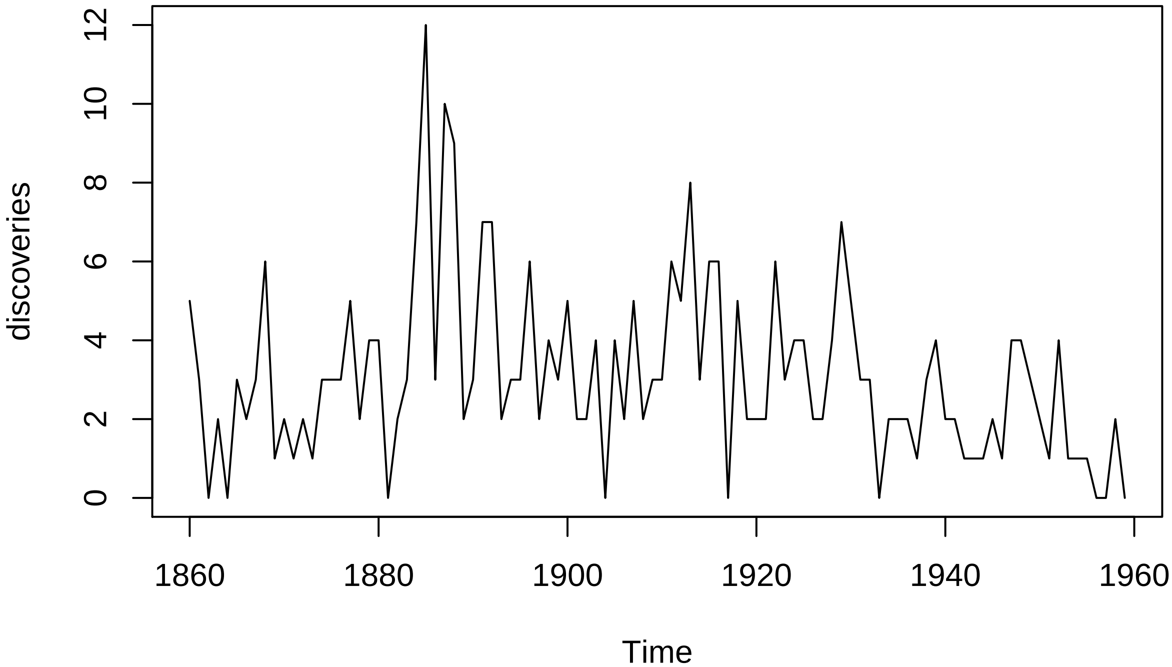
HW4 MA576

```
##3
#a
discover <- discoveries
time <- c(1860:1959)
plot(discoveries)
```



It seems that the rate of discoveries was fluctuating up and down and reches the peak aroung 1885. During the recent years, the rate becomes slower.

```
#b
mod1 <- glm(discover~time, family = poisson, data = discover)
mod2 <- update(mod1, .~. + I(time^2))
mod3 <- update(mod2, .~. + I(time^3))
mod4 <- update(mod3, .~. + I(time^4))
mod5 <- update(mod4, .~. + I(time^5))
AIClm <- function(l){
  AIC <- extractAIC(l, k = 2)[2]
  return(AIC)
}
matrix(unlist(lapply(list(mod1, mod2, mod3,mod4,mod5),AIClm)), byrow = TRUE, ncol = 1, dimnames = list(
```

```
##           AIC
## mod1 430.3225
## mod2 407.8451
## mod3 409.7340
## mod4 410.9580
## mod5 410.9930
```

