

STAT-S632

Assignment 4

John Koo

```
# packages, etc.
import::from(magrittr, `%>%`, `%<>%`, multiply_by, set_names)
dp <- loadNamespace('dplyr')
library(ggplot2)

theme_set(theme_bw())
```

Problem 1

The optimization step for Newton-Raphson is:

$$\beta^{(t+1)} = \beta^{(t)} + I^{-1}(\beta) \frac{\partial \ell}{\partial \beta}$$

Since we want to find the root of $\frac{\partial \ell}{\partial \beta}$.

$I(\beta)$ is the Hessian matrix of ℓ .

We know that $I(\beta) = -\frac{\partial^2}{\partial \beta^2} \ell(\beta)$. For Poisson regression, we have

$$\ell(\beta) = \sum_i^n \left(y_i x_i^T \beta - e^{x_i^T \beta} - \log(y_i!) \right)$$

So:

$$\frac{\partial}{\partial \beta} \ell(\beta) = \sum_i^n \left(y_i x_i^T - x_i^T e^{x_i^T \beta} \right)$$

Differentiating this again with respect to β , we obtain:

$$\begin{aligned} \frac{\partial^2}{\partial \beta^2} \ell(\beta) &= \sum_i^n x_i x_i^T e^{x_i^T \beta} \\ &= X^T \begin{bmatrix} e^{x_1^T \beta} & & & \\ & e^{x_2^T \beta} & & \\ & & \ddots & \\ & & & e^{x_n^T \beta} \end{bmatrix} X \\ &= X^T D X \end{aligned}$$

We can also say:

$$\frac{\partial \ell}{\partial \beta} = \sum_i y_i x_i^T - \sum_i x_i^T e^{x_i^T \beta}$$

$$\begin{aligned}
&= \sum_i x_i^T (y_i - e^{x_i^T \beta}) \\
&= X^T (y - \mu)
\end{aligned}$$

Where $\mu \in R^n$ and $\mu_i = e^{x_i^T \beta}$.

Putting it all together, we get:

$$\beta^{(t+1)} = \beta^{(t)} + (X^T D X)^{-1} X^T (y - \mu)$$

And using the same trick as we did for binomial regression:

$$\begin{aligned}
&= (X^T D X)^{-1} X^T D X \beta^{(t)} + (X^T D X)^{-1} X^T D D^{-1} (y - \mu) \\
&= (X^T D X)^{-1} X^T D (X \beta^{(t)} + D^{-1} (y - \mu))
\end{aligned}$$

Then we can let $z = X^T \beta^{(t)} + D^{-1} (y - \mu)$ to obtain:

