

## Handout 2: Read and Describe the Math Data

```
appsrv01.srv.cquest.utoronto.ca> ls
mathdescr1.sas      mathexplore.data
appsrv01.srv.cquest.utoronto.ca> less mathexplore.data
 1  2  2  0  78.0  65  80  39  English  Female  8   8  1
 2  2  6  2  66.0  54  75  57  English  Female 12  12  1
 3  2  4  4  80.2  77  70  62  English  Male   16  18  1
 4  2  5  2  81.7  80  67  76  English  Female  6   6  1
 5  2  4  4  86.8  87  80  86  English  Male  16  16  1
 6  2  3  1  76.7  53  75  60  English  Male  10  10  1
 7  2  3  2  85.8  86  81  54  Other    Female  6   6  1
 8  2  4  3  73.0  75  77  17  English  Male  14  16  1
 9  2  6  2  72.3  63  60   2  English  Male  14  14  1
10  2  8  6  90.3  87  88  76  English  Male  14  14  1
11  2  8  3  .      .   .   60  English  Male   2   6  1
12  2  6  4  .      .   .   61  Other    Female  1   1  1
13  .  .  .  87.2  84  83  54  English  Male   8   8  1
14  2  2  5  91.0  90  91  84  English  Male  16  16  1
15  2  3  1  72.8  53  74  .   English  Female  7   9  1
16  .  .  .  80.7  72  84  14  English  Male   8   8  1
17  2  5  0  82.5  82  85  75  Other    Female  6   6  1
18  2  4  6  91.5  95  81  94  English  Female  8   8  1
19  2  3  2  78.3  77  74  60  English  Female  8   8  1
20  .  .  .  74.5   0  85  .   English  Male  14  14  1
21  2  3  3  80.7  71  78  53  Other    Female  5  13  1
22  2  5  3  88.3  80  85  63  English  Female  8   8  1
23  2  4  2  76.8  82  64  82  Other    Female  6   6  1
```

(and so on: spacebar for another page, q for quit b for back a page, g for go to beginning of file, G for go to end of file)

Here's the input statement of mathdescr1.sas; it tells what the variables are:

```
input id course precalc calc gpa calculus english mark lang $ sex $
       nation1 nation2 sample;
```

```

/* mathdescrl.sas */
options linesize=79 noovp formdlim=' ';
title 'Gender, Ethnicity and Math performance';
title2 'Basic descriptive statistics on the exploratory sample';

proc format;
value ynfmt 0 = 'No' 1 = 'Yes';
value crsfmt 4 = 'No Resp';
value nfmt
    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 or DK';

value ncfmt
    1 = 'Asian'
    2 = 'Eastern European'
    3 = 'European not Eastern'
    4 = 'Middle-Eastern and Pakistani'
    5 = 'East Indian'
    6 = 'Other and DK' ;

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

/* Computed Variables: totscore, passed, grade, hsgpa, hscal, hsengl,
   tongue, cat1, cat2, ethnic */

totscore = precalc+calc;
if (50<=mark<=100) then passed=1; else passed=0;
if mark=0 then grade=.;
else if mark > 100 then grade=.;
else grade=mark;
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;
if lang='French' then tongue='Other ' ; else tongue=lang;
label tongue = 'Mother Tongue (Eng or Other)';

```

```

/***** Nationality According to the 2 raters *****/
if      1 <= nation1 <= 5 then cat1 = 1;
else if nation1 = 6 then cat1 = 2;
else if 7 <= nation1 <= 13 then cat1 = 3;
else if 14 <= nation1 <= 15 then cat1 = 4;
else if nation1 = 16 then cat1 = 5;
else    cat1 = 6;
if      1 <= nation2 <= 5 then cat2 = 1;
else if nation2 = 6 then cat2 = 2;
else if 7 <= nation2 <= 13 then cat2 = 3;
else if 14 <= nation2 <= 15 then cat2 = 4;
else if nation2 = 16 then cat2 = 5;
else    cat2 = 6;
if cat1=4 then ethnic=cat1; else ethnic=cat2;
/* Rater 1 knows Middle Eastern names -- otherwise believe Rater 2 */
/*****/

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'
    hscalcl = '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)'
    cat1     = 'Nationality Acc to Rater 1'
    cat2     = 'Nationality Acc to Rater 2'
    ethnic   = 'Judged Nationality of name'
    ;

format course crsfmt.;
format passed ynfmt.;
format nation1 nation2 nfmt.;
format cat1 cat2 ethnic ncfmt.;

proc freq;
title3 'Checking Agreement';
tables lang*tongue cat1*cat2/norow nocol nopercnt;

proc means;
title3 'Quantitative Variables';
var precalc calc totscore hsgpa hscalcl hsengl grade;

proc freq;
title3 'Categorical variables';
tables course sex ethnic tongue passed;

proc freq;
title3 'A few Chisquare tests';
tables (course sex ethnic tongue) * passed / nocol nopercnt chisq;

