

Basic Descriptive Statistics on the Math Data*

```
/* describemath.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 = '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'
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.;
```

* Copyright information is on the last page.

```

/*****/

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

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

proc chart;
  /* You can get crude statistical graphics -- Okay for a first look.
  There is a good variety of chart types. */
  title2 'Vertical Bar Chart of "Nationality"';
  vbar ethnic /discrete type=percent;

proc univariate normal plot;
  /* The normal option gives a test for normality. Plot gives a
  collection of line printer graphics, some of which are informative. */
  title3 'Detailed look at grade';
  var grade;

options pagesize=500;
proc freq;
  title3 'Frequency Distribution of Final Grade';
  tables grade;

/* The following shows that profs in Courses 1 and 3 bumped the marks
discontinuously, while the prof in Course 2 did something more
sophisticated.

proc freq;
  tables grade*course / norow nopercnt nopercnt;

```

Gender, Ethnicity and Math performance

1

Quantitative Variables

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

2

Categorical variables

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

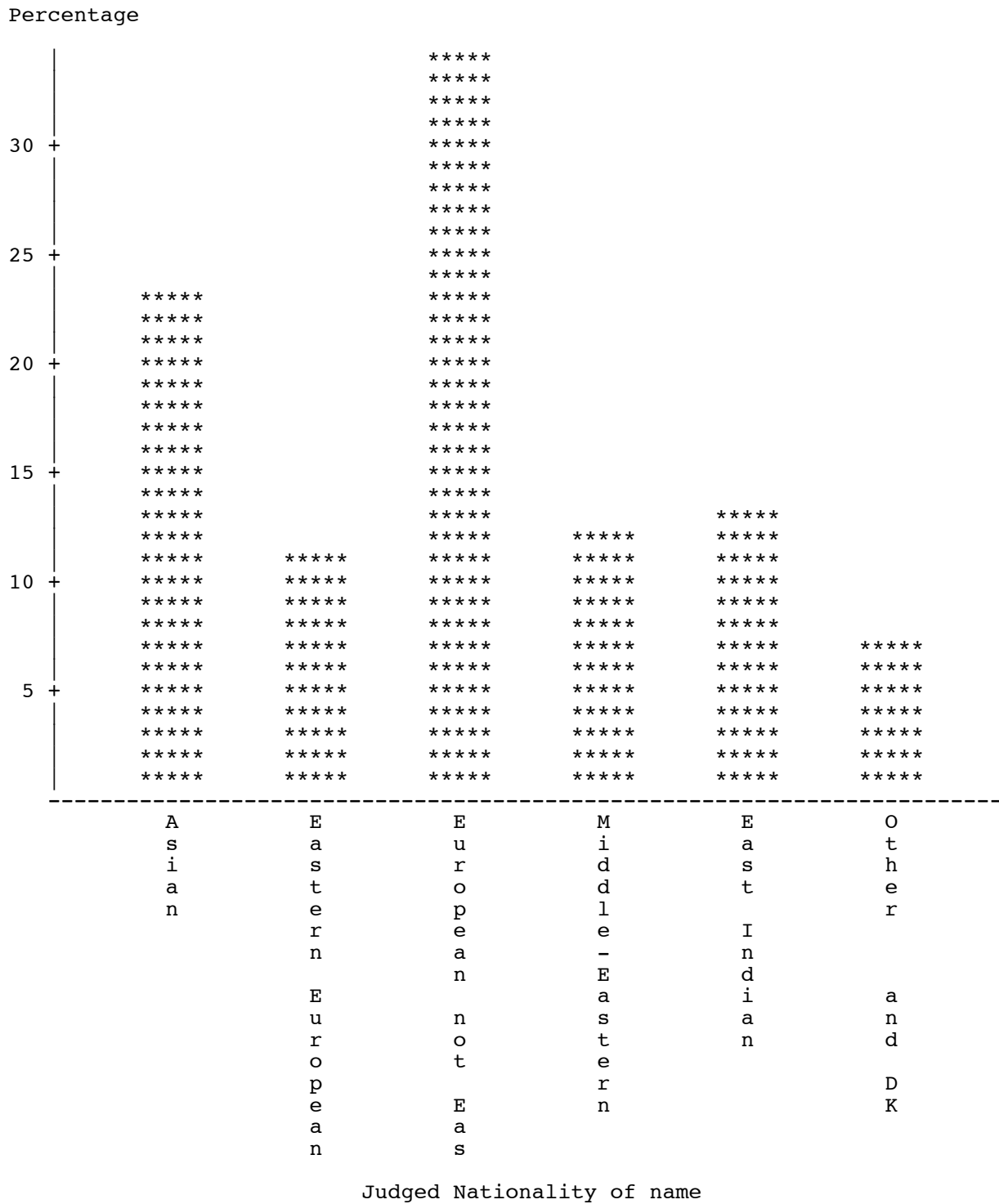
Mother Tongue (Eng or Other)

