

First look at the Math Data*

```

bash-3.2$ ls
bash-3.2$ curl
http://www.utstat.toronto.edu/~brunner/appliedf13/code_n_data/lecture/exploremath.d
ata > exploremath.data
  % Total    % Received % Xferd  Average Speed   Time    Time       Time  Current
   100 44004    100 44004    0     0    2458k      0  --:--:--  --:--:--  --:--:--  13.9M
bash-3.2$ ls
exploremath.data
bash-3.2$ less exploremath.data

```

Line	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	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

```
bash-3.2$ emacs math0.sas
```

* Copyright information is on the last page.

```
/* math0.sas */
options linesize=79 noovp formdlim=' ';
title 'Gender, Ethnicity and Math performance: Exploratory Sample';
title2 'Read Data, Take a first Look';

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

proc freq;
  title3 'Frequency Distributions of all variables, even continuous';
  tables _all_; /* Could have said id -- sample instead of _all_ */
```

```
bash-3.2$ ls
exploremath.data  math0.sas
bash-3.2$ sas math0
bash-3.2$ ls
exploremath.data  math0.log          math0.lst          math0.sas
```

```
bash-3.2$ less math0.log
```

1 The SAS System
07:33 Sunday, October 20, 2013

NOTE: Copyright (c) 2002-2010 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software 9.3 (TS1M0)
Licensed to UNIVERSITY OF TORONTO/COMPUTING & COMMUNICATIONS, Site 70072784.
NOTE: This session is executing on the SunOS 5.10 (SUN 64) platform.

You are running SAS 9. Some SAS 8 files will be automatically converted by the V9 engine; others are incompatible. Please see <http://support.sas.com/rnd/migration/planning/platform/64bit.html>

PROC MIGRATE will preserve current SAS file attributes and is recommended for converting all your SAS libraries from any SAS 8 release to SAS 9. For details and examples, please see <http://support.sas.com/rnd/migration/index.html>

This message is contained in the SAS news file, and is presented upon initialization. Edit the file "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 1.70 seconds
cpu time 0.20 seconds

```
1 /* math0.sas */
2 options linesize=79 noovp formdlim=' ';
3 title 'Gender, Ethnicity and Math performance: Exploratory Sample';
4 title2 'Read Data, Take a first Look';
5
6 data math;
7   infile 'exploremath.data';
8   input id course precalc calc gpa calculus english mark lang $
8   ! sex $
9         nation1 nation2 sample;
10
```

NOTE: The infile 'exploremath.data' is:
Filename=/u/brunner/2101f13/lecture/math/test/exploremath.data,
Owner Name=brunner,Group Name=dos,
Access Permission=rw-r--r--,
Last Modified=Sun Oct 20 07:20:31 2013,
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: The data set WORK.MATH has 579 observations and 13 variables.

NOTE: DATA statement used (Total process time):
real time 0.32 seconds
cpu time 0.06 seconds

```
11          proc freq;
```

```
2                                The SAS System
```

```
12          title3 'Frequency Distributions of all variables, even  
12          ! continuous';  
13          tables _all_; /* Could have said id -- sample instead of _all_  
13          ! */  
14  
15  
16  
17  
18
```

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

NOTE: The PROCEDURE FREQ printed pages 1-22.

NOTE: PROCEDURE FREQ used (Total process time):
real time 0.83 seconds
cpu time 0.20 seconds

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

NOTE: The SAS System used:
real time 2.98 seconds
cpu time 0.47 seconds

```
bash-3.2$
```

```
bash-3.2$ less math0.lst
```

Gender, Ethnicity and Math performance: Exploratory Sample
 Read Data, Take a first Look
 Frequency Distributions of all variables, even continuous

1

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
5	1	0.17	5	0.86
6	1	0.17	6	1.04
7	1	0.17	7	1.21
8	1	0.17	8	1.38
9	1	0.17	9	1.55
10	1	0.17	10	1.73

Skipping

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

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
4	9	1.88	480	100.00

Frequency Missing = 99

precalc	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	4	0.83	4	0.83
1	13	2.71	17	3.54
2	53	11.04	70	14.58
3	84	17.50	154	32.08
4	96	20.00	250	52.08
5	104	21.67	354	73.75
6	72	15.00	426	88.75
7	33	6.88	459	95.63
8	14	2.92	473	98.54
9	7	1.46	480	100.00

Frequency Missing = 99

The FREQ Procedure

calc	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	55	11.46	55	11.46
1	77	16.04	132	27.50
2	88	18.33	220	45.83
3	58	12.08	278	57.92
4	65	13.54	343	71.46
5	36	7.50	379	78.96
6	39	8.13	418	87.08
7	25	5.21	443	92.29
8	22	4.58	465	96.88
9	11	2.29	476	99.17
10	2	0.42	478	99.58
11	2	0.42	480	100.00

