

附录 C 自测题解答与偶数题答案

第 1 章

2. a. 10 个个体
b. 5 个变量：价格（美元）、操作系统、显示器大小（英寸）、电池寿命（小时）和 CPU 厂商
c. 分类变量：操作系统和 CPU 厂商
数量变量：价格（美元）、显示器大小（英寸）和电池寿命（小时）
d. 价格、显示器大小和电池寿命是比例尺度，操作系统和 CPU 厂商是名义尺度。
3. a. 582.90 美元
b. Windows 操作系统：723.20 美元
Android 操作系统：428.5 美元
c. 20%
d. 40%
4. a. 8 个个体
b. 分类变量：语音质量、手机在底座上
数量变量：价格、综合得分和通话时间
c. 价格是比例尺度，综合得分是间隔尺度，语音质量是名义尺度，手机在底座上是名义尺度，通话时间是比例尺度。
6. a. 分类型
b. 数量型
c. 分类型
d. 数据型
e. 分类型
8. a. 762
b. 分类型
c. 百分比
d. $0.67 \times 762 = 510.54$ ；510 或 511 人说希望该修正案通过。
10. a. 分类型
b. 百分比
c. 15%
d. 支持反对
12. a. 所有到夏威夷的观光者
b. 是
c. 问题 1 和 4 是数量型数据，问题 2 和 3 是分类型数据。
13. a. 谷歌的年收入
b. 数量型

- c. 时间序列
d. 谷歌的年收入随时间而增长
14. a. 绘制每家公司的时间序列曲线
b. Hertz 在 2007 ~ 2008 年领先；Avis 增加且与现在 Hertz 类似；逐年下降
c. 截面数据的条形图；条形图的高度为 Hertz 290, Dollar 108, Avis 270。
16. 回答这个练习题依赖于更新的常规普通汽油每加仑平均价格的时间序列，如图 1-1 所示。
18. a. 67%
b. 612
c. 分类型
20. a. 43% 的经理认为自己在股票市场上操作极佳；21% 的经理认为医疗健康类股票极有可能是接下来 12 个月股票市场的主导板块。
b. 对于投资经理总体，12 个月的预期回报率为 11.2%。
c. 投资经理总体认为科技股和电信股大约需要 2.5 年才恢复上涨。
22. a. 总体由在北卡罗来纳州夏洛特的连锁店的所有消费者组成。
b. 食品连锁店用于搜集数据的一些方式为可以调查进入或离开商店的消费者；用邮寄给有会员卡的消费者的邮件来调查；可以在消费者结账时发一份调查问卷；将给予完成一份简单的在线调查的消费者赠券，使其在下次购物时有 5% 的折扣。
24. a. 正确
b. 不正确
c. 正确
d. 不正确
e. 不正确

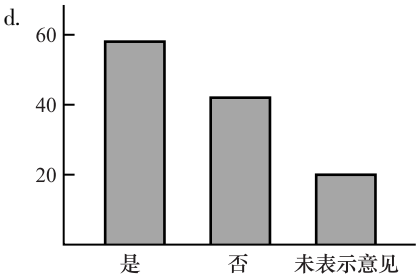
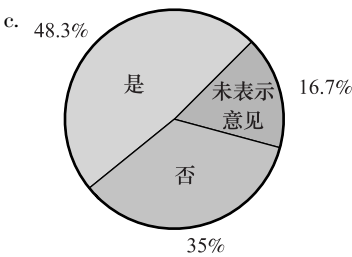
第 2 章

2. a. 0.20
b. 40
c/d.

组别	频数	百分数频数
A	44	22
B	36	18
C	80	40
D	40	20
合计	200	100

3. a. $360^\circ \times 58/120 = 174^\circ$

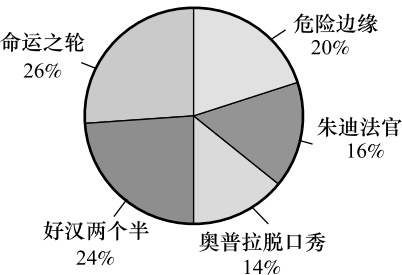
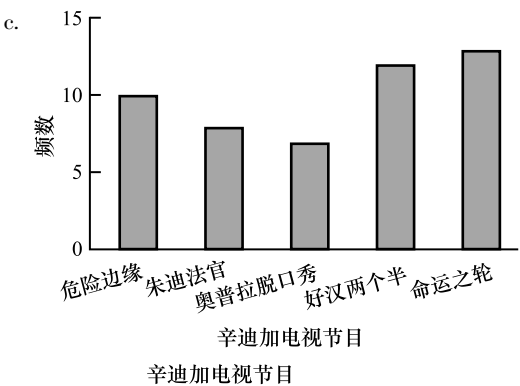
b. $360^\circ \times 42/120 = 126^\circ$



4. a. 这些数据分类型

b.

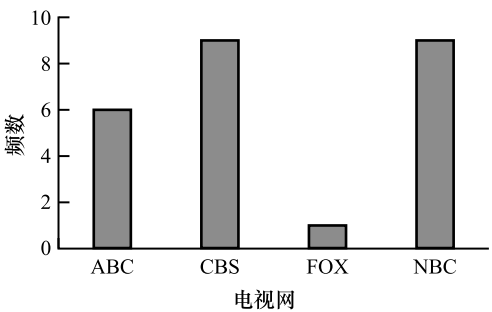
电视节目	频数	百分数频数
危险边缘	10	20
朱迪法官	8	16
奥普拉脱口秀	7	14
好汉两个半	12	24
命运之轮	13	26
合计	50	100



d. 《命运之论》拥有最多的电视观众，其次是《好汉两个半》。

6. a.

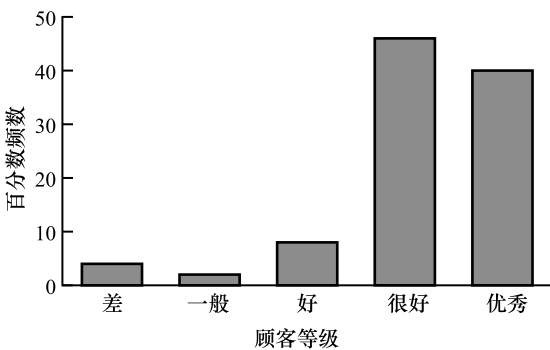
电视网	频数	百分数频数
ABC	6	24
CBS	9	36
FOX	1	4
NBC	9	36
合计	25	100



b. 根据这些数据，NBC 和 CBS 并列第一，在前 25 中各有 9（36%）个；ABC 有 6（24%）个居第三；比较年轻的电视网 FOX 有 1（4%）个。

7. a.

等级	频数	百分数频数
优秀	20	40
很好	23	46
好	4	8
一般	1	2
差	2	4
合计	50	100



管理者应该对这些结果满意，86% 的评级是很好、优秀。

b. 检查从差到一般这三个评级的解释，以识别低评级的原因。

8. a.

位置	频数	相对频数
投手 (P)	17	0.309
接手 (H)	4	0.073
一垒手 (1)	5	0.091
二垒手 (2)	4	0.073
三垒手 (3)	2	0.036
游击手 (S)	5	0.091

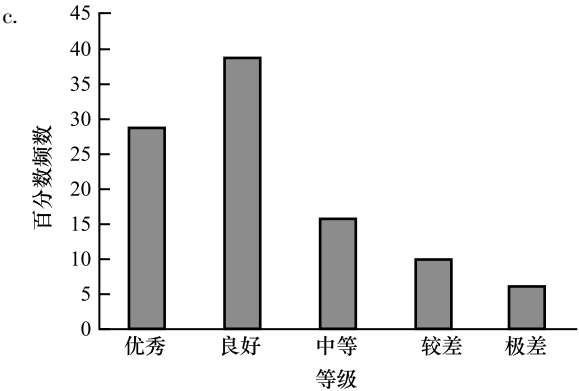
(续)

位置	频数	相对频数
左外场手 (L)	6	0.109
中外场手 (C)	5	0.091
右外场手 (R)	<u>7</u>	<u>0.127</u>
合计	55	1.000

- b. 投手 (P)
c. 三垒手 (3)
d. 右外场手 (R)
e. 内场人数是 16 人, 外场人数是 18 人。

10. a/b.

等级	频数	百分数频数
优秀	187	29
良好	252	39
中等	107	16
较差	62	10
极差	<u>41</u>	<u>6</u>
合计	649	100



- d. 在阿纳海姆喜来登酒店的客人中有 $29\% + 39\% = 68\%$ 给酒店的评级为优秀或良好。但是, 有 $10\% + 6\% = 16\%$ 的客人给酒店的评级为较差或极差。
e. 加利福尼亚州迪士尼大酒店的百分数频数分布如下: 优秀 (48%)、良好 (31%) 和中等 (12%)。加利福尼亚州迪士尼大酒店的客人中有 $48\% + 31\% = 79\%$ 给酒店的评级为优秀或良好。同时, 有 $6\% + 3\% = 9\%$ 的客人给酒店的评级为较差或极差。

12.

组	累积频数	累积相对频数
≤ 19	10	0.20
≤ 29	24	0.48
≤ 39	41	0.82
≤ 49	48	0.96
≤ 59	50	1.00

14. b/c.

组	频数	百分数频数
6.0 ~ 7.9	4	20
8.0 ~ 9.9	2	10
10.0 ~ 11.9	8	40
12.0 ~ 13.9	3	15
14.0 ~ 15.9	<u>3</u>	<u>15</u>
合计	20	100

15. 叶单位 = 0.1

6	3
7	5 5 7
8	1 3 4 8
9	3 6
10	0 4 5
11	3

16. 叶单位 = 10

11	6
12	0 2
13	0 6 7
14	2 2 7
15	5
16	0 2 8
17	0 2 3

17. a/b.

等待时间	频数	相对频数
0 ~ 4	4	0.20
5 ~ 9	8	0.40
10 ~ 14	5	0.25
15 ~ 19	2	0.10
20 ~ 24	<u>1</u>	<u>0.05</u>
合计	20	1.00

c/d.

等待时间	累积频数	累积相对频数
≤ 4	4	0.20
≤ 9	12	0.60
≤ 14	17	0.85
≤ 19	19	0.95
≤ 24	20	1.00

e. $12/20 = 0.60$

18. a.

PPG	频数
10 ~ 12	1
12 ~ 14	3
14 ~ 16	7
16 ~ 18	19
18 ~ 20	9
20 ~ 22	4
22 ~ 24	2
24 ~ 26	0
26 ~ 28	3
28 ~ 30	<u>2</u>
合计	50

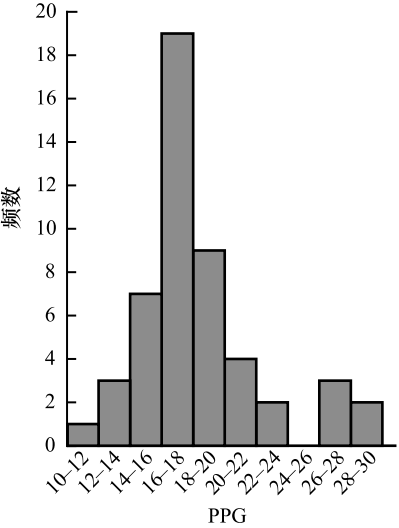
b.

PPG	相对频数
10 ~ 12	0.02
12 ~ 14	0.06
14 ~ 16	0.14
16 ~ 18	0.38
18 ~ 20	0.18
20 ~ 22	0.08
22 ~ 24	0.04
24 ~ 26	0.00
26 ~ 28	0.06
28 ~ 30	<u>0.04</u>
合计	1.00

c.

PPG	累积百分数频数
< 12	2
< 14	8
< 16	22
< 18	60
< 20	78
< 22	86
< 24	90
< 26	90
< 28	96
< 30	100

d. 呈现右偏



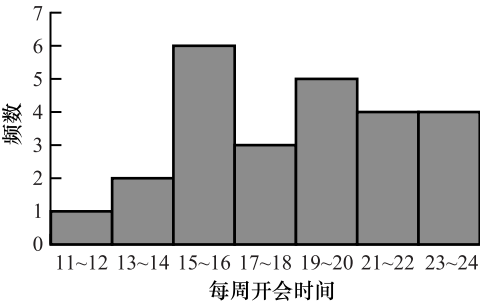
e. $(11/50) \times 100 = 22\%$

20. a. 最小值 12, 最大值 23

b.

每周开会时间	频数	百分数频数
11 ~ 12	1	4
13 ~ 14	2	8
15 ~ 16	6	24
17 ~ 18	3	12
19 ~ 20	5	20
21 ~ 22	4	16
23 ~ 24	<u>4</u>	<u>16</u>
合计	25	100

c.

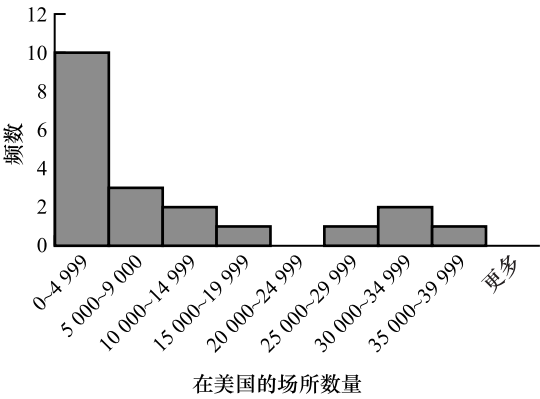


分布略微左偏

22. a.

在美国场所数量	频数	百分数频数
0 ~ 4 999	10	50
5 000 ~ 9 999	3	15
10 000 ~ 14 999	2	10
15 000 ~ 19 999	1	5
20 000 ~ 24 999	0	0
25 000 ~ 29 999	1	5
30 000 ~ 34 999	2	10
35 000 ~ 39 999	<u>1</u>	<u>5</u>
合计	20	100

b.



c. 分布右偏; 大多数特许加盟商 (50% + 15% + 15% = 80%) 场所数量小于 20 000; 麦当劳、Subway 和 7-

Eleven 有最多的场所数量。

24. 中位数起薪

4	6	8							
5	1	2	3	3	5	6	8	8	
6	0	1	1	1	2	2			
7	1	2	5						

职业生涯中期中位数收入

8	0	0	4						
9	3	3	5	6	7				
10	5	6	6						
11	0	1	4	4	4				
12	2	3	6						

职业生涯中期中位数收入比中位数起薪分布的范围要宽，同时正如期望的那样，职业生涯中期中位数收入比中位数起薪要高。职业生涯中期中位数收入大多在 93 000 ~ 114 000 美元，而中位数起薪大多在 51 000 ~ 62 000 美元。

26. a.

2	1	4							
2	6	7							
3	0	1	1	1	2	3			
3	5	6	7	7					
4	0	0	3	3	3	3	4	4	
4	6	6	7	9					
5	0	0	0	2	2				
5	5	6	7	9					
6	1	4							
6	6								
7	2								

b. 40 ~ 44 有 9 人

c. 43 有 5 人

27. a.

		y		合计
		1	2	
x	A	5	0	5
	B	11	2	13
	C	2	10	12
	合计	18	12	30

b.

		y		合计
		1	2	
x	A	100.0	0.0	100.0
	B	84.6	15.4	100.0
	C	16.7	83.3	100.0

c.

		y	
		1	2
x	A	27.8	0.0
	B	61.1	16.7
	C	11.1	83.3
	合计	100.0	100.0

d. A 总是在 $y = 1$ 时出现

B 最常在 $y = 1$ 时出现

C 最常在 $y = 2$ 时出现

28. a.

		y				合计
		20 ~ 39	40 ~ 59	60 ~ 79	80 ~ 100	
x	10 ~ 29			1	4	5
	30 ~ 49	2		4		6
	50 ~ 69	1	3	1		5
	70 ~ 90	4				4
	合计	7	3	6	4	20

b.

		y				合计
		20 ~ 39	40 ~ 59	60 ~ 79	80 ~ 100	
x	10 ~ 29			20.0	80.0	100
	30 ~ 49	33.3		66.7		100
	50 ~ 69	20.0	60.0	20.2		100
	70 ~ 90	100.0				100
	合计	100.0	100.0	100.0	100.0	100

c.

		y				合计
		20 ~ 39	40 ~ 59	60 ~ 79	80 ~ 100	
x	10 ~ 29	0.0	0.0	16.7	100.0	0.0
	30 ~ 49	28.6	0.0	66.7		0.0
	50 ~ 69	14.3	100.0	16.7		0.0
	70 ~ 90	57.1	0.0	0.0		0.0
	合计	100.0	100.0	100.0	100.0	100.0

d. x 较大值对应着 y 较小值，负相关

30. a.

		年份					总计
		1988 ~ 1992	1993 ~ 1997	1998 ~ 2002	2003 ~ 2007	2008 ~ 2012	
平均时速	130 ~ 139.9	16.7	0.0	0.0	33.3	50.0	100
	140 ~ 149.9	25.0	25.0	12.5	25.0	12.5	100
	150 ~ 159.9	0.0	50.0	16.7	16.7	16.7	100
	160 ~ 169.9	50.0	0.0	50.0	0.0	0.0	100
	170 ~ 179.9	0.0	0.0	100.0	0.0	0.0	100
	合计	100.0	100.0	100.0	100.0	100.0	100

b. 显示在 2003 年之前大多数较快的平均时速赢得时间。这可能是由于考虑驾驶员的安全、观众的安全、对环境的影响以及比赛期间的燃料消耗等新的规则出现。

e. 家庭收入在 100 000 美元及以上的家庭中，来自南部的百分数为 32.25%。来自南部的家庭中，家庭收入在 100 000 美元及以上的百分数为 17.57%。两个百分数是不同的，因为计算百分数的基础不同，第一个（行百分数）的基础是收入水平，第二个（列百分数）的基础是地区。

34. a.

行业	品牌收益 (10 亿美元)						总计
	0 ~ 25	25 ~ 50	50 ~ 75	75 ~ 100	100 ~ 125	125 ~ 150	
汽车和奢侈品	10	1	1		1	2	15
包装消费品	12						12
金融服务	2	4	2	2	2	2	14
其他	13	5	3	2	2	1	26
科技	4	4	4	1	2		15
总计	41	14	10	5	7	5	82

b.

品牌收益 (10 亿美元)	频数
0 ~ 25	41
25 ~ 50	14
50 ~ 75	10
75 ~ 100	5
100 ~ 125	7
125 ~ 150	5
总计	82

c. 包装消费品有最低的品牌收益，样本数据中 12 个包装消费品品牌的品牌收益都低于 250 亿美元。大约 57%（14 个中的 8 个）的金融服务品牌的品牌收益大于或等于 500 亿美元，47%（15 个中的 7 个）的科技品牌的品牌收益至少是 500 亿美元。

d.

行业	1 年间价值的变化率 (%)						总计
	-60 ~ -41	-40 ~ -21	-20 ~ -1	0 ~ 19	20 ~ 39	40 ~ 60	
汽车和奢侈品				11	4		15
包装消费品			2	10			12
金融服务		1	6	7			14
其他			2	20	4		26
科技	1	3	4	4	2	1	15
总计	1	4	14	52	10	1	82

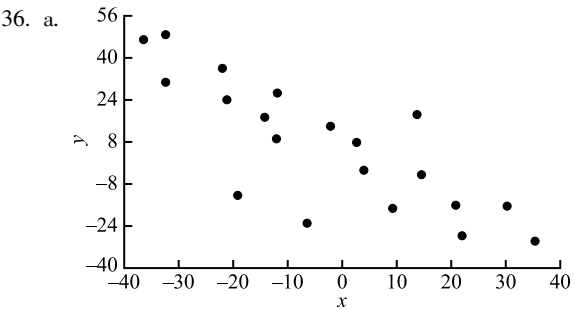
e.

1 年间价值的变化率 (%)	频数
-60 ~ -41	1
-40 ~ -21	4
-20 ~ -1	14

(续)

1 年间价值的变化率 (%)	频数
0 ~ 19	52
20 ~ 39	10
40 ~ 60	1
总计	82

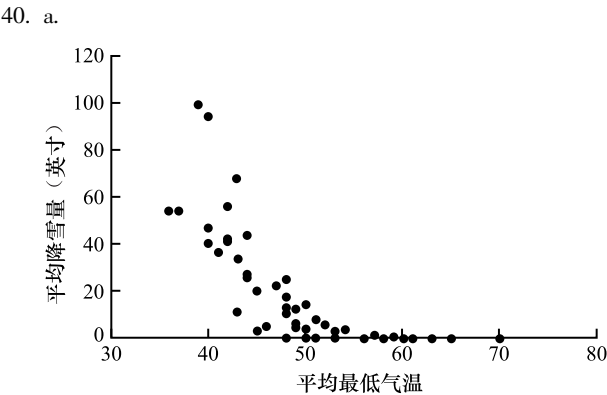
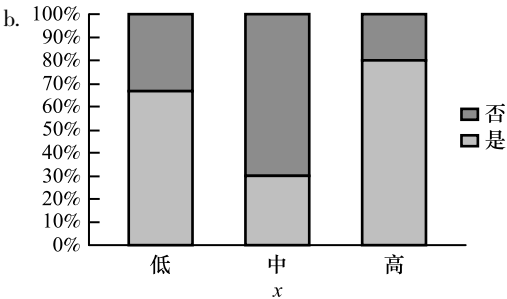
f. 所有汽车和奢侈品品牌与 1 年间价值的变化率正相关。科技品牌的变异最大。



b. x 和 y 之间负相关，y 随着 x 增大而减小。

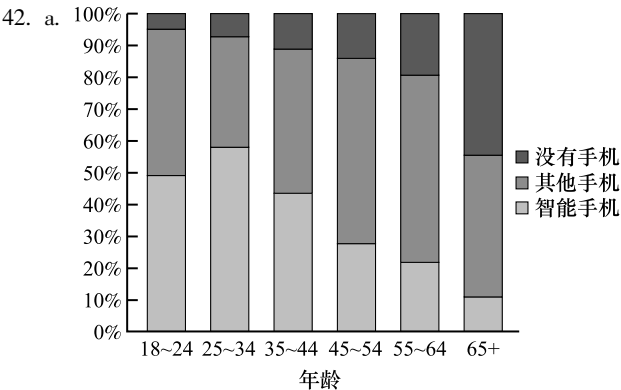
38. a.

	是	否	总计
低	66.667	33.333	100
中	30.000	70.000	100
高	80.000	20.000	100



b. 较低的平均最低气温似乎导致较高的降雪量。

c. 纽约州的布法罗和罗彻斯特这两个城市的平均降雪量接近 100 英寸，这两个城市都在纽约州的大湖附近。



- b. 智能手机拥有者在 25 ~ 34 岁增加后, 随着年龄的增加而减少; 没有手机的比例随着年龄的增加而增加; 拥有其他手机的比例在整个年龄组很少变化。
- c. 除非新设备取代智能手机, 我们期待智能手机拥有者随着年龄将变得不敏感; 由于当前拥有者将变老, 以及设备将变为必需品而非奢侈品, 这应该是真的。

44. a.

SAT 分数	频数
800 ~ 999	1
1 000 ~ 1 199	3
1 200 ~ 1 399	6
1 400 ~ 1 599	10
1 600 ~ 1 799	7
1 800 ~ 1 999	2
2 000 ~ 2 199	1
合计	30

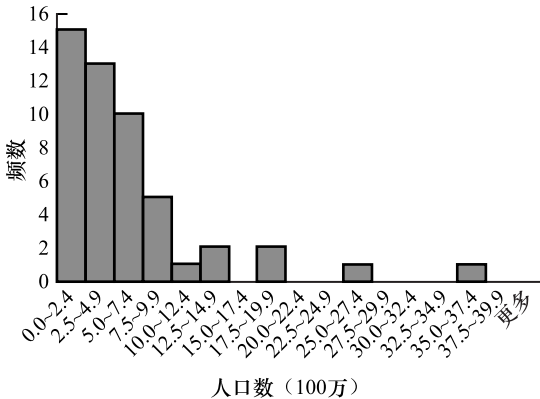
- b. 几乎对称
- c. 33% 的分数位于 1 400 ~ 1 599; 低于 800 和高于 2 200 的分数是异常情况; 平均分接近或略微超过 1 500。

46. a.

人口数 (100 万)	频数	相对频数
0.0 ~ 2.4	15	30.0
2.5 ~ 4.9	13	26.0
5.0 ~ 7.4	10	20.0
7.5 ~ 9.9	5	10.0
10.0 ~ 12.4	1	2.0
12.5 ~ 14.9	2	4.0
15.0 ~ 17.4	0	0.0
17.5 ~ 19.9	2	4.0
20.0 ~ 22.4	0	0.0
22.5 ~ 24.9	0	0.0
25.0 ~ 27.4	1	2.0
27.5 ~ 29.9	0	0.0
30.0 ~ 32.4	0	0.0

(续)

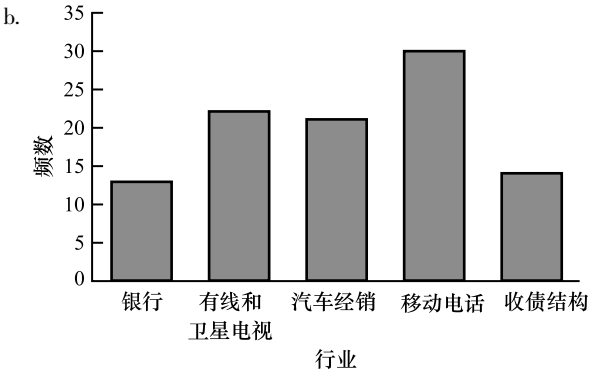
人口数 (100 万)	频数	相对频数
32.5 ~ 34.9	0	0.0
35.0 ~ 37.4	1	2.0
37.5 ~ 39.9	0	0.0
更多	0	0.0



- b. 严重右偏态
- c. 15 个州 (30%) 的人口数低于 250 万, 超过一半州 (28 个州, 56%) 的人口数低于 500 万, 只有 7 个州 (加利福尼亚州、佛罗里达州、伊利诺伊州、纽约州、俄亥俄州、宾夕法尼亚州和得克萨斯州) 的人口数高于 1 000 万, 人口最多的州是加利福尼亚州 (3 730 万), 人口最少的州是佛蒙特州和怀俄明州 (60 万)。

48. a.

