Example J Factorial experiment on cycles to failure of worsted yarn

Description of data

In an unpublished report to the Technical Committee, International Wool Textile Organization, A. Barella and A. Sust gave the data in the first four columns of `Table J.1`, concerning the number of cycles to failure of lengths of worsted yarn under cycles of repeated loading. The three factors which varied over levels specified in coded form in the first three columns, are

x_1, length of test specimen (250, 300, 350 mm)
x_2, amplitude of loading cycle (8, 9, 10 mm)
x_3, load (40, 45, 50 mm)

Table J.1. Cycles to failure, transformed values, fitted value and residuals

x_1	x_2	x_3	cycles obs	log(cycles) obs	log(cycles) fitted	log(cycles) resid
-1	-1	-1	674	6.51	6.52	-0.01
-1	-1	0	370	5.91	6.11	-0.19
-1	-1	1	292	5.68	5.74	-0.06
-1	0	-1	338	5.82	5.85	-0.03
-1	0	0	266	5.58	5.44	0.14
-1	0	1	210	5.35	5.07	0.28
-1	1	-1	170	5.14	5.26	-0.12
-1	1	0	118	4.77	4.85	-0.07
-1	1	1	90	4.50	4.48	0.02
0	-1	-1	1414	7.25	7.42	-0.17
0	-1	0	1198	7.09	7.01	0.08
0	-1	1	634	6.45	6.64	-0.19
0	0	-1	1022	6.93	6.76	0.17
0	0	0	620	6.43	6.34	0.09
0	0	1	438	6.08	5.97	0.11
0	1	-1	442	6.09	6.16	-0.07
0	1	0	332	5.81	5.75	0.06
0	1	1	220	5.39	5.38	0.02
1	-1	-1	3636	8.20	8.18	0.01
1	-1	0	3184	8.07	7.77	0.29
1	-1	1	2000	7.60	7.40	0.20
1	0	-1	1568	7.36	7.52	-0.16
1	0	0	1070	6.98	7.11	-0.13
1	0	1	566	6.34	6.74	-0.40
1	1	-1	1140	7.04	6.92	0.12
1	1	0	884	6.78	6.51	0.27
1	1	1	360	5.89	6.14	-0.26

x_1, length; x_2, Amplitude of loading cycle; x_3, load
Note: Some values in the last column are slightly different from text.

General considerations

There are number of reasons why use of log cycles to failure is likely to be the most effective way of analyzing these data. Firstly, relationships of the type $y \propto x_1^{\beta_1} x_2^{\beta_2} x_3^{\beta_3}$ are quite commonly found in the physical sciences as reasonably close approximations to empirical behavior. Secondly, the resulting parameters β_1, β_2 and β_3 are dimensionless and thus, especially if they are close to simple integers, relatively easy to interpret. Thirdly, provided that the signs of β_1, β_2 and β_3 are appropriate, sensible limiting behavior as the x's tends to 0 and infinity is achieved. All these points concern the form of the systematic variation.

As for the random variation, again a log transformation is likely to be sensible. Cycles to failure vary over a very wide range (by a factor of over 40 in fact) and the amount of random variation is likely to increase with the mean cycles to failure. More specifically, under one of the physically simplest hypothesis the effect of changing factor levels is to multiply the ‘lifetime’ of a particular individual by a constant. This, sometimes called the central assumption of accelerated life testing, implies that the coefficient of variation of cycles to failure, Y, is constant and thus that the standard deviation of $\log(Y)$ is constant.

While these two lines of argument suggest on general grounds taking $\log(Y)$ as response and $\log(x_1)$, $\log(x_2)$ and $\log(x_3)$ as explanatory variables, of course an empirical test of the suitability of this analysis is still needed.

The balanced nature of the experimental design has two closely related consequences. One is that least-squares fitting of various models representing 1^{st}–degree, 2^{nd}–degree, etc., regression of $\log(Y)$ on $\log(x_1)$, $\log(x_2)$ and $\log(x_3)$ is computationally very simple. The other is that the general form of the systematic variation can be studied very easily and directly from appropriate mean values collected in two-way and one-way tables, as for other forms of balanced factorial experiment. While the final summary of conclusions is likely to be primarily in terms of a fitted regression equation, the explanatory variables being quantitative in nature, critical inspection of two-way tables is all the same an important intermediate step in the analysis.

The analysis

Table J.2 gives two-way and marginal means of log cycles to failure; throughout the natural logs are used. The two-way tables show little evidence of interaction and the marginal means show that the variation with factor levels is predominantly linear; the factor levels are not quite equally spaced in terms of $\log(x)$.

