The Math Data*

```
tuzo.utm.utoronto.ca:~% mkdir 305s14
tuzo.utm.utoronto.ca:~% cd 305s14
tuzo.utm.utoronto.ca:~/305s14% ls
tuzo.utm.utoronto.ca:~/305s14% curl http://www.utstat.utoronto.ca/~brunner/
305s14/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  66312      0 --:--:-- --:--:-- --:--:-- 3581k
tuzo.utm.utoronto.ca:~/305s14% ls
exploremath.data
tuzo.utm.utoronto.ca:~/305s14% 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

```
tuzo.utm.utoronto.ca:~/305s14% emacs math0.sas
```

* Copyright information is on the last page.

```
/* math0.sas */
options linesize=79 noovp formdlim=' ' ;
title 'Prediction of Performance in First-year Calculus: 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_ */
```

```
tuzo.utm.utoronto.ca:~/305s14% ls
exploremath.data  math0.sas
tuzo.utm.utoronto.ca:~/305s14% sas math0
tuzo.utm.utoronto.ca:~/305s14% ls
exploremath.data  math0.log  math0.lst  math0.sas
tuzo.utm.utoronto.ca:~/305s14% less math0.log
1                                                       The SAS System
                      11:41 Sunday, December 29, 2013

NOTE: Copyright (c) 2002-2010 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software 9.3 (TS1M1)
      Licensed to UNIVERSITY OF TORONTO/COMPUTING & COMMUNICATIONS, Site 7007278
5.
NOTE: This session is executing on the Linux 2.6.9-67.ELsmp (LINUX) platform.



NOTE: Updated analytical products:

SAS/STAT 9.3_M1, SAS/ETS 9.3_M1, SAS/OR 9.3_M1

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           0.09 seconds
      cpu time            0.03 seconds
```

```
1          /* math0.sas */
2          options linesize=79 noovp formdlim=' ' ;
3          title 'Prediction of Performance in First-year Calculus: Exploratory
3        !  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=/student/cslec/brunnerj/305s14/exploremath.data,
      Owner Name=brunnerj,Group Name=lecturers,
      Access Permission=rw-------,
      Last Modified=Sun Dec 29 11:30:18 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):
^L2                                The SAS System

      real time           0.05 seconds
      cpu time            0.01 seconds


11         proc freq;
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           0.18 seconds
      cpu time            0.03 seconds


NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           0.49 seconds
      cpu time            0.07 seconds

tuzo.utm.utoronto.ca:~/305s14% less math0.lst
```

Read Data, Take a first Look

Frequency Distributions of all variables, even continuous

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

Read Data, Take a first Look

Frequency Distributions of all variables, even continuous

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

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

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

Prediction of Performance in First-year Calculus: Exploratory Sample 20
Read Data, Take a first Look
Frequency Distributions of all variables, even continuous

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

Prediction of Performance in First-year Calculus: Exploratory Sample 22
Read Data, Take a first Look
Frequency Distributions of all variables, even continuous

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

```
/* mathread.sas   Just read the data and do basic transformations */
options linesize=79 noovp formdlim='_' nodate;
title 'Prediction of Performance in First-year Calculus';

proc format;
     value ynfmt  0 = 'No'  1 = 'Yes';
     value crsfmt 1 = 'Catch-up' 2 = 'Mainstrm' 3 = 'Elite' 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 mathex;
     infile 'exploremath.data';
     input id course precalc calc gpa calculus english mark lang $ sex $
           nation1 nation2 sample;

/* Computed Variables: totscore, passed, grade, hsgpa, hscalc, 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 hscalc = 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;
     /* Eliminate 'No Resp' for course */
     if course=4 then course2=.; else course2=course;
     /*****************************************************************/

     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'
           hscalc = '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.;
```

```
/* Dummy variables: gender, mtongue and c1-c3 */
if sex = 'Female' then gender=1; else if sex = 'Male' then gender=0;
label gender = '0=Male, 1=Female';
if tongue = 'English' then mtongue=1; else if tongue='Other' then mtongue=0;
label mtongue = '0=Other, 1=English';

/* Only use 2 of these if the model has an intercept! */
if course=. then c1=.; else if course=1 then c1=1; else c1=0;
if course=. then c2=.; else if course=2 then c2=1; else c2=0;
if course=. then c3=.; else if course=3 then c3=1; else c3=0;
label c1 = 'Catch-up' c2 = 'Mainstream' c3 = 'Elite';
```

