Topic 12: Split-plot design and its relatives

Due Thursday, March 5, at the beginning of lab. Include your SAS codes, answer all parts of the questions completely, and interpret all results. To ensure maximum points for yourself, invest some time in presenting your answers in a concise, organized, and clear manner.

Question 1 [50 points]

A baker would like to study how the rise of baked bread is affected by three oven temperature, two yeast strains, and two different kneading times. To perform the experiment the baker chose 3 ovens in the kitchen. On day one the baker randomly assigned an oven temperature of 320°F, 335°F, and 350°F to each oven. The baker prepared four doughs. Each dough was prepared using one of four different recipes, which consisted of different combinations of one of two yeasts strains (yeast 1 or yeast 2) and one of two knead times (10 or 5 minutes). The 4 dough types were split into three portions and one of each dough type was placed to cook on the three ovens. The order in which each dough type was positioned in the oven was completely random. After the dough was cooked the baker took out the bread and measured the volume of each bread in cm² (i.e. rise of bread). The baker repeated this for two more days. Days are considered as a blocking factor. The data collected is summarized in the table below:

```
data hw12_1_2;
Do Day = 1 to 3;
   Do Temp = '320', '335', '350';
      Do trtmt = 'Y110', 'Y210', 'Y15', 'Y25';
         Input volume @@;
         Output;
      End;
   End;
End;
Cards;
1138  1143  1176  1170
1415  1435  1335  1399
1217  1161  1226  1294
1220  1203  1172  1198
1442  1412  1348  1409
1204  1288  1317  1305
1128  1110  1117  1147
1293  1271  1185  1316
1174  1170  1231  1277
;
Proc print;
Proc GLM data = hw12_1_2;
Class Day Temp trtmt;
Model volume = Day Temp Day*Temp
               trtmt trtmt*Temp;
Random Day Day*Temp / test;
Proc Sort data = hw12_1_2;
by Temp;
Proc GLM data = hw12_1_2;
Class Day trtmt;
Model volume = Day trtmt;
Means trtmt;
```
contrast 'Y1 vs Y2' trtmt 1 1 -1 -1;
contrast 'K10 vs K5' trtmt 1 -1 1 -1;
contrast 'Y by K int.' trtmt 1 -1 -1 1;
by Temp;
Proc Sort data = hw12_1_2;
by trtmt;
Proc GLM data = hw12_1_2;
Class Day Temp;
Model volume = Day Temp;
Means temp;
contrast 'linear' Temp -1 0 1;
contrast 'quadratic' Temp 1 -2 1;
by trtmt;
run; quit;

1.1 Describe in detail the design of this experiment please specify the experimental unit for the main plot and the experimental unit for the sub-plot(s) [see appendix at the end of this problem set].

<table>
<thead>
<tr>
<th>Class Variable</th>
<th>Block or Treatment</th>
<th>Number of Levels</th>
<th>Fixed or Random</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Block</td>
<td>3</td>
<td>Random</td>
<td>Days</td>
</tr>
<tr>
<td>2</td>
<td>Treatment</td>
<td>3</td>
<td>Fixed</td>
<td>Oven temperatures</td>
</tr>
<tr>
<td>3</td>
<td>Treatment</td>
<td>4</td>
<td>Fixed</td>
<td>Four recipes</td>
</tr>
</tbody>
</table>

