There are totally 38 points in the exam. The students with score higher than or equal to 35 points will receive 35 points. Please write down your name and student ID number.

NAME: ________________________________
ID: ________________________________
1. (10 points). The data reported 12 core samples from petroleum reservoirs, where each core sample was measured by 4 cross-sections. Each core sample was measured for permeability, and each cross-section has total area of cores, total perimeter of cores, and shaped. Therefore, there are totally \( n = 48 \) observations. The variables are: area (based on an image in pixels out of \( 256 \times 256 \)), peri (perimeter in pixels), shape, and perm (permeability). We choose area as the response variable and the rest three as explanatory variables. It is known that the MSE of the model with all of the three explanatory variables is \( \hat{\sigma}^2 = 1689242 \). The SAS output for the model with only peri and perm explanatory variables is given below.

Root MSE 1302.5457 R-Square 0.7745

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimates</th>
<th>Variance</th>
<th>Inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>144.5678</td>
<td>771.0387</td>
<td>0.19</td>
</tr>
<tr>
<td>peri</td>
<td>1</td>
<td>2.1850</td>
<td>0.1969</td>
<td>11.10</td>
</tr>
<tr>
<td>perm</td>
<td>1</td>
<td>2.8466</td>
<td>0.6438</td>
<td>4.42</td>
</tr>
</tbody>
</table>

(a) (2 points). Write down the model assumption and the fitted regression line based on the SAS output.

Solution: The model assumption is

\[
\text{area}_i = \beta_0 + \text{peri}_i \beta_1 + \text{perm}_i \beta_2 + \epsilon_i,
\]

where \( \epsilon_i \sim \text{iid} N(0, \sigma^2) \). Based on the SAS output, we have \( \hat{\beta}_0 = 144.5678 \), \( \hat{\beta}_1 = 2.1850 \), \( \hat{\beta}_2 = 2.8466 \), and \( \hat{\sigma} = 1302.5457 \). Therefore, the fitted regression line is

\[
\text{area} = 144.5678 + 2.1850 \text{peri} + 2.8466 \text{perm}.
\]

(b) (2 points). Test the significance of shape in the model with all of the three explanatory variables.

Solution: Using \( R^2 = 0.7745 \), \( \hat{\sigma}^2 = 1302.5457 \), and \( n = 48 \), we have \( SSE = 1302.5457^2 \times 45 = 76348139 \), \( SST = SSE / (1 - R^2) = 338572676 \). The SSE of the model with all of the three explanatory variables is \( SSE(F) = 1689242 \times 44 = 74326648 \). Therefore, we have the F-statistic

\[
F^* = \frac{76348139 - 74326648}{1302.5457^2} = 1.1915.
\]

Based on \( F_{1,44} \)-distribution which provides \( F_{0.05,1,44} = 4.062 \), we conclude that shape is not significant in the model with all of the three explanatory variables.

(c) (2 points). Explain the reason why the VIF values are the same in the SAS output.

Solution: According to the definition of VIF, we need consider the inverse of the correlation matrix between peri and perm, which has the form of

\[
\mathbf{A} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},
\]

where \( \rho \) is the correlation between peri and perm. The VIF value is the diagonal of \( \mathbf{A} \). It is clear that the two values are equal since they are both equal to \( 1/(1 - \rho^2) \).
(d) (2 points). Compute the type III sum of squares of the two variables based on the SAS output.

**Solution:** The F-value of the type III sum of squares are \( 11.10^2 = 123.21 \) for peri and \( 4.42^2 = 19.5364 \) for peri. Using the value of MSE, we have \( SS3(peri) = MS3(peri) = 123.21 \times 1302.5457^2 = 209041203 \) and \( SSE(perm) = MS3(perm) = 19.5364 \times 1302.5457^2 = 33146120. \)

(e) (2 points). Compute the type I sum of squares of the two variables based on the SAS output.

