第一章
商务分析简介

商务分析

商务分析是对以下要素的应用：

- 数据
- 信息技术
- 统计分析
- 量化方法
- 计算机模型

→以帮助经理人对商业运作有更好的洞见，从而做出更好的基于事实的决策。

商务分析

- 数据+商务分析→更好的业务洞见和决策
  - 微策略公司(MicroStrategy)的一项研究显示，全球企业正在利用数据:
    - 提高效率和生产力(64%)
    - 实现更有效的决策(56%)
    - 推动更好的财务业绩(51%)

商务分析思维

- 数据思维
- 业务思维（商业思维）
  - “一切业务数据化，一切数据业务化”—阿里巴巴

商务分析范围

- **描述性分析(Descriptive analytics)**：利用数据理解过去和现在企业的表现，以辅助决策者做出理智的决策
- **预测性分析(Predictive analytics)**：对历史数据进行分析，发现数据中存在的模式和关系，将这种模式和关系投射到未来，从而对未来进行预测
- **决策性分析(或规范性分析 Prescriptive analytics)**：制定目标函数（最大化利润、最小化成本等），基于此制定最优决策

例 1.1：零售降价决策

- 大部分零售店在季末都进行降价清库存活动
- 核心问题：何时开始降价，降价幅度多高可以最大化收益？
- 商务分析的潜在应用：
  - 描述性分析：检查类似产品（价格、销售量、广告）的历史数据
  - 预测性分析：预测价格变动一定量的时候的销量变化
  - 决策性分析：找到最优的降价幅度和广告策略以最大化销售收入或利润

用商务分析解决问题步骤

1. 认识到问题
2. 定义问题
3. 结构化问题
4. 分析问题
5. 对结果进行解读并做决策
6. 实施决策

认识到问题

- 当目前发生的状况与预期之间有差距时，有“问题”存在
  - 例：与竞争对手比成本太高

定义问题

- 过高的成本是因为选址不合理，或车辆调度问题
- 定义问题中可能遇到的困难：
  - 大量可能的原因
  - 产生问题的人和希望解决问题的人不是同一个
  - 目标存在互斥性（成本、质量、速度）
  - 时间限制

问题结构化

- 明确目标（最小化运输成本）
  - 提出可能的决策（选址地点）
  - 明确各限制条件（配送时间限制、资金限制等）

分析问题

- 找到结构化问题的解
  - 实验工具
  - 优化工具
  - 风险分析

解读结果、做决策

- 模型是对现实的抽象，忽略了现实中很多细节
- 管理人员必须了解模型的限制、假设，在做最终决策时考虑这些因素

实施决策

- 把模型的结果在现实中进行实施
  - 整合资源
  - 激励员工
  - 消除改变的阻力
  - 改变公司政策
  - 建立部门间的信任

第二章：数据

商务分析中的数据

- **数据（Data）**：通过一些测量过程获得的数字（或文字、图表）结果。
- **信息（Information）**：分析数据的结果，即从数据中抽取的可用于支持评估和决策制定的有意义的部分。

数据集

- **数据集**- 数据的集合.
  - 例:营销调研结果，历史股票价格，生产过程中抽取样本的尺寸
  - 一个数据集文件一般是一个二维表格，其中每一行代表一条记录或一个个体，每一列代表一个属性(fields, or attributes)或变量（variable）

例1.3：一个销售交易数据集

观测值

	A	B	C	D	E	F	G	H
1	Sales Transactions: July 14							
2								
3	Cust ID	Region	Payment	Transaction Code	Source	Amount	Product	Time Of Day
4	10001	East	Paypal	93816545	Web	$20.19	DVD	22:19
5	10002	West	Credit	74083490	Web	$17.85	DVD	13:27
6	10003	North	Credit	64942368	Web	$23.98	DVD	14:27
7	10004	West	Paypal	70560957	Email	$23.51	Book	15:38
8	10005	South	Credit	35208817	Web	$15.33	Book	15:21
9	10006	West	Paypal	20978903	Email	$17.30	DVD	13:11
10	10007	East	Credit	80103311	Web	$177.72	Book	21:59
11	10008	West	Credit	14132683	Web	$21.76	Book	4:04
12	10009	West	Paypal	40128225	Web	$15.92	DVD	19:35
13	10010	South	Paypal	49073721	Web	$23.39	DVD	13:26

记录、个体

属性、变量

大数据

- 大数据：传统数据处理应用软件不足以处理的大或复杂的数据集
  - 数据量大（Volume）--记录和变量特别多的数据集
  - 数据种类多(Variety)
  - 数据价值密度低（Value）
  - 数据产生和处理速度快（Velocity）
- 初始的分析基于小数据

数据测量尺度

- 每个变量的观测值需要用一些测量尺度来度量。
- 不同的测量尺度决定了数据中的信息量是不同的，并且需要用不同的方法去分析这些数据。

数据测量尺度

- **名义尺度** –数据只展示类别信息
- **顺序尺度** – 数据展示了顺序等级
- **间隔尺度**– 数值间的距离按某一固定度量单位显示，可比较（最常见的类型）
- **比率尺度** – 距离可比较，此外还有绝对零点的定距数据，数值之间的比率也有意义

例 1.4:数据分类

应付账款时长

	A	B	C	D	E	F	G	H	I	J
1	**Purchase Orders**									
2										
3	**Supplier**	**Order No.**	**Item No.**	**Item Description**	**Item Cost**	**Quantity**	**Cost per order**	**A/P Terms (Months)**	**Order Date**	**Arrival Date**
4	Hulkey Fasteners	Aug11001	1122	Airframe fasteners	$ 4.25	19,500	$ 82,875.00	30	08/05/11	08/13/11
5	Alum Sheeting	Aug11002	1243	Airframe fasteners	$ 4.25	10,000	$ 42,500.00	30	08/08/11	08/14/11
6	Fast-Tie Aerospace	Aug11003	5462	Shielded Cable/ft.	$ 1.05	23,000	$ 24,150.00	30	08/10/11	08/15/11
7	Fast-Tie Aerospace	Aug11004	5462	Shielded Cable/ft.	$ 1.05	21,500	$ 22,575.00	30	08/15/11	08/22/11
8	Steelpin Inc.	Aug11005	5319	Shielded Cable/ft.	$ 1.10	17,500	$ 19,250.00	30	08/20/11	08/31/11
9	Fast-Tie Aerospace	Aug11006	5462	Shielded Cable/ft.	$ 1.05	22,500	$ 23,625.00	30	08/20/11	08/26/11
10	Steelpin Inc.	Aug11007	4312	Bolt-nut package	$ 3.75	4,250	$ 15,937.50	30	08/25/11	09/01/11
11	Durrable Products	Aug11008	7258	Pressure Gauge	$ 90.00	100	$ 9,000.00	45	08/25/11	08/28/11
12	Fast-Tie Aerospace	Aug11009	6321	O-Ring	$ 2.45	1,300	$ 3,185.00	30	08/25/11	09/04/11
13	Fast-Tie Aerospace	Aug11010	5462	Shielded Cable/ft.	$ 1.05	22,500	$ 23,625.00	30	08/25/11	09/02/11
14	Steelpin Inc.	Aug11011	5319	Shielded Cable/ft.	$ 1.10	18,100	$ 19,910.00	30	08/25/11	09/05/11
15	Hulkey Fasteners	Aug11012	3166	Electrical Connector	$ 1.25	5,600	$ 7,000.00	30	08/25/11	08/29/11

名义　顺序　名义　名义　比率　比率　比率　比率　间隔　间隔

数据来源

- 一手数据：自己调查（访谈、询问、问卷等方式）得来的数据内容。
- 二手数据：从各有关方面（例如国家统计局等）间接得到的数据内容。

一手数据

- 探索性数据收集：
  - 目的：形成最初的预见和洞查，例如销量为什么下降了。
  - 方法：焦点小组，深度访谈

一手数据

- 描述性数据收集：
  - 目的：产生相关顾客群的特征的数据，例如顾客为我们的产品花了多少钱，为竞争对手的产品花了多少钱，买我们品牌产品的顾客有什么特征。
  - 方法：问卷调研

问卷调研

- 广泛使用在几乎所有500强公司中
- 用于收集顾客的态度、满意度、购买习惯
- 数据有助于对顾客进行分类、制定营销策略

测量量表

- 量表：测量工具
  - 定类量表（收集的数据类型为名义数据）
  - 定序量表（收集的数据类型为顺序数据）
  - 定距量表（收集的数据类型为间隔数据）
  - 定比量表（收集的数据类型为比率数据）

测量量表

- 定类
  - 下面的饮料中，你喜欢哪一个？（选出所有符合条件的）：

1. 可乐__
2. 雪碧__
3. 气泡水__
4. 果汁__

测量量表

- 定序
  - 根据你的喜好程度对这些饮料排序（最喜欢的饮料=1，最不喜欢的饮料=4）：

可乐__

雪碧__

气泡水__

果汁__

测量量表

- 定距
  - 请你在量表中合适的位置标出你对每种饮料的喜好程度

	非常讨厌	讨厌	喜欢	非常喜欢
可乐				
雪碧				
气泡水				
果汁				

测量量表

- 定距
  - 请移动鼠标至合适的位置，表明你对可乐的喜爱程度

测量量表

- 定比
  - 请将100分分配给下面的饮料，用来表示你对它们的喜好程度：

可乐__

雪碧__

气泡水__

果汁__

测量量表

量表	基本比较	例子	平均测量
定类	同一性	男-女、使用-不使用	众数
定序	有序性	品牌偏好、质量等级	中位数
定距	定距比较	对品牌态度	均值
定比	绝对数量比较	单位销售量、购买数量	几何平均数

问卷调研优缺点

- 问卷调研的优点：低成本 、易实施，是了解客户的很好的方式
- 问卷调研的缺点：难以获得无偏的回答；如何选择回答问卷的合适人群（可结合焦点小组）
- 参考书：
  - *《营销调研方法论基础》，北京大学出版社*
  - *《市场营销研究：应用导向》，电子工业出版社*

一手数据

- 因果性数据收集：
  - 目的：验证因果关系
    - 例如改变登录界面是否有助于更多的客户注册登录？
    - 把儿童食品包装盒设计得矮一些不易被小孩碰翻，是否能增进家长对此产品的态度评价？
  - 方法：试验
    - AB测试

因果性数据收集

- AB测试不可行时，进行观察性因果研究
  - 实验单元数很少
  - 所测试的项目不合伦理（科学的进展超前于人类的伦理规范—查理.卓别林）
  - 。。。。。。
- 因果关系推断
  - 相关性
  - 时间顺序
  - 没有共同的驱动因素

一手数据—总结

- 探索性数据收集
  - 研究问题偏探索性（ambiguous problems）
    - 产品的销量降低了，为什么？
- 描述性数据收集
  - 研究问题偏描述性（aware of problem）
    - 什么样的顾客在买我们的产品？哪些人在买竞争对手的产品？
- 因果性数据收集
  - 研究问题的因果性较强（problem clearly defined）
    - 如果我改变登录页面的设置，是否会有更多的人购买产品？

