Шпаргалка по R

# Линейная регрессия

* Оценить регрессию, заданную формулой и присвоить объекту reg:

reg <- lm(формула)

* Обобщающая таблица с результатами регрессии:

summary(reg)

* Вектор оцененных коэффициентов регрессии:

coefficients(reg) или coef(reg)

* Векторы расчетных значений и остатков:

fitted(reg), residuals(reg) или resid(reg)

* Сумма квадратов остатков, стандартная ошибка регрессии (корень из несмещенной дисперсии), количество наблюдений в регрессии и количество степеней свободы:

deviance(reg), sigma(reg), nobs(reg), df.residual(reg)

* Фрейм с переменными регрессии:

model.frame(reg)

* Матрица регрессоров, зависимая переменная:

model.matrix(reg), model.response(model.frame(reg))

или использовать дополнительные опции reg <- lm(формула, y=TRUE, x=TRUE)

reg$x, reg$y

* Таблица с доверительными интервалами для коэффициентов регрессии:

confint(reg)

* Информационный критерий Акаике, байесовский информационный критерий (Шварца) и значение логарифмической функции правдоподобия:

AIC(reg), BIC(reg), logLik(reg)

* Линия регерессии для парной регрессии:

abline(reg)

* Прогнозы по регрессии:

predict(reg, newdata = фрейм с данными для прогноза)

* Нормированные и стьюдентизированные остатки:

rstandard(reg), rstudent(reg)

df <- read.table(header=TRUE,text=  
"y x  
 3 2  
 -9 5  
 -8 0  
 -2 0  
 3 -4  
")  
  
n <- nrow(df)  
reg <- lm(y ~ x, data=df)  
summary(reg)

##   
## Call:  
## lm(formula = y ~ x, data = df)  
##   
## Residuals:  
## 1 2 3 4 5   
## 7.000e+00 -2.000e+00 -6.000e+00 -8.882e-16 1.000e+00   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -2.0000 2.5000 -0.8 0.482  
## x -1.0000 0.8333 -1.2 0.316  
##   
## Residual standard error: 5.477 on 3 degrees of freedom  
## Multiple R-squared: 0.3243, Adjusted R-squared: 0.0991   
## F-statistic: 1.44 on 1 and 3 DF, p-value: 0.3163

model.matrix(reg)

## (Intercept) x  
## 1 1 2  
## 2 1 5  
## 3 1 0  
## 4 1 0  
## 5 1 -4  
## attr(,"assign")  
## [1] 0 1

model.response(model.frame(reg))

## 1 2 3 4 5   
## 3 -9 -8 -2 3

coef(reg)

## (Intercept) x   
## -2 -1

sigma(reg)^2

## [1] 30

var(df$y)

## [1] 33.3

nobs(reg)

## [1] 5

df.residual(reg)

## [1] 3

#deviance(lm(y ~ 1, data=df)/(5-1)  
vcov(reg)

## (Intercept) x  
## (Intercept) 6.2500000 -0.4166667  
## x -0.4166667 0.6944444

1-sigma(reg)^2/var(df$y)

## [1] 0.0990991

confint(reg)

## 2.5 % 97.5 %  
## (Intercept) -9.956116 5.956116  
## x -3.652039 1.652039

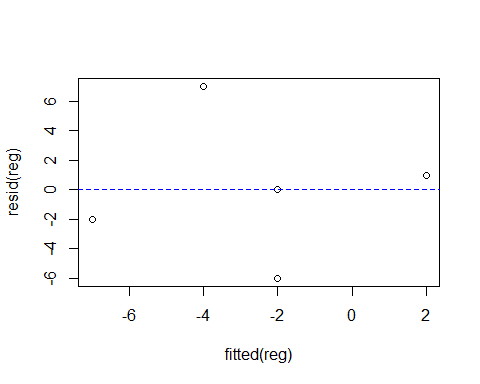
-qt(0.025,3) #-qt(0.05,3)

## [1] 3.182446

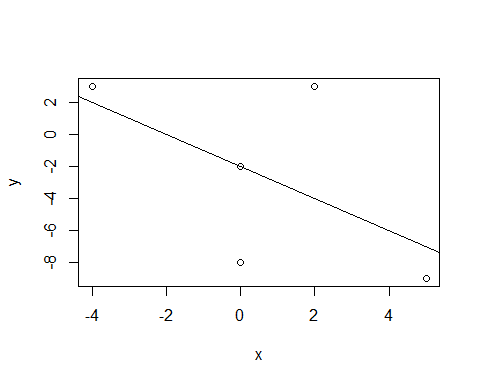
deviance(reg)

## [1] 90

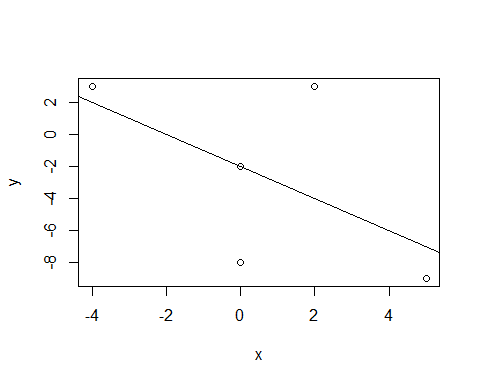
plot(resid(reg) ~ fitted(reg))  
abline(h=0, lty=2, col = 'blue')



plot(y ~ x, data=df)  
abline(reg)



plot(formula(reg), data=df)  
abline(reg)



#anova(reg)  
#summary(aov(reg))  
  
#qqnorm(resid(reg))