Table J.2 Two-way and one-way means

Load x_3	Amplitude of loading cycle x_2: -1	0	1
-1	7.32	6.70	6.09
0	7.02	6.33	5.79
1	6.58	5.92	5.26
mean	6.97	6.32	5.71

Load x_3	Length of specimen x_1: -1	0	1	mean
-1	5.82	6.76	7.53	6.70
0	5.42	6.44	7.28	6.38
1	5.17	5.98	6.61	5.92
mean	5.47	6.39	7.14	

Amplitude of loading cycle x_2	Length of specimen x_1: -1	0	1	mean
-1	6.03	6.93	7.96	6.97
0	5.58	6.48	6.89	6.32
1	4.80	5.76	6.57	5.71
mean	5.47	6.39	7.14	

Extraction of linear components of main effects, equivalent to the linear model

$$\log(Y) = \alpha + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \beta_3 \log(x_3) + \varepsilon \qquad \text{(J.1)}$$

is done either by direct least-squares fitting, or equivalently by extracting the linear regression component from the marginal means. There results

$$\hat{\beta}_1 = 4.950, \qquad \hat{\beta}_2 = -5.654, \qquad \hat{\beta}_3 = -3.503$$

As already noted, inspection of Table J.1 shows that the model in Equation (J.1) is likely to account for most of the systematic variation. To examine this in more detail, six more degrees of freedom have been isolated, i.e. six more parameters added to equation (J.1). These are respectively linear-by-linear interactions, i.e. product terms such as $\beta_{23}\log(x_2)\log(x_3)$ and pure quadratic terms such as $\beta_{11}(\log x_1)^2$, taken for convenience in a form orthogonalized with respect to the parameters in Equation (J.1).

Table J.3 gives the analysis of variance. The total contribution of quadratic terms has a mean square rather less than that for residual. It is immaterial whether the error of the estimates (J.2) is obtained via the residual mean square from the linear model (J.1) or from the residual mean square of the extended model with quadratic terms. To be slightly cautious, the second and rather larger values has been taken, giving a residual standard deviation of 0.1941and estimated standard error for (J.2) of

$$0.2745,\ 0.4118,\ 0.4118$$

with 17 degrees of freedom.

Table J.3 Analysis of variance

	df	ss	ms
Length of test specimen,x1 (linear)	1	12.5141	
Amplitude of loading cycle x2 (linear)	1	7.1695	
Load, x3(linear)	1	2.7524	
Ampl.(linear)*load(linear)	1	0.0050	
Load(linear*length(linear)	1	0.0517	
Length(linear)*ampl.(linear)	1	0.0203	
Length(quadratic)	1	0.0015	
Amplitude(quadratic)	1	0.0007	
Load(quadratic)	1	0.0478	
Total second-degree terms	6	0.1270	0.021167
Residual	17	0.6407	0.0377
Total	26	22.5630	

The near equality of $\hat{\beta}_1$ and $-\hat{\beta}_2$ suggests that the dependence on x_1 and x_2 can be expressed in terms of x_2 / x_1, and this is particularly appealing on dimensional grounds because both x_2 and x_1 are lengths. The composite variable x_2 / x_1 is the fractional extension of the loading cycle. It would, however, not be correct to argue that by dimensional analysis any dependence can only be on the dimensionless variable x_2 / x_1, because there are other lengths implied in the problem, notably the mean fiber length.

The data are in fact quite closely fitted by the simple relationship

$$y \propto \left(\frac{x_2}{x_1}\right)^{-5} {x_3}^{-7/2}$$

It would be instructive to compare the residual standard deviation of 0.194, corresponding to a coefficient of variation of about 20%, with any value that might be available for repeat tests under the same conditions. A graph of the residuals versus fitted values gives no evidence that the error of log cycles to failure varies with the mean response: thus the data seem reasonably consistent with the central assumption of accelerated life testing.

