

商务统计分析（第2版）习题参考答案

第1章 数据与统计分析

思考题

1.1.-1.4.略

1.5. 指出下列变量的类型。

- (1) 身高：数值型变量。
- (2) 性别：类别变量。
- (3) 产品等级：顺序变量。
- (4) 每月网购次数：离散型变量。
- (5) 商品购买时的支付方式（现金、银行卡、支付宝、微信）：类别变量。

练习题

1.6. 一家研究机构从西安高校应届本科毕业生中随机抽取 2000 人作为样本进行调查，发现其中 20%的毕业生月收入在 12000 元以上。

(1) 这一研究的总体？

答案：西安高校应届本科毕业生

(2) 月收入属于类别变量、顺序变量还是数值型变量？

答案：数值型变量

(3) 这一研究涉及的是截面数据还是时间序列数据？

答案：截面数据

1.7. 一家大型食品零售商收到许多客户关于 A 品牌糕点礼盒分量不足的抱怨。因而，该零售商对 A 糕点礼盒进行检查，从新进的一万盒糕点中抽查了 100 个糕点礼盒，每一个的质量精确到小数点后三位，装满的糕点礼盒应为 2.253kg。

(1) 描述总体。

答案：新进的一万盒糕点

(2) 描述研究变量。

答案：糕点礼盒的质量

(3) 描述样本。

答案：新进糕点中的 100 个糕点礼盒

(4) 描述推断。

答案： $100 \text{ 个糕点礼盒的重量} = 2.253\text{kg/个} \times 100\text{个} = 225.3\text{kg}$

1.8. 一客运公司想对长途汽车的年平均客运收入进行估算，为此从公司名下的所有长途汽车中随机抽取 100 辆研究，这 100 辆长途汽车的年平均客运收入为 22.73 万元/辆。

请回答：

(1) 这一研究的总体。

答案：该公司名下所有长途汽车的集合

(2) 该研究关心的变量是什么。

答案：该公司名下长途汽车的年客运收入

(3) 该调查使用的是描述统计方法还是推断统计方法？

答案：推断统计方法

第2章 数据的图表描述

思考题

2.1. 常用的定性数据和定量数据的图示方法各有哪些？

答案：定性数据：频数（相对频数、百分数频数）分布、条形图、饼形图。

定量数据：除上述方法外，还有直方图、累积频数（相对频数、百分数频数）分布、茎叶图。

2.2. 条形图和直方图有什么区别？

答案：首先，条形图是用条形的长度表示各类别频数的多少，其宽度(表示类别)则是固定的；直方图是用面积表示各组频数的多少，矩形的高度表示每一组的频数或频率，宽度则表示各组的组距，因此其高度与宽度均有意义。

其次，由于很多定量数据具有连续性，直方图的各矩形通常是连续排列的，而条形图则是分开排列的。

最后，条形图主要用于展示定性数据，而直方图则主要用于展示定量数据。

2.3. 饼形图和条形图使用时有什么选择标准？

答案：饼形图和条形图功能类似，大多数情形中，如果旨在比较各定性变量观测值所占百分比的相对区别时，我们多使用提供更多详实信息的饼形图；若是更关注每组观测值的频数比较，则会优先选择条形图。

2.4. 茎叶图和直方图相比有什么优点？两者的应用场合是什么？

答案：直方图可以直观看清数据分布的总体趋势，但是无法反映原始的数据内容。茎叶图则不同，它不仅可以看出数据的分布，还能保留原始数据的信息。制作茎叶图不需要对数据进行分组，特别是当数据量较少时，用茎叶图更容易观察数据的分布。茎叶图的数据可以随时记录，随时添加，方便记录与表示。

茎叶图和直方图的应用取决于总体中的个体数。当总体中的个体取值较少时，用茎叶图估计总体的分布。并且，茎叶图只方便记录较少的数据。当总体中的个体取值较多时，将样本数据恰当分组，选择直方图。