**Solution:** For perm, we have \( SS1(perm) = SS3(perm) = 33146120. \) For peri, we have \( SS1(peri) = SSM - SS1(perm) = 74326648 - 33146120 = 41180528. \)

2. (10 points). The removal of ammoniacal nitrogen is an important aspect of treatment of leachate at landfill sites. The rate of removal (in % per day) is recorded for several days for each of several treatment methods. The summary of the data is given in the following table.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Sample Size</th>
<th>Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>4.43</td>
<td>1.25</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>3.25</td>
<td>0.99</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>6.43</td>
<td>1.74</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>5.53</td>
<td>1.01</td>
</tr>
</tbody>
</table>

(a) (2 points). Complete the following ANOVA table.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>47.644</td>
<td>15.881</td>
<td>9.580</td>
</tr>
<tr>
<td>Error</td>
<td>29</td>
<td>48.075</td>
<td>1.6578</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>32</td>
<td>95.719</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We have \( \bar{y} = \frac{\sum_{i=1}^{4} n_i \bar{y}_i}{\sum_{i=1}^{4} n_i} = 5.023 \) and the MSE \( \hat{\sigma}^2 = \frac{\sum_{i=1}^{4} (n_i - 1) s_i^2}{\sum_{i=1}^{4} (n_i - 1)} = 1.6578. \) Therefore, we have \( SSM = \sum_{i=1}^{4} n_i (\bar{y}_i - \bar{y}.)^2 = 47.644 \) and \( SSE = 29\hat{\sigma}^2 = 48.075. \) Therefore, we have the following table.

(b) (2 points). Write down the baseline model (treatment 4 is the baseline). Compute the estimate of parameters and their standard errors (but not \( \sigma^2 \)).

**Solution:** The baseline model is

\[
y_{ij} = \mu + \alpha_i + \epsilon_{ij},
\]

where \( \alpha_4 = 0 \) and \( \epsilon_{ij} \sim iid \ N(0, \sigma^2). \) We have \( \hat{\mu} = 5.53 \) with standard error \( s(\hat{\mu}) = \hat{\sigma} / \sqrt{10} = 0.4072, \hat{\alpha}_1 = 4.43 - 5.53 = -1.10 \) with standard error \( s(\hat{\alpha}_1) = \hat{\sigma} \sqrt{1/6 + 1/10} = \)
3. (10 points). The article “Cellulose Acetate Microspheres Prepared by O/W Emulsification and Solvent Evaporation Methods” (K. Soppinmath, A Kulkarni, et al., Journal of Microencapsulation, 2001, 811-817) describes a study on the effects of the concentrations of polyvinyl alcohol (PVAL) and dichloromethane (DCM) on the encapsulation efficiency in a process that produces microspheres containing the drug ibuprofen. There were three concentrations of PVAL (measure in units of % w/v) and three of DCM (in mL). The results presented in the following table. There are three replications in each combination. It is known that the MSE of the interaction effect ANOVA model is 7.9164, and the MSE of the main effect model is 13.7315.

0.6649, $\hat{\alpha}_2 = (3.25 - 5.53) = -2.28$ with standard error $s(\hat{\alpha}_2) = \hat{\sigma} \sqrt{1/8 + 1/10} = 0.6017$, and $\hat{\alpha}_3 = 6.43 - 5.53 = 0.90$ with standard error $s(\hat{\alpha}_2) = \hat{\sigma} \sqrt{1/9 + 1/10} = 0.5916$.

(c) (2 points). Write down the factor effect model. Compute the estimate of parameters and their standard errors (but not $\sigma^2$).

Solution: The factor effect model is

$$y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

where $\sum_{i=1}^4 \tau_i = 0$ and $\epsilon_{ij} \sim iid N(0, \sigma^2)$. We have $\mu = (4.43 + 3.25 + 6.43 + 5.53)/4 = 4.91$ with standard error $s(\hat{\mu}) = (\hat{\sigma}/4) \sqrt{1/6 + 1/8 + 1/9 + 1/10} = 0.2282$, $\hat{\tau}_1 = 4.43 - 4.91 = -0.48$ with standard error $s(\hat{\tau}_1) = (\hat{\sigma}/4) \sqrt{9/6 + 1/8 + 1/9 + 1/10} = 0.4362$, $\hat{\tau}_2 = 3.25 - 4.91 = -1.66$ with standard error $s(\hat{\tau}_2) = (\hat{\sigma}/4) \sqrt{1/6 + 9/8 + 1/9 + 1/10} = 0.3946$, $\hat{\tau}_3 = 6.43 - 4.91 = 1.52$ with standard error $s(\hat{\tau}_3) = (\hat{\sigma}/4) \sqrt{1/6 + 1/8 + 9/9 + 1/10} = 0.3797$, and $\hat{\tau}_4 = 5.53 - 4.91 = -0.62$ with standard error $s(\hat{\tau}_4) = (\hat{\sigma}/4) \sqrt{1/6 + 1/8 + 1/9 + 9/10} = 0.3674$.

