



STK 572

Manajemen Data Statistik

Tim Dosen:

Dr. Farit Muhammad Affendi

Dr. Agus M Soleh

2016

Agregasi Data

Dr. Agus M Soleh

agusms@apps.ipb.ac.id



Kegunaan

- Memberikan deskripsi data
- Identifikasi dari anomali data
- Tahapan validasi dan verifikasi data

PROC TABULATE

- PROC TABULATE merupakan prosedur yang digunakan untuk menampilkan statistic deskriptif dalam bentuk tabular.

- Syntax

```
PROC TABULATE <options>;  
CLASS variables </ options>;  
VAR variables </ options>;  
TABLE <page> ,  
<row> ,  
column  
</ options> ;  
... other statements ... ;  
RUN;
```



Option PROC TABULATE

Data=	Specifies what input data to use.
Out=	Specifies the name of the output dataset to store calculated values
Format=	Option specifies a format to use for each cell in the table. Best12.2 is the default.
Formchar='...'	This specifies what line characters to use when drawing the table (more on this in a minute).
NoSeps	This option eliminates horizontal separators in the table (only affects traditional SAS monospace output destination).
Order=	Unformatted/Data/Formatted/Freq – orders how the CLASS values appear in the table
Missing	Tells SAS to treat missing values as valid
Style= Contents=	Used with ODS specifications
Exclusive Classdata=dset	To specify exact combinations of data to include



Option VAR Statement

Style=	ODS style element definitions. Example might be to change the justification or the font.
Weight=	specify another variable that will weight the values of the variable with the following exceptions: (0 or <0 = counts observation in total number of observations, blank= exclude observation entirely)



Option TABLE Statement

Box =	Text and style for the empty box in the upper left corner.
Condense	Print multiple pages to the same physical page.
NoContinued	Suppress the continuation message
MissText	If a cell is blank, this text will print instead
PrintMiss	Print CLASS variable values, even if there is not data for them (this only works if somewhere there is at least one observation with that value.
Indent=	Number of spaces to indent nested row headings.
RTSpace =	Number of positions to allow for the row headings.
Style=[options]	Specify ODS style elements for various parts of the table.



Tabel Satu Dimensi

```
PROC TABULATE DATA=data;  
  VAR Income;  
  TABLE Income;  
RUN;
```

Income
Sum
1563354.00



Tabel Satu Dimensi

- Menambahkan Statistik:

```
PROC TABULATE DATA=data;  
    VAR Income;  
    TABLE Income*(N Mean);  
RUN;
```

Income	
N	Mean
30.00	52111.80

Descriptive Statistics	Quantile Statistics
COLPCTN	MEDIAN P50
PCTSUM	P1
COLPCTSUM	Q3 P75
MAX	P90
ROWPCTN	P95
MEAN	P5
ROWPCTSUM	P10
MIN	P99
STDDEV / STD	Q1 P25
N	QRANGE
STDERR	
NMISS	
SUM	Hypothesis Testing
PAGEPCTSUM	ProbT
PCTN	T
VAR	



CLASS STATEMENT

- Mendeskripsikan berdasarkan kategori
- Mirip dengan VAR



Option CLASS Statements

Ascending/Descending	Specify the order the CLASS variables values are displayed
Missing	Consider missing values valid with special missing values treated separately.
MLF	Enables use of multi-level formatting with overlapping ranges (ex – by state and by region at the same time)
Order=	Groups levels of CLASS variables in the order specified: <ul style="list-style-type: none">• Internal (default) – use actual values in data• Data – same order the data is already sorted in• Formatted – use the formatted data values• Freq – highest counts first
Style=[options]	Give ODS style element definitions to these variables
PreLoadFMT	This will preload a format and will also (if other options are also specified), display all values in the table even if there are no observations present with some of the values.
exclusive	Will exclude from the table all combinations of CLASS variables not present in the data (normally used with the preloadfmt option.
groupinternal	Used to group values together by their internal values, not formatted.



CLASS Statements

```
PROC TABULATE data=one;  
  CLASS GENDER;  
  VAR income;  
  TABLE income * (N Mean)  
         INCOME * MEAN * GENDER;  
RUN;
```

Income		Income	
		Mean	
Income		Gender	
		Female	Male
N	Mean	Female	Male
30.00	52111.80	52000.69	52238.79



Tabel Dua Dimensi

```
PROC TABULATE data=one;  
  CLASS gender;  
  VAR income;  
  TABLE income * (N Mean) ,  
           gender;  
RUN;
```

		Gender	
		Female	Male
Income	N	16.00	14.00
	Mean	52000.69	52238.79



Menambahkan Statistik

```
PROC TABULATE data=one;
  CLASS gender;
  VAR income;
  TABLE gender,
         income * (N Mean Max) ;
RUN;
```

