I have neither given nor received unauthorized assistance on this exam.

KEY

______________________________
Name Signed                    Date

______________________________
Name Printed
1. An IE 361 project group studied the operation of a cut-off machine for cutting 304 stainless steel tubing. The value of

\[ y = \text{the number of tubes cut before failure of a carbide cutting insert} \]

was recorded for several inserts at each of 4 feed rates. The counts and some summary statistics are:

<table>
<thead>
<tr>
<th>Feed Rate #1</th>
<th>Feed Rate #2</th>
<th>Feed Rate #3</th>
<th>Feed Rate #4</th>
</tr>
</thead>
<tbody>
<tr>
<td>125,129,146</td>
<td>135,130,176</td>
<td>194,183,166</td>
<td>176,187,204</td>
</tr>
<tr>
<td>( \bar{y}_1 = 133.3 )</td>
<td>( \bar{y}_2 = 147.0 )</td>
<td>( \bar{y}_3 = 181.0 )</td>
<td>( \bar{y}_4 = 189.0 )</td>
</tr>
<tr>
<td>( s_1 = 11.2 )</td>
<td>( s_2 = 25.2 )</td>
<td>( s_3 = 14.1 )</td>
<td>( s_4 = 14.1 )</td>
</tr>
</tbody>
</table>

Initially, consider ONLY the information about Feed Rate #1, and assume that number-of-cuts for this feed rate is approximately normally distributed.

5 pts a) Give two-sided 95% confidence limits for the standard deviation of number of tubes cut by an insert at Feed Rate #1. (Plug in completely, but you need not simplify.)

\[
\frac{\chi^2_{L}}{n}, \frac{\chi^2_{U}}{n} \leq \frac{\chi^2_{\alpha/2}}{n}
\]

This is

\[
11.2 \sqrt{\frac{3-1}{7.378}}, 11.2 \sqrt{\frac{3-1}{0.051}}
\]

\((\text{units are "tubes cut"})\)

5 pts b) Give two-sided 95% confidence limits for the mean number of tubes cut by an insert at Feed Rate #1. (Plug in completely, but you need not simplify.)

\[
\bar{y} \pm t \frac{s}{\sqrt{n}}
\]

This is

\[
133.3 \pm 4.303 \frac{11.2}{\sqrt{3}}
\]

upper 2.5% point

\[
133.3 \pm 27.8 \sqrt{t_2} (\text{units are "tubes cut"})
\]

5 pts c) Give endpoints of a two-sided interval that you are 95% sure will contain the number of tubes cut by one of these inserts at Feed Rate #1 tomorrow. (Plug in completely, but you need not simplify.)

\[
\bar{y} \pm ts \sqrt{1 + \frac{1}{n}}
\]

This is

\[
133.3 \pm 4.303 \left(11.2\right) \sqrt{1 + \frac{1}{3}}
\]

\[
133.3 \pm 55.4 (\text{tubes cut})
\]
Now consider the results from ALL of Feed Rates #1 through #4.

5 pts  **d)** For \( y_{ij} \) = number of tubes cut by insert \( j \) at feed rate \( i \), below is a normal plot of all 12 values 

\[
\left( y_{ij} - \bar{y}_j \right) / \left( s_{\text{pooled}} \sqrt{\frac{2}{3}} \right).
\]

Say what it indicates to you about the appropriateness of analyzing these experimental results based on the one-way normal model.

Regardless of how you answered d) proceed under the assumptions of the one-way normal model.

5 pts  **e)** Give an estimate of the standard deviation of the number of tubes cut by an insert for any single Feed Rate. (A single number will suffice here. You need _not_ make confidence limits.)

\[
S^2_{\text{pooled}} \approx \frac{(3-1)(1.2)^2 + (3-1)(25.2)^2 + (3-1)(14.1)^2 + (3-1)(14.1)^2}{(3-1) + (3-1) + (3-1) + (3-1)}
\]

\[
= 289.53
\]

So \( S_{\text{pooled}} = \sqrt{289.53} = 17.0 \) (tubes cut)

5 pts  **f)** Use your answer to e) and give a lower 95% confidence limit for the difference in mean cuts made by an insert under Feed Rates #4 and #1, \( \mu_4 - \mu_1 \). (Plug in completely, but you need not simplify.)

