

ورود داده

(SAS)

تدوین: مرکز تحلیل آماری خوارزمی

www.kharazmi-statistics.ir

مرکز آماری خوارزمی

سیستم SAS برای تولید یک مجموعه ی داده امکانات فراوانی دارد. در بسیاری از موارد داده های مورد نظر به طور خام، پیشاپیش و با روش های گوناگون ذخیره شده اند. SAS می تواند داده ها را در انواع پرونده ها، از جمله EXCEL, DARABASE, ASCII و LOTUS بخواند و به مجموعه های داده های SAS تبدیل کند. البته می توانیم داده ها را مستقیماً در محیط های گوناگونی که در SAS وجود دارد نیز وارد کنیم.

مطالبی که در این متن بیان شده است عبارتند از:

معرفی قوانین و محدودیت های زبان SAS

ورود داده

۱. روش مستقیم

۲. خواندن داده های پشت سر هم

۳. استفاده از حلقه ی DO

فراخوانی داده ها از منابع خارجی

معرفی قوانین و محدودیت های زبان SAS

زبان نرم افزار SAS زبان دشواری نیست. اما همچون دیگر نرم افزارها دارای قوانین و محدودیت هایی در برنامه نویسی است. این نرم افزار به بزرگی و کوچکی حروف حساس نیست. تمامی دستورات یا فرامین در SAS به "؛" (سمی کالن) ختم می شود. این علامت به معنای علامت نقطه "." در زبان انگلیسی است. عدم قرار دادن علامت "؛" در انتهای جملات به معنای عدم تکمیل و خاتمه ی فرمان است. یک گام داده با عبارت DATA و گاه رویه با عبارت PROC آغاز می شود. گام DATA شامل گزاره هایی است که یک مجموعه داده های SAS را برای تحلیل در گام های بعدی تولید می کند. گزاره های این گام بستگی به نوع ورودی دارد، هرچند در این گام همواره دو کلمه DATA و INPUT وجود دارد. گام PROC برای نوشتن دستورالعملهای SAS می باشد که با PROC یعنی مخفف PROCEDURE مشخص می شود که روی داده ها کاری انجام می شود. این قسمت با کلمه PROC شروع می شود. در هنگام ایجاد مجموع های داده و نیز متغیرها می بایست آنها را نام

گذاری کرده و هنگام فراخوانی از همان نام ها استفاده نمود. این نام ها نباید با عدد شروع شوند و نیز نباید دارای فاصله، خط فاصله و علامت درصد باشند. اگر فهرست دنباله ای از متغیرها مد نظر است می توان تنها به نام اولی و آخری اکتفا نمود و در میان آنها از خط فاصله یا علامت های "--" یا "_" استفاده نمود.

ورود داده

برای خواندن داده در نرم افزار SAS از سه روش زیر استفاده می شود.

۱. روش مستقیم:

این روش که عموماً از آن استفاده می شود به صورت زیر است:

در اینجا ۲۰ مشاهده از افراد مختلف آورده شده است. متغیرها شامل سن، قد و وزن افراد است. در این روش داده ها به طور مستقیم وارد نرم افزار می شود. در ادامه توضیحات مربوط به هر یک از خطوط و فرامین بیان شده است. دقت داشته باشید که انتهای هر یک از فرامین علامت ";" نوشته شود.

data kharazmi;
input age height weight;
datalines;
 23 159 59
 28 160 60
 26 175 59
 39 190 90
 40 158 59
 41 152 69
 49 169 67
 49 185 59
 50 172 53
 50 190 70
 52 149 85
 53 165 53
 53 149 59
 53 176 57
 54 184 64
 56 173 78
 56 180 62
 57 172 61
 57 170 59
 58 163 54
 ;
proc print data=kharazmi;
run;

در خط اول عبارت **data** نوشته شده و سپس نام فایل مورد نظر با توجه به محدودیت هایی که در بالا بیان شد تایپ می گردد.

در این خط نام متغیرها بیان می شود. ابتدا عبارت **input** تایپ شده و سپس نام تمامی متغیرها که با فاصله از هم جدا شده اند نوشته می شود. برای متغیرهای رشته ای بعد از اسم متغیر از علامت \$ استفاده می کنیم.

با نوشتن عبارت **datalines** گام داده به اتمام میرسد. در اینجا می توان به جای این عبارت از فرمان **cards** نیز استفاده کرد.


داده ها در این قسمت نوشته می شود. توجه داشته باشید که برای تمامی داده ها از یک ; استفاده می شود و این ; نشان دهنده اتمام گام داده است. هر داده شامل سه مشاهده است. هر داده در یک خط نوشته شده و مشاهدات به وسیله ی فاصله از هم جدا می شوند.