```
/* math1.sas */
%include 'mathread.sas'; /* Reads in data definition file. */
title2 'Basic descriptives, and illustrate ods pdf output';
/*****************************************************************/

ods pdf file='math1.pdf';  /* rtf and html are also available. */


proc means;
     title3 'Quantitative Variables';
     var precalc calc totscore hsgpa hscalc hsengl grade;
run; /* Need this after every proc with ods. */

proc freq;
     title3 'Categorical variables';
     tables course sex ethnic tongue passed;
run; /* Need this after every proc with ods. */

ods pdf close; /* Must close the output file. */
```

math1.pdf

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
precalc	Number precalculus correct	480	4.4020833	1.7621620	0	9.0000000
calc	Number calculus correct	480	3.3187500	2.4918369	0	11.0000000
totscore	Total # right on diagnostic test	480	7.7208333	3.7036888	0	20.0000000
hsgpa	High School GPA	466	79.2690987	6.0384921	65.0000000	97.3000000
hscalc	HS Calculus	448	75.4441964	12.1918487	50.0000000	100.0000000
hsengl	HS English	480	75.8354167	8.7180371	50.0000000	96.0000000
grade	Final mark (if any)	393	58.7379135	19.1831935	1.0000000	99.0000000

course	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Catch-up	59	12.29	59	12.29
Mainstrm	373	77.71	432	90.00
Elite	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 minimum grade of 1 is unnaturally low. Could have followed up on it, like this:

```
proc univariate;
     title3 'Take a closer look at grade';
     var grade;

proc sort;
     by grade;

proc print;
     var id grade;
```

Actually, I noticed the low minimum grade because it caused trouble with the residuals in the regression analysis.

```
/* math2.sas  */
%include 'mathread.sas';
title2 'Predicting Grade with multiple regression';

proc freq;
     title3 'Check dummy variables';
     tables sex*gender / norow nocol nopercent missing;
     tables tongue*mtongue / norow nocol nopercent missing;
     tables (c1-c3) * course / norow nocol nopercent missing;

proc reg;
     title3 'Initial model';
     model grade = hsgpa hscalc hsengl precalc calc c1 c3 gender mtongue;
     Course: test c1=c3=0;
     Diagnostic_Test:  test  precalc=0, calc=0;
     MaybeDropThese: test calc=c1=c3=gender=0;

proc reg;
     title3 'Model 2 (Drop course and gender, combine diagnostic test)';
     model grade = hsgpa hscalc hsengl totscore mtongue;
     /* Creating new SAS data set math2, with Studentized deleted residuals,
        predicted Y values, and lower and upper 95% prediction intervals. */
     output out=math2 rstudent = delstudres
                      predicted = yhat
                      lcl = lower95 ucl = upper95;

proc univariate;
     title3 'Look at Studentized deleted residuals';
     var delstudres;

proc iml;
     title3 'Critical values for Studentized deleted residuals';
     dfe = 281; alpha = 0.05; df = dfe-1;
     Uncorrected = tinv(1-alpha/2, df);
     print "Uncorrected critical value of t = " Uncorrected;
     print "Alpha = " alpha ", df = " df;
     n = 287; a = alpha/n;
     Corrected = tinv(1-a/2, dfe-1);
     print "Bonferroni corected critical value of t = " Corrected;
     print "Joint alpha = " alpha ", corrected for " n " tests, df = " df;
     print " ";

proc plot;
     title3 'Y by Y-hat';
     plot grade * yhat; /* Y by X */
```

```
proc sort;
     by grade;
proc print;
     title3 'Locate that low mark';
     var id grade yhat delstudres;

/* Decide to throw out marks less than 15 (4 observations). They probably
   didn't show for the final exam. */

data math3;
     set mathex; /* Now math3 is identical to oiginal data set mathex */
     if grade ge 15; /* Deletes the rest */

proc reg;
     title3 'Model 2 again';
     model grade = hsgpa hscalc hsengl totscore mtongue;
     /* Creating new SAS data set math4, with Studentized deleted residuals,
        predicted Y values, and lower and upper 95% prediction intervals. */
     output out=math4 rstudent = delstudres
                      predicted = yhat
                      lcl = lower95 ucl = upper95;

proc univariate plot;
     title3 'Look at Studentized deleted residuals again';
     var delstudres;

proc plot;
     title3 'Y by Y-hat again';
     plot grade * yhat; /* Y by X */

proc plot;
     title3 'Studentized deleted residuals by variables in the model';
     plot delstudres * (hsgpa hscalc hsengl totscore mtongue);

/* The suggestion of higher variance among English speakers is interesting. */

proc plot;
     title3 'Studentized deleted residuals by variables not in the model';
     plot delstudres * (gender course ethnic);

proc reg;
     title3 'Try putting Gender back in';
     model grade = hsgpa hscalc hsengl totscore mtongue gender;

