Midterm 1 Practice Problems

1. What are the fundamental principles of experimental design? What are their advantages?

2. In a regression model, there are 6 independent variables \( (x_1, x_2, \ldots, x_6) \) and one dependent variable \( y \). Fitting the full model,

model \( y=x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_6 \)

gives the output

```
Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>6</td>
<td>722.54361</td>
<td>120.42393</td>
<td>22.43</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>24</td>
<td>128.83794</td>
<td>5.36825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>30</td>
<td>851.38154</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE          2.31695  R-Square    0.8487
Dependent Mean    47.37581  Adj R-Sq    0.8108
Coeff Var         4.89057
```

a) What is \( R^2 \)? How to interpret \( R^2 \).
b) One wants to test if both \( x_5 \) and \( x_6 \) are significant. A model without \( x_5 \) and \( x_6 \) is fitted and part of the output is given below:

```
Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>695.14669</td>
<td>173.78667</td>
<td>28.92</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>26</td>
<td>156.23485</td>
<td>6.00903</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>30</td>
<td>851.38154</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE          2.45133  R-Square    0.8165
Dependent Mean    47.37581  Adj R-Sq    0.7883
Coeff Var         5.17423
```
Based on the above outputs, use a formal test for $H_0: \beta_5 = \beta_6 = 0$.

3. Define type I error, type II error and power.

4. An experiment was conducted to study the effect of quality control level (1=low, 2=moderate and 3=high) on the productivity improvement of some manufacturing lines. The experiment was completely randomized. There are 6 observations for each level. The sample means are

$$\bar{y}_1 = 6.983, \bar{y}_2 = 8.05, \bar{y}_3 = 9.20.$$  

And the corrected total sum of squares (SST) is 26.091.

a) Construct the ANOVA table.
b) Is there any difference between the quality control levels? State the null hypothesis and use $\alpha = 5\%$.
c) Use Tukey's procedure to perform pairwise comparison. What is the critical difference? Report the results.
d) What appears to be the nature of the relationship between quality control level and productivity improvements? An analyst used the orthogonal polynomial contrasts to study the relationship as follows

\begin{verbatim}
contrast "linear" level -1 0 1;
contrast "quadratic" level 1 -2 1;
================
Contrast  DF  Contrast SS  Mean Square  F Value  Pr > F
linear    1  14.74083333  14.74083333 19.49   0.0005
quadratic 1  0.00694444  0.00694444  0.01   0.9249
\end{verbatim}

What is your conclusion?

5. Four catalysts that affect the concentration of one component in a three-component liquid solution are being investigated. The following concentrations are obtained.

\begin{center}
\begin{tabular}{c c c c c c}
\hline
\textbf{catalyst} & \textbf{concentrations} \\
\hline
1 & 12.3 & 15.4 & 13.2 & 15.5 & 14.7 \\
2 & 8.5 & 7.8 & 9.3 & 8.6 & 9.0 \\
3 & 11.0 & 12.1 & 11.7 & 10.5 & 9.5 \\
4 & 17.5 & 18.3 & 15.7 & 14.4 & 20.0 \\
\hline
\end{tabular}
\end{center}

The ANOVA table is reported as follows

\begin{verbatim}
\end{verbatim}

2
Source | DF | Squares  | Mean Square | F Value | Pr > F
--- | --- | ------- | ----------- | ------ | -----
Model  | 3  | 209.4100000 | 69.8033333 | 34.15  | <.0001
Error   | 16 | 32.7000000  | 2.0437500  |  |  
C.Total | 19 | 242.1100000 |  |

a) What is your conclusion about the catalyst's effects on concentration?
b) A residual plot is generated to check model assumptions.

![Residual Plot]

which assumption appears to be violated? Discuss its influence on the results from ANOVA.
c) What remedies can be suggested?
d) The summary statistics are given in the following.

<table>
<thead>
<tr>
<th>Level of trt</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>14.2200000</td>
<td>1.41315250</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>8.6400000</td>
<td>0.56833089</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>10.9600000</td>
<td>1.02371871</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>17.1800000</td>
<td>2.19248717</td>
</tr>
</tbody>
</table>

Log(Std Dev) is plotted against log(Mean), and a simple regression line is also fitted and reported.
Based on the plot, what transformation you would use to stabilize the variances?

6. Four different washing solutions are being compared to study their effectiveness in retarding bacteria growth in 5-gallon milk containers. The analysis is done in a laboratory and only four trials can be conducted on any day. Four solutions are randomly applied to four containers. Observations are taken for four days, and the data are shown here.

<table>
<thead>
<tr>
<th>solution</th>
<th>days</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>22</td>
<td>18</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>24</td>
<td>17</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

a) Is it a completely randomized design or a randomized complete block design? why?
b) Suppose days are treated as blocks, the ANOVA table is obtained as follows,

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>6</td>
<td>2265.375000</td>
<td>377.562500</td>
<td>42.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>79.562500</td>
<td>8.840278</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.Total</td>
<td>15</td>
<td>2344.937500</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square : 0.966071  
Coeff Var : 17.55431  
Root MSE  : 2.973260  
Mean     : 16.93750
Test if there is any difference between the solutions (use $\alpha = 5\%$). Comment on the blocking strategy in this experiment.

c) Suppose the experimenter is interested in the following three contrasts:

\begin{align*}
C_1: & & 1 & -1 & 0 & 0 \\
C_2: & & 0 & 0 & 1 & -1 \\
C_3: & & 1 & 1 & -1 & -1
\end{align*}

Are these contrasts mutually orthogonal to each other? why?

d) Part of SAS output is given below

\begin{tabular}{llllll}
\hline
Contrast & DF & Contrast SS & Mean Square & F Value & Pr > F \\
\hline
C1 & 1 & 10.124 & 10.125 & 1.15 & 0.3124 \\
C2 & 1 & 24.500 & 24.500 & 2.77 & 0.1303 \\
\hline
\end{tabular}

test if C3 is significant (use $\alpha = 5\%$).

e) Calculate the critical differences for pairwise comparison in Tukey’s and Duncan’s procedures. Comments on their overall error rates and testing powers.

f) A residual vs. predicted plot is included here. Based on the plot, how do you think about the additivity assumption? What is the formal procedure to test if there exists some special interaction between treatment effects and blocking effects.
Remark: The practice problems are not necessarily comprehensive. It just shows you the type of problems you will have in exam. The homework problems are also important during your preparation.