(d) (2 points). Provide an $F$-test to assess whether the variances of the first and the second treatments are equal. State the null hypothesis, the test statistics, and your conclusion.

Solution: The F-value is $F^* = 1.25^2/0.99^2 = 1.594$. Based $F_{5, 7}$ distribution which provides $F_{0.025, 5, 7} = 5.285$ and $F_{0.075, 5, 7} = 0.146$, we conclude that the variances of the two levels are equal since $1.594 \in [0.146, 5.285]$.

(e) (2 points). Based on the method used in the previous part and the Bonferroni approach, describe a method to evaluate whether the variances of the four treatments are equal.

Solution: We can test $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$, which can be considered by pairs. To apply the Bonferroni approach, we consider six tests: $\sigma_1^2 = \sigma_2^2$, $\sigma_1^2 = \sigma_3^2$, $\sigma_1^2 = \sigma_4^2$, $\sigma_2^2 = \sigma_3^2$, $\sigma_2^2 = \sigma_4^2$, and $\sigma_3^2 = \sigma_4^2$. Therefore, we calculate their $(1 - 0.05/6)% = 99.17\%$ Bonferroni confidence intervals. If all of the ratios $\sigma_i^2/\sigma_j^2$ belong to the confidence intervals, then we conclude that the variances are all equal; otherwise, we conclude they are not all equal.
(a) (2 points). Write down the assumption of the main effect model with the zero-sum constraint. Compute the estimates of model parameters.

**Solution:** The main effect model is

\[ y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \]

for \( i, j, k = 1, 2, 3 \), where \( \sum_{i=1}^{3} \alpha_i = \sum_{j=1}^{3} \beta_j = 0 \), and \( \epsilon_{ijk} \sim iid \ N(0, \sigma^2) \). The estimates of parameters are \( \hat{\mu} = 86.17 \), \( \hat{\alpha}_1 = 88.23 - 86.17 = 2.06 \), \( \hat{\beta}_2 = 83.23 - 86.17 = -2.94 \), \( \hat{\beta}_1 = 76.13 - 86.17 = -10.04 \), and \( \hat{\sigma}^2 = 13.7315 \), where \( \hat{\alpha}_3 = -(\hat{\alpha}_1 + \hat{\beta}_2) \) and \( \hat{\beta}_3 = -(\hat{\beta}_1 + \hat{\beta}_2) \).

(b) (2 points). Test whether the interaction effect is significant. You need to write down the null hypothesis, the test statistic, testing rule, and conclusion.

**Solution:** The interaction effect model is

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}, \epsilon_{ijk} \sim iid \ N(0, \sigma^2), \]

where \( \sum_{i=1}^{3} \alpha_i = \sum_{j=1}^{3} \beta_j = \sum_{i=1}^{3} (\alpha\beta)_{ij} = 0 \). The null hypothesis is

\[ H_0 : (\alpha\beta)_{ij} = 0 \]

for all \( i, j = 1, 2, 3 \). The consider the an \( F \)-statistic, which is

\[ F^* = \frac{(22(13.7315) - 18(7.9164))/4}{7.9164} = 5.0401. \]

Comparing it with \( F_{0.05,4,18} = 2.9277 \), we conclude that \( F^* \) is significant. Therefore, we reject \( H_0 \) and conclude that the interaction effect is significant.

(c) (2 points). Propose a column-effect model (using PVAL true values in the interaction effect) and a linear-by-linear association model. Write down the model assumptions of the two models.

**Solution:** The column effect model is

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_j X_i^\alpha + \epsilon_{ijk}, \epsilon_{ijk} \sim iid \ N(0, \sigma^2), \]

where \( \sum_{i=1}^{3} \alpha_i = \sum_{j=1}^{3} \beta_j = \sum_{j=1}^{3} \gamma_j = 0 \) and \( X_i^\alpha = 0.5, 1.0, 2.0 \) for the first row, the second row, and the third row, respectively.

