1. In 1928 Karl Mannheim argued that people are most impressionistic during their late teens and early twenties. Subsequent research has shown that the age at which people are sensitive to (or impressed by) events differs according to the nature of the events. For example, one body of research has shown that it is around age 19 when people are most sensitive to (and thus more likely to remember) historical events (such as the bombing of Hiroshima).

You are using data collected by the National Opinion Research Council in 1974. Your sample consists of 123 Americans, who were asked (among other things),

"How vividly do you remember the bombing of Hiroshima?"

You name this variable "recall". (Possible values on this variable are 1=do not remember the Hiroshima bombing, 2=barely at all, 3=somewhat vividly, 4=very vividly, and 5=extremely vividly.) Note: You should assume that this variable is a ratio-level measure.

Since you also have data on each respondent's year of birth (yob) and you know that Hiroshima was bombed in 1945, you can calculate each respondents' age at the time of Hiroshima (hiroage) as follows:

\[ \text{compute } \text{hiroage} = 1945 - \text{yob}. \]

a. What complete and reduced regression models (i.e., regression equations) would you use to evaluate whether Americans are more likely to have recalled the bombing of Hiroshima if they were 19 years old in 1945, than if they were older or younger than 19 at the time. (Do NOT attempt to find parameter estimates [i.e., no numbers please].) What does Mannheim's theory suggest will be the signs of all slopes in the complete model. (Explain your answer. Keep in mind that Mannheim's theory implies nothing about the marginal effect of "hiroage" on "recall".)

b. The computer output from the two regressions described in part a, contains the following information:

<table>
<thead>
<tr>
<th>Model Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

For the reduced model:
For the complete model:

<table>
<thead>
<tr>
<th>Model</th>
<th>R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.19836</td>
</tr>
</tbody>
</table>

Which regression model provides the most parsimonious depiction of Americans' recall of the Hiroshima bombing? (Use the .05 level of significance.)

c. After performing the above analysis, you learn that older Americans are more likely to recall historical events than younger ones. To evaluate these "maturity effects" you must create a new variable that measures age. Give SPSS, R, or SAS code that would create a variable (age) that measures each respondent's age at the time of the survey. What is the correlation between "age" (the variable you have just created) and "hiroage"? (Explain your answer. Do NOT provide a mathematical proof.)

d. How might you adjust "recall" such that Americans' recall of the Hiroshima bombing no longer varies according to the linear effects of the variable, "age"? (Hint: No need for numbers here. Just explain how you would go about doing this.)

e. A colleague suggests that you regress "recall" on both "hiroage" (to detect the linear effect of subjects' ages at the time of the bombing of Hiroshima) and "age" (to detect the maturity effects described above). Explain why this would NOT be a good idea.

2. Consider the following three variables from a sample of 126 ISU students:

ANTICIPATE (A) = "joy in knowing STAT 401 is almost at an end" on a scale from 0 = no joy to 100 = extreme joy

NEAREND (N) = "number of STAT 401 classes left to be attended" on a scale from 0 = no more classes to attend to 45 = forty-five more classes left to be attended

REQUIRED (R) = "if STAT 401 was taken as a required course" on a scale from 0 = not required to 1 = required

Some data: \(r_{AN} = -.22\) \(r_{AR} = -.58\) \(r_{NR} = .60\)

Some regression models:

Model a: \(\hat{A_a} = \hat{a_a} + \hat{b_{a1}}N\)