二手数据

- **数据库** – 一系列相关数据集
  - 例：一家零售店有客户信息的数据集、每件产品销售量的数据集等
- 常见的数据库
  - 企业内部数据库：产品销售情况、运营指标、人力资源考核、网络点击数据（访客的地点、访问时间、停留长度、访问路径、搜索产品、浏览产品、点开广告、最终购买、阅读点评等）
  - 盈利机构数据库：淘系数据等

https://tianchi.aliyun.com/?spm=5176.12282016.J_3941670930.9.34946d92cQGW7I

https://www.daas-auto.com/supermarket_data_De/727.html

https://www.datatang.com/

  - 政府机构数据库：经济统计年鉴

https://sufe.libguides.com/az.php?t=23126

一手数据 OR 二手数据

- 一手数据获得成本较高
- 根据研究目的，如有二手数据可以使用时，优先使用二手数据

数据可视化

- **数据可视化**– 用有意义的方式展示数据以增强商业洞见的过程
  - 数据可视化工具为经理人提供了简易掌握的分析能力，以减少对专业IT人员的依赖。因为信息变得更直观，在组织中增加了信息共享与合作

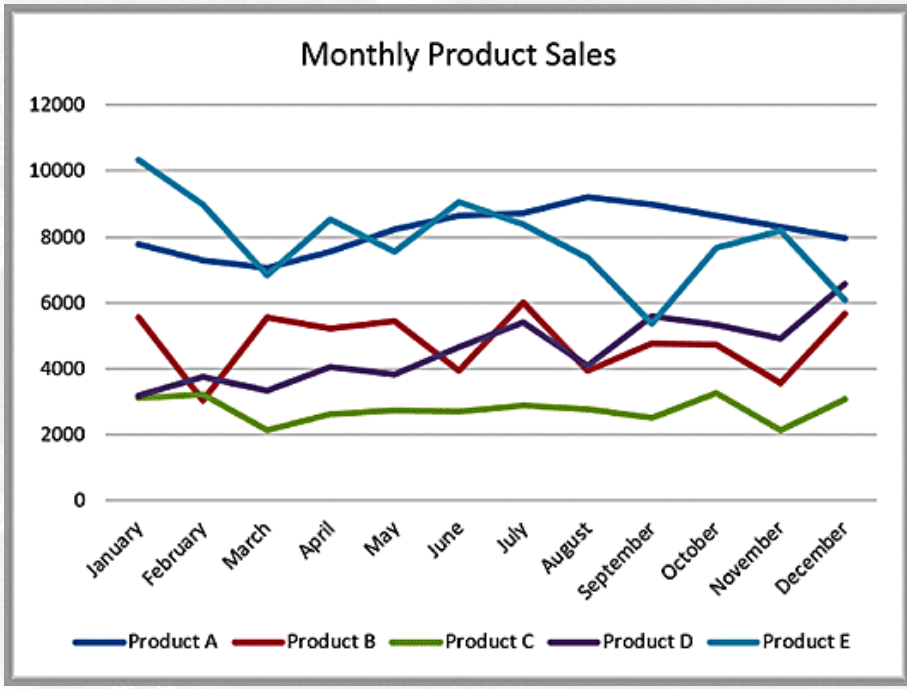
例 3.1：表格数据与可视化数据

- 表格数据提供了精确的数值信息，例如各产品每月销量，可以对此进行比较或计算
  - 例如，通过表格中数据的计算，发现A产品在二月份销量降了6.7%（1 - B3/B2）。但表格数据难以提供更宏观的结论（big picture）。

	A	B	C	D	E	F
1	Month	Product A	Product B	Product C	Product D	Product E
2	January	7792	5554	3105	3168	10350
3	February	7268	3024	3228	3751	8965
4	March	7049	5543	2147	3319	6827
5	April	7560	5232	2636	4057	8544
6	May	8233	5450	2726	3837	7535
7	June	8629	3943	2705	4664	9070
8	July	8702	5991	2891	5418	8389
9	August	9215	3920	2782	4085	7367
10	September	8986	4753	2524	5575	5377
11	October	8654	4746	3258	5333	7645
12	November	8315	3566	2144	4924	8173
13	December	7978	5670	3071	6563	6088

例3.1：表格数据与可视化数据

- 可视化数据提供了更宏观的直观信息：
  - 比较四种产品的总体销量（产品C销量最低）
  - 识别趋势性（产品D销量在增长，产品C销量较稳定，产品B销量波动较大，产品E销量在九月份有大幅下降）

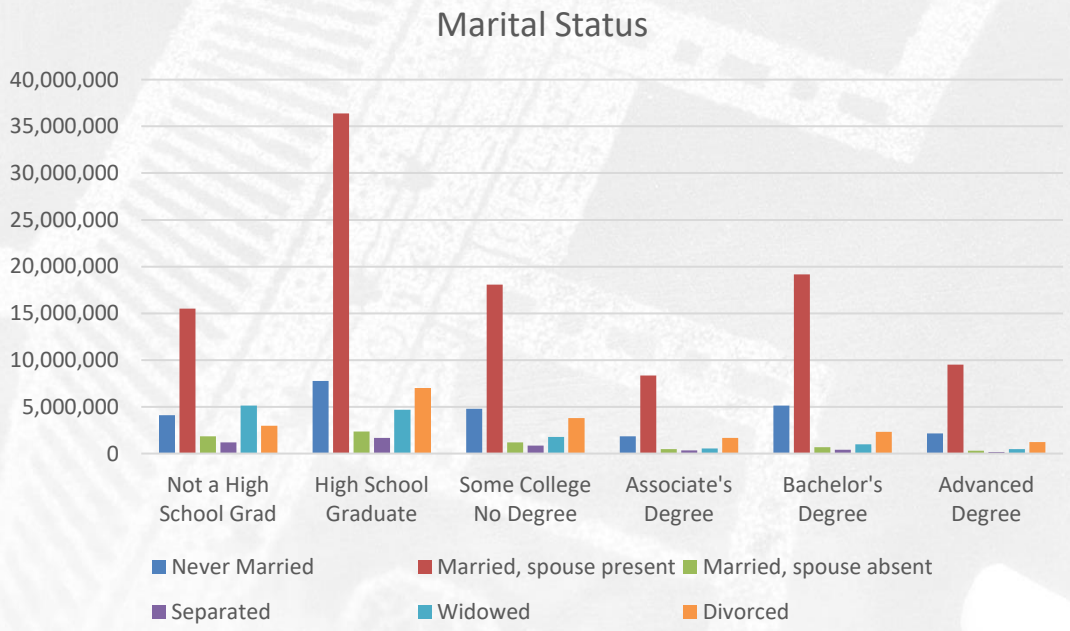


分类型数据（名义数据）

- 柱状图和条形图：比较不同类别的数量
- 柱状图：纵向；条形图：横向
  - 簇状柱形图：比较不同类型的数量
  - 堆积柱形图：比较不同类型的数量并查看其对总和的贡献
  - 百分比堆积柱形图：比较不同类型的数量并显示其在总体中的百分比

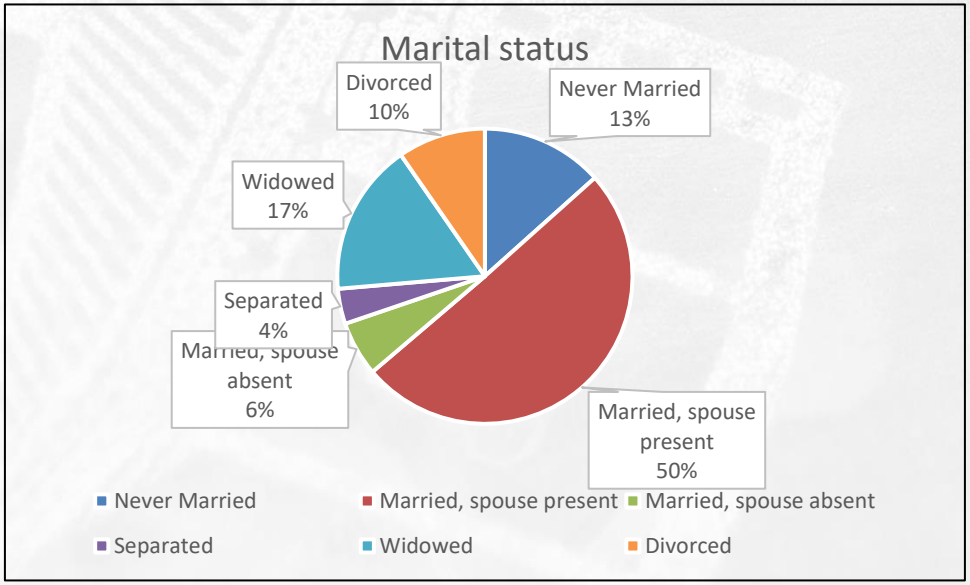
例3.2：绘制柱状图

例3.3 绘制饼图：“Census Education Data”记录了人口普查数据中的教育情况。查看不同学历中各婚姻状况分布情况。选中A19:C24。



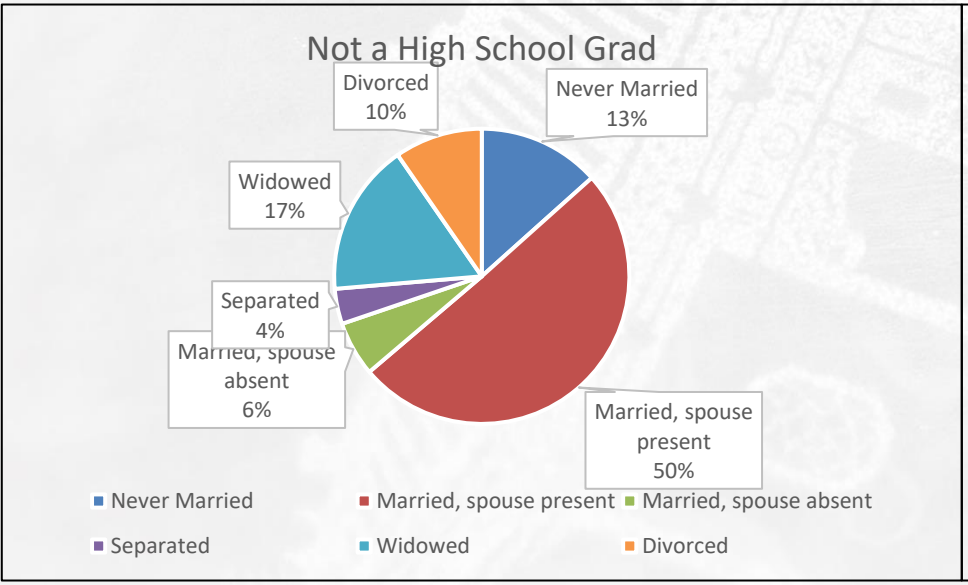
分类型数据

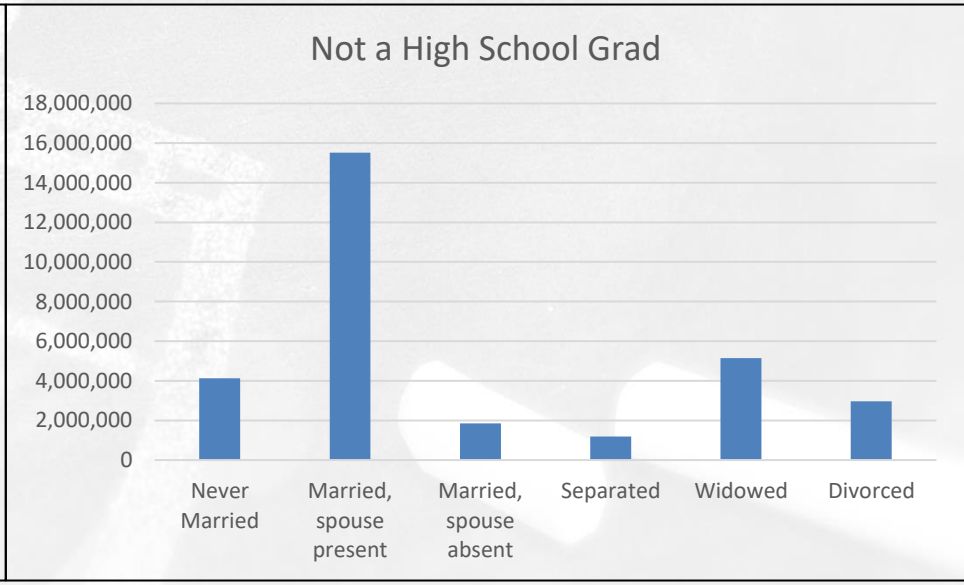
- 饼图
- 饼图展示了各类别所占百分比。
  - 例3.3 绘制饼图：”Census Education Data”记录了人口普查数据中的教育情况。查看高中以下学历中各婚姻状况分布情况。选中A19:B24。



