Computer Handout Two

```
/*******  senic2.sas  ***********/
%include 'senicdef.sas'; /* Effectively, Copy the file senicdef.sas to here */

title2 'Elementary statistical tests';
proc freq;
     title3 'Use proc freq to do chi-square test of independence';
     tables region*medschl / nocol nopercent expected chisq;
proc ttest;
     title3 'Independent t-test with proc ttest';
     class medschl;
     var infrisk age ;
proc glm;
     title3 'One-way anova and followups with proc glm';
     class region;
     model infrisk=region;
     means region;
     means region/ bon;  /* The second means statement gives
                            simultaneous confidence intervals */
proc plot;
     title3 'Scatterplots with proc corr';
     plot infrisk * nurses
          infrisk * nurses = medschl;
proc corr;
     title3 'Correlation matrix with proc corr';
     var stay -- nbeds census nurses service;
proc reg;
     title3 'Simple regression with proc reg';
     model infrisk=nurses;
```

Just show the one-factor output from `proc glm`

```
        Study of the Efficacy of Nosocomial Infection Control (SENIC)        3
                        Elementary statistical tests
                   One-way anova and followups with proc glm

                              The GLM Procedure

                           Class Level Information

          Class         Levels    Values

          region             4    North Central Northeast South West


                    Number of Observations Read         113
                    Number of Observations Used         112
```

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	16.4009623	5.4669874	3.22	0.0255
Error	108	183.1533234	1.6958641		
Corrected Total	111	199.5542857			

R-Square	Coeff Var	Root MSE	infrisk Mean
0.082188	29.98610	1.302253	4.342857

Source	DF	Type I SS	Mean Square	F Value	Pr > F
region	3	16.40096230	5.46698743	3.22	0.0255

Source	DF	Type III SS	Mean Square	F Value	Pr > F
region	3	16.40096230	5.46698743	3.22	0.0255

The GLM Procedure

Level of region	N	infrisk Mean	infrisk Std Dev
North Central	32	4.39375000	1.33921920
Northeast	28	4.87857143	1.28678674
South	36	3.86388889	1.42751588
West	16	4.38125000	0.87652248

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	108
Error Mean Square	1.695864
Critical Value of t	2.68772

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

region Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Northeast - North Central	0.4848	-0.4209	1.3906	
Northeast - West	0.4973	-0.5996	1.5942	
Northeast - South	1.0147	0.1327	1.8966	***
North Central - Northeast	-0.4848	-1.3906	0.4209	
North Central - West	0.0125	-1.0592	1.0842	
North Central - South	0.5299	-0.3205	1.3802	
West - Northeast	-0.4973	-1.5942	0.5996	
West - North Central	-0.0125	-1.0842	1.0592	
West - South	0.5174	-0.5343	1.5690	
South - Northeast	-1.0147	-1.8966	-0.1327	***
South - North Central	-0.5299	-1.3802	0.3205	
South - West	-0.5174	-1.5690	0.5343	

```
/*******  senic3.sas  ***********/
%include 'senicdef.sas'; /* Effectively, Copy the file senicdef.sas to here */

proc glm;
     title3 'Med School and infection risk ignoring hospital size';
     class medschl;
     model infrisk=medschl;
     means medschl;

proc glm;
     title3 'Med School and infection risk controlling for hospital size';
     class medschl;
     model infrisk = census nbeds medschl;
     lsmeans medschl;

proc glm;
     title3 'One-factor ancova on Region with proc glm';
     class region;
     model infrisk = census nbeds region;
     lsmeans region / pdiff  adjust=bon;
     /* Bonferroni adjustment of the p-value is just to multiply the
        p-value from the one-at-a-time test by the number of tests --
        six, in this case. */

proc reg;
     title3 'One-factor ancova on Region with proc reg';
     model infrisk = census nbeds r1 r2 r4;
     region: test r1=r2=r4=0;
     /* Pairwise comparisons with custom tests: Comparisons with South
        are part of the default output, but do them again for completeness. */
     NCvsNE: test r1=r2;
     NCvsS : test r2=0;
     NCvsW : test r2=r4;
     NEvsS : test r1=0;
     NEvsW : test r1=r4;
     SvsW  : test r4=0;
     /* For Bonferroni multiple comparisons, believe any
        p < 0.05/6 = 0.0083   */

proc glm;
     title3 'Two-factor ANOVA with proc glm';
     class region medschl;
     model infrisk = region|medschl;
     means region|medschl;

proc reg;
     title3 'Two-factor ANCOVA with proc reg';
     model infrisk =  census nbeds reg1-reg3 ms1 mr1-mr3;
     regtest:  test reg1=reg2=reg3=0;
     mstest:   test ms1=0;
     m_by_r:   test mr1=mr2=mr3=0;

proc glm;
     title3 'Two-factor Analysis of Covariance with proc glm';
     class region medschl;
     model infrisk = census nbeds region|medschl;
     lsmeans region|medschl;
```

```
proc reg;
     title3 'One big regression, no interactions';
     model infrisk = stay age census nbeds nurses service
                     r1-r3 mschool
                     xratio culratio ;
     size: test census=nbeds=nurses = 0;
     region:  test r1-r2=r3=0;
     monitor: test xratio=culratio=0;
```