```

```

appsrv01.srv.cquest.utoronto.ca> ls
mathdescr1.sas      mathexplore.data
appsrv01.srv.cquest.utoronto.ca> sas mathdescr1
appsrv01.srv.cquest.utoronto.ca> ls
mathdescr1.log      mathdescr1.lst      mathdescr1.sas      mathexplore.data
appsrv01.srv.cquest.utoronto.ca> less mathdescr1.log

```

```

appsrv01.srv.cquest.utoronto.ca> cat mathdescr1.lst

```

Gender, Ethnicity and Math performance 1  
 Basic descriptive statistics on the exploratory sample  
 Checking Agreement  
 22:48 Sunday, September 23, 2007

The FREQ Procedure

Table of lang by tongue

lang(Mother Tongue)			
tongue(Mother Tongue (Eng or Other))			
Frequency	English	Other	Total
English	402	0	402
French	0	5	5
Other	0	144	144
Total	402	149	551

Frequency Missing = 28

Table of cat1 by cat2

cat1(Nationality Acc to Rater 1)							cat2(Nationality Acc to Rater 2)							
Frequency	Asian	Eastern European	European not Eastern	Middle-Eastern and Pakistani	East Indian	Other and DK	Total	Asian	Eastern European	European not Eastern	Middle-Eastern and Pakistani	East Indian	Other and DK	Total
Asian	113	2	2	2	5	2	126							
Eastern European	1	50	4	1	3	2	61							
European not Eastern	9	8	183	1	1	2	204							
Middle-Eastern and Pakistani	0	2	1	53	8	2	66							
East Indian	0	1	2	1	67	1	72							
Other and DK	8	2	4	1	2	33	50							
Total	131	65	196	59	86	42	579							

Gender, Ethnicity and Math performance 2  
 Basic descriptive statistics on the exploratory sample  
 Quantitative Variables  
 22:48 Sunday, September 23, 2007

The MEANS Procedure

Variable	Label	N	Mean	Std Dev
precalc	Number precalculus correct	480	4.4020833	1.7621620
calc	Number calculus correct	480	3.3187500	2.4918369
totscore	Total # right on diagnostic test	480	7.7208333	3.7036888
hsgpa	High School GPA	466	79.2690987	6.0384921
hscal	HS Calculus	448	75.4441964	12.1918487
hsengl	HS English	480	75.8354167	8.7180371
grade	Final mark (if any)	393	58.7379135	19.1831935

Variable	Label	Minimum	Maximum
precalc	Number precalculus correct	0	9.0000000
calc	Number calculus correct	0	11.0000000
totscore	Total # right on diagnostic test	0	20.0000000
hsgpa	High School GPA	65.0000000	97.3000000
hscal	HS Calculus	50.0000000	100.0000000
hsengl	HS English	50.0000000	96.0000000
grade	Final mark (if any)	1.0000000	99.0000000

Gender, Ethnicity and Math performance 3  
 Basic descriptive statistics on the exploratory sample  
 Categorical variables  
 22:48 Sunday, September 23, 2007

The FREQ Procedure

course	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	59	12.29	59	12.29
2	373	77.71	432	90.00
3	39	8.13	471	98.13
No Resp	9	1.88	480	100.00

Frequency Missing = 99

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

Frequency Missing = 28

Judged Nationality of name

ethnic	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Asian	131	22.63	131	22.63
Eastern European	63	10.88	194	33.51
European not Eastern	195	33.68	389	67.18
Middle-Eastern and Pakistani	72	12.44	461	79.62
East Indian	78	13.47	539	93.09
Other and DK	40	6.91	579	100.00

Mother Tongue (Eng or Other)

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

Frequency Missing = 28

Passed the course

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

Gender, Ethnicity and Math performance  
 Basic descriptive statistics on the exploratory sample  
 A few Chisquare tests

4

22:48 Sunday, September 23, 2007

The FREQ Procedure

Table of course by passed

course	passed(Passed the course)		
Frequency	No	Yes	Total
Row Pct			
1	44 74.58	15 25.42	59
2	149 39.95	224 60.05	373
3	8 20.51	31 79.49	39
No Resp	9 100.00	0 0.00	9
Total	210	270	480

Frequency Missing = 99

Statistics for Table of course by passed

Statistic	DF	Value	Prob
Chi-Square	3	45.1035	<.0001
Likelihood Ratio Chi-Square	3	49.5188	<.0001
Mantel-Haenszel Chi-Square	1	8.8752	0.0029
Phi Coefficient		0.3065	
Contingency Coefficient		0.2931	
Cramer's V		0.3065	

Effective Sample Size = 480

Frequency Missing = 99

WARNING: 17% of the data are missing.