饼图与柱状图

- 柱状图中更易显示出各类别的多少
- 饼图对比例的显示更直观
- 饼图不适用于类别太多的情况





上海财经大学 商學院 COLLEGE OF BUSINESS SHANGHAI UNIVERSITY OF FINANCE AND ECONOMICS

汇总数据OR原始数据

- “Census Education Data”展示的是汇总数据
- “Purchase Orders”数据集展示的是原始数据

	A	B
100	Item Description	Frequency
101	Airframe fasteners	14
102	Bolt-nut package	11
103	Control Panel	4
104	Door Decal	2
105	Electrical Connector	8
106	Gasket	10
107	Hatch Decal	2
108	Machined Valve	4
109	O-Ring	12
110	Panel Decal	1
111	Pressure Gauge	7
112	Shielded Cable/ft.	11
113	Side Panel	8

频数

数据透视表

- 快速生成感兴趣的变量的汇总信息
  - “Purchase orders”中，“插入”-“数据透视表”，拖动感兴趣的变量至行、列中。“值”为需要汇总的量，可以是计数、加总、求平均值等。

行标签	计数项:Supplier
Airframe fasteners	14
Bolt-nut package	11
Control Panel	4
Door Decal	2
Electrical Connector	8
Gasket	10
Hatch Decal	2
Machined Valve	4
O-Ring	12
Panel Decal	1
Pressure Gauge	7
Shielded Cable/ft.	11
Side Panel	8
总计	94

数据透视图

- 数据透视图将数据透视表的结果可视化
- 在“数据透视表分析”下选择数据透视图
- 行标签处可以进行筛选

频率分布

- 频率是各类别数目占总数的比例
  - 如果一组数据有n个观测值，第i类的频率为

$$\frac{i\text{类的频数}}{n}$$

  - 频率一般用百分比表示
  - 频率分布用表格展示了各类别的频率
  - 值—值汇总方式—值显示方式

	A	B	C
100	Item Description	Frequency	Relative Frequency
101	Airframe fasteners	14	0.1489
102	Bolt-nut package	11	0.1170
103	Control Panel	4	0.0426
104	Door Decal	2	0.0213
105	Electrical Connector	8	0.0851
106	Gasket	10	0.1064
107	Hatch Decal	2	0.0213
108	Machined Valve	4	0.0426
109	O-Ring	12	0.1277
110	Panel Decal	1	0.0106
111	Pressure Gauge	7	0.0745
112	Shielded Cable/ft.	11	0.1170
113	Side Panel	8	0.0851
114	Total	94	1.0000

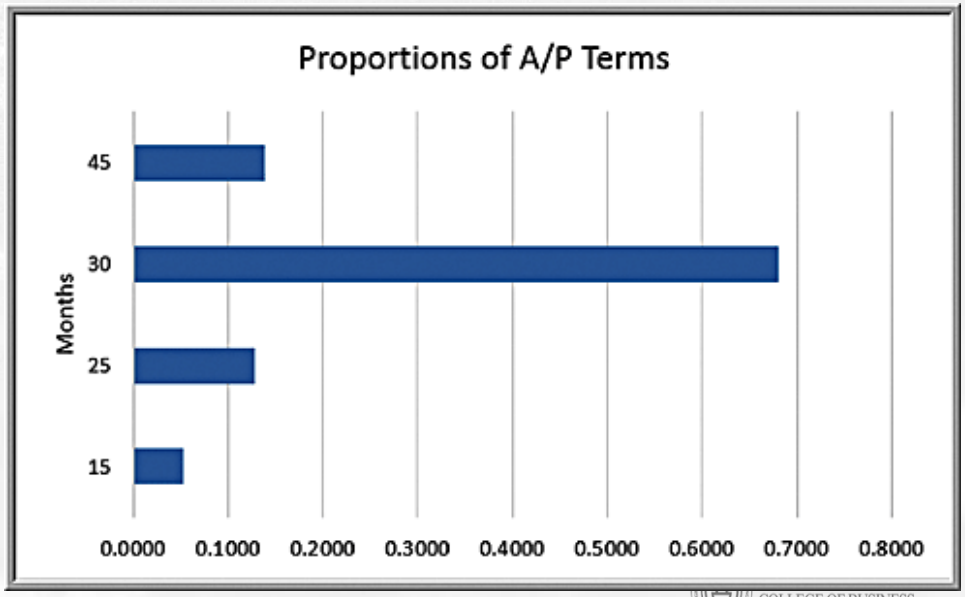
数量型数据（比率数据）

- 直方图：用柱状图展示的数量型数据频数分布图
  - 数据-数据分析-直方图
  - （文件-选项-加载项-分析工具库-转到-分析工具库-确定）
  - 定义“接收区域”：例如A/P terms 接收区域为15，25，30，45

数量型数据的频数分布

- 只有少量几个数值，与名义数据频数分布一样
  - 例3.4， A/P terms取值只有15，25，30，45

	A	B	C
117	A/P Terms	Frequency	Relative Frequency
118	15	5	0.0532
119	25	12	0.1277
120	30	64	0.6809
121	45	13	0.1383
122	Total	94	1.0000



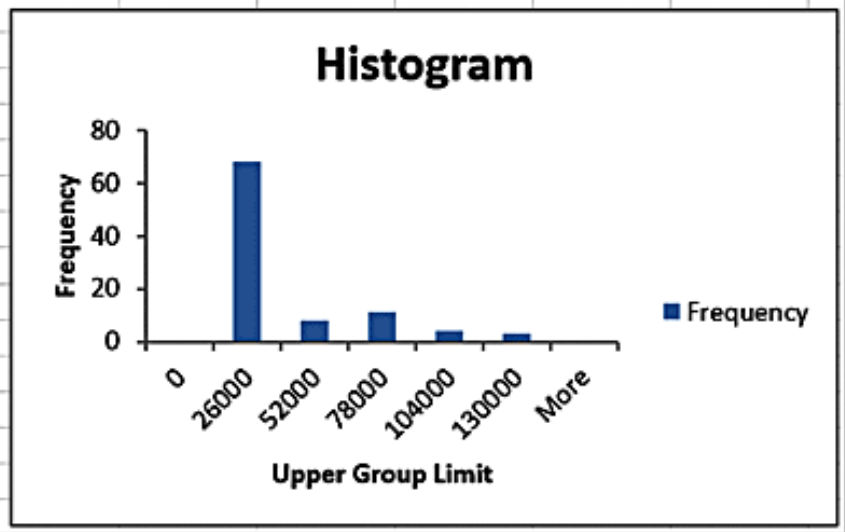
直方图

- 数据取值较多时如何定义接收区域？
  - 先确定分为几组（5-15组较合适）
  - 根据组数确定每组宽度
    - 每组宽度=（UL-LL）/组数
      - UL：大于等于数据中最大值的某一整数
      - LL:小于等于数据最小值的某一整数
  - 确定每组的上限和下限

例3.5 绘制Cost per Order直方图

- 数据范围为68.75 至127,500（排序可得）；
- LL设为0 ， UL设为130,000
- 选择5组，则每组宽度为 (130,000 – 0) / 5 = 26,000
- 接收区域为：0， 26000，52000，78000，104000，13000
  - 注意，以第二组为例，统计的是cost per order大于0小于等于26000的订数数量

	A	B
1	*Upper Group Limit*	*Frequency*
2	0	0
3	26000	68
4	52000	8
5	78000	11
6	104000	4
7	130000	3
8	More	0
9		
10		
11		
12		
13		
14		
15		

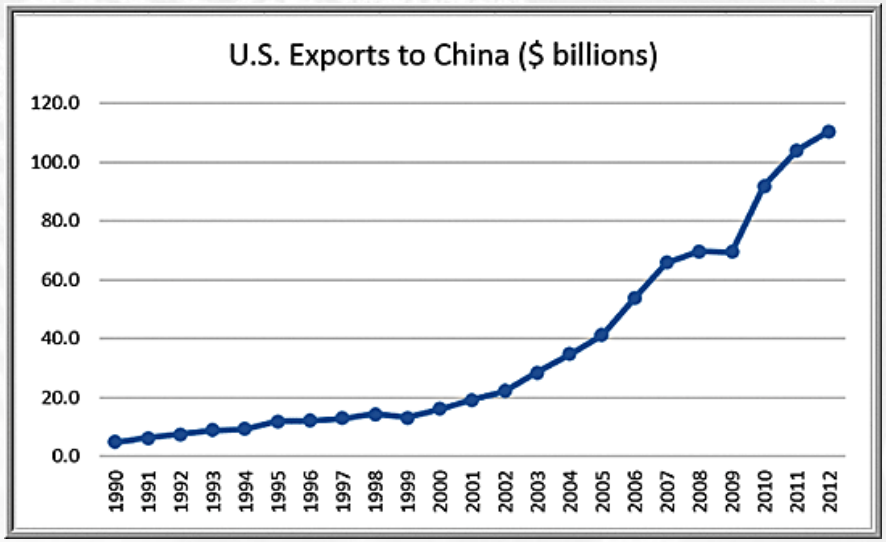


帕累托分析

- 意大利经济学家Vilfredo Pareto在1906年发现意大利大部分财富由很少数的人拥有。
- 对“Bicycle Inventory”数据集中每个产品的价值做帕累托分析