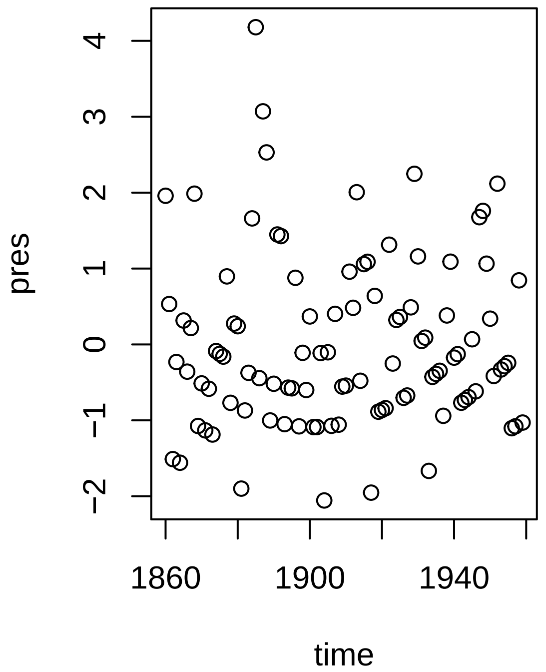
```
summary(mod2)
```

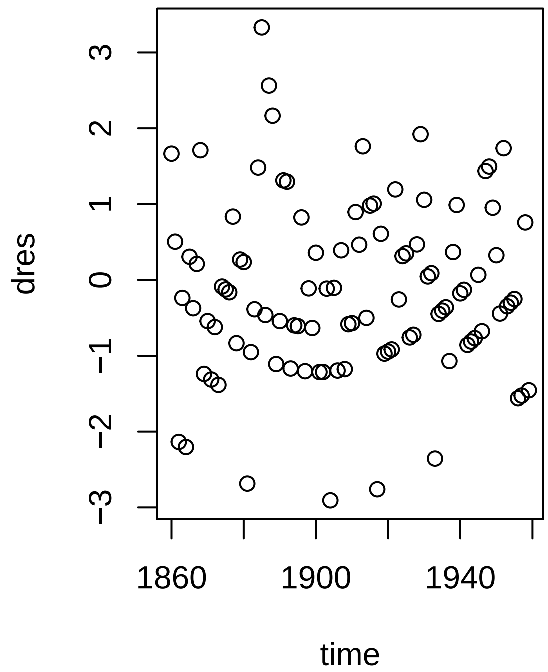
```
##
```

```
## Call:
## glm(formula = discover ~ time + I(time^2), family = poisson,
##     data = discover)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9066  -0.8397  -0.2544   0.4776   3.3303
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.482e+03  3.163e+02  -4.685 2.79e-06 ***
## time         1.561e+00  3.318e-01   4.705 2.54e-06 ***
## I(time^2)   -4.106e-04  8.699e-05  -4.720 2.35e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 164.68  on 99  degrees of freedom
## Residual deviance: 132.84  on 97  degrees of freedom
## AIC: 407.85
##
## Number of Fisher Scoring iterations: 5
```

