

# Reading and Checking the Math Data

```
tuzo.utm.utoronto.ca:~/test% ls
checkmath.sas
tuzo.utm.utoronto.ca:~/test% curl
http://fisher.utstat.toronto.edu/~brunner/442s08/code_n_data/lecture/exploremath.data >
exploremath.data
```

```
  % Total      % Received % Xferd  Average Speed   Time    Time       Time  Current
                        Dload  Upload    Total   Spent    Left   Speed
100 44004  100 44004    0     0   151k      0  --:--:--  --:--:--  --:--:--  8367k
```

```
tuzo.utm.utoronto.ca:~/test% ls
checkmath.sas  exploremath.data
tuzo.utm.utoronto.ca:~/test% sas checkmath
tuzo.utm.utoronto.ca:~/test% ls
checkmath.log  checkmath.lst  checkmath.sas  exploremath.data
tuzo.utm.utoronto.ca:~/test% less exploremath.data
```

1	2	2	0	78.0	65	80	39	English	Female	3	3	1
2	2	6	2	66.0	54	75	57	English	Female	3	3	1
3	2	4	4	80.2	77	70	62	English	Male	5	6	1
4	2	5	2	81.7	80	67	76	English	Female	2	2	1
5	2	4	4	86.8	87	80	86	English	Male	5	5	1
6	2	3	1	76.7	53	75	60	English	Male	3	3	1
7	2	3	2	85.8	86	81	54	Other	Female	2	2	1
8	2	4	3	73.0	75	77	17	English	Male	4	5	1
9	2	6	2	72.3	63	60	2	English	Male	4	4	1
10	2	8	6	90.3	87	88	76	English	Male	4	4	1
11	2	8	3	.	.	.	60	English	Male	1	2	1
12	2	6	4	.	.	.	61	Other	Female	1	1	1
13	.	.	.	87.2	84	83	54	English	Male	3	3	1
14	2	2	5	91.0	90	91	84	English	Male	5	5	1
15	2	3	1	72.8	53	74	.	English	Female	3	3	1
16	.	.	.	80.7	72	84	14	English	Male	3	3	1
17	2	5	0	82.5	82	85	75	Other	Female	2	2	1
18	2	4	6	91.5	95	81	94	English	Female	3	3	1
19	2	3	2	78.3	77	74	60	English	Female	3	3	1
20	.	.	.	74.5	0	85	.	English	Male	4	4	1
21	2	3	3	80.7	71	78	53	Other	Female	1	3	1
22	2	5	3	88.3	80	85	63	English	Female	3	3	1
23	2	4	2	76.8	82	64	82	Other	Female	2	2	1

Skipping ....

570	2	5	4	84.8	88	68	80	English	Male	1	1	1
571	2	4	3	78.3	83	84	56	English	Male	4	2	1
572	2	6	3	88.3	81	90	70	English	Female	5	5	1
573	2	3	1	.	.	.	.	English	Male	3	3	1
574	2	5	9	77.0	73	79	60	English	Female	2	2	1
575	.	.	.	78.7	80	73	.	English	Female	6	3	1
576	2	5	2	80.7	80	70	50	Other	Male	1	1	1
577	2	4	2	80.7	56	81	50	English	Female	2	2	1
578	2	4	3	.	.	.	78	Other	Female	4	4	1
579	1	6	1	82.2	80	86	61	English	Female	2	2	1

```

/* checkmath.sas */
options linesize=79 noovp formdlim=' ';
title 'Gender, Ethnicity and Math performance';
title2 'Data definition and checking on the exploratory sample';
/* Of course there was a LOT more checking than this. */

proc format;
  value ynfmt  0 = 'No'  1 = 'Yes';
  value crsfmt  4 = 'No Resp';
  value nfmt
    1 = 'Asian'
    2 = 'Eastern European'
    3 = 'European not Eastern'
    4 = 'Middle-Eastern and Pakistani'
    5 = 'East Indian'
    6 = 'Other and DK' ;

data math;
  infile 'exploremath.data';
  input id course precalc calc gpa calculus english mark lang $ sex $
        nation1 nation2 sample;

/* Computed Variables: totscore, passed, grade, hsgpa, hscal, hsengl,
   tongue, ethnic */

  totscore = precalc+calc;
  if (50<=mark<=100) then passed=1; else passed=0;
  /* Some missing final marks were zero, and 998=SDF and 999=WDR */
  if mark=0 then grade=.;
  else if mark > 100 then grade=.;
  else grade=mark;
  /* Missing HS marks were zeros */
  if 65 le gpa le 100 then hsgpa = gpa; /* Else missing is automatic */
  if 0 < calculus < 101 then hscal = calculus;
  if 0 < english < 101 then hsengl = english;
  /* There were just a few French speakers */
  if lang='French' then tongue='Other '; else tongue=lang;
  label tongue = 'Mother Tongue (Eng or Other)';
  /* Rater 1 knows Middle Eastern names -- otherwise believe Rater 2 */
  if nation1=4 then ethnic=nation1; else ethnic=nation2;

  /*****/

label
  precalc = 'Number precalculus correct'
  calc = 'Number calculus correct'
  totscore = 'Total # right on diagnostic test'
  passed = 'Passed the course'
  grade = 'Final mark (if any)'
  hsgpa = 'High School GPA'
  hscal = 'HS Calculus'
  hsengl = 'HS English'
  lang = 'Mother Tongue'
  nation1 = 'Nationality of name acc to rater1'
  nation2 = 'Nationality of name acc to rater2'
  tongue = 'Mother Tongue (Eng or Other)'

