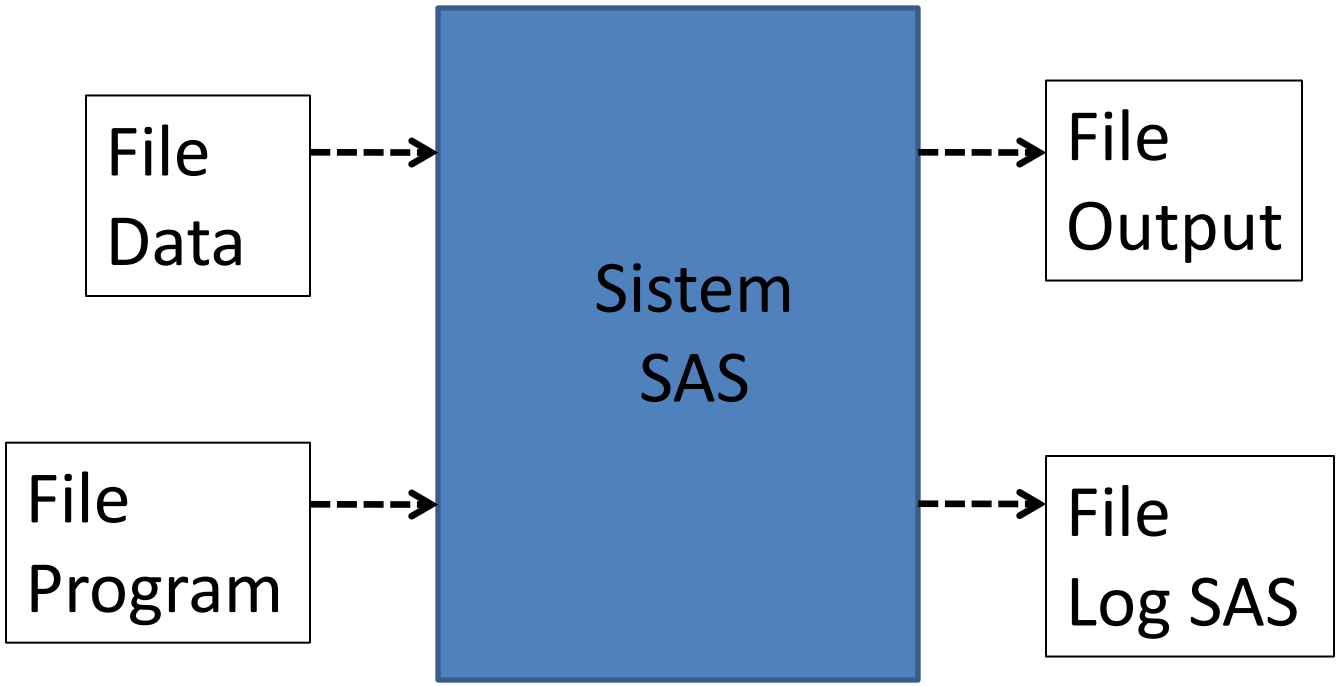
# Kemampuan SAS

- Penyimpanan Data
  - Gugus Data SAS
- Modifikasi Data
  - Fungsi-fungsi khusus (SQRT, ABS, dll)
- Penanganan File (Data)
  - Subsetting, Concatinating, Merging
- Pembuatan *Report*
  - Tabel, Chart (histogram, pie)
- Analisis Statistika
  - Deskriptif, Perbandingan dua nilai tengah, Anova, Analisis Regresi
- Penyusunan Sistem Informasi
  - Sistem informasi manajemen

# Komponen Pengolahan Data

- Paket SAS *(SAS System)*
- File Program SAS
  - Pernyataan-Pernyataan SAS
- File Data
  - Data input
- File Output
  - Hasil proses
- File Log
  - Informasi tentang proses

# Komponen Pengolahan Data



# Program SAS

- Tahapan DATA (DATA Steps)
  - Baca data (INPUT)
  - Modifikasi data
  - Penanganan data
  - Membuat gugus data baru
- Tahapan PROC (PROC Steps)
  - Memanggil satu atau lebih prosedur (PROC) dan option-optionnya

bahasa SAS tidak sensitif. bahasanya compiler bukan intrepreter kyk r (per baris bisa dieksekusi)

Tahapan data -> diawali penyiapan data

-> Menghasilkan gugus data sas

-> Tahapan data bisa ada 1, 2, dst

-> Diakhiri dengan ';'

-> Spasi & enter berpengaruh?

