STAT 495 SECOND MIDTERM EXAMINATION SOLUTION

Problem 1. During a measles outbreak, a clinic conducted a study on 65 infants for the presence of measles symptoms. The subjects were evaluated by a doctor during four visits with a one-week interval between consecutive visits. The data also contain information on children's gender, how many days born prematurely, baseline age (in weeks), weight (in kg), and height (in cm). Use the SAS code and output to answer the following questions:

(a) Specify the fitted model. Give estimates of all parameters. What is the name of this model? Is the model appropriate?

This is a random intercept logistic regression model. The fitted model is

$$\frac{\widehat{P}(\boldsymbol{measles\ symptoms})}{\mathbf{1}-\ \widehat{P}(\boldsymbol{measles\ symptoms})} = \mathbf{exp}(\mathbf{0.07505} - \mathbf{0.04590}\cdot \boldsymbol{Male} + \ \mathbf{0.03925}\cdot \boldsymbol{DaysPremature}$$

$$-\mathbf{0.00116}\cdot \boldsymbol{Age} - \mathbf{0.00350}\cdot \boldsymbol{Weight} - \mathbf{0.01451}\cdot \boldsymbol{Height} + \mathbf{0.06195}\cdot \boldsymbol{Week}),$$

with $\widehat{\sigma}_u^2 = \mathbf{0.2416}$. The model is appropriate because the *p*-value for testing $H_0\colon \sigma_u^2 = 0$ against $H_1\colon \sigma_u^2 > 0$ is 0.0265 < 0.05, thus the random intercept is needed in the model.

(b) Identify significant predictors and interpret their estimated coefficients. Use an alpha of 0.05.

Gender and DaysPremature are significant at the 5% level. For male infants, the estimated odds of measles symptoms are $\mathbf{exp}(-\mathbf{0.0459})\cdot \mathbf{100\%} = \mathbf{95.51\%}$ of those for female infants. As the number of days of premature birth increases by one, the estimated odds of measles symptoms increase by $(\mathbf{exp}(\mathbf{0.03925}) - \mathbf{1})\cdot \mathbf{100\%} = \mathbf{4.00\%}$.

(c) Predict the probability of measles symptoms at week three for a male infant who was born two days prematurely, and whose baseline age was 20 weeks, weight of 6.8 kg, and height of 63.9 cm. Show by-hand calculations. Compare your result with the one in the printout.

We calculate $\mathbf{exp}(\mathbf{0.07505} - \mathbf{0.04590} + \ \mathbf{0.03925}\cdot \mathbf{2} - \mathbf{0.00116}\cdot \mathbf{20} - \mathbf{0.00350}\cdot \mathbf{6.8} - \mathbf{0.01451}\cdot \mathbf{63.8} + \mathbf{0.06195}\cdot \mathbf{3}) = \mathbf{0.507003}$. Thus,

$P^0(\boldsymbol{measles\ symptoms}) = \frac{0.507003}{1+0.507003} = \mathbf{0.336431}$. From SAS, it is 0.33617, which is practically identical.

Problem 2. A medical center specializing in treating heart diseases conducts a study on the reliability of pacemakers, devices that stimulate contractions of heart muscles by supplying timely electrical charge. Forty-five qualified patients were chosen for the study. Their gender, age at baseline, the number of co-morbid conditions at baseline, the number of daily medications taken at baseline, and the self-reported number of false alarms of pacemakers within one year are recorded. Use the R code and output to answer these questions:

(a) Write down the fitted model. Give its name.

This is a Poisson regression. The fitted model is

$$\hat{\boldsymbol{E}}(\boldsymbol{number\ of\ false\ alarms}) = \mathbf{exp}(-\mathbf{0.366966} + \mathbf{0.252922} \cdot \boldsymbol{Female} + \mathbf{0.011557} \cdot \boldsymbol{Age} + \mathbf{0.150839} \cdot \boldsymbol{NComorbidities} - \mathbf{0.009479} \cdot \boldsymbol{NMedications}).$$

(b) What predictors are significant at the 5% significance level? Interpret only the significant estimated slopes.