Use

\[
(\bar{y}_4 - \bar{y}_1) - t_{s_{\text{pooled}}} \sqrt{\frac{1}{n_1} + \frac{1}{n_4}}
\]

This is

\[
(189.0 - 133.3) - 1.860(17.0) \sqrt{\frac{2}{8}}
\]

A upon 5% pt of \( t_8 \)

\[ \leq 23.7 \] tubes cut
As it turns out, in this problem \( SSTr = 6406.25 \) and the sample variance of all 12 values \( y_i \) is 
\[ s^2 = 793.1742. \]
This means \( SStot = (n-1)s^2 = 11 \times (793.1742) = 8724.92 \) 
So \( SSE = 8724.92 - 6406.25 = 2318.67 \)

5 pts

\[ F = \frac{MSTr}{MSE} = \frac{6406.25/3}{2318.67/12} = \frac{s^2_{\text{pooled}}} {s^2} \]

\[ F = \frac{7.37}{df} = 3 \quad 8 \]

h) The following is a non-standard question (not an application of standard formulas for the \( r \)-sample problem) and will require thinking from basics: One might wish to make a prediction interval for the difference between a new value of \( y \) for Feed Rate #4 and one for Feed Rate #1, say \( y_{4\text{new}} - y_{1\text{new}} \). That can be done if one can find a "standard error" for 
\[ (y_{4\text{new}} - y_{1\text{new}}) - (\bar{y}_4 - \bar{y}_1) \]

What is the variance of this random variable? (Your answer should be some multiple of \( \sigma^2 \).) How would you estimate the corresponding standard deviation? (This answer should be a number.)

Variance: Using independence of \( y_{1\text{new}}, y_{4\text{new}}, \bar{y}_1, \bar{y}_4 \) The variance is 
\[ \frac{2}{3} \sigma^2 + \frac{2}{3} \sigma^2 + \frac{2}{3} \sigma^2 = \frac{8}{3} \sigma^2 \]

Estimate of corresponding standard deviation (a standard error):

\[ \frac{\sqrt{8/3}}{\sqrt{3}} s_{\text{pooled}} = \frac{\sqrt{8/3}}{3} (17.0) = 27.8 \]

2. A continuous random variable, \( X \), taking values between 0 and 1 has pdf \( f(x) = 4x^3 \) for \( 0 < x < 1 \).

5 pts

\[ a) \] Evaluate \( EX \) and \( P[X < .5] \).

\[ EX = \int_0^1 x \cdot f(x) \, dx = \int_0^1 x \cdot (4x^3) \, dx = \frac{4}{5} x^5 \bigg|_0^1 = \frac{4}{5} \]

\[ P[X < .5] = \int_{-10}^{.5} f(x) \, dx = \int_{0}^{.5} 4x^3 \, dx = \frac{4}{4} x^4 \bigg|_0^{.5} = .0625 \]

\[ \frac{4}{5} = .8 \]

\[ P[X < .5] = (.5)^4 = .0625 \]

4
b) The distribution in a) is one that might be used in project management analysis, and for 
$X_1$ and $X_2$ independent random variables with this distribution, the sum $X_1 + X_2$ might represent 
the total time required to complete a "two-task" project. Set up completely (but do not evaluate) a 
double integral giving $P[X_1 + X_2 > 1]$. (Hint: Begin by making a picture of the part of the ($x_1, x_2$) 
plane involved here.)

\[
\begin{align*}
    f(x_1, x_2) &= \begin{cases} 
        (4x_1^3)(4x_2^3) & \text{on } D \\
        0 & \text{elsewhere}
    \end{cases} \\
    P[X_1 + X_2 > 1] &= \int_{D} f(x_1, x_2) \, dx_1 \, dx_2 \\
    &= \int_{0}^{1} \int_{0}^{1 - x_2} 16x_1^3x_2^3 \, dx_1 \, dx_2
\end{align*}
\]

\[5 \text{ pts}\]

\[5 \text{ pts}\]

c) Suppose in the project management context mentioned in b) variables $X_1, X_2, \ldots, X_{36}$ are times 
required to complete 36 tasks and are modeled as independent variables with a common marginal 
distribution with mean $.7$ and standard deviation $.1$. Approximate the probability that the total time 
required to complete the 36 tasks exceeds 30.0. (Hint: What is the event of interest in terms of $\bar{X}$ ?)

