# Regression Model

# C7 Week1

install.packages("UsingR")

```
library(UsingR)
```

```
## Loading required package: MASS
## Loading required package: HistData
## Loading required package: Hmisc
## Loading required package: grid
## Loading required package: lattice
## Loading required package: survival
## Loading required package: splines
## Loading required package: Formula
##
## Attaching package: 'Hmisc'
##
## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
##
## Loading required package: aplpack
## Loading required package: tcltk
## Loading required package: quantreg
## Loading required package: SparseM
##
## Attaching package: 'SparseM'
##
## The following object is masked from 'package:base':
##
##     backsolve
##
##
## Attaching package: 'quantreg'
##
## The following object is masked from 'package:Hmisc':
##
##     latex
##
## The following object is masked from 'package:survival':
##
##     untangle.specials
##
##
## Attaching package: 'UsingR'
```

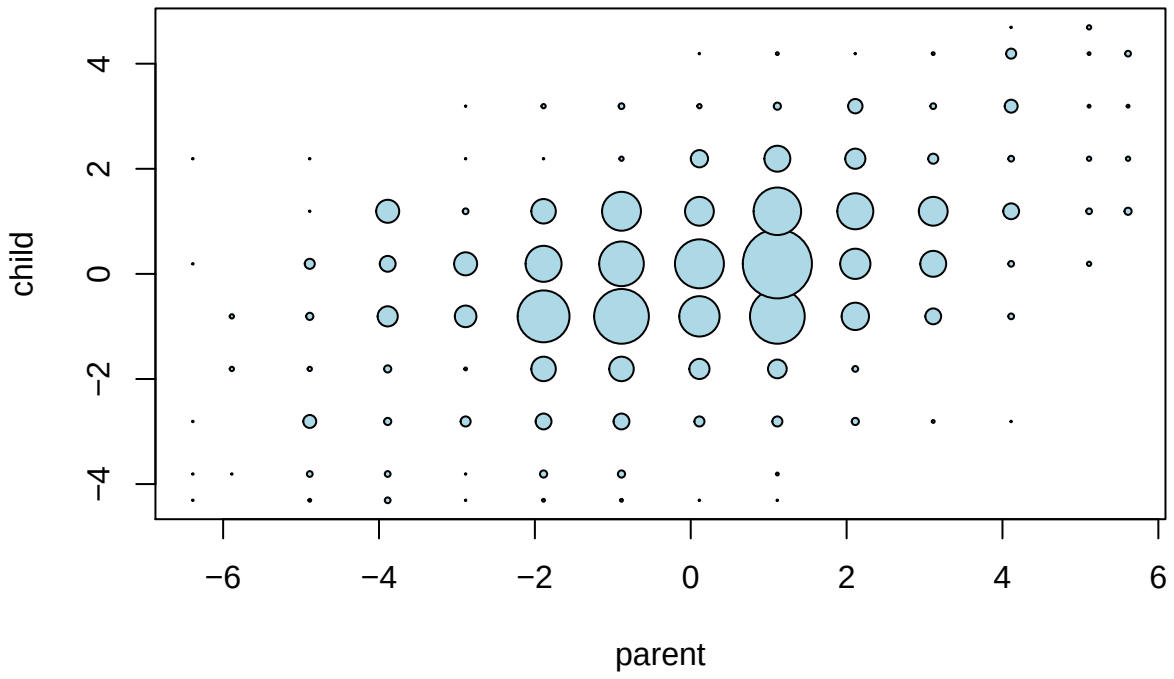
```
##
## The following object is masked from 'package:survival':
##
##     cancer
```

```
data(galton)
str(galton)
```

```
## 'data.frame':    928 obs. of  2 variables:
##  $ child : num  61.7 61.7 61.7 61.7 61.7 62.2 62.2 62.2 62.2 62.2 ...
##  $ parent: num  70.5 68.5 65.5 64.5 64 67.5 67.5 67.5 66.5 66.5 ...
```

Get1 cex

```
y<-galton$child-mean(galton$child)
x<-galton$parent-mean(galton$parent)
freqdata<-as.data.frame(table(x,y))
names(freqdata)<-c("child","parent","freq")
plot(
    as.numeric(as.vector(freqdata$parent)),
    as.numeric(as.vector(freqdata$child)),
    pch=21,col="black",bg="lightblue",
    cex=0.1*freqdata$freq,   #this command sets the size of the point
    xlab="parent",ylab="child"
    )
```



