Categorical Data Analysis: HW 5

1. The following table reports the number of beetles killed after 5 hours of exposure to gaseous carbon disulfide at a fixed set of concentration dosages.

dose	number exposed	number survived
1.691	59	53
1.724	60	47
1.755	62	44
1.784	56	28
1.811	63	11
1.837	59	6
1.861	62	1
1.884	60	0

Fit a logistic regression model, modeling the log-odds of survival in terms of dosage of the toxin to the data above. Note that here we have grouped observations, i.e., at the first dose, there are 53 successes and 6 failures. When you have data like in the table above, a convenient way to fit a logistic regression model in R is to use the following syntax:

```
fit.logit <- glm(cbind(y,n-y) ~ dose, family=binomial(link=logit), data=beetles)
```

where y is the number of beetles that survived and n is the number of beetles that were exposed to the particular toxin (and beetle is the dataset that contains the data above).

(a) Interpret the estimate for the dose effect.

(b) Construct and interpret the result of a likelihood ratio test of the hypothesis that survival is independent of the given dose.

(c) Estimate the probability of survival at a dose of 1.8.

(d) Provide and interpret a confidence interval for the probability of survival at a dose of 1.8.

2. Refer to the following Table on belief in life after death.

		Belief in Afterlife		
Race	Gender	Yes	Maybe	No
White	Female	371	49	74
	Male	250	45	71
Black	Female	64	9	15
	Male	25	5	13

(a) Fit a baseline category model and obtain the fitted model for $\log[P(Y = "Yes")/P(Y = "No")]$.

(b) Assess goodness of fit.

(c) Interpret the gender effect.

(d) Test whether belief is independent of race, given gender.

(e) Treating belief as ordinal, fit, test and interpret the gender effect in a proportional odds model.

3. Refer to the horseshoe crab data on the course webpage.

   (a) Fit a Poisson loglinear model using both weight and color (treated as nominal) to predict the *number* of satellites. Create a plot of the fitted model (similar to the plot we had for the probability of a satellite in terms of weight and color).

   (b) Test whether color is needed in the model.

   (c) Estimate and give a confidence interval for the expected number of satellites for a crab that weighs 2.4 kg and is dark.

   (d) The estimated color effects are monotone across the four color categories. Fit a model that has a linear color effect and interpret the coefficient for color if it is significant.

4. The Table below shows results when subjects were asked two questions: "Do you think a person has the right to end his or her own life if this person has an incurable disease?" and "When a person has a disease that cannot be cured, do you think doctors should be allowed to end the patient's life?" The table refers to these variables as "Suicide" and "Let Patient Die".

	Let Patient Die	
Suicide	Yes	No
Yes	1097	90
No	203	435

   (a) Perform an asymptotic test to see if the probability of the opinion on suicide and the probability of the opinion on letting a patient die are identical?

   (b) Although the sample size seems large, check the results from above using an exact test.

   (c) Construct a confidence interval for the difference between these two probabilities. (Hint: We didn't do this specifically in class, but I gave you the formula for the standard deviation of the difference, so you should be able to form a Wald confidence interval. Note: Don't use the standard deviation under the null of marginal homogeneity, which we used to derive the score statistic (=McNemar's test statistic). This standard deviation only applies if there is marginal homogeneity!)