The distribution of $\bar{X}$ is approximately normal with mean $\mu = .7$ and standard deviation $\sigma = \frac{.1}{\sqrt{36}} = .1/6$. So

\[P[\bar{X} > .83] \approx P[Z > \frac{.83 - .7}{.1/6}] = P[Z > 8] \approx 0\]

\[5 \text{ pts}\]

3. Suppose that a Poisson distribution is a sensible model for the number of nuggets of a particular 
size found processing a given amount of gravel at a gold mining site.

a) At one gold mine, nuggets are found at a rate of about 1 per 100 tons of processed gravel (so 
that, for example, the mean number found in 300 tons of material is 3.0). On a particular day 300 
tons of gravel are processed. Evaluate the probability that at least 2 nuggets are found.

For $X = \# \text{ of nuggets found}$, 
$X \sim \text{Poisson(} \lambda = 5.0 \text{)}$

\[P(X \geq 2) = 1 - P(X \leq 1) = 1 - f(0) - f(1)
\]

\[= 1 - \frac{e^{-3}(3)^0}{0!} - \frac{e^{-3}(3)^1}{1!} = .8005\]
b) 300 tons are processed at the mine in a) every day beginning on January 1. Use your answer to a) and find the probability that among the first 5 days of the year no day produces at least 2 nuggets. (If you could not do part a) you may use the incorrect answer .7 in place of a value from a).

From a) \( P[\text{no nuggets on a single day}] = 1 - .8008 = .1991 \)

Then (using a Bernoulli trials model, and thus either a binomial den for # of successes in 5 days or a geometric model for the waiting time to a first success)

\( P[\text{all 5 days produce no nuggets}] = ( .1991)^5 = .0003 \)

5 pts

c) At a mine different from the one referred to in parts a) and b) \( n = 400 \) runs of (100 tons of) gravel produce 40 runs without nuggets of the size of interest. Give 2-sided 95% confidence limits for the fraction of all runs at this site that that would fail to have nuggets of the size of interest. Then translate those limits to limits for the mean number of nuggets in runs of this size at this mine.

Limits for the long run fraction of runs with no nuggets:

Use \( \hat{p} = \frac{40}{400} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \)

This is \( \frac{40}{400} \pm 1.96 \sqrt{\frac{\frac{40}{400}(1-\frac{40}{400})}{400}} \)

i.e. \( .1 \pm .0299 \) i.e. \( (.07, .13) \)

Limits for the mean number of nuggets per run:

For \( X \sim \text{Poisson}(\lambda) \), \( f(0) = P[X = 0] = \frac{e^{-\lambda} \lambda^0}{0!} = e^{-\lambda} \) and thus \( \lambda = - \log f(0) \). So limits for \( \hat{p} = f(0) \) above become limits for \( \lambda \) by taking -logarithm and these are

\[ -\ln(.13) = 2.04 \quad \text{and} \quad -\ln(.07) = 2.66 \]

4. At the end of this exam are some JMP reports for some regression analyses of data of Lee, Posarac, and Ellis (from "An Experimental Investigation of Biodiesel Synthesis from Waste Canola Oil Using Supercritical Methanol," 2012, Fuel, Vol. 91, pp. 229-237.) These data were collected in an experimental study of how

\[ x_1 = \text{processing time (minutes - 30)} \]
\[ x_2 = \text{processing temperature (°C - 252.63158)} \]
\[ x_3 = \text{methanol/oil weight ratio (ratio -1.5)} \]

affect

\[ y' = \ln y = \ln (\% \text{ yield of methyl ester}) \]

in the processing of waste canola oil.

