# MAST30027: Modern Applied Statistics

Assignment 3 Solution

The `wavesolder` data has 48 observations of `y`, the number of defects, and seven predictor variables, `prebake`, `flux`, `speed`, `preheat`, `cooling`, `agitator` and `temp`. The experimental design only has 16 different factor combinations, which restricts what we can estimate. We have 1 intercept, 7 main effects, 21 two-ways interactions, 35 three-way interactions, etc. Of these we can only estimate 1 intercept, 7 main effects, and 7 two-way interactions. (The other interactions are in fact indistinguishable from these.)

The data is taken from Condra, Lloyd, *Reliability Improvement with Design of Experiment*. CRC Press, 2001.

Firstly we need to combine the three replicates into a single data set, and then have a look at the data.

```
> rm(list=ls())
> library(faraway)
> data(wavesolder)
> y <- c(wavesolder$y1, wavesolder$y2, wavesolder$y3) # y is the vector of responses
> wavesolder <- rbind(wavesolder, wavesolder, wavesolder) # design is replicated three times
> wavesolder <- wavesolder[-(1:3)] # remove y1, y2 and y3
> wavesolder$y <- y # add y
> par(mfrow=c(3,3), mar=c(4,4,1,1))
> plot(y ~ prebake, wavesolder)
> plot(y ~ flux, wavesolder)
> plot(y ~ speed, wavesolder)
> plot(y ~ preheat, wavesolder)
> plot(y ~ cooling, wavesolder)
> plot(y ~ agitator, wavesolder)
> plot(y ~ temp, wavesolder)
```
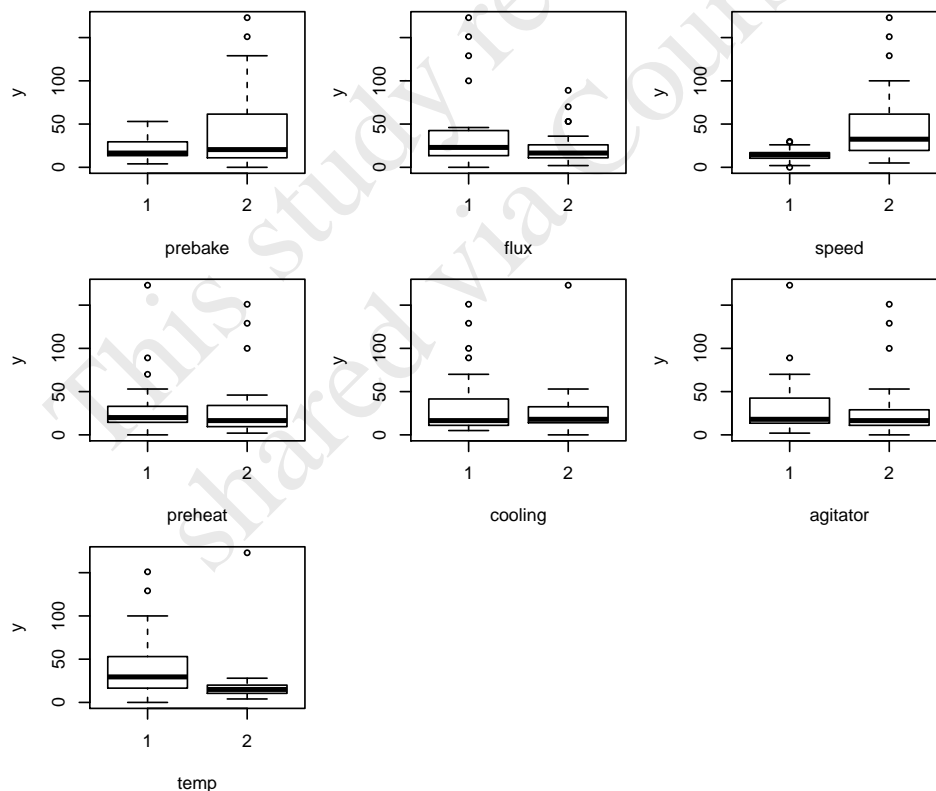

For variables `speed` and `temp` there is a clear effect on the number of defects, but for the other variables things are less clear. Given that we have count data, a Poisson regression model with a log link is a good place

to start. We want to find which factors and two-way interactions are significant. Because we can not estimate all the two-way interactions at once, I chose to do this by fitting the main effects first, then using `step` with a scope that included all the two-way interactions.