$$\beta^{(t+1)} = (X^T D X)^{-1} X^T D z$$

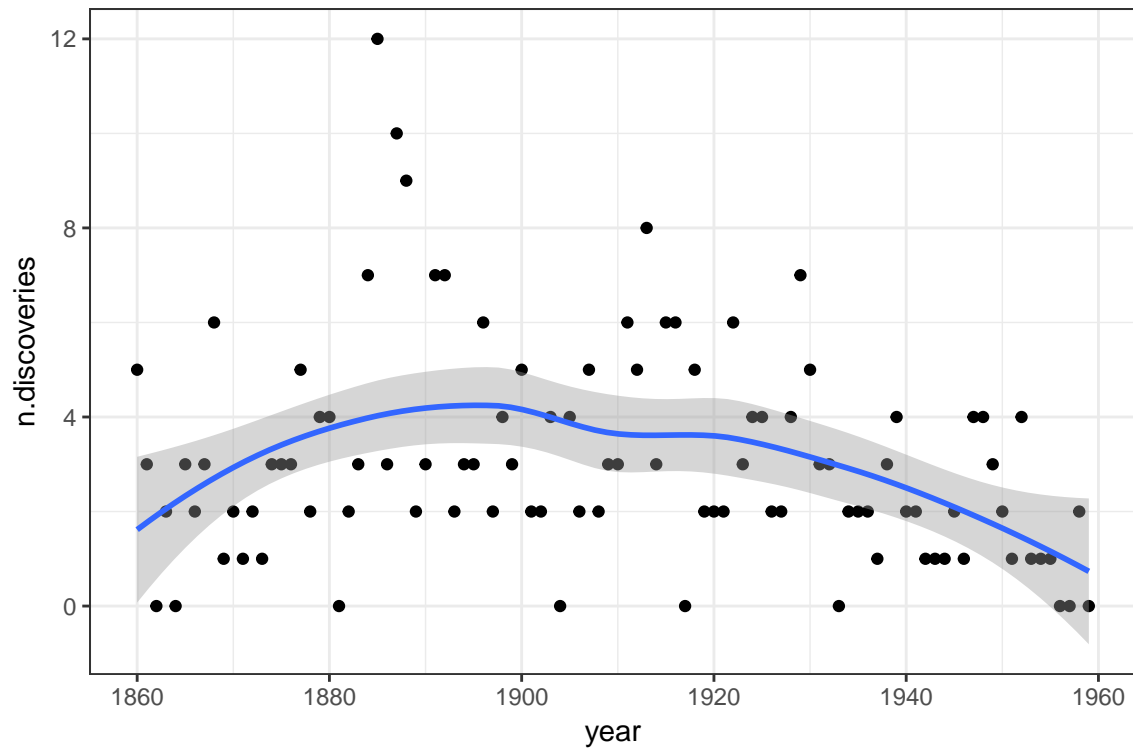
Problem 2

[From ELM 5.1]

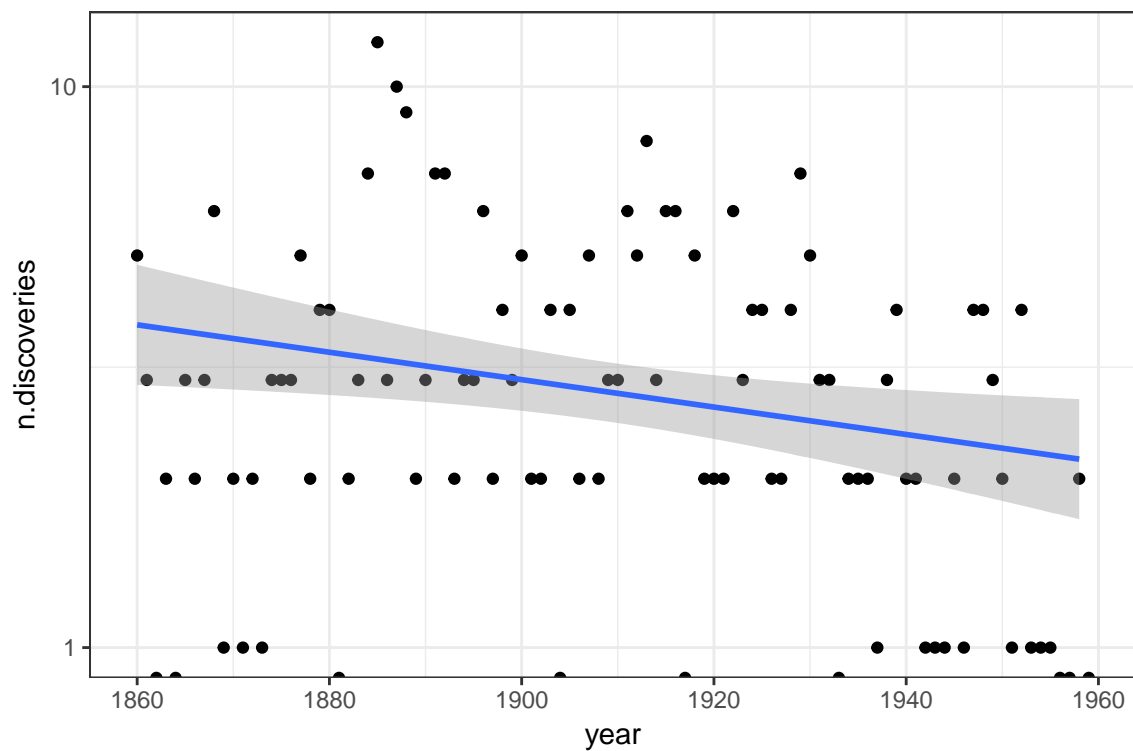
Part a

```
discoveries.df <- dplyr::data_frame(
  year = seq(1860, 1959),
  n.discoveries = discoveries
)

ggplot(discoveries.df) +
  geom_point(aes(x = year, y = n.discoveries)) +
  stat_smooth(aes(x = year, y = n.discoveries))
```



```
ggplot(discoveries.df) +  
  geom_point(aes(x = year, y = n.discoveries)) +  
  stat_smooth(aes(x = year, y = n.discoveries), method = 'lm') +  
  scale_y_log10()
```



There doesn't appear to be any definitive trend, or at least no definitive linear trend. If there is a linear trend, it would be negative.

