```
appsrv01.srv> ls
senic0.sas  senic.dat
appsrv01.srv> sas senic0
appsrv01.srv> ls
senic0.log  senic0.lst  senic0.sas  senic.dat
appsrv01.srv>
appsrv01.srv> cat senic0.log
1                                                          The SAS System
13:59 Sunday, September 5, 2007

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 Linux 2.6.8.1-smp-athlon-bk 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           0.03 seconds
      cpu time            0.00 seconds

1          /* senic0.sas */
2          data  simple;
3                infile 'senic.dat';
4                input   id stay age infrisk culratio xratio nbeds medschl
5                        region census nurses service;

NOTE: The infile 'senic.dat' is:
      File Name=/homes/students/u0/stats/brunner/senic.dat,
      Owner Name=brunner,Group Name=stats,
      Access Permission=rw-r-----,
      File Size (bytes)=5989

NOTE: 113 records were read from the infile 'senic.dat'.
      The minimum record length was 52.
      The maximum record length was 52.
NOTE: The data set WORK.SIMPLE has 113 observations and 12 variables.
NOTE: DATA statement used:
      real time           0.00 seconds
      cpu time            0.02 seconds


6          proc freq;
7                tables _all_;
8
NOTE: There were 113 observations read from the data set WORK.SIMPLE.
NOTE: The PROCEDURE FREQ printed pages 1-20.
NOTE: PROCEDURE FREQ used:
      real time           0.02 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           0.11 seconds
      cpu time            0.03 seconds

appsrv01.srv>
appsrv01.srv> less senic0.lst
                                                              The SAS System
                 13:59 Sunday, September 5, 2007   1

                                                            The FREQ Procedure

                                                                        Cumulative
    Cumulative
                                              id    Frequency     Percent     Frequency
      Percent
                                             ------------------------------------------
--------------
                                               1           1        0.88             1
        0.88
                                               2           1        0.88             2
        1.77
                                               3           1        0.88             3
        2.65
                                               4           1        0.88             4
        3.54
                                               5           1        0.88             5
        4.42
                                               6           1        0.88             6
        5.31
                                               7           1        0.88             7
        6.19
                                               8           1        0.88             8
        7.08
                                               9           1        0.88             9
        7.96
                                              10           1        0.88            10
        8.85
                                              11           1        0.88            11
        9.73
                                              12           1        0.88            12
       10.62
                                              13           1        0.88            13
       11.50
                                              14           1        0.88            14
       12.39
                                              15           1        0.88            15
       13.27
                                              16           1        0.88            16
       14.16
                                              17           1        0.88            17

:
```

```
appsrv01.srv>
appsrv01.srv> cp senic0.sas senic0.1.sas
appsrv01.srv> emacs senic0.1.sas

appsrv01.srv> cat senic0.1.sas

/* senic0.1.sas */
options linesize = 79;
data  simple;
      infile 'senic.dat';
      input   id stay age infrisk culratio xratio nbeds medschl
              region census nurses service;

      /*** sas doesn't like numeric missing value codes.  a period  .  is
                      best for missing. however ....  ***/

      if stay eq 9999 then stay = .  ;
      if age eq 9999 then age = .  ;
      if xratio eq 9999 then xratio = .  ;
      if culratio eq 9999 then culratio = .  ;
      if infrisk = 999 then infrisk  = .  ;
      if nbeds = 9 then nbeds  = .  ;
      if medschl = 9 then medschl  = .  ;
      if region = 9 then region  = .  ;
      if census = 9 then census  = .  ;
      if service = 9 then service  = .  ;
      if nurses eq (0 or .999) then nurses = .  ;

proc freq;
      tables _all_;


appsrv01.srv> sas senic0.1.lst
appsrv01.srv> less senic0.1.log
appsrv01.srv> less senic0.1.lst

                              The SAS System                              1
                                        14:21 Sunday, September 5, 2007

                            The FREQ Procedure

                                                Cumulative    Cumulative
          id    Frequency     Percent     Frequency      Percent
         -----------------------------------------------------------
           1           1        0.88             1         0.88
           2           1        0.88             2         1.77
           3           1        0.88             3         2.65
           4           1        0.88             4         3.54
           5           1        0.88             5         4.42
           6           1        0.88             6         5.31
           7           1        0.88             7         6.19
           8           1        0.88             8         7.08
           9           1        0.88             9         7.96
          10           1        0.88            10         8.85
```

11	1	0.88	11	9.73
12	1	0.88	12	10.62
13	1	0.88	13	11.50
14	1	0.88	14	12.39
15	1	0.88	15	13.27

Skipping some lines ...

