

《数据分析与处理技术》附录1
南京审计大学2015级物流管理专业

R语言可用的相关数据集

管理科学与工程学院 徐宁

版本：2017.9.5

统计数据网站

- <http://data.stats.gov.cn/> （中国国家数据）
- <http://data.un.org/> （联合国统计数据）
- <http://data.worldbank.org/indicator/> （世界银行）
- <http://www.imf.org/en/Data> （IMF）
- <https://data.oecd.org/> （OECD）
- <http://ec.europa.eu/eurostat/data/database> （欧盟数据）
- <http://www.wiod.org/home> （世界投入产出表）

案例数据集

- <http://archive.ics.uci.edu/ml/datasets.html>
- <http://www.icpsr.umich.edu/icpsrweb/ICPSR/>
- <https://usa.ipums.org/>
- <http://flowingdata.com/>

中国调查数据

- http://css.cssn.cn/css_sy/ （中国社会状况综合调查）
- <http://www.cnsda.org/index.php?r=projects/index> （中国国家调查数据库）
- <http://www.dataju.cn/Dataju/web/home>

教学案例数据

- <https://stats.idre.ucla.edu/r/>
- <https://www.stata.com/features/documentation/> (STATA)
- <https://peerj.com/collections/50-practicaldatascists/>

美国数据

- <https://www-fars.nhtsa.dot.gov/Main/index.aspx> (美国高速公路安全管理局数据)
- <http://www.iihs.org/iihs/topics>
- <https://www.eia.gov/>

R package数据……

- <https://github.com/madlogos/recharts>

R语言自带数据集

向量

euro #欧元汇率，长度为11，每个元素都有命名
landmasses #48个陆地的面积，每个都有命名
precip #长度为70的命名向量
rivers #北美141条河流长度
state.abb #美国50个州的双字母缩写
state.area #美国50个州的面积
state.name #美国50个州的全称

因子

state.division #美国50个州的分类，9个类别
state.region #美国50个州的地理分类

矩阵、数组

euro.cross #11种货币的汇率矩阵
freeny.x #每个季度影响收入四个因素的记录
state.x77 #美国50个州的八个指标
USPersonalExpenditure #5个年份在5个消费方向的数据

VADeaths #1940年弗吉尼亚州死亡率（每千人）
volcano #某火山区的地理信息（10米×10米的网格）
WorldPhones #8个区域在7个年份的电话总数
iris3 #3种鸢尾花形态数据
Titanic #泰坦尼克乘员统计
UCBAdmissions #伯克利分校1973年院系、录取和性别的频数
crimtab #3000个男性罪犯左手中指长度和身高关系
HairEyeColor #592人头发颜色、眼睛颜色和性别的频数
occupationalStatus #英国男性父子职业联系
类矩阵
eurodist #欧洲12个城市的距离矩阵，只有下三角部分
Harman23.cor #305个女孩八个形态指标的相关系数矩阵
Harman74.cor #145个儿童24个心理指标的相关系数矩阵

数据框

airquality #纽约1973年5-9月每日空气质量
anscombe #四组x-y数据，虽有相似的统计量，但实际数据差别较大
attenu #多个观测站对加利福尼亚23次地震的观测数据
attitude #30个部门在七个方面的调查结果，调查结果是同一部门35个职员赞成的百分比
beaver1 #一只海狸每10分钟的体温数据，共114条数据
beaver2 #另一只海狸每10分钟的体温数据，共100条数据
BOD #随水质的提高，生化反应对氧的需求 (mg/l) 随时间 (天) 的变化
cars #1920年代汽车速度对刹车距离的影响
chickwts #不同饮食种类对小鸡生长速度的影响
esoph #法国的一个食管癌病例对照研究
faithful #一个间歇泉的爆发时间和持续时间
Formaldehyde #两种方法测定甲醛浓度时分光光度计的读数
Freeny #每季度收入和其他四因素的记录
dating from #配对的病例对照数据，用于条件logistic回归
InsectSprays #使用不同杀虫剂时昆虫数目
iris #3种鸢尾花形态数据
LifeCycleSavings #50个国家的存款率
longley #强共线性的宏观经济数据

morley #光速测量试验数据
mtcars #32辆汽车在11个指标上的数据
OrchardSprays #使用拉丁方设计研究不同喷雾剂对蜜蜂的影响
PlantGrowth #三种处理方式对植物产量的影响
pressure #温度和气压
Puromycin #两种细胞中辅因子浓度对酶促反应的影响
quakes #1000次地震观测数据 (震级>4)
randu #在VMS1.5中使用FORTRAN中的RANDU三个一组生成随机数字，共400组。 #该随机数字有问题。在VMS2.0以上版本已修复。
rock #48块石头的形态数据
sleep #两药物的催眠效果
stackloss #化工厂将氨转为硝酸的数据
swiss #瑞士生育率和社会经济指标
ToothGrowth #VC剂量和摄入方式对豚鼠牙齿的影响
trees #树木形态指标
USArrests #美国50个州的四个犯罪率指标
USJudgeRatings #43名律师的12个评价指标
warpbreaks #织布机异常数据
women #15名女性的身高和体重

列表

state.center #美国50个州中心的经度和纬度

类数据框

ChickWeight #饮食对鸡生长的影响

CO2 #耐寒植物CO2摄取的差异

DNase #若干次试验中，DNase浓度和光密度的关系

Indometh #某药物的药物动力学数据

Loblolly #火炬松的高度、年龄和种源

Orange #桔子树生长数据

Theoph #茶碱药动学数据

时间序列数据

airmiles #美国1937-1960年客运里程营收（实际售出机位乘以飞行哩数）

AirPassengers #Box & Jenkins航空公司1949-1960年每月国际航线乘客数

austres #澳大利亚1971-1994每季度人口数（以千为单位）

Bjsales #有关销售的一个时间序列

Bjsales.lead #前一指标的先行指标（leading indicator）

co2 #1959-1997年每月大气co2浓度（ppm）

discoveries #1860-1959年每年巨大发现或发明的个数

Ideaths #1974-1979年英国每月支气管炎、肺气肿和哮喘的死亡率

fdeaths #前述死亡率的女性部分

mdeaths #前述死亡率的男性部分

freeny.y #每季度收入

JohnsonJohnson #1960-1980年每季度Johnson & Johnson股票的红利

LakeHuron #1875-1972年某一湖泊水位的记录

lh #黄体生成素水平，10分钟测量一次

lynx #1821-1934年加拿大猞猁数据

nhtemp #1912-1971年每年平均温度

Nile #1871-1970尼罗河流量

nottem #1920-1939每月大气温度

presidents #1945-1974年每季度美国总统支持率

UKDriverDeaths #1969-1984年每月英国司机死亡或严重伤害的数目

sunspot.month #1749-1997每月太阳黑子数

sunspot.year #1700-1988每年太阳黑子数

sunspots #1749-1983每月太阳黑子数

treering #归一化的树木年轮数据

UKgas #1960-1986每月英国天然气消耗

USAccDeaths #1973-1978美国每月意外死亡人数

uspop #1790-1970美国每十年一次的人口总数（百万为单位）

WWWusage #每分钟网络连接数

Seatbelts #多变量时间序列。和UKDriverDeaths时间段相同，反映更多因素。

EuStockMarkets #多变量时间序列。欧洲股市四个主要指标的每个工作日记录，共1860条记录。