放射線被ばくデータに使われる統計解析手法とその解釈

Kenichi Satoh

last update: 2019/05/24

Contact to:

- kenichi-satoh@biwako.shiga-u.ac.jp
- https://sites.google.com/view/ksatoh/english

Keywords

- Weighted linear regression, Poisson regression

Required Packages and Data Files

- dosecell.csv

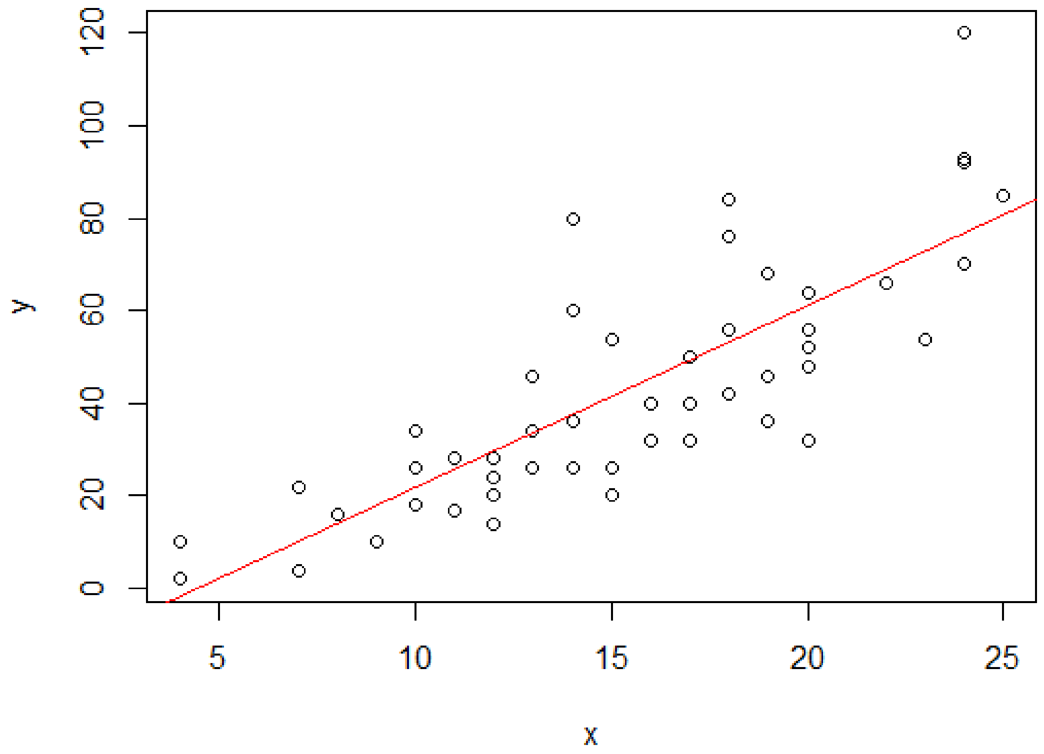
1. Simple linear regression

1.1. Scatter plot

```
head(cars)
```

```
##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
## 6     9   10
```

```
y <- cars$dist
x <- cars$speed
plot(x,y)
res <- lm(y~x)
abline(res,col="red")
```



```
print(res)

##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##     -17.579        3.932
```

- Estimated intercept is -17.579 and its slope is 3.932.

1.2. Backgroud calculation for estimates

```
mylm <- function(par) sum((y-(par[1]+par[2]*x))^2)
res <- optim(par=c(0,0),fn=mylm)
print(res$par)

## [1] -17.571729   3.931832
```

- The “par” is an argument of the function “mylm”. The first element of par, par[1] is an intercept and the second par[2] is a slope.
- The function “optim” will search the parameters, par[1] and par[2] which mimimizes the residual sum of squares expressed as “mylm”. Those initial values are given by par=c(0,0).
- The optimized regression coefficients are -17.57 and 3.93 which are almost consistent with above estimated values.