-> Run = eksekusi + **Menampilkan program**

-> Input = menampung variabel

-> Minimal ada 1 run di bagian akhir program

# Materi

## A. Contoh Program SAS

minimal ada 1 spasi

tipe data

defaultnya numeric

byte size

defaultnya 8

karakter

menyaipkan variabel

kalo ada variable lain, harus dibuat ekspresi seperti turun yg membuat variable yang sudah diinput sebelumnya

bisa : cards; lines; datalines;

untuk input data

titik koma untuk input data harus ada di enter

```
data bobot;
input nomor 1-4 nama $ 6-24 tim $ bobotsblm bobotssdh;
turun = bobotsblm - bobotssdh;
lines;
1023 Dodi Suradi          merah    189 165
1049 Amelia Serrano       kuning   145 124
1219 Alan Nance           merah    210 192
1246 Ravi Sinha           kuning   194 177
1078 Asley Makatano       merah    127 118
;
proc print noobs;
var nama bobotsblm bobotssdh turun;
proc sort data=bobot; by nomor;
run;
```

# Program SAS

- Steps may contain many statements
- Steps usually end when:
  1. Another step begins
  2. **RUN;** statement appears

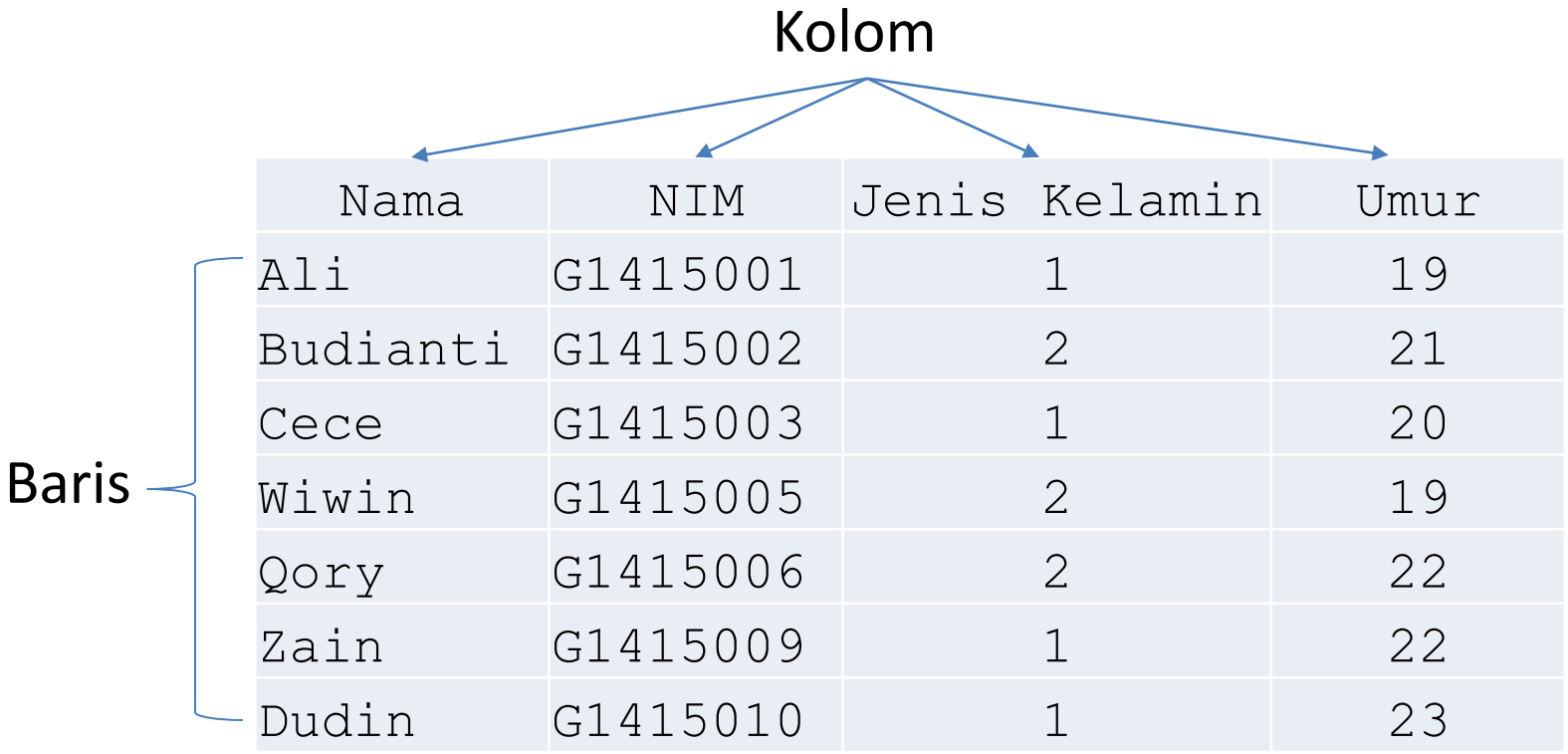
```
DATA height;
..SAS statements..
run;
PROC PRINT
  DATA=height;
run;
PROC REG;
..SAS statements..
run;
```