行业	频数	百分数频数
银行	26	13
有线和卫星电视	44	22
汽车经销	42	21
移动电话	60	30
收债机构	28	14
合计	200	100



- c. 移动电话提供商是投诉最多的公司。
- d. 百分数频数分布显示两个金融行业 (银行和收债机构) 的投诉次数几乎相同; 汽车经销及有线和卫星电视的投诉次数几乎相同。

50. a.

教育水平	百分数频数
高中毕业	$(32\,773/65\,644) \times 100 = 49.93$
学士学位	$(22\,131/65\,644) \times 100 = 33.71$
硕士学位	$(9\,003/65\,644) \times 100 = 13.71$
博士学位	$(1\,737/65\,644) \times 100 = 2.65$
总计	100.00

家长具有硕士或博士学位的百分数是 $13.71 + 2.65 = 16.36\%$ 。

b.

家庭收入 (美元)	百分数频数
25 000 以下	$(13\,128/65\,644) \times 100 = 20.00$
25 000 ~ 49 999	$(15\,499/65\,644) \times 100 = 23.61$
50 000 ~ 99 999	$(20\,548/65\,644) \times 100 = 31.30$
100 000 及以上	$(16\,469/65\,644) \times 100 = 25.09$
总计	100.00

家庭收入在 50 000 美元及以上的百分数是 $31.30 + 25.09 = 56.39\%$ 。

c.

教育水平	家庭收入 (美元)			
	25 000 以下	25 000 ~ 49 999	50 000 ~ 99 999	100 000 及以上
高中毕业	75.26	64.33	45.95	21.14
学士学位	18.92	26.87	37.31	47.46
硕士学位	5.22	7.77	14.69	24.86
博士学位	0.60	1.03	2.05	6.53
总计	100.00	100.00	100.00	100.00

家庭收入在 25 000 美元以下和家庭收入在 100 000 美元及以上的教育水平有很大的不同。

52. a.

就业增长率 (%)	规模			总计
	小型	中型	大型	
- 10 ~ 0	4	6	2	12
0 ~ 10	18	13	29	60
10 ~ 20	7	2	4	13
20 ~ 30	3	3	2	8
30 ~ 40	0	3	1	4
60 ~ 70	0	1	0	1
总计	32	28	38	98

b. 就业增长率的频数分布:

就业增长率 (%)	频数
- 10 ~ 0	12
0 ~ 10	60
10 ~ 20	13
20 ~ 30	8

(续)

就业增长率 (%)	频数
30 ~ 40	4
60 ~ 70	1
总计	98

规模的频数分布:

规模	频数
小型	32
中型	28
大型	38
总计	98

c. 列百分数的交叉分组表:

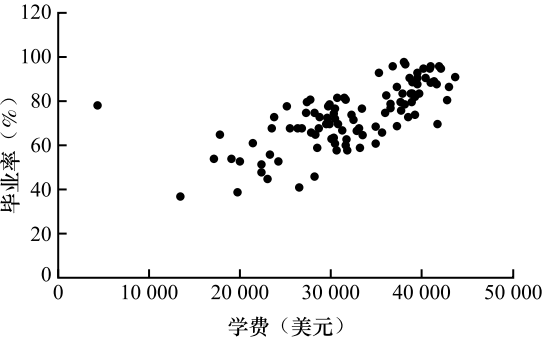
就业增长率 (%)	规模		
	小型	中型	大型
- 10 ~ 0	13	21	5
0 ~ 10	56	46	76
10 ~ 20	22	7	11
20 ~ 30	9	11	5
30 ~ 40	0	11	3
60 ~ 70	0	4	0
总计	100	100	100

d. 行百分数的交叉分组表:

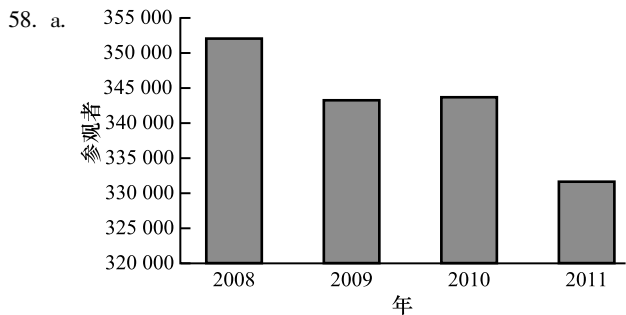
就业增长率 (%)	规模			总计
	小型	中型	大型	
- 10 ~ 0	33	50	17	100
0 ~ 10	30	22	48	100
10 ~ 20	54	15	31	100
20 ~ 30	38	38	25	100
30 ~ 40	0	75	25	100
60 ~ 70	0	100	0	100

54. c. 较古老的学院和学校有较高的毕业率趋势。

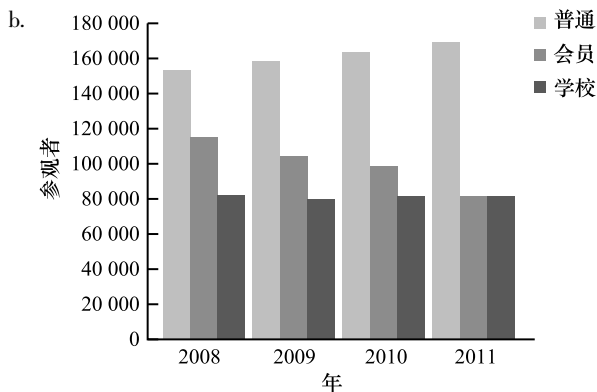
56. a.



b. 这里显示学费和毕业率之间强的正相关。



动物园参观者随着时间呈现下降趋势。



c. 普通参观者有增加的趋势,但不足以弥补会员参观者的减少,学校全体会员呈现平稳。

第3章

2. 16, 16.5

4.

时期	回报率 (%)
1	-0.060
2	-0.080
3	-0.040
4	0.020
5	0.054

过去5个时期的平均增长因子为:

$$\begin{aligned}\bar{x}_g &= \sqrt[n]{x_1 x_2 \cdots x_5} \\ &= \sqrt[5]{0.940 \times 0.920 \times 0.960 \times 1.020 \times 1.054} \\ &= \sqrt[5]{0.8925} = 0.9775\end{aligned}$$

于是,年平均增长率为 $(0.9775 - 1) \times 100\% = -2.25\%$ 。

5. 15, 20, 25, 25, 27, 28, 30, 34

$$L_{20} = \frac{p}{100} (n+1) = \frac{20}{100} (8+1) = 1.8$$

第20百分位数 $= 15 + 0.8 \times (20 - 15) = 19$

$$L_{25} = \frac{p}{100} (n+1) = \frac{25}{100} (8+1) = 2.25$$

第25百分位数 $= 20 + 0.25 \times (25 - 20) = 21.25$

$$L_{65} = \frac{p}{100} (n+1) = \frac{65}{100} (8+1) = 5.85$$

第65百分位数 $= 27 + 0.85 \times (28 - 27) = 27.85$

$$L_{75} = \frac{p}{100} (n+1) = \frac{75}{100} (8+1) = 6.75$$

第75百分位数 $= 28 + 0.75 \times (30 - 28) = 29.5$

6. 59.73, 57, 53

8. a. 中位数 = 80 或 80 000 美元,略低于《华尔街日报》报道的中位数。

b. $\bar{x} = 84$, 平均收入为 84 000 美元。15 名中层管理人员样本的样本平均收入高于中位数收入。这显示在亚特兰大市工作的中层管理人员的收入分布是右偏的。

c. 数据排序如下:

53 55 63 67 73 75 77 80
83 85 93 106 108 118 124

$$L_{25} = \frac{p}{100} (n+1) = \frac{25}{100} \times 16 = 4$$

第一四分位数或第25百分位数是第4个位置的数值或67。

$$L_{75} = \frac{p}{100} (n+1) = \frac{75}{100} \times 16 = 12$$

第三四分位数或第75百分位数是第12个位置的数值或106。

10. a. $\bar{x} = 65.9$

中位数或第50百分位数 $= 66 + 0.5 \times (67 - 66) = 66.5$

众数是61。

$$b. L_{25} = \frac{p}{100} (n+1) = \frac{25}{100} (20+1) = 5.25$$

第一四分位数或第25百分位数 = 61

$$L_{75} = \frac{p}{100} (n+1) = \frac{75}{100} (20+1) = 15.75$$

第三四分位数或第75百分位数 = 71

$$c. L_{90} = \frac{p}{100} (n+1) = \frac{90}{100} (20+1) = 18.9$$

第90百分位数 $= 78 + 0.9 \times (81 - 78) = 80.7$

90%的评级是80.7或更低;10%的评级是80.7或更高。

12. a. 收视人数的最小值为13 300万人,最大值为16 500万人。

b. 收视人数的平均数为15 040万人或近似15 000万人;中位数为15 000万人;数据是多峰的(13 600万人,14 000万人,16 100万人和16 200万人);在这种情形下,众数通常是不被报告的。

c. 首先数据按升序排列;

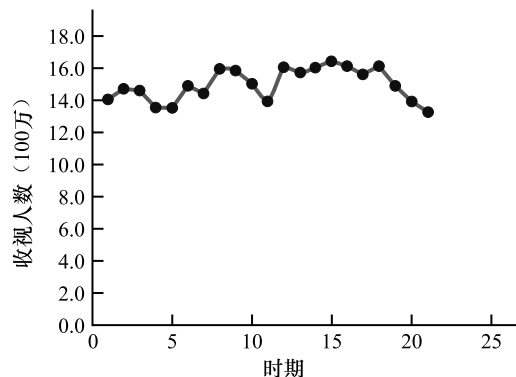
$$L_{25} = \frac{p}{100} (n+1) = \frac{25}{100} (21+1) = 5.50$$

第一四分位数或第 25 百分位数 $= 14 + 0.5 \times (14.1 - 14) = 14.05$

$$L_{75} = \frac{p}{100} (n+1) = \frac{75}{100} (21+1) = 16.5$$

第三四分位数或第 75 百分位数 $= 16 + 0.5 \times (16.1 - 16) = 16.05$

- d. 显示播出日期收视人数数据的图形如下：时期 1 对应于播出季的第 1 集，时期 2 对应于播出季的第 2 集，以此类推。



图形显示在 2011 ~ 2012 播出季《生活大爆炸》的收视人数相对平稳。

14. 对 2011 年 3 月

$$L_{25} = \frac{p}{100} (n+1) = \frac{25}{100} (50+1) = 12.75$$

第一四分位数或第 25 百分位数 $= 6.8 + 0.75 \times (6.8 - 6.8) = 6.8$

$$L_{50} = \frac{p}{100} (n+1) = \frac{50}{100} (50+1) = 25.5$$

第二四分位数或中位数 $= 8 + 0.5 \times (8 - 8) = 8$

$$L_{75} = \frac{p}{100} (n+1) = \frac{75}{100} (50+1) = 38.25$$

第三四分位数或第 75 百分位数 $= 9.4 + 0.25 \times (9.6 - 9.4) = 9.45$

对 2012 年 3 月

$$L_{25} = \frac{p}{100} (n+1) = \frac{25}{100} (50+1) = 12.75$$

第一四分位数或第 25 百分位数 $= 6.2 + 0.75 \times (6.2 - 6.2) = 6.2$

$$L_{50} = \frac{p}{100} (n+1) = \frac{50}{100} (50+1) = 25.5$$

第二四分位数或中位数 $= 7.5 + 0.5 \times (7.4 - 7.3) = 7.35$

$$L_{75} = \frac{p}{100} (n+1) = \frac{75}{100} (50+1) = 38.25$$

第三四分位数或第 75 百分位数 $= 8.6 + 0.25 \times (8.6 - 8.6) = 8.6$

如果我们将它们放在一张表格中，就很容易比较。

	2011 年 3 月	2012 年 3 月
第一四分位数	6.80	6.20
中位数	8.00	7.35
第三四分位数	9.45	8.60

2012 年 3 月的结果表明：大约有 25% 的州的失业率小于或等于 6.2%，低于 2011 年 3 月；但是，2012 年 3 月的中位数 7.35% 和第三四分位数 8.6% 都低于 2011 年 3 月相应的数值，表明整个国家的失业率在下降。

16. a.

等级分 x_i	权重 w_i
4 (A)	9
3 (B)	15
2 (C)	33
1 (D)	3
0 (F)	0
	60 学时

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{9 \times 4 + 15 \times 3 + 33 \times 2 + 3 \times 1}{9 + 15 + 33 + 3} = \frac{150}{60} = 2.5$$

- b. 能

18. 3.8, 3.7

- 20.

年	Stivers		Trippi	
	年末价值 (美元)	增长 因子	年末价值 (美元)	增长 因子
2004	11 000	1.100	5 600	1.120
2005	12 000	1.091	6 300	1.125
2006	13 000	1.083	6 900	1.095
2007	14 000	1.077	7 600	1.101
2008	15 000	1.071	8 500	1.118
2009	16 000	1.067	9 200	1.082
2010	17 000	1.063	9 900	1.076
2011	18 000	1.059	10 600	1.071

对于 Stivers 共同基金，我们有

$18\,000 = 10\,000 \times [x_1 x_2 \cdots x_8]$ ，于是 $[x_1 x_2 \cdots x_8] = 1.8$ ，

$$\bar{x}_g = \sqrt[n]{x_1 x_2 \cdots x_8} = \sqrt[8]{1.80} = 1.076\,24$$

于是，Stivers 共同基金的年平均回报率为 $(1.076\,24 - 1) \times 100 = 7.624\%$ 。

对于 Trippi 共同基金，我们有

$10\,060 = 5\,000 \times [x_1 x_2 \cdots x_8]$ ，于是 $[x_1 x_2 \cdots x_8] = 2.12$ ，

$$\bar{x}_g = \sqrt[n]{x_1 x_2 \cdots x_8} = \sqrt[8]{2.12} = 1.098\,48$$

于是，Trippi 共同基金的年平均回报率为 $(1.098\,484 - 1) \times 100 = 9.848\%$ 。

Stivers 共同基金得到很好的年平均回报率 7.6%，Trippi 共同基金获得更优异的年平均回报率 9.8%。

22. $25\,000\,000 = 10\,000\,000 \times [x_1 x_2 \cdots x_6]$, 于是 $[x_1 x_2 \cdots x_6] = 2.50$,

$$\bar{x}_g = \sqrt[n]{x_1 x_2 \cdots x_6} = \sqrt[6]{2.50} = 1.165$$

于是, 年平均增长率为 $(1.165 - 1) \times 100 = 16.5\%$ 。

24. 16, 4

25. 极差 = $34 - 15 = 19$

将数据排序为: 15, 20, 25, 25, 27, 28, 30, 34

$$L_{25} = \frac{p}{100} (n+1) = \frac{25}{100} (8+1) = 2.25$$

第一四分位数或第 25 百分位数 = $20 + 0.25 \times (20 - 15) = 21.25$

$$L_{75} = \frac{p}{100} (n+1) = \frac{75}{100} (8+1) = 6.75$$

第三四分位数或第 75 百分位数 = $28 + 0.75 \times (30 - 28) = 29.5$

$$IQR = Q_3 - Q_1 = 29.5 - 21.25 = 8.25$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{204}{8} = 25.5$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{242}{8-1} = 34.57$$

$$s = \sqrt{34.57} = 5.88$$

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
27	1.5	2.25
25	-0.5	0.25
20	-5.5	30.25
15	-10.5	110.25
30	4.5	20.25
34	8.5	72.25
28	2.5	6.25
25	-0.5	0.25
		242.00

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{242}{8-1} = 34.57$$

$$s = \sqrt{34.57} = 5.88$$

26. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{74.4}{20} = 3.72$

$$b. s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1.6515}{20-1}} = \sqrt{0.0869} = 0.2948$$

- c. 旧金山每加仑无铅汽油的平均价格比全美的平均值要高。

28. a. 发球速度的平均数为 180.95, 方差为 21.42, 标准差为 4.63。

- b. 尽管 2011 年温布尔登锦标赛发球速度前 20 名的女子单打选手的平均数稍高, 但差异很小。此外, 由 2012 年澳大利亚公开赛和 2011 年温布尔登锦标赛发球速度前 20 名的女子单打选手的方差来看, 这些差异很可能源于选手表现的随机性差异。

30. Dawson: 极差 = 2, $s = 0.67$

Clark: 极差 = 8, $s = 2.58$

32. a. 汽车: 1 960.05

百货业公司: 692.85

- b. 汽车: 481.65

百货业公司: 155.06

- c. 汽车: $2\,901 - 598 = 2\,303$

百货业公司: $1\,011 - 448 = 563$

- d. 汽车: 第一四分位数或第 25 百分位数 = $1\,714 + 0.25 \times (1\,720 - 1\,714) = 1\,715.5$

百货业公司: 第一四分位数或第 25 百分位数 = $589 + 0.25 \times (597 - 589) = 591$

汽车: 第三四分位数或第 75 百分位数 = $2\,202 + 0.75 \times (2\,254 - 2\,202) = 2\,241$

百货业公司: 第三四分位数或第 75 百分位数 = $782 + 0.75 \times (824 - 782) = 813.5$

- e. 就平均数而言, 汽车业花费更多, 与百货业公司相比有较大的标准差、较大的最大值和最小值和较大的极差。汽车业总有新车年, 可能在广告上花费较多。

34. 1/4 英里: $s = 0.0564$, 标准差系数 = 5.8%

1 英里: $s = 0.1395$, 标准差系数 = 2.9%

36. 0.20, 1.50, 0, -0.50, -2.20

37. a. $z = \frac{20-30}{5} = -2$, $z = \frac{40-30}{5} = 2$; $1 - \frac{1}{2^2} = 0.75$

至少 75%

- b. $z = \frac{15-30}{5} = -3$, $z = \frac{45-30}{5} = 3$; $1 - \frac{1}{3^2} = 0.89$

至少 89%

- c. $z = \frac{22-30}{5} = -1.6$, $z = \frac{38-30}{5} = 1.6$; $1 - \frac{1}{1.6^2} = 0.61$

至少 61%

- d. $z = \frac{18-30}{5} = -2.4$, $z = \frac{42-30}{5} = 2.4$; $1 - \frac{1}{2.4^2} = 0.83$

至少 83%

- e. $z = \frac{12-30}{5} = -3.6$, $z = \frac{48-30}{5} = 3.6$; $1 - \frac{1}{3.6^2} = 0.92$

至少 92%

38. a. 95%

- b. 几乎所有

- c. 68%

39. a. $z = 2$ 标准差

$$1 - \frac{1}{z^2} = 1 - \frac{1}{2^2} = \frac{3}{4}; \text{至少 } 75\%$$

- b. $z = 2.5$ 标准差

$$1 - \frac{1}{z^2} = 1 - \frac{1}{2.5^2} = 0.84; \text{至少 } 84\%$$

c. $z = 2$ 标准差, 经验法则: 95%

40. a. 68%

b. 81.5%

c. 2.5%

42. a. -0.67

b. 1.50

c. 都不是异常值

d. 是; $z = 8.25$

44. a. 76.5, 7

b. 16%, 2.5%

c. 12.2, 7.89; 没有异常值

46. 15, 22.5, 26, 29, 34

48. 5, 6, 8, 10, 10, 12, 15, 16, 18

最小值 = 5

$$L_{25} = \frac{p}{100} (n+1) = \frac{25}{100} (9+1) = 2.5$$

第一四分位数或第 25 百分位数 = $6 + 0.5 \times (8 - 6) = 7$

$$L_{50} = \frac{p}{100} (n+1) = \frac{50}{100} (9+1) = 5.0$$

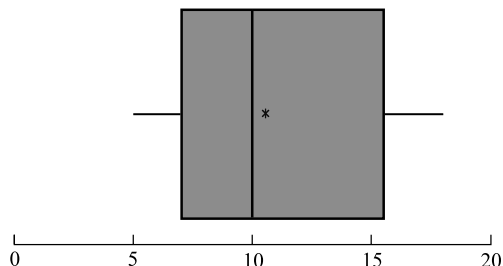
第二四分位数或中位数 = 10

$$L_{75} = \frac{p}{100} (n+1) = \frac{75}{100} (9+1) = 7.5$$

第三四分位数或第 75 百分位数 = $15 + 0.5 \times (16 - 15) = 15.5$

最大值 = 18

用 StatTools 绘制的箱形图如下:



50. a. 男子第一名比女子第一名快 43.73 分。

b. 中位数: 109.64, 131.67

男子中位数比女子中位数快 22.03 分。

c. 65.30, 83.1025, 109.64, 129.025, 148.70

109.03, 122.08, 131, 67, 147.18, 189.28

d. 男子的界限为 25.35 ~ 190.23; 没有异常值。

女子的界限为 84.43 ~ 184.83; 有 2 个异常值。

e. 女子运动员的变异较小。

51. a. 最小值 = 608

$$L_{25} = \frac{p}{100} (n+1) = \frac{25}{100} (21+1) = 5.5$$

第一四分位数或第 25 百分位数 = $1850 + 0.5 \times (1872 - 1850) = 1861$

$$L_{50} = \frac{p}{100} (n+1) = \frac{50}{100} (21+1) = 11.0$$

第二四分位数或中位数 = 4019

$$L_{75} = \frac{p}{100} (n+1) = \frac{75}{100} (21+1) = 16.5$$

第三四分位数或第 75 百分位数 = $8305 + 0.5 \times (8408 - 8305) = 8356.5$

最大值 = 14138

五数概括: 608, 1861, 4019, 8356.5, 14138

b. 界限:

$$IQR = Q_3 - Q_1 = 8356.5 - 1861 = 6495.5$$

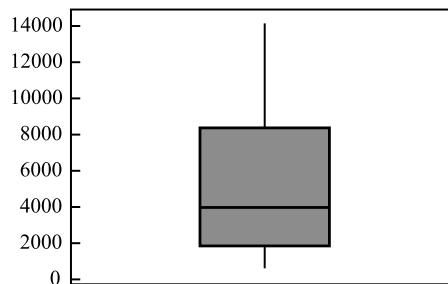
$$\text{上限: } Q_1 - 1.5 IQR = 1861 - 1.5 \times 6495.5 = -7882.25$$

$$\text{下限: } Q_3 + 1.5 IQR = 8356.5 + 1.5 \times 6495.5 = 18099.75$$

c. 无异常值, 数据均在上下限之间。

d. 是的。如果强生公司销售额的前两位数字互换了位置变为 41138, 这个销售额将上升为一个异常值; 数据值应该被复查和更正。

e. 用 Minitab 绘制的箱形图如下:



52. a. 73.5

b. 68, 71.25, 73.5, 74.75, 77

c. 界限为 66 ~ 80; 没有异常值。

d. 66, 68, 71, 73, 75; 60.5 ~ 80.5

63, 65, 66, 67.75, 69; 60.875 ~ 71.875

75, 77, 78.5, 79.75, 81; 72.875 ~ 83.875

每个移动电话服务公司都没有异常值。

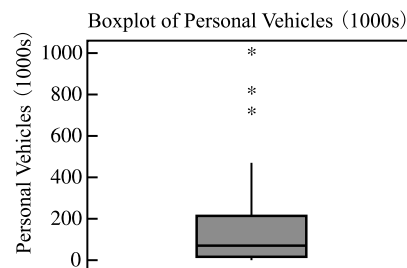
e. Verizon 得分最高, Sprint 得分最低。

54. a. 平均数 = 173.24, 中位数 (第二四分位数) = 89.5

b. 第一四分位数 = 38.5, 第三四分位数 = 232

c. 21, 38.5, 89.5, 232, 995

d. 用 Minitab 绘制的箱形图如下:



有三个异常值:

NY: Buffalo – Niagra Falls	707
TX: EI Paso	807
CA: San Ysidro	995

55. b. x 与 y 之间呈现负线性相关关系。

c.

