Applied Analysis 3

Rohan Hore

August 2024

Rodent Altruism Rats are like humans in that they dislike incarceration and physical restraint. They also dislike seeing a fellow rat in distress from physical restraint. Rats are altruistic to a certain extent: a free rat observing a restrained rat under stress may respond like the Good Samaritan and endeavour to free the restrained rat from his shackles even though there is no immediate or apparent reward in it for the free rat. Like humans, rats learn by training; on average, over time, they become more adept at restraint removal.

A study in the lab of Peggy Mason measured the response time—the time taken for the free rat to pick the lock—under a range of treatments. Each pair of rats was observed daily over a period of 12 days, which for present purposes may be taken as consecutive. Each day, one rat was restrained (the same one each day), and the other rat was free to explore in the vicinity. The observation is the time in minutes for the free rat to open the restraint. All times are truncated at 40 minutes. The first experiment consists of 96 pairs of male rats divided into six groups of 16 pairs. Each of the six treatment levels is a short-lasting drug given by injection to the free rat just before the commencement of the experiment each day. The treatment levels are: `uninjected, saline, highMDZ, lowMDZ, nadalol, propranolol`, the first two of which are control levels. The data for this experiment are available in the file `free_rat.csv`.

The data are in spreadsheet format with 96 rows, one column for treatment level, one for rat ID, and one for the response time on each of the 12 days. Although the days are presumed to be consecutive for each rat, the entire experiment was not performed in 12 days, so day 1 for rat 1 is not the same as day 1 for rat 96. Two values are missing; other values reported as 0 are not missing but are presumed to be less than 30 seconds.

Data Reading

```
#data reading
rat_data=read.csv("free_rat.csv")
head(rat_data)
```

```
##      TRT Ratid D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12
## 1 saline    R1 40 40 40 40 40 40 40 40 40  40  40  40
## 2 saline    R2 40 40 25 12  2  0  0  1  0   0   0   0
## 3 saline    R3 19 40 40 40 40  0  4  0  0   0   0   2
## 4 saline    R4 10 40 40 40 40 40 40 40 40  40  40  40
## 5 saline    R5  6 40 40 40 40 40 40 40  4  40  40   3
## 6 saline    R6 14 36 38  2  4  0  0  0  0   0   0   0
```

Possible Questions

Problem 1

Construct a plot of the data illustrating the temporal trends in mean truncated response times for each of the six treatment levels. Report any interesting observation, you see.

Answer:

```
attach(rat_data)
treatments=unique(rat_data$TRT)
aggr.data=matrix(nrow=6*12,ncol=3)
i0=0
for(i in 1:6){
  aggr.data[(i0+1):(i0+12),1]=treatments[i]
  aggr.data[(i0+1):(i0+12),2]=1:12
  aggr.data[(i0+1):(i0+12),3]=colMeans(rat_data[rat_data$TRT==treatments[i],3:14])
  i0=i0+12
}
aggr.data=as.data.frame(aggr.data)
colnames(aggr.data)=c("treatment","day","mean_response")
aggr.data$mean_response=as.numeric(aggr.data$mean_response)
aggr.data$day=as.numeric(aggr.data$day)
head(aggr.data)
```

```
##   treatment day mean_response
## 1    saline   1       33.0625
## 2    saline   2       39.7500
## 3    saline   3       38.9375
## 4    saline   4       31.5625
## 5    saline   5       33.5625
## 6    saline   6       27.8750
```

```
library(ggplot2)
ggplot(aggr.data,aes(x=day,y=mean_response))+
  geom_line()+geom_point()+
  facet_wrap(.~treatment,ncol=3)
```

