

SENIC Handout Two: Descriptive Statistics and a Few Elementary Tests

```
/* 2201senic2.sas */
title2 'Descriptive Statistics and a Few Elementary Tests';
%include '2201senicdef.sas';
options pagesize=100;

proc means n mean stddev;
  title2 'Describe continuous variables';
  var stay -- nbeds census nurses service;
proc freq;
  title2 'Describe categorical variables';
  tables agecat medschl region;

PROC TTEST;
  title3 'Independent t-test with PROC TTEST';
  CLASS MEDSCHL;
  VAR INFRISK;

PROC GLM;
  title3 'One-way ANOVA and followups with PROC GLM';
  CLASS REGION;
  MODEL INFRISK=REGION;
  means region;
  MEANS REGION/ bon;

/* Labels can get in the way. Create a new SAS data set without labels */
data without;
  set better;
  label id          = ' ' stay      = ' ' age        = ' ' infrisk = ' '
        culratio   = ' ' xratio    = ' ' nbeds      = ' ' medschl = ' '
        region     = ' ' census    = ' ' nurses     = ' ' service = ' ';

proc corr nosimple;
  var stay age nbeds census nurses xratio culratio;
  with infrisk;
```

SENIC data 1
Describe continuous variables
14:39 Sunday, February 19, 2006

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

SENIC data 2
Describe categorical variables
14:39 Sunday, February 19, 2006

The FREQ Procedure

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

medical school affiliation

medschl	Frequency	Percent	Cumulative Frequency	Cumulative 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

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

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		

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

Equality of Variances

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

The GLM Procedure

Class Level Information

Class	Levels	Values
region	4	North central Northeast South West

Number of observations 113

SENIC data 5

Describe categorical variables
 One-way ANOVA and followups with PROC GLM
 14:39 Sunday, February 19, 2006

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

SENIC data 6

Describe categorical variables
 One-way ANOVA and followups with PROC GLM
 14:39 Sunday, February 19, 2006

The GLM Procedure

Level of region	N	-----infrisk----- Mean	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

Describe categorical variables
 One-way ANOVA and followups with PROC GLM

14:39 Sunday, February 19, 2006

The GLM Procedure

Bonferroni (Dunn) t Tests for infrisk

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Alpha 0.05
 Error Degrees of Freedom 109
 Error Mean Square 1.686283
 Critical Value of t 2.68726

Comparisons significant at the 0.05 level are indicated by ***.

region Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
Northeast - North central	0.5131	-0.3815	1.4078
Northeast - West	0.5256	-0.5611	1.6124
Northeast - South	1.0430	0.1723	1.9137 ***
North central - Northeast	-0.5131	-1.4078	0.3815
North central - West	0.0125	-1.0560	1.0810
North central - South	0.5299	-0.3180	1.3777
West - Northeast	-0.5256	-1.6124	0.5611
West - North central	-0.0125	-1.0810	1.0560
West - South	0.5174	-0.5311	1.5659
South - Northeast	-1.0430	-1.9137	-0.1723 ***
South - North central	-0.5299	-1.3777	0.3180
South - West	-0.5174	-1.5659	0.5311

The CORR Procedure

1 With Variables: infrisk
 7 Variables: stay age nbeds census nurses xratio culratio

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

	stay	age	nbeds	census	nurses	xratio	culratio
infrisk	0.53344 <.0001	0.00109 0.9908	0.35976 <.0001	0.38141 <.0001	0.39398 <.0001	0.45329 <.0001	0.57914 <.0001