

SAS program steps

All SAS programs consist of a sequence of "steps". There are only two kinds of steps:

DATA step

A DATA step creates a SAS dataset (a collection of data together with a "data dictionary", which defines the variables and their properties). Data must be in the form of a SAS dataset before it can be analyzed by SAS procedures.

In the example SAS program, these lines create the dataset CLASS from raw data input:

```
DATA CLASS;
  INPUT NAME $ SEX $ AGE HEIGHT WEIGHT;
  CARDS;
JOHN      M 12 59.0  99.5
JAMES     M 12 57.3  83.0
... (more data lines)
```

PROC step

A PROCedure step calls a SAS procedure to analyse or process a SAS dataset.

In the example SAS program, these lines call two SAS procedures to analyze the CLASS dataset:

```
PROC PRINT;
PROC MEANS;
  VARIABLES HEIGHT WEIGHT;
```

A SAS program can contain any number of DATA and PROC steps. The SAS statements in each step are executed all together. Once a dataset has been created, it can be processed by any subsequent DATA or PROC step.

SAS statements

- All SAS statements start with a **keyword** (DATA, INPUT, PROC, etc.)
- **All SAS statements end with a semicolon (;)**. (The most common problem students encounter is omitting a semicolon -- SAS thinks that two statements are just one.)

```
data ersrdmtr;
input edmtr;
cards;
0.272
0.274
0.272
0.272
0.272
0.272
0.275
0.274
0.269
0.273
0.273
0.269
0.269
;
```

```
proc ttest data = ersrdmtr H0=0.27;
var edmtr;
run;
```

=====

```
data childcare;
input hrsperweek;
cards;
37
24
39
38
31
35
26
35
31
32
40
36
36
29
35
;
```

```
proc univariate data = childcare loccount mu0=30;
var hrsperweek;
run;
```

```

data rightleft;
input right left;
diff = right-left;
cards;
113 137
105 105
130 133
101 108
138 115
118 170
87 103
116 145
75 78
96 107
122 84
103 148
116 147
107 87
118 166
103 146
111 123
104 135
111 112
89 93
78 76
100 116
89 78
85 101
88 123
;

proc ttest data = rightleft;
paired right*left;
run;

proc univariate data = rightleft loccount mu0=0;
var diff;
run;

```

NOTES:

PROC UNIVARIATE calculates the sign test statistic as

$$M = (n^+ - n^-) / 2$$

where n^+ is the number of values that are greater than μ_0 , and n^- is the number of values that are less than μ_0 . Values equal to μ_0 are discarded.

The signed rank statistic S is computed as

$$S = \sum r_i^+ - n_t(n_t + 1) / 4$$

where r_i^+ is the rank of $|x_i - \mu_0|$ after discarding values of $x_i = \mu_0$, and n_t is the number of values not equal to μ_0 , and the sum is calculated for values of $x_i - \mu_0$ greater than 0. Average ranks are used for tied values.

```
data twosample;
input strgnth group $;
cards;
```

```
8      YES-CHR
13     YES-CHR
16     YES-CHR
22     YES-CHR
24     YES-CHR
25     YES-CHR
28     YES-CHR
32     YES-CHR
6      NO-CHR
7      NO-CHR
12     NO-CHR
17     NO-CHR
19     NO-CHR
20     NO-CHR
23     NO-CHR
30     NO-CHR
;
```

```
proc ttest data = twosample;
class group;
var strgnth;
run;
```

```
=====
```

```
data twosampleb;
input volume group $;
cards;
```

```
604    1
646    1
688    1
760    1
793    1
806    1
856    1
899    1
906    1
968    1
739    2
797    2
812    2
819    2
843    2
850    2
909    2
941    2
961    2
979    2
;
```

```
proc nparlway data = twosampleb wilcoxon;
class group;
var volume;
exact;
run;
```

```

data slr;
input hostility chol;
cards;
14      195
8       185
19      286
5       150
16      250
10      170
;

proc reg data = slr;
model chol = hostility / r influence clm cli clb;
run;

```

=====

```

data concrete;
input X1 X2 Y;
X1X2 = X1*X2;
cards;
21      0.65      33.55
21      0.55      47.55
7       0.65      35.00
7       0.55      35.90
28      0.60      40.90
0       0.60      39.10
14      0.70      31.55
14      0.50      48.00
14      0.60      42.30
;

proc reg data = concrete;
model y = x1 x2 x1x2 /r influence clm cli clb;
run;

```

The TTEST Procedure

Statistics

Variable	N	Lower CL	Mean	Upper CL	Lower CL	Std Dev	Upper CL	Std Err
edmtx	13	0.2708	0.272	0.2732	0.0014	0.002	0.0032	0.0005