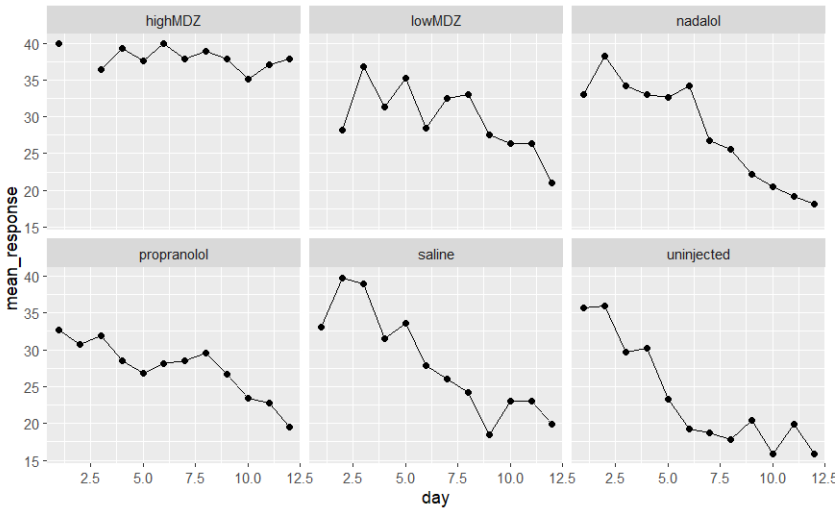

Figure 1: Caption

```
ggplot(aggr.data,aes(x=day,y=mean_response,color=treatment))+
  geom_line()+geom_point()
```

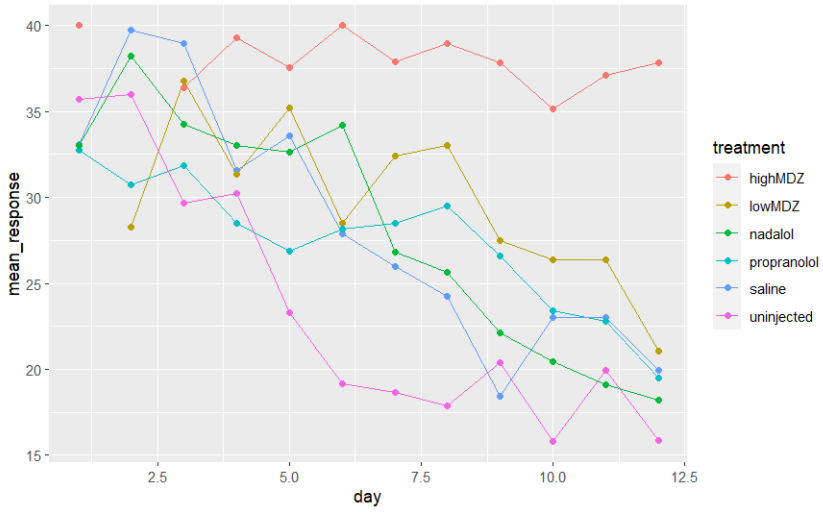

Figure 2: Caption

Problem 2

There are two control levels in the treatments. `(1) Uninjected and (2) Saline`. Does the plot above suggest that these two control levels effect the response time differently for rats? Why did the experimenter even chose to include these two control levels - what purpose does it solve?

Answer: For both control levels, the trend with time is almost similar as per visual judgement except that for saline treatment, mean response time is overall greater than that for uninjected treatment.

Even though for the saline level there is no drug effect, just the fact that we are injecting in their body it might induce some effect in the panic levels and that might effect the response times. So, the two levels have more meaning other than merely two control levels.

Problem 3

We want to fit a linear Gaussian model for studying the mean response by day and treatment. Argue whether the linear gaussian model is suitable for this data? Additionally, what model assumptions do the plot from *Problem 1* suggest?

Answer: We first need to transform the data frame slightly for easier handling.

```
trans.data=matrix(0,ncol=4,nrow=dim(rat_data)[1]*12)
i0=0
for(i in 1:dim(rat_data)[1]){
  trans.data[(i0+1):(i0+12),1]=rat_data[i,1]
  trans.data[(i0+1):(i0+12),2]=rat_data[i,2]
  trans.data[(i0+1):(i0+12),3]=1:12
  trans.data[(i0+1):(i0+12),4]=unlist(rat_data[i,3:14])
  i0=i0+12
}
trans.data=as.data.frame(trans.data)
colnames(trans.data)=c("treatment","ratID","day","response_time")
trans.data$day=as.numeric(trans.data$day)
trans.data$response_time=as.numeric(trans.data$response_time)
trans.data=trans.data[-which(is.na(trans.data$response_time)),]
head(trans.data)
```

```
##   treatment ratID day response_time
## 1    saline    R1   1            40
```

```
## 2    saline    R1   2            40
## 3    saline    R1   3            40
## 4    saline    R1   4            40
## 5    saline    R1   5            40
## 6    saline    R1   6            40
```