```

```

        ethnic = 'Judged Nationality of name'    ;
format course crsfmt.;
format passed ynfmt.;
format nation1 nation2 ethnic nfmt.;

proc freq;
title3 'Checking Agreement Between Raters';
tables nation1*nation2 / norow nocol;

proc freq;
title3 'Checking Created Categorical variables';
/* Always do this, but never bring it to the quiz. */
tables lang*tongue / norow nocol nopercent missing;
tables ethnic * (nation1 nation2) / norow nocol nopercent missing;
tables mark* passed / norow nocol nopercent missing;

options pagesize = 200; /* You can change the options at any point. */

proc print;
title3 'Checking Created Quantitative variables';
var precalc calc totscore mark grade;

proc print;
var gpa hsgpa calculus hscalc english hsengl;

```

## Here is checkmath.log

```

1
10:29 Thursday, January 3, 3008
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 5.9 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          0.72 seconds
      cpu time           0.07 seconds

```

```

1      /* checkmath.sas */
2      options nodate linesize=79 noovp formdlim=' ' nodate;
3      title 'Gender, Ethnicity and Math performance';
4      title2 'Data definition and checking on the exploratory sample';
5      /* Of course there was a LOT more checking than this. */
6
7      proc format;

```

```

8           value ynfmt  0 = 'No'  1 = 'Yes';
NOTE: Format YNFMT has been output.
9           value crsfmt  4 = 'No Resp';
NOTE: Format CRSFMT has been output.
10          value nfmt
11              1 = 'Asian'
12              2 = 'Eastern European'
13              3 = 'European not Eastern'
14              4 = 'Middle-Eastern and Pakistani'
15              5 = 'East Indian'
16              6 = 'Other and DK' ;
NOTE: Format NFMT has been output.
17

NOTE: PROCEDURE FORMAT used:
      real time          0.08 seconds
      cpu time           0.01 seconds

18          data math;
19              infile 'exploremath.data';
20              input id course precalc calc gpa calculus english mark lang $
20          ! sex $
21              nation1 nation2 sample;
22
23          /* Computed Variables: totscore, passed, grade, hsgpa, hscal,
23          ! hsengl,
24              tongue, ethnic */
25
26          totscore = precalc+calc;
27          if (50<=mark<=100) then passed=1; else passed=0;
28          /* Some missing final marks were zero, and 998=SDF and 999=WDR

2          The SAS System

28          ! */
29          if mark=0 then grade=.;
30          else if mark > 100 then grade=.;
31          else grade=mark;
32          /* Missing HS marks were zeros */
33          if 65 le gpa le 100 then hsgpa = gpa; /* Else missing is
33          ! automatic */
34          if 0 < calculus < 101 then hscal = calculus;
35          if 0 < english < 101 then hsengl = english;
36          /* There were just a few French speakers */
37          if lang='French' then tongue='Other '; else tongue=lang;
38          label tongue = 'Mother Tongue (Eng or Other)';
39          /* Rater 1 knows Middle Eastern names -- otherwise believe
39          ! Rater 2 */
40          if nation1=4 then ethnic=nation1; else ethnic=nation2;
41
42          /*****/
43
44          label
45              precalc = 'Number precalculus correct'
46              calc = 'Number calculus correct'
47              totscore = 'Total # right on diagnostic test'
48              passed = 'Passed the course'

```

```

49         grade = 'Final mark (if any)'
50         hsgpa = 'High School GPA'
51         hscal = 'HS Calculus'
52         hsengl = 'HS English'
53         lang = 'Mother Tongue'
54         nation1 = 'Nationality of name acc to rater1'
55         nation2 = 'Nationality of name acc to rater2'
56         tongue = 'Mother Tongue (Eng or Other)'
57         ethnic = 'Judged Nationality of name'      ;
58
59         format course crsfmt.;
60         format passed ynfmt.;
61         format nation1 nation2 ethnic nfmt.;
62

```

NOTE: The infile 'exploremath.data' is:  
File Name=/u/brunner/442s08/lecture/math/explore/exploremath.data,  
Owner Name=brunner,Group Name=dos,  
Access Permission=rw-r--r--,  
File Size (bytes)=44004

NOTE: 579 records were read from the infile 'exploremath.data'.  
The minimum record length was 75.  
The maximum record length was 75.

