Biomath 204 - Assignment 1

A study in a fertility clinic was conducted on n=97 non smoking males to examine factors that may influence their semen volume (svol) in ml. The investigators were primarily interested in age (in years), but also considered alcohol use (1=yes or 0=no), whether the subject had abstained from sexual activity for 2 days or more (astn2d-1=yes or 0=no) and whether the subject had hypertension (htn-1=yes or 0=no).

Use the “svol.csv” dataset and regression to determine how age, alcohol use, abstinence and hypertension are simultaneously related to semen volume (Y). You need only consider interactions (if any) with age. Report on which of the variables are related to semen volume and the magnitude and direction of their effect. Be sure to report and interpret the final regression equation and its R square value. Be sure to consider whether the relation of svol with age is linear (Suggestion- in R, can use RCS-restricted cubic spline, or can use fractional polynomials).

You do NOT have to report intermediate models, only the final model. However, in the report you may wish to comment on alternatives.

Make a table to present your “final” model and write a brief report explaining (in words) how age, alcohol use, abstention and/or having hypertension affect semen volume. Be sure to comment on whether each factor is associated with an increase or decrease in semen volume or has no effect on semen volume or if this result must be qualified or is conditional on other factors (ie interpret any significant interactions).

Be sure to label the results with the appropriate units (ml, years ...)

Suggested Table (in the style found in biomedical journals)

<table>
<thead>
<tr>
<th>Table Title</th>
<th>n=</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable</td>
<td>estimate</td>
</tr>
</tbody>
</table>

R square =