STAT 659 — Exam 1 **Name:** _____
Summer 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 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. Dietary fiber is sometimes added to heavily processed foods to make them more nutritious. Twelve individuals were given a new type of fiber-enriched crackers. They were instructed to eat the crackers and then a meal. Shortly afterward they were asked if they experienced any bloating with 10 individuals reporting that they experienced some bloating and 2 individuals experiencing no bloating.

   (a) The researchers wished to determine if there was evidence that more than 50% of individuals given the fiber-enriched crackers experience any bloating. 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 $\pi = 0.5$ helpful.

y	0	1	2	3	4	5	6	7	8	9	10	11	12
$P(y)$	0.000	0.003	0.016	0.054	0.121	0.193	0.226	0.193	0.121	0.054	0.016	0.003	0.000

   (b) Taking into account the sample size, construct a 95% confidence interval for the true proportion of individuals given the fiber-enriched cracker that experience any bloating.

The situation described on the previous page was a part of a larger experiment in which individuals were given either a new type of fiber-enriched cracker or a non-enriched cracker (assume they could not tell which type they ate). They were instructed to eat the crackers and then a meal. Shortly afterward they were asked if they experienced any bloating with the following results:

	Bloating		
Fiber Added	Yes	No	Total
Yes	10	2	12
No	6	6	12
Total	16	8	24

(c) Among 2×2 tables with the same marginal totals, construct all tables that provide stronger evidence of a positive association between fiber added to crackers and experiencing some bloating.

(d) Based on the following SAS output for the analysis of this problem, state the appropriate hypotheses, report a $P-$value and a mid $P-$value. Then provide a conclusion concerning whether there is a positive association between fiber added to crackers and experiencing some bloating at level $\alpha = 0.05$.

```
 Statistics for Table of fiber by bloat

Statistic                     DF       Value      Prob
Chi-Square                     1      3.0000    0.0833
Likelihood Ratio Chi-Square    1      3.1037    0.0781
Continuity Adj. Chi-Square     1      1.6875    0.1939
Mantel-Haenszel Chi-Square     1      2.8750    0.0900

 WARNING: 50% of the cells have expected counts less
          than 5. Chi-Square may not be a valid test.

       Fisher's Exact Test

Cell (1,1) Frequency (F)        10
Left-sided Pr <= F          0.9864
Right-sided Pr >= F         0.0965

Table Probability (P)       0.0829
Two-sided Pr <= P           0.1930
```

2. 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`). 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 between `gender` and `accept` for the three programs and (ii) the test of partial association of `gender` and `accept`, controlling for `program`,

	(i)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 (i) a common odds ratio between `gender` and `accept` and for (ii) the marginal odds ratio between `gender` and `accept`. Then comment on whether it is appropriate to use either of these intervals to summarize these tables.

i. Common odds ratio

ii. Marginal odds ratio

iii. Comment on appropriateness

3. 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`), and the number of articles written by the biochemist's mentor for the Ph. D. in the last 3 years (`ment`).

(a) The counts of the number of articles without taking into account possible predictors were examined to determine whether the Poisson distribution would fit the data. Using the estimated parameter $\hat{\lambda} = 1.6929$, the following observed frequencies and Poisson probabilities were found:

Number of articles	0	1	2	3	4	5	≥ 6	Total
Count	275	246	178	84	67	27	38	915
Poisson probability	0.1840	0.3115	0.2636	0.1488	0.0630	0.0213	0.0078	1

First, compute the contribution of the first cell (zero articles) 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 $X^2 = 246.14$ to carry out a test for goodness of fit by the Poisson distribution for these data.

Various Poisson loglinear regression models and one negative binomial regression model with log link were fit to the data. Use the accompanying SAS output to answer the remaining parts of this question.

(b) A Poisson loglinear regression model (Model A) with response `art` and all the predictors `fem,mar, kid5,phd,ment` was fit to the data. The researchers felt that the effects of `phd` and `mar` might not be useful in the model, so they fit a second Poisson loglinear 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.

(c) Use the Poisson loglinear model (Model A) with five predictors to answer this part of the problem. What is the estimated effect of being a unmarried female (`mar` = 0 and `fem` = 1) on the mean number of publications relative to being a married male (`mar` = 1 and `fem` = 0), other variables being held 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	`fem,mar,kid5,phd,ment`	1634.4	909	3314.1
2	`fem,mar,kid5,ment`	1634.6	910	3312.3
3	`fem,kid5,ment`	1640.9	911	3316.6
4	`fem,ment`	1657.0	912	3330.8
5	`ment`	1669.5	913	3341.3

(e) A negative binomial model with log link (Model C) was also fit using the same predictors as Model A. Based on the output of this model and also that of Model A, is there any evidence of lack of fit or inadequacy of the Poisson regression Model A? Give a complete explanation.