Part b

```
constant.mod <- glm(n.discoveries ~ 1, data = discoveries.df,  
                    family = poisson)  
summary(constant.mod)
```

Call:

```
glm(formula = n.discoveries ~ 1, family = poisson, data = discoveries.df)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.4900	-0.6686	-0.0571	0.4890	3.8320

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.1314	0.0568	19.92	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

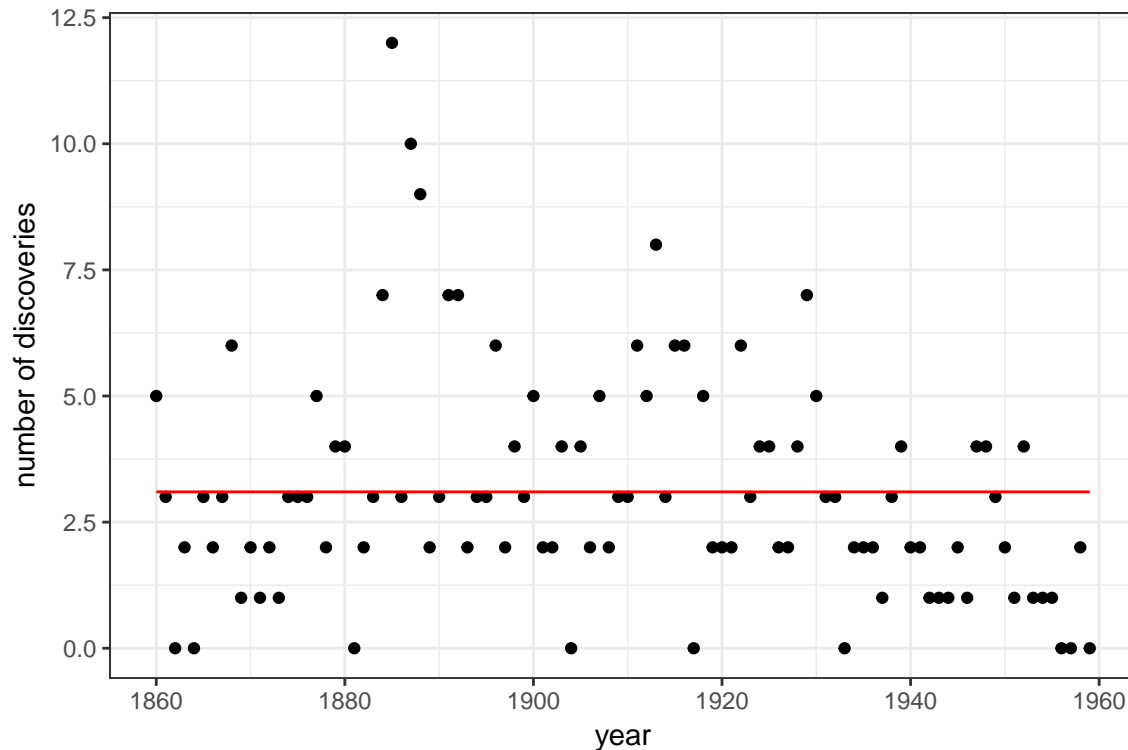
Null deviance: 164.68 on 99 degrees of freedom

Residual deviance: 164.68 on 99 degrees of freedom

AIC: 435.69

Number of Fisher Scoring iterations: 5

```
discoveries.df %<>%  
  dp$mutate(n.pred = predict(constant.mod,  
                             newdata = discoveries.df,  
                             type = 'response'))  
  
ggplot(discoveries.df) +  
  geom_point(aes(x = year, y = n.discoveries)) +  
  geom_line(aes(x = year, y = n.pred), colour = 'red') +  
  labs(y = 'number of discoveries')
```



