

tuzo > less math0.log

1
20:35 Tuesday, August 30, 2005

The SAS System

NOTE: Copyright (c) 1999-2001 by SAS Institute Inc., Cary, NC, USA.

NOTE: SAS (r) Proprietary Software Release 8.2 (TS2M0)

Licensed to UNIVERSITY OF TORONTO/COMPUTING & COMMUNICATIONS, Site 0008987001.

NOTE: This session is executing on the SunOS 5.6 platform.

This message is contained in the SAS news file, and is presented upon initialization. Edit the files "news" in the "misc/base" directory to display site-specific news and information in the program log. The command line option "-nonews" will prevent this display.

NOTE: SAS initialization used:

real time 0.78 seconds
cpu time 0.19 seconds

```
1           title 'Math Diagnostic Study: Describe Exploratory Data';  
2           title2 'Bare bones';  
3           options linesize=79;  
4           data mathex;  
5                 infile 'mexplore.dat';  
6                 input id sex $ tongue nation1 nation2  
7                         gpa english finmat alggeo hscalc  
8                         q1-q20  
9                         course grade;  
10
```

NOTE: The infile 'mexplore.dat' is:

File Name=/student/etsta313/442f05/math/mexplore.dat,
Owner Name=etsta313,Group Name=nstudent,
Access Permission=rw-r--r--,
File Size (bytes)=43919

NOTE: 579 records were read from the infile 'mexplore.dat'.

The minimum record length was 64.

The maximum record length was 81.

NOTE: The data set WORK.MATHEX has 579 observations and 32 variables.

NOTE: DATA statement used:

real time 0.39 seconds
cpu time 0.12 seconds

```
11           proc freq;  
12                 tables _all_;  
13
```

NOTE: There were 579 observations read from the data set WORK.MATHEX.

NOTE: The PROCEDURE FREQ printed pages 1-27.

NOTE: PROCEDURE FREQ used:

real time 0.50 seconds
cpu time 0.31 seconds

NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414

NOTE: The SAS System used:

real time 2.50 seconds

2 The SAS System 20:35 Tuesday, August 30, 2005

cpu time 0.64 seconds

tuzo > cat math0.lst

Math Diagnostic Study: Describe Exploratory Data 1
Bare bones 20:35 Tuesday, August 30, 2005

The FREQ Procedure

id	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1	0.17	1	0.17
2	1	0.17	2	0.35
3	1	0.17	3	0.52
4	1	0.17	4	0.69

... skipping ...

570	1	0.17	570	98.45
571	1	0.17	571	98.62
572	1	0.17	572	98.79

Math Diagnostic Study: Describe Exploratory Data 12
Bare bones 20:35 Tuesday, August 30, 2005

The FREQ Procedure

id	Frequency	Percent	Cumulative Frequency	Cumulative Percent
573	1	0.17	573	98.96
574	1	0.17	574	99.14
575	1	0.17	575	99.31
576	1	0.17	576	99.48
577	1	0.17	577	99.65
578	1	0.17	578	99.83
579	1	0.17	579	100.00

sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Female	266	48.28	266	48.28
Male	285	51.72	551	100.00

Frequency Missing = 28

tongue	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	402	72.96	402	72.96
3	149	27.04	551	100.00

Frequency Missing = 28

nation1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	73	13.30	73	13.30
2	2	0.36	75	13.66
3	12	2.19	87	15.85
4	16	2.91	103	18.76
5	23	4.19	126	22.95
6	61	11.11	187	34.06
7	35	6.38	222	40.44
8	96	17.49	318	57.92
9	7	1.28	325	59.20
10	22	4.01	347	63.21
11	8	1.46	355	64.66
12	11	2.00	366	66.67
13	25	4.55	391	71.22
14	64	11.66	455	82.88
15	2	0.36	457	83.24
16	72	13.11	529	96.36
17	2	0.36	531	96.72
18	18	3.28	549	100.00

Frequency Missing = 30

The FREQ Procedure

nation2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	96	17.49	96	17.49
2	1	0.18	97	17.67
3	8	1.46	105	19.13
4	18	3.28	123	22.40
5	8	1.46	131	23.86
6	65	11.84	196	35.70
7	27	4.92	223	40.62
8	95	17.30	318	57.92
9	11	2.00	329	59.93
10	29	5.28	358	65.21
11	9	1.64	367	66.85
12	13	2.37	380	69.22
13	12	2.19	392	71.40
14	57	10.38	449	81.79
15	2	0.36	451	82.15
16	86	15.66	537	97.81
17	5	0.91	542	98.72
18	7	1.28	549	100.00

Frequency Missing = 30

The FREQ Procedure

gpa	Frequency	Percent	Cumulative Frequency	Cumulative Percent
65	1	0.21	1	0.21
66	2	0.43	3	0.64
67.3	1	0.21	4	0.86
67.7	1	0.21	5	1.07

... skipping ...

