## Stat 545 Exam

## November 2, 2016

## Directions:

- The exam is closed book. You may not use a calculator. The exam is designed so that answers requiring numerical calculation can be done accurately enough without a calculator.
- You are allowed one 8.5 by 11 inch page of notes with writing or printing on one side only.
- Use your own paper.
- If you are unable to staple your exam sheets, please fold them to keep them together.
- To get full credit, you must justify your answer. For instance, in Problem 1, you should explain why your choice could be correct and none of the others could be.
- HAND IN YOUR EXAM PROMPTLY AT 1:50 PM!

1. [30 points] A logistic model of the form P[Y = 1|X = x] = p(x) is fit where x is a 1-dimensional continuous variable. The fitted model is

$$logit(p(x)) = -4.79 + 3.52x.$$

Below are plots of p(x) for different logistic models. Determine which one could be the plot of the given fitted model. In each plot, the vertical line corresponds to the axis where x = 0.



2. [20 points] Explain the difference between Pearson residuals and

standardized residuals. Give an example of a family (model) where the two are the same and an example where they are different.

**3. [15 points]** Define the linear (or identity) link function for a GLM, and explain why it is seldom used for the binomial or Poisson families.

4. [20 points] Below is some output from the fit of a log-linear model that is created from 3 categorical variables X, Y, and Z. Use this ouput to answer the questions that follow.

```
> summary(fit)
Call:
glm(formula = n ~ X + Y + Z + X * Y + X * Z + Y * Z, family = poisson())
Deviance Residuals:
      1
               2
                         3
                                  4
                                            5
                                                     6
                                                               7
                                                                        8
                                                                                  9
                  -2.0992 -0.9309 -0.8693
                                                1.7998
                                                        -3.0255
 1.2159
          0.2425
                                                                 -0.7586
                                                                             2.4604
Coefficients:
            Estimate Std. Error z value Pr(|z|)
                          0.1814
                                  18.251
(Intercept)
              3.3103
                                           < 2e-16 ***
                                   5.198 2.01e-07 ***
XX2
                          0.2104
              1.0939
             -0.2750
                          0.2604
                                  -1.056
ХХЗ
                                          0.29111
YY2
              0.6525
                          0.2172
                                   3.004
                                          0.00266 **
ZZ2
             -1.2815
                          0.3028
                                  -4.233 2.31e-05 ***
XX2:YY2
             -3.0822
                          0.3934
                                  -7.835 4.69e-15 ***
XX3:YY2
             -0.6343
                          0.3119
                                  -2.034
                                           0.04196 *
XX2:ZZ2
                          0.4042
                                  -2.235
             -0.9032
                                           0.02545 *
XX3:ZZ2
              0.5676
                          0.2886
                                   1.967
                                           0.04920 *
YY2:ZZ2
              1.2385
                          0.3019
                                   4.103 4.08e-05 ***
____
Signif. codes:
                0 *** 0.001 ** 0.01 * 0.05 . 0.1
                                                     1
(Dispersion parameter for poisson family taken to be 1)
```

Null deviance: 248.813 on 11 degrees of freedom Residual deviance: 31.192 on 2 degrees of freedom AIC: 106.67 Number of Fisher Scoring iterations: 5

(a) Do the results show any evidence of independence or conditional independence between any pair of the variables X, Y, Z?

(b) Write an expression for an approximate 95% confidence interval for  $\lambda_2^Z$  using the numbers from the output. You don't need to do any arithmetic.

(c) Can you give a 95% confidence interval for  $\lambda_1^X$ ?

(d) Write a few lines of R-code to compute P[Z = 1|X = 2&Y = 2].

5. [15 [points] Below is a plot of some data on a binary outcome variable Y that (possibly) depends on a predictor variable X.



Here are the results of fitting a logistic regression model to the data:

```
> summary(fit)
```

Call: glm(formula = y ~ x, family = binomial()) Deviance Residuals: Min 1Q Median ЗQ Max -1.0887 -1.0216 -0.8943 1.3662 1.4278 Coefficients: Estimate Std. Error z value Pr(>|z|)(Intercept) -0.7654 0.8878 -0.862 0.389 0.1383 0.3734 0.370 0.711 х (Dispersion parameter for binomial family taken to be 1) Null deviance: 27.910 on 20 degrees of freedom

Residual deviance: 27.772 on 19 degrees of freedom AIC: 31.772

## Number of Fisher Scoring iterations: 4

(a) Does the output of the model offer any evidence of dependence between X and Y?

(b) Do you think the data plot offers any evidence of dependence between X and Y?

(c) Suggest a better model.