```
mean(discoveries.df$n.discoveries)
```

```
[1] 3.1
```

```
unnname(exp(constant.mod$coefficients))
```

```
[1] 3.1
```

We can see that the exponential of the coefficient of the intercept model is equal to the average number of discoveries.

Part c

```
pchisq(constant.mod$deviance, constant.mod$df.residual, lower.tail = FALSE)
```

```
[1] 3.79455e-05
```

Since $P(\chi^2 > 0) \approx 0$, we have evidence that the intercept model is a not so great fit to these data. This suggests that additional coefficients might be appropriate, i.e., there is some trend in the number of discoveries over time.

Part d

```
agg.disc.df <- discoveries.df %>%
  dp$mutate(n.discoveries = ifelse(n.discoveries >= 8, '>=8', n.discoveries),
           n.discoveries = factor(n.discoveries,
                                levels = c(as.character(seq(0, 7)), '>=8')) %>%
  dp$group_by(n.discoveries) %>%
  dp$summarise(freq = n()) %>%
```

```

dp$ungroup()

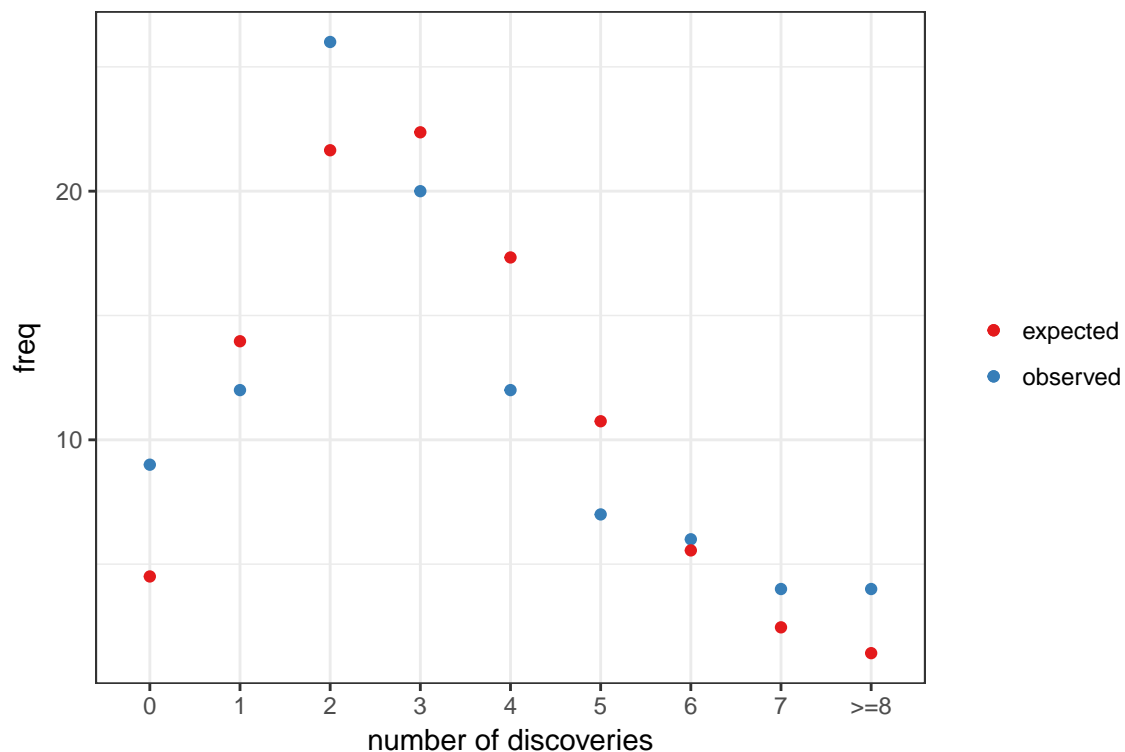
expected.freq <- c(dpois(seq(0, 7), mean(discoveries)),
  ppois(7, mean(discoveries), lower.tail = FALSE)) %>%
  multiply_by(length(discoveries)) %>%
  set_names(c(seq(0, 7), '>=8'))