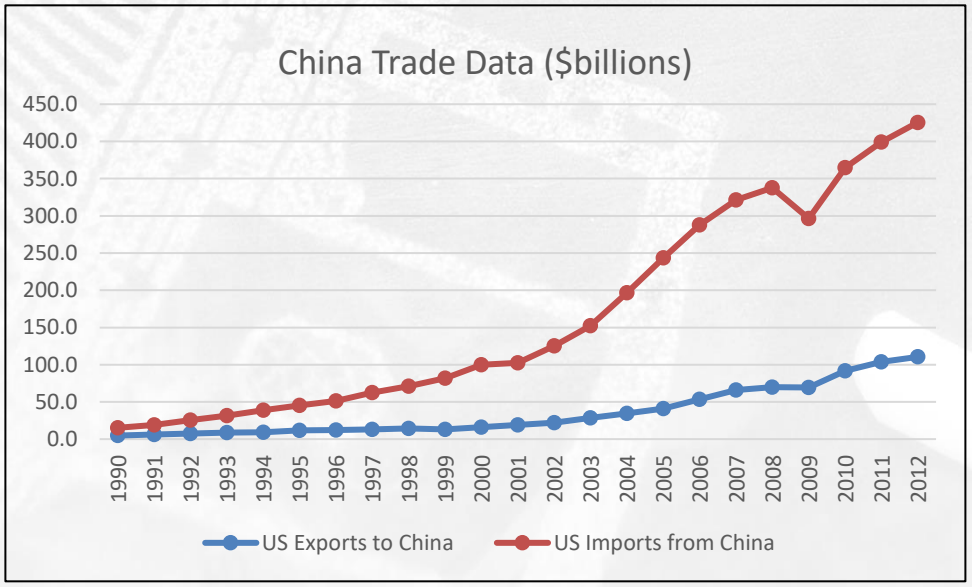
变量相关性可视化

- 折线图展示了数据的趋势
  - 注意：可以在一张折线图中展现多个数据系列，但当数量级差别较大时，显示效果较差，影响对数据的解读。因此，数量级差异大时，建议用多张图展现多个数据系列。



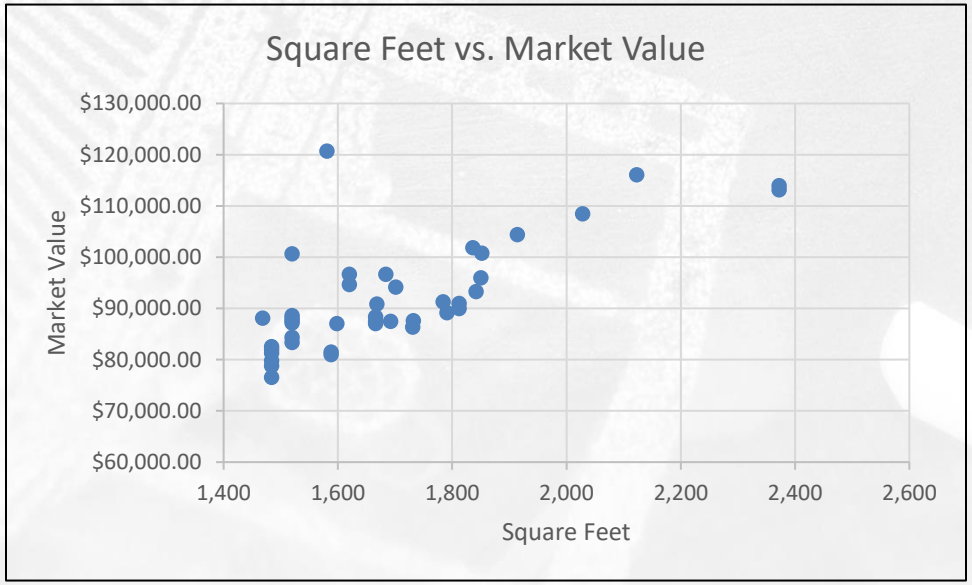
例3.6：绘制折线图

- “China Trade Data”展示了中美之间历年进出口数据。选择数据B3:C26.
- 面积图



散点图

- 散点图直观地展现了两个变量之间的关系，是分析两个变量关系的初步探索。
  - 例3.7：”Home Market Value”记录了房龄、面积、价值信息。研究面积与价值的关系。



交叉分组表

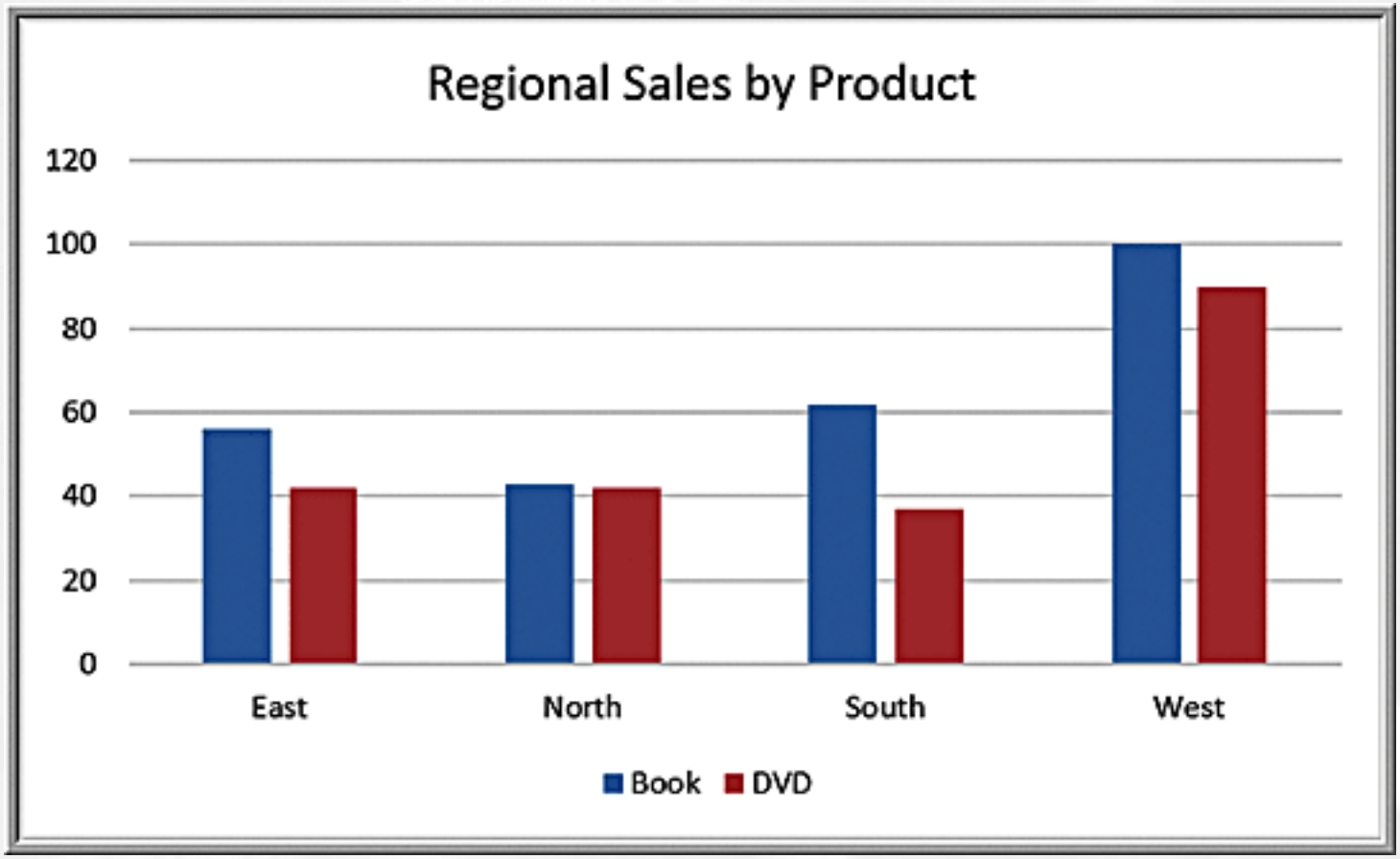
- 交叉分组表：分析两个类别变量之间关系的有效工具。例如在“Sales Transaction”数据集中，探索区域与购买产品的关系。

	A	B	C	D	E	F	G	H
1	Sales Transactions: July 14							
2								
3	Cust ID	Region	Payment	Transaction Code	Source	Amount	Product	Time Of Day
4	10001	East	Paypal	93816545	Web	$20.19	DVD	22:19
5	10002	West	Credit	74083490	Web	$17.85	DVD	13:27
6	10003	North	Credit	64942368	Web	$23.98	DVD	14:27
7	10004	West	Paypal	70560957	Email	$23.51	Book	15:38
8	10005	South	Credit	35208817	Web	$15.33	Book	15:21
9	10006	West	Paypal	20978903	Email	$17.30	DVD	13:11
10	10007	East	Credit	80103311	Web	$177.72	Book	21:59
11	10008	West	Credit	14132683	Web	$21.76	Book	4:04
12	10009	West	Paypal	40128225	Web	$15.92	DVD	19:35
13	10010	South	Paypal	49073721	Web	$23.39	DVD	13:26

Region	Book	DVD	Total
East	56	42	98
North	43	42	85
South	62	37	99
West	100	90	190
Total	261	211	472

Region	Book	DVD	Total
East	57.1%	42.9%	100.0%
North	50.6%	49.4%	100.0%
South	62.6%	37.4%	100.0%
West	52.6%	47.4%	100.0%

交叉分组表—柱状图


Regional Sales by Product
120
100
80
60
40
20
0
East
North
South
West
Book
DVD

数据透视表

- 快速制作交叉分组表并提供相关分析
  - 例3.8：“Sales Transactions”中，“插入”-“数据透视表”，拖动感兴趣的变量至行、列中。“值”为需要汇总的量，可以是计数、加总、求平均值等。

计数项:Cust ID	列标签		
行标签	Book	DVD	总计
East	56	42	98
North	43	42	85
South	62	37	99
West	100	90	190
总计	261	211	472

辛普森悖论

- 根据汇总交叉数据表和未汇总交叉数据表得出的结论可能相反（统计学家辛普森提出）
  - 例3.9：法官勒基特和肯德尔在民事庭和市政庭主持审理案件，他们判决的部分案件被提出上诉。上诉法庭对大多数上诉案件维持原来的判决，但也有部分判决被推翻。以两个变量—判决（维持或推翻）和法庭类型（民事庭或市政庭）为依据，对每位法官构建交叉分组表。

辛普森悖论

判决	法官		总计
	勒基特	肯德尔	
维持	129（86%）	110（88%）	239
推翻	21（14%）	15（12%）	36
总计（%）	150（100%）	125（100%）	275

辛普森悖论

判决	法官勒基特		总计
	民事庭	市政庭	
维持	29（91%）	100（85%）	129
推翻	3（9%）	18（15%）	21
总计（%）	32（100%）	118（100%）	150

判决	法官肯德尔		总计
	民事庭	市政庭	
维持	90（90%）	20（80%）	110
推翻	10（10%）	5（20%）	15
总计（%）	100（100%）	25（100%）	125

辛普森悖论

- 市政庭被推翻的案件比例较高
- 法官勒基特审理的案件大多数在市政庭
- 在得出结论前应该审查交叉分组表是综合还是未综合形式
- 当交叉分组表包括综合数据时，应该审查是否存在可能影响结论的隐藏变量（法庭类型）

第四章
描述性统计指标

描述性统计指标

- 数据可视化：用表格和图形的方法初步展示了数据中的信息
- 描述性统计指标：数值方法展示数据的汇总信息
  - 单变量数据统计指标
  - 多变量相关性指标

数据中心位置：算术平均数

- （算术）平均值：$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
  - Excel 函数：AVERAGE(data range)
- 平均值的性质：$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$
- 平均值易受极端值影响
  - 练习：计算Purchase Orders数据集中cost per order的平均值
  - =AVERAGE(B2:B95)

数据中心位置：几何平均数

- 几何平均数：$\bar{x}_g = \sqrt[n]{x_1 x_2 \ldots x_n} = (x_1 x_2 \ldots x_n)^{1/n}$
- Excel 函数：=GEOMEAN(data range)
- 常常用于分析财务数据的增长率