94.3	1	0.21	462	99.14
95	1	0.21	463	99.36
95.8	1	0.21	464	99.57
96.2	1	0.21	465	99.79
97.3	1	0.21	466	100.00

Frequency Missing = 113

The FREQ Procedure

english	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	9	1.84	9	1.84
50	2	0.41	11	2.25
51	1	0.20	12	2.45
52	1	0.20	13	2.66
53	1	0.20	14	2.86
54	3	0.61	17	3.48

... skipping ...

94	1	0.20	487	99.59
95	1	0.20	488	99.80
96	1	0.20	489	100.00

Frequency Missing = 90

The FREQ Procedure

finmat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
50	4	1.41	4	1.41
51	3	1.06	7	2.46
52	2	0.70	9	3.17
53	1	0.35	10	3.52
54	1	0.35	11	3.87

... skipping ...

99	3	1.06	283	99.65
100	1	0.35	284	100.00

The FREQ Procedure

Frequency Missing = 295

The FREQ Procedure

alggeo	Frequency	Percent	Cumulative Frequency	Cumulative Percent
50	14	4.03	14	4.03
51	5	1.44	19	5.48

... skipping ...

98	3	0.86	345	99.42
100	2	0.58	347	100.00

Frequency Missing = 232

The FREQ Procedure

hscalc	Frequency	Percent	Cumulative Frequency	Cumulative Percent
50	10	2.23	10	2.23
51	8	1.79	18	4.02
52	3	0.67	21	4.69
53	5	1.12	26	5.80
54	2	0.45	28	6.25

... skipping ...

100	1	0.22	448	100.00
-----	---	------	-----	--------

The FREQ Procedure

Frequency Missing = 131

q1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	151	31.46	151	31.46
1	329	68.54	480	100.00

Frequency Missing = 99

q2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	34	7.08	34	7.08
1	446	92.92	480	100.00

Frequency Missing = 99

q3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	369	76.88	369	76.88
1	111	23.13	480	100.00

Frequency Missing = 99

... skipping ...

q20	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	334	69.58	334	69.58
1	146	30.42	480	100.00

Frequency Missing = 99

course	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	65	11.23	65	11.23
2	429	74.09	494	85.32
3	36	6.22	530	91.54
4	49	8.46	579	100.00

The FREQ Procedure

grade	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1	0.25	1	0.25
2	2	0.51	3	0.76
... skipping ...				
44	2	0.51	71	18.07
45	2	0.51	73	18.58
46	8	2.04	81	20.61
47	6	1.53	87	22.14
48	1	0.25	88	22.39
50	28	7.12	116	29.52
51	11	2.80	127	32.32
52	5	1.27	132	33.59
53	6	1.53	138	35.11
54	11	2.80	149	37.91
55	10	2.54	159	40.46
... skipping ...				
96	1	0.25	391	99.49
97	1	0.25	392	99.75
99	1	0.25	393	100.00

Frequency Missing = 186

We really need some labels

```
tuzo > cp math0.sas math1.sas
tuzo > ls
math0.log      math0.sas      mexplore.dat
math0.lst      math1.sas      mreplc.dat
tuzo > emacs math1.sas

tuzo > cat math1.sas
```

```

/* math1.sas */
title 'Math Diagnostic Study: Describe Exploratory Data';
options linesize=79 pagesize=1000 noovp formdlim='_';

proc format;
  value rwfmt 0 = 'Wrong' 1 = 'Right';
  value crsfmt 4 = 'No Resp';
  value langfmt 1 = 'English' 2 = 'French' 3 = 'Other';
  value ynfmt 0 = 'No' 1 = 'Yes';
  value natfmt
    1 = 'Chinese'
    2 = 'Japanese'
    3 = 'Korean'
    4 = 'Vietnamese'
    5 = 'Other Asian'
    6 = 'Eastern European'
    7 = 'Hispanic'
    8 = 'English-speaking'
    9 = 'French'
    10 = 'Italian'
    11 = 'Greek'
    12 = 'Germanic'
    13 = 'Other European'
    14 = 'Middle-Eastern'
    15 = 'Pakistani'
    16 = 'East Indian'
    17 = 'Sub-Saharan'
    18 = 'OTHER' ;

data mathex;
  infile 'mexplore.dat';
  input id sex $ tongue nation1 nation2
        gpa english finmat alggeo hscalc
        q1-q20
        course grade;

  precalc1 = sum(of q1-q4);
  precalc2 = sum(of q5-q9);
  calcone = sum(of q10-q14);
  calctwo = sum(of q15-q20);
  precalc = precalc1 + precalc2;
  calc = calcone + calctwo;
  totscore = precalc+calc;
label
  tongue = 'Mother Tongue'
  nation1 = 'Nationality of name acc to rater1'
  nation2 = 'Nationality of name acc to rater2'
  gpa = 'High School GPA'
  english = 'Mark in HS English'
  finmat = 'Mark in HS Finite Math'
  alggeo = 'Mark in HS Algebra/Geometry'