Model b: \(\hat{A_b} = \hat{a_b} + \hat{b_{b1}}R\)
Model c:  \[ \hat{A}_c = \hat{a}_c + \hat{b}_{c1}(R - \bar{R})^2 \]
Model d:  \[ \hat{A}_d = \hat{a}_d + \hat{b}_{d1}(N - \bar{N})^2 \]
Model e:  \[ \hat{A}_e = \hat{a}_e + \hat{b}_{e1}(R - \bar{R})(N - \bar{N}) \]
Model f:  \[ \hat{A}_f = \hat{a}_f + \hat{b}_{f1}N + \hat{b}_{f2}R \]
Model g:  \[ \hat{A}_g = \hat{a}_g + \hat{b}_{g1}N + \hat{b}_{g2}(N - \bar{N})^2 \]
Model h:  \[ \hat{A}_h = \hat{a}_h + \hat{b}_{h1}N + \hat{b}_{h2}(R - \bar{R})^2 \]
Model i:  \[ \hat{A}_i = \hat{a}_i + \hat{b}_{i1}R + \hat{b}_{i2}(N - \bar{N})^2 \]
Model j:  \[ \hat{A}_j = \hat{a}_j + \hat{b}_{j1}R + \hat{b}_{j2}(R - \bar{R})^2 \]
Model k:  \[ \hat{A}_k = \hat{a}_k + \hat{b}_{k1}R + \hat{b}_{k2}(R - \bar{R})(N - \bar{N}) \]
Model l:  \[ \hat{A}_l = \hat{a}_l + \hat{b}_{l1}N + \hat{b}_{l2}(R - \bar{R})(N - \bar{N}) \]
Model m:  \[ \hat{A}_m = \hat{a}_m + \hat{b}_{m1}R + \hat{b}_{m2}N + \hat{b}_{m3}(R - \bar{R})^2 \]
Model n:  \[ \hat{A}_n = \hat{a}_n + \hat{b}_{n1}R + \hat{b}_{n2}N + \hat{b}_{n3}(N - \bar{N})^2 \]
Model o:  \[ \hat{A}_o = \hat{a}_o + \hat{b}_{o1}R + \hat{b}_{o2}N + \hat{b}_{o3}(R - \bar{R})(N - \bar{N}) \]
Model p:  \[ \hat{A}_p = \hat{a}_p + \hat{b}_{p1}R + \hat{b}_{p2}N + \hat{b}_{p3}(N - \bar{N})^2 + \hat{b}_{p4}(R - \bar{R})(N - \bar{N}) \]
Model q:  \[ \hat{A}_q = \hat{a}_q + \hat{b}_{q1}R + \hat{b}_{q2}N + \hat{b}_{q3}(R - \bar{R})^2 + \hat{b}_{q4}(R - \bar{R})(N - \bar{N}) \]
Model r:  \[ \hat{A}_r = \hat{a}_r + \hat{b}_{r1}R + \hat{b}_{r2}N + \hat{b}_{r3}(N - \bar{N})^2 + \hat{b}_{r4}(R - \bar{R})^2 + \hat{b}_{r5}(R - \bar{R})(N - \bar{N}) \]

a. Indicate which of the models (i.e., Model a through Model r [or none of these]) would be used as the complete and reduced (if applicable) models in testing each of the following research questions:

<table>
<thead>
<tr>
<th>Model(s) used</th>
<th>Complete</th>
<th>Reduced</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Does joy increase as each STAT 401 class is finished (and thus as there is one fewer STAT 401 class left to be attended)?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Does joy increase at a faster pace as there are fewer and fewer STAT 401 classes to be attended?</td>
</tr>
</tbody>
</table>
In comparison with students who take STAT 401 as an elective (i.e., not as a required course), do students who take STAT 401 as a required course exhibit a greater increase in joy for each remaining STAT 401 class to be attended?

(Hint: One of these three research questions does not call for a reduced model.)

b. Selecting from among Model a through Model r, a t-test would be preferable to which one of these regression models?

c. Seven of the regression models are misspecified (i.e., ill-formulated). Which are these?

d. Each of the seven models (referred to in part c) is misspecified for the same reason. What reason is this?

e. Referring to Model f, assume that $\hat{a}_f = 80$ and that $\hat{b}_{f1} = 1$. Give an interpretation of what each of these estimates means in words.

3. You are doing research on how well people manage their money. Some research has shown that drug abusers manage their money worse than people who do not abuse drugs. The central theoretical argument in this research is that drugs impair people's judgment and make them less capable of making good decisions regarding their families' finances. As a social worker, you find yourself working with 21 drug-free families (i.e., families in which no member abuses drugs), 11 crack families (i.e., families in which at least one member is a regular user of crack-cocaine), 5 heroin families (i.e., families in which at least one member is a regular user of heroin), and 3 amphetamine families (i.e., families in which at least one member is a regular user of amphetamines). You randomly sample 3 families from each of these four groups of families and record information on their debts. The following are data on the debt (in hundreds of dollars) of each of these 12 families:

<table>
<thead>
<tr>
<th>Drug-free</th>
<th>Crack</th>
<th>Heroin</th>
<th>Amphetamine</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>10</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: $\sum \sum (Y_{ij} - Y_{..})^2 = 120$
a. Construct an analysis of variance table that includes the Between- and Within-Group sums of squares, as well as the degrees of freedom and mean squares associated with each.

b. Test the null hypothesis that there is no significant difference in debt among the four types of families. (Use the .05 level of significance.)

c. If you were to analyze the debt data using dummy variables, how many dummy variables would you need? What values would these dummy variables take and under what conditions would they take each of these values? (Note: What is being asked is NOT the values of coefficients associated with the dummy variables, but of numbers assigned to the attributes of the dummy variables. I.e., how would you code the dummy variables?)

d. Using dummy variables give a regression equation for analyzing the debt data. Compute the numerical values of each coefficient (including the constant) for this regression equation. Show your work.

e. Say in words the meaning of each of the coefficients (including the constant) estimated in part d. (Be sure to include the appropriate numerical values in your renderings of these meanings.)