agg.disc.df %<>% dp$mutate(expected.freq)
agg.disc.df

# A tibble: 9 x 3
  n.discoveries freq expected.freq
  <fct>         <int>         <dbl>
1 0             9          4.50
2 1            12         14.0
3 2            26         21.6
4 3            20         22.4
5 4            12         17.3
6 5             7         10.7
7 6             6          5.55
8 7             4          2.46
9 >=8           4          1.42

ggplot(agg.disc.df) +
  geom_point(aes(x = n.discoveries, y = freq, colour = 'observed')) +
  geom_point(aes(x = n.discoveries, y = expected.freq, colour = 'expected')) +
  labs(x = 'number of discoveries', colour = NULL) +
  scale_colour_brewer(palette = 'Set1')

```



The two do not seem to agree. The observed seem to prefer both fewer and more discoveries while the

expected is a bit more evenly spread out.

Part e

```
chisq <-  
  sum((agg.disc.df$freq - agg.disc.df$expected.freq)^2 / agg.disc.df$expected.freq)  
pchisq(chisq, nrow(agg.disc.df) - 1, lower.tail = FALSE)
```

```
[1] 0.06924174
```

According to the χ^2 -test, there is *some* evidence that this isn't from a Poisson distribution.

We might suspect that if we do not group the ≥ 8 , the p -value would decrease.

Part f

```
quad.mod <- glm(n.discoveries ~ year + I(year ** 2), data = discoveries.df,  
               family = poisson)  
summary(quad.mod)
```

Call:

```
glm(formula = n.discoveries ~ year + I(year^2), family = poisson,  
    data = discoveries.df)
```

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 ***
year	1.561e+00	3.318e-01	4.705	2.54e-06 ***
I(year^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

The Wald test shows that the quadratic term is significant (i.e., not equal to 0). We can also try a LR test:

```
lin.mod <- glm(n.discoveries ~ year, data = discoveries.df, family = poisson)  
anova(quad.mod, lin.mod, test = 'Chi')
```

Analysis of Deviance Table

Model 1: n.discoveries ~ year + I(year^2)
Model 2: n.discoveries ~ year

```

      Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1          97      132.84
2          98      157.32 -1   -24.477 7.519e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

And the two tests agree.