Use these JMP reports as you answer the remaining questions on this exam.
5 pts a) Give fractions of raw variability in $y'$ accounted for, using first $x_2$ alone, and then the pair $(x_2, x_3)$ as predictors. Using the fact that for SLR $R^2 = (r_{x,y})^2$, we get $R^2 = (0.6205)^2 = 0.385$.

Fraction explained using $x_2$ alone: \[0.385\]

Then, since $x_2$ and $x_3$ are uncorrelated, $R^2$ for the full model containing both is the sum of the $R^2$ values for the two SLR models. SLR using only $x_3$ has $R^2 = (0.3827)^2$. So $R^2_{\text{full}} = 0.385 + (0.3827)^2 = 0.532$.

Fraction explained using both $x_2$ and $x_3$ in a single model: \[0.532\]

There is a Fit Y by X output for inference based on the model $y' = \beta_0 + \beta_1 x_1 + \epsilon$ included in the JMP reports. Use it as you answer the questions b) and c).

5 pts b) Give 95% confidence limits for the increase in mean log percent yield that accompanies a 5 minute increase in processing time. This is $5\beta_1$ that needs estimating. Limits for $5\beta_1$ are 5 times limits for $\beta_1$. Limits for $\beta_1$ are $b_1 \pm tSE b_1$, i.e. $0.0685 \pm 2.110 (0.0195)$ upper 2.5% pt of t(17) $\approx (0.0265, 0.1105)$

So limits for $5\beta_1$ are $(0.1326, 0.5525)$.

5 pts c) Give 95% prediction limits for the next log percent yield for a 30 minute processing time. (MAKE SURE YOUR ANSWER MAKES SENSE in comparison to $y'$ values in the data.) A 30 minute processing time is $x_1 = 0$. Notice too that $\bar{x}_1 = 0$!

So, $\hat{y} = b_0 + b_1 (0) = b_0$ and prediction limits are $\hat{y} \pm tS\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x-\bar{x})^2}}$ i.e. $b_0 \pm tS\sqrt{1 + \frac{1}{15}}$

Here this is $2.465 \pm 2.110 (1.033)(1.033)\sqrt{1 + \frac{1}{15}}$

There are two Fit Model reports included in the JMP reports. Use them as appropriate as you answer the questions d) through f).
d) The model \( y' = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \) is one fit to the data. For this model consider prediction of \( y' \) for \( x_1 = 0, x_2 = 0, \) and \( x_3 = 0 \). As it turns out, for this set of conditions and this model, \( SE_{y'} = .1012 \). Give 95\% two-sided prediction limits for \( y'_{\text{new}} \). (Plug in completely, but you need not simplify.)

\[
\text{For this set of conditions } \hat{y} = b_0
\]

\[
\hat{y} \pm t \sqrt{(SE_{y'})^2 + s^2} \quad \text{This is } \quad MSE
\]

\[
2.465 \pm 2.131 \sqrt{(.1012)^2 + .19475}
\]

i.e. \( 2.465 \pm .965 \)

---

e) Give the value of an F statistic and degrees of freedom for testing whether a quadratic function of \( x_1, x_2, \) and \( x_3 \) provides a detectable improvement over a linear function of these variables as a model for \( y' \).

\[
F = \frac{(SSR(\text{full}) - SSR(\text{Reduced})) / (k-1)}{MSE(\text{full})}
\]

\[
\text{Here this is } \quad F = \frac{(30.064 - 27.897) / (9-3)}{.08386}
\]

\[
F = 4.31 \quad d.f. = 6, 9
\]

---

f) What on the attached pages and in the table below suggests that there is a model "between" the linear and the full quadratic ones (with more predictors than the first and fewer than the second) that should be considered?

On JMP reports: Not all estimated coefficients for the full model have large \( |b| \) values.

