Homework 2

March 10, 2018

Data Analysis 1-The Challenger

On January 28, 1986, the space shuttle Challenger disintegrated immediately after launch (after 1 minutes and 13 seconds), killing all seven astronauts on board (the first civilian astronaut among them). The attached data set (Challenger.txt) records data on O-rings (devices that seal the field joints of the solid rocket motors, which boost the shuttle into orbit), on 23 space shuttle launches prior to the accident. After each launch, the rocket motors are recovered from the ocean for inspection and possible reuse. There were 24 launches before the Challenger, but for one flight the motors were lost at sea. The data consist of: temperature (of the joint) at take-off in F degrees (temp); leak-check pressure in psi (pres), that is the pressure used to test the O-rings after assembly of the rockets[1]; number of O-rings that failed out of $n = 6$ (the total number of primary O-rings in the shuttle). Other variables in the set, which you may ignore, count the instances of erosion, blowby, damage. Erosion is caused by excessive heat burning up the O-ring. Blowby happens when gases rush by the O-ring.

1) Before the Challenger launch, it was suspected that the pressure used in the leak-check test (which as you can see was increased from 50 to 200 psi) could cause some erosion and that the temperature could have an effect on the O-rings. Use these data to try to understand the probability of failure as a function of temperature and of temperature and pressure (which means you should compare the models as well). Use then the fitted model to estimate the probability of failure of an O-ring when the temperature was 31 F (the launch temperature on January 20, 1986).

2) Approach the previous problem with LDA, QDA, and KNN with $K = 1$. In this case, use the ungrouped data Challengerungrouped.txt.

Data Analysis 2. Titanic

The Titanic sank when it struck an iceberg on its voyage to New York City from Southampton. Of the 2201 known passengers and crew, only 711 survived. The

[1] After ignition, heat and pressure build up in the motor, the O-rings which would erode under heat are protected by a putty which is displaced toward the O-ring under pressure

data in the file titanic.txt classify the people on board the ship according to sex (male/female), age in categorical form (child/adult); class (first, second, third class passengers, crew). The data are grouped so for each age/sex/class combination totaling to the number m the number of surviving people is reported.

1) Fit a logistic regression model using sex, age, and class as predictors (they are factors). Is the model adequate, that is, does the mean function obtained matches the data well?

2) Add now to the previous term all two-factors interactions and decide if any can be removed. If the mean function you have obtained matches the data well, summarize the result. For example, how does the survival of the first class passengers differ from the third class passenegers? Did children in first class survived more than children in third?

LDA and Logistic Regression. Theoretical Questions

Solve exercises 4.2 and 4.3 in the book ESL.