50	1	0.88	50	44.25
51	1	0.88	51	45.13
52	1	0.88	52	46.02

The FREQ Procedure

id	Frequency	Percent	Cumulative Frequency	Cumulative Percent
53	1	0.88	53	46.90
54	1	0.88	54	47.79
55	1	0.88	55	48.67
56	1	0.88	56	49.56
57	1	0.88	57	50.44
58	1	0.88	58	51.33
59	1	0.88	59	52.21
60	1	0.88	60	53.10
61	1	0.88	61	53.98
62	1	0.88	62	54.87
63	1	0.88	63	55.75
64	1	0.88	64	56.64

And so on. Some of the frequency distributions are less ugly

The FREQ Procedure

medschl	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	17	15.04	17	15.04
2	96	84.96	113	100.00

region	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	29	25.66	29	25.66
2	32	28.32	61	53.98
3	36	31.86	97	85.84
4	16	14.16	113	100.00

```
/*******  senicread.sas  just reads and labels SENIC data  *************
*                                                                      *
* Reading from the data file senic.raw, in which the variables are in  *
* fixed columns and missing data are represented by blanks.            *
*                                                                      *
************************************************************************/

title 'Study of the Effectiveness of Nosocomial Infection Control';
options linesize=79 noovp formdlim='_';

proc format;  /* value labels used in data step below */
     value yesnofmt 1 = 'Yes'   2 = 'No' ;
     value regfmt  1 =  'Northeast'
                   2 =  'North Central'
                   3 =  'South'
                   4 =  'West' ;
     value acatfmt 1 = '53 & under' 2 = 'Over 53';

data senic;
     infile 'senic.raw' n=1 missover ;
                 /* n=1 lines of data per case (unnecessary here)  */
                 /* missover helps blank=missing to be done properly at end of line */
     input
         #1  id       1-5
             stay     7-11
             age      13-16
             infrisk  18-20
             culratio 22-25
             xratio   27-31
             nbeds    33-35
             medschl  37
             region   39
             census   41-43
             nurses   45-47
             service  49-52 ;
     label id       = 'Hospital identification number'
           stay     = 'Av length of hospital stay, in days'
           age      = 'Average patient age'
           infrisk  = 'Prob of acquiring infection in hospital'
           culratio = '# cultures  / # no hosp acq infect'
           xratio   = '# x-rays / # no signs of pneumonia'
           nbeds    = 'Average # beds during study period'
           medschl  = 'Medical school affiliation'
           region   = 'Region of country (usa)'
           census   = 'Aver # patients in hospital per day'
           nurses   = 'Aver # nurses during study period'
           service  = '% of 35 potential facil. & services' ;
        /* associating variables with their value labels */
     format medschl yesnofmt.;
     format region  regfmt.;

        /***** recodes, computes & ifs  *****/

     if 0<age<=53 then agecat=1;
     else if age>53 then agecat=2;
     label   agecat = 'av patient age category';
```

```
     format agecat acatfmt.;

  /*  compute ad hoc index of hospital quality  */

     quality=(2*service+nurses+nbeds+10*culratio
                  +10*xratio-2*stay)/medschl;
     if (region eq 3) then quality=quality-100;
     label quality = 'jerry''s bogus hospital quality index';

/* Commented out

proc freq;
      tables _all_;  */



/****************** senicdescr.sas ***************/
/*            Descriptive stats on SENIC Data      */
/************************************************/

%include 'senicread.sas';  /* senicread.sas reads data, etc.  */
title2 'Descriptive Statistics';

proc freq;
     title3 'Frequency distributions of categorical variables';
     tables medschl region agecat;

proc means n mean std;
     title3 'Means and SDs of quantitative variables';
     var stay -- nbeds census nurses service;
     /* single dash only works with numbered lists, like item1-item50    */

proc univariate  plot normal ; /* Plots and a test for normality */
     title3 'Describe Quantitative Variables in More Detail' ;
     var stay -- nbeds census nurses service;
```

```
          Study of the Effectiveness of Nosocomial Infection Control          1
                           Descriptive Statistics
                Frequency distributions of categorical variables
                                              14:58 Sunday, September 5, 2007

                              The FREQ Procedure

                           Medical school affiliation

                                                 Cumulative    Cumulative
        medschl    Frequency     Percent     Frequency      Percent
        ------------------------------------------------------------
            Yes           17       15.04            17        15.04
            No            96       84.96           113       100.00
```