/* Nope. */

/* How good is the prediction? */
proc sort;
     by id;

proc print;
     title3 'How good is the "prediction?"';
     var id yhat grade lower95 upper95 delstudres;
```

Part of **math2.lst** (44 pages)

Predicting Grade with multiple regression
Check dummy variables

The FREQ Procedure

Table of sex by gender

sex gender(0=Male, 1=Female)

Frequency	.	0	1	Total
	28	0	0	28
Female	0	0	266	266
Male	0	285	0	285
Total	28	285	266	579

Table of tongue by mtongue

tongue(Mother Tongue (Eng or Other))
mtongue(0=Other, 1=English)

Frequency	.	0	1	Total
	28	0	0	28
English	0	0	402	402
Other	0	149	0	149
Total	28	149	402	579

Table of c1 by course

c1(Catch-up) course

Frequency	.	Catch-up	Mainstrm	Elite	No Resp	Total
.	99	0	0	0	0	99
0	0	0	373	39	9	421
1	0	59	0	0	0	59
Total	99	59	373	39	9	579

The FREQ Procedure

Table of c2 by course

c2(Mainstream) course

Frequency	.	Catch-up	Mainstrm	Elite	No Resp	Total
.	99	0	0	0	0	99
0	0	59	0	39	9	107
1	0	0	373	0	0	373
Total	99	59	373	39	9	579

Table of c3 by course

c3(Elite) course

Frequency	.	Catch-up	Mainstrm	Elite	No Resp	Total
.	99	0	0	0	0	99
0	0	59	373	0	9	441
1	0	0	0	39	0	39
Total	99	59	373	39	9	579

Predicting Grade with multiple regression
Initial model

The REG Procedure
Model: MODEL1
Dependent Variable: grade Final mark (if any)

Number of Observations Read	579
Number of Observations Used	287
Number of Observations with Missing Values	292

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	45864	5096.00404	27.70	<.0001
Error	277	50965	183.98945		
Corrected Total	286	96829			

Root MSE	13.56427	R-Square	0.4737
Dependent Mean	60.59582	Adj R-Sq	0.4566
Coeff Var	22.38483		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-69.54902	11.39547	-6.10
hsgpa	High School GPA	1	1.57174	0.22106	7.11
hscalc	HS Calculus	1	0.25130	0.10328	2.43
hsengl	HS English	1	-0.32735	0.12640	-2.59
precalc	Number precalculus correct	1	1.73741	0.56575	3.07
calc	Number calculus correct	1	0.71167	0.37891	1.88
c1	Catch-up	1	5.61567	4.33959	1.29
c3	Elite	1	-2.74407	2.87673	-0.95
gender	0=Male, 1=Female	1	1.62805	1.71551	0.95
mtongue	0=Other, 1=English	1	-4.48951	2.13345	-2.10