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



The UNIVARIATE Procedure
Variable: grade (Final mark (if any))

Moments

N	393	Sum Weights	393
Mean	58.7379135	Sum Observations	23084
Std Deviation	19.1831935	Variance	367.994911
Skewness	-0.5121423	Kurtosis	0.35144502
Uncorrected SS	1500160	Corrected SS	144254.005
Coeff Variation	32.658963	Std Error Mean	0.9676641

Basic Statistical Measures

Location		Variability	
Mean	58.73791	Std Deviation	19.18319
Median	60.00000	Variance	367.99491
Mode	50.00000	Range	98.00000
		Interquartile Range	22.00000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 60.70073	Pr > t	<.0001
Sign	M 196.5	Pr >= M	<.0001
Signed Rank	S 38710.5	Pr >= S	<.0001

Tests for Normality

Test	--Statistic---	-----p Value-----	
Shapiro-Wilk	W 0.977228	Pr < W	<0.0001
Kolmogorov-Smirnov	D 0.100457	Pr > D	<0.0100
Cramer-von Mises	W-Sq 0.318394	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq 2.08731	Pr > A-Sq	<0.0050

Quantiles (Definition 5)

Quantile	Estimate
100% Max	99
99%	95
95%	88
90%	82
75% Q3	72
50% Median	60
25% Q1	50
10%	32
5%	22
1%	4
0% Min	1

Extreme Observations

----Lowest----		----Highest----	
Value	Obs	Value	Obs
1	176	95	268
2	319	95	340
2	9	96	452
4	498	97	486
4	70	99	286

Missing Values

Missing Value	Count	-----Percent Of-----	
		All Obs	Missing Obs
.	186	32.12	100.00

Gender, Ethnicity and Math performance
Vertical Bar Chart of "Nationality"
Frequency Distribution of Final Grade

The FREQ Procedure

Final mark (if any)

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
5	1	0.25	6	1.53
11	1	0.25	7	1.78
12	1	0.25	8	2.04
13	1	0.25	9	2.29
14	2	0.51	11	2.80
15	1	0.25	12	3.05
16	1	0.25	13	3.31
17	2	0.51	15	3.82
18	1	0.25	16	4.07
19	2	0.51	18	4.58
20	1	0.25	19	4.83
22	3	0.76	22	5.60
25	1	0.25	23	5.85
26	3	0.76	26	6.62
28	3	0.76	29	7.38
29	1	0.25	30	7.63
30	3	0.76	33	8.40
31	5	1.27	38	9.67
32	2	0.51	40	10.18
33	1	0.25	41	10.43
34	2	0.51	43	10.94
35	3	0.76	46	11.70
36	3	0.76	49	12.47
37	2	0.51	51	12.98
38	2	0.51	53	13.49
39	7	1.78	60	15.27
40	1	0.25	61	15.52
41	3	0.76	64	16.28
42	3	0.76	67	17.05
43	2	0.51	69	17.56
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
56	7	1.78	166	42.24
57	13	3.31	179	45.55
58	6	1.53	185	47.07
60	26	6.62	211	53.69
61	10	2.54	221	56.23
62	10	2.54	231	58.78
63	3	0.76	234	59.54

64	8	2.04	242	61.58
65	11	2.80	253	64.38
66	7	1.78	260	66.16
67	4	1.02	264	67.18
68	8	2.04	272	69.21
70	12	3.05	284	72.26
71	10	2.54	294	74.81
72	6	1.53	300	76.34
73	6	1.53	306	77.86
74	5	1.27	311	79.13
75	5	1.27	316	80.41
76	8	2.04	324	82.44
77	6	1.53	330	83.97
78	6	1.53	336	85.50
80	11	2.80	347	88.30
81	4	1.02	351	89.31
82	3	0.76	354	90.08
83	4	1.02	358	91.09
84	4	1.02	362	92.11
85	1	0.25	363	92.37
86	3	0.76	366	93.13
87	4	1.02	370	94.15
88	4	1.02	374	95.17
90	5	1.27	379	96.44
91	2	0.51	381	96.95
93	1	0.25	382	97.20
94	3	0.76	385	97.96
95	5	1.27	390	99.24
96	1	0.25	391	99.49
97	1	0.25	392	99.75
99	1	0.25	393	100.00

Frequency Missing = 186

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>