

HW3

Hyunjoon Rhee

3/14/2021

Problem 1

```
library(faraway)
data(salmonella)
head(salmonella)
```

```
## colonies dose
## 1      15    0
## 2      21    0
## 3      29    0
## 4      16   10
## 5      18   10
## 6      21   10
```

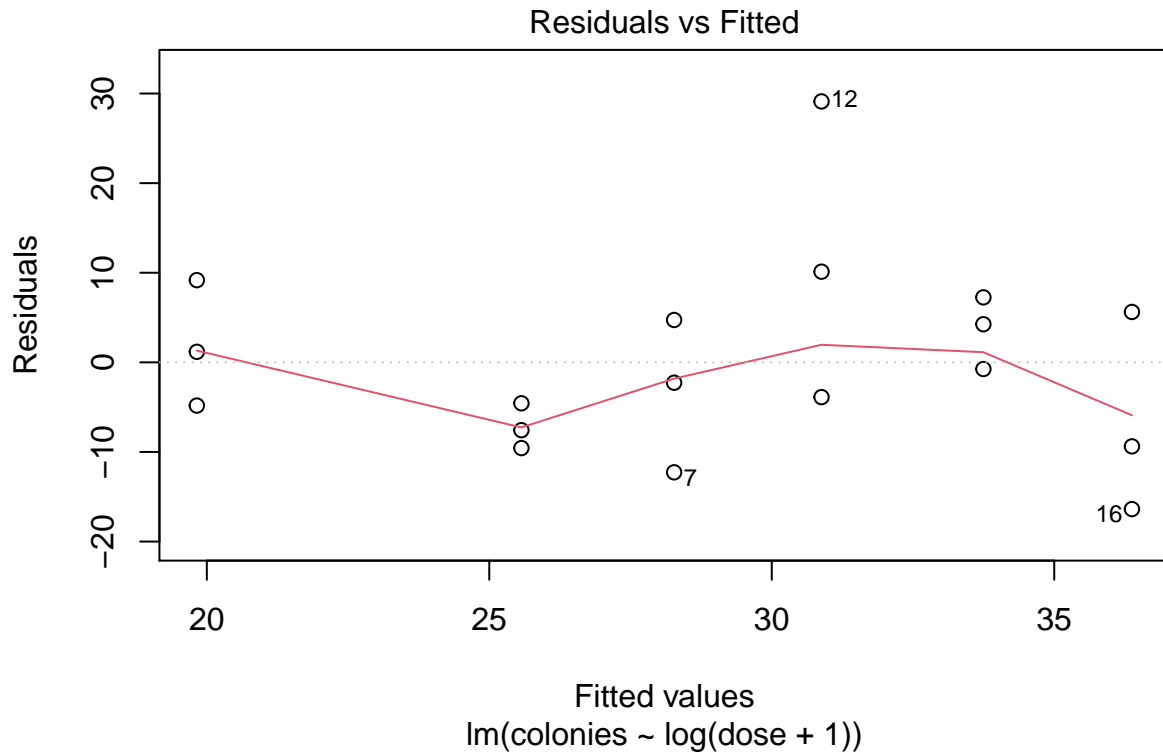
a)

```
a = lm(colonies~log(dose+1),data=salmonella)
summary(a)
```

```
##
## Call:
## lm(formula = colonies ~ log(dose + 1), data = salmonella)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.376  -6.882  -1.509   5.400  29.119
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    19.823     5.064   3.915  0.00123 **
## log(dose + 1)    2.396     1.128   2.125  0.04955 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.84 on 16 degrees of freedom
## Multiple R-squared:  0.2201, Adjusted R-squared:  0.1713
## F-statistic: 4.514 on 1 and 16 DF,  p-value: 0.04955
```

b)

```
plot(a,which=1)
```



The first plot shows the residual vs fitted.

c)

```
c = lm(colonies~factor(log(dose+1)),data=salmonella)
anova(a,c)
```

```
## Analysis of Variance Table
##
## Model 1: colonies ~ log(dose + 1)
## Model 2: colonies ~ factor(log(dose + 1))
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      16 1881.1
## 2      12 1091.3  4    789.73 2.1709 0.1342
```

Because the p value is greater than $\alpha = 0.05$ it fails to reject the null, which means that it has no lack of fit.