Parameter Estimates

Variable	Label	DF	Pr > |t|
Intercept	Intercept	1	<.0001
hsgpa	High School GPA	1	<.0001
hscalc	HS Calculus	1	0.0156
hsengl	HS English	1	0.0101
precalc	Number precalculus correct	1	0.0023
calc	Number calculus correct	1	0.0614
c1	Catch-up	1	0.1967
c3	Elite	1	0.3410
gender	0=Male, 1=Female	1	0.3434
mtongue	0=Other, 1=English	1	0.0362

Prediction of Performance in First-year Calculus 4
Predicting Grade with multiple regression
Initial model

The REG Procedure
Model: MODEL1

Test Course Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	249.36896	1.36	0.2596
Denominator	277	183.98945		

Prediction of Performance in First-year Calculus 5
Predicting Grade with multiple regression
Initial model

The REG Procedure
Model: MODEL1

Test Diagnostic_Test Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	1749.78270	9.51	0.0001
Denominator	277	183.98945		

Prediction of Performance in First-year Calculus 6
Predicting Grade with multiple regression
Initial model

The REG Procedure
Model: MODEL1

Test MaybeDropThese Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	4	297.22934	1.62	0.1705
Denominator	277	183.98945		

Predicting Grade with multiple regression
Model 2 (Drop course and gender, combine diagnostic test)

The REG Procedure
Model: MODEL1
Dependent Variable: grade Final mark (if any)

Number of Observations Read	579
Number of Observations Used	287
Number of Observations with Missing Values	292

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	44881	8976.29511	48.56	<.0001
Error	281	51948	184.86704		
Corrected Total	286	96829			

Root MSE	13.59658	R-Square	0.4635
Dependent Mean	60.59582	Adj R-Sq	0.4540
Coeff Var	22.43815		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error
Intercept	Intercept	1	-66.75516	11.25053
hsgpa	High School GPA	1	1.58918	0.22087
hscalc	HS Calculus	1	0.21759	0.10208
hsengl	HS English	1	-0.30024	0.12270
totscore	Total # right on diagnostic test	1	0.97213	0.25185
mtongue	0=Other, 1=English	1	-4.69657	2.12640

Parameter Estimates

Variable	Label	DF	t Value	Pr > |t|
Intercept	Intercept	1	-5.93	<.0001
hsgpa	High School GPA	1	7.20	<.0001