Mod2 has the smallest AIC, and order of 2 gives the most parsimonious description of the data When time is zero, the dicover is expected to be exp(-1.482e+03), very close to zero. As time increases by one year, discoveries will change to exp(-1.482e+03)*exp(1.561e+00)*exp(-4.106e-04), which is also a very small number and close to zero.

```r
#c
pres <- residuals(mod2, type = "pearson")
dres <- residuals(mod2, type = "deviance")
par(mfrow = c(1,2))
plot(time, pres)
plot(time, dres)
```






```r
dispersion <- sum((pres^2)/97)
dispersion
```

```
## [1] 1.305649
```

```r
with(mod2, cbind(res.deviance = deviance, df = df.residual,p = pchisq(deviance, df.residual, lower.tail
```

```
##      res.deviance df           p
## [1,]     132.8384 97 0.009204575
```

```r
anova(mod2, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: discover
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                         99     164.69
## time       1   7.3688        98     157.32  0.006637 **
## I(time^2)  1  24.4774        97     132.84 7.519e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

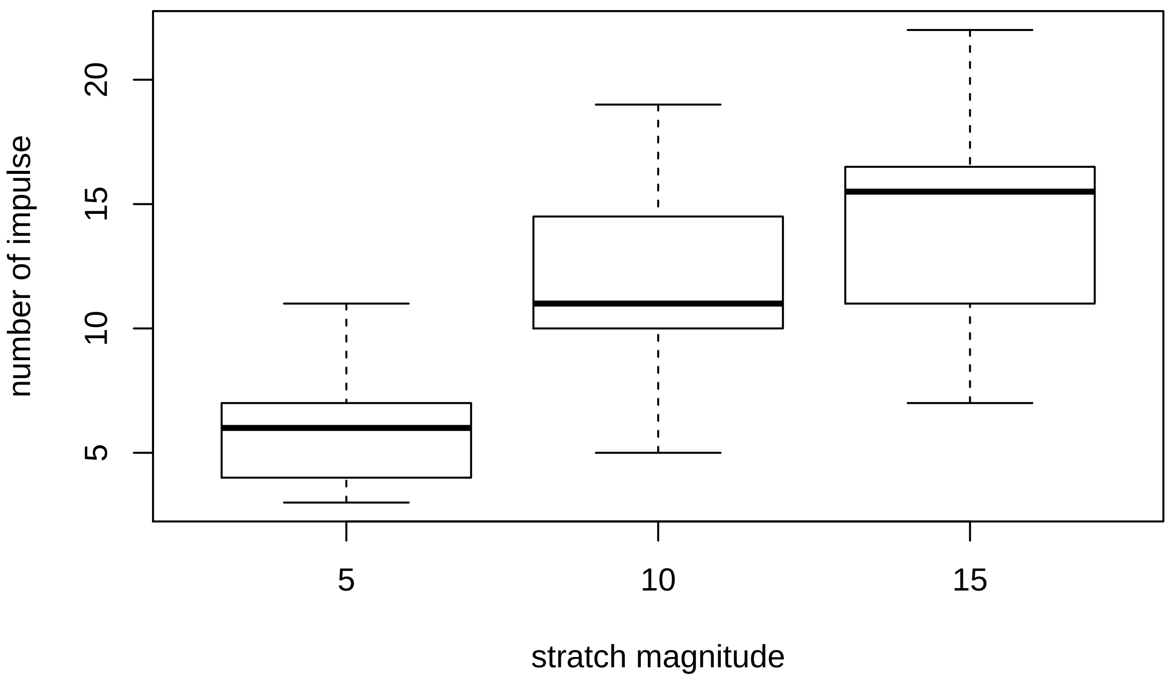
residuals plots still presents curved pattern dispersion get from the Residual deviance over 97 df is 1.369485 and my estimate by pearson residuals is 1.305649. They are all very close to 1. However, the p-value of Chi-sq test is less than 0.05. The overdispersion might be dignificant.

```r
##4
#a
stc <- read.table("stretch.dat", header = T)
```

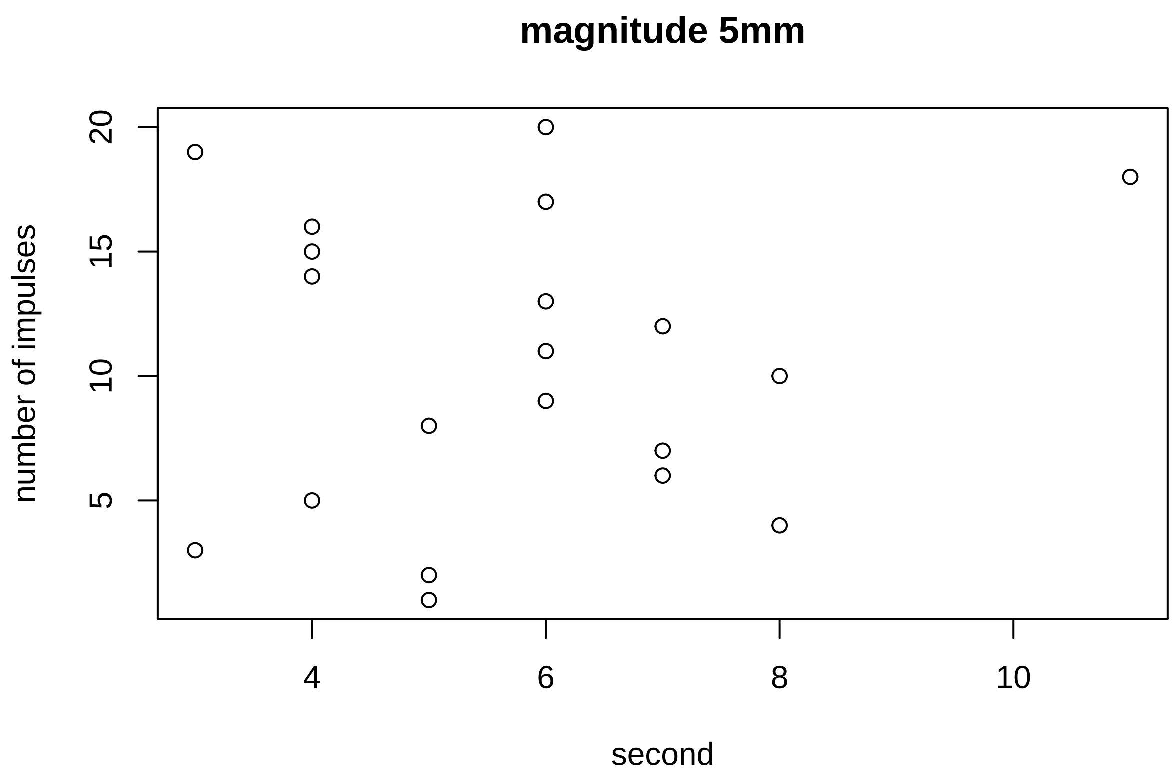
```r
attach(stc)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