Problem 2

```
library(faraway)
library(lmtest)
```

```
## Warning: package 'lmtest' was built under R version 4.0.2
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
data("gammaray")
```

```
prob2 = lm(flux ~ time, data = gammaray)
```

```
bptest(prob2)
```

```
##
```

```
## studentized Breusch-Pagan test
```

```
##
```

```
## data: prob2
```

```
## BP = 0.05492, df = 1, p-value = 0.8147
```

```
dwtest(prob2)
```

```
##
```

```
## Durbin-Watson test
```

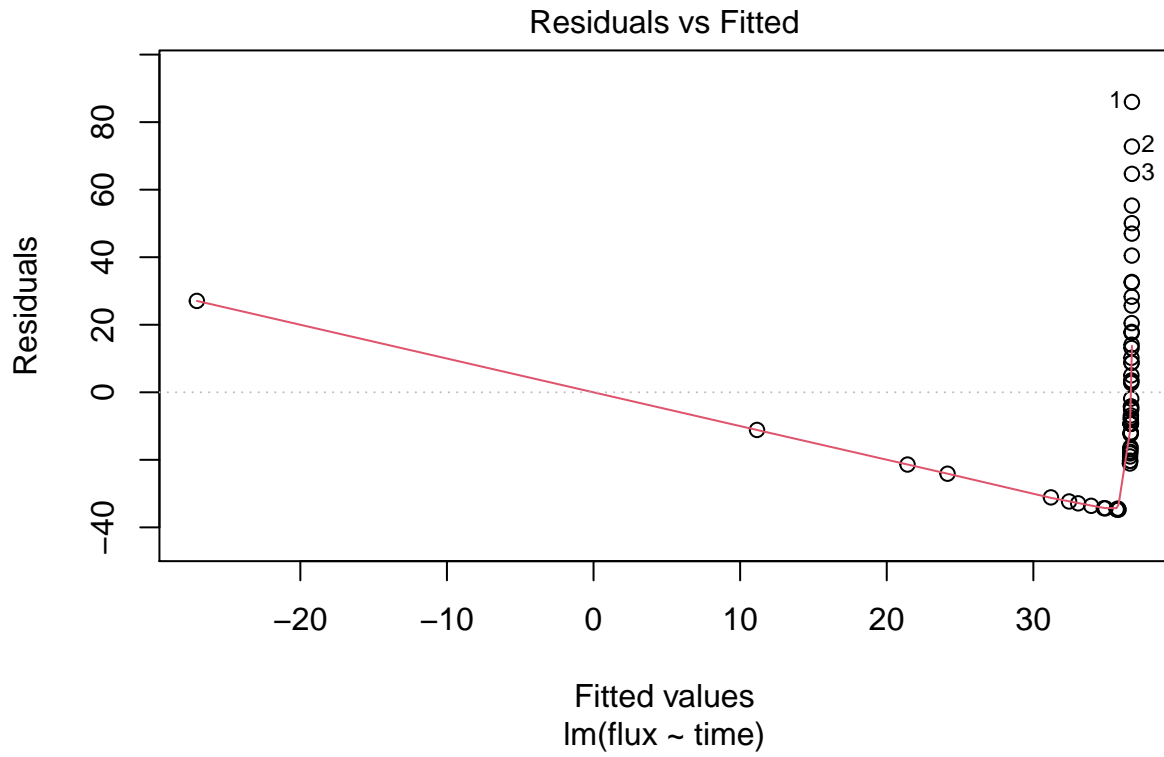
```
##
```

```
## data: prob2
```

```
## DW = 0.053965, p-value < 2.2e-16
```

