1. Make a data frame for use in this example. Include "glasslevel," "phosphorlevel," "tubetype," and "current" variables, treating the first 3 as "factor" variables. Print out the data frame and the default summary() applied to the data frame.

```r
tubetype <- c(rep(1, 3), rep(2, 3), rep(3, 3), rep(4, 3), rep(5, 3), rep(6, 3))
glasslevel <- c(rep(1, 9), rep(2, 9))
phosphorlevel <- rep(c(rep(1, 3), rep(2, 3), rep(3, 3)), 2)
tubetype <- as.factor(tubetype)
glasslevel <- as.factor(glasslevel)
phosphorlevel <- as.factor(phosphorlevel)
GlassPhosphor <- data.frame(tubetype, glasslevel, phosphorlevel, current)
GlassPhosphor
##     tubetype glasslevel phosphorlevel current
## 1         1          1             1     280
## 2         1          1             1     290
## 3         1          1             1     285
## 4         2            1             2     300
## 5         2            1             2     310
## 6         3            1             3     270
## 7         3            1             3     285
## 8         3            1             3     290
## 9         4            2             1     230
## 10        4            2             1     235
## 11        4            2             1     240
## 12        4            2             1     260
## 13        5            2             2     240
## 14        5            2             2     235
## 15        5            2             2     220
## 16        6            2             3     225
## 17        6            2             3     230
## 18        6            2             3     230
```

2. Plot the current requirement against the tubetype variable. Then plot the current requirement against the phosphorlevel variable. Why does the second of these fail to make much sense (fail to be useful)?

Since there are large variance for phosphor level against Current, and the mean for the 3 level almost the same. It tells nothing for us to analyse the data using phosphor against the current.
3 Then do a one-way analysis of the data, using \texttt{lm()} and the "factor" tubetype. Compute Ps using the one-way output and then a "margin of error" to attach to each of the \( r = 6 \) sample means, namely:
Δ = t * \frac{S_p}{\sqrt{m}}

## Call:
## lm(formula = current ~ tubetype, data = GlassPhosphor)
##
## Residuals:
##     Min      1Q  Median       3Q      Max
##-11.67   -5.00    0.00     5.00    15.00
##
## Coefficients:
##             Estimate  Std. Error t value Pr(>|t|)
## (Intercept)  262.222     1.964 133.502  < 2e-16 ***
## tubetype1     22.778     4.392   5.186 0.000227 ***
## tubetype2     39.444     4.392   8.981 1.13e-06 ***
## tubetype3     19.444     4.392   4.427 0.000825 ***
## tubetype4    -27.222     4.392  -6.198 4.60e-05 ***
## tubetype5    -17.222     4.392  -3.921 0.002031 **
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.333 on 12 degrees of freedom
## Multiple R-squared:  0.9461, Adjusted R-squared:  0.9236
## F-statistic: 42.13 on 5 and 12 DF,  p-value: 3.349e-07

1. Use the Hmisc and stats packages to make the kind of interaction plot enhanced with the ±Δ error bars illustrated in class.
In the `lm()` output, exactly how are the "estimated coefficients" related to the 6 sample means? That is, show which b's get added to make each one of the \( \bar{y}_i \)'s.

\[ b_1 + \text{intercept} = \bar{y}_1 \] same for others \( b_2, b_3, b_4, b_5 \).

For \( b_6 \): we can do \( b_1 + b_2 + b_3 + b_4 + b_5 + b_6 = 0 \) since we already know \( b_1 + b_2 + b_3 + b_4 + b_5 \) from the output of `one.out`, it is easy to calculate the \( b_6 + \text{intercept} \).

Using "set the "effects sum to zero" convention" in your notes.

```
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 262.222  1.964 133.502  < 2e-16 ***
## tubetype1     22.778  4.392   5.186 0.000227 ***
## tubetype2     39.444  4.392   8.981 1.13e-06 ***
## tubetype3     19.444  4.392   4.427 0.000825 ***
## tubetype4    -27.222  4.392  -6.198 4.60e-05 ***
## tubetype5    -17.222  4.392  -3.921 0.002031 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 8.333 on 12 degrees of freedom
## Multiple R-squared:  0.9461, Adjusted R-squared:  0.9236
## F-statistic: 42.13 on 5 and 12 DF,  p-value: 3.349e-07
```
6. Now do a two-way analysis of the current requirements using `lm()`. Exactly what are the 6 estimated coefficients produced by the `lm()` call?

\[ \text{intercept} + \text{glasslevel1} \ (\text{indicated in red value below}) \ & \ 
\text{glasslevel1} + \text{glasslevel2} = 0 \]

\[ \text{intercept} + \text{phosphorlevel1} \ & \ 
\text{phosphorlevel1} + \text{phosphorlevel2} + \text{phosphorlevel3} = 0 \]

\[ \text{intercept} + \text{phosphorlevel2} \ & \ 
\text{phosphorlevel1} + \text{phosphorlevel2} + \text{phosphorlevel3} = 0 \]

\[ \text{intercept} + \text{glasslevel1:phosphorlevel1} \]

\[ \text{intercept} + \text{glasslevel1:phosphorlevel2} \]

Using “set the "effects sum to zero" convention” in your notes.