hscalc	HS Calculus	1	2.13	0.0339
hsengl	HS English	1	-2.45	0.0150
totscore	Total # right on diagnostic test	1	3.86	0.0001
mtongue	0=Other, 1=English	1	-2.21	0.0280

The UNIVARIATE Procedure
Variable: delstudres (Studentized Residual without Current Obs)

Moments

N	287	Sum Weights	287
Mean	-0.0011951	Sum Observations	-0.3429943
Std Deviation	1.00735115	Variance	1.01475633
Skewness	-0.7188188	Kurtosis	1.03846069
Uncorrected SS	290.220721	Corrected SS	290.220311
Coeff Variation	-84289.959	Std Error Mean	0.05946206

Basic Statistical Measures

Location		Variability	
Mean	-0.00120	Std Deviation	1.00735
Median	0.11540	Variance	1.01476
Mode	.	Range	6.28032
		Interquartile Range	1.18810

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	-0.0201	Pr > |t|	0.9840
Sign	M	6.5	Pr >= |M|	0.4788
Signed Rank	S	1533	Pr >= |S|	0.2767

Quantiles (Definition 5)

Quantile	Estimate
100% Max	2.410620
99%	1.999026
95%	1.540531
90%	1.168669
75% Q3	0.683132
50% Median	0.115404
25% Q1	-0.504969
10%	-1.366051
5%	-1.892008
1%	-3.047289
0% Min	-3.869696

Prediction of Performance in First-year Calculus 9
Predicting Grade with multiple regression
Look at Studentized deleted residuals

The UNIVARIATE Procedure
Variable: delstudres (Studentized Residual without Current Obs)

Extreme Observations

------Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-3.86970	176	1.80660	489
-3.40744	9	1.87512	463
-3.04729	195	1.99903	2
-2.86101	167	2.05720	203
-2.57695	229	2.41062	146

Missing Values

Missing Value	Count	-----Percent Of----- All Obs	Missing Obs
.	292	50.43	100.00

Prediction of Performance in First-year Calculus 10
Predicting Grade with multiple regression
Critical values for Studentized deleted residuals

Uncorrected

Uncorrected critical value of t = 1.9684725

alpha df

Alpha = 0.05 , df = 280

Corrected

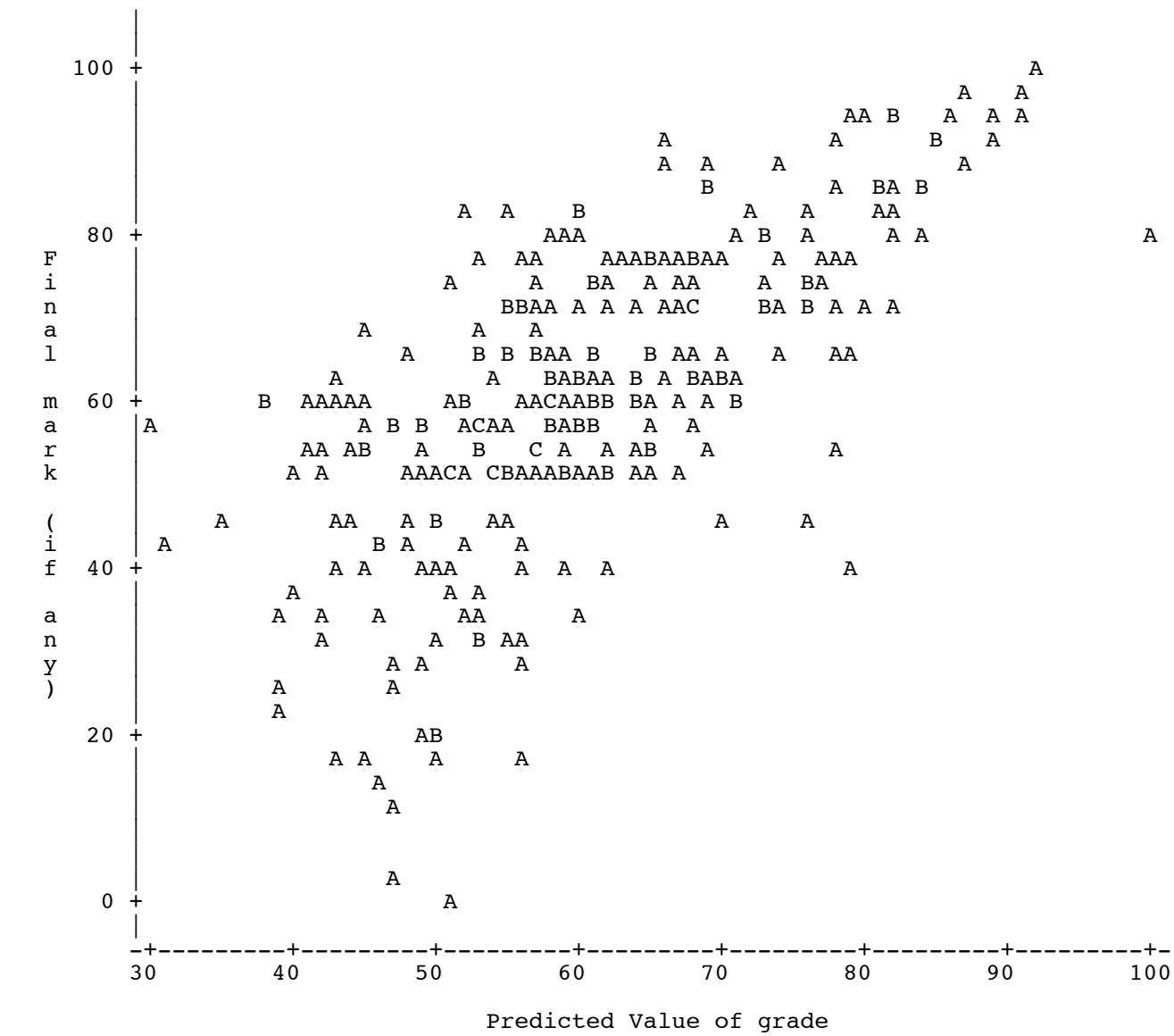
Bonferroni corected critical value of t = 3.8049308

alpha n df

Joint alpha = 0.05 , corrected for 287 tests, df = 280

Predicting Grade with multiple regression
Y by Y-hat

Plot of grade*yhat. Legend: A = 1 obs, B = 2 obs, etc.



NOTE: 292 obs had missing values.