In the table: The 6-variable model has \( R^2 \) nearly as big as the full model, a good value of \( Cp \) (\( Cp \approx p+1 \)) and \( \text{RMSE} \) nearly minimum.
<table>
<thead>
<tr>
<th>time</th>
<th>temp</th>
<th>ratio</th>
<th>yield</th>
<th>logyield</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>240</td>
<td>1</td>
<td>1.5</td>
<td>0.4054651081</td>
<td>-15</td>
<td>-12.63157895</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>240</td>
<td>1.5</td>
<td>3.2</td>
<td>1.1631508098</td>
<td>-15</td>
<td>-12.63157895</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>240</td>
<td>2</td>
<td>3.8</td>
<td>1.3360010667</td>
<td>-15</td>
<td>-12.63157895</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>255</td>
<td>1</td>
<td>2.2</td>
<td>0.7884573604</td>
<td>-15</td>
<td>2.3584210526</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>270</td>
<td>1</td>
<td>8.9</td>
<td>2.1860512767</td>
<td>-15</td>
<td>17.368421053</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>270</td>
<td>2</td>
<td>13.6</td>
<td>2.5100697927</td>
<td>-15</td>
<td>17.368421053</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>240</td>
<td>1</td>
<td>1.5</td>
<td>0.4054651081</td>
<td>0</td>
<td>-12.63157895</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>240</td>
<td>2</td>
<td>12.3</td>
<td>2.5095992024</td>
<td>0</td>
<td>-12.63157895</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>255</td>
<td>1.5</td>
<td>11.4</td>
<td>2.4336133554</td>
<td>0</td>
<td>2.3684210526</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>255</td>
<td>2</td>
<td>13.6</td>
<td>2.5100697927</td>
<td>0</td>
<td>2.3684210526</td>
</tr>
<tr>
<td>11</td>
<td>30</td>
<td>255</td>
<td>1.5</td>
<td>12.7</td>
<td>2.541601935</td>
<td>0</td>
<td>2.3684210526</td>
</tr>
<tr>
<td>12</td>
<td>30</td>
<td>255</td>
<td>2</td>
<td>18.5</td>
<td>2.917707321</td>
<td>0</td>
<td>2.3684210526</td>
</tr>
<tr>
<td>13</td>
<td>30</td>
<td>270</td>
<td>1.5</td>
<td>60.9</td>
<td>4.1092331747</td>
<td>0</td>
<td>17.368421053</td>
</tr>
<tr>
<td>14</td>
<td>45</td>
<td>240</td>
<td>1</td>
<td>4.4</td>
<td>1.4816046409</td>
<td>15</td>
<td>-12.63157895</td>
</tr>
<tr>
<td>15</td>
<td>45</td>
<td>240</td>
<td>1.5</td>
<td>16.5</td>
<td>2.8033603809</td>
<td>15</td>
<td>-12.63157895</td>
</tr>
<tr>
<td>16</td>
<td>45</td>
<td>240</td>
<td>2</td>
<td>24.5</td>
<td>3.1968731176</td>
<td>15</td>
<td>-12.63157895</td>
</tr>
<tr>
<td>17</td>
<td>45</td>
<td>255</td>
<td>1.5</td>
<td>62.8</td>
<td>4.139955035</td>
<td>15</td>
<td>2.3684210526</td>
</tr>
<tr>
<td>18</td>
<td>45</td>
<td>270</td>
<td>1</td>
<td>96.4</td>
<td>4.5685062016</td>
<td>15</td>
<td>17.368421053</td>
</tr>
<tr>
<td>19</td>
<td>45</td>
<td>270</td>
<td>2</td>
<td>102</td>
<td>4.5249728133</td>
<td>15</td>
<td>17.368421053</td>
</tr>
</tbody>
</table>

**Correlations**

<table>
<thead>
<tr>
<th></th>
<th>logyield</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
</tr>
</thead>
<tbody>
<tr>
<td>logyield</td>
<td>1.0000</td>
<td>0.6411</td>
<td><strong>0.6205</strong></td>
<td><strong>0.3827</strong></td>
</tr>
<tr>
<td>x1</td>
<td>0.6411</td>
<td>1.0000</td>
<td>-0.0000</td>
<td>0.0833</td>
</tr>
<tr>
<td>x2</td>
<td>0.6205</td>
<td>-0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>x3</td>
<td>0.3827</td>
<td>0.0833</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

*part a)*