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
4	50	-4	4	-16
6	50	-2	4	-8
11	40	3	-6	-18
3	60	-5	14	-70
16	30	8	-16	-128
40	230	0	0	-240
$\bar{x} = 8; \bar{y} = 46$				
$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1} = \frac{-240}{4} = -60$				

样本协方差显示在 x 和 y 之间存在负的线性关系。

$$d. r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-60}{5.43 \times 11.40} = -0.969$$

样本相关系数 -0.969 表明有很强的负线性关系。

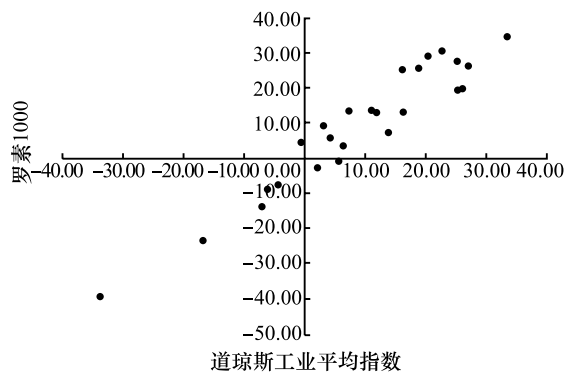
56. b. x 与 y 之间存在一个正的线性关系。

$$c. s_{xy} = 26.5$$

$$d. r_{xy} = 0.693$$

58. -0.91 ; 负相关

60. a. 道琼斯工业平均指数与罗素1000的回报率(%)



$$b. \text{道琼斯工业平均指数: } \bar{x} = \frac{\sum x_i}{n} = \frac{227.57}{25} = 9.10$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{5672.61}{24}} = 15.37$$

$$\text{罗素 1000: } \bar{x} = \frac{\sum x_i}{n} = \frac{227.29}{25} = 9.09$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{7679.81}{24}} = 17.89$$

$$c. r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{263.6110}{15.37 \times 17.89} = 0.959$$

d. 根据这个样本, 两个指数非常相似, 它们有很强的正线性关系。罗素 1000 的方差比道琼斯工业平均指数

的方差略微大一些。

62. a. 平均数为 2.95, 中位数为 3.0。

b. $L_{25} = 5.25$, 第一四分位数 = 1

$L_{75} = 15.75$, 第三四分位数 = 4.75

c. 极差为 7, 四分位数间距为 $4.75 - 1 = 3.75$ 。

d. 方差为 4.37, 标准差为 2.09。

e. 由于大多数人每周外出就餐时间相对较少, 少数家庭外出就餐较频繁, 我们将期望数据是正偏态; 偏度 0.34 表明数据有些右偏。

f. 下限为 -4.625 , 上限为 10.375 ; 数据中没有数值小于下限或大于上限, 于是没有异常值。

64. a. 对于有等待追踪系统的诊所, 病人候诊时间的平均数和中位数分别为 17.2 和 13.5; 对于没有等待追踪系统的诊所, 病人候诊时间的平均数和中位数分别为 29.1 和 23.5。

b. 对于有等待追踪系统的诊所, 病人候诊时间的方差和标准差分别为 86.2 和 9.3; 对于没有等待追踪系统的诊所, 病人候诊时间的方差和标准差分别为 275.7 和 16.6。

c. 有等待追踪系统的诊所候诊时间比没有等待追踪系统的诊所候诊时间大幅度缩短。

$$d. z = \frac{37 - 29.1}{16.6} = 0.48$$

$$e. z = \frac{37 - 17.2}{9.3} = 2.13$$

正的 z -分数显示, 两个病人的候诊时间都大于它们的样本平均数; 即使病人有相同的候诊时间, 有等待追踪系统的诊所样本中第 6 位病人的 z -分数较大, 因为这个样本有较小的平均数和较小的标准差。

f. 所有病人的 z -分数如下:

没有等待追踪系统	有等待追踪系统
-0.31	1.49
2.28	-0.67
-0.73	-0.34
-0.55	0.09
0.11	-0.56
0.90	2.13
-1.03	-0.88
-0.37	-0.45
-0.79	-0.56
0.48	-0.24

z -分数显示两个样本中没有异常值的存在。

$$66. a. \bar{x} = \frac{\sum x_i}{n} = \frac{20665}{50} = 413.3$$

比研究中的平均值略微高一些

- b. $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{69\,424.5}{49}} = 37.64$
 c. 没有异常值
68. a. 中位数 = $52.1 + 0.5 \times (52.1 - 52.1) = 52.1$
 b. -6.1%
 c. 第 25 百分位数 = 50.75, 第 75 百分位数 = 52.6
 d. 46.4, 50.75, 52.1, 52.6, 64.5
 e. 最大的家庭收入 64.5 的 z -分数 > 3 , 它是异常值。
 用四分位数间距方法, 最小观察值 46.5 和最大观测值 64.5 都被认定为异常值。
70. a. 364 间
 b. 457 美元
 c. -0.293 ; 较弱的负相关
 每晚较高的房价与较小规模的饭店相关联。
72. a. 0.268, 低或弱正相关
 b. 非常差的预示, 春训是惯例, 但不能计入常规赛和季后赛。
74. a. 60.68
 b. $s^2 = 31.23$, $s = 5.59$

第 4 章

2. $\binom{6}{3} = \frac{6!}{3!3!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1)(3 \times 2 \times 1)} = 20$

ABC	ACE	BCD	BEF
ABD	ACF	BCE	CDE
ABE	ADE	BCF	CDF
ABF	ADF	BDE	CEF
ACD	AEF	BDF	DEF

4. b. (H, H, H), (H, H, T), (H, T, H), (H, T, T),
 (T, H, H), (T, H, T), (T, T, H), (T, T, T)
 c. $1/8$
6. $P(E_1) = 0.40$, $P(E_2) = 0.26$, $P(E_3) = 0.34$
 采用相对频数方法
8. a. 计划委员会支持——市政委员会批准
 计划委员会支持——市政委员会否决
 计划委员会反对——市政委员会批准
 计划委员会反对——市政委员会否决
9. $\binom{50}{4} = \frac{50!}{4!46!} = \frac{50 \times 49 \times 48 \times 47}{4 \times 3 \times 2 \times 1} = 230\,300$
10. a. $P(\text{准点抵达}) = 0.865$
 b. $P(\text{低于 2 件}) = 3/10 = 0.30$
 c. $P(\text{多于 1 件}) = 5/10 = 0.50$
 d. $P(\text{没有准点抵达}) = 1 - P(\text{准点}) = 1 - 0.871 = 0.129$

12. a. 步骤 1: 利用计数的组合法则

$$\binom{59}{5} = \frac{59!}{5!(59-5)!} = \frac{59 \times 58 \times 57 \times 56 \times 55}{5 \times 4 \times 3 \times 2 \times 1} = 5\,006\,386$$

步骤 2: 从标有数字 1 ~ 35 的红色强力球中抽取一个红球的方法有 35 种。

强力球彩票结果的总数目为 $5\,006\,386 \times 35 = 175\,223\,510$ 。

- b. 彩票中奖的概率为 $175\,223\,510$ 分之一, 即 $1/175\,223\,510 = 0.000\,000\,057\,07$ 。
14. a. $1/4$
 b. $1/2$
 c. $3/4$
15. a. $S = \{\text{方块 A, 草花 A, 红桃 A, 黑桃 A}\}$
 b. $S = \{\text{方块 2, 方块 3, } \dots, \text{方块 10, 方块 J, 方块 Q, 方块 K, 方块 A}\}$
 c. 12 张; 4 种花色, 每种花色中的 J、Q 和 K
 d. (a) $4/52 = 1/13 = 0.08$
 (b) $13/52 = 1/4 = 0.25$
 (c) $12/52 = 0.23$

16. a. 36
 c. $1/6$
 d. $5/18$
 e. 不同意; $P(\text{奇数}) = P(\text{偶数}) = 1/2$
 f. 古典法

17. a. (4, 6), (4, 7), (4, 8)
 b. $0.05 + 0.10 + 0.15 = 0.30$
 c. (2, 8), (3, 8), (4, 8)
 d. $0.05 + 0.05 + 0.15 = 0.25$
 e. 0.15

18. a. 令 C = 公司总部位于加利福尼亚州
 $P(C) = 53/500 = 0.106$
 b. 令 N = 公司总部位于纽约州
 T = 公司总部位于得克萨斯州
 $P(N) = 50/500 = 0.100$
 $P(T) = 52/500 = 0.104$
 $P(C) + P(N) + P(T) = 0.106 + 0.100 + 0.104 = 0.31$
 c. 令 A = 公司总部位于这 8 个州之一, $P(A) = 283/500 = 0.566$

20. a.

试验结果	财务独立的年龄	人数	概率
E_1	16 ~ 20 岁	191	$19/944 = 0.020\,3$
E_2	21 ~ 24 岁	467	$467/944 = 0.494\,7$
E_3	25 ~ 27 岁	244	$244/944 = 0.258\,5$
E_4	28 岁及以上	42	$42/944 = 0.044\,5$
		944	

- b. $P(\text{年龄} < 25 \text{ 岁}) = P(E_1) + P(E_2) = 0.020\,3 + 0.494\,7 =$

- 0.697 0
- c. $P(\text{年龄} > 24 \text{ 岁}) = P(E_3) + P(E_4) = 0.258 5 + 0.044 5 = 0.303 0$
- d. 在 25 岁之前希望实现财务独立的概率等于 0.697 0。在一般的经济条件下, 这个值看上去相当高。这次调查中接受访问的青少年期望自己在相对年轻的时候就实现财务独立, 然而这种想法似乎并不现实。
22. a. 0.40, 0.40, 0.60
b. 0.80, 是
c. $A^c = \{E_3, E_4, E_5\}; C^c = \{E_1, E_4\};$
 $P(A^c) = 0.60; P(C^c) = 0.40$
d. $\{E_1, E_2, E_5\}; 0.60$
e. 0.80
23. a. $P(A) = P(E_1) + P(E_4) + P(E_6) = 0.05 + 0.25 + 0.10 = 0.40$
 $P(B) = P(E_2) + P(E_4) + P(E_7) = 0.20 + 0.25 + 0.05 = 0.50$
 $P(C) = P(E_2) + P(E_3) + P(E_5) + P(E_7) = 0.20 + 0.20 + 0.15 + 0.05 = 0.60$
b. $A \cup B = \{E_1, E_2, E_4, E_6, E_7\}$
 $P(A \cup B) = P(E_1) + P(E_2) + P(E_4) + P(E_6) + P(E_7) = 0.05 + 0.20 + 0.25 + 0.10 + 0.05 = 0.65$
c. $A \cap B = \{E_4\}; P(A \cap B) = P(E_4) = 0.25$
d. 是的, 它们是互斥的。
e. $B^c = \{E_1, E_3, E_5, E_6\}$
 $P(B^c) = P(E_1) + P(E_3) + P(E_5) + P(E_6) = 0.05 + 0.20 + 0.15 + 0.10 = 0.50$
24. a. 0.05
b. 0.70
26. a. 令 D = 国内产权基金; $P(D) = 16/25 = 0.64$
b. 令 A = 4 或者 5 星评级; $P(A) = 12/25 = 0.48$
c. 有 7 只国内产权基金的评级是 4 星, 有 2 只是 5 星。因此, 有 9 只基金是评级 4 星或者 5 星的国内产权基金。 $P(D \cap A) = 9/25 = 0.36$
d. $P(D \cup A) = P(D) + P(A) - P(D \cap A) = 0.64 + 0.48 - 0.36 = 0.76$
28. 令 $B = \{\text{由于工作原因租赁汽车}\}$
 $P = \{\text{由于个人原因租赁汽车}\}$
a. $P(B \cup P) = P(B) + P(P) - P(B \cap P) = 0.540 + 0.458 - 0.300 = 0.698$
b. $P(\text{都不是}) = 1 - 0.698 = 0.302$
30. a. $P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{0.40}{0.60} = 0.666 7$
b. $P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{0.40}{0.5} = 0.80$

c. 不是, 因为 $P(A | B) \neq P(A)$ 。

32. a.

	小汽车	轻型卡车	合计
美国	0.133 0	0.293 9	0.426 9
非美国	0.347 8	0.225 3	0.573 1
合计	0.480 8	0.519 2	1.000 0

- b. 0.426 9, 0.5731; 非美国制造商的概率更高。
0.480 8, 0.519 2; 轻型卡车的概率略高。
c. 0.311 5, 0.688 5; 轻型卡车的概率更高。
d. 0.690 9, 0.393 1; 小汽车的概率更高。
e. 0.566 1; 对于一辆轻型卡车, 它是美国制造的的概率更高。

33. a.

预定入学 学制	毕业专业			合计
	商科	工科	其他	
全日制	0.270	0.151	0.192	0.613
非全日制	0.115	0.123	0.149	0.387
合计	0.385	0.274	0.341	1.000

- b. $P(\text{商科}) = 0.385, P(\text{工科}) = 0.274, P(\text{其他}) = 0.341$, 商科专业毕业生更偏好就读 *MBA*。
- c. $P(\text{工科} | \text{全日制}) = \frac{P(\text{全日制} \cap \text{工科})}{P(\text{全日制})} = \frac{0.151}{0.613} = 0.246$
- d. $P(\text{全日制} | \text{商科}) = \frac{P(\text{全日制} \cap \text{商科})}{P(\text{商科})} = \frac{0.270}{0.385} = 0.701$
- e. 当独立的时候, 一定有
 $P(A \cap B) = P(A) P(B)$;
从 (a) 中联合概率表可得
 $P(A) = 0.613, P(B) = 0.385$
于是 $P(A) P(B) = 0.613 \times 0.385 = 0.236$ 。但是,
 $P(A \cap B) = 0.270$ 。
由于 $P(A) P(B) \neq P(A \cap B)$, 因此事件不是相互独立的。
34. a. 令 O = 航班准时抵达
 L = 航班晚点
 J = 捷蓝航空
 N = 联合航空
 U = 美国航空
已知 $P(O | J) = 0.768, P(O | N) = 0.715, P(O | U) = 0.822, P(J) = 0.30, P(N) = 0.32, P(U) = 0.38$ 根据乘法公式, 得到联合概率
 $P(J \cap O) = P(J) P(O | J) = 0.30 \times 0.768 = 0.230 4$
 $P(N \cap O) = P(N) P(O | N) = 0.32 \times 0.715 = 0.228 8$
 $P(U \cap O) = P(U) P(O | U) = 0.38 \times 0.822 = 0.312 4$
连同边际概率 $P(J) = 0.30, P(N) = 0.32, P(U) =$

0.38, 得到如下联合概率表:

	准点	延迟	合计
捷蓝航空	0.230 4	0.069 6	0.30
联合航空	0.228 8	0.091 2	0.32
美国航空	0.312 4	0.067 6	0.38
合计	0.771 6	0.228 4	1.00

b. $P(O) = 0.230\ 4 + 0.228\ 8 + 0.312\ 4 = 0.771\ 6$

c. 因为抵达 C 出口的航班中美国航空所占的百分比最高, $P(U) = 0.38$, 所以 1382 航班最可能属于美国航空。

d. 从联合概率表可知, $P(L) = 0.228\ 4$

$$P(J|L) = \frac{P(J \cap L)}{P(L)} = \frac{0.069\ 6}{0.228\ 4} = 0.304\ 7$$

$$P(N|L) = \frac{P(N \cap L)}{P(L)} = \frac{0.091\ 2}{0.228\ 4} = 0.399\ 2$$

$$P(U|L) = \frac{P(U \cap L)}{P(L)} = \frac{0.067\ 6}{0.228\ 4} = 0.296\ 1$$

1382 航班最可能属于的航空公司是联合航空, 概率为 0.399 2。该航班最不可能属于美国航空, 概率为 0.296 1。

36. a. 对于每次罚球, $P(\text{罚球命中}) = 0.93$, 因此 Jamal Crawford 两次连续罚球都命中的概率为 $P(\text{罚球命中}) \cdot P(\text{罚球命中}) = 0.93 \times 0.93 = 0.864\ 9$ 。

b. 在以下三种不同的情形下, Jamal Crawford 都可以达到“罚球至少命中一个球”:

“第一个球罚中但第二个球未罚中” “第二个球罚中但第一个球未罚中” “两球均罚中”。

由于事件“罚球未中”是事件“罚球命中”的对立事件, 有

$$P(\text{罚球未中}) = 1 - P(\text{罚球命中}) = 1 - 0.93 = 0.07$$

于是,

$$P(\text{罚球命中}) \cdot P(\text{罚球未中}) = 0.93 \times 0.07 = 0.065\ 1$$

$$P(\text{罚球未中}) \cdot P(\text{罚球命中}) = 0.07 \times 0.93 = 0.065\ 1$$

$$P(\text{罚球命中}) \cdot P(\text{罚球命中}) = 0.93 \times 0.93 = 0.864\ 9$$

得到 $P(\text{罚球至少命中一个}) = 0.995\ 1$ 。

c. 我们有两种做法。

做法 1: $P(\text{罚球未中}) \cdot P(\text{罚球未中}) = 0.07 \times 0.07 = 0.004\ 9$, 这属于直接计算。

或者做法 2: 由于事件“两次罚球均未中”是事件“两次罚球至少命中一次”的对立事件, 有

$$P(\text{罚球未中}) \cdot P(\text{罚球未中}) = 1 - 0.995\ 1 = 0.004\ 9$$

d. 对于波特兰开拓者队的每次罚球, $P(\text{罚球命中}) = 0.58$, 因此波特兰开拓者队两次连续罚球都命中的概率为 $P(\text{罚球命中}) \cdot P(\text{罚球命中}) = 0.58 \times 0.58 = 0.336\ 4$ 。

同理, 在以下三种不同的情形下, 波特兰开拓者队都可以达到“罚球至少命中一个球”:

“第一个球罚中但第二个球未罚中” “第二个球罚中但第一个球未罚中” “两球均罚中”。

由于事件“罚球未中”是事件“罚球命中”的对立事件, 有

$$P(\text{罚球未中}) = 1 - P(\text{罚球命中}) = 1 - 0.58 = 0.42$$

于是,

$$P(\text{罚球命中}) \cdot P(\text{罚球未中}) = 0.58 \times 0.42 = 0.243\ 6$$

$$P(\text{罚球未中}) \cdot P(\text{罚球命中}) = 0.42 \times 0.58 = 0.243\ 6$$

$$P(\text{罚球命中}) \cdot P(\text{罚球命中}) = 0.58 \times 0.58 = 0.336\ 4$$

得到 $P(\text{罚球至少命中一个}) = 0.823\ 6$ 。

以下两种不同的方法都可以求得波特兰开拓者队“罚球两罚均未命中”的概率, 一种是直接计算, 得

$$P(\text{罚球未中}) \cdot P(\text{罚球未中}) = 0.42 \times 0.42 = 0.176\ 4$$

另一种方法是, 考虑到事件“两次罚球均未中”是事件“两次罚球至少命中一次”的对立事件, 于是有

$$P(\text{罚球未中}) \cdot P(\text{罚球未中}) = 1 - 0.823\ 6 = 0.176\ 4$$

可见, 一个更好的策略是故意让波特兰开拓者队罚球, 但不要对 Jamal Crawford 犯规。

38. 令 $Y =$ 有大学学位

$N =$ 没有大学学位

$D =$ 学生拖欠偿还贷款

a. 由表可知, $P(Y) = 0.42$

b. 由表可知, $P(N) = 0.58$

$$c. P(D|Y) = \frac{P(D \cap Y)}{P(Y)} = \frac{0.16}{0.42} = 0.381\ 0$$

$$d. P(D|N) = \frac{P(D \cap N)}{P(N)} = \frac{0.34}{0.58} = 0.586\ 2$$

e. 获得大学学位的人拖欠偿还贷款的概率为 0.381 0, 而那些肄业未获得大学学位的人拖欠偿还贷款的概率为 0.586 2。未能获得大学学位的学生中拖欠偿还贷款的概率更高, 他们未来似乎更容易导致财务问题。

39. a. 是, 因为 $P(A_1 \cap A_2) = 0$

$$b. P(A_1 \cap B) = P(A_1)P(B|A_1) = 0.40 \times 0.20 = 0.08$$

$$P(A_2 \cap B) = P(A_2)P(B|A_2) = 0.60 \times 0.05 = 0.03$$

$$c. P(B) = P(A_1 \cap B) + P(A_2 \cap B) = 0.08 + 0.03 = 0.11$$

$$d. P(A_1|B) = 0.08/0.11 = 0.727\ 3$$

$$P(A_2|B) = 0.03/0.11 = 0.272\ 7$$

40. a. 0.10, 0.20, 0.09

b. 0.51

c. 0.26, 0.51, 0.23

42. $M =$ 拖欠支付

$D_1 =$ 不偿还债务

$D_2 =$ 偿还债务

$$P(D_1) = 0.05, P(D_2) = 0.95,$$

$$P(M | D_2) = 0.2, P(M | D_1) = 1$$

$$\begin{aligned} \text{a. } P(D_1 | M) &= \frac{P(D_1)P(M | D_1)}{P(D_1)P(M | D_1) + P(D_2)P(M | D_2)} \\ &= \frac{0.05 \times 1}{0.05 \times 1 + 0.95 \times 0.2} = \frac{0.05}{0.24} = 0.21 \end{aligned}$$

b. 是, 不偿还债务的概率超过了 0.20。

44. 令 M = 当前访问 ParFore 网站的人是女性

F = 当前访问 ParFore 网站的人是男性

D = 当前访问 ParFore 网站的人之前访问过 Dillard 网站

a. 根据历史数据, $P(F) = 0.40$ 。

b. $P(M) = 0.60, P(D | F) = 0.3, P(D | M) = 0.10$

$$\begin{aligned} P(F | D) &= \frac{P(F)P(D | F)}{P(F)P(D | F) + P(M)P(D | M)} \\ &= \frac{0.40 \times 0.30}{0.40 \times 0.30 + 0.60 \times 0.10} = 0.6667 \end{aligned}$$

46. a. $422 + 181 + 80 + 121 + 201 = 1\,005$ 名受访者

b. 频率最高的回答是 1 天及以上, 概率 = $422/1\,005 = 0.4199$ 。

c. $201/1\,005 = 0.20$

d. 回答 2 天、3 天、4 天及以上的受访者人数 = $181 + 80 + 121 = 382$, 概率 = $382/1\,005 = 0.3801$

48. a. 0.5029

b. 0.5758

c. 不独立; 由 (a) 有 $P(F) = 0.5029$, 由 (b) 有 $P(A | F) = 0.5758$; 由于 $P(F) \neq P(A | F)$, 所以事件 A 和事件 F 不独立。

50. a. 0.76

b. 0.24

52. b. 0.2022

c. 0.4618

d. 0.4005

54. a. 0.7766

b. $P(\text{赞同} | \text{年龄在 } 30 \sim 49 \text{ 岁})$

$$\begin{aligned} &= \frac{P(\text{赞同} \cap \text{年龄在 } 30 \sim 49 \text{ 岁})}{P(\text{年龄在 } 30 \sim 49 \text{ 岁})} \\ &= \frac{0.0907}{0.3180} = 0.2852 \end{aligned}$$

c. $P(\text{年龄在 } 50 \text{ 岁以上} | \text{不赞同})$

$$\begin{aligned} &= \frac{P(\text{不赞同} \cap \text{年龄在 } 50 \text{ 岁以上})}{P(\text{不赞同})} \\ &= \frac{0.4008}{0.7766} = 0.5161 \end{aligned}$$

d. 受访者对这种做法的态度与受访者的年龄是不独立的。由 (b) 有 $P(\text{赞同} | \text{年龄在 } 30 \sim 49 \text{ 岁}) = 0.2852$, 此外 $P(\text{赞同}) = 0.2234$ 。如果受访者对这种做法的态

度与受访者的年龄是独立, 那么这两个概率应该是相等的。而实际上, 它们并不相等, 数据表明受访者对这种做法的态度与受访者的年龄是不独立的。

e. 年龄在 50 岁以上的受访者比 18 ~ 29 岁年龄段的受访者更倾向于不赞同这种做法。

$$\begin{aligned} &P(\text{不赞同} | \text{年龄在 } 50 \text{ 岁以上}) \\ &= \frac{P(\text{不赞同} \cap \text{年龄在 } 50 \text{ 岁以上})}{P(\text{年龄在 } 50 \text{ 岁以上})} \end{aligned}$$

$$= \frac{0.4008}{0.4731} = 0.8472$$

$$\begin{aligned} &P(\text{不赞同} | \text{年龄在 } 18 \sim 29 \text{ 岁}) \\ &= \frac{P(\text{不赞同} \cap \text{年龄在 } 18 \sim 29 \text{ 岁})}{P(\text{年龄在 } 18 \sim 29 \text{ 岁})} \end{aligned}$$

$$= \frac{0.1485}{0.2089} = 0.7109$$

56. a. 0.25

b. 0.125

c. 0.0125

d. 0.10

e. 不是

58. a. 令 A_1 = 学生出国留学

A_2 = 学生不出国留学

F = 男学生

M = 女学生

$$P(A_1) = 0.095$$

$$P(A_2) = 1 - P(A_1) = 1 - 0.095 = 0.905$$

$$P(F | A_1) = 0.60$$

$$P(F | A_2) = 0.49$$

利用表格计算

	$P(A_i)$	$P(F A_i)$	$P(A_i \cap F)$	$P(A_i F)$
A_1	0.095	0.60	0.0570	0.1139
A_2	0.905	0.49	0.4435	0.8861
$P(F) = 0.5005$				

$$P(A_1 | F) = 0.1139$$

b.