If you try fitting all the two-way interactions R just does what it can, and you can then use `step` to see which if any are dropped. This gives a model which on the face of it looks different, because it chooses different interactions, but for this experimental design some of the two-way interactions are indistinguishable, and in fact the models are equivalent.

```
> pmod <- glm(y ~ prebake + flux + speed + preheat + cooling + agitator + temp,
+             family=poisson, data=wavesolder)
> pmod2 <- step(pmod, scope = ~ (prebake + flux + speed + preheat + cooling + agitator + temp)^2,
+               trace=0)
> summary(pmod2)

Call:
glm(formula = y ~ prebake + flux + speed + preheat + cooling +
    agitator + temp + prebake:speed + prebake:cooling + prebake:flux +
    cooling:agitator + flux:agitator, family = poisson, data = wavesolder)

Deviance Residuals:
    Min      1Q  Median      3Q     Max
-7.4791  -1.4524  -0.2013   1.0106  10.0123

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         3.05656    0.11055  27.650  < 2e-16 ***
prebake2            0.56135    0.13618   4.122 3.75e-05 ***
flux2              -0.30219    0.11745  -2.573   0.0101 *
speed2              0.44570    0.09861   4.520 6.20e-06 ***
preheat2           -0.30407    0.06269  -4.851 1.23e-06 ***
cooling2            0.63675    0.11691   5.446 5.14e-08 ***
agitator2           0.01135    0.10478   0.108   0.9137
temp2              -0.60085    0.06686  -8.987  < 2e-16 ***
prebake2:speed2     1.05679    0.13390   7.892 2.97e-15 ***
prebake2:cooling2  -0.87435    0.13453  -6.499 8.08e-11 ***
prebake2:flux2     -0.53583    0.12394  -4.323 1.54e-05 ***
cooling2:agitator2 -0.56100    0.13320  -4.212 2.53e-05 ***
flux2:agitator2     0.31588    0.13249   2.384   0.0171 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1450.52  on 47  degrees of freedom
Residual deviance:  329.89  on 35  degrees of freedom
AIC: 586.63

Number of Fisher Scoring iterations: 5
```
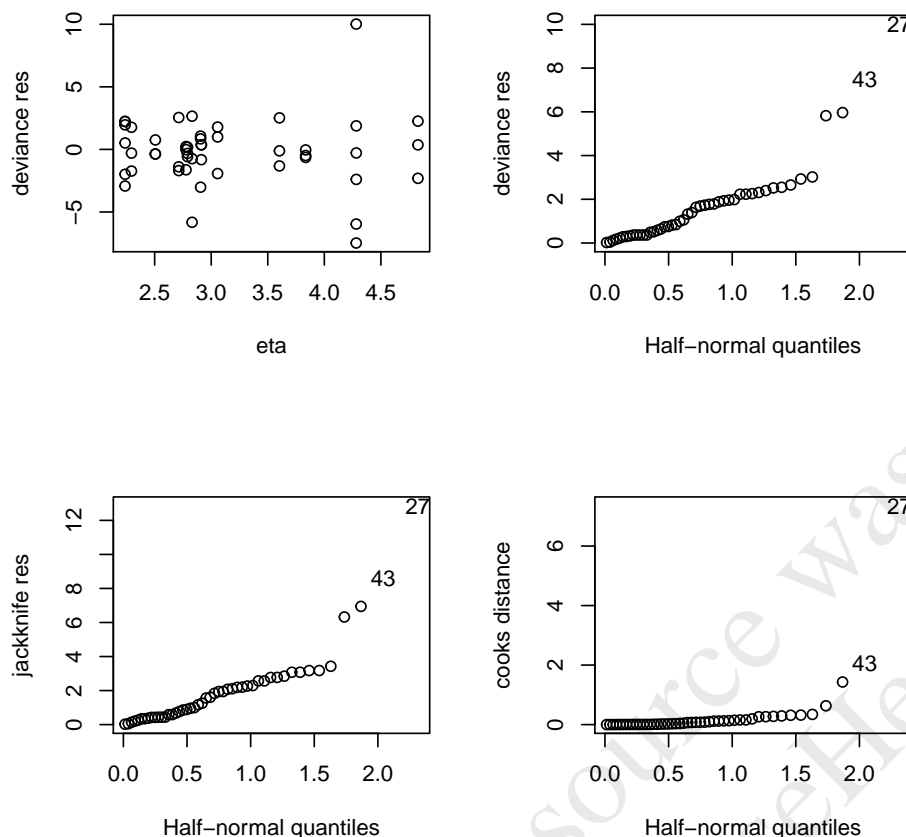
Next we perform some diagnostics. From the Cook's distance in particular, it looks like observation 27 is problematic.

```
> par(mfrow=c(2,2))
> plot(predict(pmod2), residuals(pmod2), xlab="eta", ylab="deviance res")
> halfnorm(residuals(pmod2), ylab="deviance res")
> halfnorm(rstudent(pmod2), ylab="jackknife res")
> halfnorm(cooks.distance(pmod2), ylab="cooks distance")
```