Region of country (usa)

region	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Northeast	29	25.66	29	25.66
North Central	32	28.32	61	53.98
South	36	31.86	97	85.84
West	16	14.16	113	100.00

av patient age category

agecat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
53 & under	56	49.56	56	49.56
Over 53	57	50.44	113	100.00

Descriptive Statistics
Means and SDs of quantitative variables
14:58 Sunday, September 5, 2007

The MEANS Procedure

Variable	Label	N	Mean
stay	Av length of hospital stay, in days	113	9.6483186
age	Average patient age	113	53.2318584
infrisk	Prob of acquiring infection in hospital	113	4.3548673
culratio	# cultures / # no hosp acq infect	113	15.6840708
xratio	# x-rays / # no signs of pneumonia	113	81.6300885
nbeds	Average # beds during study period	113	252.1769912
census	Aver # patients in hospital per day	113	191.3716814
nurses	Aver # nurses during study period	113	173.2477876
service	% of 35 potential facil. & services	113	43.1548673

Variable	Label	Std Dev
stay	Av length of hospital stay, in days	1.9114560
age	Average patient age	4.4616074
infrisk	Prob of acquiring infection in hospital	1.3409080
culratio	# cultures / # no hosp acq infect	10.1830441
xratio	# x-rays / # no signs of pneumonia	19.3667373
nbeds	Average # beds during study period	192.8451558
census	Aver # patients in hospital per day	153.7595639
nurses	Aver # nurses during study period	139.2653897
service	% of 35 potential facil. & services	15.2001879

The UNIVARIATE Procedure
Variable: stay (Av length of hospital stay, in days)

Moments

N	113	Sum Weights	113
Mean	9.64831858	Sum Observations	1090.26
Std Deviation	1.91145602	Variance	3.65366411
Skewness	2.06891744	Kurtosis	8.07748986
Uncorrected SS	10928.3862	Corrected SS	409.210381
Coeff Variation	19.8112863	Std Error Mean	0.17981466

Basic Statistical Measures

Location		Variability	
Mean	9.648319	Std Deviation	1.91146
Median	9.420000	Variance	3.65366
Mode	7.140000	Range	12.86000
		Interquartile Range	2.13000

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	53.65702	Pr > |t|	<.0001
Sign	M	56.5	Pr >= |M|	<.0001
Signed Rank	S	3220.5	Pr >= |S|	<.0001

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.853032	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.086861	Pr > D	0.0357
Cramer-von Mises	W-Sq	0.289727	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2.228908	Pr > A-Sq	<0.0050

Quantiles (Definition 5)

Quantile	Estimate
100% Max	19.56
99%	17.94
95%	12.07

90%	11.48
75% Q3	10.47
50% Median	9.42
25% Q1	8.34
10%	7.67
5%	7.14

The UNIVARIATE Procedure
Variable: stay (Av length of hospital stay, in days)

Quantiles (Definition 5)

Quantile	Estimate
1%	7.08
0% Min	6.70

Extreme Observations

----Lowest----		----Highest----	
Value	Obs	Value	Obs
6.70	76	12.78	13
7.08	72	13.59	34
7.13	1	13.95	104
7.14	107	17.94	112
7.14	103	19.56	47

Stem	Leaf	#	Boxplot
19	6	1	*
19			
18			
18			
17	9	1	*
17			
16			
16			
15			
15			
14			
14	0	1	0
13	6	1	|
13			|
12	8	1	|
12	01	2	|
11	556688	6	|

```
 11 0112222244              10               |
 10 577889                   6             +-----+
 10 000122223344            12             |     |
  9 55677788888899          14             |  +  |
  9 000122344444            12             *-----*
  8 556677888999999         15             |     |
  8 0001222333344           13             +-----+
  7 56667788999             11                |
  7 111114                   6                |
  6 7                        1                |
    ----+----+----+----+
```