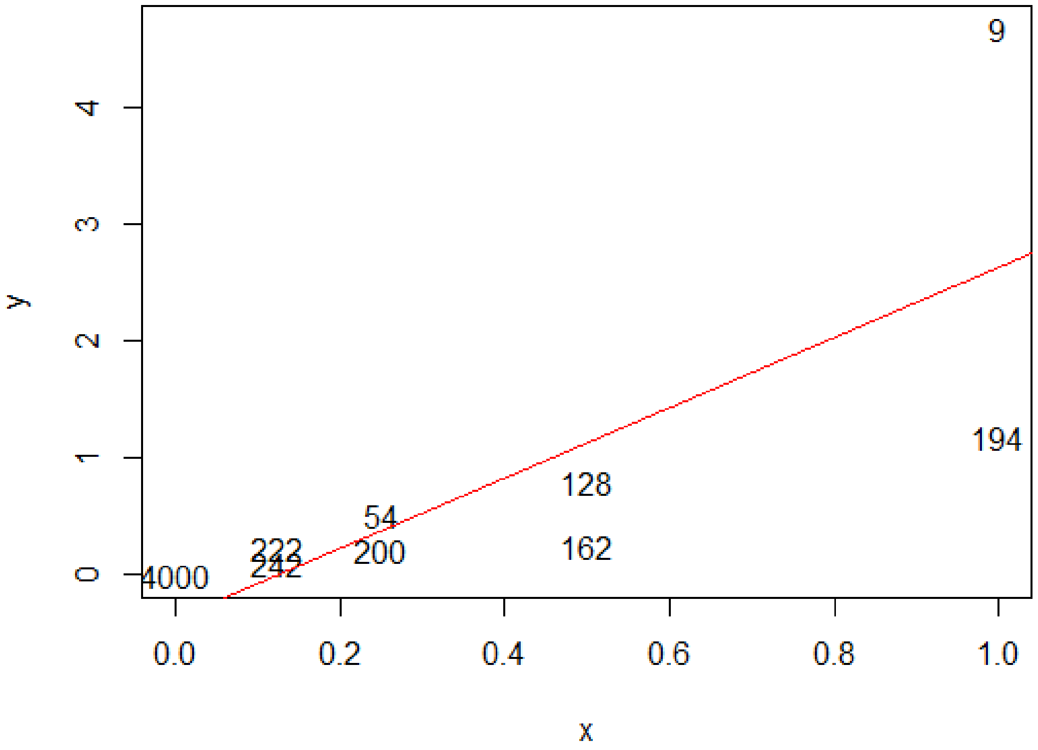
2. Weighted linear regression

2.1. Ordinal linear regression

```
d <- read.csv("dosecell.csv")
d

##    dose cell dicentric
## 1 0.000 4000         2
## 2 0.125  242        22
## 3 0.125  222        51
## 4 0.250  200        42
## 5 0.250   54        28
## 6 0.500  162        40
## 7 0.500  128       102
## 8 1.000  194       229
## 9 1.000    9        42

n <- d$cell #Each cell includes 46 chromosomes
obs <- d$dicentric
y <- obs/n # response is a chromosome abnormality rate per cell
x <- d$dose # explanatory variable is a irraddiated radiation dose (gray).
# visualization
plot(x,y,type="n")
text(x,y,n) # text shows the number of cells to be used
res <- lm(y~x) # ordinal linear regression
abline(res,col="red")
```



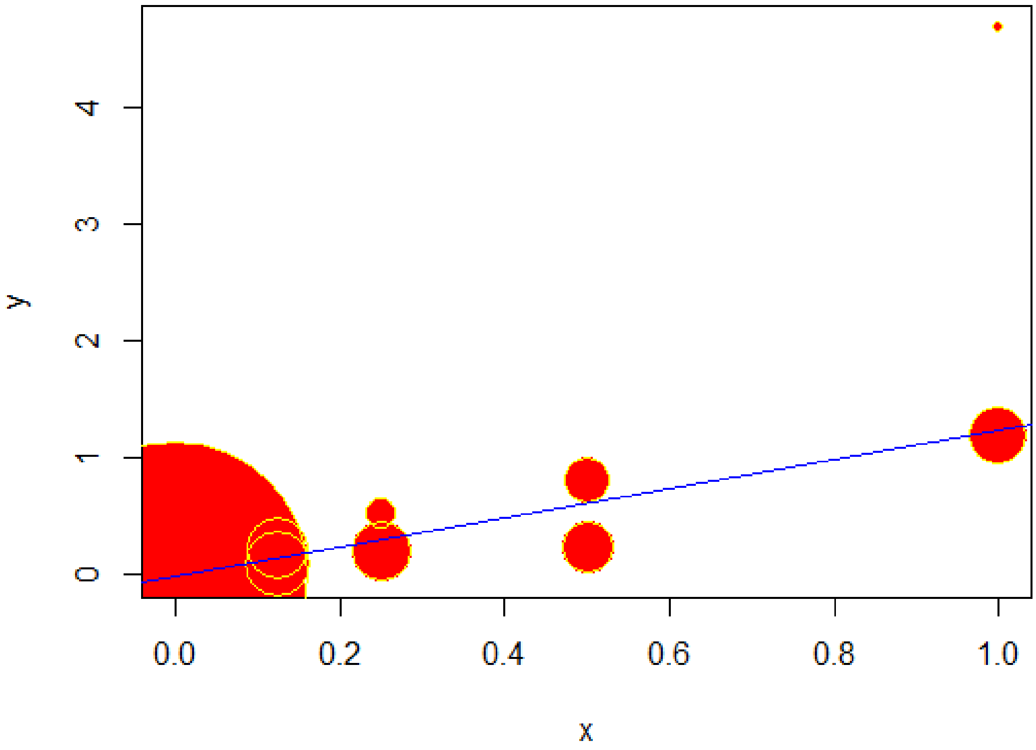
- Each point has different number of cells to be experimented.
- The point with “9” cells seems to pull the fitted line upwards even the number of obserbed cells is much less than others.