The linear-by-linear association model is

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + X_i^\alpha X_j^\beta + \epsilon_{ijk}, \epsilon_{ijk} \sim N(0, \sigma^2), \]

where \( \sum_{i=1}^{3} \alpha_i = \sum_{j=1}^{3} \beta_j = X_i^\alpha = 0.5, 1.0, 2.0 \) for the first row, the second row, and the third row, respectively, and \( X_j^\beta = 30, 40, 50 \) for the first row, the second row, and the third row, respectively.
(d) (2 Points). Write down the null hypothesis of the test about whether the interaction effect model can be reduced to the column effect model as well as the test about whether the column effect model can be reduced to the linear-by-linear association model.

Solution: To test whether the interaction effect can be reduced to column effect model, the null hypothesis is

\[ H_0 : (\alpha\beta)_{31} - (\alpha\beta)_{21} = 2((\alpha\beta)_{21} - (\alpha\beta)_{11}), (\alpha\beta)_{32} - (\alpha\beta)_{22} = 2((\alpha\beta)_{22} - (\alpha\beta)_{12}) \]

or

\[ H_0 : \frac{(\alpha\beta)_{31} - (\alpha\beta)_{21}}{(\alpha\beta)_{21} - (\alpha\beta)_{11}} = \frac{(\alpha\beta)_{32} - (\alpha\beta)_{22}}{(\alpha\beta)_{22} - (\alpha\beta)_{12}} = 2. \]

To test whether the row effect model can be reduced to the linear-by-linear association model, the null hypothesis is

\[ H_0 : \gamma_3 - \gamma_2 = \gamma_2 - \gamma_1. \]

(e) (2 points). The MSE of the column effect model is 12.3664. Test whether the interaction effect model can be reduced to the column effect model. You need to write down the test statistic, testing rule, and conclusion.

Solution: We consider an F-test. The F-value is

\[ F = \frac{[20(12.366) - 18(7.9164)]/2}{7.9164} = 6.6207. \]

Comparing it with \( F_{0.05,2,18} = 3.5548 \), we conclude that \( F^* \) is significant. Therefore, the interaction effect model cannot be reduced to the row effect model.

4. (8 points). The article ”Use of Taguchi Methods and Multiple Regression Analysis for Optima Process Development of High Energy Beam Case Hardening of Cast Iron” (M. Jean and Y. Tzeng, Surface Engineering, 2003, 150-156) describes a factorial experiment designed to determine factors in high-energy electron beams process that affect hardness in metals. Results for two factors, each with three levels are collected. Factor A is the travel speed in mm/s (10, 2, or 30). Factor B is accelerating voltage in volts (10, 25, and 50). The outcome is Vickers hardness. There are six replications for each combination. The SAS output are given, where \( SST = 543939. \) \( t_{0.025,45} = 2.0141, t_{0.0125,45} = 2.3189, t_{0.0083,45} = 2.4868, \) and \( F_{0.05,3,45} = 2.8115 \).

| Parameter    | Estimate | Standard Error | t Value | Pr > |t| |
|--------------|----------|----------------|---------|------|---|
| Intercept    | 717.0000000 B | 31.92746409  | 22.46   | <.0001 |
| speed 10     | -69.1666667 B  | 45.15225272 | -1.53   | 0.1326  |
| speed 20     | 4.1666667 B      | 45.15225272 | 0.09    | 0.9269  |
| speed 30     | 0.0000000 B      | .            | .       | .       |
| voltage 10   | 157.8333333 B    | 45.15225272 | 3.50    | 0.0011  |
| voltage 25   | 94.5000000 B     | 45.15225272 | 2.09    | 0.0420  |
| voltage 50   | 0.0000000 B      | .            | .       | .       |
| speed*voltage 10 10 | -33.5000000 B | 63.85492817 | -0.52   | 0.6024  |
(a) (2 points). Compute the MSE and $R^2$ of the fitted model.

**Solution:** Note that $\hat{\mu} = \bar{y}_{33}$. We have

$$V(\hat{\mu}) = \frac{\sigma^2}{6} \Rightarrow \hat{\sigma^2} = 6 \times 31.9274^2 = 6116.15.$$ 

Therefore,

$$R^2 = 1 - \frac{45 \times 6116.15}{543939} = 0.4940.$$ 

(b) (2 points). Compute the 95% Bonferroni joint confidence intervals for all of the differences between the main effect $A$, then the same for the main effect $B$.

