**MS5318 Final Exam Answer Sheet**

**Q1. Predict patient waiting time.** (65 pts)

(1) (10 pts) Name your model as g. Copy your R code here and attach the model summary below.

|  |
| --- |
|  |

(2) (10 pts) Based on the R output summary in (1), provide your interpretation of the coefficients corresponding to the predictors Age and Sex.

|  |
| --- |
|  |

(3) (5 pts) Does this estimated model explain statistically significant variation in the patient waiting time? **Explain your answer.**

|  |
| --- |
|  |

(4) (10 pts) Which hour of day has the lowest waiting time? Which day of week has the lowest waiting time? **Answer the questions and briefly explain why.** Note that Hour=7 represents 7 a.m. and Weekday=1 represents Monday.

|  |
| --- |
|  |

(5) (5 pts) A 24-years-old male patient arrives at the ED at 11 a.m. on Tuesday. The patient is classified as severity group 3. What is this patient’s predicted waiting time by model g? What is the 95% prediction interval using the formula we learnt in class (Keep 1 decimal place.)? **(Please provide your R code and output below.)**

|  |
| --- |
|  |

(6) (5 pts) Are there serious confounding issues with this model? Why? **(Please provide your R code and output below.)**

|  |
| --- |
|  |

(7) (5 pts) Please use the log transformed waiting time as the dependent variable and use Age, Severity Group, Hour, and Weekday as predictors. Name your new model as g1. Please provide your R code and report the value of the R-squared of model g1. *The model summary output is not needed.*

|  |
| --- |
|  |

(8) (5 pts) Recall that for multiple regression models, the error term should follow normal distributions with equal variance. Does the normal assumption hold for g1? Attach your R code and output below and answer this question.

|  |
| --- |
|  |

(9) (5 pts) Update model g1 by including the interaction term between Hour and Weekday. Name it model g2. Please provide your R code and report the value of the R-squared of model g2. The model output is not needed. Should the interaction be kept in the model? Please briefly explain your answer.

|  |
| --- |
|  |

(10) (5 pts) A 24-years-old male patient arrives at the ED at 11 a.m. on Tuesday. The patient is classified as severity group 3. What is this patient’s predicted waiting time by model g2?

|  |
| --- |
|  |

**Q2. Predict patient disposition.** (35 pts)

An emergency department (ED) physician is interested in what factors can best predict whether a patient will be admitted to the hospital or discharge home after the patient’s treatment in the ED. The predictors include patient **age** (at the time of ED visit), **sex** (male or female), **severity group** (a score indicating the patient’s urgency level), **the hour of day**, and **the day of week** upon this patient’s arrival. **Hint:** **Please pay attention to the type of the variables in your model.**

(*We have 2-3 groups of students who chose this problem as their class project. Please do not be misled as the volume of the data and the variables included in this data file are completely different. You should simply treat it as a new problem and analyze it using the models we learnt in class.*)

(1) (10 pts) Run a logistic regression to predict the probability of admission to the hospital using all the given predictors excluding their interactions. Name it model f. **Report your R code and the regression results below.**

|  |
| --- |
|  |

(2) (5 pts) Explain the effects of age on the admission probability based on the model outputs.

|  |
| --- |
|  |

(3) (5 pts) Which severity group has the highest admission probability, given that all other variables are fixed? Which hour of day in patient’s arrival time implies the highest admission probability, given that all other variables are fixed? How about the effect of the variable Weekday on the admission probability?

|  |
| --- |
|  |

(4) (5 pts) Is the entire model significant at 1% significance level? **Write down your R code and the output to justify your answer below.**

|  |
| --- |
|  |

(5) (5 pts) Note that to capture the nonlinear effect of age on the probability of admission, we categorize the Age variable into five groups to create a new variable AgeGroup as follows: 0<= age <18 years, 18<= age <40 years, 40<= age <55 years, 55<= age <70 years, and age>=70 years. Use AgeGroup instead of Age to re-run the model. Name it model f1. What is your interpretation of the effect of AgeGroup on the admission probability? **Copy your R code and the model summary to justify your answer below.**

|  |
| --- |
|  |

(6) (5 pts) A 24-years-old male patient arrives at the ED at 11 a.m. on Tuesday. The patient is classified as severity group 3. What is this patient’s predicted admission probability by model f1?

|  |
| --- |
|  |