例4.1 某基金的年回报率

年	回报率（%）	增长因子
1	-22.1	0.779
2	28.7	1.287
3	10.9	1.109
4	4.9	1.049
5	15.8	1.158
6	5.5	1.055
7	-37.0	0.630
8	26.5	1.265
9	15.1	1.151
10	2.1	1.021

$\bar{x}_g = 1.029$ $\bar{x} = 1.050$

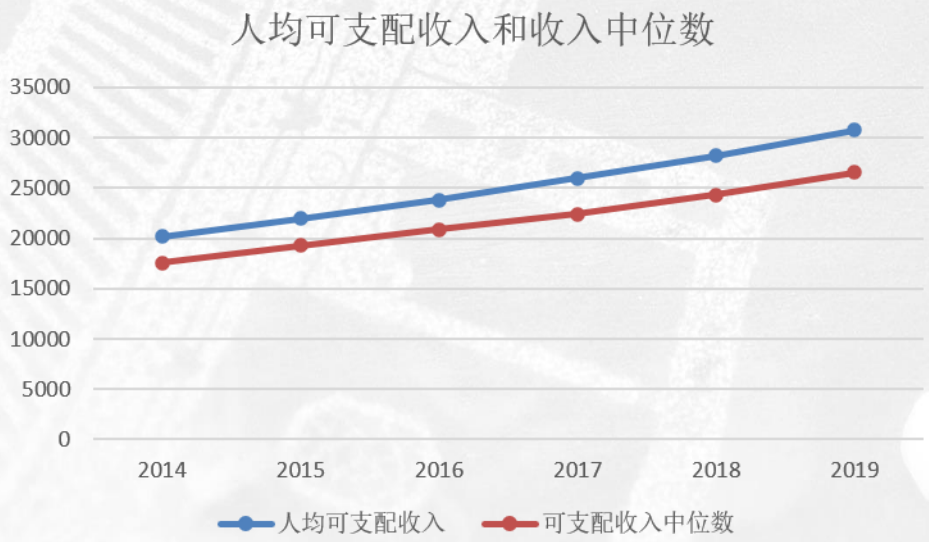
数据中心位置：中位数

- 中位数：数据从小到大排列，位于中间位置的值
  - 一半的数据比中位数小，一半数据比中位数大
  - 有奇数个观测值，中位数是中间位置的值
  - 有偶数个观测值，中位数是中间两个位置值的平均值
  - Excel 函数:=MEDIAN(data range)
- 中位数不易受到极端值的影响
  - 练习：Purchase Orders数据集中计算cost per order的中位数
  - =MEDIAN(B2:B94)

例4.2 居民可支配收入

居民人均可支配收入和收入中位数（元）

	2014	2015	2016	2017	2018	2019
均值	20167	21966	23821	25974	28228	30733
中位数	17570	19281	20883	22408	24336	26523



- 有效推进共同富裕

百分位数和四分位数

- 百分位数：第p百分位数把数据分割为两个部分，大约有p%的观测值比第p百分位数小；而大约有（100-p)%的观测值比第p百分位数大。
  - 将数据从小到大排序，第p百分数位置
  - $L_p = \frac{p}{100}(n+1)$
  - Excel函数：=PERCENTILE.EXC(data range, p%)
- 四分位数
  - Q_1--第一四分位数，第25百分位数
  - Q_2--第二四分位数，第50百分位数（中位数）
  - Q_3--第三四分位数，第75百分位数

例 4.3：电脑修理时长

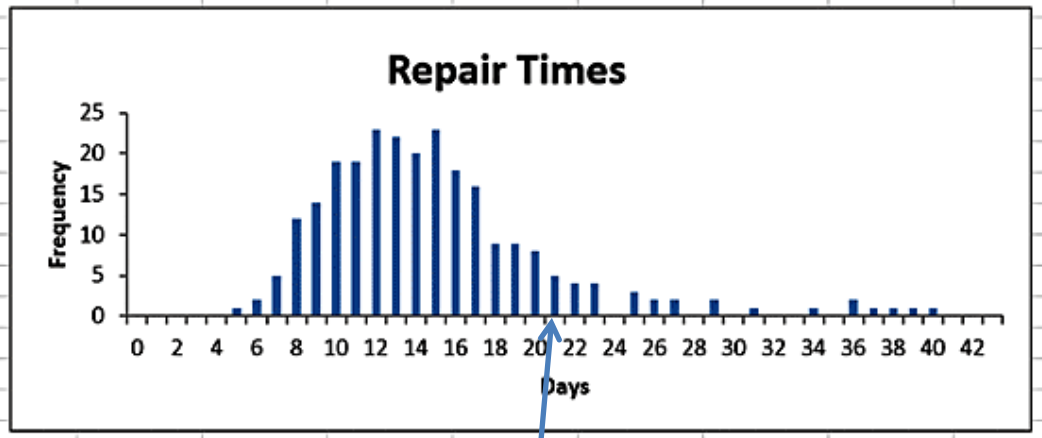
Computer Repair Times数据集 记录250个客户的电脑修理时间

- 当顾客询问大约需要等多久才能取到电脑时，如何回答？
- 中位数修理时间2周；均值大约15天.

	A	B
1	Computer Repair Times	
2		
3	Sample	Repair Time (Days)
4	1	18
5	2	15
6	3	17
250	247	31
251	248	6
252	249	17
253	250	13
254		
255	Mean	14.912
256	Median	14
257	Mode	15

例4.3 电脑修理时长

	A	B	C	D
1	Computer Repair Times			
2				
3			Relative	Cumulative
4	Days	Frequency	Frequency	Percentage
5	0	0	0.000	0.0%
6	1	0	0.000	0.0%
7	2	0	0.000	0.0%
8	3	0	0.000	0.0%
9	4	0	0.000	0.0%
10	5	1	0.004	0.4%
11	6	2	0.008	1.2%
12	7	5	0.020	3.2%
13	8	12	0.048	8.0%
14	9	14	0.056	13.6%
15	10	19	0.076	21.2%
16	11	19	0.076	28.8%
17	12	23	0.092	38.0%
18	13	22	0.088	46.8%



90% 在三周之内修理完

分类型数据位置信息

- 比例，记作p，是某一分类型数据（次品、错误）所占比例.
  - 练习：Purchase Orders数据集中计算供应商为Spacetime Technologies的订单比例
  - 数据透视表

	A	B	C	D	E	F	G	H	I	J
1	**Purchase Orders**									
2										
3	**Supplier**	**Order No.**	**Item No.**	**Item Description**	**Item Cost**	**Quantity**	**Cost per order**	**A/P Terms (Months)**	**Order Date**	**Arrival Date**
4	Spacetime Technologies	A0111	6489	O-Ring	$ 3.00	900	$ 2,700.00	25	10/10/11	10/18/11
5	Steelpin Inc.	A0115	5319	Shielded Cable/ft.	$ 1.10	17,500	$ 19,250.00	30	08/20/11	08/31/11
6	Steelpin Inc.	A0123	4312	Bolt-nut package	$ 3.75	4,250	$ 15,937.50	30	08/25/11	09/01/11
7	Steelpin Inc.	A0204	5319	Shielded Cable/ft.	$ 1.10	16,500	$ 18,150.00	30	09/15/11	10/05/11
8	Steelpin Inc.	A0205	5677	Side Panel	$ 195.00	120	$ 23,400.00	30	11/02/11	11/13/11
9	Steelpin Inc.	A0207	4312	Bolt-nut package	$ 3.75	4,200	$ 15,750.00	30	09/01/11	09/10/11
10	Alum Sheeting	A0223	4224	Bolt-nut package	$ 3.95	4,500	$ 17,775.00	30	10/15/11	10/20/11

数据中心位置：众数

- 众数是发生频率最高的观测值
- 众数适用于这样的数据集：包含的不同数值较少
- 通过统计频数分布可以快速找出众数
- Excel 函数：
  - 单个众数：=MODE.SNGL(data range)
  - 多个众数:选中多个单元格（同一列),=MODE.MULT(data range),ctrl+shift+enter
  - 练习：Purchase orders数据集中计算A/P Terms众数

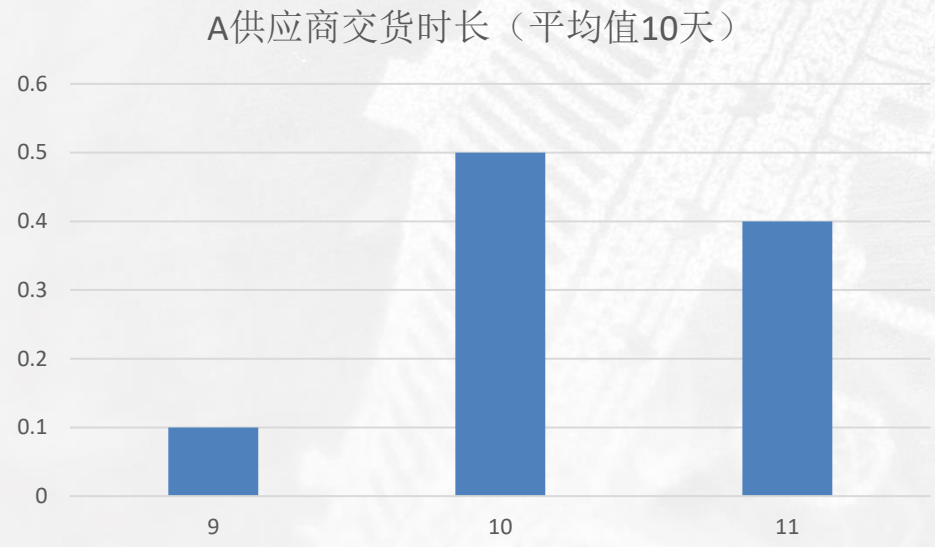
测量量表

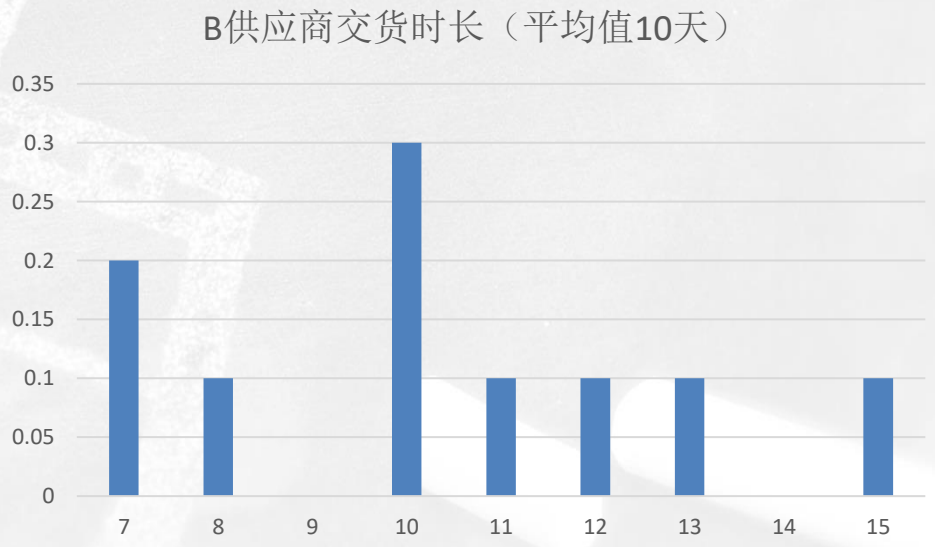
量表	基本比较	例子	平均测量
定类	同一性	男-女、使用-不使用	众数
定序	有序性	品牌偏好、质量等级	中位数
定距	定距比较	对品牌态度	均值
定比	绝对数量比较	单位销售量、购买数量	几何平均数