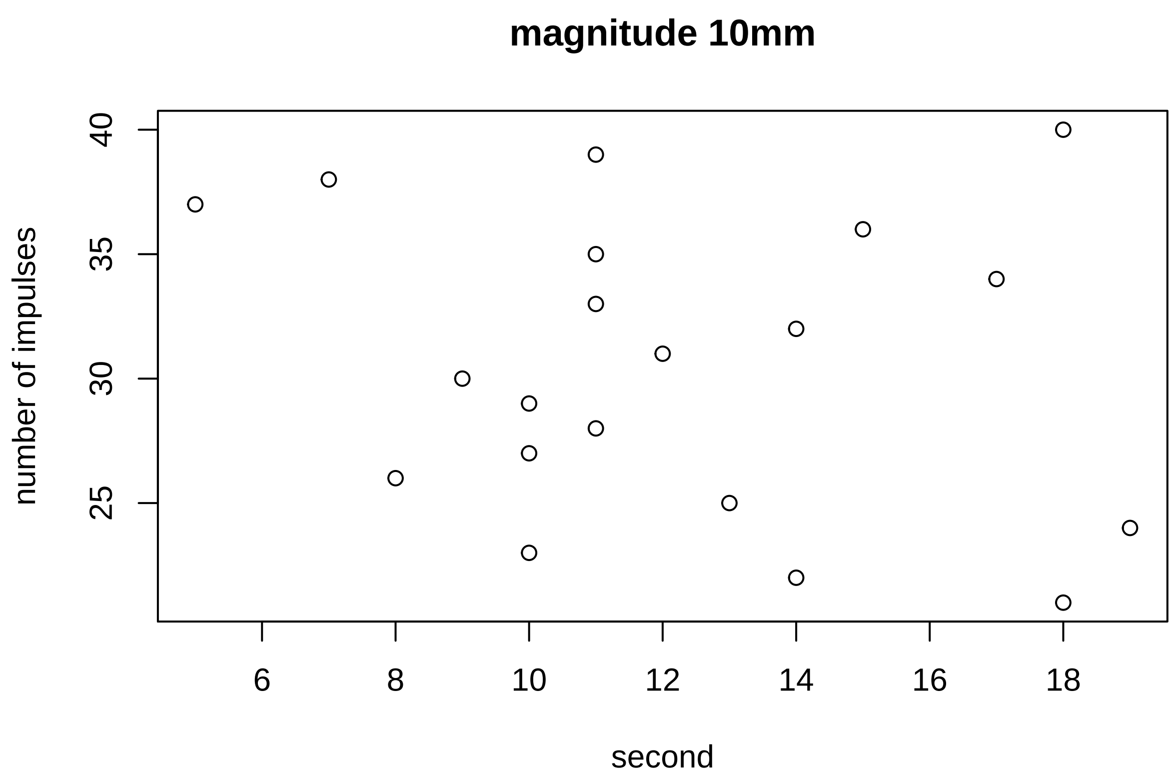
```r
Stc <- stc %>%
  group_by(Trial) %>%
  summarise(num_trial = length(Trial))
mag <- c(rep(5,20), rep(10,20), rep(15, 20))
boxplot(Stc$num_trial~mag, data = stc, xlab = "stratch magnitude", ylab = "number of impulse")
```