senic3.lst

```
        Study of the Efficacy of Nosocomial Infection Control (SENIC)        1

             Med School and infection risk ignoring hospital size

                                The GLM Procedure

                            Class Level Information

                         Class         Levels    Values

                         medschl            2    No Yes


                      Number of Observations Read         113
                      Number of Observations Used         112
```

```
                                The GLM Procedure

Dependent Variable: infrisk   Prob of acquiring infection in hospital

                                     Sum of
Source                         DF        Squares    Mean Square   F Value   Pr > F

Model                           1     11.3116108     11.3116108      6.61   0.0115

Error                         110    188.2426749      1.7112970

Corrected Total               111    199.5542857


             R-Square     Coeff Var      Root MSE    infrisk Mean

             0.056684      30.12223      1.308166        4.342857


Source                         DF      Type I SS    Mean Square   F Value   Pr > F

medschl                         1    11.31161079    11.31161079      6.61   0.0115


Source                         DF    Type III SS    Mean Square   F Value   Pr > F
```

medschl	1	11.31161079	11.31161079	6.61	0.0115

Med School and infection risk ignoring hospital size

The GLM Procedure

Level of medschl	N	infrisk Mean	infrisk Std Dev
No	95	4.20842105	1.33867696
Yes	17	5.09411765	1.11213229

Med School and infection risk controlling for hospital size

The GLM Procedure

Class Level Information

Class	Levels	Values
medschl	2	No Yes

Number of Observations Read	113
Number of Observations Used	112

Med School and infection risk *controlling for* hospital size

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	30.9555343	10.3185114	6.61	0.0004
Error	108	168.5987514	1.5610995		
Corrected Total	111	199.5542857			

R-Square	Coeff Var	Root MSE	infrisk Mean
0.155123	28.76999	1.249440	4.342857

Source	DF	Type I SS	Mean Square	F Value	Pr > F
census	1	29.85331029	29.85331029	19.12	<.0001
nbeds	1	1.09693657	1.09693657	0.70	0.4037
medschl	1	0.00528747	0.00528747	0.00	0.9537

Source	DF	Type III SS	Mean Square	F Value	Pr > F
census	1	4.21035527	4.21035527	2.70	0.1034
nbeds	1	1.10213875	1.10213875	0.71	0.4026
medschl	1	0.00528747	0.00528747	0.00	0.9537

Med School and infection risk controlling for hospital size

The GLM Procedure
Least Squares Means

medschl	infrisk LSMEAN
No	4.34655162
Yes	4.32221156

Study of the Efficacy of Nosocomial Infection Control (SENIC) 7

One-factor ancova on Region with proc glm

The GLM Procedure

Class Level Information

Class	Levels	Values
region	4	North Central Northeast South West

Number of Observations Read 113
Number of Observations Used 112

Study of the Efficacy of Nosocomial Infection Control (SENIC) 8

One-factor ancova on Region with proc glm

The GLM Procedure

Dependent Variable: infrisk Prob of acquiring infection in hospital

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	43.9194874	8.7838975	5.98	<.0001
Error	106	155.6347983	1.4682528		
Corrected Total	111	199.5542857			

