线性回归中，残差加权方法，增强回归，研究

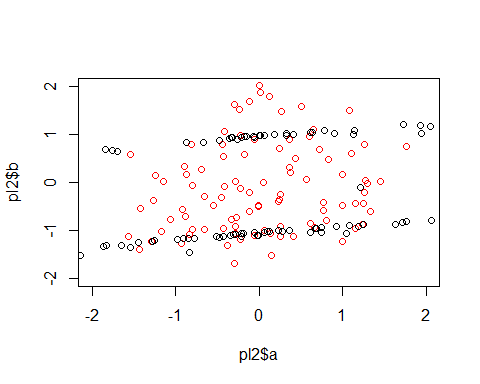
c1p

2016年8月30日

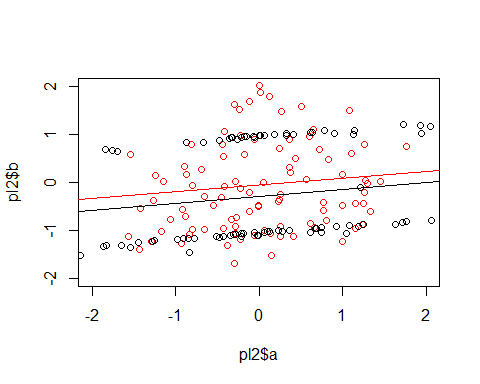
注：内容来狗熊会文章

先随机生成100个数值

set.seed(0)  
rv<-100  
a<-rnorm(rv)  
b<-rnorm(rv)  
fit<-lm(b~a)  
w<-1/(fit$residuals)^2  
a.new<-a\*sqrt(w)  
b.new<-b\*sqrt(w)  
fit2<-lm(b.new~a.new)  
pl<-data.frame(a,b,d=rep(1,length(a)))  
pl1<-data.frame(a=a.new,b=b.new,d=rep(0,length(a.new)))  
pl2<-rbind(pl,pl1)  
plot(pl2$a,pl2$b,col=factor(pl2$d),ylim=c(-2,2),xlim=c(-2,2))

 按权重相乘后数据发生明显变化

plot(pl2$a,pl2$b,col=factor(pl2$d),ylim=c(-2,2),xlim=c(-2,2))  
abline(lm(b~a),col="red")  
abline(lm(b.new~a.new))



summary(fit)

##   
## Call:  
## lm(formula = b ~ a)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.5900 -0.8153 -0.1531 0.6379 2.8379   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.04870 0.09629 -0.506 0.614  
## a 0.13879 0.10960 1.266 0.208  
##   
## Residual standard error: 0.9626 on 98 degrees of freedom  
## Multiple R-squared: 0.0161, Adjusted R-squared: 0.00606   
## F-statistic: 1.604 on 1 and 98 DF, p-value: 0.2084

summary(fit2)

##   
## Call:  
## lm(formula = b.new ~ a.new)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.9053 -0.7875 -0.7477 1.2104 1.2798   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.29403 0.10210 -2.880 0.00489 \*\*   
## a.new 0.14755 0.02288 6.449 4.27e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.021 on 98 degrees of freedom  
## Multiple R-squared: 0.2979, Adjusted R-squared: 0.2908   
## F-statistic: 41.59 on 1 and 98 DF, p-value: 4.272e-09

经过变换后R^2确实有所上升，同时Pv也有所提升。但文章显示这种方法没有依据。 个人理解这种修改数据的方法，像实在利用结果修改数据，经过修改的数据肯定更加趋近于想要的结果，如果这种思路可以应用的话完全可以再找出一个更加好的方式修改数据，结果更加接近想要的值。 总的来说这种方法是否正确还有待考证。