	$P(A_i)$	$P(M A_i)$	$P(A_i \cap M)$	$P(A_i M)$
A_1	0.095	0.40	0.0380	0.0761
A_2	0.905	0.51	0.4615	0.9239
$P(M) = 0.4995$				

$$P(A_1 | M) = 0.0761$$

c. 由上可知 $P(F) = 0.5005$, $P(M) = 0.4995$, 因此在全日制学生中男生和女生几乎是 50:50。

$$\begin{aligned}
60. \text{ a. } & P(\text{电邮广告} \mid \text{shipping!}) \\
&= \frac{P(\text{电邮广告})P(\text{shipping!} \mid \text{电邮广告})}{P(\text{电邮广告})P(\text{shipping!} \mid \text{电邮广告}) + P(\text{非电邮广告})P(\text{shipping!} \mid \text{非电邮广告})} \\
&= \frac{0.10 \times 0.051}{0.10 \times 0.051 + 0.90 \times 0.0015} = 0.791 \\
& P(\text{非电邮广告} \mid \text{shipping!})
\end{aligned}$$

$$\begin{aligned}
&= \frac{P(\text{非电邮广告})P(\text{shipping!} \mid \text{非电邮广告})}{P(\text{非电邮广告})P(\text{shipping!} \mid \text{非电邮广告}) + P(\text{电邮广告})P(\text{shipping!} \mid \text{电邮广告})} \\
&= \frac{0.10 \times 0.051}{0.90 \times 0.0015 + 0.10 \times 0.051} = 0.209
\end{aligned}$$

如果一条消息中含有“shipping!”这个词，则它是一条电邮广告的概率是相当高的，达到0.7910，于是这条消息会被表示成电邮广告。

$$\begin{aligned}
\text{b. } & P(\text{电邮广告} \mid \text{today!}) \\
&= \frac{P(\text{电邮广告})P(\text{today!} \mid \text{电邮广告})}{P(\text{电邮广告})P(\text{today!} \mid \text{电邮广告}) + P(\text{非电邮广告})P(\text{today!} \mid \text{非电邮广告})} \\
&= \frac{0.10 \times 0.045}{0.10 \times 0.045 + 0.90 \times 0.0022} = 0.694 \\
& P(\text{电邮广告} \mid \text{here!})
\end{aligned}$$

$$\begin{aligned}
&= \frac{P(\text{电邮广告})P(\text{here!} \mid \text{电邮广告})}{P(\text{电邮广告})P(\text{here!} \mid \text{电邮广告}) + P(\text{非电邮广告})P(\text{here!} \mid \text{非电邮广告})} \\
&= \frac{0.10 \times 0.034}{0.10 \times 0.034 + 0.90 \times 0.0022} = 0.632
\end{aligned}$$

由于 $P(\text{today!} \mid \text{电邮广告})$ 大于 $P(\text{here!} \mid \text{电邮广告})$ ，所以一条含有词“today!”的消息比一条含有词“here!”的消息更可能是电邮广告，含有词“today!”的消息更容易分辨是否是电邮广告。

$$\begin{aligned}
\text{c. } & P(\text{电邮广告} \mid \text{available!}) \\
&= \frac{P(\text{电邮广告})P(\text{available!} \mid \text{电邮广告})}{P(\text{电邮广告})P(\text{available!} \mid \text{电邮广告}) + P(\text{非电邮广告})P(\text{available!} \mid \text{非电邮广告})} \\
&= \frac{0.10 \times 0.014}{0.10 \times 0.014 + 0.90 \times 0.0041} = 0.275 \\
& P(\text{电邮广告} \mid \text{fingertips!}) \\
&= \frac{P(\text{电邮广告})P(\text{fingertips!} \mid \text{电邮广告})}{P(\text{电邮广告})P(\text{fingertips!} \mid \text{电邮广告}) + P(\text{非电邮广告})P(\text{fingertips!} \mid \text{非电邮广告})} \\
&= \frac{0.10 \times 0.014}{0.10 \times 0.014 + 0.90 \times 0.0011} = 0.586
\end{aligned}$$

由于 $P(\text{fingertips!} \mid \text{非电邮广告})$ 小于 $P(\text{available!} \mid \text{非电邮广告})$ ，所以一条含有词“fingertips!”的消息更可能是电邮广告；由于词“available!”在合法消息中更容易出现，因此含有词“available!”的消息更难

以分辨是否是电邮广告。

- d. 当一个词在合法消息中更不容易出现或者一个词在电邮广告消息中更容易出现时，含有这个词的消息更易于分辨是否是电邮广告。

第5章

1. a. (正, 正), (正, 反), (反, 正), (反, 反)
b. x = 两次抛掷中硬币出现正面的次数
c.

结果	x 的值
(正, 正)	2
(正, 反)	1
(反, 正)	1
(反, 反)	0

- d. 离散型; 0, 1 和 2
2. a. x = 组装产品所需时间 (以分钟计)
b. 任何正值: $x > 0$
c. 连续型
3. 令 Y = 被录用
 N = 未被录用
a. $S = \{(Y, Y, Y), (Y, Y, N), (Y, N, Y), (Y, N, N), (N, Y, Y), (N, Y, N), (N, N, Y), (N, N, N)\}$
b. 令 N = 录取人数; N 是一个离散型随机变量。
c.

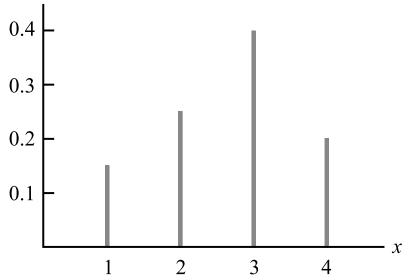
试验结果	(Y, Y, Y)	(Y, Y, N)	(Y, N, Y)	(Y, N, N)
N 的值	3	2	2	1
试验结果	(N, Y, Y)	(N, Y, N)	(N, N, Y)	(N, N, N)
N 的值	2	1	1	0

4. $x = 0, 1, 2, \dots, 9$
6. a. 0, 1, 2, \dots , 20; 离散型
b. 0, 1, 2, \dots ; 离散型
c. 0, 1, 2, \dots , 50; 离散型
d. $0 \leq x \leq 8$; 连续型
e. $x > 0$; 连续型
7. a. 对所有 x 的值, $f(x) \geq 0$
 $\sum f(x) = 1$; 因此这是一个有效的概率分布。
b. $x = 30$ 的概率是 $f(30) = 0.25$ 。
c. $x \leq 25$ 的概率是 $f(20) + f(25) = 0.20 + 0.15 = 0.35$ 。
d. $x > 30$ 的概率是 $f(35) = 0.40$ 。

8. a.

x	$f(x)$
1	$3/20 = 0.15$
2	$5/20 = 0.25$
3	$8/20 = 0.40$
4	$4/20 = 0.20$
合计 1.00	

b. $f(x)$



c. 对于 $x = 1, 2, 3, 4$, 有 $f(x) \geq 0$

$$\sum f(x) = 1$$

10. a.

x	1	2	3	4	5
$f(x)$	0.05	0.09	0.03	0.42	0.41

b.

x	1	2	3	4	5
$f(x)$	0.04	0.10	0.12	0.46	0.28

c. 0.83

d. 0.28

e. 高级经理的整体工作满意度更高

12. a. 是的; $f(x) \geq 0$, $\sum f(x) = 1$

$$b. f(500\,000) + f(600\,000) = 0.10 + 0.05 = 0.15$$

$$c. f(100\,000) = 0.10$$

14. a. 0.05

b. 0.70

c. 0.40

16. a.

y	$f(y)$	$yf(y)$
2	0.20	0.40
4	0.30	1.20
7	0.40	2.80
8	<u>0.10</u>	<u>0.80</u>
合计	1.00	5.20

$$E(y) = \mu = 5.20$$

b.

Y	$y - \mu$	$(y - \mu)^2$	$f(y)$	$\frac{(y - \mu)^2}{f(y)}$
2	-3.20	10.24	0.20	2.048
4	-1.20	1.44	0.30	0.432
7	1.80	3.24	0.40	1.296
8	2.80	7.84	0.10	<u>0.784</u>
合计				4.560

$$\text{Var}(y) = 4.56$$

$$\sigma = \sqrt{4.56} = 2.14$$

18. a/b.

x	$f(x)$	$xf(x)$	$x - \mu$	$(x - \mu)^2$	$\frac{(x - \mu)^2}{f(x)}$
0	0.218 8	0.000 0	-1.182 5	1.398 2	0.306 0
1	0.548 4	0.548 4	-0.182 5	0.033 3	0.018 3
2	0.124 1	0.248 3	0.817 5	0.668 4	0.083 0
3	0.048 9	0.146 6	1.817 5	3.303 5	0.161 4
4	<u>0.059 8</u>	<u>0.239 3</u>	2.817 5	7.938 6	<u>0.474 9</u>
合计	1.000 0	1.182 5			1.043 5

\uparrow
 $E(x)$

\uparrow
 $\text{Var}(x)$

c/d.

y	$f(y)$	$yf(y)$	$y - \mu$	$(y - \mu)^2$	$\frac{(y - \mu)^2}{f(y)}$
0	0.249 7	0.000 0	-1.218 0	1.483 5	0.370 4
1	0.481 6	0.481 6	-0.218 0	0.047 5	0.022 9
2	0.140 1	0.280 1	0.782 0	0.611 5	0.085 6
3	0.058 3	0.174 9	1.782 0	3.175 5	0.185 1
4	<u>0.070 3</u>	<u>0.281 4</u>	2.782 0	7.739 5	<u>0.544 4</u>
合计	1.000 0	1.218 0			1.208 5

\uparrow
 $E(y)$

\uparrow
 $\text{Var}(y)$

e. 过去3个月中, 自有住房停水6小时及以上的次数的数学期望为1.182 5, 比租赁的期望值略小, 租赁住房停水6小时及以上的次数的数学期望为1.218 0。自有住房的波动性(1.043 5)小于租赁住房的波动性(1.208 5)。

20. a. 430

b. -90; 客户关心的是防护大损失的费用。

22. a. 445

b. 亏损1 250 美元

24. a. 中型: 145; 大型: 140

b. 中型: 2 725; 大型: 12 400

$$25. a. E(x) = 0.2 \times 50 + 0.5 \times 30 + 0.3 \times 40 = 37$$

$$E(y) = 0.2 \times 80 + 0.5 \times 50 + 0.3 \times 60 = 59$$

$$\text{Var}(x) = 0.2 \times (50 - 37)^2 + 0.5 \times (30 - 37)^2 + 0.3 \times$$

$$(40 - 37)^2 = 61$$

$$\text{Var}(y) = 0.2 \times (80 - 59)^2 + 0.5 \times (50 - 59)^2 + 0.3 \times (60 - 59)^2 = 129$$

b.

$x + y$	$f(x + y)$
130	0.20
80	0.50
100	0.30

c.

$x + y$	$f(x + y)$	$(x + y) - E(x + y)$	$[x + y - E(x + y)]^2$	$f(x + y)$
130	0.20	26	676	0.20
80	0.50	40	1600	0.50
100	0.30	4	16	0.30
$E(x + y) = 96$		$\text{Var}(x + y) = 364$		

$$d. \sigma_{xy} = [\text{Var}(x + y) - \text{Var}(x) - \text{Var}(y)]/2 = (364 - 61 - 129)/2 = 87$$

在 (a) 中计算得到 $\text{Var}(x) = 61$, $\text{Var}(y) = 129$, 因此

$$\sigma_x = \sqrt{61} = 7.8102, \sigma_y = \sqrt{129} = 11.3578$$

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{87}{7.8102 \times 11.3578} = 0.98$$

随机变量 x 和 y 是正相关的; 协方差和相关系数都是正的, 并且它们的确是高度相关的; 相关系数几乎为 1。

$$e. \text{Var}(x + y) = \text{Var}(x) + \text{Var}(y) + 2\sigma_{xy} = 61 + 129 + 2 \times 87 = 364$$

$$\text{Var}(x) + \text{Var}(y) = 61 + 129 = 190$$

随机变量 x 和 y 的和的方差比两个方差的和大 2 倍的协方差, 即 $2 \times 87 = 174$; 这时变量之间是正相关的, 协方差是正的。只要两个变量之间是正相关的, 那么随机变量和的方差就一定大于两个随机变量各自的方差之和。

26. a. 5%, 1%, 股票 1 的风险更大。

b. 42.25 美元, 25.00 美元

c. 5.825, 2.236

d. 6.875%, 3.329%

e. -0.6, 强负相关关系。

27. a. 将表中频数除以餐馆总数就得到如下联合分布表, 不同价格和品质的菜品的二元概率见表, 即二元概率分布。比如, 质量评级为 2, 价格评级为 3 的菜品的概率 $f(2, 3) = 0.18$; 质量 x 的边际概率分布在最右侧的列, 价格 y 的边际概率分布在最下侧的行。

品质 (x)	菜品价格 (y)			合计
	1	2	3	
1	0.14	0.13	0.01	0.28
2	0.11	0.21	0.18	0.50
3	0.01	0.05	0.16	0.22
合计	0.26	0.39	0.35	1

$$b. E(x) = 0.28 \times 1 + 0.50 \times 2 + 0.22 \times 3 = 1.94$$

$$\text{Var}(x) = 0.28 \times (1 - 1.94)^2 + 0.50 \times (2 - 1.94)^2 + 0.22 \times (3 - 1.94)^2 = 0.4964$$

$$c. E(y) = 0.26 \times 1 + 0.39 \times 2 + 0.35 \times 3 = 2.09$$

$$\text{Var}(y) = 0.26 \times (1 - 2.09)^2 + 0.39 \times (2 - 2.09)^2 + 0.35 \times (3 - 2.09)^2 = 0.6019$$

$$d. \sigma_{xy} = [\text{Var}(x + y) - \text{Var}(x) - \text{Var}(y)]/2 = (1.6691 - 0.4964 - 0.6019 - 61 - 129)/2 = 0.2854$$

因为协方差 $\sigma_{xy} = 0.2854$ 是正值, 所以正如我们所期望的那样, 质量评级越高, 就餐价格也越高。

$$e. \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{0.2854}{\sqrt{0.4964} \times \sqrt{0.6019}} = 0.5221$$

相关系数等于 0.5221, 具有中等程度的正相关关系。找到一家质量评级高且就餐价格低的餐厅并不容易, 但不是不可能, 此处这样的餐厅就有 3 家, 因此 $f(3, 1) = 0.01$ 。

28. a. $E(y) = 45.30$ 美元, $\text{Var}(y) = 3.81$, $\sigma_y = 1.95$ 美元

b. $E(x) = 90.50$ 美元, $\text{Var}(x) = 24.75$, $\sigma_x = 4.97$ 美元

c.

z	$f(z)$
128	0.05
130	0.20
133	0.20
138	0.25
140	0.20
143	0.10
合计	1.00

d. 135.8 美元, 21.26 美元, 4.61 美元

e. 不独立, $\sigma_{xy} = -3.65$

f. 不是, 制造数量较少。

30. a. 333.486, -1.974

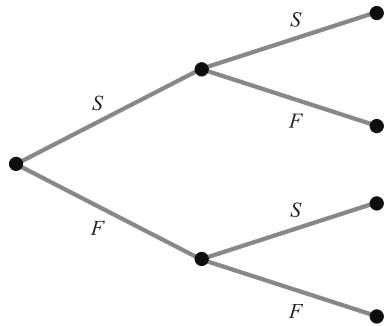
b. 9.055%, 19.89%

c. 9.425%, 11.63%

d. 7.238%, 4.94%

e. 激进的投资: 50% 的核心债券, 50% 的 REITS
保守的投资: 80% 的核心债券, 20% 的 REITS

31. a.



$$b. f(1) = \binom{2}{1} 0.4^1 0.6^1 = \frac{2!}{1!1!} 0.4 \times 0.6 = 0.48$$

$$c. f(0) = \binom{2}{0} 0.4^0 0.6^2 = \frac{2!}{0!2!} 1 \times 0.36 = 0.36$$

$$d. f(2) = \binom{2}{2} 0.4^2 0.6^0 = \frac{2!}{0!2!} 0.16 \times 0.1 = 0.16$$

$$e. P(x \geq 1) = f(1) + f(2) = 0.48 + 0.16 = 0.64$$

$$f. E(x) = np = 2 \times 0.4 = 0.8$$

$$\text{Var}(x) = np(1-p) = 2 \times 0.4 \times 0.6 = 0.48$$

$$\sigma = \sqrt{0.48} = 0.6928$$

32. a. 0.3487

b. 0.1937

c. 0.9298

d. 0.6513

e. 1

f. 0.9, 0.95

34. a. 是。因为青少年是随机选取的，各次试验中的 p 相同，并且试验是独立的。是否用潘多拉传媒公司的线上服务听音乐是每次试验的两种结果。

二项分布中 $n = 10$ 和 $p = 0.35$

$$f(x) = \frac{10!}{x!(10-x)!} 0.35^x (1-0.35)^{10-x}$$

$$b. f(0) = \frac{10!}{0!(10-0)!} 0.35^0 (1-0.35)^{10-0} = 0.0135$$

$$c. f(4) = \frac{10!}{4!(10-4)!} 0.35^4 (1-0.35)^{10-4} = 0.2377$$

$$d. P(x \geq 2) = 1 - f(0) - f(1)$$

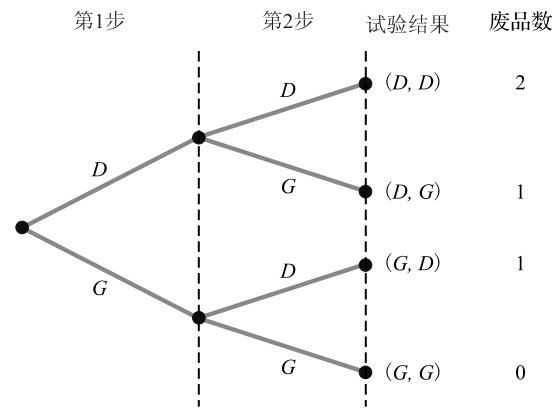
由 (b) 已知, $f(0) = 0.0135$

$$f(1) = \frac{10!}{1!(10-1)!} 0.35^1 (1-0.35)^{10-1} = 0.0725$$

36. a. 每个被选取的零件为废品的概率都是 0.03；零件的选取必须是独立的。

b. 令 D = 废品

G = 正品



c. 恰好发现一个废品的试验结果有 2 种。

$$d. P(\text{没有废品}) = 0.97 \times 0.97 = 0.9409$$

$$P(\text{恰好有一件废品}) = 2 \times 0.03 \times 0.97 = 0.0582$$

$$P(\text{恰好有两件废品}) = 0.03 \times 0.03 = 0.0009$$

38. a. 0.90

b. 0.99

c. 0.999

d. 是

40. a. 是。因为 18 ~ 34 岁与父母共同生活的人是随机选取的，各次试验中的 p 相同，并且试验是独立的。是否分担家庭开支是每次试验的两种结果。

二项分布中 $n = 15$ 和 $p = 0.75$

$$f(x) = \frac{15!}{x!(15-x)!} 0.75^x (1-0.75)^{15-x}$$

b. 15 个人都不分担家庭开支的概率 = 0.0000

$$c. \text{概率} = 0.1651 + 0.2252 + 0.2252 + 0.1559 + 0.0668 + 0.0134 = 0.8561$$

$$42. a. f(4) = \frac{20!}{4!(20-4)!} 0.30^4 (0.70)^{20-4}$$

$$b. P(x \geq 2) = 1 - f(0) - f(1)$$

$$f(0) = \frac{20!}{0!(20-0)!} 0.30^0 (0.70)^{20-0} = 0.0008$$

$$f(1) = \frac{20!}{1!(20-1)!} 0.30^1 (0.70)^{20-1} = 0.0068$$

$$P(x \geq 2) = 1 - f(0) - f(1) = 1 - (0.0008 + 0.0068) = 0.9924$$

$$c. E(x) = np = 20 \times 0.30 = 6$$

$$d. \text{Var}(x) = np(1-p) = 20 \times 0.30 \times 0.70 = 4.2$$

$$\sigma = \sqrt{4.2} = 2.0499$$

$$44. a. f(x) = \frac{3^x e^{-3}}{x!}$$

b. 0.2241

c. 0.1494

d. 0.8008

45. a. $f(x) = \frac{2^x e^{-2}}{x!}$

b. 3 单位长度的时间段中 $\mu = 6$

c. $f(x) = \frac{6^x e^{-6}}{x!}$

d. $f(2) = \frac{2^2 e^{-2}}{2!} = \frac{4 \times 0.1353}{2} = 0.2706$

e. $f(6) = \frac{6^6 e^{-6}}{6!} = 0.1606$

f. $f(5) = \frac{4^5 e^{-4}}{5!} = 0.1563$

46. a. 0.1952

b. 0.1048

c. 0.0183

d. 0.0907

48. a. 15 分钟里的均值为 $14.4/4 = 3.6$ 。

$f(0) = \frac{3.6^0 e^{-3.6}}{0!} = e^{-3.6} = 0.0273$

b. 概率 $= 1 - f(0) = 1 - 0.0273 = 0.9727$

c. 概率 $= 1 - [f(0) + f(1) + f(2) + f(3)]$
 $= 1 - [0.0273 + 0.0984 + 0.0223 + 0.1771 + 0.2125]$
 $= 0.4847$

50. a. 6 月份每天 $\mu = 18/30 = 0.6$

b. $f(0) = \frac{0.6^0 e^{-0.6}}{0!} = 0.5488$

c. $f(1) = \frac{0.6^1 e^{-0.6}}{1!} = 0.3293$

d. $P(\text{大于 } 1) = 1 - f(0) - f(1) = 1 - 0.5488 - 0.3293 = 0.1219$

52. a. $f(1) = \frac{\binom{3}{1} \binom{10-3}{4-1}}{\binom{10}{4}} = \frac{\left(\frac{3!}{1!2!}\right) \times \left(\frac{7!}{3!4!}\right)}{\left(\frac{10!}{4!6!}\right)} = \frac{3 \times 35}{210}$
 $= 0.50$

b. $f(2) = \frac{\binom{3}{2} \binom{10-3}{2-2}}{\binom{10}{2}} = \frac{3 \times 1}{45} = 0.067$

c. $f(0) = \frac{\binom{3}{0} \binom{10-3}{2-0}}{\binom{10}{2}} = \frac{1 \times 21}{45} = 0.4667$

d. $f(2) = \frac{\binom{3}{2} \binom{10-3}{4-2}}{\binom{10}{4}} = \frac{3 \times 21}{210} = 0.30$

e. $x = 4$ 大于 $r = 3$, 因此, $f(4) = 0$ 。

54. a. 0.5250

b. 0.8167

56. $N = 60, n = 10$

a. $r = 20, x = 0$

$$f(0) = \frac{\binom{20}{0} \binom{40}{10}}{\binom{60}{10}} = \frac{1 \times \left(\frac{40!}{10!30!}\right)}{\left(\frac{60!}{10!50!}\right)} = \left(\frac{40!}{10!30!}\right) \left(\frac{10!50!}{60!}\right)$$

$$= \frac{40 \cdot 39 \cdot 38 \cdot 37 \cdot 36 \cdot 35 \cdot 34 \cdot 33 \cdot 32 \cdot 31}{60 \cdot 59 \cdot 58 \cdot 57 \cdot 56 \cdot 55 \cdot 54 \cdot 53 \cdot 52 \cdot 51}$$

$$= 0.0112$$

b. $r = 20, x = 1$

$$f(1) = \frac{\binom{20}{1} \binom{40}{9}}{\binom{60}{10}} = 20 \left(\frac{40!}{9!31!}\right) \left(\frac{10!50!}{60!}\right) = 0.0725$$

c. $1 - f(0) - f(1) = 1 - 0.0112 - 0.0725 = 0.9163$

d. 与样本中有一个夏威夷工厂员工的概率相等; 0.0725

58. a. 0.2917

b. 0.0083

c. 0.5250, 0.1750; 1 家银行

d. 0.7083

e. 0.9, 0.49, 0.70

60. a.

x	$f(x)$
1	0.150
2	0.050
3	0.075
4	0.050
5	0.125
6	0.050
7	0.100
8	0.125
9	0.125
10	0.150
合计	1.000

b. 优质服务的概率 $= 0.125 + 0.150 = 0.275$

c. $E(x) = 5.925, \text{Var}(x) = 9.6694$

d. 新车代理商在等候时间上的评级为优质的概率为 $2/7 = 0.2857$ 。对于剩下的 $40 - 7 = 33$ 家服务商, 有 9 家的质量评级为优质, 这对应于概率 $9/33 = 0.2727$ 。