Descriptive Statistics

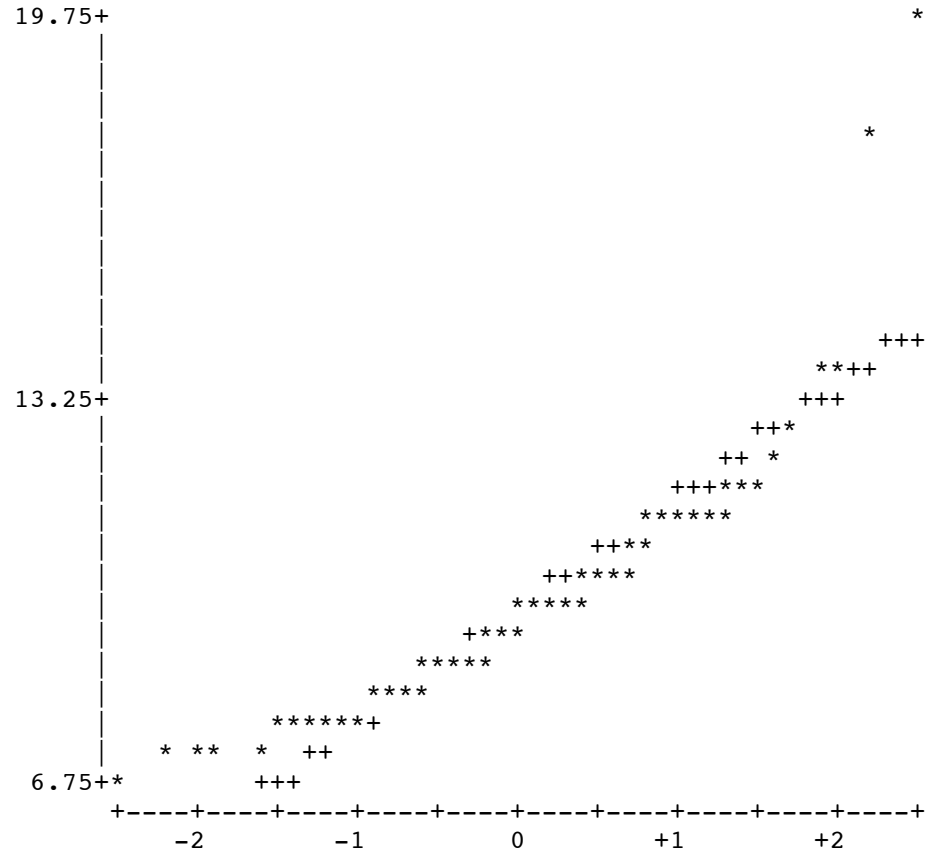
Describe Quantitative Variables in More Detail

14:58 Sunday, September 5, 2007

The UNIVARIATE Procedure

Variable: stay (Av length of hospital stay, in days)

Normal Probability Plot



```
 /****************** basicsenic.sas ***************/
/*             Elementary tests on SENIC Data      */
/*************************************************/
%include 'senicread.sas';  /* senicread.sas reads data, etc.  */
title2 'Elementary tests on SENIC Data';

proc freq;
     title3 'Use proc freq to do crosstabs with chisquare test';
     tables region*medschl / nocol nopercent expected chisq;
proc ttest;
     title3 'T-test: Less risk at Hospitals with Med School Affiliation?';
     class medschl;
     var infrisk age ;
proc glm;
     title3 'One-way anova with proc glm';
     class region;
     model infrisk=region;
     means region / ;
proc plot;
     title3 'Scatterplot';
     plot infrisk * nurses
          infrisk * nurses = medschl;
proc corr;
     title3 'Correlation Matrix';
     var stay -- nbeds census nurses service;
proc glm;
     title3 'Simple regression with proc glm';
     model infrisk=nurses;


cat basicsenic.lst
```

```
          Study of the Effectiveness of Nosocomial Infection Control          1
                        Elementary tests on SENIC Data
                    Use proc freq to do crosstabs with chisquare test
                                                15:13 Sunday, September 5, 2007

                              The FREQ Procedure

                            Table of region by medschl

                region(Region of country (usa))
                                medschl(Medical school affiliation)
                Frequency     |
                Expected      |
                Row Pct       |Yes     |No      |  Total
                --------------+--------+--------+
                Northeast     |      6 |     23 |     29
                              | 4.3628 | 24.637 |
                              |  20.69 |  79.31 |
                --------------+--------+--------+
```

