

# STA 223, HW3

Bohao Zou

February 10, 2021

## 1 Set 2

5.

For the type  $(X_{i1}, X_{i2}, Y_i)_{i=1 \dots n}$

$$\sum_{i=1}^n (y_i - (\beta_0 + \beta_1(x_{i1} - x) + \beta_2(x_{i2} - x)))^2 K\left(\frac{\|X - \vec{X}_i\|}{h}\right)$$

For p-dimensional predictors, we have:

$$\sum_{i=1}^n (y_i - \beta_0 - \vec{\beta}^T \vec{x}_p)^2 K\left(\frac{\|X - \vec{X}_i\|}{h}\right)$$

If the number of predictors  $p$  is large, then the performance of this function is worse. We need to add more points into our dataset. But the data will be more sparse. However, if there is not enough data points to support this function in some of interval, the variance of this function will be very high.

7 (c).

First, we need to calculate the variance of  $\hat{g}(x)$ .

$$Var(\hat{g}(x)) = Var\left(\sum_{i=1}^n W_i(x) Y_i\right) = \sum_{i=1}^n W_i(x)^2 \sigma^2$$

We can build a function  $T$ , Let  $T = var(\hat{g}(x)) = \sum_{i=1}^n W_i(x)^2 \sigma^2 - \lambda(\sum_{i=1}^n W_i(x) - 1)$ . We can find the derivative of  $T$  to  $W_i$ .  $\frac{\partial T}{\partial W_j} = 2\sigma^2 W_j(x) - \lambda = 0$ . From this we can get  $W_j(x) = \frac{\lambda}{2\sigma^2}$ . By this property  $\sum_{j=1}^n W_j(x) = 1$ , we can know  $W_j(x) = \frac{1}{n}$ . This implies it is a rectangular estimator. I would not choose this in practice. This is because this kernel is equal probability for all the points in the interval. This may conduct non-smoothing in practice.

## 2 Set 3

### 1.

The pmf of poisson is:

$$f(y, \lambda) = \frac{\lambda^y e^{-\lambda}}{y!} = \exp\{y \log(\lambda) - \lambda - \log(y!)\}$$
$$\theta = \log(\lambda), \phi = 1, b(\theta) = \lambda = e^\theta, c(y, \theta) = -\log(y!)$$

The link function is  $\eta = \log(\mu) = \log(\lambda)$

From this we can know  $\theta = \eta$ . The log likelihood function is:

$$l(\theta, y) = \sum_{i=1}^n [y_i \theta_i - e^{\theta_i} - \log(y_i!)]$$

Because we know  $\theta = \eta$ , we can have:

$$l(\beta, y) = \sum_{i=1}^n [y_i (\sum_j x_{ij} \beta_j) - e^{\sum_j x_{ij} \beta_j} - \log(y_i!)]$$

So, for one element of score vector:

$$\begin{aligned} u_r &= \frac{\partial l(\beta, y)}{\partial \beta_r} \\ &= \sum_{i=1}^n [y_i x_{ir} - e^{\sum_j x_{ij} \beta_j} x_{ir}] \\ &= \sum_{i=1}^n [(y_i - \lambda_i) x_{ir}] \end{aligned}$$

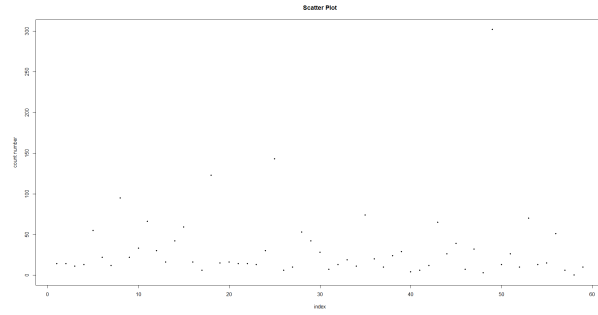
For one element of Hessian matrix  $h_{rs}$ , we know that  $h_{rs} = \frac{\partial^2 l}{\partial \beta_r \partial \beta_s} = \frac{\partial u_r}{\partial \beta_s}$ , we can have:

$$\begin{aligned} h_{rs} &= \frac{\partial^2 l}{\partial \beta_r \partial \beta_s} = \frac{\partial u_r}{\partial \beta_s} \\ &= \sum_{i=1}^n [e^{\sum_j x_{ij} \beta_j} x_{is} x_{ir}] \\ &= \sum_{i=1}^n [\lambda_i x_{is} x_{ir}] \end{aligned}$$