# Program SAS

- SAS reads data sets ***one observation at a time***-implicit looping
- SAS executes steps line by line
- Be sure to enter statements in correct order

# Program SAS

## Struktur Data




| Nama | NIM | Jenis Kelamin | Umur |
|---|---|---|---|
| Ali | G1415001 | 1 | 19 |
| Budianti | G1415002 | 2 | 21 |
| Cece | G1415003 | 1 | 20 |
| Wiwin | G1415005 | 2 | 19 |
| Qory | G1415006 | 2 | 22 |
| Zain | G1415009 | 1 | 22 |
| Dudin | G1415010 | 1 | 23 |

# Penyiapan Program SAS

```
         1         2         3
1234567890123456789012345678900
data mhs;
input nama $ nim $ jk$ umur;
datalines;
Ali G1415001 1 19
Budianti G1415003 1 20
:
:
;
proc print;
run;
```

# Penyiapan Program SAS

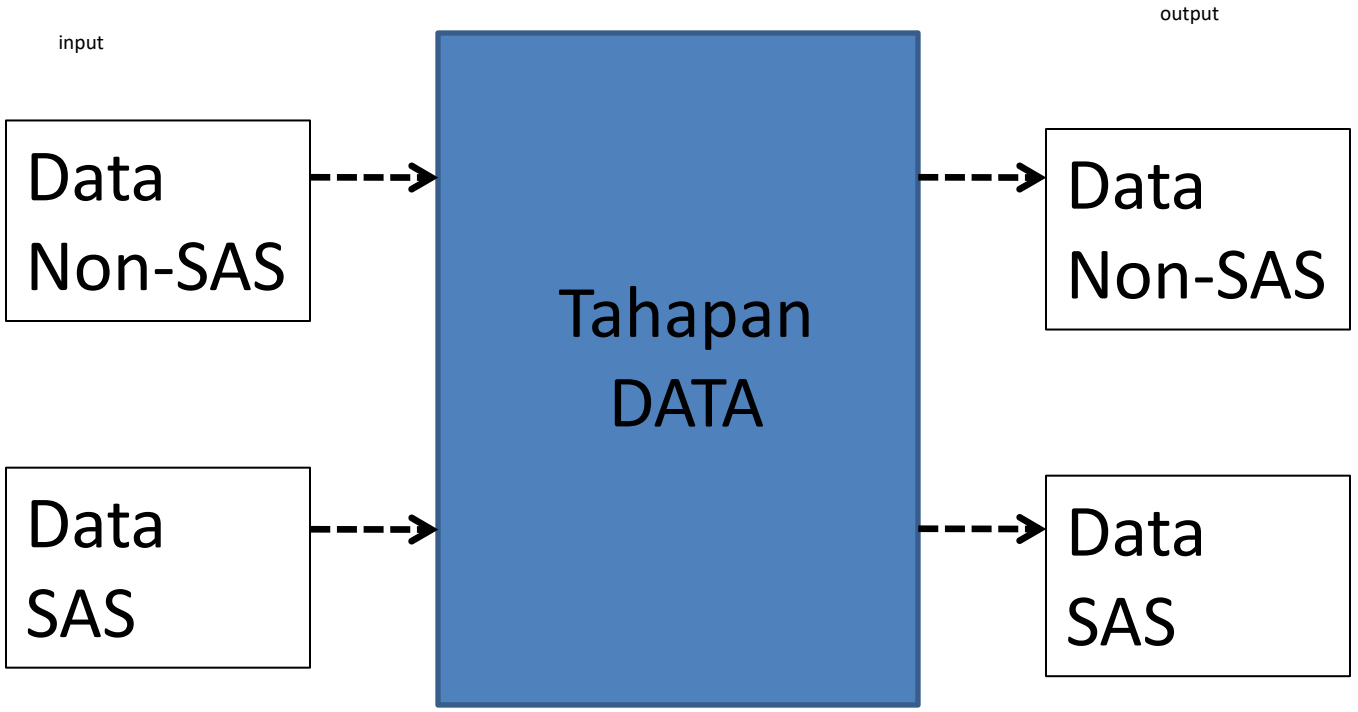
```
data mhs; input nama $ nim $ jk$ umur;
datalines;
Ali G1415001 1 19
Budianti G1415003 1 20
:
:
;
proc print; run;
```

