Final Exam

1. I generated data from the GSS (General Social Survey) website for the years 1972-2014 tabulating whether or not the respondent would vote for a female for president (1=Yes, 0=No), the respondent's religious affiliation (1=Fundamentalist, 2=Evangelical, 3=Mainline, 4=Liberal, 5=None, 6=Other) and terminal degree (0=Less than High School, 1=High School, 2=Junior College, 3=Bachelors, 4=Grad School). The data set is stored in sheet Q1 of Excel workbook Final2019.xlsx with variable names President, Religion, Degree. It is available on the website. Before proceeding, drop Religion=6, and recode Degree=4 and Degree=2 as Degree=3 (these categories have consistently low counts otherwise).

   (a) Using backward elimination, select an appropriate loglinear model for the data set.

   (b) Based on your model, write an expression for the log odds ratio of favoring a female president for each religious denomination given level of education. Does it depend on level of education? Now interpret this odds ratio for Evangelicals versus No Religion. What is the odds ratio for Mainline versus Fundamentalists?

   (c) Assume that President is the response and Religion and Degree are explanatory variables. How would that affect your model selection in (a)? Make any appropriate adjustments.

   (d) Based on your model in (c), fit the compatible logistic regression model and confirm that comparable terms in the two models have identical Type III tests and parameter estimates.

2. The following table contains counts of the number of offspring born in successive years to ewes.

	Number of Lambs			
Number of Lambs	Year 1			
Year 2	0	1	2	Row Total
0	57	50	2	109
1	28	55	5	88
2	8	14	8	30
Column Total	93	119	15	227

   (a) Test whether a model of marginal homogeneity fits the data.

   (b) Test a model of symmetry. If the model does not fit, discuss the lack of symmetry.

   (c) Test a quasi-symmetry model and interpret.

   (d) Compute Cohen's kappa for the model and interpret.

3. At a large low-level radioactive waste site, pairs of downgradient (case) and upgradient (control) groundwater monitoring wells. For each well, the presence ($Y_{ij} = 1$) or absence ($Y_{ij} = 0$) of gross alpha particle activity was recorded (Note: background radioactivity is often present in otherwise neutral soils and groundwater).

	Downgradient	
Upgradient	Present	Absent
Present	80	5
Absent	35	160

(a) Fit a conditional model to the data in PROC LOGISTIC. Interpret the odds ratio for the well effect. Is it significant?

4. Feral hog disturbance (D=Disturbance, N=No Disturbance) was measured bimonthly for subplots in three different habitats (CTS=Cypress/Tupelo Slough, MS=Muck Swamp, BHF=Bottomland Hardwood Forest) in a floodplain. Some of the subplots were sited in portions of the floodplain with active feral hog control, while others were in areas without control. The table appears below.

		Response pattern			
Habitat	Hog Treatment	DD	ND	DN	NN
CTS	None	30	15	20	85
	Active	25	10	30	90
MS	None	40	20	15	70
	Active	30	15	40	60
BHF	None	20	20	25	100
	Active	10	10	40	90

(a) Using either PROC GENMOD or PROC GLIMMIX, fit a marginal model including effects Habitat, Month, Treatment, and a Treatment by Month interaction.

(b) Assess the impact of the feral hog control program in the first bimonthly measurement period (fix the habitat). Repeat for the second bimonthly measurement period.

(c) Discuss habitat effects.

5. (a) For the same data set, construct a random effects model in either PROC GLIMMIX or PROC NLMIXED.

(b) Compare parameter estimates for the marginal effects model in (4) to the conditional model in (5).