T-Tests

Variable	DF	t Value	Pr > t
edmtx	12	3.68	0.0031

The UNIVARIATE Procedure

Variable: hrspwweek

Moments

N	15	Sum Weights	15
Mean	33.6	Sum Observations	504
Std Deviation	4.67210567	Variance	21.8285714
Skewness	-0.714141	Kurtosis	-0.1868991
Uncorrected SS	17240	Corrected SS	305.6
Coeff Variation	13.9050764	Std Error Mean	1.2063325

Basic Statistical Measures

Location	Mean	Std Deviation	Variability
	33.60000	4.67211	
	Median	35.00000	21.82857
	Mode	35.00000	18.00000
		Interquartile Range	6.00000

Tests for Location: Mu0=30

Test	-Statistic-	-p Value-
Student's t	t 2.984252	Pr > t 0.0099
Sigm	N 4.5	Pr >= M 0.0352
Signed Rank	S 43	Pr >= S 0.0121

Location Counts: Mu0=30.00

Count	Value
Num Obs > Mu0	12
Num Obs = Mu0	15
Num Obs < Mu0	3

Quantiles (Definition 5)

Quantile	Estimate
100% Max	40
95%	40
90%	40
75% Q3	39
50% Median	35
25% Q1	31
5%	24
1%	24
0% Min	24

Extreme Observations

Value	Obs	Value	Obs
24	2	36	13
26	7	37	1
29	14	38	4
31	9	39	3
31	5	40	11

The TTEST Procedure

Statistics

	Lower CL	Upper CL	Lower CL	Upper CL
Difference	N	Mean	Mean	Std Dev
right - left	25	-22.79	-13.32	17.909
				22.936
				31.907
				4.5672

T-Tests

Difference	DF	t Value	Pr > t
right - left	24	-2.90	0.0078

The UNIVARIATE Procedure

Variable: diff

Moments

	25	Sum Weights	25
N	-13.32	Sum Observations	-332
Mean	22.935979	Variance	526.06
Std Deviation	0.23145165	Kurtosis	-0.2310687
Skewness	17061	Corrected SS	12625.44
Uncorrected SS	-172.19218	Std Error Mean	4.58719958
Coef Variation			

Basic Statistical Measures

Location	Mean	Std Deviation	22.93600
	-13.3200	Variance	526.06000
	-12.0000	Range	90.00000
	-16.0000	Interquartile Range	30.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t -2.90373	Pr > t 0.0078
Sign	M -7	Pr >= M 0.0065
Signed Rank	S -93.5	Pr >= S 0.0048

Location Counts: Mu0=0.00

Count	Value
Num Obs > Mu0	5
Num Obs = Mu0	24
Num Obs < Mu0	19

The TTEST Procedure

Statistics

	Lower CL	Upper CL	Lower CL	Upper CL
Variable	N	Mean	Mean	Std Dev
group	8	9.9471	16.75	23.553
NO-CHR	8	14.267	21	27.733
CHR	8	-12.93	-4.25	4.4315
Diff (1-2)				5.9269
				8.0954
				12.767
				4.0477

T-Tests

Variable	Method	Variances	DF	t Value	Pr > t
group	Pooled	Equal	14	-1.05	0.3115
group	Satterthwaite	Unequal	14	-1.05	0.3115

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
group	Folded F	7	7	1.02	0.9789

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable volume

Classified by Variable group

group	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	10	86.0	105.0	13.228757	8.60
2	10	124.0	105.0	13.228757	12.40

Wilcoxon Two-Sample Test

Statistic (S)	86.0000
Normal Approximation	
Z	-1.3985
One-Sided Pr < 2	0.0810
Two-Sided Pr > Z	0.1620
t Approximation	
One-Sided Pr < 2	0.0890
Two-Sided Pr > Z	0.1781

Exact Test

One-Sided Pr <= S	0.0827
Two-Sided Pr >= S - Mean	0.1655

Z includes a continuity correction of 0.5.

Kruskal-Wallis Test

Chi-Square	2.0629
DF	1
Pr > Chi-Square	0.1509

The REG Procedure
Model: MODEL1
Dependent Variable: chol
Number of Observations Read 6
Number of Observations Used 6

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	1	11541	11541	25.90
Error	4	1789.1045	447.27536	0.0071
Corrected Total	5	13330		

Root MSE 21.14889 R-Square 0.8658
Dependent Mean 206.00000 Adj R-Sq 0.8322
Coef Var 10.26645

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	95% Confidence Limits
Intercept	1	96.26087	23.26517	4.14	0.0144	31.66641 160.85533
hostility	1	9.14493	1.80031	5.08	0.0071	4.14646 14.14340