Obs	id	grade	yhat	delstudres
1	15	.	37.4432	.
2	20	.	.	.
3	24	.	56.1867	.
4	27	.	.	.
5	30	.	.	.
6	31	.	68.9236	.
7	35	.	58.2510	.
8	40	.	.	.

Skipping …

184	569	.	44.2105	.
185	573	.	.	.
186	575	.	.	.
187	176	1	51.4408	-3.86970
188	9	2	46.9164	-3.40744
189	319	2	.	.
190	70	4	.	.
191	498	4	.	.
192	512	5	.	.
193	546	11	.	.
194	229	12	46.5314	-2.57695
195	88	13	46.4532	-2.51160
196	16	14	.	.
197	93	14	.	.
198	482	15	.	.
199	336	16	42.8064	-2.00408
200	8	17	44.5636	-2.05608

Skipping …

576	340	95	90.7773	0.32023
577	452	96	91.1178	0.36484
578	486	97	86.8452	0.75748
579	286	99	92.2925	0.50354

Predicting Grade with multiple regression
Model 2 again

The REG Procedure
Model: MODEL1
Dependent Variable: grade Final mark (if any)

Number of Observations Read	382
Number of Observations Used	283
Number of Observations with Missing Values	99

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	40293	8058.60757	49.87	<.0001
Error	277	44762	161.59432		
Corrected Total	282	85055			

Root MSE	12.71198	R-Square	0.4737
Dependent Mean	61.35336	Adj R-Sq	0.4642
Coeff Var	20.71928		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error
Intercept	Intercept	1	-55.11553	10.69456
hsgpa	High School GPA	1	1.52200	0.20737
hscalc	HS Calculus	1	0.21010	0.09578
hsengl	HS English	1	-0.35569	0.11549
totscore	Total # right on diagnostic test	1	0.92392	0.23752
mtongue	0=Other, 1=English	1	-5.09846	2.00241

Parameter Estimates

Variable	Label	DF	t Value	Pr > |t|
Intercept	Intercept	1	-5.15	<.0001
hsgpa	High School GPA	1	7.34	<.0001
hscalc	HS Calculus	1	2.19	0.0291
hsengl	HS English	1	-3.08	0.0023
totscore	Total # right on diagnostic test	1	3.89	0.0001
mtongue	0=Other, 1=English	1	-2.55	0.0114

skipping …

The UNIVARIATE Procedure
Variable: delstudres (Studentized Residual without Current Obs)

Extreme Observations

------Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-3.16749	119	1.73285	321
-3.13210	101	1.91499	301
-2.88341	27	1.98929	2
-2.48009	167	2.13409	123
-2.46917	214	2.47721	88

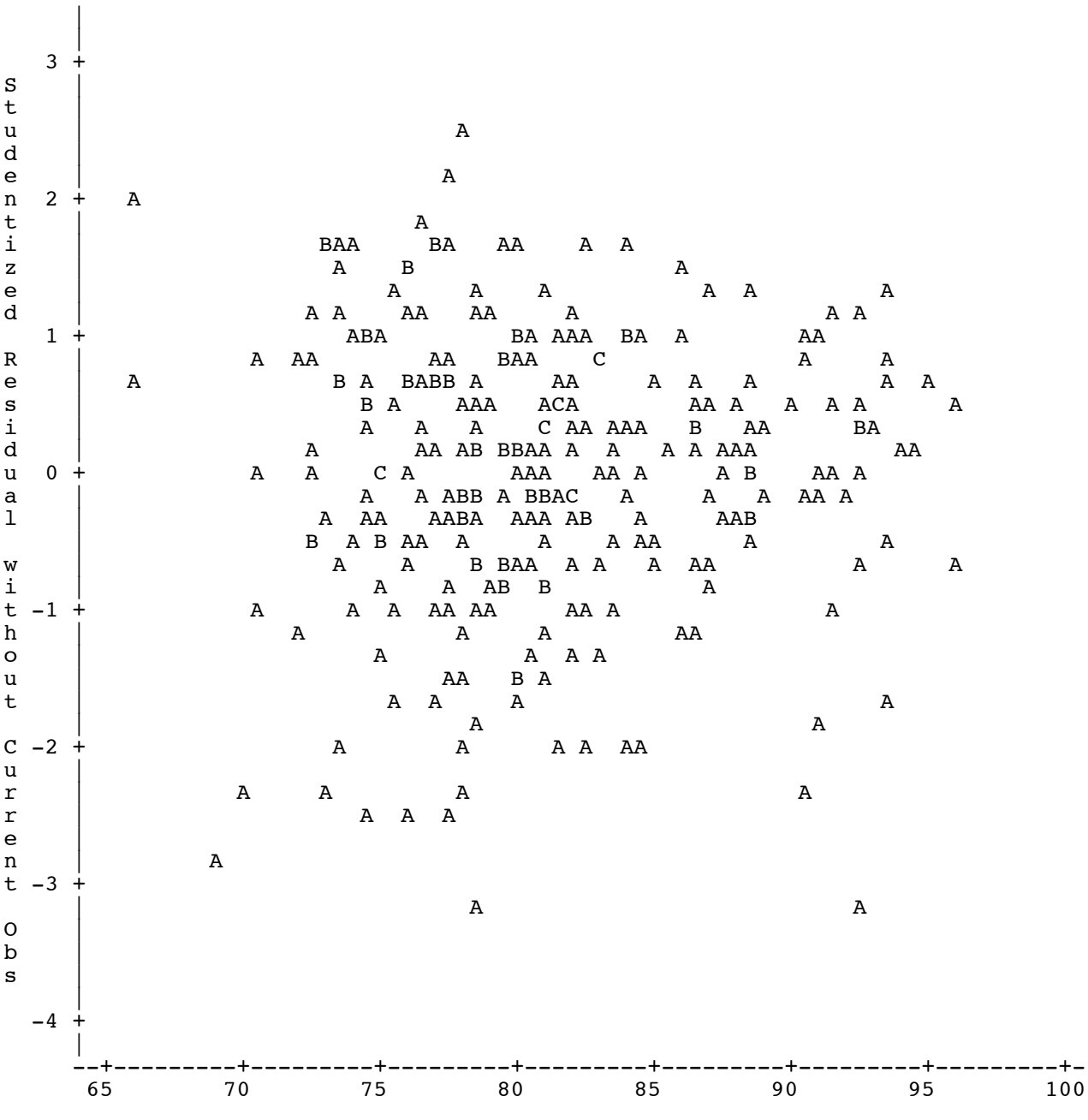
Missing Values

Missing Value	Count	-----Percent Of----- All Obs	Missing Obs
.	99	25.92	100.00