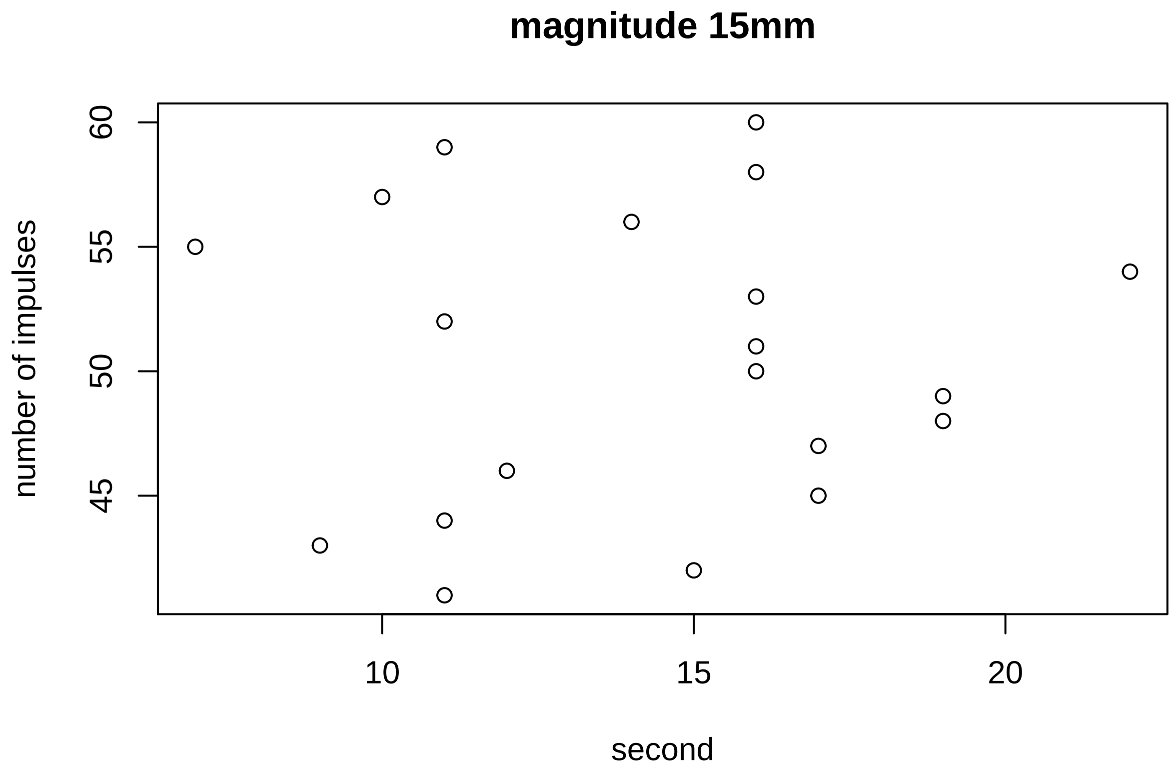

```r
secondinterval <- 1:60
plot(Stc$num_trial[1:20], secondinterval[1:20], main = "magnitude 5mm", xlab = "second", ylab = "number
```




```r
plot(Stc$num_trial[21:40], secondinterval[21:40], main = "magnitude 10mm", xlab = "second", ylab = "numb
```




```r
plot(Stc$num_trial[41:60], secondinterval[41:60], main = "magnitude 15mm", xlab = "second", ylab = "numb
```




```
#b
m1 <- glm(Stc$num_trial~mag, family = poisson)
mag_f <- factor(mag, ordered = F)
m2 <- glm(Stc$num_trial~mag_f, family = poisson)
summary(m1)
```

```
##
## Call:
## glm(formula = Stc$num_trial ~ mag, family = poisson)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.36645  -0.91061  -0.04635   0.44639   2.47577
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.50228    0.12055  12.462  < 2e-16 ***
## mag          0.08149    0.01006   8.099 5.53e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 135.919  on 59  degrees of freedom
## Residual deviance:  67.589  on 58  degrees of freedom
## AIC: 318.29
##
## Number of Fisher Scoring iterations: 4
```

```r
summary(m2)
```

```
##
## Call:
## glm(formula = Stc$num_trial ~ mag_f, family = poisson)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3283  -0.7725   0.0302   0.5075   1.9420
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.74920    0.09325  18.758  < 2e-16 ***
## mag_f10      0.74813    0.11319   6.610 3.85e-11 ***
## mag_f15      0.90756    0.11047   8.215  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 135.919  on 59  degrees of freedom
## Residual deviance:  55.684  on 57  degrees of freedom
## AIC: 308.39
##
## Number of Fisher Scoring iterations: 4
```