North Central	7	25	32
	4.8142	27.186	
	21.88	78.13	
South	2	34	36
	5.4159	30.584	
	5.56	94.44	
West	2	14	16
	2.4071	13.593	
	12.50	87.50	
Total	17	96	113

Statistics for Table of region by medschl

Statistic	DF	Value	Prob
Chi-Square	3	4.5084	0.2115
Likelihood Ratio Chi-Square	3	5.0108	0.1710
Mantel-Haenszel Chi-Square	1	2.3105	0.1285
Phi Coefficient		0.1997	
Contingency Coefficient		0.1959	
Cramer's V		0.1997	

WARNING: 38% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 113

The TTEST Procedure

Statistics

Variable	medschl	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev
infrisk	Yes	17	4.5223	5.0941	5.6659	0.8283	1.1121
infrisk	No	96	3.9524	4.224	4.4955	1.1738	1.3403
infrisk	Diff (1-2)		0.1872	0.8702	1.5531	1.1579	1.3099
age	Yes	17	50.061	51.7	53.339	2.3745	3.1883
age	No	96	52.569	53.503	54.437	4.0383	4.611
age	Diff (1-2)		-4.115	-1.803	0.509	3.9198	4.4342

Statistics

Variable	medschl	Upper CL Std Dev	Std Err	Minimum	Maximum
infrisk	Yes	1.6926	0.2697	2.9	7.7
infrisk	No	1.5622	0.1368	1.3	7.8
infrisk	Diff (1-2)	1.5081	0.3447		
age	Yes	4.8523	0.7733	44.2	56.8
age	No	5.3746	0.4706	38.8	65.9
age	Diff (1-2)	5.1053	1.1668		

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
infrisk	Pooled	Equal	111	2.52	0.0130
infrisk	Satterthwaite	Unequal	25	2.88	0.0081
age	Pooled	Equal	111	-1.55	0.1251
age	Satterthwaite	Unequal	29.4	-1.99	0.0557

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
infrisk	Folded F	95	16	1.45	0.4026
age	Folded F	95	16	2.09	0.0957

The GLM Procedure

Class Level Information

Class	Levels	Values
region	4	North Central Northeast South West

Number of observations 113

The GLM Procedure

Dependent Variable: infrisk Prob of acquiring infection in hospital

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	17.5750218	5.8583406	3.47	0.0186
Error	109	183.8048012	1.6862826		
Corrected Total	112	201.3798230			