2.2. Put the weights as the number of samples in each point

```
mylm <- function(par) sum(n*(y-(par[1]+par[2]*x))^2)
res <- optim(par=c(0,0),fn=mylm)
print(res$par)

## [1] -0.004852846  1.241992855

# visualization
plot(x,y,cex=sqrt(n)*0.3,pch=19,col="red")
points(x,y,cex=sqrt(n)*0.3,pch=21,col="yellow")
abline(res$par,col="blue")
```



- The area of circle is in proportion to the number of cells.
- The number of cells, n can be used as the number of samples in each point. For example, n*(y-(par[1]+par[2]*x))^2 corresponds to the residual sum of squares for n overlapped points at the same location (x,y).
- In case where n is one, it corresponds to an ordinal one sample case.

2.3. Weihted linear regression by lm

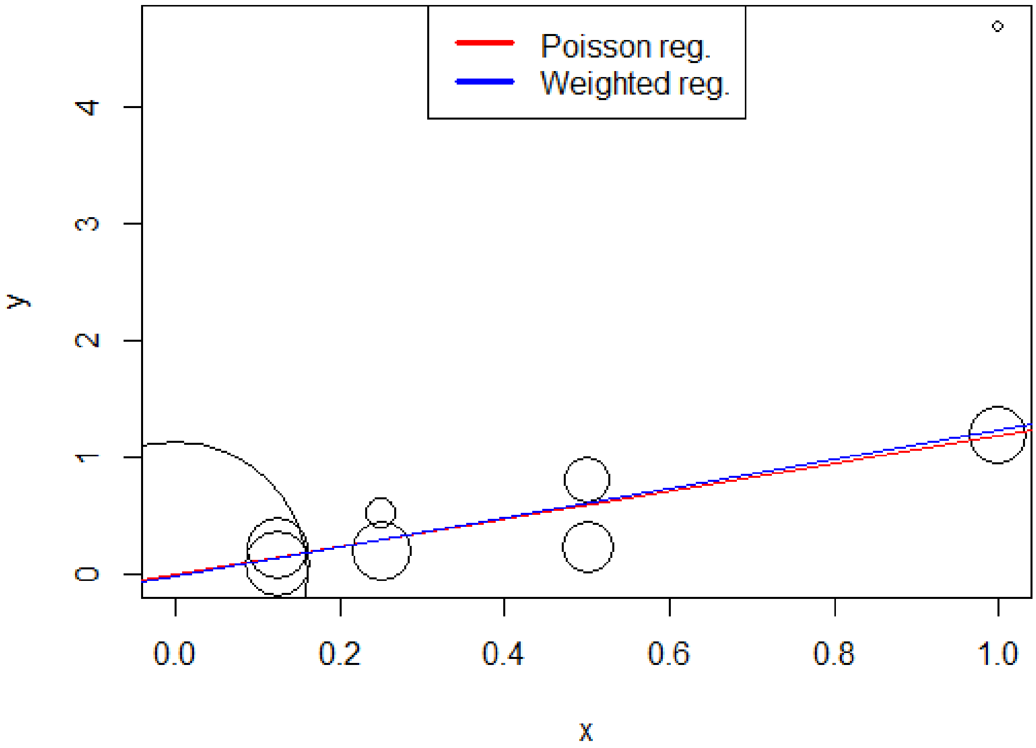
```
res <- lm(y~x,weights=n)
coef.w <- res$coefficients
print(coef.w)

##  (Intercept)            x
## -0.004818945  1.241984076
```

- Weighted linear regression can be implemented by setting “weights” option of “lm” function.

2.4. Relation to Poisson regression

```
# y/n=a+bx <=> y = an+bnx
res <- glm(obs~0+n+I(n*x),
         family=poisson(link=identity))
coef.poi <- res$coefficients
# visualization
plot(x,y,cex=sqrt(n)*0.3,pch=1)
abline(coef.poi,col="red")
abline(coef.w,col="blue")
legend("top",lwd=3,col=c("red","blue"),
  legend=c("Poisson reg.","Weighted reg."))
```



- Both resultant linear curves is quite similar.