Marketing clothing in Malaysia is challenging. Given the country’s ethnic diversity, clothiers must be culturally sensitive in what they market. Half of the country’s population consists of Malays (i.e., indigenous Moslems), for whom clothes are body and hair coverings. Nearly a quarter of the remaining citizens are of Chinese origin—a group heavily involved in business, and who typically wear more Western fashions. Then there are the Indians—7% of the population that tends to dress almost exclusively in its own ethnic dress (e.g., the sari, for women). As an entrepreneur in scarves, you are interested in marketing your head-coverings in ways that appeal to the broadest spectrum of Malaysia’s citizens.

After securing a contract to sell your scarves at the country’s five Metrojaya department stores, you administer on-site surveys to a stratified random sample of 15 Malay, 15 ethnic Chinese, and 15 ethnic Indian women who purchased your scarves in these stores last week. Your hypothesis is that the less affluent a shopper is, the more likely she will buy scarves that match her ethnicity. You ask the 45 women in your sample to indicate their annual income and their ethnicity. In addition, you measure the extent to which the scarf (or scarves) that each woman purchased matches her ethnicity. Accordingly, your data are as follows:

INCOME (I): The shopper’s annual income (in thousands of Malaysian Ringgits, where $1 [one US dollar] equals about 3 Ringgits)

ETHNICITY (E): 0=Malay, 1=ethnic Chinese, 2=ethnic Indian

MATCH (M): The extent to which the scarf (or scarves) purchased by a shopper match her ethnicity, on a range from 1=“no match” to 100=“perfect match”

a. Which of the above three variables would be the appropriate one to use as the dependent variable in a regression model for testing your hypothesis? [weight 1]

MATCH (or M)
b. After plotting your data you notice that your hypothesis is only true among shoppers whose annual incomes were above 60,000 Ringgits. When shoppers had incomes less than this, they consistently bought scarves matching their ethnicity. In the space provided below, sketch the just-described pattern of data that you found in your plot. (Hint: Be sure to specify which variables are dependent [from part a] and independent.) [weight 2]

![Graph showing the relationship between income and scarf matching]

Dependent Variable: \( M \)  
Independent Variable: \( I \)

\[ \hat{M} = \hat{a} + \hat{b}_1 I + \hat{b}_2 (I - \bar{I})^2 \]

c. What (complete) regression equation would you use to model the pattern of data sketched in part b? (Hint: A new variable [i.e., one other than I, E, and M] must be constructed for this equation. Be sure to show how you would calculate values on this new variable.) [weight 2]

\[ \hat{M} = \hat{a} + \hat{b}_1 I + \hat{b}_2 (I - \bar{I})^2 \]

d. Assuming that the median annual income among your sample of 45 shoppers equals 60,000 Ringgits and that your sketch in part b accurately reflects the distribution of your data, what would be the signs of all slopes in the regression model in part c? [weight 2]

\[ \hat{b}_1 < 0 \] (overall negative)  
\[ \hat{b}_2 < 0 \] (frown)
e. The next step in your analysis is to evaluate whether or not shoppers of any ethnicity (e.g., Malay, Chinese, or Indian) are more likely than those of another ethnicity to purchase scarves matching their ethnicity. Although the overall average score on MATCH is 50 (and its overall variance equals 100), the average MATCH score was 60 among Malays, 40 among ethnic Chinese, and 45 among ethnic Indians. At the .05 significance level, test the null hypothesis that shoppers’ average MATCH scores (namely, 60, 40, and 45) are the same among the three ethnicities. Are the three means statistically different? Explain your answer.

\[ SS_{TOTAL} = (n-1)\hat{\sigma}_M^2 (45 - 1)100 = 4,400 = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (M_{ij} - \mu_{..})^2 \]

Overall: \[ \mu_{..} = 50 \]

Malay: \[ \mu_{1..} = 60 \]

Chinese: \[ \mu_{2..} = 40 \]

Indian: \[ \mu_{3..} = 45 \]

\[ H_0 : \mu_1 = \mu_2 = \mu_3 \]

\[ H_A : \mu_i = \mu_j \text{ for some } i \neq j \]

\[ \alpha = .05 \]

\[ SS_{BETWEEN} = \sum_{i=1}^{m} n_i (M_{i..} - \mu_{..})^2 = 15(60 - 50)^2 + 15(40 - 50)^2 + 15(45 - 50)^2 \]

\[ = 1,500 + 1,500 + 375 = 3,375 \]

\[ SS_{WITHIN} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (M_{ij} - M_{i..})^2 = SS_{TOTAL} - SS_{BETWEEN} = 4,400 - 3,375 = 1,075 \]

\[ F_{m-1, n-m} = \frac{SS_{BETWEEN} / (m-1)}{SS_{WITHIN} / (n-m)} = \frac{3,375 / (3-1)}{1,075 / (45-3)} = \frac{1,687.50}{24.40} = 69.15 \]

Since \[ F_{42,.05}^2 < F_{40,.05}^2 = 3.23 < 69.15 \], Ho is rejected.

