Homework #5

April 17, 2017

1.

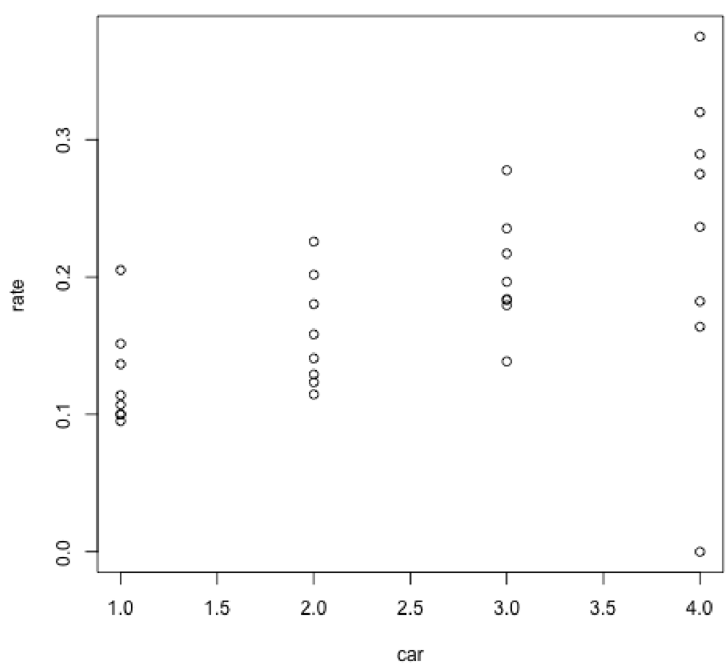
(a) $L(\beta) = \prod_{i=1}^{N} e^{e^{-\beta_1 - \sum_{j=2}^{J} x_{ij}\beta_j}} e^{Y_i(\beta_1 + \sum_{j=2}^{J} x_{ij}\beta_j)} / Y_i!$
$l(\beta) = \sum_{i=1}^{N} e^{-\beta_1 - \sum_{j=2}^{J} x_{ij}\beta_j} + Y_i(\beta_1 + \sum_{j=2}^{J} x_{ij}\beta_j)$
$U_1 = \frac{\partial l(\beta)}{\partial \beta_1} = \sum_{i=1}^{N} -e^{-\beta_1 - \sum_{j=2}^{J} x_{ij}\beta_j} + Y_i = \sum_{i=1}^{N} Y_i - \mu_i$

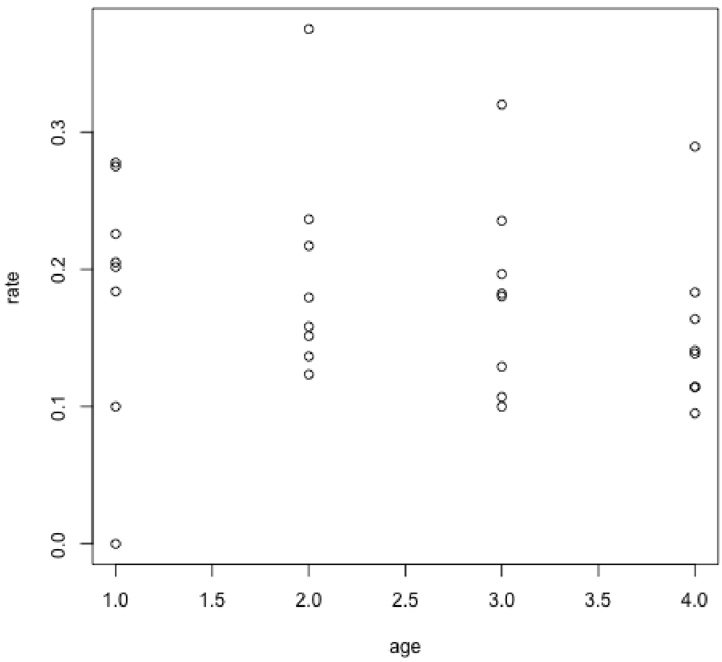
(b) We must have $U_1 = 0$, So $\sum_{i=1}^{N} \hat{\mu}_i = \sum_{i=1}^{N} Y_i$.

