STAT 659 — Final Exam **Name:** _____
Spring 2015

INSTRUCTIONS FOR THE STUDENT:

1. You have exactly 2 hours to complete the exam.
2. There are 11 pages including this cover sheet and 3 pages of SAS output.
3. Each lettered part of a question is worth 7 points unless otherwise marked.
4. Please answer all questions.
5. Show all your work on the test booklet.
6. Do not discuss or provide any information to any one concerning any of the questions on this exam or your solutions until I post the solutions.
7. You may use a calculator that does not have the capability of phoning, texting, or accessing the internet and three $8\frac{1}{2} \times 11$ formula sheets (you may use both sides). Do not use the textbook or class notes.
8. Carry out tests at level 0.05 unless otherwise stated.
9. Be sure to clearly state the hypotheses, the test statistic and its value, and conclusion for all tests.

I attest that I spent no more than 2 hours to complete the exam. I used only the materials described above. I did not receive assistance from anyone during the taking of this exam.

Student's Signature_____

INSTRUCTIONS FOR PROCTOR:

(1) Record the time at which the student starts the exam: _____

(2) Record the time at which the student ends the exam: _____

(3) Immediately after the student completes the exam, please scan the exam to a .pdf file and have the student upload it to webassign.

(4) Collect all portions of this exam at its conclusion. Do not allow them to take any portion with them.

(5) Please keep these materials until May 18, at which time you may either dispose of them or return them to the student.

I attest that the student has followed all the INSTRUCTIONS FOR THE STUDENT listed above and that the exam was scanned into a pdf and uploaded to webassign in my presence:

Proctor's Signature_____

Some Chi-Squared Percentiles

	Right-Tail Probability			
df	0.100	0.050	0.025	0.010
1	2.71	3.84	5.02	6.63
2	4.61	5.99	7.38	9.21
3	6.25	7.81	9.35	11.34
4	7.78	9.49	11.14	13.28
5	9.24	11.07	12.83	15.09
6	10.64	12.59	14.45	16.81
7	12.02	14.07	16.01	18.48
8	13.36	15.51	17.53	20.09
9	14.68	16.92	19.02	21.67
10	15.99	18.31	20.48	23.21

Some Normal Percentiles

Right-Tail Probability			
0.100	0.050	0.025	0.010
1.282	1.645	1.960	2.326

1. In the 2008 General Social Survey, males were asked the questions "How proud are you of America's economic achievements?" and "How proud are you of America's scientific and technological achievements?" resulting in the following table:

	Science and Technology			
Economic	Very Proud	Somewhat Proud	Somewhat Proud	Not Proud At All
Very Proud	369	59	6	1
Somewhat Proud	226	238	10	3
Not Very Proud	60	67	14	2
Not Proud At All	7	16	3	2

(a) The survey responses were dichotomized by combining the *Very Proud* and *Somewhat Proud* categories into the *Proud* category and the other two categories into the *Not Proud* category. The resulting 2×2 table appears below. Test whether the proportion of males rating economics as *Proud* differs from the proportion rating science and technology as *Proud*.

	Science and Technology		
Economic	Proud	Not Proud	Total
Proud	892	20	912
Not Proud	150	21	171
Total	1042	41	1083

(b) Using the data in a), estimate the marginal odds ratio for rating economics as *Proud* relative to rating science and technology as *Proud*. Then obtain the conditional maximum likelihood estimate for the corresponding conditional odds ratio in the subject specific model.

- Marginal odds ratio

- Conditional odds ratio

(c) Several models were fit to the data in the original four-by-four table resulting in the deviances and degrees of freedom in the following table:

Model	Deviance	Degrees of Freedom
Independence	210.4	9
Symmetry	218.0	6
Quasi-independence	10.5	5
Ordinal quasi-symmetry	8.89	5
Marginal homogeneity	197.9	3
Quasi-symmetry	3.42	3

Discuss the fit of these models and select the most appropriate model(s).

(d) Carry out tests for marginal homogeity in two ways:

- Test the fit of the model labeled "Marginal homogeneity".

- Assuming the quasi-symmetry model holds, carry out a test for marginal homogeneity of the data.