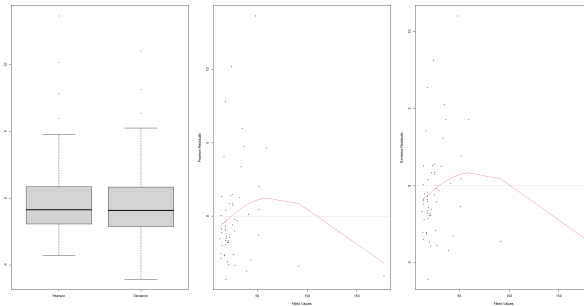
The observed information matrix is the negative of Hessian matrix, i.e.  $I_{obs} = -H$ . The expected information matrix is  $I = E[I_{obs}] = -E[H] = -H$  since  $h_{rs}$  does not depend on  $y$ .

### 7

The scatter plot to detect outlier is this:



From this plot we can find a outlier and remove it from our dataset. We can use the boxplot and Pearson&Deviance residual plot to check the goodness of fit.



From this plot we can know that there exist lack of fit in our model because there exist systematic pattern in those two types of residuals plots. The estimated coefficients are showed below:

|             | Estimated  | Std.Error | Z-value | P-value  |
|-------------|------------|-----------|---------|----------|
| (Intercept) | 1.9426534  | 0.1382858 | 14.048  | <2e-16   |
| baseline    | 0.0228025  | 0.0008308 | 27.447  | <2e-16   |
| age         | 0.02268    | 0.0040317 | 5.625   | 1.85E-08 |
| treatment1  | -0.1471737 | 0.0535225 | -2.75   | 0.00596  |

The researcher would be more interested in "Active Treatment". From this table we can know active a treatment would reduce the number of epileptic seizures during four two-week periods.

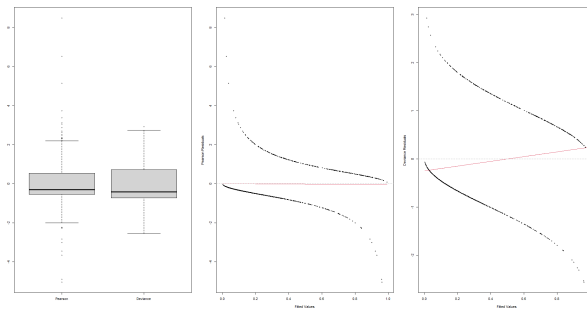
**8**

**(a)**

The estimated coefficients for model with all predictors are showed below:

| term        | estimate  | std.error | statistic  | p.value  |
|-------------|-----------|-----------|------------|----------|
| (Intercept) | -8.404696 | 0.716636  | -11.727987 | 0        |
| V1          | 0.123182  | 0.032078  | 3.84014    | 0.000123 |
| V2          | 0.035164  | 0.003709  | 9.481393   | 0        |
| V3          | -0.013296 | 0.005234  | -2.540416  | 0.011072 |
| V4          | 0.000619  | 0.006899  | 0.089713   | 0.928515 |
| V5          | -0.001192 | 0.000901  | -1.322309  | 0.186065 |
| V6          | 0.089701  | 0.015088  | 5.945334   | 0        |
| V7          | 0.94518   | 0.299147  | 3.159578   | 0.00158  |
| V8          | 0.014869  | 0.009335  | 1.592858   | 0.111192 |

The zero in table of p.value means the p-value is too small to appear. The diagnostics for this model is showed below:



The Runs test for Pearson residuals and Deviance residuals are all accepted  $H_0$ . So, this model have a good of fit.