NOTE: Missing values were generated as a result of performing an operation on  
missing values.  
Each place is given by: (Number of times) at (Line):(Column).  
99 at 26:24

NOTE: The data set WORK.MATH has 579 observations and 21 variables.

NOTE: DATA statement used:  
real time 0.30 seconds  
cpu time 0.02 seconds

```

63         proc freq;

```

```

3                 The SAS System

```

```

64         title3 'Checking Agreement Between Raters';
65         tables nation1*nation2 / norow nocol;
66

```

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

NOTE: The PROCEDURE FREQ printed page 1.

NOTE: PROCEDURE FREQ used:  
real time 0.22 seconds  
cpu time 0.00 seconds

```

67         proc freq;
68         title3 'Checking Created Categorical variables';
69         /* Always do this, but never bring it to the quiz. */
70         tables lang*tongue / norow nocol nopercent missing;
71         tables ethnic * (nation1 nation2) / norow nocol nopercent
71         ! missing;
72         tables mark* passed / norow nocol nopercent missing;
73
74         options pagesize = 200; /* You can change the options at any point.

```

74 ! \*/  
75

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

NOTE: The PROCEDURE FREQ printed pages 2-3.

NOTE: PROCEDURE FREQ used:

real time	0.01 seconds
cpu time	0.01 seconds

76 proc print;  
77 title3 'Checking Created Quantitative variables';  
78 var precalc calc totscore mark grade;  
79

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

NOTE: The PROCEDURE PRINT printed pages 4-6.

NOTE: PROCEDURE PRINT used:

real time	0.06 seconds
cpu time	0.01 seconds

80 proc print;  
81 var gpa hsgpa calculus hscalc english hsengl;  
82

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

NOTE: The PROCEDURE PRINT printed pages 7-9.

NOTE: PROCEDURE PRINT used:

real time	0.00 seconds
cpu time	0.01 seconds

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

NOTE: The SAS System used:

real time	1.82 seconds
cpu time	0.14 seconds

# Here is checkmath.lst

Gender, Ethnicity and Math performance  
 Data definition and checking on the exploratory sample  
 Checking Agreement Between Raters

1

## The FREQ Procedure

Table of nation1 by nation2

nation1(Nationality of name acc to rater1)		nation2(Nationality of name acc to rater2)					Total
Frequency	Percent	Asian	Eastern European	European not Eastern	Middle-Eastern and Pakistani	East Indian	
Asian	113 19.52	2 0.35	2 0.35	2 0.35	5 0.86	2 0.35	126 21.76
Eastern European	1 0.17	50 8.64	4 0.69	1 0.17	3 0.52	2 0.35	61 10.54
European not Eastern	9 1.55	8 1.38	183 31.61	1 0.17	1 0.17	2 0.35	204 35.23
Middle-Eastern and Pakistani	0 0.00	2 0.35	1 0.17	53 9.15	8 1.38	2 0.35	66 11.40
East Indian	0 0.00	1 0.17	2 0.35	1 0.17	67 11.57	1 0.17	72 12.44
Other and DK	8 1.38	2 0.35	4 0.69	1 0.17	2 0.35	33 5.70	50 8.64
<b>Total</b>	<b>131 22.63</b>	<b>65 11.23</b>	<b>196 33.85</b>	<b>59 10.19</b>	<b>86 14.85</b>	<b>42 7.25</b>	<b>579 100.00</b>

Gender, Ethnicity and Math performance  
 Data definition and checking on the exploratory sample  
 Checking Created Categorical variables

2

The FREQ Procedure

Table of lang by tongue

lang(Mother Tongue)		tongue(Mother Tongue (Eng or Other))				
Frequency		English	Other		Total	
		28	0	0		28
English		0	402	0		402
French		0	0	5		5
Other		0	0	144		144
Total		28	402	149		579