```
                Histogram                  #             Boxplot
 2.25+*                                    2                |
     .********                             15               |
     .*************                        26               |
     .**************************           53             +-----+
     .************************             49             *-----*
     .*******************************      63             |  +  |
     .*****************                    33             +-----+
     .*********                            17               |
     .*******                              14               |
     .****                                  8               0
     .*                                     1               0
-3.25+*                                     2               0
      ----+----+----+----+----+----+--
      * may represent up to 2 counts
```

Skipping …

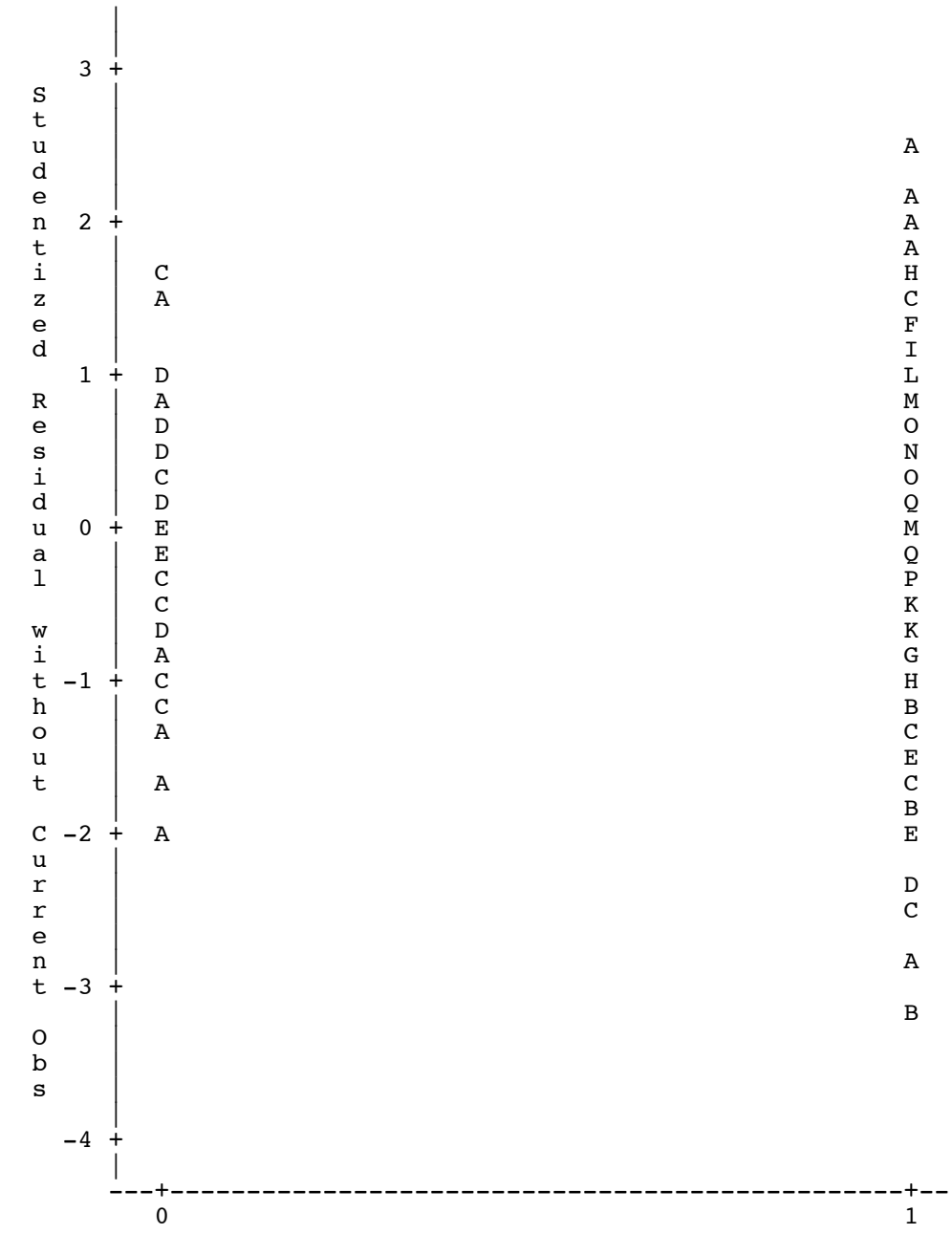
Plot of delstudres*hsgpa. Legend: A = 1 obs, B = 2 obs, etc.



NOTE: 99 obs had missing values.

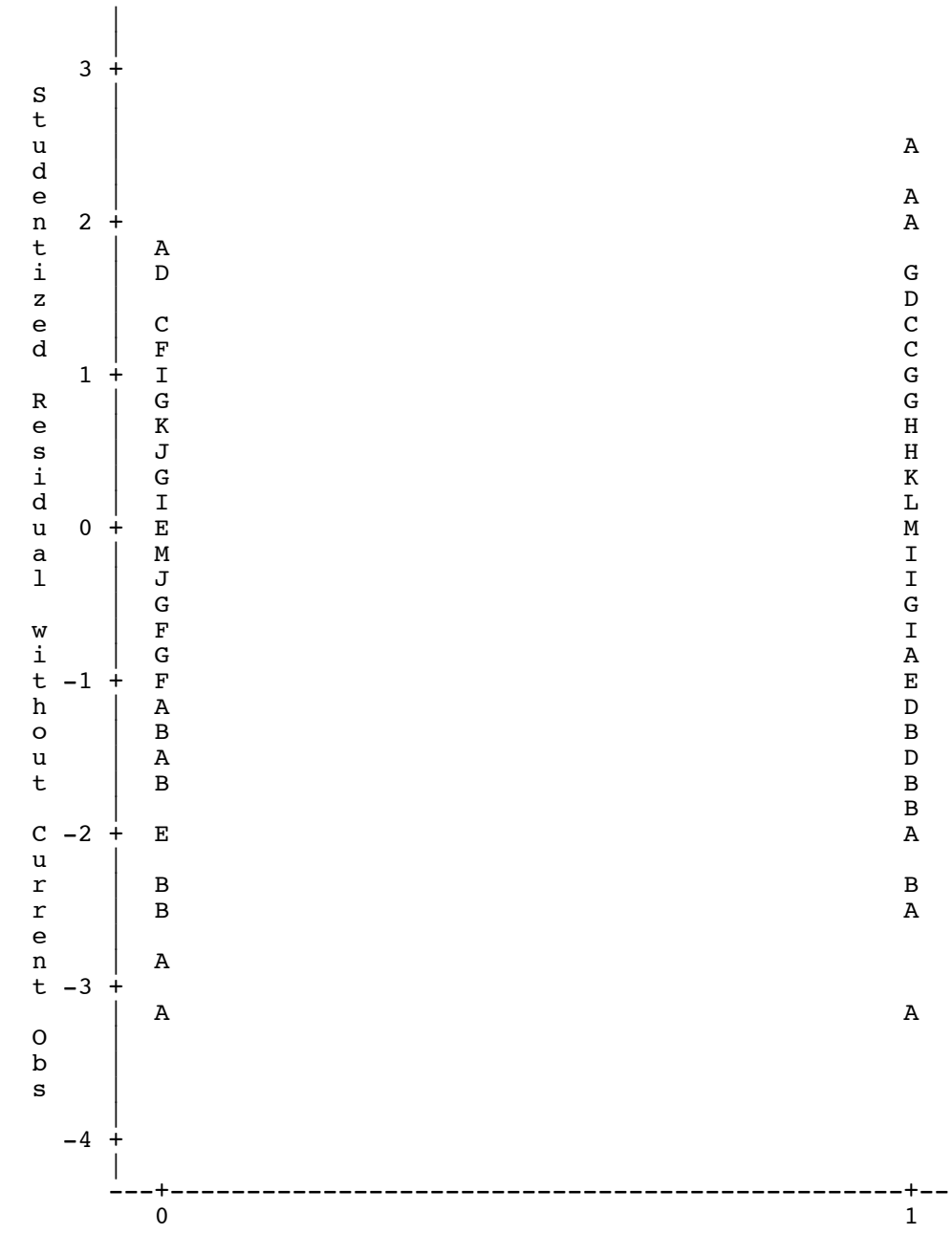
Skipping …

Plot of delstudres*mtongue. Legend: A = 1 obs, B = 2 obs, etc.



NOTE: 99 obs had missing values.

Plot of delstudres*gender. Legend: A = 1 obs, B = 2 obs, etc.



NOTE: 99 obs had missing values.

Skipping ...

Obs	id	yhat	grade	lower95	upper95	delstudres
1	1	45.552	39	20.2995	70.804	-0.51948
2	2	32.298	57	6.8320	57.765	1.98929
3	3	60.522	62	35.3854	85.658	0.11661
4	4	63.578	76	38.3479	88.808	0.98530
5	5	69.111	86	43.9607	94.261	1.33724
6	6	44.678	60	19.3175	70.039	1.22305
7	7	69.350	54	43.9464	94.753	-1.22751
8	8	45.729	17	20.4941	70.965	-2.29691
9	10	77.136	76	51.8792	102.393	-0.09005
10	11	.	60	.	.	.
11	12	.	61	.	.	.
12	13	.	54	.	.	.