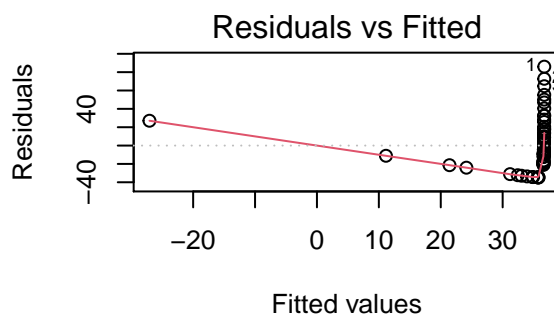
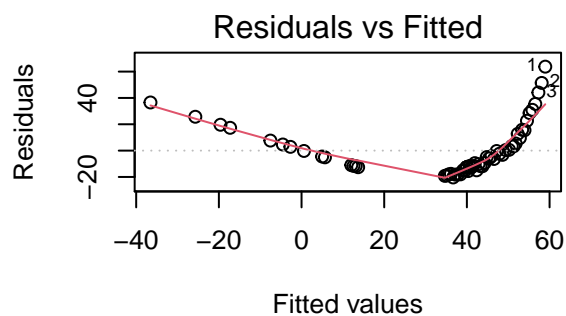
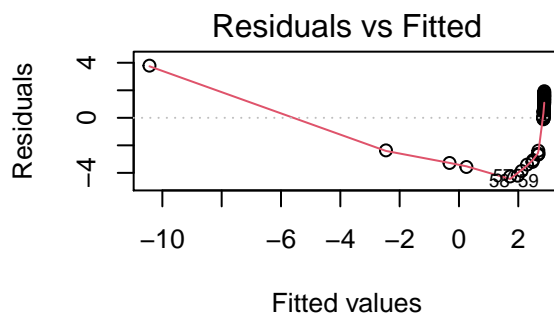
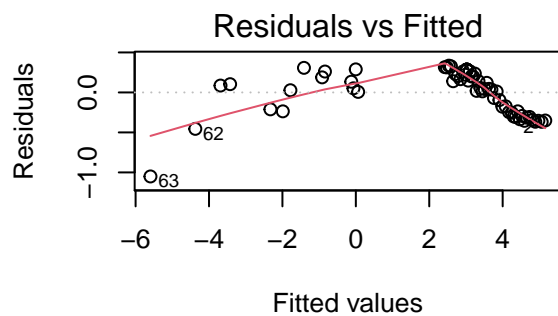
```
## alternative hypothesis: true autocorrelation is greater than 0
```

```
plot(prob2,which=1)
```



a)

```
prob21 = lm(log(flux) ~ log(time), data = gammaray)
prob22 = lm(log(flux) ~ time, data = gammaray)
prob23 = lm(flux ~ log(time), data = gammaray)
prob24 = lm(flux ~ time, data = gammaray)
par(mfrow=c(2,2))
plot(prob21,which=1)
plot(prob22,which=1)
plot(prob23,which=1)
plot(prob24,which=1)
```



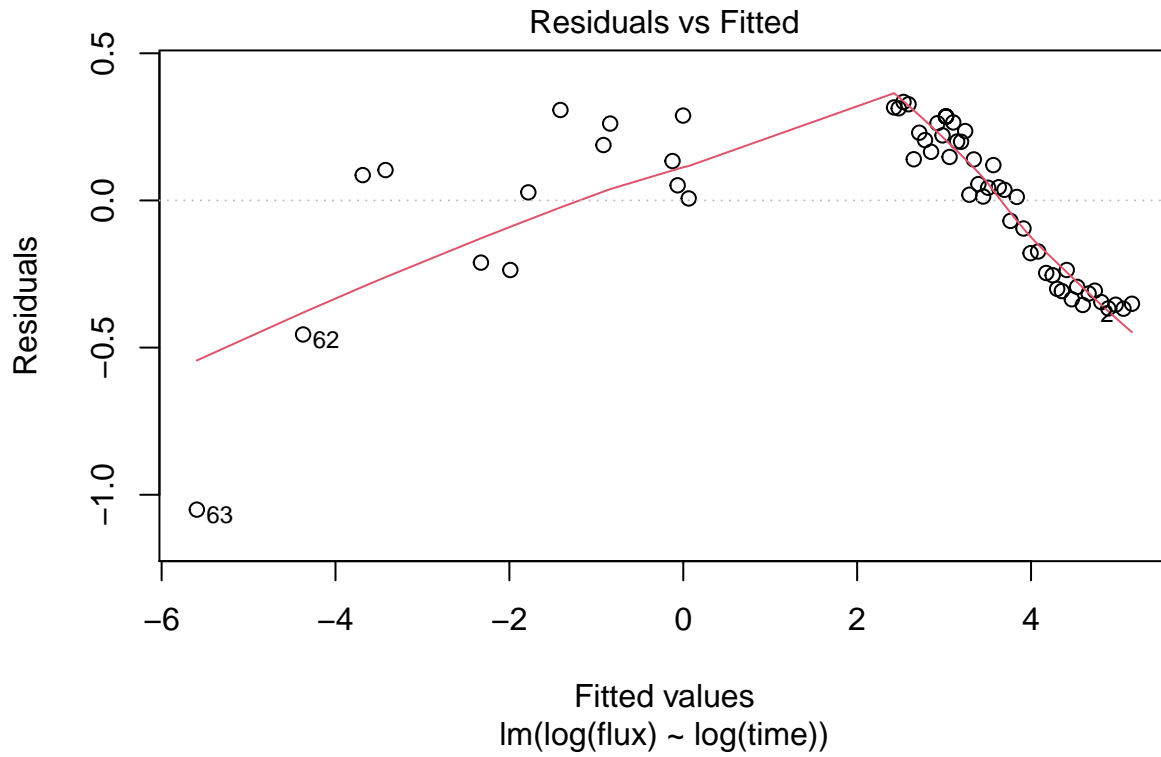
```
prob2a = lm(log(flux) ~ log(time), data = gammaray)
bptest(prob2a)
```

```
##
## studentized Breusch-Pagan test
##
## data: prob2a
## BP = 7.2356, df = 1, p-value = 0.007147
```

```
dwtest(prob2a)
```

```
##
## Durbin-Watson test
##
## data: prob2a
## DW = 0.2978, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```
plot(prob2a, which=1)
```