**Coefficients:**

|                      | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------------|----------|------------|---------|----------|
| (Intercept)          | 262.222  | 1.964      | 133.50  | < 2e-16  *** |
| glasslevel1          | 27.222   | 1.964      | 13.86   | 9.57e-09 *** |
| phosphorlevel1       | -2.222   | 2.778      | -0.80   | 0.43926 |
| phosphorlevel2       | 11.111   | 2.778      | 4.00    | 0.00176 ** |
| glasslevel1:phosphorlevel1 | -2.222 | 2.778     | -0.80   | 0.43926 |
| glasslevel1:phosphorlevel2 | 1.111   | 2.778      | 0.40    | 0.69619 |

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.333 on 12 degrees of freedom

Multiple R-squared: 0.9461, Adjusted R-squared: 0.9236

F-statistic: 42.13 on 5 and 12 DF, p-value: 3.349e-07

7. Find a p-value for testing $H_0: \alpha_{ij} = 0$ (i.e. that there are no interactions between glass and phosphor) on the two-way output. In rough terms, how is it related to the confidence intervals for Glass-Phosphor interactions presented in lecture?

The F-statistic 0.32 and the p-value is 0.732158 >0.05, indicating interactions between glass and phosphor.

```r
anova(lm(current~glasslevel+phosphorlevel+glasslevel:phosphorlevel, 
          data=GlassPhosphor))
```

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>glasslevel</td>
<td>1</td>
<td>13338.9</td>
<td>13338.9</td>
<td>192.08</td>
<td>9.568e-09 ***</td>
</tr>
<tr>
<td>phosphorlevel</td>
<td>2</td>
<td>1244.4</td>
<td>622.2</td>
<td>8.96</td>
<td>0.004162 **</td>
</tr>
<tr>
<td>glasslevel:phosphorlevel</td>
<td>2</td>
<td>44.4</td>
<td>22.2</td>
<td>0.32</td>
<td>0.732158</td>
</tr>
<tr>
<td>Residuals</td>
<td>12</td>
<td>833.3</td>
<td>69.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8. Redo the one-way analysis of the current requirements data using dummy variables and a multiple linear regression viewpoint. (See Section 9.3.2 of Vardeman and Jobe.) How are the coefficients produced related to those on the one-way output? How is $S_{sp}$ related to $S_p$?
The problem is solved with the same procedure of problem 5.

\[ S_p = \sqrt{S_{sp}} = 8.333 \]

## Coefficients:

|             | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 262.222  | 1.964      | 133.502 | < 2e-16  *** |
| dumT1       | 22.778   | 4.392      | 5.186   | 0.000227 *** |
| dumT2       | 39.444   | 4.392      | 8.981   | 1.13e-06 *** |
| dumT3       | 19.444   | 4.392      | 4.427   | 0.000825 *** |
| dumT4       | -27.222  | 4.392      | -6.198  | 4.60e-05 *** |
| dumT5       | -17.222  | 4.392      | -3.921  | 0.002031 ** |

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dumT1</td>
<td>1</td>
<td>5400.0</td>
<td>5400.0</td>
<td>77.760</td>
<td>1.36e-06 ***</td>
</tr>
<tr>
<td>dumT2</td>
<td>1</td>
<td>4355.6</td>
<td>4355.6</td>
<td>62.720</td>
<td>4.16e-06 ***</td>
</tr>
<tr>
<td>dumT3</td>
<td>1</td>
<td>277.8</td>
<td>277.8</td>
<td>4.000</td>
<td>0.068655 .</td>
</tr>
<tr>
<td>dumT4</td>
<td>1</td>
<td>3526.7</td>
<td>3526.7</td>
<td>50.784</td>
<td>1.20e-05 ***</td>
</tr>
<tr>
<td>dumT5</td>
<td>1</td>
<td>1067.8</td>
<td>1067.8</td>
<td>15.376</td>
<td>0.002031 **</td>
</tr>
<tr>
<td>Residuals</td>
<td>12</td>
<td>833.3</td>
<td>69.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. **Redo the two-way analysis of the current requirements data using dummy variables and a multiple linear regression viewpoint. How are the coefficients produced related to those on the two-way output? How is \( S_{sp} \) related to \( S_{sp} \)?**

The problem is solved with the same procedure of problem 6.

\[ S_p = \sqrt{S_{sp}} = 8.333 \]

## Coefficients:

|             | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 262.222  | 1.964      | 133.50  | < 2e-16  *** |
| dumA1       | 27.222   | 1.964      | 13.86   | 9.57e-09 *** |
| dumB1       | -2.222   | 2.778      | -0.80   | 0.43926 |
| dumB2       | 11.111   | 2.778      | 4.00    | 0.00176 ** |
| dumA1:dumB1 | -2.222   | 2.778      | -0.80   | 0.43926 |
| dumA1:dumB2 | 1.111    | 2.778      | 0.40    | 0.69619 |

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 8.333 on 12 degrees of freedom
Multiple R-squared:  0.9461, Adjusted R-squared:  0.9236
F-statistic: 42.13 on 5 and 12 DF,  p-value: 3.349e-07


Analysis of Variance Table

Response: current

Df  Sum Sq Mean Sq F value   Pr(>F)
dumA1  1 13338.9 13338.9  192.08 9.568e-09 ***
dumB1  1   133.3   133.3    1.92  0.191070

dumB2  1  1111.1  1111.1   16.00  0.001762 **
dumA1:dumB1  1    33.3    33.3    0.48  0.501611
dumA1:dumB2  1   11.1    11.1    0.16  0.696185
Residuals 12   833.3    69.4

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1