Frequency Missing = 99

The FREQ Procedure

gpa	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	25	5.09	25	5.09
65	1	0.20	26	5.30
66	2	0.41	28	5.70
67.3	1	0.20	29	5.91
67.7	1	0.20	30	6.11
68	1	0.20	31	6.31
69	2	0.41	33	6.72
69.2	3	0.61	36	7.33
69.3	1	0.20	37	7.54
69.5	1	0.20	38	7.74
69.7	1	0.20	39	7.94
69.8	4	0.81	43	8.76
70	2	0.41	45	9.16
70.3	1	0.20	46	9.37
70.5	2	0.41	48	9.78
70.7	4	0.81	52	10.59
70.8	2	0.41	54	11.00
71.2	3	0.61	57	11.61
71.3	2	0.41	59	12.02
71.5	5	1.02	64	13.03
71.8	2	0.41	66	13.44
72.2	2	0.41	68	13.85
72.3	3	0.61	71	14.46
72.5	5	1.02	76	15.48
72.7	3	0.61	79	16.09
72.8	6	1.22	85	17.31
73	4	0.81	89	18.13
73.2	4	0.81	93	18.94
73.3	6	1.22	99	20.16

73.5	5	1.02	104	21.18
73.7	3	0.61	107	21.79
73.8	6	1.22	113	23.01
74	3	0.61	116	23.63
74.2	2	0.41	118	24.03
74.3	2	0.41	120	24.44
74.5	7	1.43	127	25.87
74.7	4	0.81	131	26.68
74.8	4	0.81	135	27.49
75	5	1.02	140	28.51
75.2	6	1.22	146	29.74
75.3	2	0.41	148	30.14
75.5	6	1.22	154	31.36
75.7	3	0.61	157	31.98
75.8	7	1.43	164	33.40
76	6	1.22	170	34.62
76.2	10	2.04	180	36.66
76.3	5	1.02	185	37.68
76.5	7	1.43	192	39.10
76.7	7	1.43	199	40.53
76.8	6	1.22	205	41.75
77	7	1.43	212	43.18

Gender, Ethnicity and Math performance: Exploratory Sample
 Read Data, Take a first Look
 Frequency Distributions of all variables, even continuous

15

The FREQ Procedure

gpa	Frequency	Percent	Cumulative Frequency	Cumulative Percent
77.2	9	1.83	221	45.01
77.3	4	0.81	225	45.82
77.5	8	1.63	233	47.45
77.7	5	1.02	238	48.47

Skipping

94.3	1	0.20	487	99.19
95	1	0.20	488	99.39
95.8	1	0.20	489	99.59
96.2	1	0.20	490	99.80
97.3	1	0.20	491	100.00

Frequency Missing = 88

Gender, Ethnicity and Math performance: Exploratory Sample
 Read Data, Take a first Look
 Frequency Distributions of all variables, even continuous

17

The FREQ Procedure

calculus	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	29	5.91	29	5.91
50	10	2.04	39	7.94
51	8	1.63	47	9.57
52	3	0.61	50	10.18
53	5	1.02	55	11.20
54	2	0.41	57	11.61
55	7	1.43	64	13.03
56	3	0.61	67	13.65
57	4	0.81	71	14.46
58	4	0.81	75	15.27
59	1	0.20	76	15.48
60	14	2.85	90	18.33
61	8	1.63	98	19.96
62	9	1.83	107	21.79
63	8	1.63	115	23.42
64	4	0.81	119	24.24
65	14	2.85	133	27.09
66	7	1.43	140	28.51

Skipping

98	3	0.61	475	96.74
99	1	0.20	476	96.95
100	1	0.20	477	97.15
999	14	2.85	491	100.00

Frequency Missing = 88

Gender, Ethnicity and Math performance: Exploratory Sample
 Read Data, Take a first Look
 Frequency Distributions of all variables, even continuous

19

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

Skipping

95	1	0.20	488	99.80
96	1	0.20	489	100.00

Frequency Missing = 90

Gender, Ethnicity and Math performance: Exploratory Sample
 Read Data, Take a first Look
 Frequency Distributions of all variables, even continuous

20

The FREQ Procedure

mark	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	2	0.47	2	0.47
1	1	0.24	3	0.71
2	2	0.47	5	1.18
4	2	0.47	7	1.66

Skipping

97	1	0.24	394	93.36
99	1	0.24	395	93.60
998	14	3.32	409	96.92
999	13	3.08	422	100.00

Frequency Missing = 157

lang	Frequency	Percent	Cumulative Frequency	Cumulative Percent
English	402	72.96	402	72.96
French	5	0.91	407	73.87
Other	144	26.13	551	100.00

Frequency Missing = 28

Gender, Ethnicity and Math performance: Exploratory Sample
 Read Data, Take a first Look
 Frequency Distributions of all variables, even continuous

22

The FREQ Procedure

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

Frequency Missing = 28

nation1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	126	21.76	126	21.76
2	61	10.54	187	32.30
3	204	35.23	391	67.53
4	66	11.40	457	78.93
5	72	12.44	529	91.36
6	50	8.64	579	100.00

nation2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	131	22.63	131	22.63
2	65	11.23	196	33.85
3	196	33.85	392	67.70
4	59	10.19	451	77.89
5	86	14.85	537	92.75
6	42	7.25	579	100.00

sample	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	579	100.00	579	100.00

Issues

Course=4 means No Response

GPA = 0 means missing

Calculus 0=missing, 999=WDR? Anyway not a real grade

English 0=missing

Mark Maybe 0=missing, 998=SDF and 999=WDR

This handout was prepared by Jerry Brunner, Department of Statistics, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The OpenOffice.org document is available from the course website:

<http://www.utstat.toronto.edu/~brunner/oldclass/appliedf13>