Statistics 512: Solution to Homework#12

For the next three problems use the Rehabilitation therapy data of problem 22.11 (CH25PR11.DAT) in the text (also see the description in Problem 16.9).

1. Analyze this data using a one-way ANOVA model, ignoring patient age. Summarize your conclusions from this analysis.

Solution: The factor of prior physical fitness is significant in this analysis, with a p-value of $4.13 \times 10^{-5}$. From the graph of days of physical therapy against fitness, we see that variation among fitness levels is large compared to the variation within fitness level, so that it is not surprising that fitness is a significant factor. The $R^2$ is 0.62.

Dependent Variable: days

<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>2</td>
<td>672.000000</td>
<td>336.000000</td>
<td>16.96</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>21</td>
<td>416.000000</td>
<td>19.809524</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>23</td>
<td>1088.000000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square Coeff Var Root MSE days Mean
0.617647 13.90872 4.450789 32.00000

Source fitness  DF  Type I SS Mean Square F Value Pr > F
2 672.000000 336.0000000 16.96 <.0001

Source fitness  DF  Type III SS Mean Square F Value Pr > F
2 672.000000 336.0000000 16.96 <.0001

Figure 1: Plots for Problem 1

The qqplot of residuals (right plot on Figure 1) shows no obvious departures from normality. Residual plots (Figure 2) do not indicate any obvious problems with variance.

Figure 2: Residual plots for Problem 1

2. Analyze the data using a one-way ANCOVA model with patient age as a covariate. Show appropriate graphs and summarize your conclusions from this analysis.

Solution: Both age and fitness are significant in this analysis ($p < 10^{-16}$ and $p = 1.11 \times 10^{-16}$, respectively). All three fitness levels were significantly different from one another, with “below average” fitness levels having the longest average successful physical therapy and “above average” fitness levels having the shortest. The $R^2$ for this analysis is 0.994.

Dependent Variable: days

<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>3</td>
<td>1081.834252</td>
<td>360.611417</td>
<td>1169.72</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Error 20 6.165748 0.308287
Corrected Total 23 1088.000000

R-Square Coeff Var Root MSE days Mean
0.994333 1.735114 0.555236 32.000000

Source DF Type I SS Mean Square F Value Pr > F
age 1 835.7505470 835.7505470 2710.95 <.0001
fitness 2 246.0837050 123.0418525 399.11 <.0001

Source DF Type III SS Mean Square F Value Pr > F
age 1 409.8342521 409.8342521 1329.39 <.0001
fitness 2 246.0837050 123.0418525 399.11 <.0001

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

LSMEAN
fitness days LSMEAN Number
1 34.9504643 1
2 33.1030856 2
3 26.2275715 3

Least Squares Means for effect fitness
Pr > |t| for H0: LS Mean(i)=LS Mean(j)

Dependent Variable: days

i/j 1 2 3
1 <.0001 <.0001
2 <.0001 <.0001
3 <.0001 <.0001

To test the assumption of equal slopes, we re-run with the interaction term:

Sum of
Source DF Squares Mean Square F Value Pr > F
Model 3 1048.421843 349.473948 176.60 <.0001
Error 20 39.578157 1.978908
Corrected Total 23 1088.000000

R-Square Coeff Var Root MSE days Mean
0.963623 4.396052 1.406737 32.000000

Source DF Type I SS Mean Square F Value Pr > F
age 1 835.7505470 835.7505470 422.33 <.0001
fitness 1 210.6765502 210.6765502 106.46 <.0001
age*fitness 1 1.9947453 1.9947453 1.01 0.3274

Source DF Type III SS Mean Square F Value Pr > F
age 1 54.74027210 54.74027210 27.86 <.0001
fitness 1 17.01643528 17.01643528 8.60 0.0082
age*fitness 1 1.99474527 1.99474527 1.01 0.3274

Since the p-value is 0.328 > 0.05, there is insufficient evidence that the lines have
different slopes, so our assumption of equal slopes is reasonable.

3. Explain any differences in your conclusions from the two analyses. (You should say what those differences are and also explain why they happened.)

**Solution:** At first glance, there seems to be little difference between the conclusions of the two analyses. The effect of prior physical status becomes more significant with the covariate included ($p = 1.11 \times 10^{-16}$) than when it is not included ($p = 4.13 \times 10^{-5}$). However, several important differences should be noted with the inclusion of the covariate age. The $MSE$ falls from a value of 19.809 to a value of 0.3083; as a result, predictions about the completion of physical activity should be $\sqrt{\frac{19.809}{0.3083}} \approx 24$ times more accurate when age is known. (It is also worthwhile to note $R^2$ jumps from 0.618 to 0.994 with the addition of the covariate.) Finally, the differences of means for all three groups, while all significant, drop from values of 6, 8, and 14 (below average – average, average – above average, and below average – above average, respectively) to respective values of 1.847, 6.876, and 8.723.

**For the next two problems use the Coil winding data of Problem 25.9 (CH24PR09.DAT) in the text.**

4. Analyze this data using the random effects model. Test the null hypothesis that the mean coil winding characteristic is the same in all machines (i.e. test whether $\sigma^2_\mu = 0$). Interpret the results of your analysis.

**Solution:** The null hypothesis that all machines have the same mean ($\sigma^2_\mu = 0$) is rejected with $p = 1.54 \times 10^{-9}$. We conclude that coil-winding characteristics vary with each machine.

\[
\begin{array}{llllll}
\text{Source} & \text{DF} & \text{Squares} & \text{Mean Square} & \text{F Value} & \text{Pr > F} \\
\text{Model} & 3 & 602.5000000 & 200.8333333 & 28.09 & <.0001 \\
\text{Error} & 36 & 257.4000000 & 7.1500000 \\
\text{Corrected Total} & 39 & 859.9000000 \\
\end{array}
\]

\[
\begin{array}{llllll}
\text{R-Square} & \text{Coeff Var} & \text{Root MSE} & \text{winding Mean} \\
0.700663 & 1.304047 & 2.673948 & 205.0600 \\
\end{array}
\]

\[
\begin{array}{llllll}
\text{Source} & \text{DF} & \text{Type I SS} & \text{Mean Square} & \text{F Value} & \text{Pr > F} \\
\text{machine} & 3 & 602.5000000 & 200.8333333 & 28.09 & <.0001 \\
\end{array}
\]

\[
\begin{array}{llllll}
\text{Source} & \text{Type III Expected Mean Square} \\
\text{machine} & \text{Var(Error) + 10 Var(machine)} \\
\end{array}
\]

5. Give a point estimate of the intraclass correlation coefficient $\frac{\sigma^2_\mu}{\sigma^2_\mu + \sigma^2}$.

**Solution:** The proc varcomp output is

\[
\text{MIVQUE(0) Estimates}
\]

\[
\begin{array}{ll}
\text{Variance Component} & \text{winding} \\
\end{array}
\]
Var(machine)  19.36833
Var(Error)  7.15000

As a result, we have that

\[ \hat{\sigma}_\mu^2 = 19.36 \]
\[ \hat{\sigma}^2 = 7.15 \]

and thus

\[ \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2} = \frac{19.36}{19.36 + 7.15} = 0.730 \]

6. Consider a two-way ANOVA with the fixed-effect factor “prices” and random-effect factor “color scheme”. When specifying the model, please state explicitly whether you are considering a restricted mixed model, or an unrestricted mixed model. Please note the test statistics may be different for the two different models.