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

INSTRUCTIONS FOR THE STUDENT:

1. You have exactly 70 minutes to complete the exam.
2. There are 12 pages including this cover sheet and 5 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 August 1, 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 of the labor force participation by women in the United States was carried out using data from 1976. The sample consisted of 753 white, married women between the ages of 30 and 60 years old. The response variable is `y` = 1 (`inLF`) if the wife was in the labor force and = 0 (`NOtinLF`) if the wife was not in the labor force. The predictors that were measured were `k5` = the number of children 5 years old or under, `k618` = the number of children from 6 to 18 years old, `age` of the wife in years, `wc` = 1 if the wife attended college and = 0 otherwise, `hc` = 1 if the husband attended college and = 0 otherwise, `lwg` = the logarithm of the wife's estimated wages, and `inc` = the family's estimated income excluding the wife's income. A logistic regression model was fit to these data with `y` as the response and predictors `k5, k618, age, wc,hc, lwg, inc`. Use the accompanying SAS output to help you answer the parts to this problem.

   (a) Consider the model above. Is there any indication that we could eliminate any of the predictors? Explain your reasoning.

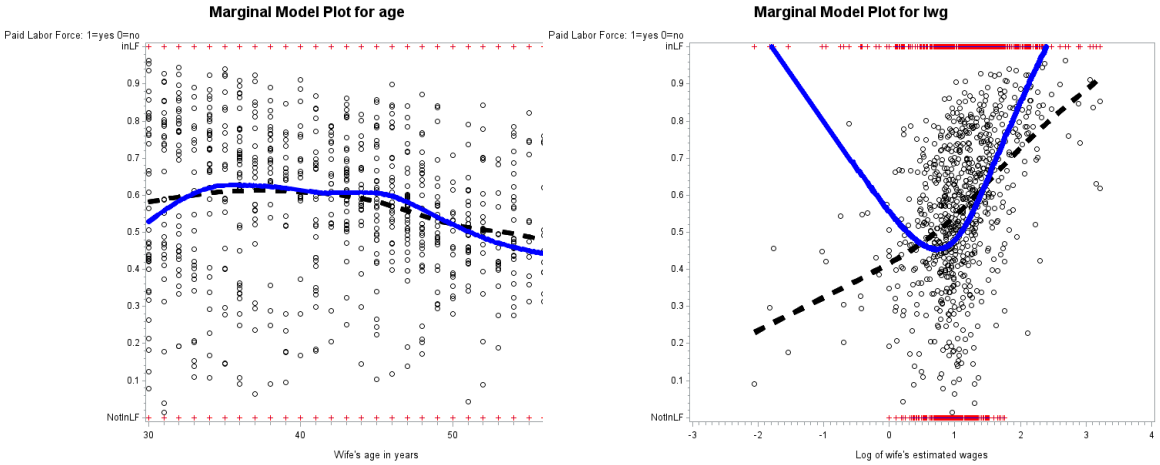
   (b) Based on the attached SAS output for the model above, is there any indication of lack of fit for the model. Explain your answer.

(c) What is the effect of the wife having attended college on the estimated odds of the wife being in the labor force, keeping all other variables constant?

(d) A classification table was formed using the cutoff $\pi_0 = 0.568$, where the observed proportion in this sample of married women who were in the labor force during the year was $0.568 = 428/753$. Use this table to estimate the sensitivity, specificity, and proportion of correct classifications based on this cutoff.

	Prediction, $\pi_0 = 0.568$		
Actual	$\hat{y} = 1$	$\hat{y} = 0$	Total
$y = 1$	290	138	428
$y = 0$	105	220	325

(e) The researchers wished to check whether the effects of `age` and `lwg` might be nonlinear. Marginal model plots for the variable `age` and for `lwg` were made for the model that includes all the predictors. The solid line corresponds to the blue line in lecture where the responses were smoothed. The dashed line corresponds to the black line in lecture where the predicted values were smoothed. Interpret these plots.



2. Long (1990) investigated models to relate the number of publications (`art`) produced by Ph. D. biochemists as a function of gender (`fem` = 1 if female), marital status (`mar` = 1 if married), number of children (`kid5` = 0, 1, 2, or 3), a numerical rating of the prestige of the institution where the biochemist obtained the Ph. D. (`phd`) ranging from 0.75 to 4.62, and the number of articles written by the biochemist's mentor for the Ph. D. in the last 3 years (`ment`) ranging from 0 to 77. In the first test, we examined using Poisson loglinear models to analyze these data. Here we will define a new variable `pubs` which equals 0 when `art`= 0, equals 1 when `art`= 1, and equals 2 when `art`$\geq$ 2. A proportional odds cumulative logit model and a baseline-category logit model with `pubs`= 0 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 each of the two models, carry out the test of whether the variable `mar` is useful in the model, given that the other variables are present.

- Proportional odds model

- Baseline category model

(c) Using the proportional odds model, report the setting of `fem,mar,kid5,phd,ment` at which the Ph. D. biochemist is most likely to have zero publications. Explain your answer.

(d) A baseline-category logit model with `pubs` = 0 being the baseline category relating `pubs` to x = `ment` was fit to the data. The model resulted in the two estimated logits, $\log(\hat{\pi}_2/\hat{\pi}_0) = -0.351 + 0.092x$ and $\log(\hat{\pi}_1/\hat{\pi}_0) = -0.531 + 0.062x$. Derive the prediction equation for $\log(\hat{\pi}_2/\hat{\pi}_1)$ and determine the range of x for which $\hat{\pi}_1 > \hat{\pi}_2$.

3. A random sample of 3688 applicants for vocational education programs in major northeastern school districts was obtained. Each student was classified according to program applied for (`program`), gender (`gender`), and whether or not the student was accepted (`accept`). Logit models relating the response `accept` to the predictors `program` 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 between `gender` and `accept` for the three programs.

(b) Assuming that the homogeneous association model is appropriate, carry out a test of partial association of `gender` and `accept`, controlling for `program`.

(c) Estimate the odds ratio for being accepted for a female applying to the plumbing program relative to a male applying to the cosmetology program using: (i) the homogeneous association model and (ii) the model with interaction.

i. Homogeneous association model

ii. Model with interaction