Because of truncation at 40 and clustering at 0, the histogram looks very different. We here opt for the boxcox transformation and can use the optimally transformed response.

```
hist(trans.data$response_time)
library(MASS)
b=boxcox(lm(trans.data$response_time+1~1))
plot(b)
(lambda=b$x[which.max(b$y)])
trans.data$response_time=(trans.data$response_time^lambda-1)/lambda
```

```
## [1] 1.070707
```

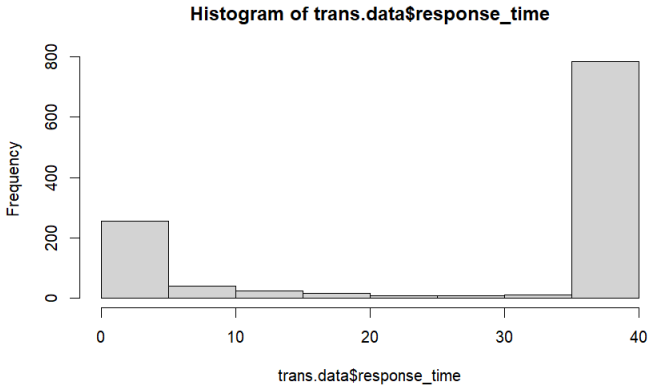

Figure 3: Caption

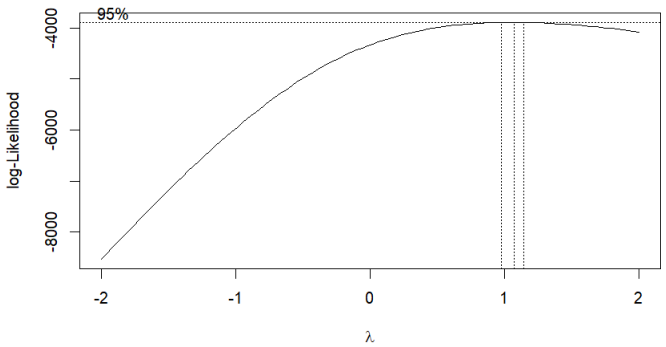

Figure 4: Caption

qqplot is a nice addition here

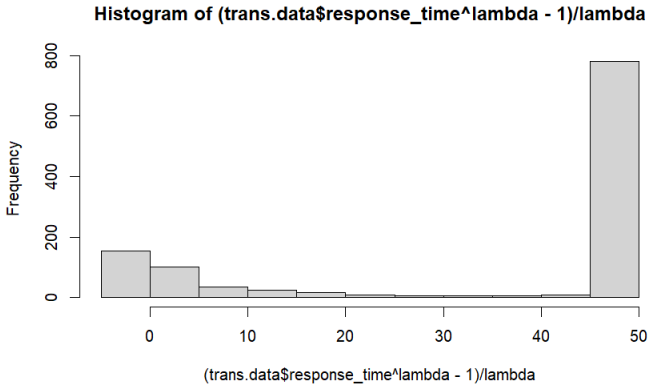

Figure 5: Caption

Problem 4

Fit the previously discussed model on the data and provide its interpretation. Report the estimates , their significance and also discuss how one should interpret them.