**(b)**

The selectors of AIC is showed below:

| term        | estimate  | std.error | statistic  | p.value  |
|-------------|-----------|-----------|------------|----------|
| (Intercept) | -8.405136 | 0.716703  | -11.727498 | 0        |
| V1          | 0.123172  | 0.032069  | 3.840884   | 0.000123 |
| V2          | 0.035112  | 0.003662  | 9.587038   | 0        |
| V3          | -0.013214 | 0.005154  | -2.563913  | 0.01035  |
| V5          | -0.001157 | 0.000814  | -1.421142  | 0.155275 |
| V6          | 0.090089  | 0.014462  | 6.229371   | 0        |
| V7          | 0.947595  | 0.298006  | 3.179783   | 0.001474 |
| V8          | 0.014789  | 0.00929   | 1.591965   | 0.111393 |

The selectors of BIC is showed below:

| term        | estimate  | std.error | statistic  | p.value  |
|-------------|-----------|-----------|------------|----------|
| (Intercept) | -8.415851 | 0.656908  | -12.811313 | 0        |
| V1          | 0.141926  | 0.027105  | 5.236104   | 0        |
| V2          | 0.033826  | 0.003345  | 10.11169   | 0        |
| V6          | 0.078097  | 0.013771  | 5.67114    | 0        |
| V7          | 0.901294  | 0.291696  | 3.089834   | 0.002003 |

From those results we can know the predictors in the final model which selected by BIC is smaller than AIC.

### 2.0.1 (c)

From the final model of AIC, we can know "V1,2,3,4,5,6,7" are all significant under  $\alpha = 0.05$ . We can say diabetes have high correlated with those factors. "V1,2,6,7" can increase the risk to get diabetes. "V3,5" can decrease the risk to get diabetes.

## 9

### (a)

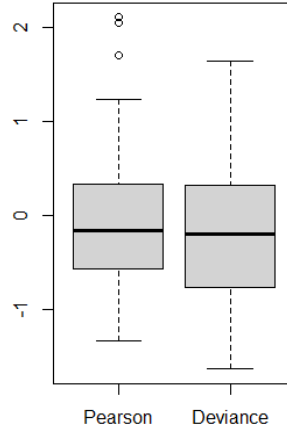
The deviance table is:

| term      | df | Deviance   | Resid..Df | Resid..Dev | p.value  |
|-----------|----|------------|-----------|------------|----------|
| NULL      | NA | NA         | 64        | 322.526768 | NA       |
| Dust      | 2  | 252.108219 | 62        | 70.418549  | 0        |
| Race      | 1  | 2.453634   | 61        | 67.964915  | 0.117253 |
| Smoker    | 1  | 11.793614  | 60        | 56.1713    | 0.000594 |
| Sex       | 1  | 1.8e-05    | 59        | 56.171282  | 0.996593 |
| EmpLength | 1  | 12.261884  | 58        | 43.909398  | 0.000462 |

From this table we should include Dust, Smoker and EmpLength. The AIC table shows below:

| term        | estimate  | std.error | statistic  | p.value  |
|-------------|-----------|-----------|------------|----------|
| (Intercept) | -2.139866 | 0.212967  | -10.047889 | 0        |
| Dust2       | -2.551772 | 0.261355  | -9.763621  | 0        |
| Dust3       | -2.715364 | 0.189822  | -14.304778 | 0        |
| Smoker2     | -0.630283 | 0.190316  | -3.311769  | 0.000927 |
| EmpLength   | 0.33486   | 0.089772  | 3.730116   | 0.000191 |

From this table, we can know the selectors that selected by AIC are the same with Deviance table. The boxplot of deviance and Pearson residuals are showed below:



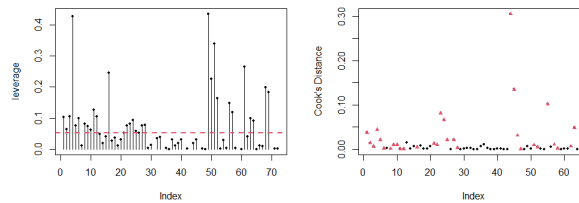
A high dust workplace can increase the risk to get Byssinosis. Smoking can increase the risk to get Byssinosis. The length of employment can increase the risk to get Byssinosis.

(b)

The Smoking and Length of employment should be considered as one-sided test. This is because in our common sense, smoking and the long time of employment would increase the risk to get some lung diseases. Those factors should be positive. The hypotheses of smoking is  $H_0 : \beta_{smoking2} \geq 0, H_1 : \beta_{smoking2} < 0$ . The hypotheses of Length of employment is  $H_0 : \beta_{Emplength} \leq 0, H_1 : \beta_{Emplength} > 0$ . The result is that under significant level 0.05, the predictors of smoking and Length of employment are all significant.

