STAT 4130: Homework 4

Due: 2023-07-26

Question 1

The data in the table below give the percentage share of market of a particular brand of canned peaches (y_t) for the past 15 months and the relative selling price (x_t).

Market Share and Price of Canned Peaches

t	x_t	y_t	t	x_t	y_t
1	100	15.93	9	85	16.60
2	98	16.26	10	83	17.16
3	100	15.94	11	81	17.77
4	89	16.81	12	79	18.05
5	95	15.67	13	90	16.78
6	87	16.47	14	77	18.17
7	93	15.66	15	78	17.25
8	82	16.94			

a. Fit a simple linear regression model to these data. Plot the residuals versus time. Is there any indication of autocorrelation?
b. Use the Durbin–Watson test to determine if there is positive autocorrelation in the errors. What are your conclusions?
c. Use one iteration of the Cochrane–Orcutt procedure to estimate the regression coefficients. Find the standard errors of these regression coefficients.
d. Is there positive autocorrelation remaining after the first iteration? Would you conclude that the iterative parameter estimation technique has been successful?

Question 2

The market research department of a soft drink manufacturer is investigating the effectiveness of a price discount coupon on the purchase of a two-liter beverage product. A sample of 5500 customers was given coupons for varying price discounts between 5 and 25 cents. The response variable was the number of coupons in each price discount category redeemed after one month. The data are shown below.

Discount, x	Sample Size, n	Number Redeemed, r
5	500	100
7	500	122
9	500	147
11	500	176
13	500	211
15	500	244
17	500	277
19	500	310
21	500	343
23	500	372
25	500	391

```
# You can directly copy this into your homework
data <- tibble(
  discount = c(5,7,9,11,13,15,17,19,21,23,25),
  size = c(500,500,500,500,500,500,500,500,500,500,500),
  redeem = c(100,122,147,176,211,244,277,310,343,372,391)
)
```

a. Fit a logistic regression model to the data. Use a simple linear regression model as the structure for the linear predictor ($\eta = \beta_0 + \beta_1 x$).
b. Does the model deviance indicate that the logistic regression model from part a is adequate?
c. Draw a graph of the data and the fitted logistic regression model.
d. Expand the linear predictor to include a quadratic term. Is there any evidence that this quadratic term is required in the model?
e. Draw a graph of this new model on the same plot that you prepared in part c. Does the expanded model visually provide a better fit to the data than the original model from part a?
f. Find approximate 95% confidence intervals on the model parameters for the quadratic logistic regression model from part d.

Question 3

Myers [1990] presents data on the number of fractures (frac) that occur in the upper seams of coal mines in the Appalachian region of western Virginia. Four regressors were reported: inb = inner burden thickness (feet), the shortest distance between seam floor and the lower seam; extrp = percent extraction of the lower previously mined seam; seamh = lower seam height (feet); and time = time (years) that the mine has been in operation. The data are shown below.

```
library(locfit)  #if you don't have this package, try install.packages('locfit')
data(mine)
mine
```

```
##    frac inb extrp seamh time
## 1     2  50    70    52  1.0
## 2     1 230    65    42  6.0
## 3     0 125    70    45  1.0
## 4     4  75    65    68  0.5
## 5     1  70    65    53  0.5
## 6     2  65    70    46  3.0
## 7     0  65    60    62  1.0
## 8     0 350    60    54  0.5
## 9     4 350    90    54  0.5
```

```
## 10    4 160    80    38  0.0
## 11    1 145    65    38 10.0
## 12    4 145    85    38  0.0
## 13    1 180    70    42  2.0
## 14    5  43    80    40  0.0
## 15    2  42    85    51 12.0
## 16    5  42    85    51  0.0
## 17    5  45    85    42  0.0
## 18    5  83    85    48 10.0
## 19    0 300    65    68 10.0
## 20    5 190    90    84  6.0
## 21    1 145    90    54 12.0
## 22    1 510    80    57 10.0
## 23    3  65    75    68  5.0
## 24    3 470    90    90  9.0
## 25    2 300    80   165  9.0
## 26    2 275    90    40  4.0
## 27    0 420    50    44 17.0
## 28    1  65    80    48 15.0
## 29    5  40    75    51 15.0
## 30    2 900    90    48 35.0
## 31    3  95    88    36 20.0
## 32    3  40    85    57 10.0
## 33    3 140    90    38  7.0
## 34    0 150    50    44  5.0
## 35    0  80    60    96  5.0
## 36    2  80    85    96  5.0
## 37    0 145    65    72  9.0
## 38    0 100    65    72  9.0
## 39    3 150    80    48  3.0
## 40    2 150    80    48  0.0
## 41    3 210    75    42  2.0
## 42    5  11    75    42  0.0
## 43    0 100    65    60 25.0
## 44    3  50    88    60 20.0
```

a. Fit a Poisson regression model to these data using the log link.
b. Does the model deviance indicate that the model from part a is satisfactory?
c. Find approximate 95% Wald confidence intervals on the model parameters.
d. Reconsider the model. Remove any regressors from the original model that you think might be unimportant and comment on your findings.
e. Construct plots of the deviance residuals from part a and d and comment on the plots. Does the model appear satisfactory from a residual analysis viewpoint?