Table of ethnic by nation1

ethnic(Judged Nationality of name)		nation1(Nationality of name acc to rater1)						
Frequency		Asian	Eastern European	European not Eastern	Middle-Eastern and Pakistani	East Indian	Other and DK	Total
Asian		113	1	9	0	0	8	131
Eastern European		2	50	8	0	1	2	63
European not Eastern		2	4	183	0	2	4	195
Middle-Eastern and Pakistani		2	1	1	66	1	1	72
East Indian		5	3	1	0	67	2	78
Other and DK		2	2	2	0	1	33	40
Total		126	61	204	66	72	50	579



Table of ethnic by nation2

ethnic(Judged Nationality of name)							Total
nation2(Nationality of name acc to rater2)							
Frequency	Asian	Eastern European	European not Eastern	Middle-Eastern and Pakistani	East Indian	Other and DK	
Asian	131	0	0	0	0	0	131
Eastern European	0	63	0	0	0	0	63
European not Eastern	0	0	195	0	0	0	195
Middle-Eastern and Pakistani	0	2	1	59	8	2	72
East Indian	0	0	0	0	78	0	78
Other and DK	0	0	0	0	0	40	40
Total	131	65	196	59	86	42	579

Gender, Ethnicity and Math performance  
 Data definition and checking on the exploratory sample  
 Checking Created Categorical variables

3

The FREQ Procedure

Table of mark by passed

mark				passed(Passed the course)	
Frequency	No	Yes	Total		
.	157	0	157		
0	2	0	2		
1	1	0	1		
2	2	0	2		
4	2	0	2		
5	1	0	1		
11	1	0	1		
12	1	0	1		

13	1	0	1
14	2	0	2
15	1	0	1
16	1	0	1
17	2	0	2
18	1	0	1
19	2	0	2
20	1	0	1
22	3	0	3
25	1	0	1
26	3	0	3
28	3	0	3
29	1	0	1
30	3	0	3
31	5	0	5
32	2	0	2
33	1	0	1
34	2	0	2
35	3	0	3
36	3	0	3
37	2	0	2
38	2	0	2
39	7	0	7
40	1	0	1
41	3	0	3
42	3	0	3
43	2	0	2
44	2	0	2
45	2	0	2

46	8	0	8
47	6	0	6
48	1	0	1
50	0	28	28
51	0	11	11
52	0	5	5
53	0	6	6
54	0	11	11
55	0	10	10
56	0	7	7
57	0	13	13
58	0	6	6
60	0	26	26
61	0	10	10
62	0	10	10
63	0	3	3
64	0	8	8
65	0	11	11
66	0	7	7
67	0	4	4
68	0	8	8
70	0	12	12
71	0	10	10
72	0	6	6
73	0	6	6
74	0	5	5
75	0	5	5
76	0	8	8
77	0	6	6

78	0	6	6
80	0	11	11
81	0	4	4
82	0	3	3
83	0	4	4
84	0	4	4
85	0	1	1
86	0	3	3
87	0	4	4
88	0	4	4
90	0	5	5
91	0	2	2
93	0	1	1
94	0	3	3
95	0	5	5
96	0	1	1
97	0	1	1
99	0	1	1
998	14	0	14
999	13	0	13
Total	274	305	579

Gender, Ethnicity and Math performance  
 Data definition and checking on the exploratory sample  
 Checking Created Quantitative variables

Obs	precalc	calc	totscore	mark	grade
1	2	0	2	39	39
2	6	2	8	57	57
3	4	4	8	62	62
4	5	2	7	76	76
5	4	4	8	86	86
6	3	1	4	60	60
7	3	2	5	54	54
8	4	3	7	17	17
9	6	2	8	2	2
10	8	6	14	76	76
11	8	3	11	60	60
12	6	4	10	61	61
13	.	.	.	54	54
14	2	5	7	84	84
15	3	1	4	.	.
16	.	.	.	14	14
17	5	0	5	75	75
18	4	6	10	94	94
19	3	2	5	60	60
20	.	.	.	.	.
21	3	3	6	53	53
22	5	3	8	63	63
23	4	2	6	82	82
24	3	3	6	.	.
25	6	0	6	67	67
26	2	6	8	55	55
27	8	9	17	.	.
28	6	3	9	77	77
29	5	2	7	80	80
30	3	2	5	.	.
31	7	1	8	.	.
32	3	2	5	26	26
33	4	7	11	81	81
34	5	8	13	78	78
35	7	6	13	.	.
36	7	8	15	17	17
37	6	4	10	80	80
38	3	5	8	66	66
39	4	5	9	29	29
40	2	2	4	.	.
41	3	1	4	.	.
42	.	.	.	0	.
43	6	4	10	60	60
44	2	0	2	.	.
45	5	5	10	53	53
46	3	3	6	26	26
47	1	0	1	.	.
48	.	.	.	50	50

Skipping the rest