Additional Notes

(1) The text says, “6 more degrees of freedom have been isolated, i.e. six more parameters added to equation (J.1). These are respectively linear-by-linear interactions, i.e. product terms and pure quadratic

terms, taken for convenience in a form orthogonalized with respect to the parameters in equation (J.1)." The extended model stated here can be expressed as,

$$\begin{aligned}\log(Y) = \alpha + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \beta_3 \log(x_3) + \beta_{23}\big(\log(x_2) - \overline{\log(x_2)}\big)\big(\log(x_3) - \overline{\log(x_3)}\big) \\ + \beta_{31}\big(\log(x_3) - \overline{\log(x_3)}\big)\big(\log(x_1) - \overline{\log(x_1)}\big) + \beta_{12}\big(\log(x_1) - \overline{\log(x_1)}\big)\big(\log(x_2) - \overline{\log(x_2)}\big) \\ +\beta_{11}\big(\log(x_1) - \overline{\log(x_1)}\big)^2 + \beta_{22}\big(\log(x_2) - \overline{\log(x_2)}\big)^2 + \beta_{33}\big(\log(x_3) - \overline{\log(x_3)}\big)^2 + \varepsilon\end{aligned}$$

The logarithms of each factor levels, for example, for factor A the log levels `log(250)`, `log(300)`, `log(350)`, are not equally spaced. It is very complicated to fit orthogonal polynomials when observations of predictor variables are not equally spaced. Fortunately, I have verified that `R` fits orthogonal polynomials automatically. See supplements for more detailed discussion.

(2) The text says, "The total contribution of quadratic terms has a mean square rather less than that for residual. It is immaterial whether the error of the estimates (J.2) is obtained via the residual mean square from the linear model (J.1) or from the residual mean square of the extended model with quadratic terms." To understand this point, see the following table for a comparison of the standard errors of the estimates from model (J.1) and from the extended model stated above.

Coeff	Linear model		Extended model	
	Estimate	Std.Error	Estimate	Std.Error
β_1	4.9504	0.2557	4.9424	0.2745
β_2	-5.6537	0.3858	-5.6591	0.4118
β_3	-3.5030	0.3858	-3.5477	0.4118

(3) The text gives the following relationship

$$y \propto \left(\frac{x_2}{x_1}\right)^{-5} {x_3}^{-7/2}$$

What the text means here actually regards a question of how to interpret the results meaningfully. How to get this simple relationship? From (J.1) we have the fitted function

$$\begin{aligned}\log(\hat{Y}) &= c + 4.9504 \log(x_1) - 5.6537 \log(x_2) - 3.530 \log(x_3) \\ &\approx c + 5 \log(x_1) - 5 \log(x_2) - 3.5 \log(x_3) \\ &= c - 5 \log\left(\frac{x_2}{x_1}\right) - 3.5 \log(x_3) \propto \left(\frac{x_2}{x_1}\right)^{-5} {x_3}^{-7/2}\end{aligned}$$

Appendix: R Code

```
########################
# R code for Example J
########################

#-----------------------------------------------------
# Producing fitted values and residuals in Table J.1
#-----------------------------------------------------

x1 = c(rep(-1,9), rep(0,9), rep(1,9))
x2 = rep(c(rep(-1,3), rep(0,3), rep(1,3)),3)
x3 = rep(c(-1,0,1), 9)
cycles = c(674,370,292,338,266,210,170,118,90,1414,1198,634,1022,620,
           438,442,332,220,3636,3184,2000,1568,1070,566,1140,884,360)
len = x1 * 50 + 300
amp = x2 * 1  + 9
load = x3 * 5 + 45
loglen = log(len)
logamp = log(amp)
logload = log(load)
logcyc.obs = log(cycles)

# fit the model (J.1)

logcyc.lm = lm(logcyc.obs ~ loglen + logamp + logload)
logcyc.fit = logcyc.lm$fitted               # fitted values
resid = logcyc.obs - logcyc.fit             # residuals

# the data frame corresponding to Table J.1

J1 = data.frame(x1, x2, x3, cycles, logcyc.obs, logcyc.fit, resid)

# coefficients of multiple regression

summary(logcyc.lm)
```

```
> summary(logcyc.lm)
Call:
lm(formula = logcyc.obs ~ loglen + logamp + logload)

Residuals:
     Min       1Q   Median       3Q      Max
-0.39866 -0.12636  0.01380  0.11180  0.29365

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.8646     2.2334   1.730    0.097 .
loglen        4.9504     0.2557  19.363 9.78e-16 ***
logamp       -5.6537     0.3858 -14.656 3.72e-13 ***
logload      -3.5030     0.3858  -9.081 4.56e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1827 on 23 degrees of freedom
Multiple R-Squared: 0.9669,     Adjusted R-squared: 0.9626
F-statistic: 224.1 on 3 and 23 DF,  p-value: < 2.2e-16
```