2

We refit the model omitting observation 27, and we see that the fit does change substantially, as now only four interactions are included. (We don't pay too much attention to which ones, because two-thirds of them are indistinguishable from the other third.)

```
> pmod3 <- glm(y ~ prebake + flux + speed + preheat + cooling +
+              agitator + temp + preheat:agitator + prebake:cooling + preheat:cooling +
+              cooling:agitator + prebake:temp, family=poisson, subset=-27, data=wavesolder)
> pmod4 <- step(pmod3, scope = ~ (prebake + flux + speed + preheat + cooling + agitator + temp)^2,
+              trace=0)
> summary(pmod4)

Call:
glm(formula = y ~ prebake + flux + speed + preheat + cooling +
    agitator + temp + preheat:agitator + prebake:cooling + cooling:agitator +
    flux:cooling, family = poisson, data = wavesolder, subset = -27)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.1382  -1.1562   0.1280   0.9379   3.7482

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      3.12852    0.10861  28.806  < 2e-16 ***
prebake2         0.84362    0.09119   9.251  < 2e-16 ***
flux2           -0.50503    0.08594  -5.877 4.18e-09 ***
speed2           0.80630    0.06875  11.727  < 2e-16 ***
preheat2        -0.49449    0.09629  -5.136 2.81e-07 ***
cooling2         0.33309    0.13221   2.519 0.011757 *
agitator2       -0.21245    0.10944  -1.941 0.052234 .
temp2           -0.75540    0.06836 -11.051  < 2e-16 ***
```

3

```
preheat2:agitator2   0.76089    0.13456   5.655 1.56e-08 ***
prebake2:cooling2   -1.18692    0.13669  -8.683  < 2e-16 ***
cooling2:agitator2  -0.32559    0.13442  -2.422 0.015429 *
flux2:cooling2       0.45211    0.12778   3.538 0.000403 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1137.32  on 46  degrees of freedom
Residual deviance:  135.62  on 35  degrees of freedom
AIC: 383.37


Number of Fisher Scoring iterations: 5
```

The residuals for the restricted model look OK, but the deviance is quite high.

```
> pchisq(deviance(pmod4), pmod4$df.residual, lower.tail=FALSE)
```

```
[1] 8.964612e-14
```

```
> par(mfrow=c(2,2))
> plot(predict(pmod4), residuals(pmod4), xlab="eta", ylab="deviance res")
> halfnorm(residuals(pmod4), ylab="deviance res")
> halfnorm(rstudent(pmod4), ylab="jackknife res")
> halfnorm(cooks.distance(pmod4), ylab="cooks distance")
```
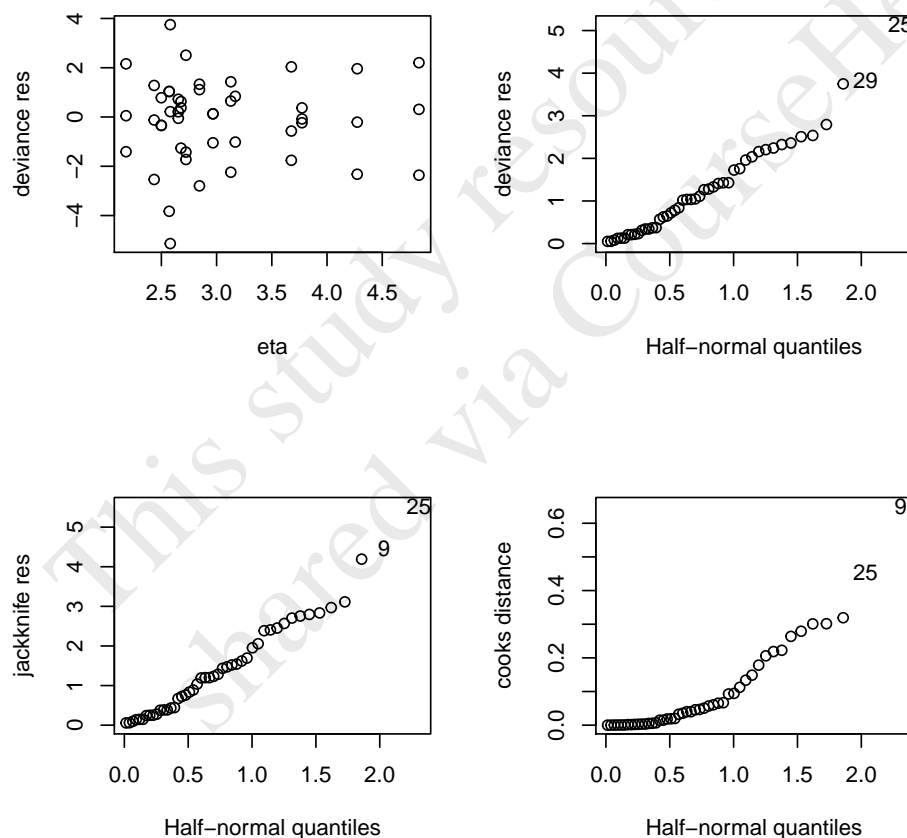

