STAT 659 — Exam 1 **Name:** _____
Summer 2014

INSTRUCTIONS FOR THE STUDENT:

1. You have exactly 70 minutes to complete the exam.
2. There are 11 pages including this cover sheet and 4 pages of SAS output.
3. Each lettered part of a question is worth 8 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 one $8\frac{1}{2} \times 11$ formula sheet (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 70 minutes 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 June 30, 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. A study was carried out to investigate the relationship between passive smoking and lung cancer. The study defined *passive smoking* as exposure to the cigarette smoke of a spouse who smoked at least one cigarette a day for at least 6 months. The study included 235 nonsmoking subjects with cancer and 235 nonsmoking control subjects matched to the case patients on the basis of sex and age. The passive-smoking status of each subject was ascertained resulting in the following table:

	Disease Status	
Passive-smoking status	Cancer Patients	Controls
Yes	120	80
No	115	155

   (a) Can you use these data to compare the proportions of subjects who suffer from cancer for those exposed to passive smoking with those not exposed to passive smoking using the relative risk? Explain why or why not.

   (b) Construct a 95% confidence interval for the odds ratio of suffering from cancer for subjects exposed to passing smoking relative to subjects not exposed to passive smoking. Interpet this estimate. Under what circumstances is this a good approximation to the relative risk?

2. Researchers are interested in the relationship between gender (`gender`= `male or female`) and depression (`depress`=`yes or no`). Since level of education is thought to be related to depression, the level of education in years of schooling (`educ`=`low` for less than 12 years or =`high` for 12 years or more) was recorded for each subject in the study. A random sample of subjects from a population known to be prone to depression was taken. Two-way tables for `gender` and `depress` were formed for subjects falling into the two education categories. Use the accompanying SAS output to help you answer this problem.

   (a) Is Simpson's paradox present for these data? Explain why or why not.

   (b) Determine the values of the test statistics, the $P-$values, and the conclusions for (i) the test of equal odds ratios for between `gender` and `depress` for the two levels of education and (ii) the test of partial association of `gender` and `depress`, controlling for level of education.

	(i) the test of equal odds ratios	(ii) the test of partial association
Value of test statistic		
$P-$value		
Conclusion		

(c) Report a 95% confidence interval for a common odds ratio between `gender` and `depress`. Comment on whether it is appropriate to use this interval to summarize these tables.

3. A study is carried out on an experimental treatment with respect to its effectiveness in reducing joint pain in patients with arthritis. The current use of nonsteroidal anti-inflamatory drugs (NSAIDs) has been found to be effective in 60% of such patients. A trial of 12 patients was carried out and 10 of them found that the experimental treatment reduced joint pain. Carry out a test to determine if the new treatment is more effective than NSAIDs in reducing joint pain. State the hypotheses, report the $P-$value and mid $P-$value and interpret. You might find the following table of the binomial distribution with $n = 12$ and $\pi = 0.6$ helpful.

y	0	1	2	3	4	5	6	7	8	9	10	11	12
P(y)	0.000	0.000	0.003	0.013	0.042	0.101	0.177	0.227	0.213	0.142	0.064	0.017	0.002

4. The German Social Survey provides data on the number of children (`child`) for women in addition to their age (`age`) in years, nationality (`nation`= 1, if German, = 0, otherwise), duration of school education (`dur` ranging from 5 to 23 years), university degree (`univ`= 1 if yes, = 0 if no), and religion (`relig`= $1, \ldots, 6$ representing answers to the question "God is the most important in man." with 1 = strongly agree and 6 = strongly disagree). Loglinear Poisson regression models were fit to these data. Use the accompanying SAS output to help you answer the following questions.

   (a) The counts of the number of children for each family without taking into account possible predictors were examined to determine whether the Poisson distribution would fit the data. Using the estimated parameter $\hat{\lambda} = 1.57296$, the following observed frequencies and Poisson probabilities were found:

Number of children	0	1	2	3	4	≥ 5	Total
Count	398	455	575	227	65	41	1761
Poisson probability	0.2074	0.3263	0.2566	0.1345	0.0529	0.0222	1

   First, compute the contribution of the first cell (zero children) to the chi-squared statistic for testing the goodness of fit of the Poisson distribution. Then use the information that the value of the chi-squared statistic equals $X^2 = 70.3774$ to carry out a test for goodness of fit by the Poisson distribution for these data.

   (b) A model (Model A) with the predictors `age,nation,dur,univ,relig` was fit to the data. The researchers felt that the effects of age and duration of education might be nonlinear, so they fit a second model (Model B) including the predictors in the first model as well as three additional polynomial terms in age (`age2,age3,age4`) and a quadratic term in duration (`dur2`). Carry out a likelihood ratio test to determine whether the additional polynomial terms significantly improve the fit of the model.

(c) Use the Poisson loglinear model (Model A) without the additional polynomial terms in `age` and `dur` to answer this part of the problem. What is the effect of having a university degree on the mean number of children for a women, keeping all other variables constant?

(d) Several Poisson regression models with log link were fit to the data. At each step, the least significant predictor was eliminated from the model, and a model with the remaining predictors was fit to the data. Based on information in the table below, select a reasonable Poisson regression model. Explain your reasoning.

Model	Predictors	Deviance	DF	AIC_C
1	`age,age2,age3,age4,dur,dur2,nation,relig,univ`	1718.6	1747	5197.0
2	`age,age2,age3,age4,dur,dur2,relig,univ`	1718.8	1748	5195.2
3	`age,age2,age3,age4,dur,dur2,univ`	1725.6	1753	5191.8
4	`age,age2,age3,age4,dur,univ`	1733.8	1754	5198.0
5	`age,age2,age3,age4,dur`	1740.0	1755	5202.2
6	`age,age3,age4,dur`	1753.7	1756	5213.8

5. Researchers used data from the 2006 General Social Survey to examine the relationship of happiness and income. Respondents were asked, "Taken all together, would you say that you are very happy, pretty happy, or not so happy?" The table cross-classifies this response with family income, here measured as the response to the question, "Compared with American families in general, would you say that your income is below average, average, or above average?" Use the accompanying SAS output to help you answer this question.

   (a) Assume that the respondents to the 2006 General Social Survey represent a random sample from the American population. Construct a 95% confidence interval for the proportion of all Americans that consider themselves *very happy*.

   (b) Determine the values of the test statistics, the $P-$values, and the conclusions for the tests of independence of income and happiness versus (i) a general alternative and (ii) an alternative that takes into account the ordering of the categories.

	(i) general alternative	(ii) ordered alternative
Value of test statistic		
$P-$value		
Conclusion		