(c)

The Cook's distance and leverage plots is showed below:



A high leverage value or high Cook's distance are considered this point as a potential outlier point. In practice, if leverage value bigger than  $\frac{2p}{n}$ , then the i-th case is identified as influential points. If Cook's distance bigger than  $\frac{4}{n-p}$ , the i-th case will be considered as potential influential case. By using leverage value, there are 10 instances treated as potential outliers. By using Cook's distance, there are 5 instances treated as potential outliers.

(d)

$$\beta = [\beta_0, \beta_1, \dots, \beta_{p-1}]$$
$$A = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = \text{Identity Matrix}$$
$$H_0 : A\beta = 0$$

The df is the number of coefficients in model. I will add the interaction between Dust and Length of employment. Because I think there exist some relation between those two variables.

The coefficients of the final model is:

| term           | estimate  | std.error | statistic | p.value  |
|----------------|-----------|-----------|-----------|----------|
| (Intercept)    | -0.622994 | 0.524431  | -1.187941 | 0.234857 |
| Dust           | -1.177688 | 0.25472   | -4.623469 | 4e-06    |
| Smoker         | -0.675549 | 0.188976  | -3.574787 | 0.000351 |
| EmpLength      | 0.566226  | 0.210753  | 2.686687  | 0.007216 |
| Dust:EmpLength | -0.139612 | 0.114173  | -1.22281  | 0.221401 |

From this table we can know a high dust workplace can increase the risk to get Byssinosis. Smoking can increase the risk to get Byssinosis. The length of employment can increase the risk to get Byssinosis. The interaction between Dust and Length of employment is not significant under 0.05. This means there is no interaction between those predictors.

## Set 4

2

(a)

$$E[Y] = E[E[Y|Z]] = E[Z] = \frac{\kappa}{\rho}$$

(b)

$$\begin{aligned} \text{Var}(Y) &= E[\text{Var}(Y|Z)] + \text{Var}(E[Y|Z]) \\ &= E[Z] + \text{Var}(Z) \\ &= \frac{\kappa}{\rho} + \frac{\kappa}{\rho^2} \\ &= \frac{\kappa(\rho + 1)}{\rho^2} \end{aligned}$$

(c)

Because  $E[Y] = \mu = \frac{\kappa}{\rho}$ , we can have  $\kappa = \mu\rho$ . Plug into the variance we can have  $\text{Var}(Y) = \frac{\mu(\rho+1)}{\rho}$ . The limitation  $\lim_{\rho \rightarrow \infty} \frac{\mu(\rho+1)}{\rho} = \mu = E[Y]$ . This means when the case of small and large  $\rho$ , the variance of  $Y$  is equal with the expectation of  $Y$ .

### 3

Aranda-Ordaz link family:

$$g(\mu_i, \alpha) = g(\pi_i, \alpha) = \log\left\{\frac{(1 - \pi_i)^{-\alpha} - 1}{\alpha}\right\}$$

With the Taylor expansion, we have:

$$\begin{aligned} g(\pi_i, \alpha) &\approx g(\pi_i, \alpha_0) + (\alpha - \alpha_0) \frac{\partial g(\pi_i, \alpha)}{\partial \alpha} \\ &= \eta_i + (\alpha - \alpha_0) \gamma_i \end{aligned}$$

For  $g$  function in above and fits  $\hat{\pi}_i$ , we have:

$$\begin{aligned} \gamma_i &= \frac{\partial g(\pi_i, \alpha)}{\partial \alpha} \\ &= -\frac{\alpha}{(1 - \pi_i)^{-\alpha} - 1} \frac{(1 - \pi_i)^{-\alpha} - 1 + \alpha(1 - \pi_i)^{-\alpha} \log(1 - \pi_i)}{\alpha^2} \\ &= -\left(\frac{1}{\alpha} + \frac{(1 - \pi_i)^{-\alpha} \log(1 - \pi_i)}{(1 - \pi_i)^{-\alpha} - 1}\right) \\ &= \frac{\log(1 - \hat{\pi}_i)}{(1 - \hat{\pi}_i)^{\alpha} - 1} - \frac{1}{\alpha} \end{aligned}$$

For the special case  $\alpha_0 = 1$ , we have:

$$\gamma_i = -\left[1 + \frac{\log(1 - \hat{\pi}_i)}{\hat{\pi}_i}\right]$$