```r
anova(m1,m2, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Stc$num_trial ~ mag
## Model 2: Stc$num_trial ~ mag_f
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1        58     67.589
## 2        57     55.684  1   11.905 0.0005598 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For model of covariate, all parameters are significant at 95% level. Intercept means if magnitude is zero, the impulse number would be exp(1.50228) which is about 4. As magnitude increaes by one, impulse number would be increase to exp(1.50228) * exp(0.08149), which is about 5.

For model of factors, all parameters are significant at 95% level. Intercept means if magnitude is 5mm, the impulse number would be exp(1.7492) which is about 6. As magnitude goes to 10mm level, impulse number would be exp(1.7492) * exp(0.74813), which is about 12. As magnitude goes to 15mm level, impulse number would be exp(1.7492) * exp(0.90756), which is about 14.

The analysis of deviance table suggests that factors model is preferred even though both models' parameters are all significant at 95% level. For covariate model, the dispersion is about 67.589 /58 = 1.165328 and 55.684 / 57 = 0.9769123 for the factors model. The covariate model has slight overdispersion and the factor model has little underdispersion, but they are all very close to one. The following is the chi-sq test.Ttheir p-values are all larger than 0.05, the overdispersion and underdispersion might be not dignificant

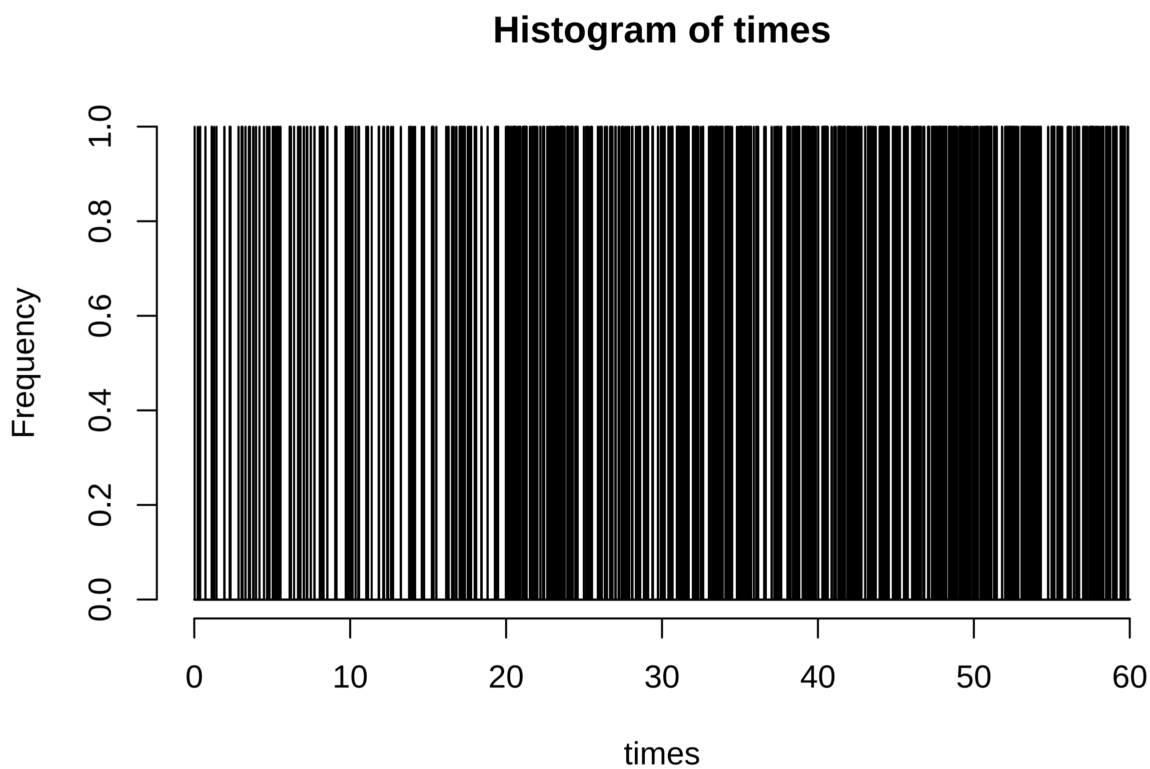
```r
with(m1, cbind(res.deviance = deviance, df = df.residual, p = pchisq(deviance, df.residual, lower.tail=
```