Subsamples? NO

1.2 Perform the appropriate ANOVA and describe if there are significant interactions between temperature and recipe.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>2</td>
<td>51422</td>
<td>25711</td>
<td>6.28</td>
<td>0.0583</td>
</tr>
<tr>
<td>* Temp</td>
<td>2</td>
<td>230623</td>
<td>115311</td>
<td>28.18</td>
<td>0.0044</td>
</tr>
<tr>
<td>Error: MS(Day*Temp)</td>
<td>4</td>
<td>16369</td>
<td>4092.152778</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* This test assumes one or more other fixed effects are zero.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day*Temp</td>
<td>4</td>
<td>16369</td>
<td>4092.152778</td>
<td>6.20</td>
<td>0.0026</td>
</tr>
<tr>
<td>* trtmt</td>
<td>3</td>
<td>10417</td>
<td>3472.407407</td>
<td>5.26</td>
<td>0.0088</td>
</tr>
<tr>
<td>Temp*trtmt</td>
<td>6</td>
<td>25382</td>
<td>4230.407407</td>
<td>6.41</td>
<td>0.0010</td>
</tr>
<tr>
<td>Error: MS(Error)</td>
<td>18</td>
<td>11878</td>
<td>659.907407</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* This test assumes one or more other fixed effects are zero.

The ANOVA results suggest that there are significant interactions between temperature and the recipe ($P = 0.0010$).
1.3 For each recipe, test if bread rise has a linear and/or a quadratic component with increasing temperature.

<table>
<thead>
<tr>
<th>Contrast</th>
<th>DF</th>
<th>Contrast SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td>1</td>
<td>1980.16667</td>
<td>1980.16667</td>
<td>1.30</td>
<td>0.3176</td>
</tr>
<tr>
<td>quadratic</td>
<td>1</td>
<td>82553.38889</td>
<td>82553.38889</td>
<td>54.25</td>
<td>0.0018</td>
</tr>
</tbody>
</table>

When the dough was prepared using yeast 1 and 10 minutes of kneading the rise in bread had no linear component but did have a quadratic component ($P = 0.3176$ and $P = 0.0018$, respectively).

![Bread Rise Y1 with 10 min of Kneading]

When the dough was prepared using yeast 1 and 5 minutes of kneading the rise in bread had a linear component and very close to having a significant quadratic component ($P = 0.0462$ and $P = 0.0570$).
When the dough was prepared using yeast 2 and 10 minutes of kneading the rise in bread had no linear component and a quadratic component ($P = 0.2454$ and $P = 0.0050$, respectively).
When the dough was prepared using yeast 2 and 5 minutes of kneading the rise in bread had a linear component and a significant quadratic component ($P = 0.0020$ and $P = 0.0006$).

![Bread Rise Y2 with 5 min of Kneading](image_url)

1.4 For each temperature test if yeast one is different from yeast 2, if kneading time 10 min is different from kneading time 5 minutes, and if there is a significant interaction between yeast type and kneading time.

<table>
<thead>
<tr>
<th>Temp=320</th>
<th>Contrast</th>
<th>DF</th>
<th>Contrast SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1 vs Y2</td>
<td>1</td>
<td>33.3333333</td>
<td>33.3333333</td>
<td>0.08</td>
<td>0.7810</td>
<td></td>
</tr>
<tr>
<td>K10 vs K5</td>
<td>1</td>
<td>120.3333333</td>
<td>120.3333333</td>
<td>0.31</td>
<td>0.6006</td>
<td></td>
</tr>
<tr>
<td>Y by K int.</td>
<td>1</td>
<td>533.3333333</td>
<td>533.3333333</td>
<td>1.35</td>
<td>0.2890</td>
<td></td>
</tr>
</tbody>
</table>

When the oven temperature was set at 320°F there was no difference in bread rise between the two yeast, the two kneading times, and there was no significant interaction between yeast and kneading time ($P = 0.7810$, $P = 0.6006$, and $P = 0.2890$, respectively).

<table>
<thead>
<tr>
<th>Temp=335</th>
<th>Contrast</th>
<th>DF</th>
<th>Contrast SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1 vs Y2</td>
<td>1</td>
<td>4181.333333</td>
<td>4181.333333</td>
<td>9.47</td>
<td>0.0217</td>
<td></td>
</tr>
<tr>
<td>K10 vs K5</td>
<td>1</td>
<td>6348.000000</td>
<td>6348.000000</td>
<td>14.38</td>
<td>0.0090</td>
<td></td>
</tr>
<tr>
<td>Y by K int.</td>
<td>1</td>
<td>6912.000000</td>
<td>6912.000000</td>
<td>15.66</td>
<td>0.0075</td>
<td></td>
</tr>
</tbody>
</table>

When the oven temperature was set at 335°F there was a significant difference in bread rise between the two yeast, the two kneading times, and there was a significant interaction between yeast and kneading time ($P = 0.0217$, $P = 0.0090$, and $P = 0.0075$, respectively).