The means are statistically different at the .05 level.
At this point you drop the 15 ethnic Indian shoppers from your analysis, leaving you with a sample of 30 Malaysian shoppers for the remainder of the exam. Note that the ETHNICITY (E) variable is now a dummy variable for which 0=Malay and 1=ethnic Chinese. The following output will be useful in completing the remaining parts of this exam. It is SPSS output from the regression of MATCH on INCOME and ETHNICITY:

### Regression

#### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.540</td>
<td>.292</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), I, E

#### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
</tr>
<tr>
<td>I</td>
<td>-1.222</td>
</tr>
<tr>
<td>E</td>
<td>-9.045</td>
</tr>
</tbody>
</table>

a. Dependent Variable: M

f. Using the .05 significance level and the above output, test the (alternative) hypothesis that after adjusting shoppers’ incomes as if all shoppers were Malay, the extent to which affluent shoppers’ scarves match their ethnicity is less than it is for nonaffluent shoppers. [weight 4]

\[
\begin{align*}
H_0 : & b_I = 0 \\
H_0 : & b_I < 0
\end{align*}
\]

\[
\begin{align*}
t_{n-k-1} = & \frac{\hat{b}_I - b_0}{\hat{\sigma}_{\hat{b}_I}} = \frac{-1.222 - 0}{.441} = -2.771 = t_{30-2-1}
\end{align*}
\]

Since \(-t_{27,.05} = -1.703 > -2.771\), Ho is rejected.

After adjusting shoppers’ incomes for (what I am assuming here is) the greater affluence of ethnic Chinese shoppers relative to Malay shoppers, there is a significant decline (at the .05 level) in the ethnicity-scarf match among shoppers with increasingly higher incomes.
g. What is the $P$-value associated with the finding in part f? [weight 2]

\[ \Pr(t_{27} < -2.771) = \Pr(t_{27} > 2.771) = .005 \]

h. Part f refers to “adjusting shoppers’ incomes as if all shoppers were Malay.” Explain how to create a new variable, INCADJ, that is shoppers’ incomes adjusted as if all 30 shoppers were Malay (i.e., such that Chinese shoppers’ incomes are reduced by the amount of income they earn more than Malay shoppers). (Hint: There is no need to calculate any numbers here. Merely explain the strategy you would use to make this adjustment.) [weight 1]

First, regress $I$ on $E$ and obtain the slope, $\hat{b}_E$, from the following regression model:

\[ \hat{I} = \hat{\alpha} + \hat{b}_E E \]

Second, compute $I_{adj}$ as follows:

\[ I_{adj} = I - \hat{b}_E E \]
i. In the space provided below, sketch how M, I, and E would be related among the 30 shoppers if there is a greater tendency among Chinese-origin than among Malay shoppers for affluence to undermine their purchases of ethnicity-matching scarves. (Hint: Again be sure to specify which variables are dependent and independent.) [weight 2]

![Graph showing relationship between M, I, and E]

j. What complete and reduced regression models would you use to test whether or not your data on the 30 Malay and Chinese-origin shoppers correspond to (i.e., fit the pattern drawn in) the sketch made in part i? (Hint: Answering this question requires the construction of a new variable, which you may call NEWVAR [or N]. Be sure to show how this variable was constructed.) [weight 2]

New variable: \( N = (I - \bar{I}) \times (E - \bar{E}) \)

Complete: \( \hat{M} = \hat{a} + \hat{b}_1 I + \hat{b}_2 E + \hat{b}_3 N \)

Reduced: \( \hat{M}' = \hat{a}' + \hat{b}_1' I + \hat{b}_2' E \)
k. A constructed variable was used as an independent variable in the complete model described in part j. What would be the sign of the slope associated with this constructed variable if your data on the 30 Malay and Chinese-origin shoppers correspond to the sketch made in part i? Show how you obtained your answer. [weight 2]

If my data fit the theory, values from the reduced model (in part j) would plot as two parallel lines as follows:

![Diagram showing two parallel lines with arrows indicating shifts.]

Arrows in the sketch indicate how these lines would shift if the complete model were estimated. Accordingly, the theory table would look as follows:

<table>
<thead>
<tr>
<th></th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 (Malay)</td>
</tr>
<tr>
<td>affluent</td>
<td>+</td>
</tr>
<tr>
<td>nonaffluent</td>
<td>–</td>
</tr>
</tbody>
</table>

Given that \( N = (I - \bar{I}) \times (E - \bar{E}) \), the signs of corresponding values on this variable would be as follows:

<table>
<thead>
<tr>
<th></th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 (Malay)</td>
</tr>
<tr>
<td>affluent</td>
<td>–</td>
</tr>
<tr>
<td>nonaffluent</td>
<td>+</td>
</tr>
</tbody>
</table>

Since the signs in the theory and measure tables are opposites, the sign of \( \hat{b}_3 \) would be negative if my data correspond to the sketch in part i.