2.5. 常用的两个变量数据的描述方法有哪些？

答案：两个变量数据的表格描述：交叉分组表。

两个变量数据的图形描述：散点图、线图（时间序列图）、复合条形图与结构条形图、环形图。

2.6. 如何创建有效的图形？

答案：(1) 标题要清晰、简明。

(2) 图形保持简洁，当能用二维表示时不要用三维表示。

(3) 坐标轴有清楚的标记，并给出测量的单位。

(4) 如果使用颜色区分类别，要确保颜色是不同的。

(5) 如果使用多种颜色或线型，用图例来标明时，要将图例靠近所表示的数据。

练习题

2.7. 为推广试验新的网站，知乎非常关注其网页导航设计的简便程度。随机选择 200 个普通的网络用户，要求他们利用新网页进行一项网络搜索。每个人需要对网络导航操作的简便程度进行评价，具体分为较差、较好、良好和非常好四个等级。调查结果如表 2-24 所示。

表 2-24 200 名网络用户对知乎的评价统计

等级	人数
非常好	102
良好	58
较好	30
较差	10

请选择恰当的图形来显示这些信息。

答案：（1）条形图 2-1 可以反映评价等级的分布。

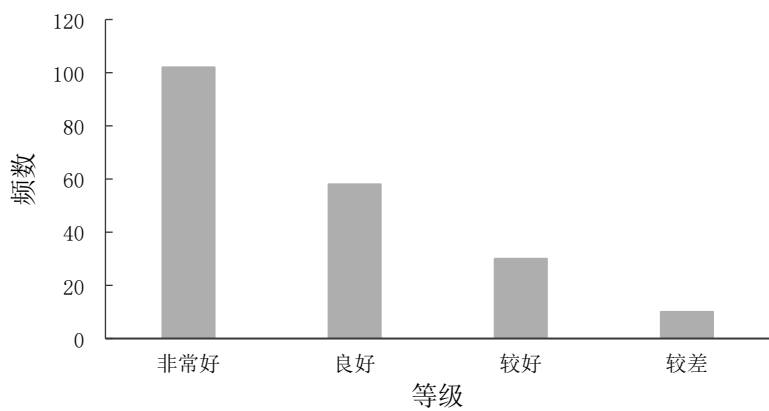


图 2-1 评价等级条形图

（2）饼形图 2-2 可以反映评价等级的构成。

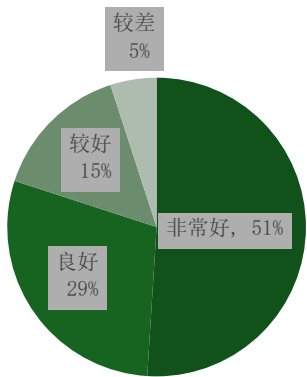


图 2-2 评价等级饼形图

2.8. 表 2-25 是一个大众汽车集团某个月出售的 180 辆汽车的销售利润表（单位美元）。

表 2-25 大众汽车的销售利润（美元）

1387	2148	2201	963	820	2230	3043	2584	2370
------	------	------	-----	-----	------	------	------	------

1754	2207	996	1298	1266	2341	1059	2666	2637
1817	2252	2813	1410	1741	3292	1674	2991	1426
1040	1428	323	1553	1772	1108	1807	934	2944
1273	1889	352	1648	1932	1295	2056	2063	2147
1529	1166	482	2071	2350	1344	2236	2083	1973
3082	1320	1144	2116	2422	1906	2928	2856	2502
1951	2265	1485	1500	2446	1952	1269	2989	783
2692	1323	1509	1549	369	2070	1717	910	1538
1206	1761	1638	2348	978	2454	1797	1536	2339
1342	1919	1961	2498	1238	1606	1955	1957	2700
443	2357	2127	294	1818	1680	2199	2240	2222
754	2866	2430	1115	1824	1827	2482	2695	2597
1621	732	1704	1124	1907	1915	2701	1325	2742
870	1464	1876	1532	1938	2084	3210	2250	1837
1174	1626	2010	1688	1940	2639	377	2279	2842
1412	1761	2165	1822	2197	842	1220	2626	2434
1809	1915	2231	1897	2646	1963	1401	1501	1640
2415	2119	2389	2445	1461	2059	2175	1752	1821
1546	1766	335	2886	1731	2338	1118	2058	2487