<table>
<thead>
<tr>
<th>Temp=350</th>
<th>Contrast</th>
<th>DF</th>
<th>Contrast SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1 vs Y2</td>
<td>1</td>
<td>1323.000000</td>
<td>1323.000000</td>
<td>1.16</td>
<td>0.3236</td>
<td></td>
</tr>
</tbody>
</table>
When the oven temperature was set at 350°F there was no significant difference in bread rise between the two yeast, there was a significant difference between the two kneading times, and there was no significant interaction between yeast and kneading time ($P = 0.0217$, $P = 0.0090$, and $P = 0.0075$, respectively).

1.5 By using the LSD test, test if there is a significant bread rise difference between recipe 4 at 350°F compared to recipe 1 at 335°F? If you were to do this test using the Tukey-Kramer procedure, what would you change in the procedure you did for the LSD test?

To test if there is a significant difference between recipe 4 at 350°F and recipe 1 at 350°F it is necessary to do a mixed comparison. To do this we must calculate a mixed MSE and a mixed t-value.

The mixed MSE can be calculated with the following formula:

$$MSE_{mix} = \frac{(b-1)MSE_B + MSE_A}{b}$$

Where $b =$ levels of trtmt $= 4$, $MSE_B =$ MSE for trtmt $= 659.91$, $MSE_A =$ MSE for temperature $= 4092.2$

$$MSE_{mix} \cdot \frac{(4-1) 659.91 + 4092.2}{4} = 1517.97$$

The mixed t-value can be calculated with the following formula:

$$t_{mix} = \frac{(b-1)t_BMSE_B + t_A MSE_A}{(b-1)MSE_B + MSE_A}$$

Where $b =$ levels of trtmt $= 4$, $MSE_B =$ MSE for trtmt $= 659.91$, $MSE_A =$ MSE for temperature $= 4092.2$, $t_B =$ critical t-value given 18 df, $t_A =$ critical t-value given 4 df.

$$t_{mix} = \frac{(4-1) * 2.1009 * 659.91 + 2.7764 * 4092.2}{(4-1)659.91 + 4092.2} = 2.55615$$

The minimum significant difference between the two treatments in question is:

$$LSD_{0.05} = t_{mix} \sqrt{\frac{2MSE_{mix}}{r}}$$

Where $r =$ number of replications

$$LSD_{0.05} = 2.55615 \sqrt{\frac{2 \cdot 1517.97}{3}} = 81.3154$$

The difference between recipe 4 at 350°F compared to recipe 1 at 335°F

$$Recipe 1 335 - Recipe 4 350 = 1383.33 - 1292 = 91.33$$

$91.33 > 81.3154$; therefore we conclude that there is a significant difference between recipe 4 at 350°F and recipe 1 at 335°F.

If we were to perform a Tukey-Kramer test instead of and LSD test you would have to make two slight adjustments [no points off if you failed to state the second of these]:
a. First, you would use Studentized critical values (\(q_{\alpha}\), e.g. from Table A.8) in Step 2 above.

b. You would drop the \(\sqrt{2}\) from the MSD formula in Step 3 because the table of Studentized critical values has the \(\sqrt{2}\) built in.

1.6 If the initial interaction between temperature and dough recipe would have not been significant and you were trying to test the lineal and quadratic response of temperature. Would this be correct?