Output Statistics

Obs	Variable	Dependent Value	Predicted Value	Std Error	95% CI Mean	95% CI Predict	Residual
1	195.0000	224.2899	9.3547	198.3170	250.2627	160.0834	288.4964
2	185.0000	169.4203	11.2429	138.2049	200.6357	102.9200	235.9206
3	286.0000	270.0145	15.2762	227.6011	312.4279	197.5788	342.4492
4	150.0000	141.9855	15.2762	99.5721	184.3989	69.5508	214.4302
5	230.0000	242.5797	11.2429	211.3643	273.7951	176.0734	309.0800
6	170.0000	187.7101	9.3547	161.7373	213.6830	123.5036	251.9166

Output Statistics

Obs	Std Error	Student	Cook's D	Hat	Diag	Cov	Ratio	DFBETs
1	18.967	-1.544	***	0.290	-2.1044	0.1957	0.3605	-1.0379
2	17.913	0.870	*	0.149	0.8365	0.2826	1.6284	0.5250
3	14.626	1.093	**	0.652	1.1302	0.5217	1.8285	1.1805
4	14.626	0.548	*	0.154	0.4934	0.5217	0.5154	0.5154
5	17.913	0.414	*	0.034	0.3667	0.2826	2.2701	0.2302
6	18.967	-0.934	*	0.106	-0.9144	0.1957	1.3518	-0.4510

Output Statistics

Obs	Intercept	hostility
1	0.0155	-0.3395
2	-0.4619	-0.3363
3	-0.6567	-0.9739
4	0.5029	-0.4252
5	-0.0713	-0.1474
6	-0.3156	0.1736

Sum of Residuals 0
Sum of Squared Residuals 1789.1045
Predicted Residual SS (PRESS) 3787.42120

The REG Procedure
Model: MODEL1
Dependent Variable: Y
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	249.17583	83.05861	14.15	0.0071
Error	5	29.34917	5.86983		
Corrected Total	8	278.52500			

Root MSE 2.42277 R-Square 0.8946
Dependent Mean 39.31667 Adj R-Sq 0.8314
Coef Var 6.16221

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	95% Confidence Limits
Intercept	1	6.21667	30.30351	0.21	0.8455	-71.68099 84.11432
X1	1	5.77857	2.07907	2.78	0.0389	0.43416 11.12296
X2	1	51.33333	50.43406	1.02	0.3355	180.97621 -0.46009
X1X2	1	-9.35714	3.46111	-2.70	0.0426	-18.25420 0.00000

Output Statistics

Obs	Variable	Dependent Value	Predicted Value	Std Error	95% CI Mean	95% CI Predict	Residual
1	33.5500	33.2083	1.7601	28.6838	37.7328	25.5104	40.9663
2	47.5500	47.7250	1.7601	43.2005	52.2495	40.0271	55.4229
3	35.0000	37.4883	1.7601	32.9338	29.7604	45.1563	-2.4583
4	35.0000	38.8750	1.7601	34.3505	43.3995	31.1771	46.5729
5	40.9000	41.6167	1.6152	37.4647	45.7686	34.1316	49.1017
6	39.1000	37.0267	1.6152	32.8647	41.1686	29.5316	44.5017
7	31.5500	31.3500	1.6152	27.1980	35.5020	23.8649	38.8331
8	48.0000	47.2833	1.6152	43.1314	51.4353	39.7983	54.7664
9	42.5000	39.3167	0.8076	37.2407	41.3926	32.7518	45.8815

Output Statistics

Obs	Std Error	Student	Cook's D	Hat	Diag	Cov	Ratio	DFBETs
1	1.665	0.205	*	0.012	0.1843	0.5278	4.9980	0.1949
2	1.665	-0.105	*	0.003	-0.0941	0.5278	5.1245	-0.0995
3	1.665	-1.477	**	0.609	-1.7587	0.5278	0.5228	-2.8106
4	1.665	-1.787	**	0.032	-2.6586	0.5278	0.0882	-2.9226
5	1.806	-0.397	*	0.032	-0.3607	0.4444	3.8694	1.0774
6	1.806	1.154	*	0.266	1.2046	0.4444	1.2742	1.0774
7	1.806	0.111	*	0.002	0.0992	0.4444	4.3516	0.0887
8	1.806	0.397	*	0.032	0.3607	0.4444	3.8694	0.3226
9	2.284	1.306	*	0.053	1.4392	0.1111	0.5175	0.5088

Output Statistics

Obs	Intercept	X1	X2	X1X2
1	0.1960	-0.1392	-0.1674	0.1341
2	0.8554	-0.0763	-0.0548	0.0685
3	1.3754	-1.2428	-1.4343	1.2790
4	-2.2511	2.1683	2.1683	-1.9046
5	0.0986	-0.0134	0.0986	-0.0080
6	0.0986	-0.0448	0.0986	-0.0080
7	-0.8281	-0.0080	0.0986	-0.0080
8	0.0986	-0.0080	0.0986	-0.0080
9	0.0986	-0.0080	0.0986	-0.0080

Sum of Residuals 0
Sum of Squared Residuals 29.34917
Predicted Residual SS (PRESS) 98.23676