Quiz 4. AMS 597

Name: ______________________ SBU ID: ______________________

The quiz is due at the end of the lecture by 11:20am – please submit no later than 11:30am.
Please email your completed quiz to your TA at: song.jiecheng@stonybrook.edu

Please include (1) R code
(2) Output from R
(3) Answers to all the questions asked

Please keep yourself on Zoom video until you have emailed your solutions.

General Linear Model with the Math Performance Data

The accompanying csv file contains data including student math scores in three school periods, demographic, social and school related features. Each case is a student, and the variables are:

1 school – student’s school (binary: each student is from one of the two schools, ’GP’ or ’MS’)
2 sex – student’s sex (binary: ’F’ – female or ’M’ – male)
3 age – student’s age (numeric: from 15 to 22)
4 address – student’s home address type (binary: ’U’ – urban or ’R’ – rural)
5 famsize – family size (binary: ’LE3’ – less or equal to 3 or ’GT3’ – greater than 3)
6 Pstatus – parent’s cohabitation status (binary: ’T’ – living together or ’A’ – apart)
7 Medu – mother’s education (numeric: 0 – none, 1 – primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
8 Fedu – father’s education (numeric: 0 – none, 1 – primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
9 Mjob – mother’s job (nominal: ’teacher’, ’health’ care related, civil ’services’ (e.g. administrative or police), ’at_home’ or ’other’)
10 Fjob – father’s job (nominal: ’teacher’, ’health’ care related, civil ’services’ (e.g. administrative or police), ’at_home’ or ’other’)
11 reason – reason to choose this school (nominal: close to ’home’, school ’reputation’, ’course’ preference or ’other’)
12 guardian – student’s guardian (nominal: ’mother’, ’father’ or ’other’)
13 traveltime – home to school travel time (numeric: 1 – <15 min., 2 – 15 to 30 min., 3 – 30 min. to 1 hour, or 4 – >1 hour)
14 studytime – weekly study time (numeric: 1 – <2 hours, 2 – 2 to 5 hours, 3 – 5 to 10 hours, or 4 – >10 hours)
15 failures – number of past class failures (numeric: n if 1<=n<3, else 4)
16 schoolsup – extra educational support (binary: yes or no)
17 famsup – family educational support (binary: yes or no)
18 paid – extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19 activities – extra-curricular activities (binary: yes or no)
20 nursery – attended nursery school (binary: yes or no)
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>higher</td>
<td>wants to take higher education (binary: yes or no)</td>
<td>binary: yes or no</td>
</tr>
<tr>
<td>internet</td>
<td>Internet access at home (binary: yes or no)</td>
<td>binary: yes or no</td>
</tr>
<tr>
<td>romantic</td>
<td>with a romantic relationship (binary: yes or no)</td>
<td>binary: yes or no</td>
</tr>
<tr>
<td>famrel</td>
<td>quality of family relationships (numeric: from 1 - very bad to 5 - excellent)</td>
<td>numeric: from 1 to 5</td>
</tr>
<tr>
<td>freetime</td>
<td>free time after school (numeric: from 1 - very low to 5 - very high)</td>
<td>numeric: from 1 to 5</td>
</tr>
<tr>
<td>goout</td>
<td>going out with friends (numeric: from 1 - very low to 5 - very high)</td>
<td>numeric: from 1 to 5</td>
</tr>
<tr>
<td>Dalc</td>
<td>workday alcohol consumption (numeric: from 1 - very low to 5 - very high)</td>
<td>numeric: from 1 to 5</td>
</tr>
<tr>
<td>Walc</td>
<td>weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)</td>
<td>numeric: from 1 to 5</td>
</tr>
<tr>
<td>health</td>
<td>current health status (numeric: from 1 - very bad to 5 - very good)</td>
<td>numeric: from 1 to 5</td>
</tr>
<tr>
<td>absences</td>
<td>number of school absences (numeric: from 0 to 93)</td>
<td>numeric: from 0 to 93</td>
</tr>
</tbody>
</table>

# each student’s Math exam scores in three periods:
1. G1 - first period grade (numeric: from 0 to 20)
2. G2 - second period grade (numeric: from 0 to 20)
3. G3 - third period grade (numeric: from 0 to 20)

1. Please find a model that best predicts the students’ average math score [G=(G1+G2+G3)/3] using the stepwise variable selection method. Please provide model goodness-of-fit index. Please perform all model diagnostics necessary.

2. Please find a model that best predicts the students’ average math score [G=(G1+G2+G3)/3] using the best subset variable selection method. Please provide model goodness-of-fit index. Please perform all model diagnostics necessary.

3. Are the final models in 1 & 2 above, the same? Please discuss any limitations/imperfections your models might have. How can you improve your models?