# Program SAS

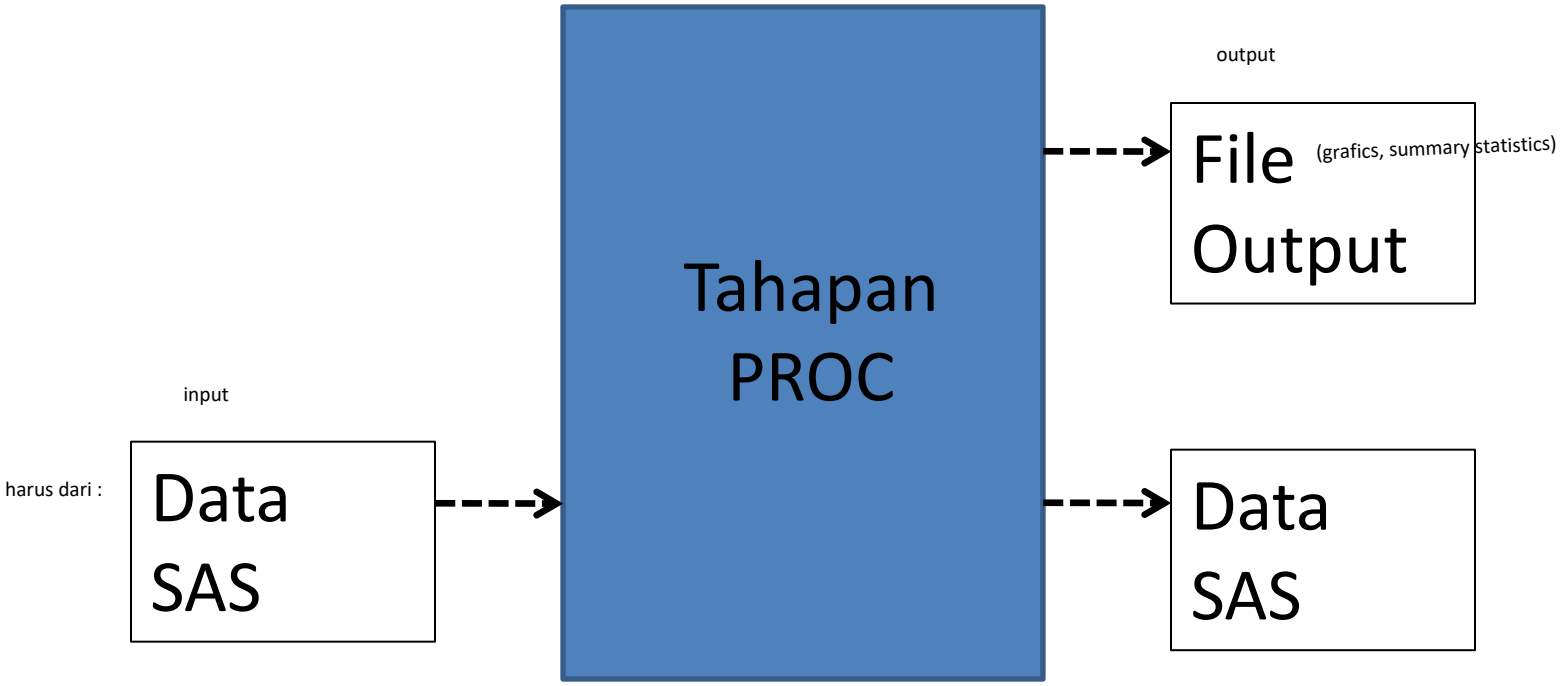
```
/*Tahapan DATA*/
data mhs;
input nama $ nim $ jk $ umur;
datalines;
Ali  G1415001  1  19
Budianti  G1415002  2  21
Cece  G1415003  1  20
Wiwin  G1415005  2  19
Qory  G1415006  2  22
Zain  G1415009  1  22
Dudin  G1415010  1  23
;
run;
```

```
/*Tahapan PROC*/
proc print;
proc means;
  var umur;
proc freq;
  table jk;
run;
```

