SAS- regression procedure

2013/10/04
FE522 고급금융계량분석
TA: 이은미

참고자료

- Proc REG Summary: Proc REG 문법과 옵션
  - http://www.psych.yorku.ca/lab/sas/sasregr.htm
- Annotated Output of Regressions: Proc REG 결과해석하기
  - http://www.ats.ucla.edu/stat/sas/output/reg.htm
- SAS Regression (UCLA webbook):
  - http://www.ats.ucla.edu/stat/sas/webbooks/reg/default.htm
- SAS online user guide
  - http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#reg_toc.htm
- 기타
  - 결과 그림 그래픽으로 출력 관련 참고: http://support.sas.com/rnd/base/topics/statgraph/192-31-updated.pdf

Example

DATA 입력 부분

```
data class;
      input Name $ Height Wght Age;
      datalines;
   Alfred   69.0 112.5 14
   Alice    56.5  84.0 13
   Barbara 65.3   98.0 13
   Carol    62.8 102.5 14
   Henry    63.5 102.5 14
   James    57.3  83.0 12
   Jane     59.8  84.5 12
   Janet    62.5 112.5 15
   Jeffrey 62.5   84.0 13
   John     59.0  99.5 12
   Joyce    51.3  50.5 11
   Judy     64.3  90.0 14
   Louise   56.3  77.0 12
   Mary     66.5 112.0 15
   Philip   72.0 150.0 16
   Robert   64.8 128.0 12
   Ronald   67.0 133.0 15
   Thomas   57.5  85.0 11
   William 66.5 112.0 15
   ;
```

Regression Procedure 명령문

```
proc reg data=class;
          model Wght = Height;
run;
quit;
```

소스코드 출처:
http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_introreg_sect003.htm

Proc REG (http://www.psych.yorku.ca/lab/sas/sasregr.htm)

```
PROC REG options;
        MODEL dependents=regressors / options;
        VAR variables;
        FREQ variable;
        WEIGHT variable;
        ID variable;
        OUTPUT OUT=SASdataset keyword=names...;
        PLOT yvariable*xvariable = symbol ...;
        RESTRICT linear_equation,...;
        TEST linear_equation,...;
        MTEST linear_equation,...;
        BY variables;
```

Reg procedure를 위한 데이터셋

- $y_t = \alpha_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \epsilon_t$ 꼴의 회귀분석
  - 여러 개의 y (여러 개의 포트폴리오 수익률 등)
  - 설명 가능성이 있는 여러 개의 변수 x1, x2, x3
- SAS dataset
  - Observation: $(y_{1,t}, \ldots, y_{p,t}, x_{1,t}, \ldots, x_{q,t})$ for t=1,2,…,T
  - 다음과 같은 형태의 dataset을 만든다. (xls 혹은 sas)

obs	y1	…	yp	X1	…	xq
1	…	…	…	…	…	…
2	…	…	…	…	…	…
3	…	…	…	…	…	…
4	…	…	…	…	…	…
…	…	…	…	…	…	…
T	…	…	…	…	…	…

  - 예)

VIEWTABLE: Work.Ds1

	datem	mkt_rf	SMB	HML	RF	small_low	small_2	small_high	big_low	big_2	big_high	lmkt	lsmb	lhml
1	1926-07-31	2,65	-2,39	-2,57	0,22	1,53	0,71	0,18	5,74	1,91	1,94	.	.	.
2	1926-08-31	2,59	-1,27	4,58	0,25	-0,59	2,39	5,57	2,75	2,68	5,74	2,65	-2,39	-2,57
3	1926-09-30	0,37	-1,25	-0,09	0,23	-2,32	-0,04	-0,47	1,42	0,09	-0,6	2,59	-1,27	4,58
4	1926-10-31	-3,45	-0,02	1,02	0,32	-3,83	-4,67	-1,47	-3,64	-2,33	-3,96	0,37	-1,25	-0,09
5	1926-11-30	2,43	-0,24	-0,63	0,31	2,92	3,81	1,8	3,26	2,85	3,13	-3,45	-0,02	1,02
6	1926-12-31	2,74	-0,12	0,16	0,28	2,4	1,86	3,27	2,87	2,69	2,33	2,43	-0,24	-0,63
7	1927-01-31	-0,19	-0,2	4,62	0,25	-0,75	-0,15	4,18	-0,23	0,01	4,09	2,74	-0,12	0,16
8	1927-02-28	4,17	0,04	3,04	0,26	3,19	6,97	6,49	5,12	3,51	7,91	-0,19	-0,2	4,62
9	1927-03-31	0,06	-1,7	-2,67	0,3	-1,76	-1,26	-2,92	1,68	-0,03	-2,5	4,17	0,04	3,04
10	1927-04-30	0,44	0,32	0,75	0,25	-1,16	0,3	3,85	3,14	-0,76	-0,36	0,06	-1,7	-2,67

회귀분석 돌려보기

* 가장 기본적인 regression procedure 문법 ;

```
proc reg data=ds1;
        model small_low = lmkt;
        model small_low = lmkt lhml;
run;
quit;
```

* Estimation 결과를 데이터셋으로 저장. Tableout option은 statistics를 저장;

```
proc reg data=ds1 outest=est tableout;
        model small_low = lmkt lhml;
run;
quit;
```

회귀분석 돌려보기

```
* 여러 개의 설명변수 포함;
proc reg data=ds1;
        model small_low big_high = lmkt;
        model small_low big_high = lmkt lsmb;
        model small_low big_high = lmkt lhml;
run;
quit;
```

```
* Option 활용, regression 결과 residual 저장하기,
ds2: ds1의 columns와 'resid_bh'(label은 Residual) column 포함 ;
proc reg data=ds1;
        model big_high = lmkt;
        output out=ds2 r=resid_bh;
run;
quit;
```

(Optional) Dataset 다루기

```
* data merge 예시;
data ds1;
        merge fffactor ffret(rename=(date=datem));
        datem = intnx('month',datem,0,'e');
        lmkt = lag(mkt_rf);
        lsmb = lag(smb);
        lhml = lag(hml);
        by datem;
run;
```