SDS 439 - Homework 05

Due Mar 31, 1:00 pm

Batteries

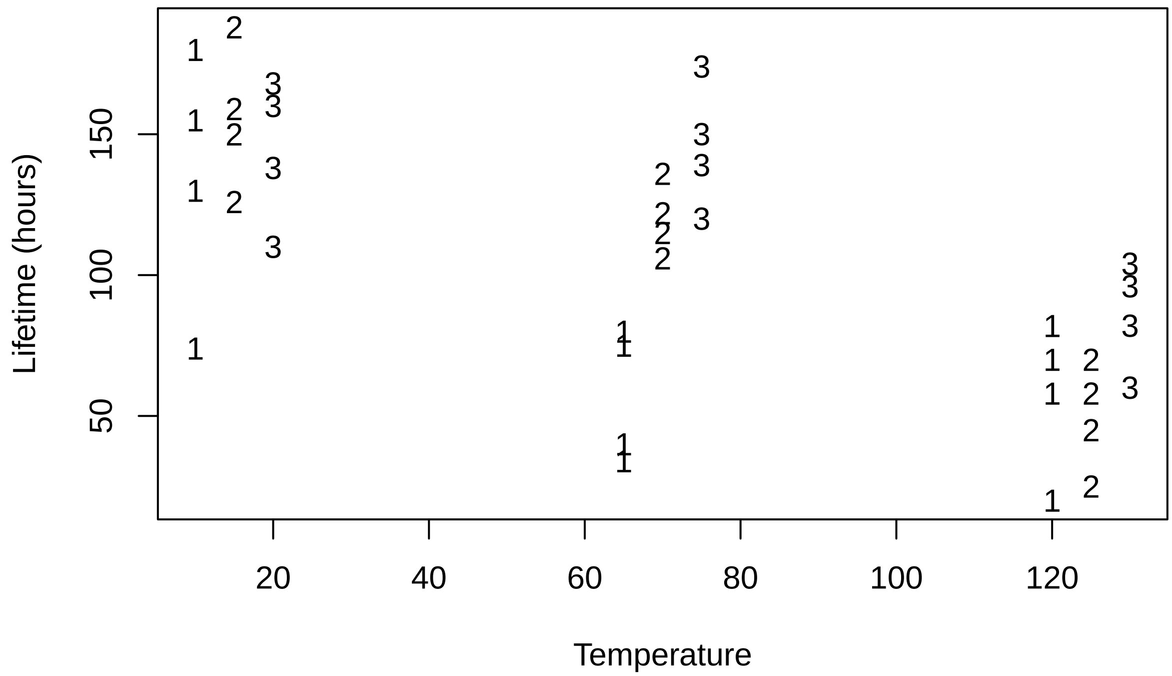
We have a dataset measuring the lifetimes of three different battery types at three different temperatures. At each type x temperature combination, there are four independent observations.

1. Load the data and make an informative plot, showing information about the type, temperature, and lifetimes

```r
dat <- read.table( "../datasets/battery.txt", header = TRUE )
head(dat)
```

```
##   type temp life
## 1    1   15  130
## 2    1   15  155
## 3    1   15   74
## 4    1   15  180
## 5    1   70   34
## 6    1   70   40
```

```r
plot(
    dat$temp + dat$type*5 - 10,
    dat$life, type = "n", xlab = "Temperature",
    ylab = "Lifetime (hours)"
)
text( dat$temp + dat$type*5 - 10, dat$life, dat$type )
```



2. Fit the two-factor additive model, treating lifetime as the response, and type and temperature as the two factors, and print the table.

```r
m1 <- lm( life ~ as.factor(type) + as.factor(temp), data = dat )
summary(m1)
```

```
##
## Call:
## lm(formula = life ~ as.factor(type) + as.factor(temp), data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -54.389 -21.681   2.694  17.215  57.528
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           122.47      11.17  10.965 3.39e-12 ***
## as.factor(type)2       25.17      12.24   2.057  0.04819 *
## as.factor(type)3       41.92      12.24   3.426  0.00175 **
## as.factor(temp)70     -37.25      12.24  -3.044  0.00472 **
## as.factor(temp)125    -80.67      12.24  -6.593 2.30e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.97 on 31 degrees of freedom
## Multiple R-squared:  0.6414, Adjusted R-squared:  0.5951
## F-statistic: 13.86 on 4 and 31 DF,  p-value: 1.367e-06
```

3. Give a brief description of why each parameter estimate "makes sense".

```r
# (Intercept) = 122.47 is expected lifetime for type 1 at 15 degrees, which
# makes sense because that's roughly the average for that combination
# type2 = 25.17 makes sense because battery type 2 generally does better
# than battery type 1
```

```r
# type3 = 41.92 makes sense because battery type 3 does better than both
# battery type 2 and battery type 1
# temp70 = -37.25 makes sense because the lifetimes are generally shorter
# at 70 degrees than at 15 degrees
# temp125 = -80.67 makes sense because the lifetimes are the shortest
# at 125 degrees
```