Answer: Even after boxcox transformation, the histogram looks nothing close to a gaussian distribution. Hence, we will model the raw response time itself. The model coefficients will be easier to interpret that way.

```
library(nlme)
model.1=lme(response_time~day*treatment,data=trans.data,random=~1|ratID,method="ML")
summary(model.1)
model.1d=lme(response_time~day,data=trans.data,random=~1|ratID,method="ML")
anova(model.1d,model.1)
model.1t=lme(response_time~treatment,data=trans.data,random=~1|ratID,method="ML")
anova(model.1t,model.1)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: trans.data
##        AIC      BIC    logLik
##   9045.419 9116.084 -4508.709
##
## Random effects:
##  Formula: ~1 | ratID
##         (Intercept) Residual
## StdDev:    10.70931 10.98634
##
## Fixed effects:  response_time ~ day * treatment
##                               Value Std.Error   DF   t-value p-value
## (Intercept)               37.98331  3.189744 1048 11.907951  0.0000
## day                       -0.04549  0.232004 1048 -0.196089  0.8446
## treatmentlowMDZ           -3.53390  4.512641   90 -0.783111  0.4356
## treatmentnadalol           1.84529  4.506188   90  0.409501  0.6831
## treatmentpropranolol      -4.23520  4.506188   90 -0.939864  0.3498
## treatmentsaline            1.70987  4.506188   90  0.379450  0.7052
## treatmentuninjected       -2.47573  4.506188   90 -0.549408  0.5841
## day:treatmentlowMDZ       -0.68195  0.328497 1048 -2.075976  0.0381
## day:treatmentnadalol      -1.75346  0.327316 1048 -5.357086  0.0000
## day:treatmentpropranolol -0.92457  0.327316 1048 -2.824698  0.0048
```

```
## day:treatmentsaline      -1.71019  0.327316 1048 -5.224892  0.0000
## day:treatmentuninjected  -1.79301  0.327316 1048 -5.477930  0.0000
##  Correlation:
##                          (Intr) day    trtMDZ trtmntnd trtmntp trtmnts trtmntnn dy:MDZ dy:trtmntnd
## day                      -0.475
## treatmentlowMDZ          -0.707  0.335
## treatmentnadalol         -0.708  0.336  0.500
## treatmentpropranolol     -0.708  0.336  0.500  0.501
## treatmentsaline          -0.708  0.336  0.500  0.501    0.501
## treatmentuninjected      -0.708  0.336  0.500  0.501    0.501   0.501
## day:treatmentlowMDZ       0.335 -0.706 -0.475 -0.237   -0.237  -0.237  -0.237
## day:treatmentnadalol      0.336 -0.709 -0.238 -0.473   -0.238  -0.238  -0.238    0.501
## day:treatmentpropranolol  0.336 -0.709 -0.238 -0.238   -0.473  -0.238  -0.238    0.501  0.502
## day:treatmentsaline       0.336 -0.709 -0.238 -0.238   -0.238  -0.473  -0.238    0.501  0.502
## day:treatmentuninjected   0.336 -0.709 -0.238 -0.238   -0.238  -0.238  -0.473    0.501  0.502
##                          dy:trtmntp dy:trtmnts
## day
## treatmentlowMDZ
## treatmentnadalol
## treatmentpropranolol
## treatmentsaline
## treatmentuninjected
## day:treatmentlowMDZ
## day:treatmentnadalol
## day:treatmentpropranolol
## day:treatmentsaline       0.502
## day:treatmentuninjected   0.502      0.502
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.51177635 -0.50588233  0.03146399  0.56522868  2.83272548
##
## Number of Observations: 1150
## Number of Groups: 96
```

Is the effect of treatment statistically significant?

```
\begin{verbatim}
##          Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## model.1d     1  4 9087.940 9108.130 -4539.970
## model.1      2 14 9045.419 9116.084 -4508.709 1 vs 2 62.52068  <.0001
```

Is the effect of day statistically significant?

```
##          Model df      AIC      BIC    logLik   Test L.Ratio p-value
## model.1t     1  8 9226.255 9266.635 -4605.127
## model.1      2 14 9045.419 9116.084 -4508.709 1 vs 2 192.836  <.0001
```

Both the fixed effects of day and treatment are significant. As per the fitted model coefficients, there is a slight decreasing trend in response time with day. The model treats high MDZ as the base level and we can interpret the coefficients as follows: if we compare high mDZ and low MDZ, high MDZ has overall higher levels of response time by 3.5 and decreases with time in a slower rate (about 0.68)

Problem 5

Observe that the same rat was observed on 12 consecutive days and it might be possible that the residual effect of drug in one day is carried forward to the next day measurements. Does your model above already take care of that issue? If not, how will you fix it?