```

```

hscalc = 'Mark in HS Calculus'
precalc1 = 'Precalculus 1 (bc1) subscale'
precalc2 = 'Precalculus 2 (bc2) subscale'
precalc = 'Number precalculus correct'
calcone = 'Calculus 1 (c1) subscale'
calctwo = 'Calculus 2 (c2) subscale'
calc = 'Number calculus correct'
totalscore = 'Total # right on diagnostic test'
course = 'Which university calculus course'
grade = 'Mark in university calculus';

```

```
/* Associate variables with printing formats */
```

```
format q1-q20 rwfmt.;
format course crsfmt.;
format tongue langfmt.;
format nation1 nation2 natfmt.;
```

```
if (50<=grade<=100) then passed=1; else passed=0;
label passed = 'Passed the course';
format passed ynfmt.;
```

```
proc freq;
  tables _all_;
```

```
tuzo > sas math1
tuzo > cat math1.lst
```

Just show a part of the output this time.

Mother Tongue				
tongue	Frequency	Percent	Cumulative Frequency	Cumulative Percent
English	402	72.96	402	72.96
Other	149	27.04	551	100.00

Frequency Missing = 28

Nationality of name acc to rater1

nation1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Chinese	73	13.30	73	13.30
Japanese	2	0.36	75	13.66
Korean	12	2.19	87	15.85
Vietnamese	16	2.91	103	18.76
Other Asian	23	4.19	126	22.95
Eastern European	61	11.11	187	34.06
Hispanic	35	6.38	222	40.44
English-speaking	96	17.49	318	57.92
French	7	1.28	325	59.20
Italian	22	4.01	347	63.21
Greek	8	1.46	355	64.66
Germanic	11	2.00	366	66.67
Other European	25	4.55	391	71.22
Middle-Eastern	64	11.66	455	82.88
Pakistani	2	0.36	457	83.24
East Indian	72	13.11	529	96.36
Sub-Saharan	2	0.36	531	96.72
OTHER	18	3.28	549	100.00

Frequency Missing = 30

Nationality of name acc to rater2

nation2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Chinese	96	17.49	96	17.49
Japanese	1	0.18	97	17.67
Korean	8	1.46	105	19.13
Vietnamese	18	3.28	123	22.40
Other Asian	8	1.46	131	23.86
Eastern European	65	11.84	196	35.70
Hispanic	27	4.92	223	40.62
English-speaking	95	17.30	318	57.92
French	11	2.00	329	59.93
Italian	29	5.28	358	65.21
Greek	9	1.64	367	66.85
Germanic	13	2.37	380	69.22
Other European	12	2.19	392	71.40
Middle-Eastern	57	10.38	449	81.79
Pakistani	2	0.36	451	82.15
East Indian	86	15.66	537	97.81
Sub-Saharan	5	0.91	542	98.72
OTHER	7	1.28	549	100.00

Frequency Missing = 30

There is still a problem with mark in HS English.

Mark in HS English

english	Frequency	Percent	Cumulative Frequency	Cumulative Percent
50	4	0.41	4	0.41
51	2	0.21	6	0.62
52	1	0.10	7	0.73
53	2	0.21	9	0.93
54	6	0.62	15	1.56

... skipping ...

q1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Wrong	151	31.46	151	31.46
Right	329	68.54	480	100.00

Frequency Missing = 99

q2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Wrong	34	7.08	34	7.08
Right	446	92.92	480	100.00

Frequency Missing = 99

... skipping ...

Cumulative Cumulative

Which university calculus course

course	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	65	11.23	65	11.23
2	429	74.09	494	85.32
3	36	6.22	530	91.54
No Resp	49	8.46	579	100.00

Mark in university calculus

grade	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1	0.25	1	0.25
2	2	0.51	3	0.76
4	2	0.51	5	1.27

... skipping ...