R-Square	Coeff Var	Root MSE	infrisk Mean
0.220088	27.90133	1.211715	4.342857

Source	DF	Type I SS	Mean Square	F Value	Pr > F
census	1	29.85331029	29.85331029	20.33	<.0001
nbeds	1	1.09693657	1.09693657	0.75	0.3893
region	3	12.96924057	4.32308019	2.94	0.0363

Source	DF	Type III SS	Mean Square	F Value	Pr > F
census	1	2.77285459	2.77285459	1.89	0.1723
nbeds	1	0.48650669	0.48650669	0.33	0.5661
region	3	12.96924057	4.32308019	2.94	0.0363

One-factor ancova on Region with proc glm

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

region	infrisk LSMEAN	LSMEAN Number
North Central	4.34865502	1
Northeast	4.75064564	2
South	3.89627373	3
West	4.62244420	4

Least Squares Means for effect region
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: infrisk

i/j	1	2	3	4
1		1.0000	0.7728	1.0000
2	1.0000		0.0432	1.0000
3	0.7728	0.0432		0.3130
4	1.0000	1.0000	0.3130	

One-factor ancova on Region with proc reg

The REG Procedure
Model: MODEL1
Dependent Variable: infrisk Prob of acquiring infection in hospital

Number of Observations Read	113
Number of Observations Used	112
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	43.91949	8.78390	5.98	<.0001
Error	106	155.63480	1.46825		
Corrected Total	111	199.55429			

Root MSE	1.21171	R-Square	0.2201
Dependent Mean	4.34286	Adj R-Sq	0.1833
Coeff Var	27.90133		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error
Intercept	Intercept	1	3.29733	0.24988
census	Aver # patients in hospital per day	1	0.00554	0.00403
nbeds	Average # beds during study period	1	-0.00184	0.00319
r1		1	0.85437	0.31176
r2		1	0.45238	0.29552
r4		1	0.72617	0.36976

Parameter Estimates

Variable	Label	DF	t Value	Pr > |t|
Intercept	Intercept	1	13.20	<.0001
census	Aver # patients in hospital per day	1	1.37	0.1723
nbeds	Average # beds during study period	1	-0.58	0.5661
r1		1	2.74	0.0072
r2		1	1.53	0.1288
r4		1	1.96	0.0522

One-factor ancova on Region with proc reg

The REG Procedure
Model: MODEL1

Test region Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	4.32308	2.94	0.0363
Denominator	106	1.46825		

Deleting page headers for pairwise comparisons ...

Test NCvsNE Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	2.29187	1.56	0.2143
Denominator	106	1.46825		

Test NCvsS Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	3.44059	2.34	0.1288
Denominator	106	1.46825		

Test NCvsW Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.77173	0.53	0.4701
Denominator	106	1.46825		

Test NEvsS Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	11.02704	7.51	0.0072
Denominator	106	1.46825		

Test NEvsW Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.15054	0.10	0.7494
Denominator	106	1.46825		

Test SvsW Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	5.66292	3.86	0.0522
Denominator	106	1.46825		

Study of the Efficacy of Nosocomial Infection Control (SENIC) 18

Two-factor ANOVA with proc glm

The GLM Procedure

Class Level Information

Class	Levels	Values
region	4	North Central Northeast South West
medschl	2	No Yes

Number of Observations Read 113
Number of Observations Used 112

Study of the Efficacy of Nosocomial Infection Control (SENIC) 19

Two-factor ANOVA with proc glm

The GLM Procedure

Dependent Variable: infrisk Prob of acquiring infection in hospital

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	29.2019453	4.1717065	2.55	0.0184
Error	104	170.3523404	1.6380033		
Corrected Total	111	199.5542857			