R-Square	Coeff Var	Root MSE	infrisk Mean
0.087273	29.81881	1.298569	4.354867

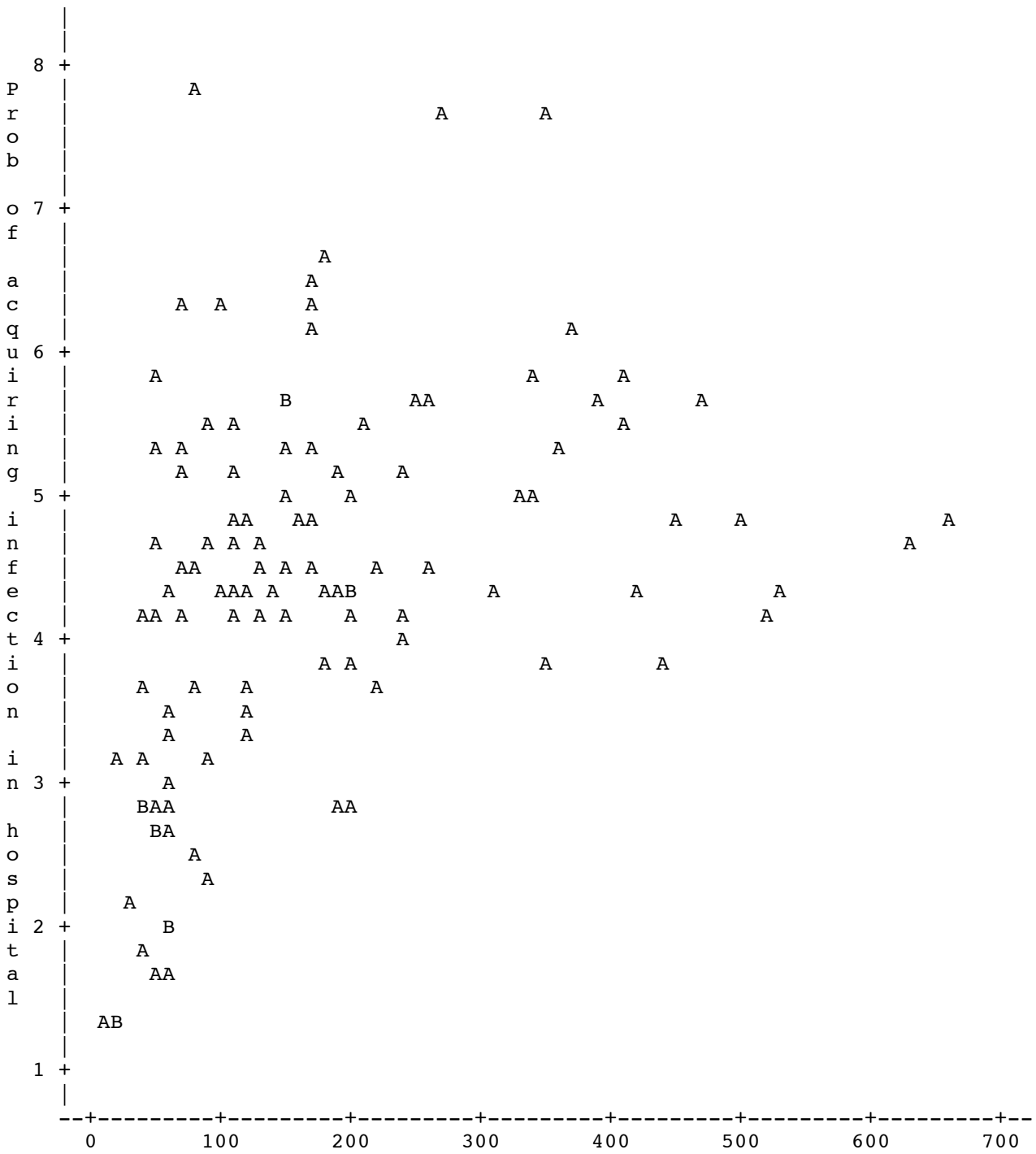
Source	DF	Type I SS	Mean Square	F Value	Pr > F
region	3	17.57502176	5.85834059	3.47	0.0186

Source	DF	Type III SS	Mean Square	F Value	Pr > F
region	3	17.57502176	5.85834059	3.47	0.0186