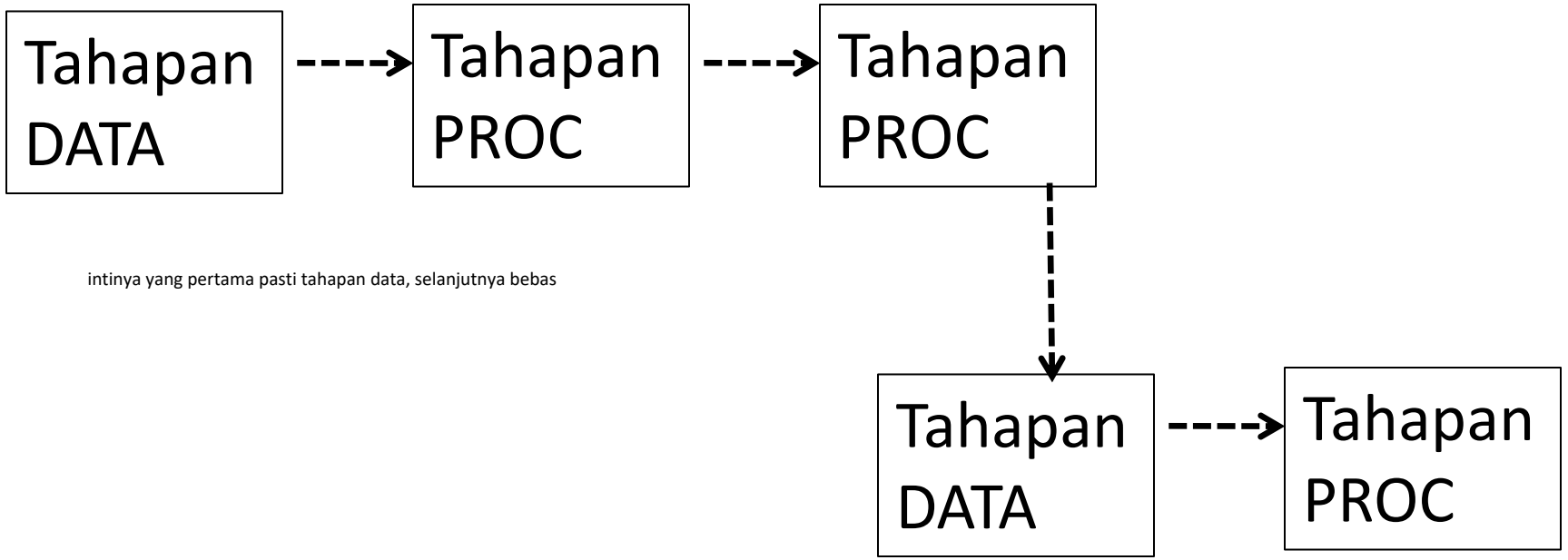
# Urutan Proses Program SAS



# Urutan Proses Program SAS



# Urutan Proses Program SAS




intinya yang pertama pasti tahapan data, selanjutnya bebas

# Tipe Data SAS

- Data Non-SAS
  - Tipe *text* (ASCII)*, excel, dbase,* dll
  - Input pada Tahapan DATA
  - Tidak dapat digunakan langsung pada Tahapan PROC

- Data SAS (Gugus Data SAS – *SAS Dataset*)
  - Hasil Tahapan DATA dengan/tanpa PROC
  - Hanya diproses dengan Sistem SAS

# Gugus Data SAS

- Specially formatted “worksheet”; current extension is ***.sas7bdat** format nama file extention data sas versi 7 (terbaru)
- Variables represented by *Columns*
- Observations represented by *Rows*
- Two Data Types: *Numeric* and *Character*
- Default variable length is 8

| Nama | NIM | Jenis Kelamin | Umur |
|---|---|---|---|
| Ali | G1415001 | 1 | 19 |
| Budianti | G1415002 | 2 | 21 |
| Cece | G1415003 | 1 | 20 |
| Wiwin | G1415005 | 2 | 19 |
| Qory | G1415006 | 2 | 22 |
| Zain | G1415009 | 1 | 22 |
| Dudin | G1415010 | 1 | 23 |

# Gugus Data SAS

- Choose data type based on how you use the variable
- Example: NRP could be character

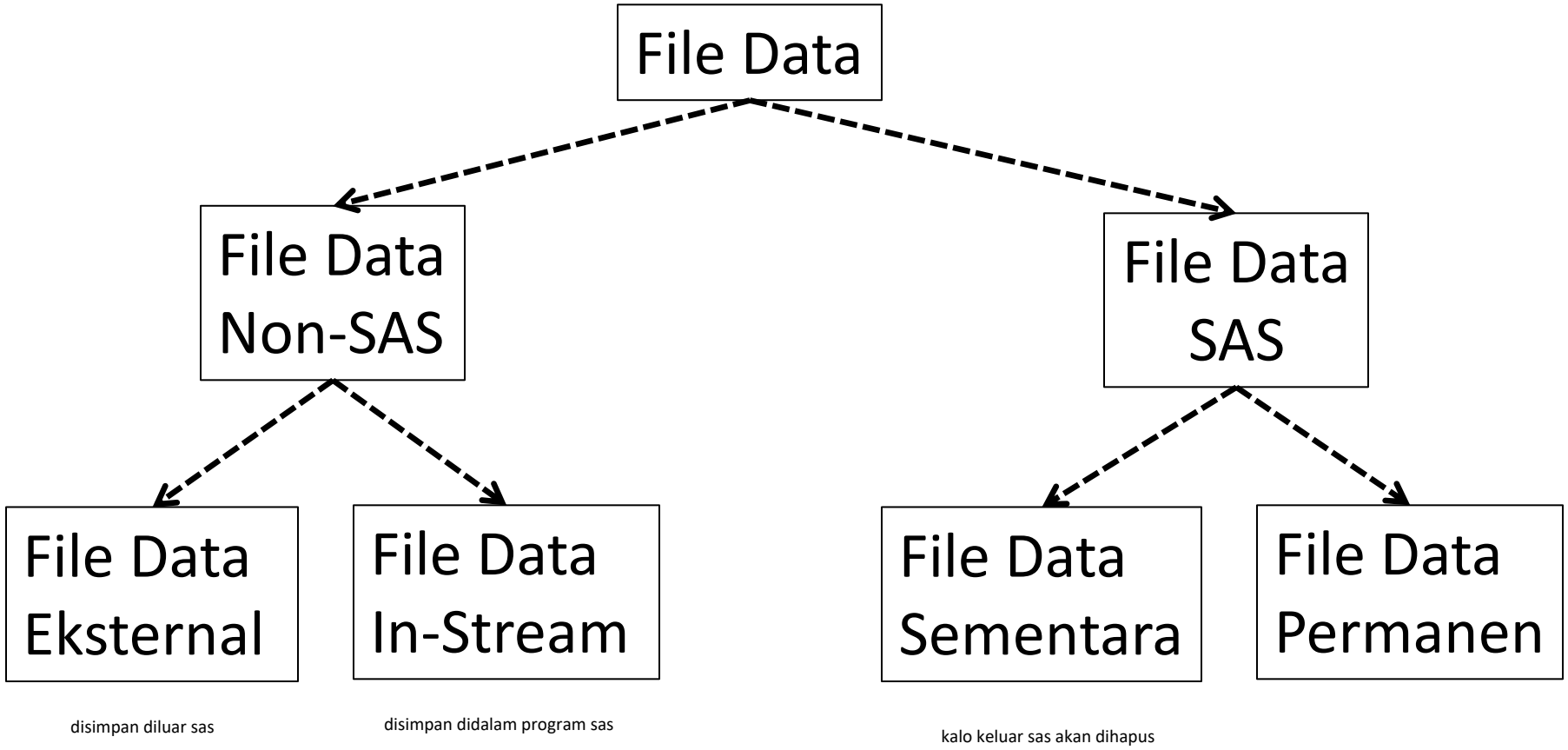
angka yang terlalu besar itu akan disimpan sebagai float.
kapasitas integer : 11digit? (max integer = 2M)

