

## 成绩评定

- ❖ 平时（作业、考勤、课堂表现） 15%
- ❖ 章节测试（课程报告） 10%
- ❖ 半期考试 25%
- ❖ 期末考试 50%

— 刘 颖 —

## 问题

房价数据：加拿大Windsor 和Essex在1987年7—9月售出的546套住房的观测值，一共12个变量，如下：

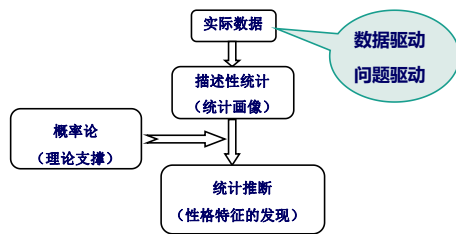
变量名	说明	类型	变量名	说明	类型
sell	房价	数量	rec	是否有娱乐室	分类
lot	面积	数量	ffin	是否有地下室	分类
bdms	卧室数目	数量	ghw	是否用煤气热水	分类
fb	卫生间数目	数量	ca	是否有中央空调	分类
sty	层数	数量	gar	停车位数目	数量
drv	是否有车道	分类	reg	社区是否好	分类

从数据中发现关系 —— 统计学的奥秘！！

— 刘 颖 —

SWJTU

## 对数据的认识



思考：对股票数据的分析（股民，财经评论员，政府智囊团）

— 刘 颖 —

SWJTU

## 第一章

## 描述性统计



## 中心位置的描述

总体（Population）：研究对象的全体元素的集合

个体（individual;object）：组成总体的每个元素

样本（Sample）



— 刘 颖 —

SWJTU

❖ 均值（Mean）

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

❖ 中位数（Median）

$$M_e = \begin{cases} x_{(\frac{n+1}{2})} & n \text{ 为奇数} \\ \frac{1}{2} (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & n \text{ 为偶数} \end{cases}$$

❖ 四分位数

$$Q_L \quad Q_M \quad Q_U$$

— 刘 颖 —

SWJTU

## 离散程度的描述

❖ 极差 (Range)  $R = \max - \min$

❖ 四分位数间距  $f_s = Q_U - Q_L$

❖ 样本方差和样本标准差

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{S_{xx}}{n-1} \quad S = \sqrt{S^2}$$



$S_{xx} = \dots\dots$

— 刘 航 —

SWJTU

## 图表工具

❖ 茎叶图 (Stem-and-Leaf Display)

❖ 直方图 (Histogram)

❖ 散点图 (Scatter plot)

❖ 折线图 (Line Chart)

❖ 箱线图 (Boxplot)



— 刘 航 —

SWJTU

## 茎 叶 图

The decimal point is 1 digit(s) to the right of the |

4 | 1 7 9

5 | 0 4 7

6 | 0 2 2 2 3 4 4 6 7

7 | 1 2 2 3 3 4 6

8 | 0 1 1 5 5

9 | 0 4 5

茎: 十位

叶: 个位

— 刘 航 —

SWJTU

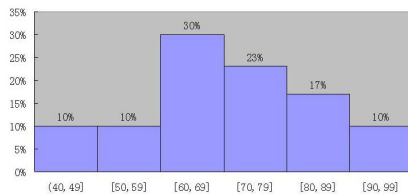
## 茎 叶 图

乙班	7 5 3	4	1 7 9	甲班
	9 7 4 4 0 5	0 4 7		
	7 6 4 3 2 6	0 2 2 2 3 4 4 6 7		
8 8 6 4	4 3 3 2 1 7	1 2 2 3 3 4 6		
	7 6 6 2 1 8	0 1 1 5 5		
	4 3 0	9	0 4 5	

— 刘 航 —

SWJTU

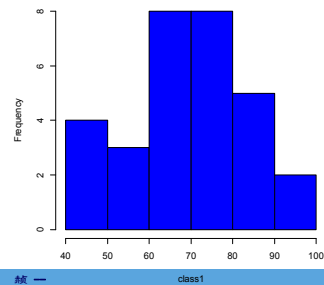
## 直 方 图 (Excel)