R-Square	Coeff Var	Root MSE	infrisk Mean
0.146336	29.47012	1.279845	4.342857

Source	DF	Type I SS	Mean Square	F Value	Pr > F
region	3	16.40096230	5.46698743	3.34	0.0222
medschl	1	7.35408469	7.35408469	4.49	0.0365
region*medschl	3	5.44689831	1.81563277	1.11	0.3492

Source	DF	Type III SS	Mean Square	F Value	Pr > F
region	3	6.18741311	2.06247104	1.26	0.2924
medschl	1	6.83414433	6.83414433	4.17	0.0436
region*medschl	3	5.44689831	1.81563277	1.11	0.3492

Two-factor ANOVA with proc glm

The GLM Procedure

Level of region	N	infrisk Mean	infrisk Std Dev
North Central	32	4.39375000	1.33921920
Northeast	28	4.87857143	1.28678674
South	36	3.86388889	1.42751588
West	16	4.38125000	0.87652248

Level of medschl	N	infrisk Mean	infrisk Std Dev
No	95	4.20842105	1.33867696
Yes	17	5.09411765	1.11213229

Level of region	Level of medschl	N	infrisk Mean	infrisk Std Dev
North Central	No	25	4.32800000	1.41554465
North Central	Yes	7	4.62857143	1.08122505
Northeast	No	22	4.65454545	1.24775122
Northeast	Yes	6	5.70000000	1.17132404
South	No	34	3.75588235	1.39440248
South	Yes	2	5.70000000	0.14142136
West	No	14	4.39285714	0.93353281
West	Yes	2	4.30000000	0.42426407

Two-factor ANCOVA with proc reg

The REG Procedure
Model: MODEL1
Dependent Variable: infrisk Prob of acquiring infection in hospital

Number of Observations Read	113
Number of Observations Used	112
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	53.01240	5.89027	4.10	0.0002
Error	102	146.54188	1.43669		
Corrected Total	111	199.55429			

Root MSE	1.19862	R-Square	0.2657
Dependent Mean	4.34286	Adj R-Sq	0.2009
Coeff Var	27.59976		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error
Intercept	Intercept	1	3.62049	0.32159
census	Aver # patients in hospital per day	1	0.00448	0.00416
nbeds	Average # beds during study period	1	-0.00044697	0.00324
reg1		1	0.41091	0.27705
reg2		1	-0.28123	0.26025
reg3		1	0.10584	0.36263
ms1		1	-0.10686	0.22186
mr1		1	0.14977	0.26880
mr2		1	-0.32513	0.25947
mr3		1	0.75013	0.36540

Parameter Estimates

Variable	Label	DF	t Value	Pr > |t|
Intercept	Intercept	1	11.26	<.0001
census	Aver # patients in hospital per day	1	1.08	0.2839
nbeds	Average # beds during study period	1	-0.14	0.8905
reg1		1	1.48	0.1411
reg2		1	-1.08	0.2824
reg3		1	0.29	0.7710
ms1		1	-0.48	0.6311
mr1		1	0.56	0.5786
mr2		1	-1.25	0.2131
mr3		1	2.05	0.0426

Two-factor ANCOVA with proc reg

The REG Procedure
Model: MODEL1

Test regtest Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	1.62429	1.13	0.3404
Denominator	102	1.43669		

Two-factor ANCOVA with proc reg

The REG Procedure
Model: MODEL1

Test mstest Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.33327	0.23	0.6311
Denominator	102	1.43669		

Two-factor ANCOVA with proc reg

The REG Procedure
Model: MODEL1

Test m_by_r Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	2.87451	2.00	0.1186
Denominator	102	1.43669		

Two-factor Analysis of Covariance with proc glm

The GLM Procedure

Class Level Information

Class	Levels	Values
region	4	North Central Northeast South West
medschl	2	No Yes

Number of Observations Read	113
Number of Observations Used	112

Two-factor Analysis of Covariance with proc glm

The GLM Procedure

Dependent Variable: infrisk Prob of acquiring infection in hospital

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	53.0124017	5.8902669	4.10	0.0002
Error	102	146.5418841	1.4366851		
Corrected Total	111	199.5542857			