- (1) 根据表 2-25 绘制频数分布表。
(2) 根据表 2-25 绘制直方图。
(3) 基于直方图，你能得出哪些蕴含的信息？

答案：(1) 频数分布表如表 2-1 所示。

表 2-1 销售利润频数分布表

销售利润（美元）	频数
200~600	8
600~1000	11
1000~1400	23
1400~1800	38
1800~2200	45
2200~2600	32
2600~3000	19
3000~3400	4
合计	180

- (2) 组数：8；组距= $\frac{3292-294}{8} \approx 400$ 。直方图如图 2-3 所示。

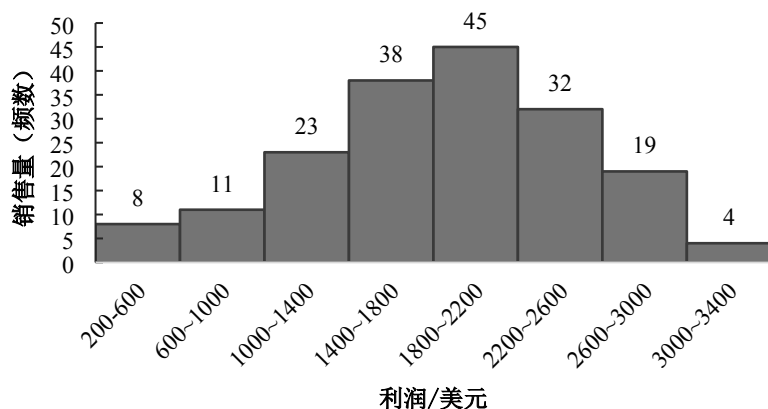


图 2-3 销售利润的直方图

(3) a.销售利润集中在 1000~3000 美元, 157 辆 (87%) 汽车的销售利润均在这一范围。

b.频数最多的组是 1800~2200 美元这一组, 这一组的中间值是 2000 美元, 所以可以说售出一辆汽车的典型利润是 2000 美元。

2.9. 表 2-26 是某运动员最近 13 次的训练得分, 绘制茎叶图, 并说明运动员的发挥稳定程度。

表 2-26 某运动员的训练得分统计

12	25	31	49	44	24	37
31	36	36	15	50	39	

答案: 茎叶图如图 2-4 所示。

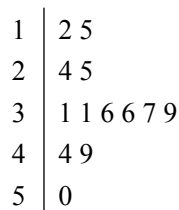


图 2-4 运动员成绩的茎叶图

运动员的得分大都在 20 和 40 之间, 分数较对称, 集中程度高, 说明其发挥比较稳定。

2.10. 基金类型分为成长型和价值型两个类别, 风险水平可分为低、中和高三个类别。表 2-27 是 316 只基金风险水平总结表。

表 2-27 316 只基金风险水平统计

基金风险水平	成长型基金	价值型基金
低	143	69
中	74	17
高	10	3

(1) 根据表 2-27 绘制基金类型和风险水平的交叉分组表。

(2) 根据表 2-27 分别绘制基于总和百分比的交叉分组表、基于行和百分比的交叉分组表和基于列和百分比的交叉分组表, 并进行描述性分析。

(3) 选择合适的图表将表 2-27 中的基金样本数据可视化, 并进行描述性分析。

答案: (1) 基金类型和风险水平的交叉分组表如表 2-2 所示。

表 2-2 基金类型和基金风险水平的交叉分组表

基金类型	风险水平			合计
	低	中	高	

成长型	143	74	10	227
价值型	69	17	3	89
总计	212	91	13	316

(2) 基于总和百分比的交叉分组表、基于行和百分比的交叉分组表和基于列和百分比的交叉分组表分别如表 2-3、2-4 和 2-5 所示。

表 2-3 基金类型和基金风险水平的基于总和百分比的交叉分组表

基金类型	风险水平			合计
	低	中	高	
成长型	45.25%	23.42%	3.16%	71.83%
价值型	21.84%	5.38%	0.95%	28.17%
总计	67.09%	28.80%	4.11%	100.00%

表 2-4 基金类型和基金风险水平的基于行和百分比的交叉分组表

基金类型	风险水平			合计
	低	中	高	
成长型	63.00%	32.60%	4.40%	100.00%
价值型	77.53%	19.10%	3.37%	100.00%
总计	67.09%	28.80%	4.11%	100.00%