No. The correct error term is missing!

```plaintext
Class Day Temp trtmt;
Model volume = Day Temp Day*Temp trtmt trtmt*Temp;
Test h=Temp e=Day*Temp
   contrast 'linear' Temp -1 0 1 / e=Day*Temp;
   contrast 'quadratic' Temp 1 -2 1 / e=Day*Temp;
```

**Question 2** [50 points]

An investigator would like to test the effect of 3 new prescription drugs on the heart rate of humans. The investigator has picked 24 volunteers perform the experiment. One of the three drugs was randomly assigned to each volunteer in such a way that 8 volunteers received the same drug. After taking the medication the heart rate of the volunteers was measured every hour up to four hours. The data is summarized in the table below:

```plaintext
data hw12_2;
Do Rep = 1 to 8;
   Do Drug = 1 to 3;
      Do Time = 1 to 4;
      Input hr @@;
      Output;
   End;
End;
Cards;
72 87 84 81 79 91 86 83 73 81 77 79
79 86 89 85 84 89 96 86 72 69 71 80
73 86 82 77 76 88 87 79 91 95 96 91
72 84 86 72 80 89 90 74 87 87 84 77
69 83 78 69 70 85 82 71 79 80 76 75
75 84 86 80 82 87 90 83 69 69 70 68
63 74 78 71 67 79 82 77 81 75 73 75
71 77 77 71 76 80 83 77 78 77 72 70
;
Proc Print;
Proc GLM data = hw12_2;
Class Rep Drug Time;
Model hr = Drug Drug*Rep Time Drug*Time;
Random Drug*Rep / test;
Proc GPlot;
axis1 offset=(5 pct,5 pct);
axis2 offset = (5 pct,5 pct);
symbol1 i=stdmtj v=none color=BLUE;
```
symbol2 i=std1mtj v=none color=BLACK;
symbol3 i=std1mtj v=none color=GREEN;
symbol3 i=std1mtj v=none color=GREEN;
plot hr * Time = Drug /;

Run;

Data hw12_22;
input rep drug $ t1 t2 t3 t4;
cards;
1 Drug1 72 87 84 81
2 Drug1 79 86 89 85
3 Drug1 73 86 82 77
4 Drug1 72 84 86 72
5 Drug1 69 83 78 69
6 Drug1 75 84 86 80
7 Drug1 63 74 78 71
8 Drug1 71 77 77 71
1 Drug2 79 91 86 83
2 Drug2 84 89 96 86
3 Drug2 76 88 87 79
4 Drug2 80 89 90 74
5 Drug2 70 85 82 71
6 Drug2 82 87 90 83
7 Drug2 67 79 82 77
8 Drug2 76 80 83 77
1 Drug3 73 81 77 79
2 Drug3 72 69 71 80
3 Drug3 91 95 96 91
4 Drug3 87 87 84 77
5 Drug3 79 80 76 75
6 Drug3 69 69 70 68
7 Drug3 81 75 73 75
8 Drug3 78 77 72 70
;
Proc GLM data = hw12_22;
Class drug;
Model t1 t2 t3 t4 = drug / NoUni;
REPEATED time / NoM PRINTE;
REPEATED time polynomial / summary;
REPEATED time Mean(1) / summary;
REPEATED time Helmert / summary;
run; quit;

2.1 Describe in detail the design of this experiment [see appendix at the end of this problem set].

<table>
<thead>
<tr>
<th>Design:</th>
<th>CRD with repeated measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response Variable:</td>
<td>Heart rate</td>
</tr>
<tr>
<td>Experimental Unit:</td>
<td>Volunteer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class Variable</th>
<th>Block or Treatment</th>
<th>Number of Levels</th>
<th>Fixed or Random</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Treatment</td>
<td>3</td>
<td>Fixed</td>
<td>Prescription Drug</td>
</tr>
<tr>
<td>2</td>
<td>Treatment</td>
<td>4</td>
<td>Fixed</td>
<td>Hours</td>
</tr>
</tbody>
</table>