```
##      res.deviance df         p
## [1,]      67.5893 58 0.1822919
```

```r
with(m2, cbind(res.deviance = deviance, df = df.residual, p = pchisq(deviance, df.residual, lower.tail='
```

```
##      res.deviance df         p
## [1,]      55.6842 57 0.4754463
```

```r
#c
times <- (Trial + SpikeTimes - 1)
mag_tt <- c(rep(5,20000), rep(10,20000), rep(15, 20000))
spike1ms <- hist(times, breaks = seq(0,60,0.001))$counts
bins <- hist(times, breaks = seq(0,60,0.001))$breaks[-1]
```




```r
history <- rep(bins[1:1000],60)
m3 <- glm(spike1ms~mag_tt+history, poisson)
summary(m3)
```

```
##
## Call:
## glm(formula = spike1ms ~ mag_tt + history, family = poisson)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.2172  -0.1638  -0.1393  -0.1182   3.0025
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.96796    0.13423 -37.011  < 2e-16 ***
## mag_tt       0.08149    0.01006   8.099 5.53e-16 ***
## history     -0.94850    0.13968  -6.791 1.12e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 5833.2  on 59999  degrees of freedom
## Residual deviance: 5717.8  on 59997  degrees of freedom
## AIC: 7009.8
##
## Number of Fisher Scoring iterations: 7
```

```
with(m3, cbind(res.deviance = deviance, df = df.residual, p = pchisq(deviance, df.residual, lower.tail=
```

```
##      res.deviance    df p
## [1,]     5717.762 59997 0
```

```
anova(m3, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: spike1ms
##
## Terms added sequentially (first to last)
##
##
##         Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                    59999     5833.2
## mag_tt   1   68.329     59998     5764.9 < 2.2e-16 ***
## history  1   47.147     59997     5717.8 6.586e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(dp <-sum(residuals(m3,type="pearson")^2)/m3$df.res)
```

```
## [1] 0.9795825
```

```
summary (m3, dispersion=dp)
```

```
##
## Call:
## glm(formula = spike1ms ~ mag_tt + history, family = poisson)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.2172  -0.1638  -0.1393  -0.1182   3.0025
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.967959   0.132850 -37.395  < 2e-16 ***
## mag_tt       0.081492   0.009958   8.183 2.76e-16 ***
## history     -0.948497   0.138244  -6.861 6.84e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 0.9795825)
##
##     Null deviance: 5833.2  on 59999  degrees of freedom
```

```
## Residual deviance: 5717.8  on 59997  degrees of freedom
## AIC: 7009.8
##
## Number of Fisher Scoring iterations: 7
```

Under 95% confidence level, the rate does not depends on time. Intercept means if magnitude and time is zero, the impulse number would be exp(-5.450402) which is close 0. As magnitude increaes by one, impulse number would be increase to exp(-5.450402) * exp(0.099332), which is still very close to zero. As times increaes by one, impulse number would be increase to exp(-5.450402) * exp(-0.004460), which is very close to zero. The underdispersion is significant.

