Midterm Two

Instructions: This is a closed-book and closed-notes exam. You may use 3 two-sided help sheets of letter size. You may use a calculator (cell-phones may NOT be used as a calculator). The total score is 25 points. **Give concise but detailed answers for full credit.**

1. A survey was taken to see if a person’s purchases based on infomercials on television differed by the level of several different factors. One study considered the two factors “household income” and “marital status.” Household income was categorized into four categories: (1) under $30,000, (2) $30,000- $50,000, (3) $50,000- $100,000, and (4) over $100,000. Martial status was categorized into three levels: A, Single; B, married; and C, divorced/separated/widowed. For each of the 12 cells, 10 people were surveyed and reported their estimated past purchases (in dollars) that were based on infomercials on television. The data are summarized as follows (cell mean and the standard deviation, SD, of the 10 replications):

<table>
<thead>
<tr>
<th>Martial Status (MS)</th>
<th>Household Income (HI)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean = 56</td>
<td>Mean = 73</td>
<td>Mean = 64</td>
<td>Mean = 62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD = 33</td>
<td>SD = 39</td>
<td>SD = 45</td>
<td>SD = 44</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>63.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70.50</td>
<td></td>
<td></td>
<td>76.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>56.50</td>
</tr>
<tr>
<td>B</td>
<td>Mean = 63</td>
<td>Mean = 80</td>
<td>Mean = 74</td>
<td>Mean = 65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD = 35</td>
<td>SD = 23</td>
<td>SD = 40</td>
<td>SD = 20</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Mean = 68</td>
<td>Mean = 81</td>
<td>Mean = 85</td>
<td>Mean = 73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD = 41</td>
<td>SD = 32</td>
<td>SD = 19</td>
<td>SD = 27</td>
<td></td>
</tr>
</tbody>
</table>

a. [3 points] What is this design? Identify the response variable, factors and the type of each factor, fixed or random. Also identify blocking factor if there is any.

Answer: 3x4 factorial design (2-way Anova model) with 10 replicates response var. = person’s purchase factors = MS. and H.I. ; both are fixed effect (No blocking factors)
b. [5 points] Fill up all *'s in the ANOVA table below (MS denotes Martial Status and HI denotes Household Income):

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>2</td>
<td>3381.67</td>
<td>1690.83</td>
<td>1.44</td>
</tr>
<tr>
<td>HI</td>
<td>3</td>
<td>4566.67</td>
<td>1522.22</td>
<td>1.29</td>
</tr>
<tr>
<td>MS*HI</td>
<td>6</td>
<td>578.33</td>
<td>96.39</td>
<td>0.082</td>
</tr>
<tr>
<td>Error</td>
<td>*</td>
<td>127080</td>
<td>1176.67</td>
<td></td>
</tr>
</tbody>
</table>

\[
\text{Answer: } \quad \text{DF}_{\text{Error}} = \text{DF}_{\text{Total}} - (2+3+6) = (12 \times 10 - 1) - 11 = 108
\]

\[
\text{SS}_{\text{MS}} = 10 \times 4 \sum_{i=1}^{3} \left( \overline{y}_{i..} - \overline{y}_{..} \right)^2,
\]

where the row means \( \overline{y}_{i..} \) are

\[
\overline{y}_{1..} = \frac{56 + 73 + 64 + 62}{4} = 63.75
\]

\[
\overline{y}_{2..} = \frac{63 + 80 + 74 + 65}{4} = 70.50
\]

\[
\overline{y}_{3..} = \frac{68 + 81 + 85 + 73}{4} = 76.75
\]

\[
\Rightarrow \overline{y}_{..} = \frac{\overline{y}_{1..} + \overline{y}_{2..} + \overline{y}_{3..}}{3} = 70.33
\]

\[
\Rightarrow \text{SS}_{\text{MS}} = 3381.67
\]

\[
\text{MS} = \frac{\text{SS}}{\text{DF}} \quad \text{and} \quad F = \frac{\text{MS}_{\text{MS}}}{\text{MS}_{\text{Error}}} \quad \text{(fixed-effect model)}
\]

c. [3 points] Make a profile plot (also called interaction plot) with Household Income on the horizontal axis. From the plot, do you see a significant interaction effect between Household Income and Martial Status? Explain your answer.