رویه ی **print** جهت فرمان دادن به نرم افزار برای نمایش داده ها است. و فرمان **run** جهت اجرای فرامین تایپ شده توسط کاربر.

پس از انجام مراحل بالا با طی مسیر Run > submit یا علامت  در نوار ابزار برنامه اجرا می شود و داده ها

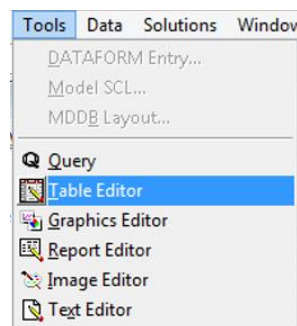
The SAS System			
Obs	age	height	weight
1	23	159	59
2	28	160	60
3	26	175	59
4	39	190	90
5	40	158	59
6	41	152	69
7	49	169	67
8	49	185	59
9	50	172	53
10	50	190	70
11	52	149	85
12	53	165	53
13	53	149	59
14	53	176	57
15	54	184	64
16	56	173	78
17	56	180	62
18	57	172	61
19	57	170	59
20	58	163	54

در صفحه ی output نمایش داده می شود. اگر روند صحیح پیش رود خروجی نرم افزار چیزی مشابه تصویر روبرو خواهد بود. در صورت ننوشتن رویه ی print داده ها در نرم افزار تعریف شده ولی نمایش و چاپ داده نمی شود.

جهت پاک کردن فرامین تایپ شده و یا خروجی های ارائه شده توسط نرم افزار در حالی که صفحه ی مورد را انتخاب کرده اید می توانید از منوی edit گزینه ی clear all و یا علامت  در منوی نوار ابزار استفاده نمایید.

در این روش نقاط "گمشده" (missing) را با نقطه "." مشخص می کنیم. در ضمن لازم به یادآوری است که متغیر های از نوع رشته ای با علامت \$ مشخص می شوند. در محیط SAS به هر واحد آزمایشی یک مشاهده (observation) و به هر یک از صفات مورد اندازه گیری، یک متغیر (variable) گفته می شود. داده هایی که در SAS خوانده می شود عبارتست از یک ماتریس عددی که سطرهای آن را تعداد واحدهای آزمایشی و ستون های آن را تعداد صفات مورد اندازه گیری تشکیل می دهند. دستور INPUT نام متغیرها و همچنین نوع آنها (متنی یا عددی) را تعیین می کند.

یکی دیگر از روش های ورود داده در این نرم افزار به روش مستقیم وزود داده در یک محیط جدولی است. برای ورود داده ها در SAS کافی است از منوی Tools زیر منوی Table Editor را انتخاب نمایید. پس کلیک بر روی Table Editor پنجره ای با محیط جدولی باز می شود که می توانید داده ها را در آن وارد کنید. با دبل کلیک کردن بر روی هر یک از سلول ها می توان محتوای آن را پر کرد.



برای حرکت در سلول ها باید از علامت های نشانگر و یا نشانگر موس استفاده نمایید. این محیط از نظر شناسایی کارکترهای رشته ای و عددی هوشمند است. یعنی اگر در سطرهای اول و دوم یک ستون متغیر عددی وارد نمایید و در سطر بعدی از متغیر رشته ای (حروفی) استفاده کنیم نرم افزار به صورت هوشمند متوجه

	gender	color
1	f	green
2	m	blue
3	m	green
4	f	green
5	m	blue

شده و تغییر رشته ای وارد کرده را نادیده گرفته و آن سلول را خالی می گذارد.

۲. خواندن داده های پشت سر هم

برای پشت سر هم خواندن داده ها می توانید از علامت @@ استفاده کنید. برای مثال داده های شامل ۱۰ مشاهده در نظر بگیرید. عبارات به صورت روبرو در نرم افزار نوشته می شود.

```
data amar;
input gender number @@;
datalines;
1 20 2 16 1 16 1 17 1 17 2 15 2 18 2 14 1 20 2 15
;
proc print data=amar;
run;
```

همانطور که در عبارت روبرو مشاهده می کنید در قسمت input نام دو مشاهده با فاصله آورده شده است و علامت @@ فرمان خواندن داده ها پشت سر هم است. و داده ها به صورتی که مشاهده می کنید نوشته می شود.

The SAS System		
Obs	gender	number
1	1	20
2	2	16
3	1	16
4	1	17
5	1	17
6	2	15
7	2	18
8	2	14
9	1	20
10	2	15

سایر عبارات در قسمت قبل توضیح داده شده است. خروجی به صورت روبرو است. ستون اول شماره مشاهده ستون دوم جنسیت و ستون سوم نمره کسب شده می باشد.