Answer: The observation for a single rat on two close by days will be highly correlated than two observations on two far away days. Hence, instead of a variance structure taking care of the rat specific block effect, we can use an autoregressive block matrix

$$V_{ij} = \sigma_r^2 \phi^{|d(i)-d(j)|} \times 1\{r(i) = r(j)\}$$

where $r(i)$ and $d(i)$ correspondingly denote the rat-ID and day of observation i. We can allow for such complicate correlation structure using nlme package in R.

```
model.2=lme(response_time~day*treatment,data=trans.data,random=~1|ratID,
correlation = corAR1(form=~day|ratID),method="ML")
summary(model.2)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: trans.data
##        AIC      BIC   logLik
##   8806.359 8882.072 -4388.18
##
## Random effects:
##  Formula: ~1 | ratID
##         (Intercept) Residual
## StdDev:    8.785108 12.41189
##
## Correlation Structure: ARMA(1,0)
##  Formula: ~day | ratID
##  Parameter estimate(s):
##     Phi1
## 0.555089
## Fixed effects:  response_time ~ day * treatment
##                               Value Std.Error   DF   t-value p-value
## (Intercept)               38.61768  3.604851 1048 10.712697  0.0000
## day                       -0.11085  0.369470 1048 -0.300013  0.7642
## treatmentlowMDZ           -4.80279  5.110204   90 -0.939842  0.3498
## treatmentnadalol          -0.48055  5.096926   90 -0.094283  0.9251
## treatmentpropranolol      -4.44642  5.096926   90 -0.872373  0.3853
## treatmentsaline           -0.79037  5.096926   90 -0.155069  0.8771
## treatmentuninjected       -2.83085  5.096926   90 -0.555403  0.5800
## day:treatmentlowMDZ       -0.63618  0.523989 1048 -1.214099  0.2250
## day:treatmentnadalol      -1.49519  0.522408 1048 -2.862116  0.0043
## day:treatmentpropranolol -0.95921  0.522408 1048 -1.836140  0.0666
## day:treatmentsaline       -1.40495  0.522408 1048 -2.689378  0.0073
## day:treatmentuninjected  -1.71172  0.522408 1048 -3.276604  0.0011
##  Correlation:
##                           (Intr) day    trtMDZ trtmntnd trtmntp trtmnts trtmntnn dy:MDZ dy:trtmntnd
## day                       -0.667
## treatmentlowMDZ           -0.705  0.470
## treatmentnadalol          -0.707  0.471  0.499
## treatmentpropranolol      -0.707  0.471  0.499  0.500
## treatmentsaline           -0.707  0.471  0.499  0.500    0.500
## treatmentuninjected       -0.707  0.471  0.499  0.500    0.500   0.500
```

```
## day:treatmentlowMDZ       0.470 -0.705 -0.668 -0.332   -0.332  -0.332  -0.332
## day:treatmentnadalol      0.471 -0.707 -0.333 -0.666   -0.333  -0.333  -0.333    0.499
## day:treatmentpropranolol  0.471 -0.707 -0.333 -0.333   -0.666  -0.333  -0.333    0.499  0.500
## day:treatmentsaline       0.471 -0.707 -0.333 -0.333   -0.333  -0.666  -0.333    0.499  0.500
## day:treatmentuninjected   0.471 -0.707 -0.333 -0.333   -0.333  -0.333  -0.666    0.499  0.500
##                          dy:trtmntp dy:trtmnts
## day
## treatmentlowMDZ
## treatmentnadalol
## treatmentpropranolol
## treatmentsaline
## treatmentuninjected
## day:treatmentlowMDZ
## day:treatmentnadalol
## day:treatmentpropranolol
## day:treatmentsaline       0.500
## day:treatmentuninjected   0.500      0.500
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.9150292 -0.6972115  0.1322362  0.7185762  2.0405232
##
## Number of Observations: 1150
## Number of Groups: 96
```

Problem 6

The plots in *Problem 1* should have indicated that the effect of time is different on response time for different treatment. Can you do a statistical test to check whether the effect is actually different across different treatments. Clearly mention the models and test procedures you use.