62. a. 0.05; 0.40; 在航站楼未购物乘客人数为 0 名。

b. $E(x) = 1.14, \text{Var}(x) = 0.3804$

x	$f(x)$
0	0.05
1	0.20
2	0.20
	1.00

c. 0.5, 0.45

d. $E(t) = 1.64$, $\text{Var}(t) = 0.5504$

t	$f(t)$
1	0.50
2	0.38
3	0.10
4	0.02
	1.00

e. $\sigma_{xy} = -0.14$, $\rho_{xy} = -0.3384$, 负相关关系

64. a. $f(3) = \binom{20}{3} 0.53^3 \times 0.47^{17} = 0.0005$

b. $n = 20$, $x = 0$

$$P(x \leq 5) = f(0) + f(1) + f(2) + f(3) + f(4) + f(5) = 0.4952$$

c. $E(x) = np = 2000 \times 0.49 = 980$

$$d. E(x) = np = 2000 \times 0.36 = 720, \sigma^2 = np(1-p) = 2000 \times 0.36 \times 0.64 = 460.8$$

$$\sigma = \sqrt{460.8} = 21.4663$$

66. a. 0.9510

b. 0.0480

c. 0.0490

68. a. $E(x) = np = 200 \times 0.235 = 47$

$$b. \sigma = \sqrt{np(1-p)} = \sqrt{200 \times 0.235 \times 0.765} = 5.9962$$

c. 5.9962

70. 0.1912

72. a. 0.2240

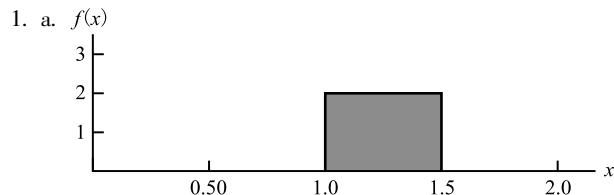
b. 0.5767

74. a. 0.4667

b. 0.4667

c. 0.0667

第6章



b. $P(x = 1.25) = 0$; 由于任一单个点的曲线下面积为 0, 所以任意单个点的概率是 0。

c. $P(1.0 \leq x \leq 1.25) = 2 \times 0.25 = 0.50$

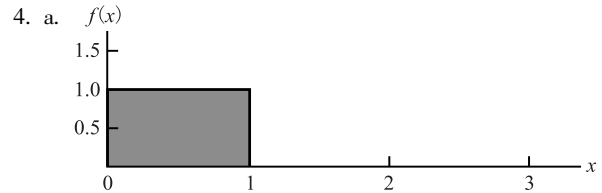
d. $P(1.20 < x < 1.5) = 2 \times 0.30 = 0.60$

2. b. 0.50

c. 0.60

d. 15

e. 8.33



b. $P(0.25 < x < 0.75) = 1 \times 0.50 = 0.50$

c. $P(x \leq 0.30) = 1 \times 0.30 = 0.30$

d. $P(x > 0.60) = 1 \times 0.40 = 0.40$

6. a. 均匀概率密度函数; $a = 136 - 1/2 \times 160 = 56$, $b = 136 + 1/2 \times 160 = 216$

b. $P(100 \leq x \leq 200) = (200 - 100) \times 0.00625 = 0.6250$

c. $P(x \geq 150) = (216 - 150) \times 0.00625 = 0.4125$

d. $P(x \leq 80) = (80 - 56) \times 0.00625 = 0.1500$

10. a. 0.9332

b. 0.8413

c. 0.0919

d. 0.4938

12. a. 0.2967

b. 0.4418

c. 0.3300

d. 0.5910

e. 0.8849

f. 0.2389

13. a. $P(-1.98 \leq z \leq 0.49) = P(z \leq 0.49) - P(z < -1.98) = 0.6879 - 0.0239 = 0.6640$

b. $P(0.52 \leq z \leq 1.22) = P(z \leq 1.22) - P(z < 0.52) = 0.8888 - 0.6985 = 0.1903$

c. $P(-1.75 \leq z \leq -1.04) = P(z \leq -1.04) - P(z < -1.75) = 0.1492 - 0.0401 = 0.1091$

14. a. $z = 1.96$ b. $z = 1.96$ c. $z = 0.61$ d. $z = 1.12$ e. $z = 0.44$ f. $z = 0.44$ 15. a. 与累积概率 0.2119 相对应的 $z = -0.80$

b. 计算 $0.9030/2 = 0.4515$; 与累积概率 $0.5000 + 0.4515 = 0.9515$ 相对应的 $z = 1.66$

c. 计算 $0.2052/2 = 0.1026$; 与 z 相对应的累积概率为 $0.5000 + 0.1026 = 0.6026$, 因此, $z = 0.26$ 。

d. 与累积概率 0.9948 相对应的 $z = 2.56$

e. z 左侧区域的面积等于 $1 - 0.6915 = 0.3085$, 因此, $z = -0.50$ 。

16. a. $z = 2.33$

b. $z = 1.96$

c. $z = 1.645$

d. $z = 1.28$

18. $\mu = 14.4$ 和 $\sigma = 4.4$

a. 当 $x = 20$ 时, $z = \frac{20 - 14.4}{4.4} = 1.27$

$$P(z \leq 1.27) = 0.8980$$

$$P(x \geq 20) = 1.000 - 0.8980 = 0.1020$$

b. 当 $x = 10$ 时, $z = \frac{10 - 14.4}{4.4} = -1.00$

$$P(z \leq -1.00) = 0.1587$$

$$P(x \leq 10) = 0.1587$$

c. z 值为 1.28, 截去大约 10% 的上侧面积

$$x = 14.4 + 4.4 \times 1.28 = 20.03$$

当国内股票基金的股票收益率达到或者超过 20.03% 时, 其排名能进入前 10%。

20. a. 0.1788

b. 69.15%

c. 0.0495

22. a. 0.6553

b. 13.05 小时

c. 0.9838

24. a. $z = \frac{x - \mu}{\sigma} = \frac{400 - 749}{225} = -1.55 = 1.27$

$$P(x < 400) = P(z < -1.55) = 0.0606$$

b. $z = \frac{x - \mu}{\sigma} = \frac{800 - 749}{225} = 0.23$

$$P(x \geq 800) = P(z \geq 0.23) = 1 - P(z \leq 0.23) \\ = 1 - 0.5910 = 0.4090$$

c. 当 $x = 1000$ 时, $z = \frac{x - \mu}{\sigma} = \frac{1000 - 749}{225} = 1.12$

当 $x = 500$ 时, $z = \frac{x - \mu}{\sigma} = \frac{500 - 749}{225} = -1.11$

$$P(500 \leq z \leq 1000) = P(z \leq 1.12) - P(z \leq -1.11) \\ = 0.8686 - 0.1335 = 0.7351$$

d. 当 $z = 1.645$ 时, 上侧 5% 或者面积 $= 1 - 0.05 = 0.95$

$$x = \mu + z\sigma = 749 + 1.645 \times 225 = 1119 \text{ (美元)}$$

26. a. $\mu = np = 100 \times 0.20 = 20$

$$\sigma^2 = np(1 - p) = 100 \times 0.20 \times 0.80 = 16$$

$$\sigma = \sqrt{16} = 4$$

b. 是, 因为 $np = 20$ 和 $n(1 - p) = 80$

c. $P(23.5 \leq x \leq 24.5)$

$$z = \frac{24.5 - 20}{4} = 1.13 P(z \leq 1.13) = 0.8708$$

$$z = \frac{23.5 - 20}{4} = 0.88 P(z \leq 0.88) = 0.8106$$

$$P(23.5 \leq x \leq 24.5) = P(0.88 \leq z \leq 1.13) = 0.8708 - 0.8106 = 0.0602$$

d. $P(17.5 \leq x \leq 22.5)$

$$z = \frac{22.5 - 20}{4} = 0.63 P(z \leq 0.63) = 0.7357$$

$$z = \frac{17.5 - 20}{4} = -0.63 P(z \leq -0.63) = 0.2643$$

$$P(17.5 \leq x \leq 22.5) = P(-0.63 \leq z \leq 0.63) = 0.7357 - 0.2643 = 0.4714$$

e. $P(x \leq 15.5)$

$$z = \frac{15.5 - 20}{4} = -1.13 P(z \leq -1.13) = 0.1292$$

$$P(x \leq 15.5) = P(z \leq -1.13) = 0.1292$$

28. a. $\mu = np = 250 \times 0.20 = 50$

b. $\sigma^2 = np(1 - p) = 250 \times 0.20 \times (1 - 0.2) = 40$

$$\sigma = \sqrt{40} = 6.3246$$

$$P(x < 40) = P(x \leq 39.5)$$

$$z = \frac{x - \mu}{\sigma} = \frac{39.5 - 50}{6.3246} = -1.66, \text{ 面积} = 0.0485$$

$$P(x \leq 39.5) = 0.0485$$

c. $P(55 \leq x \leq 60) = P(54.5 \leq x \leq 60.5)$

$$z = \frac{x - \mu}{\sigma} = \frac{54.5 - 50}{6.3246} = 0.71, \text{ 面积} = 0.7611$$

$$z = \frac{x - \mu}{\sigma} = \frac{60.5 - 50}{6.3246} = 1.66, \text{ 面积} = 0.9515$$

$$P(54.5 \leq x \leq 60.5) = 0.9515 - 0.7611 = 0.1904$$

d. $P(x \geq 70) = P(x \geq 69.5)$

$$z = \frac{x - \mu}{\sigma} = \frac{69.5 - 50}{6.3246} = 3.08, \text{ 面积} = 0.9990$$

$$P(x \geq 69.5) = 1 - 0.9990 = 0.0010$$

30. a. 144

b. 0.1841

c. 0.9943

32. a. 0.5276

b. 0.3935

c. 0.4724

d. 0.1341

33. a. $P(x \leq x_0) = 1 - e^{-x_0/3}$

b. $P(x \leq 2) = 1 - e^{-2/3} = 1 - 0.5134 = 0.4866$

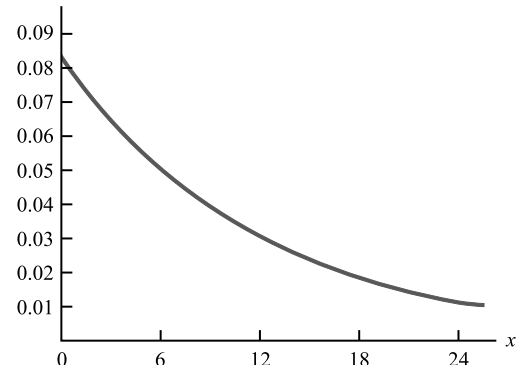
c. $P(x \geq 3) = 1 - P(x \leq 3) \\ = 1 - (1 - e^{-3/3}) \\ = e^{-1} = 0.3679$

d. $P(x \leq 5) = 1 - e^{-5/3} = 1 - 0.1889 = 0.8111$

e. $P(2 \leq x \leq 5) = P(x \leq 5) - P(x \leq 2) \\ = 0.8111 - 0.4866 = 0.3245$

34. a. $f(x) = 0.05e^{-0.05x}$

b. 0.5276

- c. 0.367 9
d. 0.510 5
35. a. $f(x)$
- 
- b. $P(x \leq 12) = 1 - e^{-12/12} = 1 - 0.367 9 = 0.632 1$
c. $P(x \leq 6) = 1 - e^{-6/12} = 1 - 0.606 5 = 0.393 5$
d. $P(x \geq 30) = 1 - P(x < 30) = 1 - (1 - e^{-30/12}) = 0.082 1$
36. a. $P(x \leq 1) = 1 - e^{-1/2} = 1 - e^{-0.5} = 1 - 0.606 5 = 0.393 5$
b. $P(x \leq 2) = 1 - e^{-2/2} = 1 - e^{-1.0} = 1 - 0.367 9 = 0.632 1$
 $P(1 \leq x \leq 2) = P(x \leq 2) - P(x \leq 1) = 0.632 1 - 0.367 9 = 0.264 2$
c. 该用户的电缆维修服务时间将超过 4 小时。
 $P(x > 4) = 1 - P(x \leq 4) = 1 - (1 - e^{-4/2}) = e^{-2.0} = 0.135 3$
38. a. 两个电话之间的时间间隔服从指数分布, $\mu = \frac{60 \text{ 分钟/小时}}{1.6 \text{ 个/小时}} = 37.5 \text{ 分钟/个}$
b. 指数分布密度函数 $f(x) = \frac{1}{37.5} e^{-x/37.5}$, $x \geq 0$, 其中 x 是两个 911 电话的时间间隔。
c. $P(x < 60) = 1 - e^{-60/37.5} = 1 - 0.201 9 = 0.798 1$
d. $P(x \geq 30) = 1 - P(x \leq 30) = 1 - (1 - e^{-30/37.5}) = 1 - 0.550 7 = 0.449 3$
e. $P(5 \leq x \leq 20) = (1 - e^{-20/37.5}) - (1 - e^{-5/37.5}) = 0.413 4 - 0.124 8 = 0.288 6$
40. a. 16 312 美元
b. 7.64%
c. 22 948 美元
42. a. $\sigma = 25.531 9$
b. 0.940 1
c. 706 或者更多
44. a. 0.022 8
b. 50 美元
46. a. 38.3%
b. 3.59% 的人的分数较好, 96.41% 的人的分数较差
c. 38.21%
48. $\mu = 19.23$ 盎司
50. a. 损失 240 美元

- b. 0.178 8
c. 0.355 7
d. 0.059 4

52. a. 1/7 分钟
b. $7e^{-7x}$
c. 0.000 9
d. 0.246 6
54. a. 2 分钟
b. 0.221 2
c. 0.393 5
d. 0.082 1

第 7 章

1. a. AB, AC, AD, AE, BC, BD, BE, CD, CE, DE
b. 10 个样本, 每个出现的概率为 1/10
c. E 和 C, 因为 8 和 0 没有用; 5 对应于 E; 7 没有用; 5 已经被选入样本, 所以跳过; 3 对应于 C; 样本容量 2 已经满足, 所以就不必再选 2 所对应的 B 了。
2. 22, 147, 229, 289
3. 459, 147, 385, 113, 340, 401, 215, 2, 33, 348
4. a. 美国南方贝尔, 美国 LSI Logic, 美国通用电气
b. 120
6. 2 782, 493, 825, 1 807, 289
8. 埃克森美孚公司, 雪弗龙公司, Travelers 保险公司, 微软, 辉瑞制药, 英特尔
10. a. 有限总体; b. 无限总体; c. 无限总体; d. 有限总体; e. 无限总体;
11. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{54}{6} = 9$
b. $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$
 $\sum (x_i - \bar{x})^2 = (-4)^2 + (-1)^2 + 1^2 + (-2)^2 + 2^2 + 5^2 = 48$
 $s = \sqrt{\frac{48}{6-1}} = 3.1$
12. a. 0.50
b. 0.366 7
13. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{465}{5} = 93$

	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
	94	+1	1
	100	+7	49
	85	-8	64
	94	+1	1
	92	-1	1
合计	465	0	116

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{116}{4}} = 5.39$$

14. a. 0.05

b. 0.425

c. 0.20

16. a. 50 岁及以上的所有美国人

b. 0.821 6

c. 315

d. 0.831 0

e. 目标总体与抽样总体一致; 若限定为 AARP 成员, 则推断将会颇有争议。

18. a. 200

b. 5

c. 正态概率分布, $E(\bar{x}) = 200$, $\sigma_{\bar{x}} = 5$ d. \bar{x} 的概率分布

19. a. 抽样分布服从正态分布:

$$E(\bar{x}) = \mu = 200$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{100}} = 5$$

对于 ± 5 , $195 \leq \bar{x} \leq 205$

根据标准正态概率表,

当 $\bar{x} = 205$ 时,

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{5}{5} = 1$$

$$P(z \leq 1) = 0.841 3$$

当 $\bar{x} = 195$ 时,

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{-5}{5} = -1$$

$$P(z \leq -1) = 0.158 7$$

$$P(195 \leq \bar{x} \leq 205) = 0.841 3 - 0.158 7 = 0.682 6$$

b. 对于 ± 10 , $190 \leq \bar{x} \leq 210$

根据标准正态概率表,

当 $\bar{x} = 210$ 时,

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{10}{5} = 2$$

$$P(z \leq 2) = 0.977 2$$

当 $\bar{x} = 190$ 时,

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{-10}{5} = -2$$

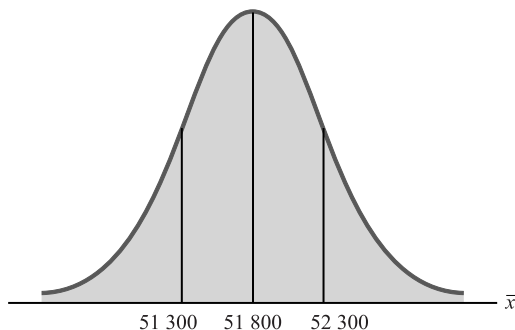
$$P(z \leq -2) = 0.022 8$$

$$P(190 \leq \bar{x} \leq 210) = 0.977 2 - 0.022 8 = 0.954 4$$

20. 3.54, 2.50, 2.04, 1.77

随着 n 的增加, $\sigma_{\bar{x}}$ 减小22. a. 服从正态概率分布, $E(\bar{x}) = 51 800$ 和 $\sigma_{\bar{x}} = 516.40$ b. $\sigma_{\bar{x}}$ 减小到 365.15c. 随着 n 的增加, $\sigma_{\bar{x}}$ 减小

23. a.



$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4 000}{\sqrt{60}} = 516.40$$

$$\text{当 } \bar{x} = 52 300 \text{ 时, } z = \frac{52 300 - 51 800}{516.40} = 0.97$$

$$P(\bar{x} \leq 52 300) = P(z \leq 0.97) = 0.834 0$$

$$\text{当 } \bar{x} = 51 300 \text{ 时, } z = \frac{51 300 - 51 800}{516.40} = -0.97$$

$$P(\bar{x} < 51 300) = P(z < -0.97) = 0.166 0$$

$$P(51 300 \leq \bar{x} \leq 52 300) = 0.834 0 - 0.166 0 = 0.668 0$$

$$\text{b. } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4 000}{\sqrt{120}} = 365.15$$

$$\text{当 } \bar{x} = 52 300 \text{ 时, } z = \frac{52 300 - 51 800}{365.15} = 1.37$$

$$P(\bar{x} \leq 52 300) = P(z \leq 1.37) = 0.914 7$$

$$\text{当 } \bar{x} = 51 300 \text{ 时, } z = \frac{51 300 - 51 800}{365.15} = -1.37$$

$$P(\bar{x} < 51 300) = P(z < -1.37) = 0.085 3$$

$$P(51 300 \leq \bar{x} \leq 52 300) = 0.914 7 - 0.085 3 = 0.829 4$$

24. a. 服从正态概率分布, $E(\bar{x}) = 17.5$ 和 $\sigma_{\bar{x}} = 0.57$

b. 0.919 8

c. 0.621 2

26. a. 0.354 4, 0.444 8, 0.593 4, 0.904 4

b. 样本均值与总体均值越接近, 概率越高

28. a. 服从正态概率分布, $E(\bar{x}) = 22$ 和 $\sigma_{\bar{x}} = 0.730 3$

b. 0.829 4

c. 0.907 0

d. (c) 部分纽约州, 因为样本容量更大。

30. a. $n/N = 0.01$, 不能

b. 1.29, 1.30, 几乎没有差别

c. 0.876 4

32. a. $E(\bar{p}) = 0.40$

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.40 \times 0.60}{200}} = 0.034 6$$

对于 ± 0.03 , 意味着 $0.37 \leq \bar{p} \leq 0.43$

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{0.03}{0.034 6} = 0.87$$

$$P(0.37 \leq \bar{p} \leq 0.43) = P(-0.87 \leq z \leq 0.87)$$

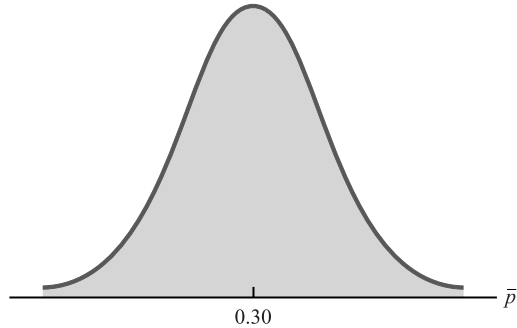
$$= 0.807 8 - 0.192 2 = 0.615 6$$

$$b. z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{0.05}{0.0346} = 1.44$$

$$P(0.35 \leq \bar{p} \leq 0.45) = P(-1.44 \leq z \leq 1.44) \\ = 0.9251 - 0.0749 = 0.8502$$

34. a. 0.6156
b. 0.7814
c. 0.9488
d. 0.9942
e. n 越大, 概率越大

35. a.



$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.30 \times 0.70}{100}} = 0.0458$$

因为 $np = 100 \times 0.30 = 30$ 和 $n(1-p) = 100 \times 0.70 = 70$ 均大于 5, 所以采用正态分布是合适的。

- b. $P\{0.20 \leq \bar{p} \leq 0.40\} = ?$

$$z = \frac{0.040 - 0.030}{0.0458} = 2.18$$

$$P(0.20 \leq \bar{p} \leq 0.40) = P(-2.18 \leq z \leq 2.18) \\ = 0.9854 - 0.0146 = 0.9708$$

c. $P\{0.25 \leq \bar{p} \leq 0.35\} = ?$

$$z = \frac{0.035 - 0.030}{0.0458} = 1.09$$

$$P(0.25 \leq \bar{p} \leq 0.35) = P(-1.09 \leq z \leq 1.09) \\ = 0.8621 - 0.1379 = 0.7242$$

36. a. 服从 $E(\bar{p}) = 0.55$ 和 $\sigma_{\bar{p}} = 0.0352$ 的正态概率分布
b. 0.8444
c. 服从 $E(\bar{p}) = 0.45$ 和 $\sigma_{\bar{p}} = 0.0352$ 的正态概率分布
d. 0.8444
e. 相同, 因为两部分中的标准误差是相同的。
f. 0.9556, 由于样本容量增加降低了标准误, 使得概率增加。
38. a. 服从 $E(\bar{p}) = 0.42$ 和 $\sigma_{\bar{p}} = 0.0285$ 的正态概率分布
b. 0.7062
c. 0.9198
d. 概率将增大
40. a. 服从 $E(\bar{p}) = 0.76$ 和 $\sigma_{\bar{p}} = 0.0214$ 的正态概率分布
b. 0.8384
c. 0.9452

42. a. LMI Aerospace, Alpha & Omega Semiconductor, Olympic Steel, Kimball International, International Shipholding
b. 抽取与 (a) 中同样的企业是极其不可能的。
44. a. 服从 $E(\bar{x}) = 406$ 和 $\sigma_{\bar{x}} = 10$ 的正态概率分布
b. 0.8664
c. $z = -2.60, 0.0047$, 是的
46. a. 955
b. 0.50
c. 0.7062
d. 0.8230
48. a. 625
b. 0.7888
50. a. 服从 $E(\bar{p}) = 0.15$ 和 $\sigma_{\bar{p}} = 0.0230$ 的正态概率分布
b. 0.9182
c. 0.6156
52. a. 0.8882
b. 0.0233
54. a. 48
b. 正态概率分布, $E(\bar{p}) = 0.25$, $\sigma_{\bar{p}} = 0.0625$
c. 0.2119

第 8 章

2. 利用 $\bar{x} \pm z_{\alpha/2}(\sigma/\sqrt{n})$ 得
a. $32 \pm 1.645(6/\sqrt{50})$
 $32 \pm 1.4; (30.6, 33.4)$
b. $32 \pm 1.96(6/\sqrt{50})$
 $32 \pm 1.66; (30.34, 33.66)$
c. $32 \pm 2.576(6/\sqrt{50})$
 $32 \pm 2.19; (29.81, 34.19)$
4. 54
5. a. 在 99% 的置信度下, $z_{\alpha/2} = z_{0.005} = 2.576$, 边际误差是
 $2.576\sigma/\sqrt{n} = 2.576(6 \times \sqrt{64}) = 1.93$
b. 置信区间为 21.52 ± 1.93 或者 $(19.59, 23.45)$ 。
6. $(39.13, 41.49)$
8. a. 总体分布至少近似服从正态分布
b. 3.41
c. 4.48
10. a. $(3388, 3584)$
b. $(3370, 3602)$
c. $(3333, 3639)$
d. 随着置信度的增大, 置信区间的宽度增加
12. a. 2.179
b. -1.676
c. 2.457

