

# 1. 客户价值分析

# 安装功能包

```
install.packages('PerformanceAnalytics')
```

```
install.packages('DAAG')
```

# 读入数据文件 (数据文件需放在 Rstudio 的工作目录下, 默认为<我的文档>文件夹)

```
a<-read.csv('c1.csv')
```

# 修改数据头名称为英文

```
names(a)<-c('y','x1','x2','x3')
```

# 简单描述统计

```
summary(a)
```

#使用 PerformanceAnalytics 包进行更全面的描述统计和变量相关分析

```
library('PerformanceAnalytics')
```

```
chart.Correlation(a)
```

```
cbind(y=table.Stats(a[[1]]),x1=table.Stats(a[[2]]),x2=table.Stats(a[[3]]),x3=table.Stats(a[[4]]))
```

#简单线性回归模型, 输出显著性检验和拟合优度检验结果以及回归系数的置信区间

```
lm <- lm(y~x1+x2+x3,data=a)
```

```
summary(lm)
```

```
confint(lm)
```

#使用 DAAG 包进行 VIF 检验

```
library('DAAG')
```

```
vif(lm)
```

# 回归诊断: (1)残差图; (2) Q-Q 图; (3)标准化残差方根散点图; (4) Cook 距离图

```
par(mfrow=c(2,2)); plot(lm)
```

## 2. 商品房定价

# 安装功能包

```
install.packages('dplyr')
```

# 读入数据文件 (数据文件需放在 Rstudio 的工作目录下, 默认为<我的文档>文件夹)

```
a<-read.csv('c2.csv')
```

# 修改数据头名称为英文

```
names(a)=c("y","x1","x2","x3","x4")
```

# 对因变量做对数变换

```
a$y=log(a$y)
```

# 使用 dplyr 对因变量做分组描述分析

```
library('dplyr')
```

```
group_by(a,x1) %>%
```

```
summarise(N=n(),mean=mean(y),SD=sd(y),min=min(y),med=median(y),max=max(y))
```

```
group_by(a,x2) %>%
```

```
summarise(N=n(),mean=mean(y),SD=sd(y),min=min(y),med=median(y),max=max(y))
```

```
group_by(a,x3) %>%
```

```
summarise(N=n(),mean=mean(y),SD=sd(y),min=min(y),med=median(y),max=max(y))
```

```
group_by(a,x4) %>%
```

```
summarise(N=n(),mean=mean(y),SD=sd(y),min=min(y),med=median(y),max=max(y))
```

#方差分析, 对全因素模型(x1, x2, x3, x4, x1 与 x2 交互作用)进行全局及各因素的 F 检验

```
lm<-lm(y~x1*x2+x3+x4,data=a)
```

```
anova(lm)
```

```
summary(lm)
```

#方差分析, 对去除不显著因素的模型(x1, x2, x3)进行全局及各因素的 F 检验

```
lm1<-lm(y~x1+x2+x3,data=a)
```

```
anova(lm1)
```

```
summary(lm1)
```

### 3. 上市企业 ST

# 读入数据文件 (数据文件需放在 Rstudio 的工作目录下, 默认为<我的文档>文件夹)

```
a<-read.csv('c3.csv')
```

# 描述分析和盒装图对比

```
summary(a)
```

```
boxplot(ARA~ST,xlab="ST Status",ylab="ARA",data=a)
```

# 逻辑回归

```
glm=glm(ST~ARA+ASSET+ATO+GROWTH+LEV+ROA+SHARE,family=binomial(link=logit),data=a)
```

```
summary(glm)
```

# 根据逻辑回归模型对原数据进行预测

```
pred=predict(glm,a) #获取 logit 预测值
```

```
prob=exp(pred)/(1+exp(pred)) # 计算概率
```

# 以概率阈值\alpha =0.5 预测是否 ST 并计算混淆矩阵

```
y1=1*(prob>0.5)
```

```
table(a$ST,y1)
```

```
confusion(a$ST,y1)
```

# 以概率阈值\alpha =0.0526 预测是否 ST 并计算混淆矩阵

```
y2=1*(prob>0.0526)
```

```
table(a$ST,y2)
```

```
install.packages('DAAG')
```

```
library('DAAG')
```

```
confusion(a$ST,y2)
```