Answer: We can use the simple model from problem 4 with rat-specific random effect and can test for the significance of the interaction term between treatment and day effects. We can opt for a simple likelihood ratio test. The log-likelihood ratio is significantly large for a χ^2_1 distribution, hence the effect is significant here.

```
model.3=lme(response_time~day+treatment,data=trans.data,random=~1|ratID,method="ML")
summary(model.3)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: trans.data
##       AIC      BIC   logLik
##   9084.82 9130.247 -4533.41
##
## Random effects:
##  Formula: ~1 | ratID
##         (Intercept) Residual
## StdDev:    10.69033 11.24653
##
## Fixed effects:  response_time ~ day + treatment
##                           Value Std.Error   DF    t-value p-value
## (Intercept)           45.46647  2.872157 1053  15.830077  0.0000
## day                   -1.19228  0.096473 1053 -12.358607  0.0000
## treatmentlowMDZ       -7.98145  3.963075   90  -2.013954  0.0470
## treatmentnadalol      -9.58125  3.962604   90  -2.417917  0.0176
```

```
## treatmentpropranolol -10.27396  3.962604   90  -2.592728  0.0111
## treatmentsaline       -9.43541  3.962604   90  -2.381115  0.0194
## treatmentuninjected  -14.15937  3.962604   90  -3.573249  0.0006
##  Correlation:
##                      (Intr) day    trtMDZ trtmntnd trtmntp trtmnts
## day                  -0.219
## treatmentlowMDZ      -0.690  0.000
## treatmentnadalol     -0.690  0.001  0.500
## treatmentpropranolol -0.690  0.001  0.500  0.500
## treatmentsaline      -0.690  0.001  0.500  0.500    0.500
## treatmentuninjected  -0.690  0.001  0.500  0.500    0.500   0.500
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.25241197 -0.49329157  0.07015039  0.60021478  2.79051915
##
## Number of Observations: 1150
## Number of Groups: 96
```

```
anova(model.3,model.1)
```

```
##         Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## model.3     1  9 9084.820 9130.247 -4533.410
## model.1     2 14 9045.419 9116.084 -4508.709 1 vs 2 49.40069  <.0001
```

Problem 7

In the conduct of this experiment, it appears that the rats must first become familiar with their handler. Otherwise they 'freak out' when injected by a stranger. Assume that one graduate student can become friendly with up to eight pairs of rats at once, and that it is feasible for one hard-working student to perform eight experiments per day. The student can repeat this performance once per month on different batches of rats over a three-month period. Accordingly, four graduate students are needed to perform the experiment on 96 rat pairs over a 3-month period. Handler information is not available in the data provided.

How would you allocate rats to students if you were in charge? Explain how this information would affect your analysis of the data if it were available. Discuss other factors that could influence the design of an experiment such as this, and also the analysis of data.

Answer: Equal representation of treatments in each block (each handler), so that any effect of treatment doesn't get confounded with the effect of handlers.

Problem 8

What do you think is the impact of the truncation of response times on the results you generated? Describe in detail a model that that you think would be best for analyzing these data. Report parameter estimates and their interpretation.

Answer: Almost 67% of the response times are 40 i.e. truncated. So, one possible solution is to binarize the response time. If the response time is less than 40, rat responded successfully and otherwise, it didn't.

```
mean(trans.data$response_time==40)
trans.data=cbind(trans.data,0)
trans.data[,5]=as.numeric(trans.data$response_time<40)
colnames(trans.data)=c("treatment","ratID","day","response_time","bin_response")
head(trans.data)
```

```
## [1] 0.6773913
```

The data with binarized response time looks as follows

```
##   treatment ratID day response_time bin_response
## 1    saline    R1   1            40            0
## 2    saline    R1   2            40            0
## 3    saline    R1   3            40            0
## 4    saline    R1   4            40            0
## 5    saline    R1   5            40            0
## 6    saline    R1   6            40            0
```