— 刘 航 —

SWJTU

## 直 方 图 (R)

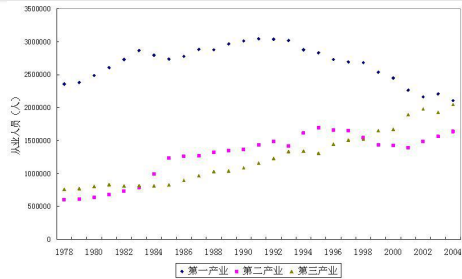


— 刘 航 —

class1

SWJTU

## 散点图

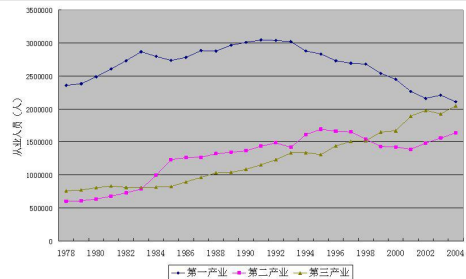


成都市1978—2004年分产业从业人员的总人数散点图

— 刘 航 —

SWJTU

## 折线图

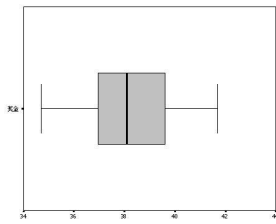


成都市1978—2004年分产业从业人数的变化趋势折线图

— 刘 航 —

SWJTU

## 箱线图 (Boxplot)

最小值  $x_{(1)}$  下四分位数  $Q_L$  中位数  $M_e$  上四分位数  $Q_U$  最大值  $x_{(n)}$ 

❖ 四分位数间距

$$f_s = Q_U - Q_L$$

❖ 离群值 (Outlier)

$$1.5f_s < d < 3f_s$$

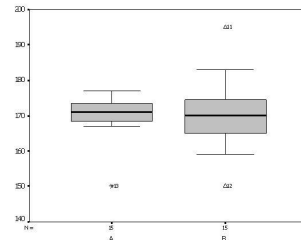
❖ 极值 (Extreme Value)

$$> 3f_s$$

— 刘 航 —

SWJTU

A	171	167	168	173	172	170	169	173	169	168	174	177	150	175	174
B	165	174	162	180	159	183	165	172	175	165	195	150	170	165	170



— 刘 航 —

SWJTU

```

RGui (64-bit)
文件 编辑 查看 其他 程序包 窗口 帮助
R Console
R是自由软件，不承担任何担保。
在某些条件下你可以再发自由散布。
用 'license()' 或 'licence()' 来查看散布的详细条件。

R是个合作计划，有许多人为了它做出了贡献。
用 'contributors()' 来查看合作的详细情况。
用 'citation()' 会告诉你如何在出版物中正确地引用R或R程序包。

用 'demo()' 来运行一些示范程序。用 'help()' 来阅读在线帮助文件，或
用 'help.start()' 通过浏览器来查看帮助文件。
用 'q()' 退出R。

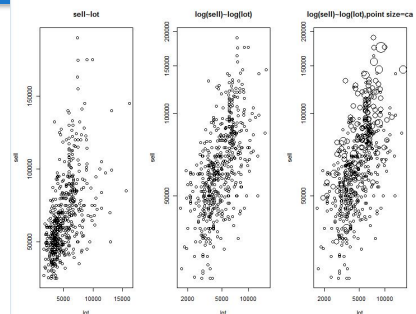
[原来保存的工作空间已还原]

> w<-read.csv("w.csv")#读入数据
> par(mfrow=c(1,3))
> plot(sell~lot,w_main="sell~lot")#原始变量散点图
> plot(log(sell~lot,w_log="xy",main="log(sell)~log(lot)")#对数变换
> plot(sell~lot,w_log="xy",cex=0.5)
> title("log(sell)~log(lot),point size=cex")#气泡图
> par(mfrow=c(1,1))
> |

```

— 刘 航 —

SWJTU



— 刘 航 —

SWJTU

```

> par(mfrow=c(1,3))
> plot(sell~lot, log="xy", pch=ca+1, w)
> legend("topleft", c("ca=0", "ca=1"), pch=1:2)
> title("log(sell)~log(lot), point shape=ca")
> boxplot(sell~fb, w, main="sell~fb")
> boxplot(sell~ca, w, main="sell~ca")
> par(mfrow=c(1,1))

```

— 刘 航 —

SWJTU