- 低级别的量表的平均测量指标也可以用于高级别量表

数据变异程度

- 除了位置的度量以外，人们往往还需要考虑变异程度，即数据离散程度。





极差

- 极差（range）：变量的最大观测值与最小观测值之差
- Excel函数：=MAX(*data range*) – MIN(*data range*).
- 极差易受极端值影响
  - 练习：Purchase Orders数据集中计算cost per oerder变量的极差

四分位差

- 四分位差(interquartile range, IQR):第三四分位数Q_3-第一四分位数Q_1
- 克服异常值的影响
  - 练习：Purchase Orders数据集中计算cost per oerder变量的四分位差

方差

- 极差和四分位差：只利用了一部分数据的信息
- 方差：数据离开平均值的距离（离差）平方的平均
- 总体方差：$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$
  - Excel中：=VAR.P(data range)
- 样本方差：$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$
  - Excel中：=VAR.S(data range)
  - 注意分母的差异
  - 练习：计算purchase orders数据集中变量cost per order的方差，注意方差的单位

标准差

- 标准差(standard deviation)：方差的平方根，单位与数据单位一致，因此应用更广泛
- 总体标准差：$\sigma=\sqrt{\sigma^2}$
  - Excel中：=STDEV.P(data range)
- 样本标准差：$s=\sqrt{s^2}$
  - Excel中：=STDEV.S(data range)
  - 练习：计算purchase orders数据集中变量cost per order的标准差
- 财务分析中用标准差度量风险

标准差系数

- 标准差系数（coefficient of variation, CV）$=\frac{\text{标准差}}{\text{平均值}}*100\%$
- 适用于比较具有不同标准差和不同平均数的变量的变异程度
- 较好地衡量了相对投资回报的投资风险
  - 财务风险分析中常用1/CV衡量单位风险的投资回报（越高越好）
  - 股票的夏普值

例4.5 股票的投资风险

- *Closing Stock Prices*数据集
  - Intel (INTC)投资风险较高
  - 指数基金投资风险最低

	A	B	C	D	E	F
1	**Closing Stock Prices**					
2						
3	**Date**	**IBM**	**INTC**	**CSCO**	**GE**	**DJ Industrials Index**
4	9/3/2010	$127.58	$18.43	$21.04	$15.39	10447.93
5	9/7/2010	$125.95	$18.12	$20.58	$15.44	10340.69
6	9/8/2010	$126.08	$17.90	$20.64	$15.70	10387.01
22	9/30/2010	$134.14	$19.20	$21.90	$16.25	10788.05
23	10/1/2010	$135.64	$19.32	$21.91	$16.36	10829.68
24	**Mean**	$130.93	$18.81	$21.50	$16.20	$10,639.98
25	**Standard Deviation**	$3.22	$0.50	$0.52	$0.35	$171.94
26	**Coefficient of Variation**	0.025	0.027	0.024	0.022	0.016

数据标准化

- 标准化之后的值，通过记为z值，提供了观测值离开均值几倍标准差的信息.
- 第i个观测值的z值为$z_i = \frac{x_i - \bar{x}}{s}$
  - z值=1代表观测值在均值右边一倍标准差位置
  - z值=-1.5代表观测值在均值左边1.5倍标准差位置
  - Excel 函数 :=STANDARDIZE(x,mean,standard_dev)

例4.6 计算z值

- *Purchase Orders*数据集中将变量Cost per order进行标准化

	A	B	C
1	Observation	Cost per order	z-score
2	x1	$2,700.00	-0.79
3	x2	$19,250.00	-0.24
4	x3	$15,937.50	-0.35
5	x4	$18,150.00	-0.27
6	x5	$23,400.00	-0.10
91	x90	$6,750.00	-0.65
92	x91	$16,625.00	-0.32
93	x92	$74,375.00	1.61
94	x93	$72,250.00	1.54
95	x94	$6,562.50	-0.66
96			
97	Mean	$26,295.32	
98	Standard Deviation	$29,842.83	

=(B2 - B97)/B98, or
=STANDARDIZE(B2,B97,B98).

车贝晓夫定理

- 车贝晓夫定理：与平均数的距离在k（$k > 1$）标准差之内的数据占了至少$1 - 1/k^2$
  - k = 2：至少3/4的数据在均值附近2倍标准差之内
  - k = 3：至少8/9 的数据在均值附近3倍标准差之内
    - 练习：商务分析班上的同学期末考试平均成绩为70分，标准差为5分。请问有多少学生的考试成绩在60-80分？有多少同学的考试成绩在58-82分？
    - 练习：在任一数据集中验证车贝晓夫定理。
      - 数据“筛选”功能
      - COUNT函数应用
      - 数据透视表

经验法则

- 现实中很多数据的直方图呈现钟形
- 具有钟形分布的数据
  - 大约 68% 的观测值在均值附近一倍标准差内
  - 大约 95% 的观测值在均值附近两倍标准差内
  - 大约 99.7% 的观测值在均值附近三倍标准差内
- 注意，车贝晓夫“定理”适用于任一分布形状的数据

异常值检测

- 异常大或异常小的观测值称为异常值（outlier）
- 检查异常值的反常原因
  - 利用z值检测异常值：z值小于-3或大于3
  - 利用四位分数检测异常值：
    - 下限=$Q_1 - 1.5IQR$
    - 上限=$Q_3 + 1.5IQR$
  - 注意：异常值不一定是错误值；两种方法的上下限可以不同，可以用任一种或两种方法
    - 练习：在purchase orders数据集中用两种方法检测cost per order的异常值

变量相关性统计指标

- 协方差和相关系数
  - 样本协方差：$s_{xy}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}$
    - Excel函数：=COVARIANCE.S(array1,array2)
  - 相关系数（不受数据量纲影响）：$r_{xy}=\frac{s_{xy}}{s_x s_y}$
    - Excel函数：=CORREL(array1,array2)
    - 相关系数的取值在-1到1之间。如何证明？
    - 正负性与协方差一致

协方差和相关系数计算

Colleges and Universities 数据集：数据-数据分析
相关系数

	A	B	C	D	E	F
1		*Median SAT*	*Acceptance Rate*	*Expenditures/Student*	*Top 10% HS*	*Graduation %*
2	Median SAT	1				
3	Acceptance Rate	-0.601901959	1			
4	Expenditures/Student	0.572741729	-0.284254415	1		
5	Top 10% HS	0.503467995	-0.609720972	0.505782049	1	
6	Graduation %	0.564146827	-0.55037751	0.042503514	0.138612667	1

注意：输入的数据必须位于相邻列

相关系数解释

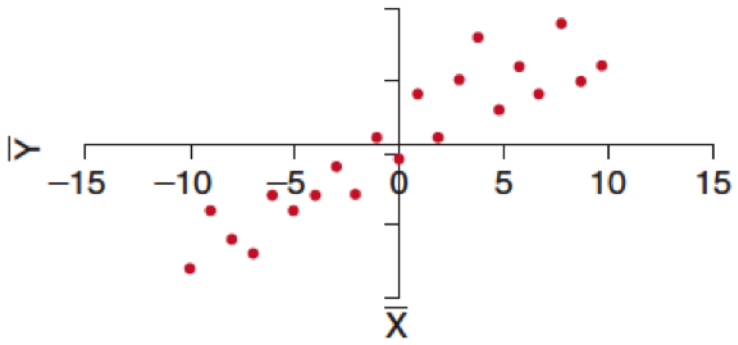
x_i	y_i
5	10
10	30
15	50

$$r_{xy} = 1$$

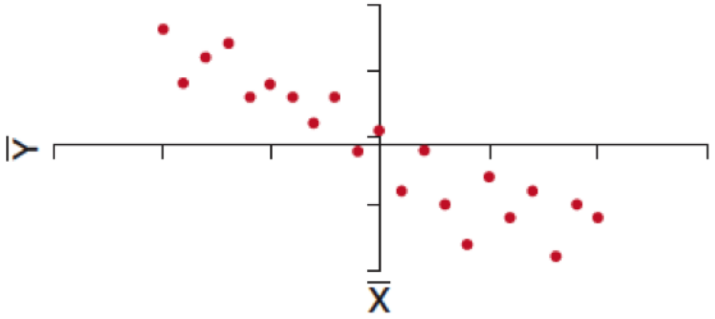
x_i	y_i
5	-10
10	-30
15	-50

$$r_{xy} = -1$$

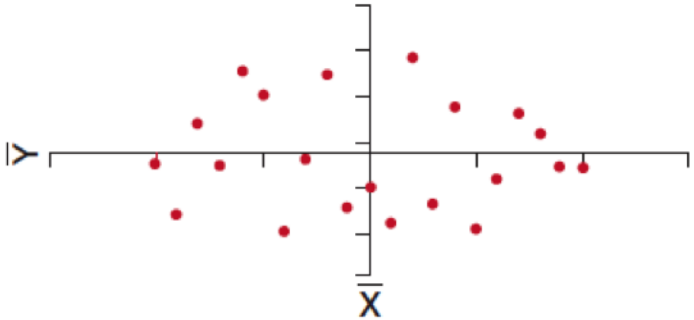
散点图与相关系数



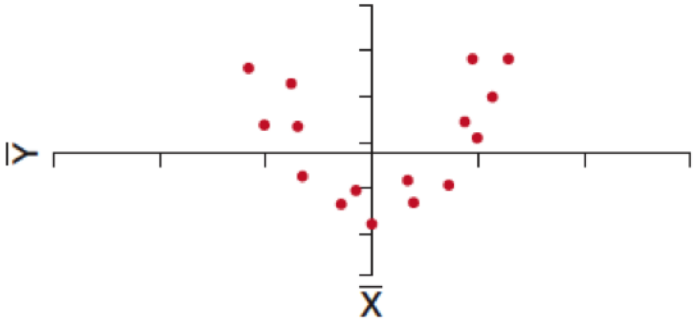
(a) Positive Correlation



(b) Negative Correlation



(c) No Correlation



(d) A Nonlinear Relationship with No Linear Correlation

相关系数展现了两个变量的线性相关程度，结合散点图更好地了解两个变量的相关性

第五章 预测方法

预测方法

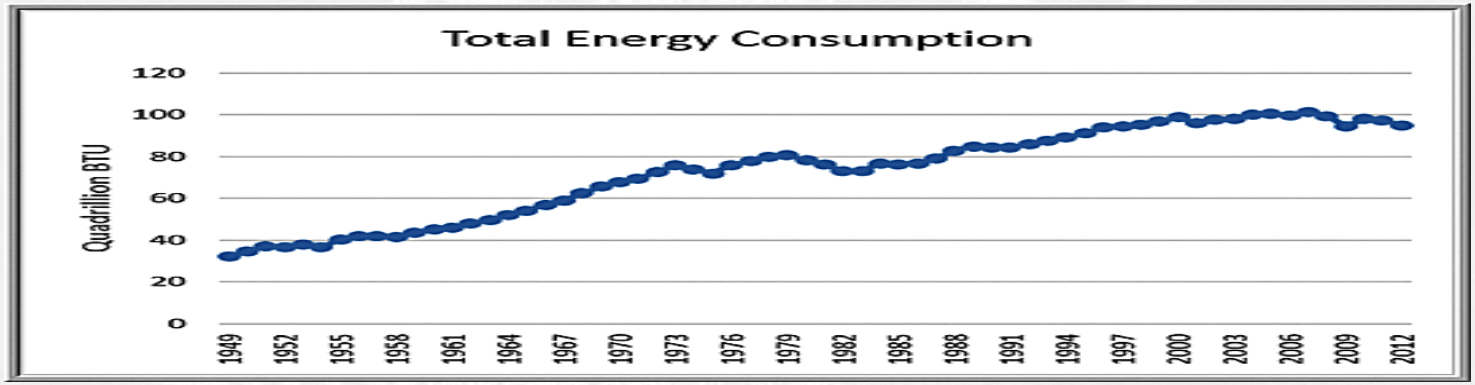
- 经理人需要基于对未来的预测做出决策
- 三大类预测方法：
  - 主观预测
  - 基于时间序列的预测模型
  - 基于特征的预测模型