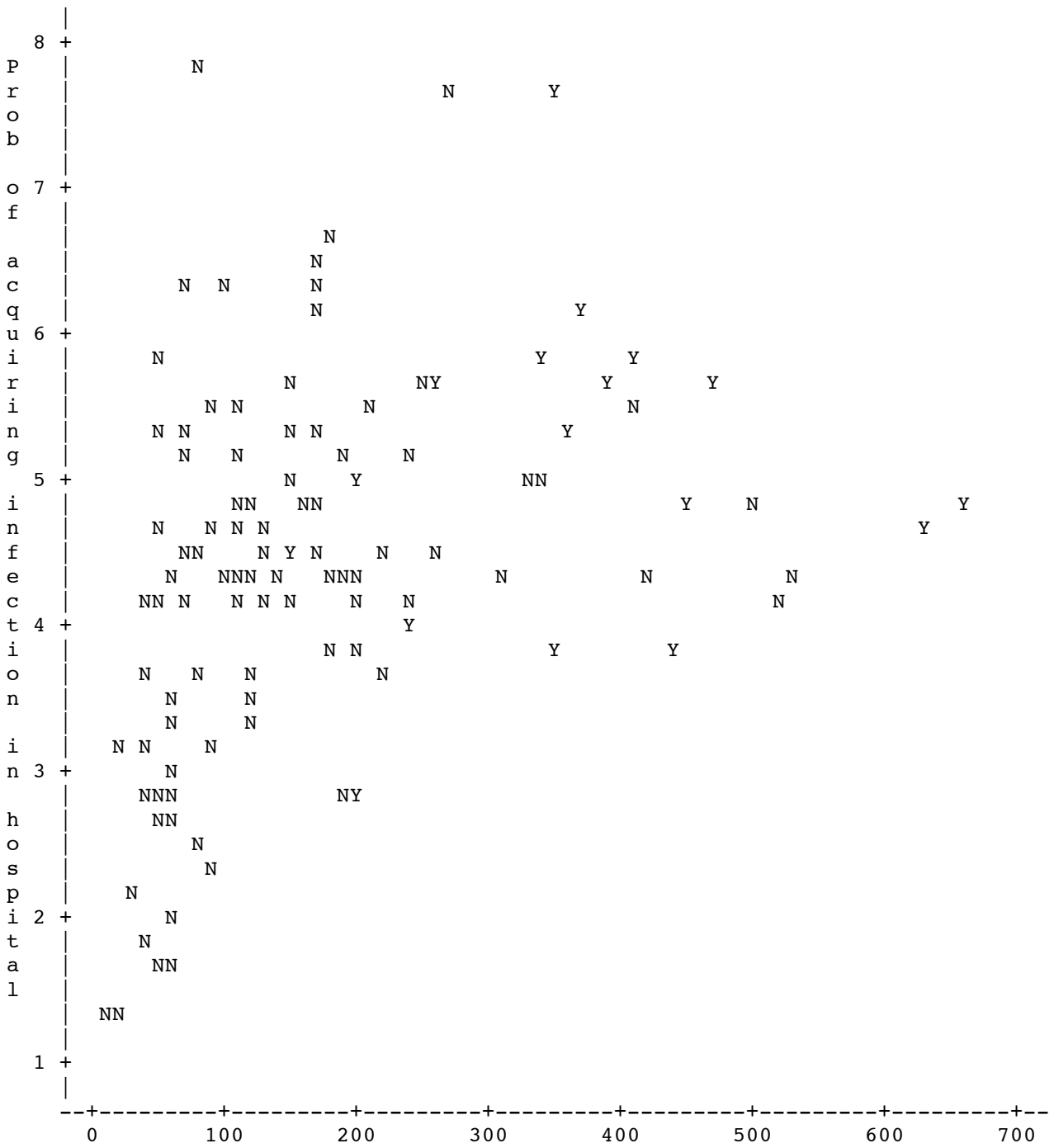
The GLM Procedure

Level of region	N	infrisk Mean	infrisk Std Dev
North Central	32	4.39375000	1.33921920
Northeast	29	4.90689655	1.27277285
South	36	3.86388889	1.42751588
West	16	4.38125000	0.87652248

Plot of infrisk*nurses. Legend: A = 1 obs, B = 2 obs, etc.



Plot of infrisk*nurses. Symbol is value of medschl.



NOTE: 6 obs hidden.

The CORR Procedure

9 Variables: stay age infrisk culratio xratio nbeds census nurses service

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
stay	113	9.64832	1.91146	1090	6.70000	19.56000
age	113	53.23186	4.46161	6015	38.80000	65.90000
infrisk	113	4.35487	1.34091	492.10000	1.30000	7.80000
culratio	113	15.68407	10.18304	1772	1.60000	60.50000
xratio	113	81.63009	19.36674	9224	39.60000	133.50000
nbeds	113	252.17699	192.84516	28496	29.00000	835.00000
census	113	191.37168	153.75956	21625	20.00000	791.00000
nurses	113	173.24779	139.26539	19577	14.00000	656.00000
service	113	43.15487	15.20019	4877	5.70000	80.00000

Simple Statistics

Variable	Label
stay	Av length of hospital stay, in days
age	Average patient age
infrisk	Prob of acquiring infection in hospital
culratio	# cultures / # no hosp acq infect
xratio	# x-rays / # no signs of pneumonia
nbeds	Average # beds during study period
census	Aver # patients in hospital per day
nurses	Aver # nurses during study period
service	% of 35 potential facil. & services