4. Fit the two-factor interaction model and print the summary table

```r
m2 <- lm( life ~ as.factor(type)*as.factor(temp), data = dat )
summary(m2)
```

```
##
## Call:
## lm(formula = life ~ as.factor(type) * as.factor(temp), data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -60.750 -14.625   1.375  17.937  45.250
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          134.75      12.99  10.371 6.46e-11 ***
## as.factor(type)2                      21.00      18.37   1.143 0.263107
## as.factor(type)3                       9.25      18.37   0.503 0.618747
## as.factor(temp)70                    -77.50      18.37  -4.218 0.000248 ***
## as.factor(temp)125                   -77.25      18.37  -4.204 0.000257 ***
## as.factor(type)2:as.factor(temp)70    41.50      25.98   1.597 0.121886
## as.factor(type)3:as.factor(temp)70    79.25      25.98   3.050 0.005083 **
## as.factor(type)2:as.factor(temp)125  -29.00      25.98  -1.116 0.274242
## as.factor(type)3:as.factor(temp)125   18.75      25.98   0.722 0.476759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.98 on 27 degrees of freedom
## Multiple R-squared:  0.7652, Adjusted R-squared:  0.6956
## F-statistic:    11 on 8 and 27 DF,  p-value: 9.426e-07
```

5. The largest interaction term is between type 3 and temperature 70. Give an interpretation for this term and explain why it makes sense that this is the largest interaction.

```r
# This is [ (type 3, temp 70) - (type 1, temp 70) ] -
#         [ (type 3, temp 15) - (type 1, temp 15) ]
# It makes sense that this estimate is large, because these two types have
# roughly the same lifetimes at 15 degrees, but type 3 outperforms type 1
# at 70 degrees by a large margin.
```

6. Now fit a factor numeric interaction model, where type is the factor, and temperature is a numeric covariate, and print the model summary

```r
m3 <- lm( life ~ as.factor(type)*temp, data = dat )
summary(m3)
```

```
##
## Call:
## lm(formula = life ~ as.factor(type) * temp, data = dat)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.167 -14.458   1.479  14.135  58.208
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           132.3258    15.6188   8.472 1.87e-09 ***
## as.factor(type)2       43.6212    22.0883   1.975 0.057550 .
## as.factor(type)3       29.9848    22.0883   1.358 0.184747
## temp                   -0.7023     0.1878  -3.739 0.000777 ***
## as.factor(type)2:temp  -0.2636     0.2656  -0.993 0.328825
## as.factor(type)3:temp   0.1705     0.2656   0.642 0.525881
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.22 on 30 degrees of freedom
## Multiple R-squared:  0.6702, Adjusted R-squared:  0.6153
## F-statistic: 12.19 on 5 and 30 DF,  p-value: 1.678e-06
```

7. What are the slopes for the three types, and which type has the worst performance with respect to increasing temperature?

```r
# type 1
m3$coefficients["temp"]
```

```
##       temp
## -0.7022727
```

```r
# type 2
m3$coefficients["temp"] + m3$coefficients["as.factor(type)2:temp"]
```

```
##       temp
## -0.9659091
```

```r
# type 3
m3$coefficients["temp"] + m3$coefficients["as.factor(type)3:temp"]
```

```
##       temp
## -0.5318182
```

8. The factor-factor interaction model is different from the factor-numeric interaction model. Explain how they are different and which of the two models is more flexible.

```r
# The factor-numeric interaction model assumes that the effect of
# temperature is linear for each type. Since the temperatures are
# spaced out evenly, this means that the effect of changing from 15 to
# 70 degrees is assumed to be the same as the effect of changing from
# 70 to 125 degrees. The factor-factor interaction model does not impose
# this constraint on the effect of temperature, so it is more flexible.
```

9. Fit a model that has type as factor, and temperature as numeric, but include linear and quadratic effects of temperature, and the linear and quadratic terms both interact with type. Print the model summary.

```r
m4 <- lm( life ~ as.factor(type)*temp + as.factor(type)*I(temp^2), data = dat)
summary(m4)
```

```
##
```

```
## Call:
## lm(formula = life ~ as.factor(type) * temp + as.factor(type) *
##     I(temp^2), data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -60.750 -14.625   1.375  17.938  45.250
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                169.380165  20.567656   8.235 7.66e-09 ***
## as.factor(type)2            -9.756198  29.087058  -0.335  0.73991
## as.factor(type)3           -36.617769  29.087058  -1.259  0.21884
## temp                        -2.501446   0.755148  -3.313  0.00264 **
## I(temp^2)                    0.012851   0.005260   2.443  0.02139 *
## as.factor(type)2:temp        2.328099   1.067941   2.180  0.03815 *
## as.factor(type)3:temp        3.404339   1.067941   3.188  0.00361 **
## as.factor(type)2:I(temp^2)  -0.018512   0.007439  -2.488  0.01929 *
## as.factor(type)3:I(temp^2)  -0.023099   0.007439  -3.105  0.00443 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.98 on 27 degrees of freedom
## Multiple R-squared:  0.7652, Adjusted R-squared:  0.6956
## F-statistic:    11 on 8 and 27 DF,  p-value: 9.426e-07
```

10. Which of the previous models is this model equivalant to? If you've implemented it correctly, it should be equivalent to one of the other models you've fit.

```
# It's equivalent to the factor-factor interaction model. This is because
# the model assumes that each type gets its own quadratic without any
# constraints on how the quadratics from different types are related to one
# another. This allows the quadratic model to match the factor-factor
# interaction fitted values exactly, since a quadratic can be fit perfectly to
# any three points.
```