Gender and number of co-morbid conditions are significant predictors at the 5% level. For female patients, the estimated mean number of false alarms is $\mathbf{exp}(\mathbf{0.252922}) \cdot \mathbf{100\%} = \mathbf{128.7783\%}$ of that for male patients. As the number of co-morbid conditions increases by one, the estimated mean number of false alarms increases by $(\mathbf{exp}(\mathbf{0.150839}) - \mathbf{1}) \cdot \mathbf{100\%} = \mathbf{16.28094\%}$.

(c) Use the fitted model to predict the number of false alarms of the pacemaker that a 72-year-old woman will experience within one year if, at the baseline, she has two co-morbid conditions and is taking five medications daily. Show your computations. Compare your result to the one in the output.

$$\boldsymbol{number\ of\ false\ alarms}^{\mathbf{0}} = \mathbf{exp}(-\mathbf{0.366966} + \mathbf{0.252922} + \mathbf{0.011557} \cdot \mathbf{72} + \mathbf{0.150839} \cdot \mathbf{2} - \mathbf{0.009479} \cdot \mathbf{5}) = \mathbf{2.644132}.$$

From the output, the predicted value is 2.644127, which is the same.

(d) What is the name of the fitted model? Write it down explicitly. What linear and loess terms are significant at the 5 % level?

This is a nonparametric Poisson regression model. The fitted model is

$$\hat{\boldsymbol{E}}(\boldsymbol{number\ of\ false\ alarms}) = \mathbf{exp}(-\mathbf{1.43695} + \mathbf{0.55932} \cdot \boldsymbol{Female} + \mathbf{0.01905} \cdot \boldsymbol{Age} + \boldsymbol{loess}(\boldsymbol{Age}) + \mathbf{0.20149} \cdot \boldsymbol{NComorbidities} + \boldsymbol{loess}(\boldsymbol{NComorbidities}) + \mathbf{0.04604} \cdot \boldsymbol{NMedications} + \boldsymbol{loess}(\boldsymbol{NMedications}))$$

where loess() denotes detrended loess functions.

Problem 3. Epidemiologists in a clinic specializing in sports injuries conduct a study on the efficacy of physical therapy in the rehabilitation of runners. The runners are made to run on a treadmill during three sessions with two weeks between consecutive sessions. The data contain the runner's gender, age, oxygen intake (in ml per kg body weight per minute), run time (time to run 1 mile, in minutes), and pulse (average heart rate while running). Use the SAS code and output to provide your answers to the following questions:

(a) Give the name of the model that was fitted to these data. Write down the fitted model. Specify all estimated parameters. Is this model appropriate? Give a quantitative explanation.

This is a random slope and intercept model for normal response. The fitted model is $\widehat{E}(pulse) = 172.88 + 4.7249 \cdot female - 0.1747 \cdot age - 0.9634 \cdot oxygen + 0.4824 \cdot runtime + 6.0839 \cdot run\,number$, with $\widehat{\sigma}^2_{u_1} = 37.5141, \widehat{\sigma}_{u_1 u_2} = -38.6973, \widehat{\sigma}^2_{u_2} = 33.0452$, and $\widehat{\sigma}^2 = 21.0330$. The model is appropriate because the distribution of pulse is normal as supported by the bell-shaped curve on the histogram and *p*-values larger than 0.05 for the normality tests. In addition, the variances of the random slope and intercept and the covariance term are significant which validates the use of the model.

(b) What predictors are significant at the 5% level? Give an interpretation of the estimated significant regression coefficients.

Gender, oxygen intake, and run number are significant predictors at the 5% level. For females, the estimated average pulse is 4.7249 points higher than that for males. As the oxygen intake increases by one unit, the estimated mean pulse decreases by 0.9634 points. From run to run, the estimated average pulse increases by 6.0839 points.

(c) Predict the pulse rate at the third run for a 22-year-old male runner whose oxygen intake is 40.2 ml per kg, and whose run time for one mile is 10.3 minutes. Calculate the value by hand and compare it to the one computed in SAS.

$pulse^0 = 172.88 - 0.1747 \cdot 22 - 0.9634 \cdot 40.2 + 0.4824 \cdot 10.3 + 6.0839 \cdot 3 = 153.5283$, which is very close to the outputted value of 153.531.