STAT 659 — Exam 2 **Name:** _____
Spring 2015

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 two $8\frac{1}{2} \times 11$ 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 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 April 20, 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 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. Logit models relating `depress` to `smoke` and `gender` were fit to the data. Use the accompanying SAS output to help you answer this problem.

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

   (b) Carry out a test of partial association of `smoke` and `depress`, controlling for gender.

(c) Estimate the odds ratio between `gender` and `depress` for a regular smoker (`smoke=yes`) using: (i) the homogeneous association model and (ii) the model with interaction.

2. Exposure to asbestos dust is known to increase the risk of several serious diseases. Asbestos exposure was measured on 83 Navy workers and classified into three ordered levels: low exposure, action level, and above the legal limit. It was expected that exposure would be different for two tasks: removing insulation and removing tiles. Exposure was also thought to depend on the nature of the ventilation: ordinary ventilation using fans (labelled "general") or negative pressure using a pump and air filter. A proportional odds model and a baseline-category model with low exposure being the baseline were fit to the data. Using the accompanying SAS output to help you answer this question.

   (a) Is the assumption of a proportional odds model reasonable? Explain. Be sure to take into account the nature of the response as well as any relevant tests.

(b) For the baseline category model, estimate the probability that a worker has low exposure (i.e., $P(Y = 1)$) when he/she removes tile with ordinary ventilation.

(c) For the proportional odds model, estimate the probability that a worker has low exposure (i.e., $P(Y = 1)$) when he/she removes tile with ordinary ventilation. Is this value similar to that in part (b)?

(d) According to the proportional odds model, which combination of task and ventilation leads to the highest exposure and which combination leads to the lowest exposure? Explain.

3. Researchers studied the incidence of lower respiratory infections in 284 children over a year. Explanatory variables included passive smoking (`passive` = 1 if yes), socioeconomic status (`ses`, 3 categories), crowding (`crowding` = 1 if yes), race (`race` = 1 if black, = 0 if white), exposure time (`risk`), and age (`agegroup`, three categories). In the first test, you investigated a Poisson regression model with response variable, the number times that the child had a lower respiratory infection during the year. Now we consider various logistic regression models with response variable `lri`= 1 if the child had one or more lower respiratory infections during the year and = 0, otherwise.

   (a) A logit model (Model A) with the predictors `passive, crowding, ses, race, agegroup, risk` was fit to the data. The researchers felt that the effects of `race` and `ses` might not be useful in the model, so they fit a second logit model (Model B) omitting these two variables from the model. Carry out a likelihood ratio test to determine whether it is appropriate to simultaneously omit the two effects from Model A.

(b) Use the logistic regression model (Model A) with all the predictors to answer this part of the problem. What is the estimated effect of simultaneously being in a crowded setting (`crowding` $= 1$) and being exposed to passive smoke (`passive` $= 1$) on the odds of having at least one lower respiratory infection relative to being in a noncrowded setting (`crowding` $= 0$) and not being exposed to passive smoke (`passive` $= 0$), keeping all other variables constant?

(c) A classification table was formed using Model A and the cutoff $\pi_0 = 0.407$, where the proportion of children with one or more lower respiratory infections during the year was $0.407 = 114/284$. Use this table to estimate the sensitivity, specificity, and proportion of correct classifications based on this cutoff.

	Prediction, $\pi_0 = 0.407$		
Actual	$\hat{y} = 1$	$\hat{y} = 0$	Total
$y = 1$	62	52	114
$y = 0$	66	104	170

(d) Several logistic regression models 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 logistic regression model. Explain your reasoning.

Model	Predictors	Deviance	DF	AIC_C	Dev2	DF2
1	`passive,crowding,ses,race,agegroup,risk`	353.6	275	371.6	255.0	182
2	`passive,crowding,ses,agegroup,risk`	353.6	276	369.6	255.0	183
3	`passive,crowding,agegroup,risk`	354.4	278	366.4	255.7	185
4	`crowding,agegroup,risk`	355.5	279	365.9	257.3	186
5	`crowding,risk`	366.8	281	372.8	268.1	188
6	`crowding`	369.2	282	373.2	270.5	189

(e) Based on the output of Model A, is there any evidence of lack of fit or inadequacy of the logistic regression Model A? Give a complete explanation.