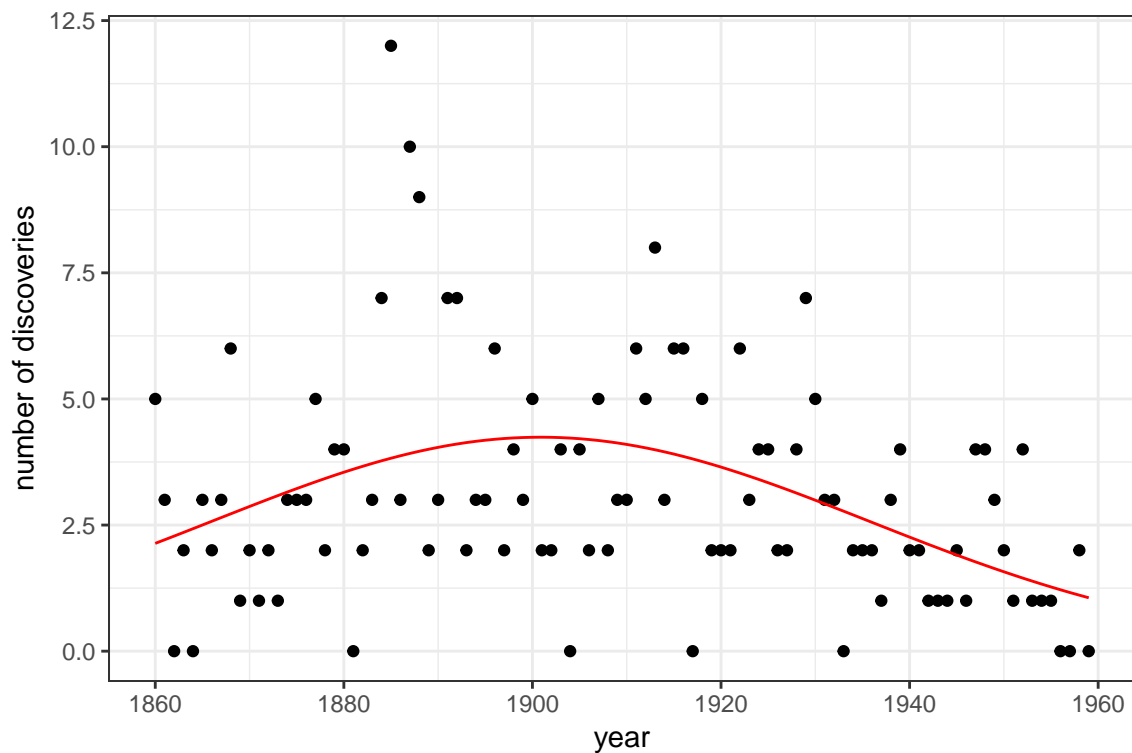
Part g

```

discoveries.df %<>%
  dp$mutate(n.pred = predict(quad.mod,
                             newdata = discoveries.df,
                             type = 'response'))

ggplot(discoveries.df) +
  geom_point(aes(x = year, y = n.discoveries)) +
  geom_line(aes(x = year, y = n.pred), colour = 'red') +
  labs(y = 'number of discoveries')

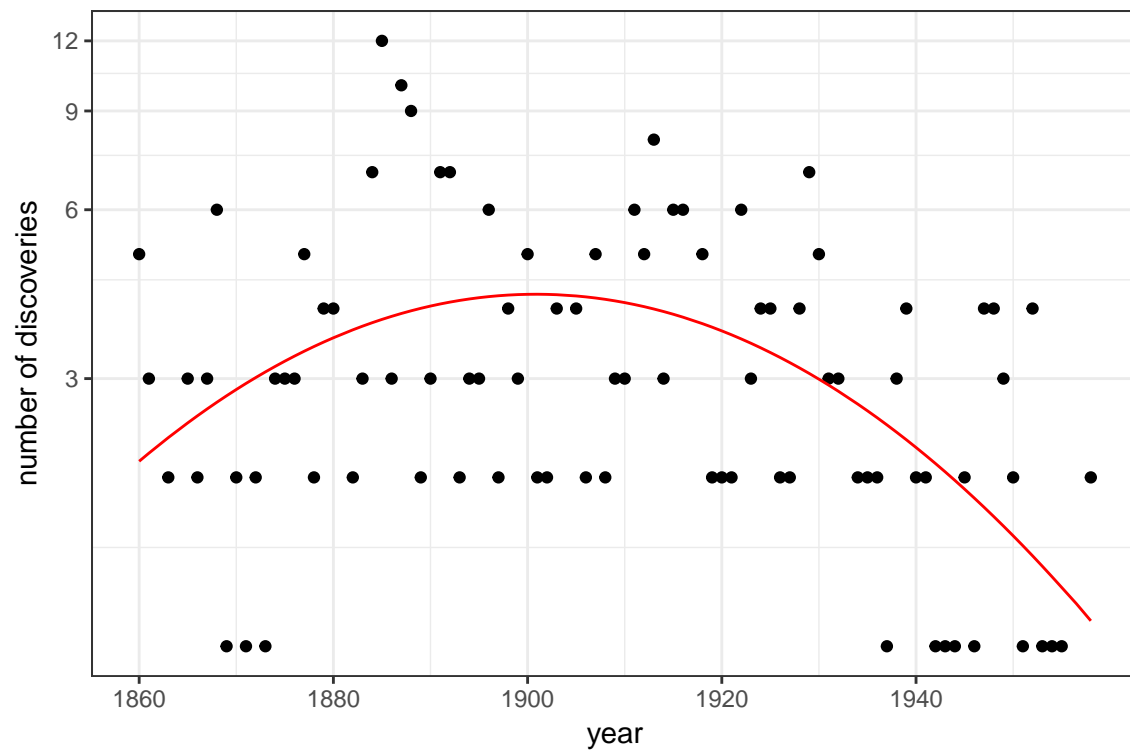
```



```

ggplot(discoveries.df %>% dp$filter(n.discoveries > 0)) +
  geom_point(aes(x = year, y = n.discoveries)) +
  geom_line(aes(x = year, y = n.pred), colour = 'red') +
  labs(y = 'number of discoveries') +
  coord_trans(y = 'log10')

```

The second plot compares $X\beta$ to x_1 , and here we should see a quadratic relationship.