Pearson Correlation Coefficients, N = 113
Prob > |r| under H0: Rho=0

	stay	age	infrisk	culratio
stay	1.00000	0.18891	0.53344	0.33354
Av length of hospital stay, in days		0.0451	<.0001	0.0003
age	0.18891	1.00000	0.00109	-0.23593
Average patient age	0.0451		0.9908	0.0119
infrisk	0.53344	0.00109	1.00000	0.57914
Prob of acquiring infection in hospital	<.0001	0.9908		<.0001
culratio	0.33354	-0.23593	0.57914	1.00000
# cultures / # no hosp acq infect	0.0003	0.0119	<.0001	
xratio	0.38236	-0.01893	0.45329	0.42707
# x-rays / # no signs of pneumonia	<.0001	0.8423	<.0001	<.0001

Elementary tests on SENIC Data
Correlation Matrix
15:13 Sunday, September 5, 2007

The CORR Procedure

Pearson Correlation Coefficients, N = 113
Prob > |r| under H0: Rho=0

	stay	age	infrisk	culratio
nbeds	0.40924	-0.05884	0.35976	0.14879
Average # beds during study period	<.0001	0.5359	<.0001	0.1158
census	0.47389	-0.05477	0.38141	0.15288
Aver # patients in hospital per day	<.0001	0.5645	<.0001	0.1060
nurses	0.34037	-0.08294	0.39398	0.20742
Aver # nurses during study period	0.0002	0.3824	<.0001	0.0275
service	0.35572	-0.04064	0.41303	0.18239
% of 35 potential facil. & services	0.0001	0.6691	<.0001	0.0532

Pearson Correlation Coefficients, N = 113
Prob > |r| under H0: Rho=0

	xratio	nbeds	census	nurses
stay	0.38236	0.40924	0.47389	0.34037
Av length of hospital stay, in days	<.0001	<.0001	<.0001	0.0002
age	-0.01893	-0.05884	-0.05477	-0.08294

Average patient age	0.8423	0.5359	0.5645	0.3824
infrisk	0.45329	0.35976	0.38141	0.39398
Prob of acquiring infection in hospital	<.0001	<.0001	<.0001	<.0001
culratio	0.42707	0.14879	0.15288	0.20742
# cultures / # no hosp acq infect	<.0001	0.1158	0.1060	0.0275
xratio	1.00000	0.04597	0.06304	0.07754
# x-rays / # no signs of pneumonia		0.6288	0.5071	0.4143
nbeds	0.04597	1.00000	0.98100	0.91550
Average # beds during study period	0.6288		<.0001	<.0001
census	0.06304	0.98100	1.00000	0.90790
Aver # patients in hospital per day	0.5071	<.0001		<.0001
nurses	0.07754	0.91550	0.90790	1.00000
Aver # nurses during study period	0.4143	<.0001	<.0001	
service	0.11199	0.79481	0.77835	0.78375
% of 35 potential facil. & services	0.2376	<.0001	<.0001	<.0001

Study of the Effectiveness of Nosocomial Infection Control 13
Elementary tests on SENIC Data
Correlation Matrix
15:13 Sunday, September 5, 2007

The CORR Procedure

Pearson Correlation Coefficients, N = 113
Prob > |r| under H0: Rho=0

	service
stay	0.35572
Av length of hospital stay, in days	0.0001
age	-0.04064
Average patient age	0.6691
infrisk	0.41303
Prob of acquiring infection in hospital	<.0001
culratio	0.18239
# cultures / # no hosp acq infect	0.0532
xratio	0.11199
# x-rays / # no signs of pneumonia	0.2376
nbeds	0.79481
Average # beds during study period	<.0001

census Aver # patients in hospital per day	0.77835 <.0001
nurses Aver # nurses during study period	0.78375 <.0001
service % of 35 potential facil. & services	1.00000

The GLM Procedure

Dependent Variable: infrisk Prob of acquiring infection in hospital

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	31.2584372	31.2584372	20.40	<.0001
Error	111	170.1213858	1.5326251		
Corrected Total	112	201.3798230			

R-Square	Coeff Var	Root MSE	infrisk Mean
0.155221	28.42779	1.237992	4.354867

Source	DF	Type I SS	Mean Square	F Value	Pr > F
nurses	1	31.25843716	31.25843716	20.40	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
nurses	1	31.25843716	31.25843716	20.40	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	3.697664892	0.18638716	19.84	<.0001
nurses	0.003793424	0.00083997	4.52	<.0001