2. In the homework, you studied a data set from the MBTI Step II National Sample (collected and compiled by CPP, Inc.) on whether a sample of individuals report drinking alcohol frequently ($Y = 1$ if yes, $= 0$ if no) and on the four binary scales of the MyersBriggs personality test: Extroversion/Introversion (E/I), Sensing/iNtuitive (S/N), Thinking/Feeling (T/F) and Judging/Perceiving (J/P). The 16 predictor combinations correspond to the 16 personality types: ESTJ, ESTP, ESFJ, ESFP, ENTJ, ENTP, ENFJ, ENFP, ISTJ, ISTP, ISFJ, ISFP, INTJ, INTP, INFJ, INFP. You fit logistic regression models using the four scales as predictors of π = the probability of drinking alcohol frequently.

   (a) Various logistic regresion models including interactions were fit to the data. Use the following table to choose the most appropriate models among those included in the table.

Model	G^2	AIC	df
Main effects only	11.15	637.5	11
Main effect and all 2-factor interactions	3.74	642.1	5
Main effects and all 2- and 3-factor interactions	0.14	646.5	1
Saturated model	0.00	648.3	0

   (b) Starting with the main effects model, six models were fit where each 2-way interaction term was included one at a time with the main effects. Since the `EI*JP` and `SN*TF` interactions were the only ones that had $p-$values smaller than 0.15 in their models, a new model with all the main effects and the two interactions `EI*JP, SN*TF` were fit to the data. The model had a deviance of 5.23 with 9 degrees of freedom. Carry out likelihood ratio tests to determine which of the three models you prefer among (i) this model, (ii) the model with all the main effects, and (iii) the model with all the main effects and two-way interactions.

(c) For this problem, we use only the first letter to identify the predictor (i.e., `E=EI,J=JP,S=SN,T=TF`). Construct the equivalent loglinear model for (i) the logit model with main effects only and (ii) the logit model studied in part (b) with predictors `E,J,S,T,E*J,S*T`.

(d) Using the SAS output for the logit model with main effects only, estimate the odds ratio for being alcoholic for a person of personality type ESTJ (i.e., `EI=E,SN=S,` etc.) relative to a person with personality type INTJ.

(e) Using the SAS output for the logit model with main effects only, determine which personality type has the smallest $\hat{\pi}$ and explain why this is the case based on the model parameter estimates.

(f) We now look at the relationship between the TF scale and Y in two ways.

i. Use this table below to estimate the marginal odds ratio between the thinking and feeling personality types and frequent alcohol drinking.

	Drink Alcohol Frequently		
Personality	Yes	No	Total
Thinking	49	351	400
Feeling	48	602	650
Total	97	953	1050

ii. Use the logit model with main effects only to estimate the conditional odds ratio between the thinking and feeling personality types and frequent alcohol drinking keeping all the other three scales constant.

3. A study was carried out to investigate the relationship between regular smoking and major depressive disorder. A sample of 3213 individuals were classified according to regular smoking habit (`smoke`), major depressive disorder (`depress`), and gender (`gender`), a variable also thought to be related to major depressive disorder. Various loglinear models identified by the first letter of the explanatory variable were fit to the data. Use the accompanying SAS output to help you answer this problem. Use the following table to help you answer this problem.

Model	G^2	df
(D,G,S)	159.2	4
(G,DS)	136.4	3
(D,GS)	57.1	3
(S,DG)	135.1	3
(DG,GS)	33.0	2
(DS,GS)	34.4	2
(DG,DS)	112.3	2
(DG,DS,GS)	0.77	1
(DGS)	0.0	0

(a) Carry out a test of equal odds ratios for between `smoke` and `depress` for males and females.

(b) Assuming that the homogeneous association model holds, carry out a test of partial association of `smoke` and `depress`, controlling for `gender`.

Model	G^2	df
(D,G,S)	159.2	4
(G,DS)	136.4	3
(D,GS)	57.1	3
(S,DG)	135.1	3
(DG,GS)	33.0	2
(DS,GS)	34.4	2
(DG,DS)	112.3	2
(DG,DS,GS)	0.77	1
(DGS)	0.0	0

(c) Based on the table of deviances, which model would you recommend? Be sure to justify your answer.

(d) Estimate the odds ratio between `gender` and `depress` for a regular smoker (`smoke=yes`) using: the homogeneous association model and the saturated model. SAS output for these models is provided.

- Homogeneous association model

- Saturated model