- d. $(-1.708, 1.708)$
e. $(-2.014, 2.014)$
13. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{80}{8} = 10$
b. $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{84}{7}} = 3.464$
c. $t_{0.025} \left(\frac{s}{\sqrt{n}} \right) = 2.365 \times \frac{3.46}{\sqrt{8}} = 2.9$
d. $\bar{x} \pm t_{0.025} \left(\frac{s}{\sqrt{n}} \right)$
 $10 \pm 2.9; (7.1, 12.9)$
14. a. $(21.5, 23.5)$
b. $(21.3, 23.7)$
c. $(20.9, 24.1)$
d. 更大的边际误差和更宽的区间
15. $\bar{x} \pm t_{\alpha/2} (s/\sqrt{n})$
置信度为 90%，自由度 = 64， $t_{0.05} = 1.669$
 $19.5 \pm 1.669 \times \frac{5.2}{\sqrt{65}}$
 19.5 ± 1.08 或者 $(18.42, 20.58)$
置信度为 95%，自由度 = 64， $t_{0.025} = 1.998$
 $19.5 \pm 1.998 \times \frac{5.2}{\sqrt{65}}$
 19.5 ± 1.29 或者 $(18.21, 20.79)$
16. a. 9.706 3, 7.980 5
b. $(7.153\ 6, 12.259\ 0)$
c. 3.885 4, 1.619 4
d. $(3.367\ 4, 4.403\ 4)$
18. a. 22 周
b. 3.802 0
c. $(18.20, 25.80)$
d. 下次需要更大的 n
20. a. $\bar{x} = 2\ 551$
b. $(2\ 409.99, 2\ 692.01)$
c. 我们以 95% 的把握认为密歇根州每年汽车保费的均值高于全美平均水平。
22. a. $(21\ 613, 24\ 587)$
b. 2 848
c. 观众总人数 = 11 619 840 人，票房销售总收入 = 94 248 000 美元
24. a. σ 的计划值 = $\frac{\text{极差}}{4} = \frac{36}{4} = 9$
b. $n = \frac{z_{0.025}^2 \sigma^2}{E^2} = \frac{(1.96)^2 (9)^2}{(3)^2} = 34.57$; 取 $n = 35$
c. $n = \frac{(1.96)^2 (9)^2}{(2)^2} = 77.79$; 取 $n = 78$
25. 由 $n = \frac{z_{\alpha/2}^2 \sigma^2}{E^2}$ 得 $n = \frac{(1.96)^2 (6.84)^2}{(1.5)^2} = 79.88$; 取 $n = 80$
 $n = \frac{(1.645)^2 (6.84)^2}{(2)^2} = 31.65$; 取 $n = 32$
26. a. 25
b. 49
c. 97
28. a. 188
b. 267
c. 461
d. 随着置信度的增大，样本容量变得更大；不建议采用 99% 的置信水平
30. 1 537
31. a. $\bar{p} = \frac{100}{400} = 0.25$
b. $\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\frac{0.25 \times 0.75}{400}} = 0.021\ 7$
c. $\bar{p} \pm z_{0.025} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$
 $0.25 \pm 1.96 \times 0.021\ 7$
 $0.25 \pm 0.042\ 4$; $(0.207\ 6, 0.292\ 4)$
32. a. $(0.673\ 3, 0.726\ 7)$
b. $(0.668\ 2, 0.731\ 8)$
34. 1 068
35. a. $\bar{p} = \frac{1\ 760}{2\ 000} = 0.88$
b. 边际误差
 $z_{0.05} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 1.645 \times \sqrt{\frac{0.88 \times (1-0.88)}{2\ 000}} = 0.012\ 0$
c. 置信区间为 $0.88 \pm 0.012\ 0$ ，即 $(0.868, 0.892)$
d. 边际误差
 $z_{0.025} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 1.96 \sqrt{\frac{0.88 \times (1-0.88)}{2\ 000}} = 0.014\ 2$
95% 置信区间为 $0.88 \pm 0.014\ 2$ ，即 $(0.865\ 8, 0.894\ 2)$
36. a. 0.23
b. $(0.171\ 6, 0.288\ 4)$
38. a. 0.570 4
b. 0.081 4; $(0.489\ 0, 0.651\ 8)$
c. 1 046
39. a. $n = \frac{z_{0.025}^2 p^* (1-p^*)}{E^2} = \frac{(1.96)^2 \times 0.156 \times (1-0.156)}{0.03^2} = 562$
b. $n = \frac{z_{0.005}^2 p^* (1-p^*)}{E^2} = \frac{(2.576)^2 \times 0.156 \times (1-0.156)}{0.03^2} = 970.77$;

取 $n = 971$

40. 0.034 6, (0.485 4, 0.554 6)
42. a. 0.044 2
b. 601, 1 068, 2 401, 9 604
44. a. 4.00
b. (29.77, 37.77)
46. a. 122
b. (1 751, 1 995)
c. 1 723.16 亿美元
d. 少于 1 873 美元
48. a. (712.27 美元, 833.73 美元)
b. (172.31 美元, 201.69 美元)
c. 0.34
d. (a) 中为 60.73; (a) 中边际误差更大, 因为 (a) 中样本标准差更大。
50. 37
52. 176
54. a. 0.542 0
b. 0.050 8
c. (0.491 2, 0.592 8)
56. a. 0.22
b. (0.190 4, 0.249 6)
c. (0.384 7, 0.455 3)
d. (c) 中的置信区间更宽; 在 (c) 中样本比率接近于 0.5。
58. a. 1 267
b. 1 509
60. a. 0.310 1
b. (0.289 8, 0.330 4)
c. 8 219; 不行, 样本容量不必这么大

第 9 章

2. a. $H_0: \mu \leq 14$
 $H_a: \mu > 14$
b. 没有证据能够证明新计划增加了销售量
c. 支持研究中的假设 $\mu > 14$; 认为新计划使销售量增加
4. a. $H_0: \mu \geq 220$
 $H_a: \mu < 220$
b. 不能得到结论认为所推荐的方法能够降低成本
c. 得到结论认为所推荐的方法能够降低成本
5. a. 得到结论认为芝加哥周边地区每月用电成本的总体均值大于 104 美元, 即高于辛辛那提相应周边地区的用电成本。
b. 第一类错误是指, 当 H_0 为真时拒绝它; 当研究者得出结论认为芝加哥周边地区每月用电成本的总体均值大于 104 美元, 而实际上用电成本的总体均值等于或者小于 104 美元时, 发生这类错误。
- c. 第二类错误是指, 当 H_0 为假时却接受了它; 当研究者得出结论认为芝加哥周边地区每月用电成本的总体均值等于或者小于 104 美元, 而实际上并非如此, 这时发生第二类错误。
6. a. $H_0: \mu \leq 1$
 $H_a: \mu > 1$
b. 认为 $\mu > 1$, 而实际情况并非如此
c. 认为 $\mu \leq 1$, 而实际情况并非如此
8. a. $H_0: \mu \geq 220$
 $H_a: \mu < 220$
b. 认为 $\mu < 220$, 而实际情况并非如此
c. 认为 $\mu \geq 220$, 而实际情况并非如此
10. a. $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{26.4 - 25}{6/\sqrt{40}} = 1.48$
b. 根据正态分布表有 $z = 1.48$, 于是
 p -值 $= 1.000 0 - 0.930 6 = 0.069 4$
c. p -值 > 0.01 , 不能拒绝 H_0
d. 如果 $z \geq 2.33$, 则拒绝 H_0
 $1.48 < 2.33$, 不能拒绝 H_0
11. a. $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{14.15 - 15}{3/\sqrt{50}} = -2.00$
b. p -值 $= 2 \times 0.022 8 = 0.045 6$
c. p -值 ≤ 0.05 , 拒绝 H_0
d. 如果 $z \leq -1.96$ 或者 $z \geq 1.96$, 则拒绝 H_0
因为 $-2.00 < -1.96$, 所以拒绝 H_0
12. a. 0.105 6, 不能拒绝 H_0
b. 0.006 2, 拒绝 H_0
c. ≈ 0 , 拒绝 H_0
d. 0.796 7, 不能拒绝 H_0
14. a. 0.384 4, 不能拒绝 H_0
b. 0.007 4, 拒绝 H_0
c. 0.083 6, 不能拒绝 H_0
15. a. $H_0: \mu \geq 1 056$
 $H_a: \mu < 1 056$
b. $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{910 - 1 056}{1 600/\sqrt{400}} = -1.83$
 p -值 $= 0.033 6$
c. p -值 ≤ 0.05 , 拒绝 H_0 ; “最后一分钟”纳税人所交税款的均值少于 1 056 美元。
d. 如果 $z \leq -1.645$, 则拒绝 H_0
因为 $-1.83 < -1.645$, 所以拒绝 H_0
16. a. $H_0: \mu \leq 3 173$

- $H_a: \mu > 3.173$
 b. 0.020 7
 c. 拒绝 H_0 , 得到结论认为毕业生信用卡余额的均值增加了。
18. a. $H_0: \mu = 192$
 $H_a: \mu \neq 192$
 b. -2.23, 0.025 8
 c. 拒绝 H_0 ; 2012 年, 年轻的千禧一代在餐馆就餐的平均次数已经发生了变化。
20. a. $H_0: \mu \geq 838$
 $H_a: \mu < 838$
 b. -2.40
 c. 0.008 2
 d. 拒绝 H_0 ; 结论是中西部人均处方药的年支出费用低于东北部。
22. a. $H_0: \mu = 8$
 $H_a: \mu \neq 8$
 b. 0.170 6
 c. 不能拒绝 H_0 ; 不能得出结论认为等待时间的均值与 8 分钟有差异。
 d. (7.83, 8.97); 是
24. a. $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{17 - 18}{4.5/\sqrt{48}} = -1.54$
 b. 自由度 = $n - 1 = 47$
 下侧面积介于 0.05 和 0.10 之间
 p -值 (双侧) 介于 0.10 和 0.20 之间
 精确的 p -值 = 0.130 3
 c. p -值 > 0.05, 不能拒绝 H_0
 d. 自由度 = 47, $t_{0.025} = 2.012$
 如果 $t \leq -2.012$ 或者 $t \geq 2.012$, 则拒绝 H_0
 因为 $t = -1.54$, 所以不能拒绝 H_0
26. a. 介于 0.02 和 0.05 之间; 精确的 p -值 = 0.039 7; 拒绝 H_0
 b. 介于 0.01 和 0.02 之间; 精确的 p -值 = 0.012 5; 拒绝 H_0
 c. 介于 0.10 和 0.20 之间; 精确的 p -值 = 0.128 5; 不能拒绝 H_0
 d. 自由度 = 99, $t_{0.05} = -1.66$
 如果 $t \leq -1.66$, 则拒绝 H_0
- 因为 $-1.45 > -1.66$, 所以不能拒绝 H_0
28. a. $H_0: \mu \leq 9$
 $H_a: \mu > 9$
 b. 介于 0.005 和 0.01 之间
 精确的 p -值 = 0.007 2;
 c. 拒绝 H_0 ; 首席执行官 (CEO) 任期的均值低于 9 年。
30. a. $H_0: \mu = 6.4$
 $H_a: \mu \neq 6.4$
 b. 介于 0.10 和 0.20 之间
 精确的 p -值 = 0.126 8
 c. 不能拒绝 H_0 ; 不能得出结论认为小组的说法有误。
32. a. $H_0: \mu = 10.192$
 $H_a: \mu \neq 10.192$
 b. 介于 0.02 和 0.05 之间
 精确的 p -值 = 0.030 4
 c. 拒绝 H_0 ; 二手车市场的价格均值与全国的均值有差异。
34. a. $H_0: \mu = 2$
 $H_a: \mu \neq 2$
 b. 2.2
 c. 0.516
 d. 介于 0.20 和 0.40 之间
 精确的 p -值 = 0.253 5
 e. 不能拒绝 H_0 ; 出于估计成本的目的, 没有理由认为植树所需平均时间不是 2 小时。
36. a. $z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.68 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{300}}} = -2.80$
 p -值 = 0.002 6
 p -值 ≤ 0.05 , 拒绝 H_0
 b. $z = \frac{0.72 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{300}}} = -1.20$
 p -值 = 0.115 1
 p -值 > 0.05, 不能拒绝 H_0
 c. $z = \frac{0.70 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{300}}} = -2.00$
 p -值 = 0.022 8
 p -值 ≤ 0.05 , 拒绝 H_0
 d. $z = \frac{0.77 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{300}}} = 0.80$
 p -值 = 0.788 1
 p -值 > 0.05, 不能拒绝 H_0
38. a. $H_0: p = 0.64$

$$H_a: p \neq 0.64$$

$$b. \bar{p} = 52/100 = 0.52$$

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.52 - 0.64}{\sqrt{\frac{0.64(1-0.64)}{100}}} = -2.50$$

$$p\text{-值} = 2 \times 0.0062 = 0.0124$$

$$c. p\text{-值} \leq 0.05, \text{拒绝 } H_0$$

比率与报告的 0.64 有差异。

$$d. \text{是的, 由于 } \bar{p} = 0.52, \text{说明几乎没有人认为超市的牌子与大众名牌是一样的。}$$

$$40. a. 21$$

$$b. H_0: p \geq 0.46$$

$$H_a: p < 0.46$$

$$p\text{-值} \approx 0.0436$$

$$c. \text{是, } 0.0436$$

$$42. a. \bar{p} = 0.15$$

$$b. (0.718, 0.2282)$$

$$c. \text{休斯敦百货公司的退货比率与全国的退货比率存在显著差异。}$$

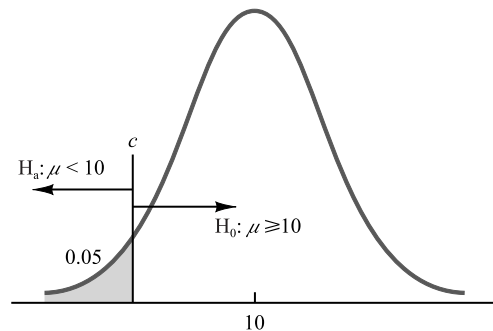
$$44. a. H_0: p \leq 0.50$$

$$H_a: p > 0.50$$

$$b. \bar{p} = 0.6133; p\text{-值} = 0.0027$$

$$c. \text{拒绝 } H_0; \text{年龄超过 55 岁的医生至少被起诉 1 次的比率更高。}$$

$$46.$$



$$c = 10 - 1.645(5/\sqrt{120}) = 9.25$$

$$\text{如果 } \bar{x} \leq 9.25, \text{则拒绝 } H_0$$

$$a. \text{当 } \mu = 9 \text{ 时,}$$

$$z = \frac{9.25 - 9}{5/\sqrt{120}} = 0.55$$

$$P(\text{拒绝 } H_0) = 1.0000 - 0.7088 = 0.2912$$

$$b. \text{第二类错误}$$

$$c. \text{当 } \mu = 8 \text{ 时,}$$

$$z = \frac{9.25 - 8}{5/\sqrt{120}} = 2.74$$

$$\beta = 1.0000 - 0.9969 = 0.0031$$

$$48. a. \text{得到 } \mu \leq 15 \text{ 的结论, 而实际情况并非如此}$$

$$b. 0.2676$$

$$c. 0.0179$$

$$49. a. H_0: \mu \geq 25$$

$$H_a: \mu < 25$$

$$\text{如果 } z \leq -2.05, \text{则拒绝 } H_0$$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 25}{3/\sqrt{30}} = -2.05$$

$$\text{解得 } \bar{x} = 23.88$$

$$\text{决策规则为如果 } \bar{x} > 23.88, \text{则接受 } H_0; \text{如果 } \bar{x} \leq 23.88, \text{则拒绝 } H_0.$$

$$b. \text{对于 } \mu = 23,$$

$$z = \frac{23.88 - 23}{3/\sqrt{30}} = 1.61$$

$$\beta = (1.0000 - 0.9463) = 0.0537$$

$$c. \text{对于 } \mu = 24,$$

$$z = \frac{23.88 - 24}{3/\sqrt{30}} = -0.22$$

$$\beta = (1.0000 - 0.4129) = 0.5871$$

$$d. \text{这种情形下不会发生第二类错误。注意, 当 } \mu = 25.5 \text{ 时, } H_0 \text{ 为真; 当 } H_0 \text{ 为假时才可能会发生第二类错误。}$$

$$50. a. \text{得到 } \mu = 28 \text{ 的结论, 而实际情况并非如此。}$$

$$b. 0.0853, 0.6179, 0.6179, 0.0853$$

$$c. 0.9147$$

$$52. 0.1151, 0.0015$$

$$\text{增大 } n \text{ 以减小 } \beta$$

$$54. n = \frac{(z_\alpha + z_\beta)^2 \cdot \sigma^2}{(\mu_0 - \mu_a)^2} = \frac{(1.645 + 1.28)^2 \cdot 5^2}{(10 - 9)^2} = 214$$

$$56. 109$$

$$57. \text{当 } \mu_0 = 400 \text{ 时, } \alpha = 0.02; z_{0.02} = 2.05$$

$$\text{当 } \mu_a = 385 \text{ 时, } \beta = 0.10; z_{0.10} = 1.28$$

$$\text{由 } \sigma = 30, \text{有}$$

$$n = \frac{(z_\alpha + z_\beta)^2 \cdot \sigma^2}{(\mu_0 - \mu_a)^2} = \frac{(2.05 + 1.28)^2 \cdot 30^2}{(400 - 385)^2} = 44.4 \text{ 或者 } 45$$

$$58. 324$$

$$60. a. H_0: \mu = 16$$

$$H_a: \mu \neq 16$$

$$b. 0.0286; \text{拒绝 } H_0, \text{重新调整生产线}$$

$$c. 0.2186; \text{不能拒绝 } H_0, \text{继续运行}$$

$$d. z = 2.19; \text{拒绝 } H_0$$

$$z = -1.23; \text{不能拒绝 } H_0$$

$$\text{是的, 相同的结论}$$

$$62. a. H_0: \mu \leq 4$$

$$H_a: \mu > 4$$

$$b. p\text{-值} = 0.0049$$

- c. 拒绝 H_0 ; 得出结论认为低收入家庭孩子每天置身于电视背景下的平均时间多于 4 小时。
64. $H_0: \mu \leq 30.8$
 $H_a: \mu > 30.8$
 $t = -1.0894$
 p -值介于 0.10 和 0.20 之间
 精确的 p -值 = 0.1408
 不能拒绝 H_0 ; 没有证据支持得出结论认为最近英国男性结婚的平均年龄比 2013 年高。
66. $t = 2.26$
 p -值介于 0.01 和 0.025 之间
 精确的 p -值 = 0.0155
 拒绝 H_0 ; 平均成本高于 125 000 美元。
68. a. $H_0: p \leq 0.80$
 $H_a: p > 0.80$
 得到结论认为乘客的安全感提高了。
 b. 不能拒绝 H_0 ; 不推荐强制使用
70. a. $H_0: p \leq 0.30$
 $H_a: p > 0.30$
 b. 0.34
 c. p -值 = 0.0401
 d. 拒绝 H_0 ; 30% 的千禧一代要么与父母同住, 要么依靠父母。
72. $H_0: p \geq 0.90$
 $H_a: p < 0.90$
 p -值 = 0.0808, 不能拒绝 H_0 ; 不能拒绝“至少 90%”的说法。
74. a. $H_0: \mu \leq 72$
 $H_a: \mu > 72$
 b. 0.2912
 c. 0.7939
 d. 0, 因为 H_0 是成立的
76. a. 45
 b. 0.0192, 0.2358, 0.7291, 0.7291, 0.2358, 0.0192

第 10 章

1. a. $\bar{x}_1 - \bar{x}_2 = 13.6 - 11.6 = 2$
 b. $z_{\alpha/2} = z_{0.05} = 1.645$

$$\bar{x}_1 - \bar{x}_2 \pm 1.645 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$2 \pm 1.645 \sqrt{\frac{2.2^2}{50} + \frac{3^2}{35}}$$

$$2 \pm 0.98; (1.02, 2.98)$$
 c. $z_{\alpha/2} = z_{0.025} = 1.96$
2. a. $z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(25.2 - 22.8) - 0}{\sqrt{\frac{5.2^2}{40} + \frac{6^2}{50}}} = 2.03$
 b. p -值 = $1.0000 - 0.9788 = 0.0212$
 c. p -值 ≤ 0.05 , 拒绝 H_0
4. a. $\bar{x}_1 - \bar{x}_2 = 85.36 - 81.40 = 3.96$
 b. $z_{0.025} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1.96 \times \sqrt{\frac{4.55^2}{37} + \frac{3.97^2}{44}} = 1.88$
 c. $3.96 \pm 1.88; (2.08, 5.84)$
6. p -值 = 0.0351
 拒绝 H_0 ; 亚特兰大的平均价格低于休斯敦的平均价格。
8. a. 拒绝 H_0 ; Rite Aid 公司改善了客户服务质量。
 b. 不能拒绝 H_0 ; 统计上没有显著差异。
 c. p -值 = 0.0336; 拒绝 H_0 ; Expedia 公司改善了客户服务质量。
 d. 1.80
 e. J. C. Penny 公司的满意度得分在统计上没有显著增加。
9. a. $\bar{x}_1 - \bar{x}_2 = 22.5 - 20.1 = 2.4$
 b. $df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{2.5^2}{20} + \frac{4.8^2}{30}\right)^2}{\frac{1}{19} \left(\frac{2.5^2}{20}\right)^2 + \frac{1}{29} \left(\frac{4.8^2}{30}\right)^2} = 45.8$
 c. $df = 45, t_{0.025} = 2.014$

$$t_{0.025} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.014 \sqrt{\frac{2.5^2}{20} + \frac{4.8^2}{30}} = 2.1$$

 d. $2.4 \pm 2.1; (0.3, 4.5)$
10. a. $t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(13.6 - 10.1) - 0}{\sqrt{\frac{5.2^2}{35} + \frac{8.5^2}{40}}} = 2.18$
 b. $df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{5.2^2}{35} + \frac{8.5^2}{40}\right)^2}{\frac{1}{34} \left(\frac{5.2^2}{35}\right)^2 + \frac{1}{39} \left(\frac{8.5^2}{40}\right)^2} = 65.7$
 取 $df = 65$
 c. $df = 65$, 上侧面积介于 0.01 和 0.025 之间; 双侧 p -值介于 0.02 和 0.05 之间
 精确的 p -值 = 0.0329
 d. p -值 ≤ 0.05 ; 拒绝 H_0
12. a. $\bar{x}_1 - \bar{x}_2 = 22.5 - 18.6 = 3.9$ 英里

$$b. df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{8.4^2}{50} + \frac{7.4^2}{40}\right)^2}{\frac{1}{49}\left(\frac{8.4^2}{50}\right)^2 + \frac{1}{39}\left(\frac{7.4^2}{40}\right)^2} = 87.1$$

取 $df=87$, $t_{0.025}=1.988$

$$3.9 \pm 1.988 \sqrt{\frac{8.4^2}{50} + \frac{7.4^2}{40}}$$

$$3.9 \pm 3.3; (0.6, 7.2)$$

14. a. $H_0: \mu_1 - \mu_2 \geq 0$

$H_a: \mu_1 - \mu_2 < 0$

b. -2.41

c. 利用 t 分布表, p -值介于 0.005 和 0.01 之间

精确的 p -值 = 0.009

d. 拒绝 H_0 ; 坦帕的护士工资较低

16. a. $H_0: \mu_1 - \mu_2 \leq 0$

$H_a: \mu_1 - \mu_2 > 0$

b. 38

c. $t=1.80$, $df=25$

利用 t 分布表, p -值介于 0.025 和 0.05 之间, 精确的 p -值 = 0.042 0

d. 拒绝 H_0 ; 结论是如果大学毕业, 平均分数更高。

18. a. $H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0$

b. 50.6 分和 52.8 分

c. p -值 > 0.40

不能拒绝 H_0 ; 不能断定, 总体均值的延迟时间会有所不同。

19. a. 1, 2, 0, 0, 2

b. $\bar{d} = \sum d_i / n = 5/5 = 1$

c. $s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{4}{5-1}} = 1$

d. $t = \frac{\bar{d} - \mu}{s_d / \sqrt{n}} = \frac{1-0}{1/\sqrt{5}} = 2.24$

$df = n - 1 = 4$

利用 t 分布表, p -值介于 0.025 和 0.05 之间

精确的 p -值 = 0.044 3

p -值 ≤ 0.05 ; 拒绝 H_0

20. a. 3, -1, 3, 5, 3, 0, 1

b. 2

c. 2.08

d. 2

e. (0.07, 3.93)

21. $H_0: \mu_d \leq 0$

$H_a: \mu_d > 0$

$\bar{d} = 0.625$, $s_d = 1.30$

$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{0.625 - 0}{1.30 / \sqrt{8}} = 1.36$

$df = n - 1 = 7$

利用 t 分布表, p -值介于 0.10 和 0.20 之间

精确的 p -值 = 0.108 0

p -值 > 0.05; 不能拒绝 H_0 ; 不能得出广告使平均潜在购买力得到提高的结论。

22. a. 3.41 美元

b. (1.67 美元, 5.15 美元)

有了很不错的增长

24. a. $H_0: \mu_d \leq 0$

$H_a: \mu_d > 0$

$\bar{d} = 23$, $t = 2.05$

p -值介于 0.025 和 0.05 之间

拒绝 H_0 ; 可以得出机票的平均价格有了一个显著增长的结论。

b. 487 美元, 464 美元

c. 机票价格增长 5%

26. a. $t = -1.42$

利用 t 分布表, p -值介于 0.10 和 0.20 之间

精确的 p -值 = 0.171 8

不能拒绝 H_0 ; 平均分无差异

b. -1.05

c. 1.28; 是的

28. a. $\bar{p}_1 - \bar{p}_2 = 0.48 - 0.36 = 0.12$

b. $\bar{p}_1 - \bar{p}_2 \pm z_{0.05} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$

$0.12 \pm 1.645 \sqrt{\frac{0.48(1-0.48)}{400} + \frac{0.36(1-0.36)}{300}}$

$0.12 \pm 0.061 4; (0.058 6, 0.181 4)$

c. $0.12 \pm 1.96 \sqrt{\frac{0.48(1-0.48)}{400} + \frac{0.36(1-0.36)}{300}}$

$0.12 \pm 0.073 1; (0.046 9, 0.193 1)$

29. a. $\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{200 \times 0.22 + 300 \times 0.16}{200 + 300} = 0.184 0$