表 2-5 基金类型和基金风险水平的基于列和百分比的交叉分组表

基金类型	风险水平			合计
	低	中	高	
成长型	67.45%	81.32%	76.92%	71.84%
价值型	32.55%	18.68%	23.08%	28.16%
总计	100.00%	100.00%	100.00%	100.00%

表 2-3 可以看出，在样本基金中，71.83%是成长型基金，28.17%是价值型基金。低、中、高风险的基金各占 67.09%、28.80%和 4.11%。

表 2-4 可以看出，成长型基金中，低、中、高风险的各占 63.00%、32.60%和 4.40%；而价值型基金中，低、中、高风险的各占 77.53%、19.10%和 3.37%。

表 2-5 可以看出，低风险的基金中，成长型基金和价值型基金各占 67.45%和 32.55%；中等风险的基金中，成长型基金和价值型基金各占 81.32%和 18.68%；高风险的基金中，成长型基金和价值型基金各占 76.92%和 23.08%。

(3)

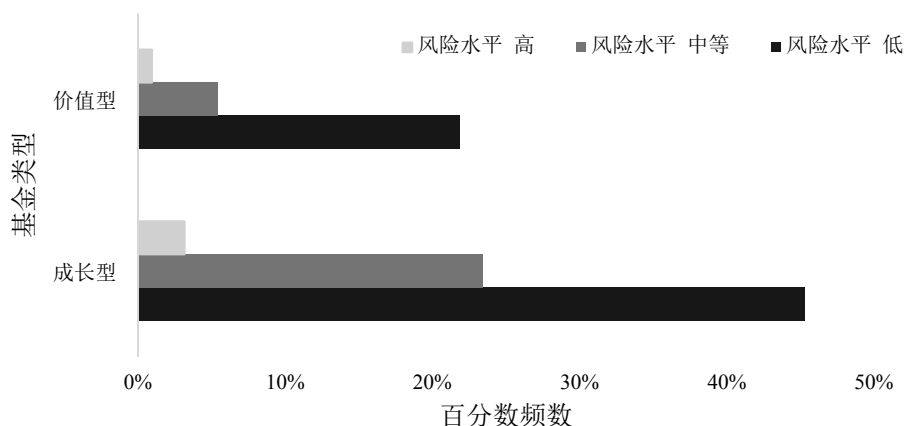


图 2-5 基金类型与风险的复合条形图

从复合条形图2-5以看到成长型基金和价值型基金中都有很大一部分是低风险的。然而，成长型基金中具有低风险和中等风险的基金占比都更大。

2.11. 表 2-28 是 30 个球队的收入和估值数据，为了迅速使球队收入（百万美元）和估值（百万美元）的关系可视化，绘制一张散点图，并说明两者之间的关系。

表 2-28 30 个球队的收入和估值统计

收入	估值	收入	估值	收入	估值
119	425	191	775	144	590
169	875	121	475	139	560
190	780	128	575	117	469
115	410	295	1350	137	565
195	1000	126	453	140	587
145	515	188	770	115	550
162	765	109	405	167	660
124	495	116	430	149	520
139	450	116	420	131	525
160	750	287	1400	122	485