- Missing data in SAS is denoted by a period (.) for numeric data and a null space for character data
- Variable attributes (Type, Length, etc) are stored with the SAS data set

# Penamaan Peubah SAS

- Names can be **32** characters or fewer
- Names must begin with a **letter** or **underscore** (_)
- Names may contain only **letters**, **number**, and **underscores**
- Names can contain **upper/lower** case letters

# Tipe Data


File Data
File Data Non-SAS
File Data SAS
File Data Eksternal
File Data In-Stream
File Data Sementara
File Data Permanen
disimpan diluar sas
disimpan didalam program sas
kalo keluar sas akan dihapus

# Tipe Data

- File Data **Eksternal**
  - File data yang terpisah dari file program SAS
- File Data **In-Stream**
  - File data dalam file program SAS
- File Data SAS **Sementara** (Temporary)
  - File data SAS yang tidak disimpan (save) dalam suatu file
- File Data SAS **Permanen**
  - File data SAS yang disimpan (save) dalam suatu file

# Program SAS dan Tipe Data

```
/* data SAS sementara*/
/* data in-stream    */
DATA stk371;
INPUT x y z;
DATALINES;
10  4  12
 9  9  15
13  8  14
;
RUN;
```

setelah input

yang sementara bisa dijadikan permanen harus gunakan libname

perhatikan susunannya

```
/* data SAS sementara*/
/* data eksternal    */
DATA stk371;
INFILE 'D:\education\statcomp\KomStat\STK371\Ganjil201920\stkdat1.txt';
INPUT x y z;
RUN;
```

sebelum input

# Program SAS dan Tipe Data

posisinya sebelum data

nlib = nama lib, bisa diganti selama tidak lebih dari 32 char

```
/* data SAS permanen */
/* data in-stream    */
LIBNAME nlib 'D:\education\statcomp\KomStat\STK371\Ganjil201920\';
DATA nlib.stk371;
INPUT x y z;
DATALINES;
10  4  12
 9  9  15
13  8  14
;
RUN;
```

# Penggunaan Paket SAS

SAS (***S**tatistical **A**nalysis **S**ystem*)