Subsamples? NO
2.2 Analyze the data from this experiment in the following three ways and discuss if the results support that there are significant differences between prescription drugs, time of measurement, and if there are significant interactions between prescription drugs and times [NOTE: if interactions are significant you do not have to look at the simple effects]:

a. Generate an ANOVA table assuming no correlation among the heart rate measurements of any given individual (i.e. an ANOVA with full degrees of freedom).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug</td>
<td>2</td>
<td>346.937500</td>
<td>173.468750</td>
<td>1.38</td>
<td>0.2738</td>
</tr>
<tr>
<td>Error: MS(Rep*Drug)</td>
<td>21</td>
<td>2642.187500</td>
<td>125.818452</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* This test assumes one or more other fixed effects are zero.

b. Generate an ANOVA table assuming perfect correlation (i.e. an ANOVA with conservative degrees of freedom).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep*Drug</td>
<td>21</td>
<td>2642.187500</td>
<td>125.818452</td>
<td>13.74</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>* Time</td>
<td>3</td>
<td>887.708333</td>
<td>295.902778</td>
<td>32.32</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Drug*Time</td>
<td>6</td>
<td>457.979167</td>
<td>76.329861</td>
<td>8.34</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error: MS(Error)</td>
<td>63</td>
<td>576.812500</td>
<td>9.155754</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* This test assumes one or more other fixed effects are zero.

The ANOVA results suggest that there are no significant differences between prescription drugs ($P = 0.2738$). There are also significant differences in heart rate between the different times ($P < 0.0001$). The interaction between prescription drug and time is significant ($P < 0.0001$).

c. Present ANOVA tables for mainplot (Diet) and subplot (Time) effects using the "Repeated" option under Proc GLM. Present the Sphericity test result and comment on its meaning.

Sphericity Tests

<table>
<thead>
<tr>
<th>Variables</th>
<th>DF</th>
<th>Mauchly's Criterion</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
</table>
Sphericity Tests

<table>
<thead>
<tr>
<th>Variables</th>
<th>DF</th>
<th>Mauchly's Criterion</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orthogonal Components</td>
<td>5</td>
<td>0.6784796</td>
<td>7.6502678</td>
<td>0.1766</td>
</tr>
</tbody>
</table>

The sphericity test suggest that the correlations among the repeated measurements are homogeneous across all individuals ($P = 0.1766$). It is valid to make conclusion of the test from the analysis of the repeated measures analysis.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>drug</td>
<td>2</td>
<td>346.937500</td>
<td>173.468750</td>
<td>1.38</td>
<td>0.2738</td>
</tr>
<tr>
<td>Error</td>
<td>21</td>
<td>2642.187500</td>
<td>125.818452</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results suggest that there are not significant differences between prescription drugs ($P = 0.2738$).

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
<th>Adj Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>3</td>
<td>887.7083333</td>
<td>295.9027778</td>
<td>32.32</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>time*drug</td>
<td>6</td>
<td>457.9791667</td>
<td>76.3298611</td>
<td>8.34</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error(time)</td>
<td>63</td>
<td>576.8125000</td>
<td>9.1557540</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results suggest that there are significant differences in heart rate between times ($P < 0.0001$) and there are significant interaction between drugs and time ($P < 0.0001$).

2.3 Present an interaction plot between the mainplot and subplot effects, putting the subplot levels along the x-axis. Use this plot to help you express in words the largest likely component of the Mainplot*Subplot interaction.
The plot demonstrates how heart rate changes over the course of the four hours for each prescription drug. The plot demonstrates that drug 3 had a considerable different effect on heart rate over time compared to drugs 1 and 2 and is most likely to be the cause of the significant interaction.

Appendix

When you are asked to "describe in detail the design of this experiment," please do so by completing the following template:

<table>
<thead>
<tr>
<th>Design:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Response Variable:</td>
<td></td>
</tr>
<tr>
<td>Experimental Unit:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class Variable</th>
<th>Block or Treatment</th>
<th>Number of Levels</th>
<th>Fixed or Random</th>
<th>Description</th>
<th>Subsamples?</th>
<th>YES / NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 ↓ n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTICE: There is a new column in the above table ("Fixed or Random"). Now, for each class variable in your model, you need to indicate if it is a "Fixed" effect or a "Random" effect.