时间序列预测模型

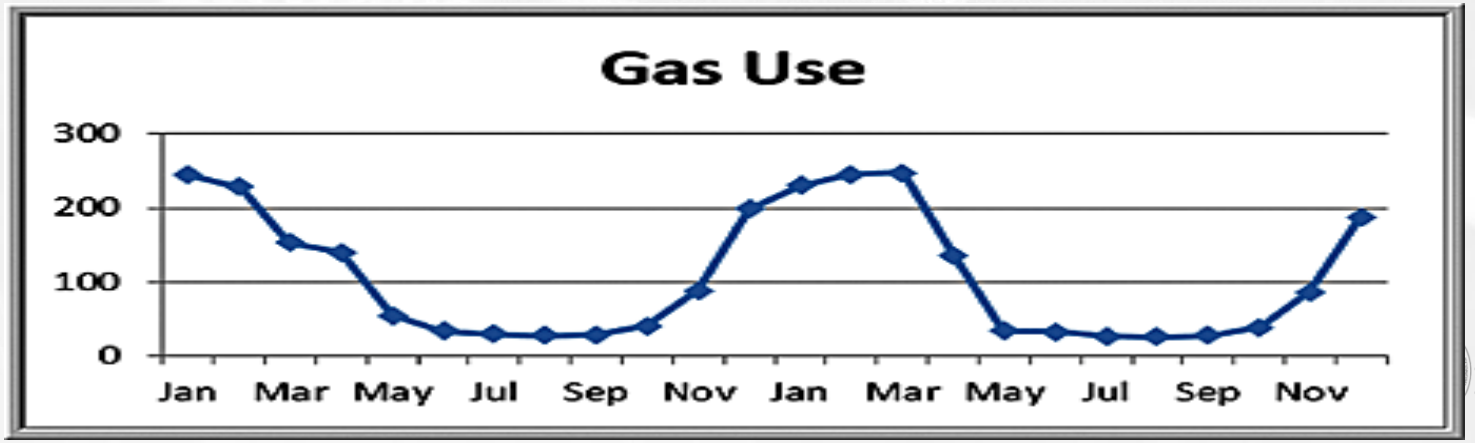
- 时间序列 - 一系列历史数据，例如每周的销售额
  - $t = 1, 2, \ldots, T$
- 时间序列中可能包括趋势性、季节性
- 平稳时间序列：没有趋势性、季节性，只有随机波动

时间序列数据

- 趋势性



- 季节性



平稳时间序列预测

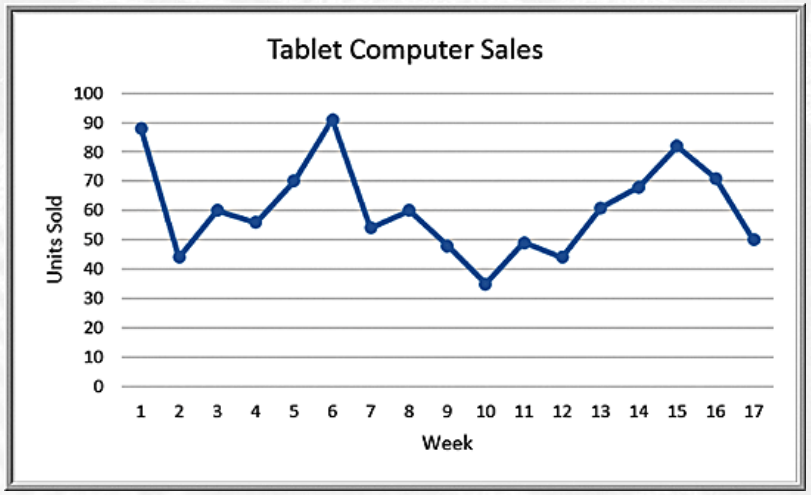
- 移动平均法
- 指数平滑法
  - 时间序列没有明显的趋势和季节性
  - 短期预测

移动平均法

- 移动平均法：对下一期的预测等于最近k期观测值的平均
  - k越大，预测越“平滑”，受极端值影响越小
  - 移动平均法的思想：通过多期的加总，把随机波动的影响去除，得到数据真正的水平

例5.1：移动平均法预测

- *Tablet Computer Sales* 数据集展示了过去17周的平板电脑销量.



- 预测第18周的销量，k=3:

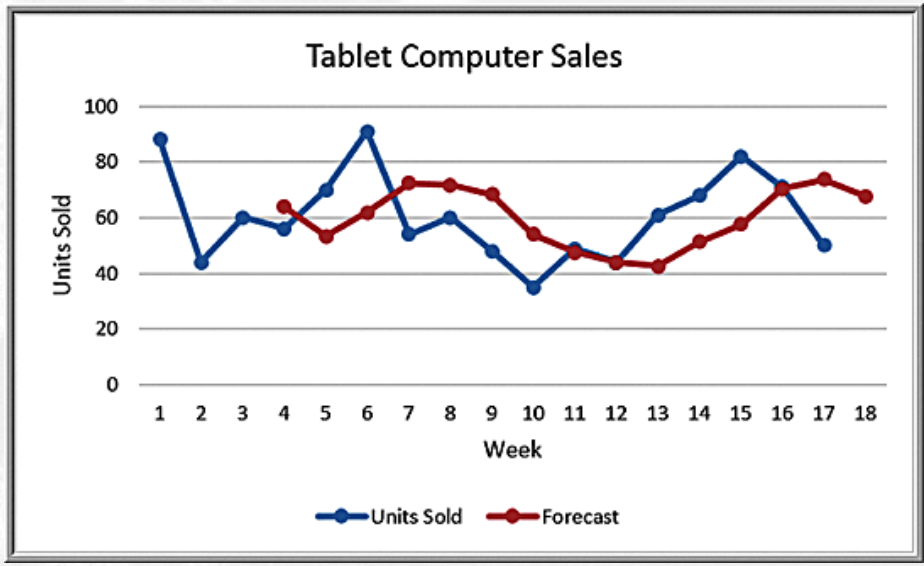
$$(82 + 71 + 50)/3 = 67.67$$

例5.1 续

	A	B	C	D	E	F
1	Tablet Computer Sales					
2			Moving Average			
3	Week	Units Sold	Forecast			
4	1	88				
5	2	44				
6	3	60				
7	4	56	64.00			
8	5	70	53.33			
9	6	91	62.00			
10	7	54	72.33			
11	8	60	71.67			
12	9	48	68.33			
13	10	35	54.00			
14	11	49	47.67			
15	12	44	44.00			
16	13	61	42.67			
17	14	68	51.33			
18	15	82	57.67			
19	16	71	70.33			
20	17	50	73.67			
21	18		67.67			
22						

Forecast for week 4
=AVERAGE(B4:B6)

Forecast for week 18
=AVERAGE(B18:B20)



尝试$k = 5, 10$

预测准确性衡量指标

- Mean absolute deviation (MAD)（平均绝对偏差）

$$MAD = \frac{\sum_{t=1}^{n} |A_t - F_t|}{n}$$

- Mean square error (MSE)（均方误差）

$$MSE = \frac{\sum_{t=1}^{n} (A_t - F_t)^2}{n}$$

- Root mean square error (RMSE)（均方误差根）

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (A_t - F_t)^2}{n}}$$

- Mean absolute percentage error (MAPE)（平均绝对百分比误差）

$$MAPE = \frac{\sum_{i=1}^{n} |\frac{A_t - F_t}{A_t}|}{n} \times 100\%$$

例5.2：比较不同预测模型预测准确度

- *Tablet Computer Sales* 数据集
- K=2，3，4
- K=2 预测结果最准确

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Tablet Computer Sales																
2			k = 2					k = 3					k = 4				
3	Week	Units Sold	Forecast	Error	Absolute Deviation	Squared Error	Absolute % Error	Forecast	Error	Absolute Deviation	Squared Error	Absolute % Error	Forecast	Error	Absolute Deviation	Squared Error	Absolute % Error
4	1	88															
5	2	44															
6	3	60	66.00	-6.00	6.00	36.00	10.00										
7	4	56	52.00	4.00	4.00	16.00	7.14	64.00	-8.00	8.00	64.00	14.29					
8	5	70	58.00	12.00	12.00	144.00	17.14	53.33	16.67	16.67	277.78	23.81	62.00	8.00	8.00	64.00	11.43
9	6	91	63.00	28.00	28.00	784.00	30.77	62.00	29.00	29.00	841.00	31.87	57.50	33.50	33.50	1122.25	36.81
10	7	54	80.50	-26.50	26.50	702.25	49.07	72.33	-18.33	18.33	336.11	33.95	69.25	-15.25	15.25	232.56	28.24
11	8	60	72.50	-12.50	12.50	156.25	20.83	71.67	-11.67	11.67	136.11	19.44	67.75	-7.75	7.75	60.06	12.92
12	9	48	57.00	-9.00	9.00	81.00	18.75	68.33	-20.33	20.33	413.44	42.36	68.75	-20.75	20.75	430.56	43.23
13	10	35	54.00	-19.00	19.00	361.00	54.29	54.00	-19.00	19.00	361.00	54.29	63.25	-28.25	28.25	798.06	80.71
14	11	49	41.50	7.50	7.50	56.25	15.31	47.67	1.33	1.33	1.78	2.72	49.25	-0.25	0.25	0.06	0.51
15	12	44	42.00	2.00	2.00	4.00	4.55	44.00	0.00	0.00	0.00	0.00	48.00	-4.00	4.00	16.00	9.09
16	13	61	46.50	14.50	14.50	210.25	23.77	42.67	18.33	18.33	336.11	30.05	44.00	17.00	17.00	289.00	27.87
17	14	68	52.50	15.50	15.50	240.25	22.79	51.33	16.67	16.67	277.78	24.51	47.25	20.75	20.75	430.56	30.51
18	15	82	64.50	17.50	17.50	306.25	21.34	57.67	24.33	24.33	592.11	29.67	55.50	26.50	26.50	702.25	32.32
19	16	71	75.00	-4.00	4.00	16.00	5.63	70.33	0.67	0.67	0.44	0.94	63.75	7.25	7.25	52.56	10.21
20	17	50	76.50	-26.50	26.50	702.25	53.00	73.67	-23.67	23.67	560.11	47.33	70.50	-20.50	20.50	420.25	41.00
21	18		60.50		13.63	254.38	23.63	67.67		14.86	299.84	25.37	67.75		16.13	355.25	28.07
22					MAD	MSE	MAPE			MAD	MSE	MAPE			MAD	MSE	MAPE