Transformation of log for both seems to show better.

b)

```
prob2b = lm(log(flux)~log(time), data=gammaray,weights = 1/abs(error))
```

c)

```
summary(prob2b)
```

```
##
## Call:
## lm(formula = log(flux) ~ log(time), data = gammaray, weights = 1/abs(error))
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -4.2948 -0.9795 -0.9323 -0.7641  6.0373
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.27506    0.35397   43.15  <2e-16 ***
## log(time)    -1.68718    0.02864  -58.91  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 1.561 on 61 degrees of freedom
## Multiple R-squared:  0.9827, Adjusted R-squared:  0.9824
## F-statistic: 3471 on 1 and 61 DF,  p-value: < 2.2e-16
```

```
bptest(prob2b)
```

```
##
## studentized Breusch-Pagan test
##
## data:  prob2b
## BP = 7.2356, df = 1, p-value = 0.007147
```

```
shapiro.test(residuals(prob2b))
```

```
##
## Shapiro-Wilk normality test
##
## data:  residuals(prob2b)
## W = 0.94836, p-value = 0.01029
```

Because the Breusch-Pagan test shows that it has p value less than 0.05, it has constant variance. Also because Shapiro test value is less than 0.05, it can be assumed that it is normal.

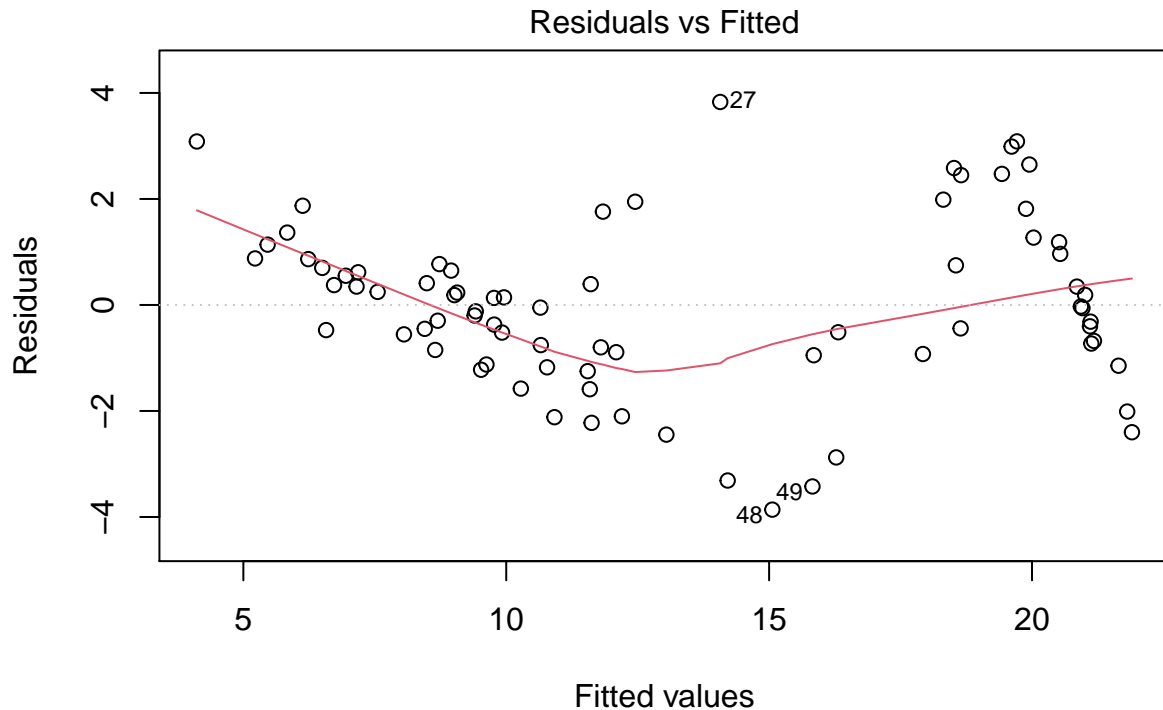
d) I believe this log transformation on both end is the best model because it both meets constant variance and normality assumption.