$z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.22 - 0.16}{\sqrt{0.184 0(1-0.184 0)\left(\frac{1}{200} + \frac{1}{300}\right)}} = 1.70$

p -值 = 1.000 0 - 0.955 4 = 0.044 6

b. p -值 ≤ 0.05 ; 拒绝 H_0

30. $\bar{p}_1 = 220/400 = 0.55$; $\bar{p}_2 = 192/400 = 0.48$

- $$\bar{p}_1 - \bar{p}_2 \pm z_{0.05} \sqrt{\frac{\bar{p}_1 (1 - \bar{p}_1)}{n_1} + \frac{\bar{p}_2 (1 - \bar{p}_2)}{n_2}}$$
- $$0.55 - 0.48 \pm 1.96 \sqrt{\frac{0.55 (1.055)}{400} + \frac{0.48 (1 - 0.48)}{400}}$$
- $$0.07 \pm 0.0691 \text{ c } (0.0009, 0.1391)$$
32. a. $\bar{p}_1 = 0.45$
b. $\bar{p}_2 = 0.35$
c. $(0.0011, 0.1989)$
34. a. $H_0: p_1 - p_2 \leq 0$
 $H_a: p_1 - p_2 > 0$
b. $\bar{p}_1 = 0.2017$
c. $\bar{p}_2 = 0.1111$
d. $p\text{-值} = 0.0179$
拒绝 H_0 ; 可以得出 2005 年钻的井比 2012 年钻的井干涸得更频繁。
36. a. $H_0: p_1 - p_2 \leq 0$
 $H_a: p_1 - p_2 > 0$
b. 0.84, 0.81
c. $p\text{-值} = 0.0094$
拒绝 H_0 ; 可以得出有了一个显著增长的结论。
d. $(0.005, 0.055)$; 由于入住率有了增长, 因此感到很满意。
38. $H_0: \mu_1 - \mu_2 = 0$
 $H_a: \mu_1 - \mu_2 \neq 0$
 $z = 2.79$
 $p\text{-值} = 0.0052$
拒绝 H_0 ; 系统之间存在显著差异。
40. a. $H_0: \mu_1 - \mu_2 \leq 0$
 $H_a: \mu_1 - \mu_2 > 0$
b. $t = 0.60, df = 57$
利用 t 分布表, $p\text{-值}$ 大于 0.20
精确的 $p\text{-值} = 0.2754$
不能拒绝 H_0 : 不能得出有佣金的基金平均收益率更高的结论。
42. a. $\bar{d} = 14$
b. $(-6.78, 34.78)$
c. $H_0: \mu_d = 0$
 $H_a: \mu_d \neq 0$
$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{14 - 0}{53.744 / \sqrt{20}} = 1.165$$

 $p\text{-值}$ 介于 0.20 和 0.40 之间
精确的 $p\text{-值}$ 是 0.2584
不能拒绝 H_0 ; 不能得出有兄弟姊妹组与无兄弟姊妹组之间的平均分存在差异的结论。
44. a. $p\text{-值} \approx 0$, 拒绝 H_0

b. $(0.0468, 0.1332)$

46. a. 0.35 和 0.47

b. 0.12 ± 0.1037 , 即 $(0.0163, 0.2237)$

c. 是的, 我们预期的租用率提高了。

第 11 章

2. $s^2 = 25$ a. 当自由度为 19 时, $\chi_{0.05}^2 = 30.144$, $\chi_{0.95}^2 = 10.117$

$$\frac{19 \times 25}{30.144} \leq \sigma^2 \leq \frac{19 \times 25}{10.117}$$

$$15.76 \leq \sigma^2 \leq 46.95$$

b. 当自由度为 19 时, $\chi_{0.025}^2 = 32.852$, $\chi_{0.975}^2 = 8.907$

$$\frac{19 \times 25}{32.852} \leq \sigma^2 \leq \frac{19 \times 25}{8.907}$$

$$14.46 \leq \sigma^2 \leq 53.33$$

c. $3.8 \leq \sigma \leq 7.3$ 4. a. $(0.22, 0.71)$ b. $(0.47, 0.84)$

6. a. 41 美元

b. 23.52

c. $(17.37, 36.40)$

8. a. 0.4748

b. 0.6891

c. $(0.2383, 1.3687)$ $(0.4882, 1.1699)$ 9. $H_0: \sigma^2 \leq 0.0004$ $H_a: \sigma^2 > 0.0004$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(30-1) \times 0.0005}{0.0004} = 36.25$$

查自由度为 29 的表, 知 $p\text{-值}$ 大于 0.10 $p\text{-值} > 0.05$; 不能拒绝 H_0

没有出现违背产品规格的情形

10. a. 84

b. 118.71

c. 10.90

d. $\chi^2 = 11.54$, $p\text{-值} > 0.20$ 不能拒绝假设 $\sigma = 12$

12. a. 0.8106

b. $\chi^2 = 9.49$ $p\text{-值}$ 大于 0.20不能拒绝 H_0 ; 不能得出与其他杂志方差不同的结论。14. a. $F = 2.4$ $p\text{-值}$ 介于 0.025 和 0.05 之间拒绝 H_0 b. $F_{0.05} = 2.2$; 拒绝 H_0

15. a. 较大的样本方差是 s_1^2
- $$F = \frac{s_1^2}{s_2^2} = \frac{8.2}{4.0} = 2.05$$
- 自由度为 20, 25
- 查表, 上侧面积介于 0.025 和 0.05 之间, 检验的双侧 p -值介于 0.05 和 0.10 之间, p -值 > 0.05 , 所以不能拒绝 H_0 。
- b. 对于双侧检验
- $$F_{\alpha/2} = F_{0.025} = 2.30$$
- 如果 $F \geq 2.30$, 拒绝 H_0
- 由于 $2.05 < 2.30$, 不能拒绝 H_0
16. $F = 1.59$
- p -值介于 0.025 和 0.05 之间
- 拒绝 H_0 ; Fidelity 基金有较大的方差。
17. a. 总体 1 是已使用了 4 年的汽车。
- $$H_0: \sigma_1^2 \leq \sigma_2^2$$
- $$H_a: \sigma_1^2 > \sigma_2^2$$
- b. $F = \frac{s_1^2}{s_2^2} = \frac{170^2}{100^2} = 2.89$
- 自由度为 25, 24
- 查表, p -值小于 0.01
- p -值 ≤ 0.01 ; 拒绝 H_0
- 结论是已经使用 4 年的汽车年维修费用的方差比已经使用 2 年的汽车年维修费用的方差大。因为我们预期较旧的车更容易发生较高的维修费用, 从而导致更大的年维修费用的方差。
18. $F = 1.44$
- p -值大于 0.20
- 不能拒绝 H_0 ; 方差之间的差异在统计上不显著。
20. $F = 5.29$
- p -值 ≈ 0
- 拒绝 H_0 ; 资深人员和管理人员的总体方差不相等。
22. a. $F = 4$
- p -值小于 0.01
- 拒绝 H_0 ; 湿滑路面上刹车距离的方差较大。
24. (10.72, 24.68)
26. a. $\chi^2 = 27.44$
- p -值介于 0.01 和 0.025 之间
- 拒绝 H_0 ; 方差超出最大要求。
- b. (0.00012, 0.00042)
28. $\chi^2 = 31.50$
- p -值介于 0.05 和 0.10 之间
- 拒绝 H_0 ; 得出总体方差大于 1 的结论。
30. a. $n = 15$
- b. (6.25, 11.13)

32. $F = 1.39$
- 不能拒绝 H_0 ; 不能得出总平均分的方差不同的结论。
34. $F = 2.08$
- p -值介于 0.05 和 0.10 之间
- 拒绝 H_0 ; 得出总体方差不等的结论。

第 12 章

1. a. $H_0: p_1 = p_2 = p_3$
- H_a : 所有总体比例不全相等
- 期望频数 (e_{ij}):

	1	2	3	合计
是	132.0	158.4	105.6	396
否	118.0	141.6	94.4	354
合计	250	300	200	750

χ^2 计算结果 $(f_{ij} - e_{ij})^2 / e_{ij}$:

	1	2	3	合计
是	2.45	0.45	0.87	3.77
否	2.75	0.50	0.98	4.22
				$\chi^2 = 7.99$

$df = k - 1 = 3 - 1 = 2$

$\chi^2 = 7.99$ 的 χ^2 表, 表明 p -值介于 0.01 和 0.025 之间, p -值 ≤ 0.05 , 拒绝 H_0 ; 所有总体比例不全相等。

2. a. $\bar{p}_1 = 150/250 = 0.60$, $\bar{p}_2 = 150/300 = 0.50$, $\bar{p}_3 = 96/200 = 0.48$
- b. 对于 1 和 2

$$\begin{aligned} CV_{12} &= \sqrt{\chi_{\alpha, k-1}^2} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} \\ &= \sqrt{5.991} \sqrt{\frac{0.60(1-0.60)}{250} + \frac{0.50(1-0.50)}{300}} \\ &= 0.1037 \end{aligned}$$

p_i	p_j	差	n_i	n_j	临界值	如果差 $> CV$, 则显著
0.60	0.50	0.10	250	300	0.1037	
0.60	0.48	0.12	250	200	0.1150	是
0.50	0.48	0.02	300	200	0.1117	

1 和 3 的比较是显著的。

4. a. $H_0: p_1 = p_2 = p_3$
- H_a : 所有总体比例不全相等
- b. 期望频数 (e_{ij}):

部件	A	B	C	合计
有缺陷	25	25	25	75
良好	475	475	475	1425
合计	500	500	500	1500

χ^2 计算结果 $(f_{ij} - e_{ij})^2 / e_{ij}^2$:

部件	A	B	C	合计
有缺陷	4.00	1.00	9.00	14.00
良好	0.21	0.05	0.47	0.74
				$\chi^2 = 14.74$

$df = k - 1 = 3 - 1 = 2$

$\chi^2 = 14.74$ 的 χ^2 表, 表明 p -值小于 0.01, p -值 ≤ 0.05 , 拒绝 H_0 ; 三家供应商没有提供相同的有缺陷部件比例。

c. $\bar{p}_1 = 15/500 = 0.03$, $\bar{p}_2 = 20/500 = 0.04$, $\bar{p}_3 = 40/500 = 0.08$

对于 A 和 B

$$CV_{12} = \sqrt{\chi^2_{\alpha, k-1}} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$$
$$= \sqrt{5.991} \sqrt{\frac{0.03(1-0.03)}{500} + \frac{0.04(1-0.04)}{500}}$$
$$= 0.0284$$

比较	p_i	p_j	差	n_i	n_j	临界值	如果差 > CV, 则显著
A 和 B	0.03	0.04	0.01	500	500	0.0284	
A 和 C	0.03	0.08	0.05	500	500	0.0351	是
B 和 C	0.04	0.08	0.04	500	500	0.0366	是

供应商 A 和供应商 B 都与供应商 C 差异显著。

6. a. 0.14, 0.09

b. $\chi^2 = 3.41$, $df = 1$

p -值介于 0.05 和 0.10 之间

拒绝 H_0 ; 得出两个办事处出错率不等的结论。

c. z 提供单侧检验的选项

8. $\chi^2 = 5.70$, $df = 4$

p -值大于 0.10

不能拒绝 H_0 ; 没有供货商质量不同的证据。

9. H_0 : 列变量与行变量独立

H_a : 列变量与行变量不独立

期望频数 (e_{ij}):

	A	B	C	合计
P	28.5	39.9	45.6	114
Q	21.5	30.1	34.4	86
合计	50	50	80	200

χ^2 计算结果 $(f_{ij} - e_{ij})^2 / e_{ij}^2$:

	A	B	C	合计
P	2.54	0.42	0.42	3.38
Q	3.36	0.56	0.56	4.48
				$\chi^2 = 7.86$

$df = (2 - 1) \times (3 - 1) = 2$

利用 χ^2 表, p -值介于 0.01 和 0.025 之间

p -值 ≤ 0.05 , 拒绝 H_0 ; 列变量与行变量不独立。

10. $\chi^2 = 19.77$, $df = 4$

p -值小于 0.005

拒绝 H_0 ; 列变量与行变量不独立。

11. a. H_0 : 机票类型与航班类型独立

H_a : 机票类型与航班类型不独立

期望频数:

$e_{11} = 35.59$ $e_{12} = 15.41$ $e_{21} = 150.73$

$e_{22} = 65.27$ $e_{31} = 455.68$ $e_{32} = 197.32$

观察频数 (f_i)	期望频数 (e_i)	$\chi^2 = (f_i - e_i)^2 / e_i$
29	35.59	1.22
22	15.41	2.82
95	150.73	20.6
121	65.27	47.59
518	455.68	18.52
135	197.32	19.68
920		$\chi^2 = 100.43$

$df = (3 - 1) \times (2 - 1) = 2$

利用 χ^2 表, p -值小于 0.005

p -值 ≤ 0.05 , 拒绝 H_0 ; 航班类型与机票类型不独立。

b. 列百分数

机票类型	航班类型	
	国内航班	国际航班
头等舱	4.5%	7.9%
商务舱	14.8%	43.5%
经济舱	80.7%	48.6%

对于国际航班, 购买头等舱和商务舱的比例较高。

12. a. $\chi^2 = 9.44$, $df = 2$

p -值小于 0.01

拒绝 H_0 ; 计划与公司类型不独立。

b.

雇用计划	私有	公有
增加员工	0.5139	0.2963
不改变	0.2639	0.3148
减少员工	0.2222	0.3889

私有企业雇佣机会较好

14. a. $\chi^2 = 6.57$, $df = 6$

p -值大于 0.10

不能拒绝 H_0 ; 不能拒绝独立性假设

b. 29%, 46% 和 25%

杰出是车主最常给出的评价。

16. a. 900

b. 影迷偏爱詹妮弗·劳伦斯 ($\bar{p}=0.2278$), 但其他三位提名人 (杰西卡·查斯坦、艾曼纽·丽娃、娜奥米·沃茨) 被影迷偏爱程度几乎相同。

c. $\chi^2 = 77.74$

p -值几乎为 0

拒绝 H_0 ; 得出人们对获得 2013 年奥斯卡最佳女演员奖的女演员的态度与回答者的年龄不独立。

18. $\chi^2 = 45.36$, $df = 4$

p -值小于 0.005

拒绝 H_0 ; 得出主持人的评价不独立

19. a. 期望频数: $e_1 = 200 \times 0.40 = 80$, $e_2 = 200 \times 0.40 = 80$, $e_3 = 200 \times 0.20 = 40$

观测频数: $f_1 = 60$, $f_2 = 120$, $f_3 = 20$

$$\begin{aligned}\chi^2 &= \frac{(60-80)^2}{80} + \frac{(120-80)^2}{80} + \frac{(20-40)^2}{40} \\ &= \frac{400}{80} + \frac{1600}{80} + \frac{400}{40} \\ &= 5 + 20 + 10 = 35\end{aligned}$$

$df = k - 1 = 2$

$\chi^2 = 35$, 表明 p -值小于 0.005

p -值 ≤ 0.01 , 拒绝 H_0 ; 总体比例不是 0.40, 0.40 和 0.20。

b. 如果 $\chi^2 \geq 9.210$, 则拒绝 H_0 ; 拒绝 H_0

20. $n = 30$, 我们等分成六类, 每类的概率为 0.1667。

$\bar{x} = 22.8$, $s = 6.27$

z	x 的临界值
-0.98	$22.8 - 0.98 \times 6.27 = 16.66$
-0.43	$22.8 - 0.43 \times 6.27 = 20.11$
0	$22.8 + 0.00 \times 6.27 = 22.80$
0.43	$22.8 + 0.43 \times 6.27 = 25.49$
0.98	$22.8 + 0.98 \times 6.27 = 28.94$

区间	观察频数	期望频数	差
小于 16.66	3	5	-2
16.66 ~ 20.11	7	5	2
20.11 ~ 22.80	5	5	0
22.80 ~ 25.49	7	5	2
25.49 ~ 28.94	3	5	-2
28.94 及以上	5	5	0

$$\chi^2 = \frac{(-2)^2}{5} + \frac{2^2}{5} + \frac{0^2}{5} + \frac{(2)^2}{5} + \frac{(-2)^2}{5} + \frac{0^2}{5} = \frac{16}{5} = 3.20$$

$df = k - 2 - 1 = 6 - 2 - 1 = 3$

$\chi^2 = 3.20$, 显示 p -值大于 0.10

p -值 > 0.05 , 不能拒绝 H_0 ; 正态分布的假设没有被拒绝。

21. H_0 : 总体比例是 0.29, 0.28, 0.25 和 0.18

H_a : 总体比例不是上面所述

期望频数: $300 \times 0.29 = 87$, $300 \times 0.28 = 84$

$300 \times 0.25 = 75$, $300 \times 0.18 = 54$

观测频数: $f_1 = 95$, $f_2 = 70$, $f_3 = 89$, $f_4 = 46$

$$\begin{aligned}\chi^2 &= \frac{(95-87)^2}{87} + \frac{(70-84)^2}{84} + \frac{(89-75)^2}{75} + \frac{(46-54)^2}{54} \\ &= 6.87\end{aligned}$$

$df = k - 1 = 4 - 1 = 3$

$\chi^2 = 6.87$, 显示 p -值介于 0.05 和 0.10 之间

p -值 > 0.05 ; 不能拒绝 H_0 ; 观众收视率没有显著变化。

22. $\chi^2 = 5.85$, $df = 5$

p -值大于 0.10

不能拒绝 H_0 ; 不能拒绝比例为所陈述的假设。

24. a. $\chi^2 = 14.33$, $df = 6$

p -值介于 0.025 和 0.05 之间

拒绝 H_0 ; 得出一周中每天的比例不全相等的结论。

b. 15.17, 11.90, 12.62, 11.19, 13.10, 16.43, 19.05

星期六和星期五的比例最高。

26. $\chi^2 = 2.8$, $df = 3$

p -值大于 0.10

不能拒绝 H_0 ; 正态分布的假设没有被拒绝。

28. a. 8.8%, 11.7%, 9.0%, 8.5%

b. $\chi^2 = 2.48$, $df = 3$

p -值大于 0.10

不能拒绝 H_0 ; 不能拒绝总体比例相等的假设。

30. a. $\chi^2 = 9.56$, $df = 2$

p -值小于 0.01

拒绝 H_0 ; 得出生活节奏与性别不独立的结论。

b. 女性较慢 75.17%, 男性较慢 67.65%

男性较快 26.47%, 女性较快 16.55%

32. $\chi^2 = 6.17$, $df = 6$

p -值大于 0.10

不能拒绝 H_0 ; 县类型与星期几独立的假设没有被拒绝。

34. $\chi^2 = 2.00$, $df = 5$

p -值大于 0.10

不能拒绝 H_0 ; 正态分布的假设不能被拒绝。

第 13 章

1. a. $\bar{\bar{x}} = (156 + 142 + 134)/3 = 144$

$$\begin{aligned}\text{SSTR} &= \sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2 \\ &= 6 \times (156 - 144)^2 + 6 \times (142 - 144)^2 \\ &\quad + 6 \times (134 - 144)^2 = 1488\end{aligned}$$

$$\text{b. MSTR} = \frac{\text{SSTR}}{k-1} = \frac{1488}{2} = 744$$

$$c. s_1^2 = 164.4, s_2^2 = 131.2, s_3^2 = 110.4$$

$$SSE = \sum_{j=1}^k (n_j - 1)s_j^2 = 5 \times 164.4 + 5 \times 131.2 + 5 \times 110.4 = 2030$$

$$d. MSE = \frac{SSE}{n_T - k} = \frac{2030}{18 - 3} = 135.3$$

e.

方差来源	平方和	自由度	均方	F	p-值
处理	1488	2	744	5.50	0.0162
误差	2030	15	135.3		
合计	3518	17			

$$f. F = \frac{MSTR}{MSE} = \frac{744}{135.3} = 5.50$$

从分子自由度为 2, 分母自由度为 15 的 F 分布表可知, p -值介于 0.01 和 0.025 之间

利用 Excel 或 Minitab, $F = 5.50$ 对应于的 p -值是 0.0162
因为 p -值 $\leq \alpha = 0.05$, 所以我们拒绝三个总体均值相等的假设。

2.

方差来源	平方和	自由度	均方	F	p-值
处理	300	4	75	14.07	0.0000
误差	160	30	5.33		
合计	460	34			

4.

方差来源	平方和	自由度	均方	F	p-值
处理	150	2	75	4.80	0.0233
误差	250	16	15.63		
合计	400	18			

因为 p -值 $\leq \alpha = 0.05$, 所以拒绝 H_0

6. 因为 p -值 $= 0.0082 \leq \alpha = 0.05$, 所以我们拒绝三个总体均值相等的原假设。

$$8. \bar{\bar{x}} = (79 + 74 + 66)/3 = 73$$

$$SSTR = \sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2 = 6 \times (79 - 73)^2 + 6 \times (74 - 73)^2 + 6 \times (66 - 73)^2 = 516$$

$$MSTR = \frac{SSTR}{k-1} = \frac{516}{2} = 258$$

$$s_1^2 = 34, s_2^2 = 20, s_3^2 = 32$$

$$SSE = \sum_{j=1}^k (n_j - 1)s_j^2 = 5 \times 34 + 5 \times 20 + 5 \times 32 = 430$$

$$MSE = \frac{SSE}{n_T - k} = \frac{430}{18 - 3} = 28.67$$

$$F = \frac{MSTR}{MSE} = \frac{258}{28.67} = 9.00$$

方差来源	平方和	自由度	均方	F	p-值
处理	516	2	258	9.00	0.003
误差	430	15	28.67		
合计	946	17			

根据分子自由度为 2, 分母自由度为 15 的 F 分布表可知, p -值小于 0.01

利用 Excel 或 Minitab, $F = 9.00$ 对应于的 p -值是 0.003

因为 p -值 $\leq \alpha = 0.05$, 所以我们拒绝三家工厂的平均考试成绩是相同的原假设。换句话说, 方差分析支持的结论是, 三家 NCP 工厂总体的平均考试成绩是不相同的。

10. p -值 $= 0.0000$

因为 p -值 $\leq \alpha = 0.05$, 所以我们拒绝三组情形的平均分数是相同的原假设。

12. p -值 $= 0.0038$

因为 p -值 $\leq \alpha = 0.05$, 所以我们拒绝三种类型的餐厅吃一顿饭的平均价格是相同的原假设。

$$13. a. \bar{\bar{x}} = (30 + 45 + 36)/3 = 37$$

$$SSTR = \sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2 = 5 \times (30 - 37)^2 + 5 \times (45 - 37)^2 + 5 \times (36 - 37)^2 = 570$$

$$MSTR = \frac{SSTR}{k-1} = \frac{570}{2} = 285$$

$$SSE = \sum_{j=1}^k (n_j - 1)s_j^2 = 4 \times 6 + 4 \times 4 + 4 \times 6.5 = 66$$

$$MSE = \frac{SSE}{n_T - k} = \frac{66}{15 - 3} = 5.5$$

$$F = \frac{MSTR}{MSE} = \frac{285}{5.5} = 51.82$$

根据分子自由度为 2, 分母自由度为 12 的 F 分布表可知, p -值小于 0.01

利用 Excel 或 Minitab, $F = 51.82$ 对应于的 p -值是 0.0000

因为 p -值 $\leq \alpha = 0.05$, 所以我们拒绝三个总体均值相等的原假设。

$$b. LSD = t_{\alpha/2} \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} = t_{0.025} \sqrt{5.5 \times \left(\frac{1}{5} + \frac{1}{5} \right)} = 2.179 \sqrt{2.2} = 3.23$$

$$|\bar{x}_1 - \bar{x}_2| = |30 - 45| = 15 > LSD, \text{ 存在显著差异}$$

$$|\bar{x}_1 - \bar{x}_3| = |30 - 36| = 6 > LSD, \text{ 存在显著差异}$$

$$|\bar{x}_2 - \bar{x}_3| = |45 - 36| = 9 > LSD, \text{ 存在显著差异}$$

$$c. \bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} = (30 - 45) \pm 2.179 \sqrt{5.5 \times \left(\frac{1}{5} + \frac{1}{5} \right)}$$

$$-15 \pm 3.23 = (-18.23, -11.77)$$

14. a. 显著; p -值=0.010 6

b. LSD = 15.34

1 和 2, 显著

1 和 3, 不显著

2 和 3, 显著

15. a.