**Solution:** The variance of $\hat{\alpha}_i - \hat{\alpha}_j$ are all identical. The formulae are

$$\hat{\alpha}_i - \hat{\alpha}_j \pm t_{0.025} \times 45.1522 = \hat{\alpha}_i - \hat{\alpha}_j \pm 2.4868 \times 45.15.$$ 

We have the 95% Bonferroni joint confidence interval for $\mu_1 - \mu_2$ is $[-185.62, 38.95]$, for $\alpha_1 - \alpha_3$ is $[-181.45, 43.18]$, and for $\alpha_2 - \alpha_3$ is $[-108.12, 116.45]$.

(c) (2 points). Compute the 95% Tukey joint confidence intervals for all of the differences between the main effect $A$, then the same for the main effect $B$.

**Solutions:** The 95% Tukey joint confidence interval is

$$\hat{\alpha}_i - \hat{\alpha}_j \pm \frac{q_{3,45}}{\sqrt{2}} \times 45.1522 = \hat{\alpha}_i - \hat{\alpha}_j \pm 3.4275 \times 45.1522.$$ 

Then, the 95% Tukey joint confidence interval for $\mu_1 - \mu_2$ is $[-182.76, 36.10]$, for $\alpha_1 - \alpha_3$ is $[-178.60, 40.26]$, and for $\alpha_2 - \alpha_3$ is $[-105.26, 113.60]$.

(d) (2 points). Let $\alpha_1, \alpha_2, \alpha_3$ be denoted as parameters for main effect of $A$. Let $L_1 = (\alpha_1 + \alpha_2)/2 - \alpha_3$, $L_2 = (2\alpha_1 + \alpha_2)/3 - \alpha_3$, $L_3 = \alpha_1 - (3\alpha_2 + \alpha_3)/4$, and $L_4 = (3\alpha_1 + 2\alpha_3)/5 - \alpha_2$. Compute the 95% Scheffe’s joint confidence interval for $L_1, L_2, L_3, \text{and } L_4$.

**Solutions:** We have $\hat{L}_1 = 32.5$, $\hat{L}_2 = -44.72$, $\hat{L}_3 = -72.29$, and $\hat{L}_4 = -45.67$. Their standard errors are $s(\hat{L}_1) = 95.78$, $s(\hat{L}_2) = 97.54$, $s(\hat{L}_3) = 99.69$, and $s(\hat{L}_4) = 96.42$. Using the formula for the interval

$$\hat{L}_i \pm \sqrt{3F_{0.05,3,45} s(\hat{L}_i)}.$$ 

We have the 95% Scheffe’s joint confidence interval for $L_1$ is $[-310.66, 245.66]$, for $L_2$ is $[-322.19, 232.75]$, for $L_3$ is $[-361.81, 217.23]$, and for $L_4$ is $[-325.69, 234.36]$. 

<table>
<thead>
<tr>
<th>speed*voltage</th>
<th>10 25</th>
<th>-71.1666667 B</th>
<th>63.85492817</th>
<th>-1.11</th>
<th>0.2710</th>
</tr>
</thead>
<tbody>
<tr>
<td>speed*voltage</td>
<td>10 50</td>
<td>0.0000000 B</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>speed*voltage</td>
<td>20 10</td>
<td>-52.5000000 B</td>
<td>63.85492817</td>
<td>-0.82</td>
<td>0.4153</td>
</tr>
<tr>
<td>speed*voltage</td>
<td>20 25</td>
<td>-31.8333333 B</td>
<td>63.85492817</td>
<td>-0.50</td>
<td>0.6205</td>
</tr>
<tr>
<td>speed*voltage</td>
<td>20 10</td>
<td>-52.5000000 B</td>
<td>63.85492817</td>
<td>-0.82</td>
<td>0.4153</td>
</tr>
<tr>
<td>speed*voltage</td>
<td>30 10</td>
<td>0.0000000 B</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>speed*voltage</td>
<td>30 25</td>
<td>0.0000000 B</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>speed*voltage</td>
<td>30 10</td>
<td>0.0000000 B</td>
<td>.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>