We check for overdispersion and find it. We then question the significance of all the terms in our model, using F-tests at the 5% significance level. Note that we drop interactions before considering main effects, and if an interaction is present we retain the corresponding main effects even if they are not significant. We find that two more interactions drop out.

```
> (phi <- sum(residuals(pmod4, type="pearson")^2/pmod4$df.residual))
```

```
[1] 3.42695
> pmod5 <- glm(y ~ prebake + flux + speed + preheat + cooling + agitator + temp +
+             preheat:agitator + cooling:agitator + cooling:prebake + cooling:flux,
+             family=quasipoisson, subset=-27, data=wavesolder)
> anova(pmod5, test="F")

Analysis of Deviance Table

Model: quasipoisson, link: log

Response: y

Terms added sequentially (first to last)


                 Df Deviance Resid. Df Resid. Dev      F    Pr(>F)
NULL                               46    1137.32
prebake           1    94.11        45    1043.21  27.4630 7.747e-06 ***
flux              1    50.90        44     992.30  14.8537 0.0004752 ***
speed             1   391.84        43     600.46 114.3404 1.427e-12 ***
preheat           1     1.89        42     598.57   0.5521 0.4624241
cooling           1    65.59        41     532.98  19.1387 0.0001042 ***
agitator          1     2.23        40     530.75   0.6517 0.4249713
temp              1   279.01        39     251.74  81.4168 1.163e-10 ***
preheat:agitator  1    24.72        38     227.02   7.2134 0.0109893 *
cooling:agitator  1     4.73        37     222.29   1.3812 0.2478307
prebake:cooling   1    74.08        36     148.21  21.6159 4.606e-05 ***
flux:cooling      1    12.59        35     135.62   3.6736 0.0634760 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> pmod6 <- glm(y ~ prebake + flux + speed + preheat + cooling + agitator + temp +
+             preheat:agitator + cooling:prebake + cooling:flux,
+             family=quasipoisson, subset=-27, data=wavesolder)
> anova(pmod6, test="F")

Analysis of Deviance Table

Model: quasipoisson, link: log

Response: y

Terms added sequentially (first to last)


                 Df Deviance Resid. Df Resid. Dev      F    Pr(>F)
NULL                               46    1137.32
prebake           1    94.11        45    1043.21  27.9592 6.222e-06 ***
flux              1    50.90        44     992.30  15.1220 0.0004166 ***
speed             1   391.84        43     600.46 116.4062 7.856e-13 ***
preheat           1     1.89        42     598.57   0.5621 0.4583017
cooling           1    65.59        41     532.98  19.4845 8.856e-05 ***
agitator          1     2.23        40     530.75   0.6634 0.4207052
temp              1   279.01        39     251.74  82.8878 7.165e-11 ***
preheat:agitator  1    24.72        38     227.02   7.3437 0.0102416 *
prebake:cooling   1    73.24        37     153.78  21.7592 4.166e-05 ***
flux:cooling      1    12.30        36     141.48   3.6534 0.0639404 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> pmod7 <- glm(y ~ prebake + flux + speed + preheat + cooling + agitator + temp +
+             preheat:agitator + cooling:prebake,
+             family=quasipoisson, subset=-27, data=wavesolder)
> anova(pmod7, test="F")
```

```
Analysis of Deviance Table

Model: quasipoisson, link: log

Response: y

Terms added sequentially (first to last)


                Df Deviance Resid. Df Resid. Dev       F    Pr(>F)
NULL                           46     1137.32
prebake          1    94.11    45     1043.21  25.9920 1.041e-05 ***
flux             1    50.90    44      992.30  14.0580 0.0006052 ***
speed            1   391.84    43      600.46 108.2158 1.547e-12 ***
preheat          1     1.89    42      598.57   0.5225 0.4743151
cooling          1    65.59    41      532.98  18.1136 0.0001363 ***
agitator         1     2.23    40      530.75   0.6168 0.4372555
temp             1   279.01    39      251.74  77.0557 1.412e-10 ***
preheat:agitator 1    24.72    38      227.02   6.8270 0.0129029 *
prebake:cooling  1    73.24    37      153.78  20.2282 6.579e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(pmod7)

Call:
glm(formula = y ~ prebake + flux + speed + preheat + cooling +
    agitator + temp + preheat:agitator + cooling:prebake, family = quasipoisson,
    data = wavesolder, subset = -27)

Deviance Residuals:
    Min      1Q   Median       3Q      Max
-5.7575  -1.2948   0.0038   1.2097   2.7561

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         3.0618     0.1928  15.883  < 2e-16 ***
prebake2            0.8237     0.1718   4.795 2.65e-05 ***
flux2              -0.3614     0.1174  -3.079  0.00390 **
speed2              0.8280     0.1295   6.393 1.85e-07 ***
preheat2           -0.4316     0.1818  -2.374  0.02287 *
cooling2            0.3587     0.1852   1.937  0.06037 .
agitator2          -0.2614     0.1684  -1.552  0.12907
temp2              -0.7473     0.1298  -5.756 1.34e-06 ***
preheat2:agitator2  0.7642     0.2565   2.980  0.00507 **
prebake2:cooling2  -1.1749     0.2608  -4.506 6.42e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 3.620902)

    Null deviance: 1137.32  on 46  degrees of freedom
Residual deviance:  153.78  on 37  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

To see the behaviour of the fitted model we consider relative rates. Recall that for a poisson regression with log link, the log rates ratio for two points is just the difference of the η values.

The fitted rate for an observation with prebake = 1, flux = 1, speed = 1, preheat = 1, cooling = 1, agitator = 1, temp = 1 is $\exp(3.0618) = 21.37$. If flux = 2 then the fitted rate decreases by a factir of $\exp(0.3614) = 1.44$. If speed = 2 then the fitted rate increases by a factor of $\exp(0.8280) = 2.29$. If temp = 2 then the fitted rate decreases by a factor of $\exp(0.7473) = 2.11$. Because they interact, we consider the effects of preheat and agitator together. The following table gives the factor by which the base rate changes for different

combinations of preheat and agitator. (Both tend to reduce the rate, but when combined their effects largely cancel out.)