13	14	71.297	84	46.0053	96.589	1.01023
14	17	62.064	75	36.6227	87.506	1.03529
15	18	79.437	94	54.1793	104.696	1.15719
16	19	53.435	60	28.2909	78.580	0.51823
17	21	60.427	53	35.0805	85.774	-0.59129
18	22	68.145	63	42.9640	93.326	-0.40667
19	23	61.782	82	36.4062	87.158	1.61816
20	25	64.257	67	39.0064	89.508	0.21738
21	26	50.281	55	25.1415	75.421	0.37235
22	28	63.475	77	38.2598	88.691	1.07247
23	29	59.945	80	34.8001	85.090	1.58968
24	32	47.961	26	22.7357	73.186	-1.74815
25	33	61.260	81	35.6400	86.881	1.59611
26	34	69.965	78	44.7483	95.182	0.63631
27	36	52.368	17	26.7948	77.941	-2.88341
28	37	.	80	.	.	.
29	38	.	66	.	.	.
30	39	.	29	.	.	.
31	43	73.187	60	47.6313	98.743	-1.06063
32	45	65.800	53	40.3965	91.204	-1.02278
33	46	40.792	26	15.5797	66.004	-1.17329
34	48	.	50	.	.	.
35	49	55.449	65	30.3397	80.559	0.75330
36	50	101.156	81	75.0253	127.287	-1.66787
37	51	63.154	51	37.9399	88.367	-0.96330
38	52	48.652	42	23.4900	73.813	-0.52547
39	53	84.971	91	59.6249	110.317	0.47987

Skipping the rest of math2.lst

Next is the proc reg output in Model2b.pdf

Number of Observations Read	382
Number of Observations Used	283
Number of Observations with Missing Values	99

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	40293	8058.60757	49.87	<.0001
Error	277	44762	161.59432		
Corrected Total	282	85055			

Root MSE	12.71198	**R-Square**	0.4737
Dependent Mean	61.35336	**Adj R-Sq**	0.4642
Coeff Var	20.71928		

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	-55.11553	10.69456	-5.15	<.0001
hsgpa	High School GPA	1	1.52200	0.20737	7.34	<.0001
hscalc	HS Calculus	1	0.21010	0.09578	2.19	0.0291
hsengl	HS English	1	-0.35569	0.11549	-3.08	0.0023
totscore	Total # right on diagnostic test	1	0.92392	0.23752	3.89	0.0001
mtongue	0=Other, 1=English	1	-5.09846	2.00241	-2.55	0.0114