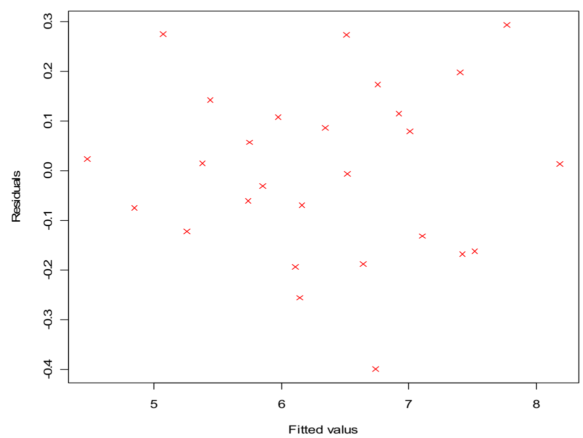
```
anova(logcyc.lm)
```

```
> anova(logcyc.lm)
Analysis of Variance Table
Response: logcyc.obs
          Df  Sum Sq Mean Sq F value    Pr(>F)
loglen     1 12.5141 12.5141 374.924 9.781e-16 ***
```

```
logamp     1  7.1695  7.1695 214.800 3.716e-13 ***
logload    1  2.7524  2.7524  82.463 4.563e-09 ***
Residuals 23  0.7677  0.0334
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# plot of residual vs fitted values
# This graph no evidence that the error of log cycles to failure varies
# with the mean response, thus the data seem reasonably consistent with
# the central assumption of accelerated life testing.

plot(logcyc.lm$fitted, logcyc.lm$residuals, pch=4, col='red',
     xlab='Fitted valus', ylab='Residuals')
```




```
#----------------------
# producing Table J.2
#----------------------

# two-way table for means(rows represent Load, column represent Amplitude)

two32 = tapply(logcyc.obs, list(factor(x3), factor(x2)), mean)

# marginal means log(cycles) relative to levels of Amplitude

one2  = tapply(logcyc.obs, list(factor(x2)), mean)

# Two-way table for means(rows represent Load, columns represent Length)

two31 = tapply(logcyc.obs, list(factor(x3), factor(x1)), mean)

# marginal means log(cycles) relative to levels of Length

one1  = tapply(logcyc.obs, list(factor(x1)), mean)

# marginal means log(cycles) relative to levels of Load

one3  = tapply(logcyc.obs, list(factor(x3)), mean)

# Two-way table for means(rows represent Amptitude, columns represent Length)
```

```
two21 = tapply(logcyc.obs, list(factor(x2), factor(x1)), mean)

print(two32);print(two31);print(two21)
print(one2);print(one1);print(one3)
```

```
> print(two32)
         -1        0        1
-1 7.322016 6.703373 6.088631
0  7.022602 6.329543 5.786759
1  6.576568 5.922640 5.259847
> print(two31)
         -1        0        1
-1 5.824025 6.758335 7.531660
0  5.422561 6.441088 7.275255
1  5.174557 5.975965 6.608534
> print(two21)
         -1        0        1
-1 6.034496 6.931545 7.955145
0  5.584550 6.480485 6.890521
1  4.802098 5.763357 6.569782
> print(one2)
      -1        0        1
6.973729 6.318519 5.711746
> print(one1)
      -1        0        1
5.473714 6.391796 7.138483
> print(one3)
      -1        0        1
6.704673 6.379635 5.919685>
```

```
#-----------------------
# reproducing Table J.3
#-----------------------

quad.lm = lm(logcyc.obs ~ loglen + logamp + logload +
            I(loglen*logamp) + I(loglen*logload) + I(logamp*logload) +
            I(loglen^2) + I(logamp^2) + I(logload^2))
summary(quad.lm)
anova(quad.lm)
```

```
> summary(quad.lm)
Call:
lm(formula = logcyc.obs ~ loglen + logamp + logload + I(loglen *
    logamp) + I(loglen * logload) + I(logamp * logload) + I(loglen^2) +
    I(logamp^2) + I(logload^2))

Residuals:
      Min        1Q    Median        3Q       Max
-0.304034 -0.112395 -0.005587  0.116521  0.266156

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         -239.1097   157.2494  -1.521    0.147
loglen                29.4930    34.6778   0.850    0.407
logamp                16.8684    36.9321   0.457    0.654
logload               74.6512    52.3750   1.425    0.172
I(loglen * logamp)    -2.1849     2.9806  -0.733    0.474
I(loglen * logload)   -3.4905     2.9806  -1.171    0.258
I(logamp * logload)   -1.6384     4.4973  -0.364    0.720
I(loglen^2)           -0.5695     2.8232  -0.202    0.843
I(logamp^2)           -0.8791     6.3898  -0.138    0.892
I(logload^2)          -7.1965     6.3898  -1.126    0.276
Residual standard error: 0.1941 on 17 degrees of freedom
Multiple R-Squared: 0.9724,     Adjusted R-squared: 0.9578
F-statistic: 66.52 on 9 and 17 DF,  p-value: 1.776e-11

> anova(quad.lm)
Analysis of Variance Table
Response: logcyc.obs
                    Df  Sum Sq Mean Sq  F value    Pr(>F)
loglen               1 12.5141 12.5141 332.0444 1.363e-12 ***