指数平滑法

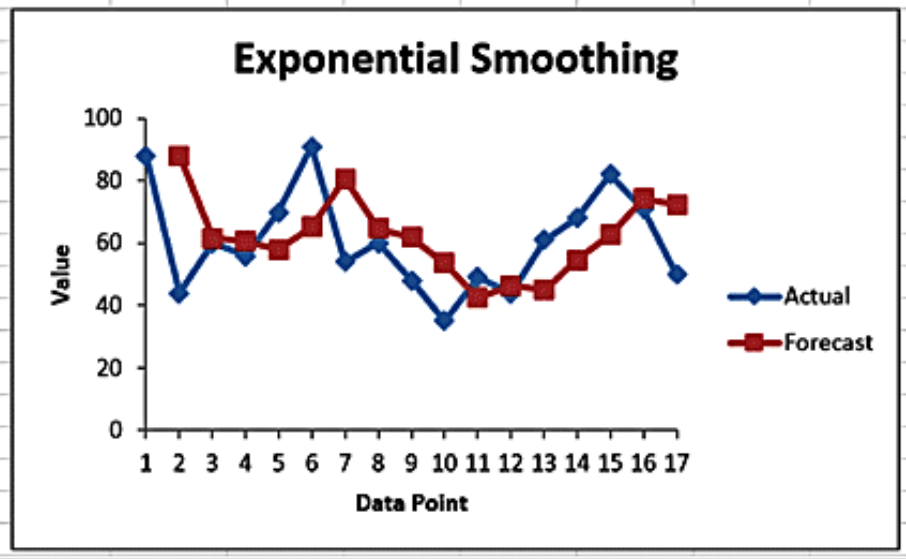
- **简单指数平滑模型：**
- $F_{t+1} = (1-\alpha)F_t + \alpha A_t = F_t + \alpha(A_t - F_t)$
  - $0 \leq \alpha \leq 1$ 平滑系数
  - $F_1 = A_1$

例5.3 指数平滑法

	A	B	C	D	E	F	G	H	I	J	K
1	Tablet Computer Sales										
2			Smoothing Constant								
3	Week	Units Sold	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
4	1	88	88.00	88.00	88.00	88.00	88.00	88.00	88.00	88.00	88.00
5	2	44	88.00	88.00	88.00	88.00	88.00	88.00	88.00	88.00	88.00
6	3	60	83.60	79.20	74.80	70.40	66.00	61.60	57.20	52.80	48.40
7	4	56	81.24	75.36	70.36	66.24	63.00	60.64	59.16	58.56	58.84
8	5	70	78.72	71.49	66.05	62.14	59.50	57.86	56.95	56.51	56.28
9	6	91	77.84	71.19	67.24	65.29	64.75	65.14	66.08	67.30	68.63
10	7	54	79.16	75.15	74.37	75.57	77.88	80.66	83.53	86.26	88.76
11	8	60	76.64	70.92	68.26	66.94	65.94	64.66	62.86	60.45	57.48
12	9	48	74.98	68.74	65.78	64.17	62.97	61.87	60.86	60.09	59.75
13	10	35	72.28	64.59	60.45	57.70	55.48	53.55	51.86	50.42	49.17
14	11	49	68.55	58.67	52.81	48.62	45.24	42.42	40.06	38.08	36.42
15	12	44	66.60	56.74	51.67	48.77	47.12	46.37	46.32	46.82	47.74
16	13	61	64.34	54.19	49.37	46.86	45.56	44.95	44.70	44.56	44.37
17	14	68	64.00	55.55	52.86	52.52	53.28	54.58	56.11	57.71	59.34
18	15	82	64.40	58.04	57.40	58.71	60.64	62.63	64.43	65.94	67.13
19	16	71	66.16	62.83	64.78	68.03	71.32	74.25	76.73	78.79	80.51
20	17	50	66.65	64.47	66.65	69.22	71.16	72.30	72.72	72.56	71.95
21	18		64.98	61.57	61.65	61.53	60.58	58.92	56.82	54.51	52.20
22		MAD	19.33	17.16	16.15	15.36	14.93	14.71	14.72	14.88	15.36
23		MSE	496.07	390.84	359.18	346.56	340.77	338.41	339.03	343.32	352.38
24		MAPE	38.28%	32.71%	30.12%	28.36%	27.54%	27.09%	27.09%	27.38%	28.23%

例5.3 续：指数平滑法

	A	B	C
1	Tablet Computer Sales		
2			Exponential Smoothing
3	Week	Units Sold	Forecast
4	1	88	#N/A
5	2	44	88.00
6	3	60	61.60
7	4	56	60.64
8	5	70	57.86
9	6	91	65.14
10	7	54	80.66
11	8	60	64.66
12	9	48	61.87
13	10	35	53.55
14	11	49	42.42
15	12	44	46.37
16	13	61	44.95
17	14	68	54.58
18	15	82	62.63
19	16	71	74.25
20	17	50	72.30



第六章 决策分析

决策模型

- 确定性模型
- 不确定性模型

确定性模型

- 确定性模型
- 例6.1：供需匹配问题
  - KDGL是一家为零售商提供物流服务的运输公司
  - 下个月，KDGL需要为一家客户提供饮料的配送
  - 客户拥有三个仓库，分别位于Los Angeles(L)，Chicago(C)，和New York City(N)；并且拥有三个分销中心，位于Denver(D)，Austin(A)，和Washington，DC(W)。
  - 每个仓库都有一定数量的饮料，必须被运走，以腾出空间存放新到的货物。

例6.1续

- 每个仓库数量

Warehouse	To Be Shipped Out (tons)
Los Angeles (L)	15
Chicago (C)	20
New York City (N)	30

- 每个分销中心都有需求必须被满足

Distribution Center	Minimum To Be Shipped In (tons)
Denver	10
Austin	13
Washington, DC	20

例6.1续

- 不同地点之间的单位运输成本不同

From / To	Denver (D)	Austin (A)	Washington, DC (W)
Los Angeles (L)	$105.00	$135.00	$153.00
Chicago (C)	$110.00	$140.00	$137.00
New York City (N)	$130.00	$132.00	$115.00

例6.1续

- KDGL的决策、目标和约束？

Minimize the Total Shipping Cost

$$105*X_{LD} + 135*X_{LA} + 153*X_{LW} + 110*X_{CD} + 140*X_{CA} + 137*X_{CW} + 130*X_{ND} + 132*X_{NA} + 115*X_{NW}$$

$X_{LD} + X_{LA} + X_{LW} = 15$ (Los Angeles supply)

$X_{CD} + X_{CA} + X_{CW} = 20$ (Chicago supply)

$X_{ND} + X_{NA} + X_{NW} = 30$ (New York supply)

$X_{LD} + X_{CD} + X_{ND} \geq 10$ (Minimum Denver demand)

$X_{LA} + X_{CA} + X_{NA} \geq 13$ (Minimum Austin demand)

$X_{LW} + X_{CW} + X_{NW} \geq 20$ (Minimum Washington, DC demand)

$X_{LD}, \ldots, X_{NW} \geq 0$ (Non-negativity)

确定性模型

- 确定性模型的好处
  - 可以解决较大规模的问题，多产品，多资源
  - 当决策者对环境有较大的控制时确定性假设有合理的地方，例如
    - 制定短期计划
    - 制定带长期合同（价格、需求、供给）的长期计划
- 确定性模型的缺点
  - 当现实中不确定性较高时，提供的决策不可靠

不确定性模型

- 仿真模拟
- 决策树
- …

决策树模型

- 面临高不确定性的决策工具，适用于单阶段或多阶段决策问题
- 单阶段决策问题：例6.2 供应商选择问题
  - IDEA售卖一种夏季的户外帐篷，必须在夏季到来前提前生产。帐篷的价格被定为150元
  - 根据以往的经验，IDEA预测在接下来的夏季中，*需求有50%的可能性是强的*，能达到10000件，*有50%的可能性是弱的*，只能卖出5000件
  - IDEA有两个备选供应商， 一个位于瑞士（S），另一个位于波兰（P）。如果IDEA与供应商签约的话，必须向供应商保证买断它所有的产能。IDEA最多只能和一个签约

决策树模型

- 两个供应商的产能、成本等参数

	S	P
capacity	5000 units	10000 units
Up-front charge by supplier	0	$50000
Unit price	$150	$150
Labor costs	$60	$30
Material costs	$40	$40
shipping	$20	$30
Unit cost	$120	$100

决策树模型

- 决策点
  - 决策者必须做决策，从这类节点引出的边表示不同的决策方案
  - IDEA选择哪个供应商，或不选择
- 事件点or机会点
  - 存在不确定事件的点，从这类节点引出的边表示不同的事件，边下的数字表示对应事件出现的概率
  - IDEA:市场是强的或弱的
- 结果点
  - 给定决策者和随机事件的结果位于树的末梢端，并在这类节点旁注明各种结果的收益损失
  - IDEA：IDEA的利润受决策者的选择和市场强弱的影响

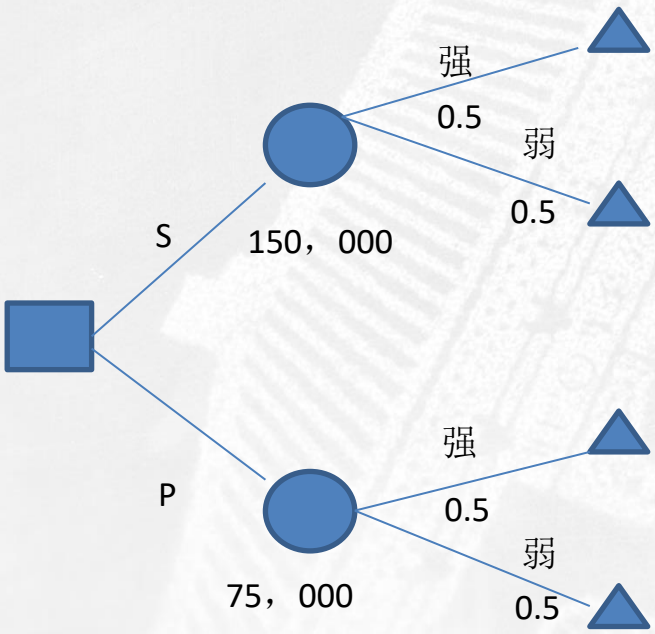
决策树模型

- 建立决策树：决策点、事件点、结果点
- 浏览决策树：了解结果值的范围
- 利用决策树做出决策
  - 从最后的结果点开始，向树根推算
  - 在事件点中，计算最小值/最大值/平均值
  - 在决策点，把不能最大化利润的决策剔除
- 评价标准：风险规避、追逐、中性

决策树模型

- 评价标准
  - “maxi-min” 策略:
    - 最大化最差情况下的收益
    - 忽略较好情况下的收益
    - 风险规避
  - “maxi-max” 策略:
    - 最大化最好情况下的收益
    - 忽略较差情况下的收益
    - 风险追逐
  - 最大化期望收益:
    - 计算平均收益
    - 考虑了各种可能情况
    - 风险中性

决策树



（150-120）*5000=150，000

（150-120）*5000=150，000

（150-100）*10000-50000=450，000

150*5000-100*10000-50000=-300，000