# Regression Model

# C7 Week1 学习笔记

这一章的笔记没有按照讲课的知识点来，只是记录了个人没太遇到过的东西。  
关于回归分析如果听了课还是不太明白，就找些基本的书来看看就行。

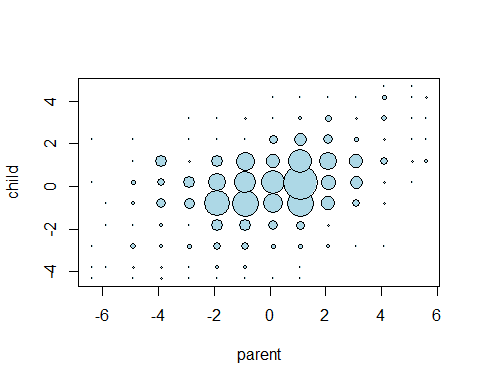
library(UsingR)

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 ...

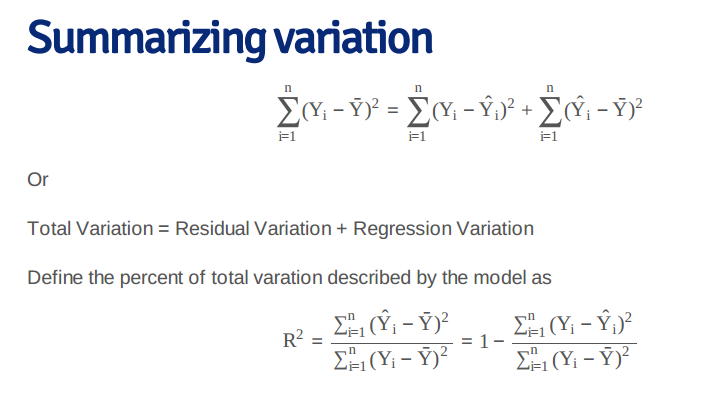
Get1：设置散点图中圆圈的大小，用cex参数

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"  
 )



## 最小二乘估计

一元回归分析的基本假定：



R语言中做回归的是lm函数

# Week2 学习笔记

## 残差residuals

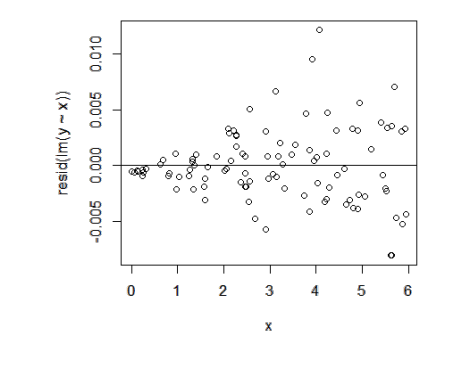
对于通过lm()函数fit的模型 fit<-lm(y~x)

其有以下的常用的属性：

Predict(),summary(),coef(),resid()

## 残差residuals

异方差：残差是与x有关的，如下的残差是随着x的变大而变大的，而不是随机分布在0附近的。



# Week4 学习笔记

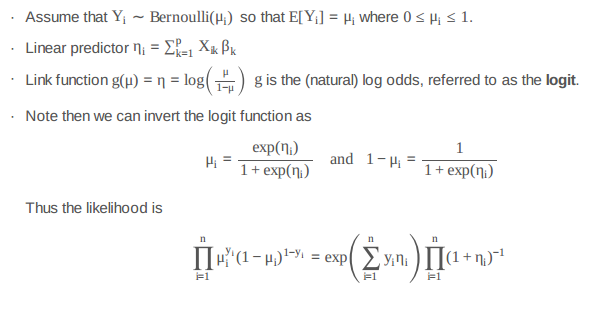
广义线性模型包含了以下主要部份：

1. 响应变量y来自指数分布族

2. 线性预测算子*η*=*Xβ*

3. 链接函数g使得*E*(*y*)=*μ*=*g*−1(*Xβ*)

比如logistic model



## binary data

1. 概率：结果为1的概率p
2. odds: *p*1−*p*
3. log odds: log(*p*1−*p*)

### 示例

#数据下载不了  
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回归

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

更多资料： <http://data.princeton.edu/R/glms.html>

## 2.count variable——possion regression

## 3. fit functions

splines

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)

# Swirl

在绘制散点图的时候，有时候由于点比较密，一些点就覆盖在了一起，但是不会显示出来。利用jitter函数，可以将紧密程度显示出来

plot(jitter(child,4) ~ parent, galton)