```
m4 <- glm(spike1ms[2:60000]~spike1ms[1:59999], poisson)
m5 <- glm(spike1ms[3:60000]~spike1ms[2:59999] + spike1ms[1:59998], poisson)
m6 <- glm(spike1ms[4:60000]~spike1ms[3:59999] + spike1ms[2:59998] + spike1ms[1:59997] , poisson)
m7 <- glm(spike1ms[5:60000]~spike1ms[4:59999] + spike1ms[3:59998] + spike1ms[2:59997] + spike1ms[1:5999
m8 <- glm(spike1ms[6:60000]~spike1ms[5:59999] + spike1ms[4:59998] + spike1ms[3:59997] + spike1ms[2:5999
AIClm <- function(l){
  AIC <- extractAIC(l, k = 2)[2]
  return(AIC)
}
matrix(unlist(lapply(list(m4, m5, m6,m7,m8),AIClm)), byrow = TRUE, ncol = 1, dimnames = list(c("m4", "m
```

```
##         AIC
## m4 7109.361
## m5 7097.332
## m6 7096.320
## m7 7098.175
## m8 7098.966
```

```
summary(m6)
```

```
##
## Call:
## glm(formula = spike1ms[4:60000] ~ spike1ms[3:59999] + spike1ms[2:59998] +
##     spike1ms[1:59997], family = poisson)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.1485  -0.1485  -0.1485  -0.1485   2.9571
##
## Coefficients:
##                    Estimate Std. Error  z value Pr(>|z|)
## (Intercept)        -4.50790    0.03953 -114.042   <2e-16 ***
## spike1ms[3:59999] -13.79468  225.47746   -0.061    0.951
## spike1ms[2:59998] -13.79468  225.47746   -0.061    0.951
## spike1ms[1:59997]  -0.85963    0.57870   -1.485    0.137
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 5833.2  on 59996  degrees of freedom
## Residual deviance: 5802.3  on 59993  degrees of freedom
## AIC: 7096.3
##
## Number of Fisher Scoring iterations: 16
```

```r
(dp2 <-sum(residuals(m6,type="pearson")^2)/m6$df.res)
```

```
## [1] 0.9679129
```

```r
summary(m6, dispersion = dp2)
```

```
##
## Call:
## glm(formula = spike1ms[4:60000] ~ spike1ms[3:59999] + spike1ms[2:59998] +
##     spike1ms[1:59997], family = poisson)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.1485  -0.1485  -0.1485  -0.1485   2.9571
##
## Coefficients:
##                    Estimate Std. Error  z value Pr(>|z|)
## (Intercept)        -4.50790    0.03889 -115.917   <2e-16 ***
## spike1ms[3:59999] -13.79468  221.83051   -0.062    0.950
## spike1ms[2:59998] -13.79468  221.83051   -0.062    0.950
## spike1ms[1:59997]  -0.85963    0.56934   -1.510    0.131
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 0.9679129)
##
##     Null deviance: 5833.2  on 59996  degrees of freedom
## Residual deviance: 5802.3  on 59993  degrees of freedom
## AIC: 7096.3
##
## Number of Fisher Scoring iterations: 16
```

```r
anova(m6, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: spike1ms[4:60000]
##
## Terms added sequentially (first to last)
##
##
##                   Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                              59996     5833.2
## spike1ms[3:59999]  1  13.8567     59995     5819.3 0.0001973 ***
## spike1ms[2:59998]  1  14.0076     59994     5805.3 0.0001821 ***
## spike1ms[1:59997]  1   2.9901     59993     5802.3 0.0837762 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The temporal dependencies exsists till lag of three. However, partial auto-correlation also exists with lag of 1 and of 2. The coefficients stand for partial auto-correlation between the impulse count(t) and impulse count(t+h). It suggests that the dispersion I get from previous parts are enlarged because covariance now taken into account.