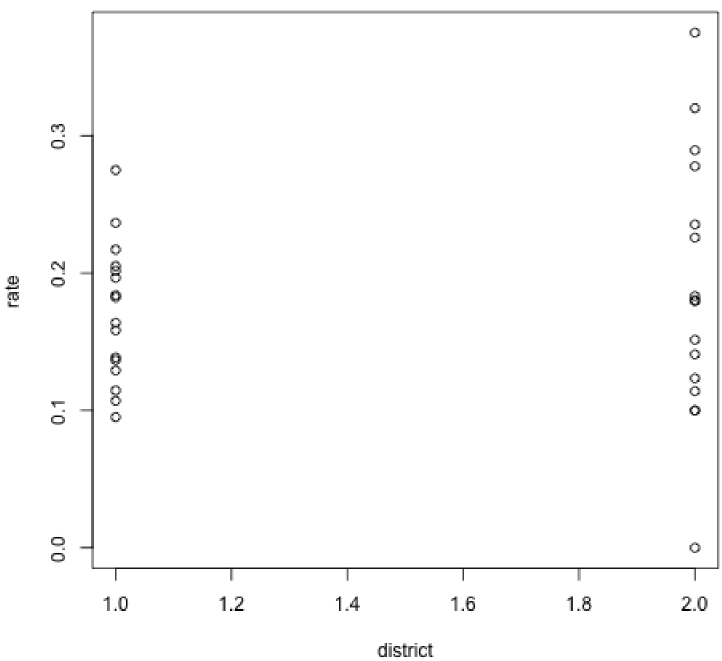
(c) $D = 2\sum_{i=1}^{N} [Y_i \log(Y_i/\hat{\mu}_i) - (Y_i - \hat{\mu}_i)]$
$D = 2\sum_{i=1}^{N} [Y_i \log(Y_i/\hat{\mu}_i)]$

2.

(a) See Table 1. There is a decreasing trend in rate when age is increasing. There is an increasing trend in rate when car is increasing. There is an increasing trend in rate when district is increasing.







(b) See Table 2.

(c) See Table 3. The model (c) is embedded in the model in (b), we can do a LRT. $LRT = 2*(14138.91 - 14129.22) = 19.39 \sim \chi^2_{19}, \chi^2_{19,0.95} = 30.14 > 19.39$, fail to reject H_0 : the model in (c) is better than the model in (b). So the conclusion is that the model in (c) is better.

(d) age: log rate ratio of insurance claims for one year increase in age, holding other covariates constant.
car: log rate ratio of insurance claims for one unit increase in car insurance category,

district=0	age=1	age=2	age=3	age=4
car=1	0.20	0.14	0.11	0.10
car=2	0.20	0.16	0.13	0.11
car=3	0.18	0.22	0.20	0.14
car=4	0.28	0.24	0.18	0.16
district=1	**age=1**	**age=2**	**age=3**	**age=4**
car=1	0.10	0.15	0.10	0.11
car=2	0.23	0.12	0.18	0.14
car=3	0.28	0.18	0.23	0.18
car=4	0.00	0.38	0.32	0.29

Table 1: Table for 2(a).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.61	0.12	-13.05	0.00
age2	-0.36	0.17	-2.12	0.03
age3	-0.64	0.18	-3.53	0.00
age4	-0.74	0.13	-5.50	0.00
car2	0.02	0.16	0.11	0.91
car3	-0.05	0.19	-0.25	0.80
car4	0.22	0.33	0.68	0.50
district1	-0.13	0.30	-0.42	0.67
age2:car2	0.10	0.21	0.49	0.62
age3:car2	0.20	0.22	0.92	0.36
age4:car2	0.16	0.17	0.97	0.33
age2:car3	0.49	0.24	2.00	0.05
age3:car3	0.66	0.25	2.68	0.01
age4:car3	0.42	0.20	2.07	0.04
age2:car4	0.34	0.38	0.89	0.37
age3:car4	0.33	0.38	0.86	0.39
age4:car4	0.32	0.34	0.95	0.34
car2:district1	0.08	0.18	0.46	0.64
car3:district1	0.13	0.19	0.66	0.51
car4:district1	0.43	0.22	1.92	0.05
age2:district1	-0.06	0.34	-0.19	0.85
age3:district1	0.27	0.32	0.84	0.40
age4:district1	0.27	0.29	0.95	0.34

Table 2: Table for 2(b)

holding other covariates constant.
district: log rate ratio of insurance claims comparing policy holders who live in London and other major cities to those who do not, holding other covariates constant.

(e) In the model in (c), the residual deviance and its degree of freedom are 24.685 and 28. We can estimate the dispersion parameter as 24.685/28=0.8816. It is less than one, so there is no over dispersion issue.

3.

(a) See Table 4. Model 2 and model 3 are identical, all three models have similar estimates.

(b) $H_0 : \beta_2 = \beta_5 = 0$
Model 1:
Wald test statistics: 18.37525$\sim \chi^2_2$
Reject H_0 at $\alpha = 0.05$.
Model 2 and 3:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.83	0.05	-33.74	0.00
age (coded as 0,1,2,3)	-0.18	0.02	-9.56	0.00
car(coded as 0,1,2,3)	0.20	0.02	9.51	0.00
district1	0.22	0.06	3.74	0.00

Table 3: Table for 2(c)

	Model 1	Model 2&3
(Intercept)	-10.79	-10.78
smoking	1.44	1.44
age	2.38	2.37
age square	-0.20	-0.20
smoking:age	-0.31	-0.31

Table 4: Table for 2(e)

Wald test statistics: 18.31824$\sim \chi_2^2$
Reject H_0 at $\alpha = 0.05$.