An old exam problem on two-factor fixed effect ANOVA

Use SAS to generate output to answer the questions below:

1. A study was performed to look at systolic blood pressure (BP) in three different diet groups:
   - strict vegetarians (SV)
   - lactovegetarians (LV) (eat dairy products)
   - normal - (standard American diet)

   Samples of size $n = 5$ from each gender were taken from each dietary group and BP was measured. Total sample size is $N = 30$. Let $y_{ijk}$ denote the BP for subject $k$ in diet group $i$ and gender $j$.

   (a) Estimate the standard deviation of BPs among people in a given dietary-gender classification group, assuming homogeneity of variance across the 6 groups.

   ```sas
   proc means data=one nway;
   class gender diet;
   var bp;
   output out=variances var=s2;
   run;
   proc print data=variances; title "variances"; run;
   proc means data=variances;
   var s2;
   run;
   ```

   (b) **True**/false: The $MS[E]$ is the average of the sample variances for the 6 treatment combinations:

   $$MS[E] = \frac{1}{6}(s_{11}^2 + s_{12}^2 + s_{21}^2 + s_{22}^2 + s_{31}^2 + s_{32}^2)$$

   where $s_{ij}^2 = \frac{1}{4} \sum_{k=1}^{5} (y_{ijk} - \bar{y}_{ij})^2$.

   (c) Test for a diet×gender interaction effect at level $\alpha = 0.05$.

   ```sas
   proc glm data=one;
   class gender diet;
   model bp=gender diet gender*diet;
   run;
   ```

   (d) Report a $p$-value for the null hypothesis that the average BPs in the three diet populations are equal. **See output from code above.**

   (e) Estimate the difference between mean BP for men and women. Report a standard error. The following statement in PROC GLM should do it:

   ```sas
   estimate "men - women" gender -1 1;
   ```

   (f) **True**/False: There is significant ($\alpha = 0.05$) evidence to suggest that population differences among diet groups are different for men than for women. This question may be addressed using the test for interaction.
(g) Let $\mu_{NOR}, \mu_{LV}, \mu_{SV}$ denote the mean BPs in the three populations after averaging over gender. Prepare a table of estimates, with estimated standard errors in parentheses:

\[
\begin{align*}
\hat{\mu}_{NOR} &= ( ) \\
\hat{\mu}_{SV} &= ( ) \\
\hat{\mu}_{LV} &= ( )
\end{align*}
\]

\texttt{lsmeans gender/stderr;}

(h) Suppose you are interested in $c = 4$ contrasts: the 3 pairwise contrasts and the difference $\theta$ between mean BP for normal eaters and the average of the mean BPs for the two vegetarian populations. Use Scheffé’s procedure to obtain simultaneous 95% confidence intervals for these four contrasts. Note that

\[
\sqrt{(3-1)F(0.95, 2, 24)MS[E]} \left( \frac{2}{10} \right) = 10.1
\]

\[
\sqrt{(3-1)F(0.95, 2, 24)MS[E]} = 22.5
\]

\[
\sqrt{\frac{1}{10} + \frac{1}{4} + \frac{1}{4}} = 0.4
\]

\[
\mu_{NOR} - \mu_{SV} : 11.0 \pm 10.1
\]

\[
\mu_{NOR} - \mu_{LV} : \pm
\]

\[
\mu_{LV} - \mu_{SV} : \pm
\]

\[
\theta : \pm
\]

A little bit of computation by hand might be necessary, but all of the ingredients can be generated with the \texttt{scheffe} option:

\texttt{means diet/scheffe;}

(i) Label each sum of squares below so that the source of variability which it quantifies is clear

\[ SS[\text{diet}], SS[\text{gender}], SS[\text{diet} \times \text{gender}], SS[E], SS[\text{Tot}] \]

<table>
<thead>
<tr>
<th>Sum of squares</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{5} (y_{ijk} - \bar{y}_{ij})^2 ]</td>
<td></td>
</tr>
<tr>
<td>[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{5} (\bar{y}<em>{i++} - \bar{y}</em>{+++})^2 ]</td>
<td></td>
</tr>
<tr>
<td>[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{5} (\bar{y}<em>{ij+} - \bar{y}</em>{i++} - \bar{y}<em>{j+} + \bar{y}</em>{+++})^2 ]</td>
<td></td>
</tr>
<tr>
<td>[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{5} (\bar{y}<em>{ij} - \bar{y}</em>{+++})^2 ]</td>
<td></td>
</tr>
<tr>
<td>[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{5} (y_{ijk} - \bar{y}_{+++})^2 ]</td>
<td></td>
</tr>
</tbody>
</table>

(j) Complete the interaction plot below, using “M” for male, “F” for female.

```{r}
output out=meanz p=bpmean;
symbol value=dot i=join;
proc gplot data=meanz;
plot bpmean*diet=gender;
run;
```