	Income		
	N	Mean	Max
Gender			
Female	16.00	52000.69	93849.00
Male	14.00	52238.79	78695.00

```
PROC TABULATE data=one;
  CLASS gender;
  VAR income;
  TABLE gender * (N Mean Max) ,
         income ;
RUN;
```

		Income
Gender		
Female	N	16.00
	Mean	52000.69
	Max	93849.00
Male	N	14.00
	Mean	52238.79
	Max	78695.00



Menambahkan Total dan Sub Total

```
PROC TABULATE data=one;
  CLASS gender fulltime educ;
  TABLE fulltime * gender ALL,
         educ * N ;
RUN;
```

		Educ			
		High School	Bachelors	Masters	Doctorate
		N	N	N	N
Fulltime	Gender				
Fulltime	Female		5.00		5.00
	Male	3.00	1.00	4.00	2.00
Parttime	Female	3.00	4.00	2.00	
	Male	3.00	1.00	3.00	
All		9.00	11.00	9.00	7.00



```

PROC TABULATE data=one;
  CLASS gender fulltime educ;
  TABLE (fulltime ALL) * gender ALL,
         educ * N ;
RUN;

```

		Educ			
		High School	Bachelors	Masters	Doctorate
		N	N	N	N
Fulltime	Gender				
Fulltime	Female		5.00		5.00
	Male	3.00	1.00	4.00	2.00
Parttime	Female	3.00	4.00	2.00	
	Male	3.00	1.00	3.00	
All	Female	3.00	9.00	2.00	5.00
	Male	6.00	2.00	7.00	2.00
All		9.00	11.00	9.00	7.00



Menambahkan Label

```
PROC TABULATE data=one;  
  CLASS gender fulltime;  
  VAR income;  
  TABLE gender = 'Gender' ALL = 'Total',  
          Fulltime = 'Employment Status' * income * mean = 'Mean' ;  
RUN;
```

	Employment Status	
	Fulltime	Parttime
	Income	Income
	Mean	Mean
Gender		
Female	58680.40	40867.83
Male	52596.56	51594.80
Total	55798.58	45743.73

```
PROC TABULATE data=one;  
  CLASS gender fulltime;  
  VAR income;  
  TABLE      gender ALL ,  
              Fulltime * income * mean ;  
  LABEL gender='Gender' Fulltime='Employment Status';  
  KEYLABEL mean='Mean' all='Total';  
RUN;
```



Menyembunyikan Label

```

PROC TABULATE data=one;
  CLASS educ gender fulltime;
  VAR income;
  TABLE educ ,
    Fulltime='Employment Status' *
    gender = ' ' *
    income *
    mean = ' ' ;
RUN;

```

	Employment Status			
	Fulltime		Parttime	
	Female	Male	Female	Male
	Income	Income	Income	Income
Educ				
High School		52506.67	44788.67	35600.00
Bachelors	42771.20		36947.00	
Masters		50729.25		62258.00
Doctorate	74589.60	56466.00		



Tabel Tiga Dimensi

```

PROC TABULATE data=one;
  CLASS gender fulltime educ;
  VAR income;
  TABLE educ='Education',
         fulltime = 'Employment Status',
         Gender * income * mean
         / BOX=_PAGE_ ;
RUN;

```

Education High School	Gender	
	Female	Male
	Income	Income
	Mean	Mean
Employment Status		
Fulltime		52506.67
Parttime	44788.67	35600.00

Education Doctorate	Gender	
	Female	Male
	Income	Income
	Mean	Mean
Employment Status		
Fulltime	74589.60	56466.00



Persentase

- PCTN, ROWPCTN, COLPCTN

```
PROC TABULATE data=one;  
  CLASS ethnic educ;  
  TABLE ethnic * PCTN,  
          Educ ;  
RUN;
```

		Educ			
		High School	Bachelors	Masters	Doctorate
Ethnic					
Asian	PctN	3.13	9.38		
Af.Amer.	PctN	6.25	3.13	6.25	3.13
Hisp.	PctN	6.25	3.13	6.25	6.25
Am. Ind.	PctN		6.25	3.13	3.13
White	PctN	9.38	9.38	9.38	6.25



Persentase

```
PROC TABULATE data=one;  
  CLASS ethnic educ;  
  TABLE ethnic * ROWPCTN,  
          Educ ;  
RUN;
```

		Educ			
		High School	Bachelors	Masters	Doctorate
Ethnic					
Asian	RowPctN	25.00	75.00		
Af. Amer.	RowPctN	33.33	16.67	33.33	16.67
Hisp.	RowPctN	28.57	14.29	28.57	28.57
Am. Ind.	RowPctN		50.00	25.00	25.00
White	RowPctN	27.27	27.27	27.27	18.18



MISSING VALUE

- Gunakan opsi MISSING dalam PROC TABULATE atau CLASS

```
PROC TABULATE data=one;  
  CLASS gender ethnic / MISSING;  
  VAR income;  
  TABLE ethnic ALL,  
           Income * (gender ALL) * n;  
RUN;
```





Selesai