

R表达式中常用的符号

表8-2 R表达式中常用的符号

符 号	用 途
~	分隔符号，左边为响应变量，右边为解释变量。例如，要通过x、z和w预测y，代码为y ~ x + z + w
+	分隔预测变量
:	表示预测变量的交互项。例如，要通过x、z及x与z的交互项预测y，代码为y ~ x + z + x:z
*	表示所有可能交互项的简洁方式。代码y~ x * z * w可展开为y ~ x + z + w + x:z + x:w + z:w + x:z:w
^	表示交互项达到某个次数。代码y ~ (x + z + w)^2可展开为y ~ x + z + w + x:z + x:w + z:w
.	表示包含除因变量外的所有变量。例如，若一个数据框包含变量x、y、z和w，代码y~.可展开为y ~ x + z + w
-	减号，表示从等式中移除某个变量。例如，y ~ (x + z + w)^2 - x:w可展开为y ~ x + z + w + x:z + z:w
-1	删除截距项。例如，表达式y ~ x - 1拟合y在x上的回归，并强制直线通过原点
I()	从算术的角度来解释括号中的元素。例如，y~x+(z+w)^2将展开为y~x+z+w+z:w。相反，代码y ~ x + I((z + w)^2)将展开为y ~ x + h，h是一个由z和w的平方和创建的新变量
function	可以在表达式中用的数学函数。例如，log(y) ~ x + z + w表示通过x、z和w来预测log(y)

★~和+

lm()输入数据必须是数据框格式  
summary()查看回归后的详细结果

```
Call:
lm(formula = weight ~ height, data = women)

Residuals:
    Min       1Q   Median       3Q      Max
-1.733 -1.133 -0.383  0.742  3.117

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -87.5167     5.9369  -14.7  1.7e-09 ***
height        3.4500     0.0911   37.9  1.1e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.5 on 13 degrees of freedom
Multiple R-squared:  0.991,    Adjusted R-squared:  0.99
F-statistic: 1.43e+03 on 1 and 13 DF,  p-value: 1.09e-14
```

- 对于拟合出的模型:
- 1. F-statistic是否小于0.05，小于代表模型拟合好
  - 2. R-squared
  - 3. 系数和截距检验是否小于0.05,小于代表系数通过检验

表8-3 对拟合线性模型非常有用的其他函数

函 数	用 途
summary()	展示拟合模型的详细结果
coefficients()	列出拟合模型的模型参数（截距项和斜率）
confint()	提供模型参数的置信区间（默认95%）
fitted()	列出拟合模型的预测值
residuals()	列出拟合模型的残差值
anova()	生成一个拟合模型的方差分析表，或者比较两个或更多拟合模型的方差分析表
vcov()	列出模型参数的协方差矩阵
AIC()	输出赤池信息统计量
plot()	生成评价拟合模型的诊断图
predict()	用拟合模型对新的数据集预测响应变量值

多元线性回归

lm(因变量~自变量1+自变量2, data=)

有交互关系的x1和x2变量可以用x1:x2组合

如果变量很多，考虑相互间交互会有很多模型，可以用逐步回归和全子集回归  
全子集回归，比较慢，把所有变量组合可能都尝试，选取AIC最小的

```
library(leaps)
```

```
leaps<-regsubsets(Murder~Population+Illiteracy+Income+Frost,data=states,nbest=4)
```

```
plot(leaps,scale="adjr2")
```

## 回归诊断

## 非线性

```
nls(formula,data=,start=c(),trace=FALSE是否显示算法执行过程)
```