logamp               1  7.1695  7.1695 190.2336 1.164e-10 ***
logload              1  2.7524  2.7524  73.0320 1.469e-07 ***
```

```
I(loglen * logamp)   1  0.0203  0.0203   0.5373    0.4735
I(loglen * logload)  1  0.0517  0.0517   1.3714    0.2577
I(logamp * logload)  1  0.0050  0.0050   0.1327    0.7201
I(loglen^2)          1  0.0015  0.0015   0.0407    0.8425
I(logamp^2)          1  0.0007  0.0007   0.0189    0.8922
I(logload^2)         1  0.0478  0.0478   1.2684    0.2757
Residuals           17  0.6407  0.0377
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
newquad.lm = lm(logcyc.obs ~ loglen + logamp + logload +
                I((loglen - mean(loglen)) * (logamp - mean(logamp))) +
                I((loglen - mean(loglen)) * (logload - mean(logload))) +
                I((logamp - mean(logamp)) * (logload - mean(logload))) +
                I((loglen - mean(loglen))^2) + I((logamp - mean(logamp))^2) +
                I((logload - mean(logload))^2) )
summary(newquad.lm)
anova(newquad.lm)
```

```
> summary(newquad.lm)

Call:
lm(formula = logcyc.obs ~ loglen + logamp + logload + I((loglen -
    mean(loglen)) * (logamp - mean(logamp))) + I((loglen - mean(loglen)) *
    (logload - mean(logload))) + I((logamp - mean(logamp)) *
    (logload - mean(logload))) + I((loglen - mean(loglen))^2) +
    I((logamp - mean(logamp))^2) + I((logload - mean(logload))^2))

Residuals:
      Min        1Q    Median        3Q       Max
-0.304034 -0.112395 -0.005587  0.116521  0.266156

Coefficients:
                                                       Estimate Std. Error t value Pr(>|t|)
(Intercept)                                              4.1697     2.4023   1.736    0.101
loglen                                                   4.9424     0.2745  18.002 1.66e-12 ***
logamp                                                  -5.6591     0.4118 -13.742 1.23e-10 ***
logload                                                 -3.5477     0.4118  -8.615 1.31e-07 ***
I((loglen - mean(loglen)) * (logamp - mean(logamp)))    -2.1849     2.9806  -0.733    0.474
I((loglen - mean(loglen)) * (logload - mean(logload)))  -3.4905     2.9806  -1.171    0.258
I((logamp - mean(logamp)) * (logload - mean(logload)))  -1.6384     4.4973  -0.364    0.720
I((loglen - mean(loglen))^2)                            -0.5695     2.8232  -0.202    0.843
I((logamp - mean(logamp))^2)                            -0.8791     6.3898  -0.138    0.892
I((logload - mean(logload))^2)                          -7.1965     6.3898  -1.126    0.276
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1941 on 17 degrees of freedom
Multiple R-Squared: 0.9724,     Adjusted R-squared: 0.9578
F-statistic: 66.52 on 9 and 17 DF,  p-value: 1.776e-11
```

```
> anova(newquad.lm)
Analysis of Variance Table

Response: logcyc.obs
                                                       Df  Sum Sq Mean Sq  F value    Pr(>F)
loglen                                                  1 12.5141 12.5141 332.0444 1.363e-12 ***
logamp                                                  1  7.1695  7.1695 190.2336 1.164e-10 ***
logload                                                 1  2.7524  2.7524  73.0320 1.469e-07 ***
I((loglen - mean(loglen)) * (logamp - mean(logamp)))    1  0.0203  0.0203   0.5373    0.4735
I((loglen - mean(loglen)) * (logload - mean(logload)))  1  0.0517  0.0517   1.3714    0.2577
I((logamp - mean(logamp)) * (logload - mean(logload)))  1  0.0050  0.0050   0.1327    0.7201
I((loglen - mean(loglen))^2)                            1  0.0015  0.0015   0.0407    0.8425
I((logamp - mean(logamp))^2)                            1  0.0007  0.0007   0.0189    0.8922
I((logload - mean(logload))^2)                          1  0.0478  0.0478   1.2684    0.2757
Residuals                                              17  0.6407  0.0377
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```