Passed the course

passed	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	274	47.32	274	47.32
Yes	305	52.68	579	100.00

Based on this output, we want to

- Recode english zero to missing
- For each math course, note whether the student took it or not.
- Count number of HS math courses

Once this is done, concatenate the data files for cleaning only .

```
tuzo > ls
math0.log  math0.sas  math1.lst  mexplore.dat
math0.lst  math1.log  math1.sas  mreplc.dat
tuzo > cat mexplore.dat mreplc.dat > mboth.dat
tuzo > cp math1.sas math2.sas
tuzo > emacs math2.sas

tuzo > cat math2.sas
```

```

/* math2.sas */
title 'Math Diagnostic Study: Describe Pooled Data';
options linesize=79 pagesize=1000 noovp formdlim='_';

proc format;
  value rwfmt 0 = 'Wrong' 1 = 'Right';
  value crsfmt 4 = 'No Resp';
  value langfmt 1 = 'English' 3 = 'Other';
  value ynfmt 0 = 'No' 1 = 'Yes';
  value natfmt
    1 = 'Chinese'
    2 = 'Japanese'
    3 = 'Korean'
    4 = 'Vietnamese'
    5 = 'Other Asian'
    6 = 'Eastern European'
    7 = 'Hispanic'
    8 = 'English-speaking'
    9 = 'French'
    10 = 'Italian'
    11 = 'Greek'
    12 = 'Germanic'
    13 = 'Other European'
    14 = 'Middle-Eastern'
    15 = 'Pakistani'
    16 = 'East Indian'
    17 = 'Sub-Saharan'
    18 = 'OTHER' ;

data mathboth;
  infile 'mboth.dat'; /* Read pooled data for cleaning */
  input id sex $ tongue nation1 nation2
        gpa english finmat alggeo hscal
        q1-q20
        course grade;
  /* Check whether credit for HS math courses */
  if 0 <= finmat <= 100 then credfm = 1; else credfm=0;
  if 0 <= alggeo <= 100 then credag = 1; else credag=0;
  if 0 <= hscal <= 100 then credcalc = 1; else credcalc=0;
  nhsmath = credfm+credag+credcalc;
  /* Diagnostic test subscales */
  precalc1 = sum(of q1-q4);
  precalc2 = sum(of q5-q9);
  calccone = sum(of q10-q14);
  calctwo = sum(of q15-q20);
  precalc = precalc1 + precalc2;
  calc = calccone + calctwo;
  totscore = precalc+calc;
  if english = 0 then english = .; /* Zero means mark not available */

```

```

label
  tongue = 'Mother Tongue'
  nation1 = 'Nationality of name acc to rater1'
  nation2 = 'Nationality of name acc to rater2'
  gpa = 'High School GPA'
  english = 'Mark in HS English'
  finmat = 'Mark in HS Finite Math'
  alggeo = 'Mark in HS Algebra/Geometry'
  hscalc = 'Mark in HS Calculus'
    credfm = 'Credit for (took?) Finite math'
    credag = 'Credit for (took?) Algebra/geometry'
    credcalc = 'Credit for (took?) HS Calculus'
    nhsmath = 'Number of HS math courses'
  precalc1 = 'Precalculus 1 (bc1) subscale'
  precalc2 = 'Precalculus 2 (bc2) subscale'
  precalc = 'Number precalculus correct'
  calcone = 'Calculus 1 (c1) subscale'
  calctwo = 'Calculus 2 (c2) subscale'
  calc = 'Number calculus correct'
  totscore = 'Total # right on diagnostic test'
  course = 'Which university calculus course'
  grade = 'Mark in university calculus';

/* Associate variables with printing formats */

format q1-q20 rwfmt.;
format course crsfmt.;
format tongue langfmt.;
format nation1 nation2 natfmt.;
  format credfm credag credcalc ynfmt.;

if (50<=grade<=100) then passed=1; else passed=0;
label passed = 'Passed the course';
format passed ynfmt.;

proc freq;
  tables _all_;

```

```
tuzo > cat math2.lst
```

Math Diagnostic Study: Describe Pooled Data 1
 21:38 Tuesday, August 30, 2005

The FREQ Procedure

id	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	2	0.17	2	0.17

... skipping ...

Mark in HS English

english	Frequency	Percent	Cumulative Frequency	Cumulative Percent
50	4	0.41	4	0.41
51	2	0.21	6	0.62
52	1	0.10	7	0.73

... skipping ...

Mark in HS Finite Math

finmat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
48	1	0.18	1	0.18
50	6	1.07	7	1.24
51	3	0.53	10	1.78

... skipping ...

Mark in HS Calculus

hscalc	Frequency	Percent	Cumulative Frequency	Cumulative Percent
43	1	0.11	1	0.11
50	21	2.33	22	2.44
51	13	1.44	35	3.88
52	5	0.55	40	4.43
53	9	1.00	49	5.43

... skipping ...

Credit for (took?) Finite math

credfm	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	595	51.38	595	51.38
Yes	563	48.62	1158	100.00

Credit for (took?) Algebra/geometry

credag	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	471	40.67	471	40.67
Yes	687	59.33	1158	100.00

Credit for (took?) HS Calculus

credcalc	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	255	22.02	255	22.02
Yes	903	77.98	1158	100.00

Number of HS math courses

nhsmath	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	197	17.01	197	17.01
1	114	9.84	311	26.86
2	502	43.35	813	70.21
3	345	29.79	1158	100.00