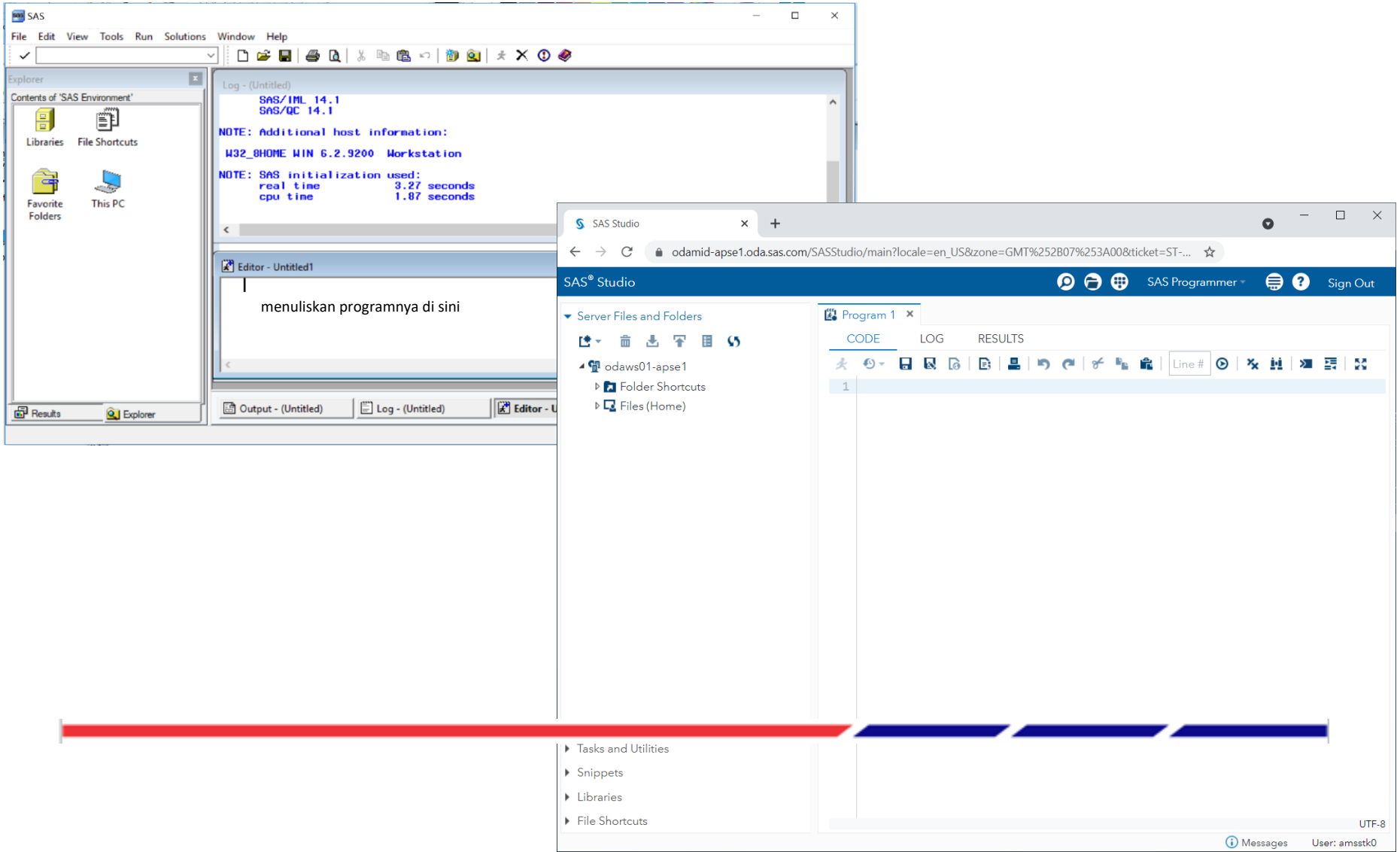
- Awal penggunaan SAS
- SAS Windows (Editor, Log, Output, Explorer, Results)
- Contoh program SAS
- Run program SAS

# Penggunaan Paket SAS

SAS dengan GUI (*graphical user interface*) akan membuka ***window*** :

- **Results** → untuk menyajikan hasil
- **Explorer** → untuk menampilkan data
- **Editor** → untuk menuliskan program SAS
- **Log** → untuk menyajikan sintaks program SAS yang telah diproses (RUN) dan catatan salah sintaks (jika ada)
- **Output** → untuk menyajikan hasil (Results window)

# Penggunaan Paket SAS



# Penggunaan Paket SAS

Contoh Program SAS

```
data test;
   do i = 1 to 100;
      x1 = normal(0);
      x2 = uniform(0);
      output;
  end;
run;
```

Tahapan DATA (*DATA Step*)

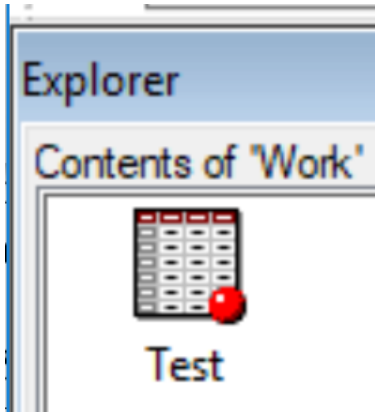
```
proc print data=test (obs=5);
run;
proc univariate data=test;
    var x1 x2;
run;
```

Tahapan PROC (*PROC Step*)

# Penggunaan Paket SAS

## Contoh Program SAS

```
data test;
   do i = 1 to 100;
      x1 = normal(0);
      x2 = uniform(0);
      output;
  end;
run;
```