答案：绘制散点图如图 2-6 所示。

从散点图 2-6 可以看出，球队的收入和估值之间表现出很强的正相关关系。也就是说，收入较少球队的估值也比较低，而收入较高球队的估值也比较高。

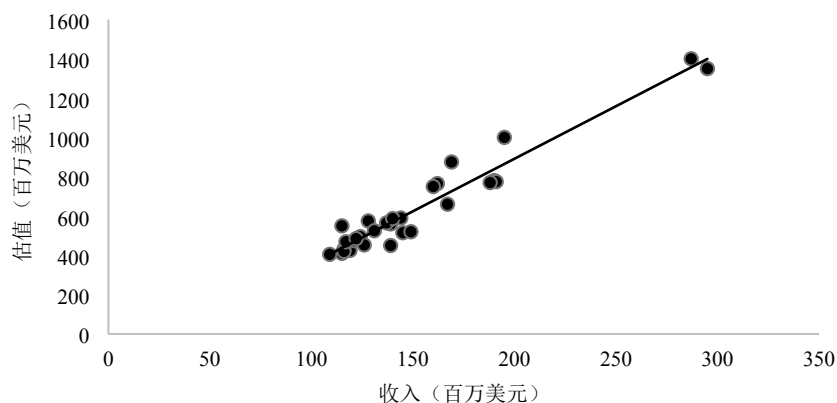


图 2-6 球队收入和估值的散点图

第 3 章 数据的数字描述

思考题

3.1. 一组数据的分布特征可以从哪些方面和用哪些统计量进行测度？

答案：位置：平均数、分位数、众数；离散程度：极差和四分位差、平均差、方差和标准差、离散系数；分布形状：偏态、峰态；相对位置：标准分数。

3.2. 对于比率数据的平均数为什么采用几何平均数？

答案：比率数据是相对数，只能对其进行连乘，连加没有意义。

3.3. 简述平均数、中位数、众数的关系以及他们各自的特点和应用场合。

答案：略

3.4. 简述极差、四分位差、平均差、方差和标准差的特点和应用场合。

答案：极差计算过程简单，但是容易受到极值的影响。四分位差在一定程度上减轻了极值的影响。但这两个统计量都没有考虑数据的分布情况，因此只适用于简单的分析数据的离散程度。平均差能够比较全面地反映数据的离散程度，但是其数学性质差，实际应用较少。方差和标准差是应用最广泛的离散统计量，考虑了数据的分布情况，相比之下标准差与原数据的单位一致，更容易进行比较。

3.5. 为什么要计算离散系数？离散系数用于哪些场合？

答案：消除变量水平高低和计量单位不同对离散程度测度值的影响；主要用于对不同组别数据离散程度的比较。

3.6. 测度一组数据分布形状的统计量有哪些？

答案：偏态、峰态等。

3.7. 标准分数有哪些用途？

答案：去除量纲、数据标准化处理、判断异常值等。

3.8. 相关系数相比于协方差有哪些优点。

答案：协方差可以得到任何值，因此很难根据协方差判定变量之间的相对关系强弱。相关系数的范围在-1 到 1 之间，因此可以用来比较变量间的相关程度强弱

3.9. 相关系数和协方差能否证明两个变量存在因果关系。

答案：不能，相关系数和协方差是两个变量之间线性相关关系的度量，而不是因果关系的度量。

练习题

3.10. 2018 年某企业精加工车间 20 名工人加工 A 零件的日产量资料如下：

按日产量分组（件）	工人人数（人）
28	2
29	4
30	7
31	5
32	2
合计	20

计算 20 名工人日产量的平均数、中位数和众数。

答案：平均数：30.05 件；中位数：30 件；众数：30 件。

3.11. 某投资银行的年利率按复利计算，2007-2018 年的年利率分组资料如下：

按年利率分组（%）	年数（年）
6	2
8	4
9	3
12	2
15	1
合计	12

计算 2007-2018 年的平均年利率。

答案：平均年利率：9.14%

3.12. 某生产流水线有前后衔接的五道工序，某日各工序产品的合格率分别为 95%、92%、90%、85%、80%。请问整条生产流水线产品的平均合格率是多少？

答案：平均合格率：88.32%

3.13. 某冰箱零售店的 10 名销售人员 8 月份销售的冰箱数量（单位：台）排序如下：

4 7 9 10 10 10 12 13 15 15，

（1）计算冰箱销售量的平均数、中位数和众数；

答案：平均数：10.5 台，中位数：10 台，众数：10 台。

（2）计算冰箱销售量的四分位数；

答案：下四分位数 9.25 台，第二四分位数（中位数）10 台，上四分位数 12.75 台

（3）计算冰箱销售量的极差和四分位差；

答案：极差 11 台，四分位差 4.5 台

（4）计算冰箱销售量的标准差；

答案：标准差 3.26

3.14. 随机抽取某地区 10 名成年人和 10 名幼儿，对其身高进行测量统计，得到结果如下（单位：cm）：

成年	168	169	173	175	182	180	176	188	173	178
幼儿	66	69	65	63	71	70	69	67	70	68

（1）哪些统计量适合用来比较成年人或幼儿的身高差异，为什么？

答案：离散系数

（2）成年人的身高差异大还是幼儿的身高差异大？为什么？

答案：成人身高的离散系数为 0.034，儿童身高的离散系数为 0.037

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

<https://d.book118.com/558014001103006052>