We can take motivation from previously fitted models and fit a generalized linear mixed model with rat specific random effects and fixed effects of treatment, day and an interaction between them. We can further test like before, whether the interaction term is even needed or not.

```
library(lme4)
model.4=glmer(bin_response~day*treatment+(1|ratID),data=trans.data,family="binomial")
summary(model.4)

model.5=glmer(bin_response~day+treatment+(1|ratID),data=trans.data,family="binomial")
anova(model.5,model.4)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
##  Family: binomial  ( logit )
## Formula: bin_response ~ day * treatment + (1 | ratID)
##    Data: trans.data
##
##      AIC      BIC   logLik deviance df.resid
##    928.6    994.2   -451.3    902.6     1137
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5556 -0.3548 -0.1009  0.2523 15.2877
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  ratID  (Intercept) 8.276    2.877
## Number of obs: 1150, groups:  ratID, 96
##
## Fixed effects:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -4.86906    1.12660  -4.322 1.55e-05 ***
## day                        0.05640    0.08959   0.630 0.528991
## treatmentlowMDZ            2.68475    1.41265   1.901 0.057367 .
## treatmentnadalol           1.43088    1.43060   1.000 0.317217
## treatmentpropranolol       2.27560    1.44415   1.576 0.115087
## treatmentsaline            0.29089    1.49454   0.195 0.845676
## treatmentuninjected        0.83522    1.48598   0.562 0.574068
## day:treatmentlowMDZ        0.05396    0.10744   0.502 0.615500
## day:treatmentnadalol       0.28105    0.11211   2.507 0.012181 *
## day:treatmentpropranolol   0.14989    0.11468   1.307 0.191193
## day:treatmentsaline        0.38357    0.12624   3.038 0.002379 **
## day:treatmentuninjected    0.45957    0.13675   3.361 0.000777 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation of Fixed Effects:
##             (Intr) day    trtMDZ trtmntnd trtmntp trtmnts trtmntnn dy:MDZ dy:trtmntnd dy:trtmntp
## day         -0.560
## trtmntlwMDZ -0.778  0.446
## tretmntndll -0.770  0.440  0.604
## trtmntprprn -0.757  0.436  0.595  0.588
## treatmntsln -0.713  0.420  0.565  0.558    0.551
## trtmntnnjct -0.702  0.422  0.560  0.553    0.548   0.529
## dy:trtmnMDZ  0.462 -0.834 -0.528 -0.364   -0.361  -0.350  -0.352
## dy:trtmntnd  0.434 -0.799 -0.349 -0.550   -0.342  -0.333  -0.338    0.667
## dy:trtmntpr  0.420 -0.781 -0.339 -0.334   -0.546  -0.325  -0.330    0.652  0.627
## dy:trtmntsl  0.358 -0.708 -0.295 -0.291   -0.291  -0.592  -0.300    0.593  0.573       0.562
## dy:trtmntnn  0.299 -0.653 -0.256 -0.251   -0.254  -0.262  -0.579    0.549  0.534       0.525
##             dy:trtmnts
## day
## trtmntlwMDZ
## tretmntndll
## trtmntprprn
## treatmntsln
## trtmntnnjct
## dy:trtmnMDZ
## dy:trtmntnd
## dy:trtmntpr
## dy:trtmntsl
## dy:trtmntnn  0.499
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.126336 (tol = 0.002, component 1)

## Data: trans.data
## Models:
## model.5: bin_response ~ day + treatment + (1 | ratID)
## model.4: bin_response ~ day * treatment + (1 | ratID)
##         npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## model.5    8 943.64 984.02 -463.82   927.64
## model.4   13 928.62 994.24 -451.31   902.62 25.018  5  0.0001382 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Additionally other than 67% of response times being 40, other 92 observations are 0. Can we suggest a better model taking this peculiarity into consideration?

Instead of binarizing, we can create an ordinal variable. Response times at 0 can be considered as category 1-fast response. Response times at 40 can be considered as category 3-slow response. Any response time in between can be considered as category 2-mid response.