

INSTRUCTIONS FOR THE STUDENT:

1. You have exactly 2 hours to complete the exam.
2. There are 10 pages including this cover sheet and 2 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 August 13, 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

df	Right-Tail Probability			
	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 clinical psychiatric study involved a sample of 200 subjects. The patients were asked to rate themselves according to their perceived severity of depression. The categories were **None**, **Minor**, **Major**. Each patient was also evaluated by a psychiatrist who also rated the patient according to the severity of depression. The following table was obtained:

Patient Rating	Psychiatrist Rating		
	None	Minor	Major
None	66	13	6
Minor	36	16	10
Major	14	12	27

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

Model	Deviance	Degrees of Freedom
Independence	46.13	4
Symmetry	14.71	3
Ordinal quasi-symmetry	3.20	2
Quasi-symmetry	0.24	1

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

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

2. In the second test, we considered data from the German Social Survey 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, ..., 6 representing answers to the question “God is the most important in man.” with 1 = strongly agree and 6 = strongly disagree). The response variable was defined to be $y = 1$ if **child** ≥ 1 and $= 0$ if **child** = 0. For this problem, we will replace the numerical variable **age** with a nominal variable **agecat** with 4 levels and consider **relig** to be a nominal variable.
- (a) Consider logistic models with response variable y (Y). Write down the equivalent loglinear models for the logistic models with the specified predictors.
- Predictors **univ** (U), **nationality** (N), **religion** (R), and **agecat** (A) with no interactions.
 - Predictors **univ** (U), **nationality** (N), **religion** (R), and **agecat** (A) and the interaction between **univ** and **agecat**.
- (b) Consider the model (YU, YA, UA, AN, UR, RN) . Draw the association diagram. Then discuss any conditional independence between Y and other variables.

3. A number of experiments in which animals were exposed to cigarette smoke that were carried out before there was clear evidence that tobacco smoke is carcinogenic. In one such experiment, 23 mice were exposed each day in a chamber filled with the smoke of one cigarette every hour for 12 hours. A comparable number of mice were kept in a similar chamber without smoke. After one year, an autopsy was carried out on each mouse that had survived for at least 2 months in the experiment. The data for the 23 treated mice and 32 control mice are in the following table.

Group	Classified Group		Total
	Tumor Present	Tumor Absent	
Treated	21	2	23
Control	19	13	32
Total	40	15	55

- (a) Given 2×2 tables with the same marginal totals, construct all tables that provide stronger evidence that tumor presence is positively associated with cigarette smoke.

- (b) Based on the attached SAS output for the analysis of the data in this problem, state the appropriate hypotheses, report a P -value and a mid P -value using exact methods for testing whether a greater proportion of mice had tumors after exposure to smoke. Then provide a conclusion concerning the proportion of mice with tumors in the two groups. $\alpha = 0.05$.

4. 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. Various loglinear models (identified by the first letter of the variables) were fit to these data. Use the following table to help you answer this problem.

Model	G^2	df
(D,G,E)	44.79	4
(D,GE)	29.68	3
(DE,G)	37.93	3
(DG,E)	20.50	3
(DG,GE)	5.39	2
(DE,GE)	22.82	2
(DG,DE)	13.64	2
(DG,DE,GE)	1.48	1
(DGE)	0.0	0

- (a) Carry out a likelihood ratio test for equality of odds ratios between **gender** and **depress** for the two levels of education.

- (b) Assuming that the homogeneous association model holds, carry out a likelihood ratio test for association of **gender** and **depress**, controlling for level of education.

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

Model	G^2	df
(D,G,E)	44.79	4
(D,GE)	29.68	3
(DE,G)	37.93	3
(DG,E)	20.50	3
(DG,GE)	5.39	2
(DE,GE)	22.82	2
(DG,DE)	13.64	2
(DG,DE,GE)	1.48	1
(DGE)	0.0	0

5. A medical researcher investigated allergic reactions to two types of penicillin, Type BT and Type G. The researcher gave skin tests for both types of penicillin to each of 500 subjects. Of these subject, 52 reacted to type G and 68 reacted to type BT.

- (a) The data could be recorded in either of two ways as exemplified in the following two tables:

Table I

Type	Type G		
BT	Yes	No	Total
Yes	40	28	68
No	12	420	432
Total	52	448	500

Table II

	React		
	Yes	No	Total
Type BT	68	432	500
Type G	52	448	500
Total	120	880	1000

The researcher wished to test for a difference in the proportions of individuals who had allergic reactions to the two types of penicillin. Which table is the more appropriate for displaying the data and why? Then carry out a test for the difference in proportions who had allergic reactions for the two types of penicillin.

Table I			
Type	Type G		Total
BT	Yes	No	
Yes	40	28	68
No	12	420	432
Total	52	448	500

Table II			
	React		Total
	Yes	No	
Type BT	68	432	500
Type G	52	448	500
Total	120	880	1000

- (b) Consider now the data as entered in Table I. Assume the subject specific model for these data. Construct a 95% confidence interval for the conditional odds ratio for having an allergic reaction for Type BT relative to Type G. Then interpret this interval.

6. In a clinical study on the healing of sports-related injuries of the knee, $n = 127$ patients were treated. The patients were randomly assigned to either a treatment group ($\text{Tr}=1$) where an anti-inflammatory spray was used or a placebo group ($\text{Tr}=0$) where a spray without active ingredients was used. The mobility of the knee was investigated in which the patient actively moved the knee. The pain y occurring during the movement was assessed on a five-point scale ranging from 1 for no pain to 5 for severe pain. The covariate age was also measured. To fit the model, age was centered using $\text{Age1} = \text{age} - \overline{\text{age}}$. It was also determined that a quadratic term in age ($\text{Age2}=(\text{age}-\overline{\text{age}})^2$) should be included in the model. A multinomial cumulative logit model with response variable y was fit to the data. Use the accompanying SAS output to help you answer this problem.
- (a) For a fixed age, obtain a 95% confidence interval for the odds ratio comparing the cumulative logit of the treatment group to that of the placebo group. The use this confidence interval for the odds ratio to explain whether or not the treatment was effective.

- (b) For an individual in the treatment group of average age ($\text{age} = \overline{\text{age}}$), estimate the probability of severe pain ($y = 5$).

- (c) For the data in this problem, 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.