Problem 3 a)

```
data(divusa)
head(divusa)
```

```
##   year divorce unemployed femlab marriage birth military
## 1 1920      8.0         5.2  22.70      92.0 117.9    3.2247
## 2 1921      7.2        11.7  22.79      83.0 119.8    3.5614
## 3 1922      6.6         6.7  22.88      79.7 111.2    2.4553
## 4 1923      7.1         2.4  22.97      85.2 110.5    2.2065
## 5 1924      7.2         5.0  23.06      80.3 110.9    2.2889
## 6 1925      7.2         3.2  23.15      79.2 106.6    2.1735
```

```
prob3 = lm(divorce ~ unemployed+femlab+marriage+birth+military, data = divusa)
plot(prob3,which=1)
```



lm(divorce ~ unemployed + femlab + marriage + birth + military)

It is hard to say that it is autocorrelated just by looking at the graph, but I believe it is autocorrelated because it looks like it is connected from one after another.

b)

```
dwtest(prob3)
```

```
##
## Durbin-Watson test
##
## data: prob3
## DW = 0.29988, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

Null hypothesis is that errors are not autocorrelated. However, since the pvalue is smaller than 0.05, we reject the null hypothesis, meaning that errors are autocorrelated.

c)

```
library(nlme)
prob3c = gls(divorce ~ unemployed+femlab+marriage+birth+military, correlation= corAR1(form=~year),method="REML")
summary(prob3c)
```

```
## Generalized least squares fit by maximum likelihood
```



```
## Model: divorce ~ unemployed + femlab + marriage + birth + military
## Data: divusa
##      AIC      BIC    logLik
## 179.9523 198.7027 -81.97613
##
## Correlation Structure: AR(1)
## Formula: ~year
## Parameter estimate(s):
##      Phi
## 0.9715486
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept) -7.059682  5.547193 -1.272658  0.2073
## unemployed   0.107643  0.045915  2.344395  0.0219
## femlab       0.312085  0.095151  3.279878  0.0016
## marriage     0.164326  0.022897  7.176766  0.0000
## birth       -0.049909  0.022012 -2.267345  0.0264
## military     0.017946  0.014271  1.257544  0.2127
##
## Correlation:
##      (Intr) unmply femlab marrig birth
## unemployed -0.420
## femlab     -0.802  0.240
## marriage   -0.516  0.607  0.307
## birth      -0.379  0.041  0.066 -0.094
## military   -0.036  0.436 -0.311  0.530  0.128
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.4509327 -0.9760939 -0.6164694  1.1375377  2.1593261
##
## Residual standard error: 2.907665
## Degrees of freedom: 77 total; 71 residual
```

d)

```
intervals(prob3c)
```

```
## Approximate 95% confidence intervals
##
## Coefficients:
##              lower      est.      upper
## (Intercept) -18.12047042 -7.05968162  4.001107165
## unemployed   0.01609101  0.10764313  0.199195251
## femlab       0.12235846  0.31208493  0.501811412
## marriage     0.11867101  0.16432630  0.209981587
## birth       -0.09380023 -0.04990919 -0.006018159
## military    -0.01050915  0.01794640  0.046401944
## attr(,"label")
## [1] "Coefficients:"
##
## Correlation structure:
```

```
##          lower      est.      upper
## Phi 0.6528097 0.9715486 0.9980192
## attr(,"label")
## [1] "Correlation structure:"
##
## Residual standard error:
##          lower      est.      upper
## 0.7974404 2.9076645 10.6020628
```

Because the confidence interval does not contain 0, we reject the null hypothesis. Therefore correlation coefficient is significant.

- e) GLS model does not change the variables that are shown significant in the lm model.
- f) Autocorrelation gives a potential reason why there might be a error in correlation, due to its time sequential properties.