R-Square	Coeff Var	Root MSE	infrisk Mean
0.265654	27.59976	1.198618	4.342857

Source	DF	Type I SS	Mean Square	F Value	Pr > F
census	1	29.85331029	29.85331029	20.78	<.0001
nbeds	1	1.09693657	1.09693657	0.76	0.3843
region	3	12.96924057	4.32308019	3.01	0.0337
medschl	1	0.46938064	0.46938064	0.33	0.5689
region*medschl	3	8.62353359	2.87451120	2.00	0.1186

Source	DF	Type III SS	Mean Square	F Value	Pr > F
census	1	1.66747644	1.66747644	1.16	0.2839
nbeds	1	0.02736713	0.02736713	0.02	0.8905
region	3	4.87288207	1.62429402	1.13	0.3404
medschl	1	0.33326705	0.33326705	0.23	0.6311
region*medschl	3	8.62353359	2.87451120	2.00	0.1186

Two-factor Analysis of Covariance with proc glm

The GLM Procedure
Least Squares Means

region	infrisk LSMEAN
North Central	4.08634201
Northeast	4.77848277
South	4.47340912
West	4.13205944

medschl	infrisk LSMEAN
No	4.47442834
Yes	4.26071833

region	medschl	infrisk LSMEAN
North Central	No	4.51832543
North Central	Yes	3.65435860
Northeast	No	4.73556984
Northeast	Yes	4.82139569
South	No	3.83013762
South	Yes	5.11668062
West	No	4.81368045
West	Yes	3.45043842

One big regression, no interactions

The REG Procedure
Model: MODEL1
Dependent Variable: infrisk Prob of acquiring infection in hospital

Number of Observations Read	113
Number of Observations Used	109
Number of Observations with Missing Values	4

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	109.88530	9.15711	11.10	<.0001
Error	96	79.22424	0.82525		
Corrected Total	108	189.10954			

Root MSE	0.90843	R-Square	0.5811
Dependent Mean	4.38073	Adj R-Sq	0.5287
Coeff Var	20.73703		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error
Intercept	Intercept	1	-0.74931	1.22190
stay	Av length of hospital stay, in days	1	0.21968	0.06989
age	Average patient age	1	0.02046	0.02212
census	Aver # patients in hospital per day	1	0.00379	0.00343
nbeds	Average # beds during study period	1	-0.00331	0.00266
nurses	Aver # nurses during study period	1	0.00153	0.00169
service	% of 35 potential facil. & services	1	0.02078	0.01012
r1		1	-0.96795	0.33329
r2		1	-0.62302	0.29915
r3		1	-0.81574	0.29081
mschool		1	-0.64256	0.32070
xratio	# x-rays / # no signs of pneumonia	1	0.00931	0.00530
culratio	# cultures / # no hosp acq infect	1	0.05623	0.01075

Parameter Estimates

Variable	Label	DF	t Value	Pr > |t|
Intercept	Intercept	1	-0.61	0.5412
stay	Av length of hospital stay, in days	1	3.14	0.0022
age	Average patient age	1	0.92	0.3575
census	Aver # patients in hospital per day	1	1.10	0.2727
nbeds	Average # beds during study period	1	-1.24	0.2168
nurses	Aver # nurses during study period	1	0.91	0.3671
service	% of 35 potential facil. & services	1	2.05	0.0428
r1		1	-2.90	0.0046
r2		1	-2.08	0.0399
r3		1	-2.81	0.0061
mschool		1	-2.00	0.0479
xratio	# x-rays / # no signs of pneumonia	1	1.76	0.0820
culratio	# cultures / # no hosp acq infect	1	5.23	<.0001

One big regression, no interactions

The REG Procedure
Model: MODEL1

Test size Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	0.76325	0.92	0.4319
Denominator	96	0.82525		

One big regression, no interactions

The REG Procedure
Model: MODEL1

Test region Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	3.86122	4.68	0.0115
Denominator	96	0.82525		

One big regression, no interactions

The REG Procedure
Model: MODEL1

Test monitor Results for Dependent Variable infrisk

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	16.10666	19.52	<.0001
Denominator	96	0.82525		