Answer:

The interaction effect seems not significant as the 3 lines are somewhat parallel.
d. Do the following problems.
   i. [1 point] By the Anova table in part b, is the interaction term significant at 5% level?
   
   ii. [2 points] Remove the interaction term and write the Anova table of the reduced model.
   
   iii. [1 point] Based on the reduced model, are the main effects of Martial Status and Household Income significant at 5% level? Show your work.

Answer:

i. \[ F = 0.082 < 1 \Rightarrow \text{Not significant} \]
   
   \[ F_{6,108;5\%} = 2.18 \text{ from Table.} \]

ii. \[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Source} & \text{DF} & \text{SS} & \text{MS} = \frac{\text{SS}}{\text{DF}} & F = \frac{\text{MS}_{\text{X}}}{\text{MSE}} \\
\hline
\text{MS} & 2 & 3381.67 & 1690.84 & 1.51 \\
\text{HI} & 3 & 4566.67 & 1522.22 & 1.36 \\
\text{Error} & 108+6 = 114 & 578335+127080 = 1119.81 & 127658.335 & \\
\hline
\end{array}
\]

iii. MS: \[ F = 1.51 < F_{2,114;5\%} = 3.07 \Rightarrow \text{The main effect of MS is not significant at 5\% level.} \]

H.I.: \[ F = 1.36 < F_{3,114;5\%} = 2.68 \Rightarrow \text{The main effect of H.I. is not significant at 5\% level.} \]
2. We wish to determine the amount of salt to put in a microwave popcorn bag so that it has the best overall acceptability. We will test 3 levels of salt: low, medium, and high. We have randomly recruited 25 volunteers to taste popcor ns, and while we expect the individuals to be reasonably consistent in their own personal ratings, we expect large volunteer to volunteer differences in overall ratings.
   a. [2 points] What design would you choose? Why?
   b. [3 points] Give the degrees of freedom for all the various terms in your model and indicate the treatment factors and the blocking factors if any. Also identify the type of each factor, fixed or random.

Answer:

a. Randomized block design. Treatment factor is the salt levels and block factor is the 25 volunteers. We must block volunteers as we expect large volunteer to volunteer differences. In addition, we don’t expect salt levels interact with volunteers.

b. | Source        | DF  |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Salt</td>
<td>3-1 = 2</td>
</tr>
<tr>
<td>Volunteer Error</td>
<td>25-1 = 24</td>
</tr>
<tr>
<td>Error</td>
<td>2 \times 24 = 48</td>
</tr>
</tbody>
</table>
3. We wish to study the effect of two food additives (plus a control for a total of three treatments) on the milk productivity of cows. We have 3 large herds available, each of a different breed, and we expect an age effect, which we make explicit by dividing cows into 3 groups: those have had 0, 1, and 2 or more previous calves. We have enough resources to study 18 animals through one breeding cycle.

   a. [2 points] What are the experimental units, response variable, treatment factors and blocking factors if any?
   b. (3503/3602 only) [3 points] Briefly describe the experimental design you would choose and explain why you choose the design.
   c. (6305 only) [3 points] Briefly describe the experimental design you would choose and treatment assignments to the experimental units. Explain why you choose the design.

   a. Units = cows.
   response var. = milk productivity
   treatment factor = food additives (plus a control)
   blocking factors = breeds and age

   b. Latin square design with 2 replicates, since the 3 factors all have 3 levels and 18 cows are not enough to run a full factorial design (3x3x3=27)

   ① breeds and age are unlikely to interact with food additives and so both can be treated as block factors
   ② a Latin square design needs only 3x3=9, so 18 cows can do two Latin squares.

   c. Since breeds and age are fixed levels, of

   \[
   \begin{array}{c|c|c}
   \text{breed} & 0 & 1 & 2+ \\
   \hline
   1 & B & C & A \\
   2 & C & A & B \\
   3 & A & B & C \\
   \end{array}
   \quad \begin{array}{c|c|c}
   \text{Age} & 0 & 1 & 2+ \\
   \hline
   1 & B & C & A \\
   2 & C & A & B \\
   3 & A & B & C \\
   \end{array}
   \]

   A: food additive 1
   B: " " 2
   C: control