The first Tuesday in November is Election Day in the United States. Prior to the election, candidates are interested in how many people will actually turn out to vote. For the 2004 Presidential Election the number of registered voters (Democratic, Republican and No Party Affiliation) and the Total number of Votes cast in 18 counties in Iowa are listed below.

<table>
<thead>
<tr>
<th>County</th>
<th>Dem Reg</th>
<th>Rep Reg</th>
<th>NP Reg</th>
<th>Total Votes</th>
<th>County</th>
<th>Dem Reg</th>
<th>Rep Reg</th>
<th>NP Reg</th>
<th>Total Votes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6011</td>
<td>4788</td>
<td>7213</td>
<td>13852</td>
<td>10</td>
<td>1639</td>
<td>2472</td>
<td>2922</td>
<td>5381</td>
</tr>
<tr>
<td>2</td>
<td>1601</td>
<td>2334</td>
<td>3288</td>
<td>5395</td>
<td>11</td>
<td>4468</td>
<td>3312</td>
<td>4387</td>
<td>9310</td>
</tr>
<tr>
<td>3</td>
<td>9764</td>
<td>8630</td>
<td>12549</td>
<td>24414</td>
<td>12</td>
<td>8159</td>
<td>9236</td>
<td>9865</td>
<td>19010</td>
</tr>
<tr>
<td>4</td>
<td>3100</td>
<td>2394</td>
<td>3748</td>
<td>6825</td>
<td>13</td>
<td>1739</td>
<td>2574</td>
<td>2787</td>
<td>5455</td>
</tr>
<tr>
<td>5</td>
<td>1406</td>
<td>3541</td>
<td>2564</td>
<td>5491</td>
<td>14</td>
<td>16528</td>
<td>16265</td>
<td>25435</td>
<td>44007</td>
</tr>
<tr>
<td>6</td>
<td>1966</td>
<td>2209</td>
<td>2836</td>
<td>5091</td>
<td>15</td>
<td>1737</td>
<td>2508</td>
<td>3364</td>
<td>5873</td>
</tr>
<tr>
<td>7</td>
<td>2676</td>
<td>3843</td>
<td>4254</td>
<td>8336</td>
<td>16</td>
<td>3352</td>
<td>4685</td>
<td>6342</td>
<td>10764</td>
</tr>
<tr>
<td>8</td>
<td>1685</td>
<td>3520</td>
<td>3188</td>
<td>5951</td>
<td>17</td>
<td>1468</td>
<td>1978</td>
<td>2077</td>
<td>4131</td>
</tr>
<tr>
<td>9</td>
<td>2025</td>
<td>1571</td>
<td>2763</td>
<td>4696</td>
<td>18</td>
<td>2010</td>
<td>3923</td>
<td>3420</td>
<td>6473</td>
</tr>
</tbody>
</table>

Refer to the JMP output, Predicting Total Votes based on Registered Voters.

a) Which explanatory variable Dem Reg, Rep Reg, or NP Reg has the highest correlation with Total Votes? What is that correlation?

NP Reg
Correlation: r = 0.9968

b) Is the correlation in a) statistically significant? Support you answer statistically.

Yes, the P-value is less than 0.0001 which is very small and indicates a statistically significant correlation.

c) What do your answers in a) and b) indicate about multicollinearity?

Nothing. Multicollinearity deals with correlation among explanatory variables not with the correlation between a response and an explanatory variable.
d) Is there multicollinearity among the explanatory variables; Dem Reg, Rep Reg and NP Reg? Support your answer statistically.

   Yes. All of the explanatory variables are highly correlated with each other.

   Dem Reg with Rep Reg  \( r = 0.9637 \)
   Dem Reg with NP Reg  \( r = 0.9809 \)
   Rep Reg with NP Reg  \( r = 0.9759 \)

e) Does NP Reg add significantly to the model that already contains Dem Reg and Rep Reg? Support your answer with the appropriate test of hypothesis.

   Yes. \( t = 8.92, \text{P-value} < 0.0001 \)
   The small P-value indicates that adding NP Reg to the model that already contains Dem Reg and Rep Reg is statistically significant.

f) Does Rep Reg add significantly to the model that already contains NP Reg? Support your answer with the appropriate test of hypothesis.

   No. \( t = 1.94, \text{P-value} = 0.0709 \)
   The P-value is greater than 0.05 which indicates that adding Rep Reg to the model with NP Reg is not statistically significant. Although the value of R\text{Square} has increased it has not increased significantly.

g) Give the Analysis of Variance table for the model that contains Rep Reg and Dem Reg as explanatory variables.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>1668506662</td>
<td>834250331</td>
<td>798.84</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>15664900</td>
<td>1044326.678</td>
<td></td>
</tr>
<tr>
<td>C. Total</td>
<td>17</td>
<td>1684165562</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
h) What is the additional Sum of Squares due to adding Dem Reg to the model that has Rep Reg?

\[ SS_{RepReg} = SS_{RepReg,NPReg} - SS_{NPReg|Rep Reg} = 1675476800 - 56033685 \]
\[ SS_{RepReg} = 1619443115 \]

\[ SS_{DemReg|RepReg} = SS_{DemReg,RepReg} - SS_{Rep Reg} = 1668500662 - 1619443115 \]
\[ SS_{DemReg|RepReg} = 49057547 \]

i) There are seven possible models that can be made using the three explanatory variables: Dem Reg alone, Rep Reg alone, NP Reg alone, Dem Reg with Rep Reg, Dem Reg with NP Reg, Rep Reg with NP Reg, and the three variable model Dem Reg, Rep Reg and NP Reg. Of these seven models is the model that includes all three variables, Dem Reg, Rep Reg and NP Reg the best model? Explain briefly.

Yes.
The model with Dem Reg, Rep Reg and NP Reg is useful (F = 3346.036, P-value < 0.0001).
Each individual variable adds significantly to the model:
Dem Reg: \( t = 6.15, \) P-value < 0.0001
Rep Reg: \( t = 2.63, \) P-value = 0.0198
NP Reg: \( t = 8.92, \) P-value < 0.0001
Because it has the most explanatory variables of the seven possible models it must have the highest value of \( R^2. \)

j) Comment on each of the three plots for the distribution of residuals. What does this indicate about the identically distributed and normally distributed errors conditions?

From the histogram, the distribution of residuals is mounded between 0 and 250 and is skewed left.
From the box plot there is a potential outlier with the residual around 1000.
From the Normal quantile plot, the points start out following the diagonal, normal model, line but then begin to stray a little bit.
Both the identically distributed and normally distributed errors conditions are in doubt.