**Log window**:

```
122  data test;
123     do i = 1 to 100;
124        x1 = normal(0);
125        x2 = uniform(0);
126        output;
127  end;
128  run;

NOTE: The data set WORK.TEST has 100 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time          0.02 seconds
      cpu time           0.03 seconds
```

# Penggunaan Paket SAS

## Contoh Program SAS

```
proc print data=test (obs=5);
run;
```

**Log window**:

```
130
131  proc print data=test (obs=5);
NOTE: Writing HTML Body file: sashtml1.htm
132  run;

NOTE: There were 5 observations read from the data set WORK.TEST.
NOTE: PROCEDURE PRINT used (Total process time):
      real time          0.51 seconds
      cpu time           0.18 seconds
```

| Obs | i | x1 | x2 |
|---|---|---|---|
| 1 | 1 | -0.26159 | 0.17880 |
| 2 | 2 | -0.16799 | 0.44690 |
| 3 | 3 | -1.38560 | 0.50751 |
| 4 | 4 | 0.26943 | 0.49078 |
| 5 | 5 | -1.61011 | 0.47845 |

# Penggunaan Paket SAS

## Contoh Program SAS

error = prgram berhenti

warning = program masih dijalankan

proc print data=test (obs=5) ← tanpa “titik koma”
run;

**LOG window**

```
139  proc print data=test (obs=5)
140  run;
     ---
     22
     202
ERROR 22-322: Syntax error, expecting one of the following: ;, BLANKLINE, CONTENTS, DATA, DOUBLE,
          GRANDTOTAL_LABEL, GRANDTOT_LABEL, GRAND_LABEL, GTOTAL_LABEL, GTOT_LABEL, HEADING,
          LABEL, N, NOOBS, NOSUMLABEL, OBS, ROUND, ROWS, SPLIT, STYLE, SUMLABEL, UNIFORM,
          WIDTH.

ERROR 202-322: The option or parameter is not recognized and will be ignored.
```

# Beberapa Pernyataan SAS

## Entering Data

| | |
|---|---|
| **libname** | An engine to connect to Microsoft files |
| **data** | Begins a data step which manipulates datasets |
| **infile** | Identifies an external raw data file to read |
| **input** | Lists variable names in the input file |
| **datalines** | Indicates internal data |
| **set** | Reads a SAS data set |
| **proc contents** | Contents of a data set |
| **proc print** | Prints observations of variables in a data set |
| **proc copy** | Copies SAS files from one location to another |

# Beberapa Pernyataan SAS

## Entering Data

| | |
|---|---|
| **libname** | An engine to connect to Microsoft files |
| **data** | Begins a data step which manipulates datasets |
| **infile** | Identifies an external raw data file to read |
| **input** | Lists variable names in the input file |
| **datalines** | Indicates internal data |
| **set** | Reads a SAS data set |
| **proc contents** | Contents of a data set |
| **proc print** | Prints observations of variables in a data set |
| **proc copy** | Copies SAS files from one location to another |

# Beberapa Pernyataan SAS

## Exploring Data

| | |
|---|---|
| **proc contents** | Contents of a SAS dataset |
| **proc print** | Displays the data |
| **proc means** | Descriptive statistics |
| **proc univariate** | More descriptive statistics |
| **proc freq** | Frequency tables, frequency charts, and crosstabs |
| **Ods** | Output delivery system, creating output in various formats |
| **proc corr** | Correlation matrix and scatterplots |
| **proc sgplot** | Produces many types of plots |
| **proc contents** | Contents of a SAS dataset |

# Beberapa Pernyataan SAS

## Modifying Data

| | |
|---|---|
| **proc format** | Creates formats (aka value labels) |
| **label** | Creates labels for variables |
| **rename** | Changes the name of a variable in a data step |
| **if then** | Executes a statement only if the condition is true |
| **functions** | Creating new variables using SAS functions |
| **merge** | Merge files |

# Beberapa Pernyataan SAS

## Managing Data

| | |
|---|---|
| **if** and **where** | Conditional statement |
| **keep** | Keeps named variables |
| **drop** | Drops named variables |
| **set** | Reads in named file(s), append |
| **proc sort** | Sorts cases in a dataset |
| **merge** | Merges files |

# Beberapa Pernyataan SAS

## Analyzing Data

| | |
|---|---|
| **proc ttest** | t-tests, including one sample, two sample and paired |
| **proc freq** | Used here for chi-squared tests |
| **proc reg** | Simple and multiple regression |
| **proc glm** | Used here for ANOVA models |
| **proc logistic** | Logistic regression |
| **proc npar1way** | Non-parametric analyses |
| **proc univariate** | Used here for signrank tests |

# Terima Kasih


# IPB University

Bogor Indonesia

Inspiring Innovation with Integrity

in Agriculture, Ocean and Biosciences for a Sustainable World