```
> newd1 <- data.frame(prebake=c("1","1","1","1"), flux=c("1","1","1","1"),
+                     speed=c("1","1","1","1"), preheat=c("1","1","2","2"),
+                     cooling=c("1","1","1","1"), agitator=c("1","2","1","2"),
+                     temp=c("1","1","1","1"))
> xtabs(exp(predict(pmod7, newdata=newd1) - pmod7$coefficients[1]) ~ preheat + agitator, newd1)

        agitator
preheat         1         2
      1 1.0000000 0.7699351
      2 0.6494687 1.0737615
```

Similarly, because prebake and cooling interact, we consider their effects jointly. (Both tend to increase the rate, but again we see that when combined their effects largely cancel out.)

```
> newd2 <- data.frame(prebake=c("1","1","2","2"), flux=c("1","1","1","1"),
+                     speed=c("1","1","1","1"), preheat=c("1","1","1","1"),
+                     cooling=c("1","2","1","2"), agitator=c("1","1","1","1"),
+                     temp=c("1","1","1","1"))
> xtabs(exp(predict(pmod7, newdata=newd2) - pmod7$coefficients[1]) ~ prebake + cooling, newd2)

        cooling
prebake         1         2
      1 1.000000 1.431474
      2 2.278928 1.007525
```

So, to minimise the rate of defects we should set flux = 2, speed = 1, temp = 2, agitator = 1, preheat = 2, cooling = 1, and prebake = 1. This is as you would expect looking at the main effects, except for agitator, for which the interaction with preheat outweighs the main effect.