۳. استفاده از حلقه ی DO

برای وارد کردن داده های جدول، می توانیم از حلقه DO استفاده کنیم. به طور مثال فرض کنید که بخواهیم داده های جدول زیر را به نرم افزار معرفی کنیم:

I \ J	۱	۲	۳	۴	۵
۱	۱	۸	۹	۷	۶
۲	۹	۴	۴	۲	۵
۳	۶	۰	۱۴	۹	۳
۴	۴	۶	۳	۷	۱

برای وارد کردن اعداد این جدول، از دو حلقه ی DO به شکل زیر استفاده می کنیم.

```
data number;
```

```
  do i=1 to 5;
```

```
    do j=1 to 4;
```

```
      input response @@;
```

```
      output;
```

```
    end;
```

```
  end;
```

```
cards;
```

```
1 8 9 7 6 9 4 4 2 5 6 0 14 9 3 4 6 3 7 1
```

```
;
```

```
proc print data=number;
```

```
run;
```

The SAS System			
Obs	i	j	response
1	1	1	1
2	1	2	8
3	1	3	9
4	1	4	7
5	2	1	6
6	2	2	9
7	2	3	4
8	2	4	4
9	3	1	2
10	3	2	5
11	3	3	6
12	3	4	0
13	4	1	14
14	4	2	9
15	4	3	3
16	4	4	4
17	5	1	6
18	5	2	3
19	5	3	7
20	5	4	1

همانطور که در تصویر بالا خروجی نرم افزار را مشاهده می کنید نرم افزار عدد ۱ را در خانه ی (۱,۱) شناخته است و مثلا عدد (۲,۴) را عدد ۴ شناخته است.

این شکل نوشته (جلوتر و عقب تر نوشتن فرامین در خطوط) تنها برای مرتب کردن هر یک از حلقه ها است و تاثیری در روند محاسبات و فرامین ندارد.

کلیدهای عملگر

در ادامه کلید های عملگری که در نرم افزار sas تعریف شده است را بیان کرده ایم. با زدن دکمه ی F9 پنجره ی <DMKEYS>keys باز می شود که تمامی دکمه ی های عملگر را معرفی می کند.

Key	Definition
F1	help
F2	reshow
F3	end; /*gsubmit buffer=default*/
F4	recall
F5	wpgm
F6	log
F7	output
F8	zoom off;submit
F9	keys
F11	command focus
F12	
SHF F1	subtop
SHF F2	
SHF F6	
SHF F7	left
SHF F8	right
SHF F9	
SHF F10	wpopup
SHF F11	
SHF F12	
CTL F1	
CTL F2	
CTL F3	
CTL F11	
CTL F12	
ALT F1	
ALT F2	
ALT F3	
ALT F11	
ALT F12	
CTL B	libname
CTL D	dir
CTL E	clear
CTL G	
CTL H	help
CTL I	options
CTL J	
CTL K	cut
CTL L	log
CTL M	mark
CTL Q	filename
CTL R	rfind
CTL T	title
CTL U	unmark
CTL W	access
CTL Y	
RMB	wpopup
SHF RMB	
CTL RMB	
MMB	
SHF MMB	
CTL MMB	

تذکر: به جای استفاده از کلیدهای عملگری که در تصویر روبرو معرفی شده است می توان به طور مستقیم از دستورهای متناظر با آنها استفاده کرد. مثلا برای استفاده از راهنمای SAS می توان به جای فشار دادن کلید F1 با وارد کردن دستور HELP پنجره HELP را باز شده و می توان به جستجوی موارد مورد نظر پرداخت.

منبع:

- کاربرد نرم افزار SAS در تجزیه و تحلیل داده های آماری، عبدالامیر محیسی، دی ماه ۱۳۸۳، وزارت جهاد کشاورزی سازمان تحقیقات و آموزش کشاورزی مرکز تحقیقات کشاورزی و منابع طبیعی لرستان
- آموزش گام به گام SAS، رسول نصیری، انتشارات فرهنگی نشر گسترش، سال ۱۳۸۴
- موسسه فرهنگی دیجیتال بهگامان www.behkaman.ir
- آمار کاربردی و زبان برنامه نویسی، دکتر حمید رضا نواب پور، انتشارات رافع، ۱۳۷۹
- آموزش کاربردی SAS، ماهبانو تاتا و سید مرتضی موسوی، انتشارات پندار پارس، پاییز ۹۰