R `lm`

### residuals

- lm() - predict,summary,resid,coef

## week2

## week3

#### dummay variable

R

relevel()

## week4

### GLM

- 1. f - 2. $\eta = X\beta$ - 3. g $E(y) = \mu = g^{-1}(X\beta)$

logistic model

### 1.binary data

#### logit

- 1 p
- odds: $\frac{p}{1-p}$
- log odds: $\log(\frac{p}{1-p})$

$$\log(\frac{p_i}{1-p_i}) = b_0 + b_1 x_i$$

```
##
setwd("F:/05Course/Data science/7.Regreesion Model")
download.file("https://dl.dropboxusercontent.com/u/7710864/data/ravensData.rda",destfile="ravensData.rda
load("ravensData.rda")
head(ravensData)
```

logit

```
logRegRavens<-glm(ravenWinNum~ravenScore,data=ravensData,family="binomial")
summary(logRegRavens)
```

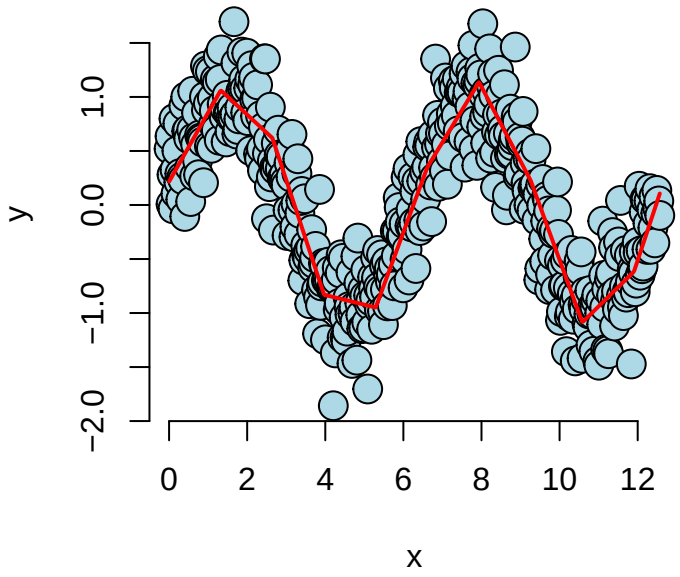
http://data.princeton.edu/R/glms.html

### 2.count variable——possion regression

### fit functions

splines ## Simulated example

```
n <- 500
x <- seq(0, 4 * pi, length = n)
y <- sin(x) + rnorm(n, sd = .3)
knots <- seq(0, 8 * pi, length = 20)
splineTerms <- sapply(knots, function(knot) (x > knot) * (x - knot))
xMat <- cbind(1, x, splineTerms)
yhat <- predict(lm(y ~ xMat - 1))
plot(x, y, frame = FALSE, pch = 21, bg = "lightblue", cex = 2)
lines(x, yhat, col = "red", lwd = 2)
```




## week 2 homework

```
## starting httpd help server ... done
```

## week3 quize

## week4 quize

library(MASS) ?shuttle str(shuttle) shuttle$use = as.numeric(shuttle$use) shuttle$use[shuttle$use==2]=0 relevel(shuttle$wind,ref="tail")

fit= glm(use~wind,data=shuttle,family="binomial") summary(fit)$coefficients[2,1] exp(summary(fit)$coefficients[2,1])

### 2

rm(list=ls()) library(MASS) str(shuttle) shuttle$use = as.numeric(shuttle$use) shuttle$use[shuttle$use==2]=0 fit=glm(use~wind+magn,data=shuttle,family="binomial")

1/exp(fit$coefficients2)

### 4

data(InsectSprays) str(InsectSprays) glm(count~spray,family="poisson",data=InsectSprays) 1/exp(0.05588)

### 6

x <- -5:5 y <- c(5.12, 3.93, 2.67, 1.87, 0.52, 0.08, 0.93, 2.05, 2.54, 3.87, 4.97)

knots <-0 splineTerms <- sapply(knots, function(knot) (x > knot) * (x - knot)) xMat <- cbind(1, x, splineTerms) lm(y ~ xMat - 1)

## swirl

library(swirl)

install_from_swirl("Regression Models")