	制造商		
	1	2	3
样本均值	23	28	21
样本方差	6.67	4.67	3.33

$$\bar{\bar{x}} = (23 + 28 + 21)/3 = 24$$

$$\begin{aligned} \text{SSTR} &= \sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2 \\ &= 4 \times (23 - 24)^2 + 4 \times (28 - 24)^2 + 4 \times (21 - 24)^2 \\ &= 104 \end{aligned}$$

$$\text{MSTR} = \frac{\text{SSTR}}{k-1} = \frac{104}{2} = 52$$

$$\begin{aligned} \text{SSE} &= \sum_{j=1}^k (n_j - 1)s_j^2 = 3 \times 6.67 + 3 \times 4.67 + 3 \times 3.33 \\ &= 44.01 \end{aligned}$$

$$\text{MSE} = \frac{\text{SSE}}{n_T - k} = \frac{44.01}{12 - 3} = 4.89$$

$$F = \frac{\text{MSTR}}{\text{MSE}} = \frac{52}{4.89} = 10.63$$

根据分子自由度为 2, 分母自由度为 9 的 F 分布表可知, p -值小于 0.01

利用 Excel 或 Minitab, $F = 10.63$ 对应于的 p -值是 0.0043

因为 p -值 $\leq \alpha = 0.05$, 所以我们拒绝三家制造商生产的机器混合一批原料的平均时间是相同的原假设。

$$\begin{aligned} \text{b. LSD} &= t_{\alpha/2} \sqrt{\text{MSE} \left(\frac{1}{n_1} + \frac{1}{n_3} \right)} \\ &= t_{0.025} \sqrt{4.89 \times \left(\frac{1}{4} + \frac{1}{4} \right)} = 2.262 \sqrt{2.45} = 3.54 \end{aligned}$$

因为 $|\bar{x}_1 - \bar{x}_3| = |23 - 21| = 2 < 3.54$, 所以在制造商 1 与制造商 3 之间似乎不存在任何显著性差异。

16. $\bar{x}_1 - \bar{x}_2 \pm \text{LSD}$

$$23 - 28 \pm 3.54$$

$$-5 \pm 3.54 = (-8.54, -1.46)$$

18. a. 显著; p -值=0.000 0

b. 显著; $2.3 > \text{LSD} = 1.19$

20. a. 显著; p -值=0.011

b. 比较北区和南区

$$|7702 - 5566| = 2136 > \text{LSD} = 1620.76; \text{存在显著差异}$$

比较北区和西区

$$|7702 - 8430| = 728 < \text{LSD} = 1620.76; \text{不存在显著差异}$$

比较南区 and 西区

$$|5566 - 8430| = 2864 > \text{LSD} = 1620.76; \text{存在显著差异}$$

21. 处理均值

$$\bar{x}_{\cdot 1} = 13.6, \bar{x}_{\cdot 2} = 11.0, \bar{x}_{\cdot 3} = 10.6$$

区组均值

$$\bar{x}_{1 \cdot} = 9, \bar{x}_{2 \cdot} = 7.67, \bar{x}_{3 \cdot} = 15.67, \bar{x}_{4 \cdot} = 18.67,$$

$$\bar{x}_{5 \cdot} = 7.67$$

总样本均值

$$\bar{\bar{x}} = 176/15 = 11.73$$

第 1 步

$$\begin{aligned} \text{SST} &= \sum_i \sum_j (x_{ij} - \bar{\bar{x}})^2 = (10 - 11.73)^2 + (9 - 11.73)^2 \\ &\quad + \cdots + (8 - 11.73)^2 = 354.93 \end{aligned}$$

第 2 步

$$\begin{aligned} \text{SSTR} &= b \sum_j (\bar{x}_{\cdot j} - \bar{\bar{x}})^2 = 5 \times [(13.6 - 11.73)^2 \\ &\quad + (11.0 - 11.73)^2 + (10.6 - 11.73)^2] = 26.53 \end{aligned}$$

第 3 步

$$\begin{aligned} \text{SSBL} &= k \sum_j (\bar{x}_{i \cdot} - \bar{\bar{x}})^2 = 3 \times [(9 - 11.73)^2 \\ &\quad + (7.67 - 11.73)^2 + (15.67 - 11.73)^2 \\ &\quad + (18.67 - 11.73)^2 + (7.67 - 11.73)^2] \\ &= 312.32 \end{aligned}$$

第 4 步

$$\begin{aligned} \text{SSE} &= \text{SST} - \text{SSTR} - \text{SSBL} \\ &= 354.93 - 26.53 - 312.32 = 16.08 \end{aligned}$$

方差来源	平方和	自由度	均方	F	p -值
处理	26.53	2	13.27	6.60	0.020 3
区组	312.32	4	78.08		
误差	16.08	8	2.01		
合计	354.93	14			

根据分子自由度为 2, 分母自由度为 8 的 F 分布表可知, p -值介于 0.01 和 0.025 之间

实际的 p -值=0.020 3

因为 p -值 $\leq \alpha = 0.05$, 所以我们拒绝三个处理的均值是相等的原假设。

22.

方差来源	平方和	自由度	均方	F	p -值
处理	310	4	77.5	17.69	0.000 5
区组	85	2	42.5		
误差	35	8	4.38		
合计	430	14			

显著; p -值 $\leq\alpha=0.05$

24. p -值 $=0.0453$

因为 p -值 $\leq\alpha=0.05$, 所以我们拒绝两种发动机分析器的平均调整时间是相同的原假设。

26. a. 显著; p -值 $=0.0231$

b. 写作部分

28. 第1步

$$\begin{aligned} SST &= \sum_i \sum_j \sum_k (x_{ijk} - \bar{\bar{x}})^2 = (135 - 111)^2 \\ &\quad + (165 - 111)^2 + \cdots + (136 - 111)^2 = 9028 \end{aligned}$$

第2步

$$\begin{aligned} SSA &= br \sum_i (\bar{x}_{i.} - \bar{\bar{x}})^2 = 3 \times 2 \times [(104 - 111)^2 \\ &\quad + (118 - 111)^2] = 588 \end{aligned}$$

第3步

$$\begin{aligned} SSB &= ar \sum_j (\bar{x}_{.j} - \bar{\bar{x}})^2 = 2 \times 2 \times [(130 - 111)^2 \\ &\quad + (97 - 111)^2 + (106 - 111)^2] = 2328 \end{aligned}$$

第4步

$$\begin{aligned} SSAB &= r \sum_i \sum_j (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}})^2 \\ &= 2 \times [(150 - 104 - 130 + 111)^2 \\ &\quad + (78 - 104 - 97 + 111)^2 + \cdots \\ &\quad + (128 - 118 - 106 + 111)^2] = 4392 \end{aligned}$$

第5步

$$\begin{aligned} SSE &= SST - SSA - SSB - SSAB = 9028 - 588 - 2328 - 4392 \\ &= 1720 \end{aligned}$$

方差来源	平方和	自由度	均方	F	p -值
因子 A	588	1	588	2.05	0.2022
因子 B	2328	2	1164	4.06	0.0767
交互作用	4392	2	2196	7.66	0.0223
误差	1720	6	286.67		
合计	9028	11			

因子 A: $F=2.05$

根据分子自由度为 1, 分母自由度为 6 的 F 分布表可知, p -值大于 0.10

利用 Excel 或 Minitab, $F=2.05$ 对应的 p -值是 0.2022

因为 p -值 $>\alpha=0.05$, 所以因子 A 是不显著的

因子 B: $F=4.06$

根据分子自由度为 2, 分母自由度为 6 的 F 分布表可知, p -值介于 0.05 和 0.10 之间

利用 Excel 或 Minitab, $F=4.06$ 对应的 p -值是 0.0767

因为 p -值 $>\alpha=0.05$, 所以因子 B 是不显著的

交互作用: $F=7.66$

根据分子自由度为 2, 分母自由度为 6 的 F 分布表可知, p -值介于 0.01 和 0.025 之间

利用 Excel 或 Minitab, $F=7.66$ 对应的 p -值是 0.0223

因为 p -值 $\leq\alpha=0.05$, 交互作用是显著的

30. 设计图案: p -值 $=0.0104$, 显著

广告尺寸: p -值 $=0.1340$, 不显著

交互作用: p -值 $=0.2519$, 不显著

32. 汽车类型: p -值 $=0.0002$, 显著

动力类型: p -值 $=0.0006$, 显著

交互作用: p -值 $=0.4229$, 不显著

34. 显著; p -值 $=0.0134$

36. 不显著; p -值 $=0.088$

38. 不显著; p -值 $=0.2455$

40. a. 显著; p -值 $=0.0175$

42. 区组对应于 12 名高尔夫球员, 处理对应于三种设计。

Minitab 的双向输出是:

方差来源	平方和	自由度	均方	F	p -值
设计	2	3032.0	1516.0	12.89	0.000
球员	11	5003.3	454.8	3.87	0.003
误差	22	2586.7	117.6		
合计	35	10622.0			

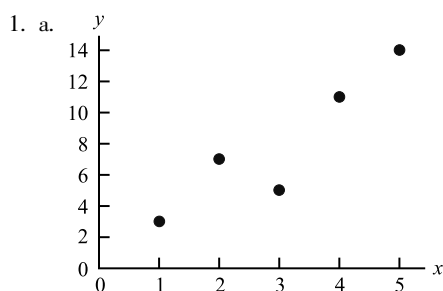
因为设计的 p -值 $=0.000$, 小于 $\alpha=0.05$, 所以在三种设计的平均击球距离之间存在显著差异。

44. 机器的类型是显著的, 因为 p -值 $=0.0226$

投料系统的类型是不显著的, 因为 p -值 $=0.7913$

交互作用是不显著的, 因为 p -值 $=0.0671$

第 14 章



b. 在 x 和 y 之间似乎存在一种正的线性关系。

c. 我们可以画出许多不同的直线作为 x 和 y 之间关系的一个线性近似; 在 (d) 中我们将根据最小二乘准则, 确定一个表示这种关系的“最佳”直线方程。

d. 为了计算斜率和 y 轴截距所需要的求和:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{15}{5} = 3, \quad \bar{y} = \frac{\sum y_i}{n} = \frac{40}{5} = 8,$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = 26, \quad \sum (x_i - \bar{x})^2 = 10$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{26}{10} = 2.6$$

$$b_0 = \bar{y} - b_1 \bar{x} = 8 - 2.6 \times 3 = 0.2$$

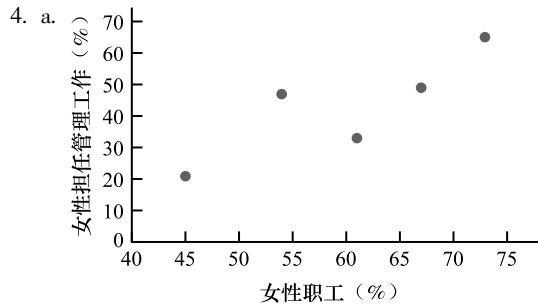
$$\hat{y} = 0.2 + 2.6x$$

$$e. \hat{y} = 0.2 + 2.6x = 0.2 + 2.6 \times 4 = 10.6$$

2. b. 在 x 和 y 之间似乎存在一种负的线性关系。

$$d. \hat{y} = 68 - 3x$$

$$e. 38$$



b. x 表示五家公司中女性职工所占的百分比, y 表示五家公司中女性担任管理工作的百分比, 在 x 和 y 之间似乎存在一种正的线性关系。

c. 我们可以画出许多不同的直线作为 x 和 y 之间关系的一个线性近似; 在 (d) 中我们将根据最小二乘准则, 确定一个表示这种关系的“最佳”直线方程。

$$d. \bar{x} = \frac{\sum x_i}{n} = \frac{300}{5} = 60, \bar{y} = \frac{\sum y_i}{n} = \frac{215}{5} = 43$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = 624, \sum (x_i - \bar{x})^2 = 480$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{624}{480} = 1.3$$

$$b_0 = \bar{y} - b_1 \bar{x} = 43 - 1.3 \times 60 = -35$$

$$\hat{y} = -35 + 1.3x$$

$$e. \hat{y} = -35 + 1.3x = -35 + 1.3 \times 60 = 43\%$$

$$6. c. \hat{y} = -70.391 + 17.175x$$

$$e. 43.8 \text{ 或近似 } 44\%$$

$$8. c. \hat{y} = 0.2046 + 0.9077x$$

$$d. 3.29 \text{ 或近似 } 3.3$$

$$10. c. \hat{y} = -167.81 + 2.7149x$$

e. 合理

$$12. c. \hat{y} = 17.49 + 1.0334x$$

d. 150 美元

14. a. 散点图表明 x (工作距离) 与 y (缺勤天数) 存在负的线性关系。

$$b. \hat{y} = 8.0978 - 0.3442x$$

$$c. \hat{y} = 8.0978 - 0.3442 \times 5 = 6.4 \text{ 或近似 } 6 \text{ 天}$$

$$15. a. \hat{y}_i = 0.2 + 2.6x_i \text{ 和 } \bar{y} = 8$$

x_i	y_i	\hat{y}_i	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	3	2.8	0.2	0.04	-5	25
2	7	5.4	1.6	2.56	-1	1

(续)

x_i	y_i	\hat{y}_i	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
3	5	8.0	-3.0	9.00	-3	9
4	11	10.6	0.4	0.16	3	9
5	14	13.2	0.8	<u>0.64</u>	6	<u>36</u>
				SSE = 12.40	SST = 80	
SSR = SST - SSE = 80 - 12.4 = 67.6						

$$b. r^2 = \frac{SSR}{SST} = \frac{67.6}{80} = 0.845$$

最小二乘直线给出了一个好的拟合; y 的 84.5% 的变异性能被最小二乘直线解释。

$$c. r = \sqrt{0.845} = +0.9192$$

$$16. a. SSE = 230, SST = 1850, SSR = 1620$$

$$b. r^2 = 0.876$$

$$c. r_{xy} = -0.936$$

$$18. a. \bar{x} = \frac{\sum x_i}{n} = \frac{600}{6} = 100, \bar{y} = \frac{\sum y_i}{n} = \frac{330}{6} = 55$$

$$SST = \sum (y_i - \bar{y})^2 = 1800, SSE = \sum (y_i - \hat{y}_i)^2 = 287.624$$

$$SSR = SST - SSE = 1800 - 287.624 = 1512.376$$

$$b. r^2 = \frac{SSR}{SST} = \frac{1512.376}{1800} = 0.84$$

$$c. r = \sqrt{r^2} = \sqrt{0.84} = 0.917$$

$$20. a. \hat{y} = 28574 - 1439x$$

$$b. r^2 = 0.864$$

$$c. 6989 \text{ 美元}$$

$$22. a. r^2 = \frac{SSR}{SST} = \frac{9524.97}{10568} = 0.9013$$

b. 估计的回归方程给出了一个非常好的拟合; 因变量的大约 90% 的变异性被两变量之间的线性关系解释。

c. $r = \sqrt{r^2} = \sqrt{0.9013} = 0.95$, 这反映了两变量之间存在强的线性关系。

$$23. a. s^2 = MSE = \frac{SSE}{n-2} = \frac{12.4}{3} = 4.133$$

$$b. s = \sqrt{MSE} = \sqrt{4.133} = 2.033$$

$$c. \sum (x_i - \bar{x})^2 = 10$$

$$s_{b_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{2.033}{\sqrt{10}} = 0.643$$

$$d. t = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{2.6 - 0}{0.643} = 4.044$$

根据自由度为 3 的 t 分布表可知, $t = 4.044$ 的上侧面积介于 0.01 和 0.025 之间

于是 p -值介于 0.02 和 0.05 之间

利用 Excel 或 Minitab, 对应于 $t = 4.044$ 的 p -值是 0.0272

因为 p -值 $\leq \alpha = 0.05$, 所以我们拒绝原假设 $H_0: \beta_1 = 0$

$$e. \text{MSR} = \frac{\text{SSR}}{1} = 67.6$$

$$F = \frac{\text{MSR}}{\text{MSE}} = \frac{67.6}{4.133} = 16.36$$

根据分子自由度为 1, 分母自由度为 3 的 F 分布表可知, p -值介于 0.25 和 0.05 之间

利用 Excel 或 Minitab, 对应于 $F = 16.36$ 的 p -值是 0.027 2

因为 p -值 $\leq \alpha = 0.05$, 所以我们拒绝原假设 $H_0: \beta_1 = 0$

方差来源	平方和	自由度	均方	F	p -值
回归	67.6	1	67.6	16.36	0.027 2
误差	12.4	3	4.133		
合计	80.0	4			

24. a. 76.666 7

b. 8.756 0

c. 0.652 6

d. 显著; p -值 = 0.019 3

e. 显著; p -值 = 0.019 3

26. a. 第 18 题给出了估计的回归方程是 $\hat{y} = 23.194 + 0.318x_0$ 。

在第 18 题的答案中, 我们得到了 $\text{SSE} = 287.624$ 。

$$s^2 = \text{MSE} = \text{SSE} / (n - 2) = 287.624 / 4 = 71.906$$

$$s = \sqrt{\text{MSE}} = \sqrt{71.906} = 8.479 7$$

$$\sum (x_i - \bar{x})^2 = 14 950$$

$$s_{b_1} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{8.479 7}{\sqrt{14 950}} = 0.069 4$$

$$t = \frac{b_1}{s_{b_1}} = \frac{0.318}{0.069 4} = 4.58$$

根据自由度为 4 的 t 分布表可知, $t = 4.58$ 的上侧面积介于 0.005 和 0.01 之间

p -值介于 0.01 和 0.02 之间

利用 Excel, 对应于 $t = 4.58$ 的 p -值是 0.010

因为 p -值 $\leq \alpha$, 所以我们拒绝原假设 $H_0: \beta_1 = 0$; 在价格和整体评价得分之间存在一个显著的关系

b. 在第 18 题的中答案, 我们得到了 $\text{SSR} = 1 512.376$

$$\text{MSR} = \text{SSR} / 1 = 1512.376 / 1 = 1 512.376$$

$$F = \text{MSR} / \text{MSE} = 1512.376 / 71.906 = 21.03$$

根据分子自由度为 1, 分母自由度为 4 的 F 分布表可知, p -值介于 0.01 和 0.025 之间

利用 Excel, 对应于 $F = 21.03$ 的 p -值是 0.010

因为 p -值 $\leq \alpha$, 所以我们拒绝原假设 $H_0: \beta_1 = 0$

c.

方差来源	平方和	自由度	均方	F	p -值
回归	1 512.376	1	1 512.376	21.03	0.010
误差	287.624	4	71.906		
合计	1 800	5			

28. 它们是相关的; p -值 = 0.000

30. 根据自由度为 4 的 t 分布表可知, t 的上侧面积低于 0.005, p -值低于 0.01。在服务中的汽车与年度收入之间存在一个显著的关系。

32. a. $s = 2.033$

$$\bar{x} = 3, \quad \sum (x_i - \bar{x})^2 = 10$$

$$s_{\hat{y}^*} = s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 2.033 \times \sqrt{\frac{1}{5} + \frac{(4 - 3)^2}{10}} = 1.11$$

$$b. \hat{y}^* = 0.2 + 2.6x^* = 0.2 + 2.6 \times 4 = 10.6$$

$$\hat{y}^* \pm t_{\alpha/2} s_{\hat{y}^*}$$

$$10.6 \pm 3.182 \times 1.11$$

$$10.6 \pm 3.53 \text{ 或 } (7.07, 14.13)$$

$$c. s_{\text{pred}} = s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 2.033 \times \sqrt{1 + \frac{1}{5} + \frac{(4 - 3)^2}{10}} = 2.32$$

$$d. \hat{y}^* \pm t_{\alpha/2} s_{\text{pred}}$$

$$10.6 \pm 3.182 \times 2.32$$

$$10.6 \pm 7.38 \text{ 或 } (3.22, 17.98)$$

34. 置信区间: (8.65, 21.15)

预测区间: (-4.50, 41.30)

35. a. $\hat{y}^* = 2 090.5 + 581.1x^* = 2 090.5 + 581.1 \times 3 = 3 833.8$

$$b. s = \sqrt{\text{MSE}} = \sqrt{21284} = 145.89$$

$$\bar{x} = 3.2, \quad \sum (x_i - \bar{x})^2 = 0.74$$

$$s_{\hat{y}^*} = s \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 145.89 \times \sqrt{\frac{1}{6} + \frac{(3 - 3.2)^2}{0.74}} = 68.54$$

$$\hat{y}^* \pm t_{\alpha/2} s_{\hat{y}^*}$$

$$3 833.8 \pm 2.776 \times 68.54 = 3 833.8 \pm 190.27 \text{ 或 } (3 643.53 \text{ 美元}, 4 024.07 \text{ 美元})$$

$$c. s_{\text{pred}} = s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 145.89 \times \sqrt{1 + \frac{1}{6} + \frac{(3 - 3.2)^2}{0.74}} = 161.19$$

$$\hat{y}^* \pm t_{\alpha/2} s_{\text{pred}}$$

$$3 833.8 \pm 2.776 \times 161.19 = 3 833.8 \pm 447.46 \text{ 或 } (3 386.34 \text{ 美元}, 4 281.26 \text{ 美元})$$

d. 正如所料, 预测区间比置信区间宽。这是基于这样的事实, 预测一名 GPA 为 3.0 的新学生的月薪比估计所有 GPA 为 3.0 学生的平均月薪要困难得多。

36. a. (112 190 美元, 119 810 美元)

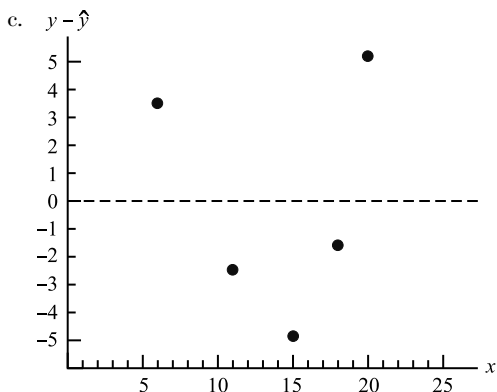
b. (104 710 美元, 127 290 美元)

38. a. 5 046.67 美元

- b. (3 815.10 美元, 6 278.24 美元)
 c. 没有超出规定的范围
40. a. 9
 b. $\hat{y} = 20.0 + 7.21x$
 c. 1.362 6
 d. $SSE = SST - SSR = 51\,984.1 - 41\,587.3 = 10\,396.8$
 $MSE = 10\,396.8/7 = 1\,485.3$
 $F = \frac{MSR}{MSE} = \frac{41\,587.3}{1\,485.3} = 28.0$
 根据分子自由度为 1, 分母自由度为 7 的 F 分布表可知, p -值小于 0.01
 利用 Excel 或 Minitab, 对应于 $F = 28.0$ 的 p -值是 0.0011
 因为 p -值 $\leq \alpha = 0.05$, 所以我们拒绝原假设 $H_0: \beta_1 = 0$
- e. $\hat{y} = 20.0 + 7.21 \times 50 = 380.5$ 或 380 500 美元
42. a. $\hat{y} = 80.0 + 50.0x$
 b. 30
 c. 显著; p -值 = 0.000
 d. 680 000 美元
44. b. 是
 c. $\hat{y} = 2\,044.38 - 28.35 \times \text{重量}$
 d. 显著; p -值 = 0.000。
 e. 0.774; 是一个好的拟合
45. a. $\bar{x} = \frac{\sum x_i}{n} = \frac{70}{5} = 14$, $\bar{y} = \frac{\sum y_i}{n} = \frac{76}{5} = 15.2$,
 $\sum (x_i - \bar{x})(y_i - \bar{y}) = 200$, $\sum (x_i - \bar{x})^2 = 126$
 $b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{200}{126} = 1.587\,3$
 $b_0 = \bar{y} - b_1\bar{x} = 15.2 - 1.587\,3 \times 14 = -7.022\,2$
 $\hat{y} = -7.02 + 1.59x$

b.

x_i	y_i	\hat{y}_i	$y_i - \hat{y}_i$
6	6	2.52	3.48
11	8	10.47	-2.47
15	12	16.83	-4.83
18	20	21.60	-1.60
20	30	24.78	5.22



仅利用 5 个观测值来判断假定是否成立, 这很困难。
 但是, 残差图的弯曲程度已经表明, 误差项的假定没有得到满足。这些数据的散点图还表明, 在 x 和 y 之间可能存在一种曲线关系。

d. $s^2 = 23.78$

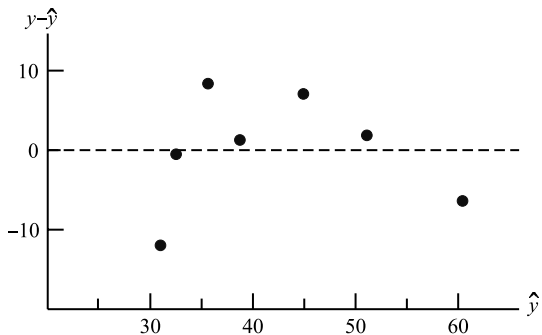
$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} = \frac{1}{5} + \frac{(x_i - 14)^2}{126}$$

x_i	h_i	$s_{y_i - \hat{y}_i}$	$y_i - \hat{y}_i$	标准化残差
6	0.707 9	2.64	3.48	1.32
11	0.271 4	4.16	-2.47	-0.59
15	0.207 9	4.34	-4.83	-1.11
18	0.327 0	4.00	-1.60	-0.40
20	0.485 7	3.50	5.22	1.49

- e. 关于 \hat{y} 的标准化残差图和最初的残差图有同样的图形。正如在 (c) 所陈述的那样, 观察到的弯曲程度表明, 关于误差项的假定没有得到满足。
46. a. $\hat{y} = 2.32 + 0.64x$
 b. 没有得到满足; 对于较大的 x 的值, 方差似乎在增加。
47. a. 设 x 表示广告费支出, y 表示收入
 $\hat{y} = 29.4 + 1.55x$
 b. $SST = 1\,002$, $SSE = 310.28$, $SSR = 691.72$
 $MSR = \frac{SSR}{1} = 691.72$
 $MSE = \frac{SSE}{n-2} = \frac{310.28}{5} = 62.055\,4$
 $F = \frac{MSR}{MSE} = \frac{691.72}{62.055\,4} = 11.15$
 根据分子自由度为 1, 分母自由度为 5 的 F 分布表可知, p -值介于 0.01 和 0.025 之间
 利用 Excel 或 Minitab, p -值 = 0.020 6
 因为 p -值 $\leq \alpha = 0.05$, 所以我们的结论是两变量是相关的

c.

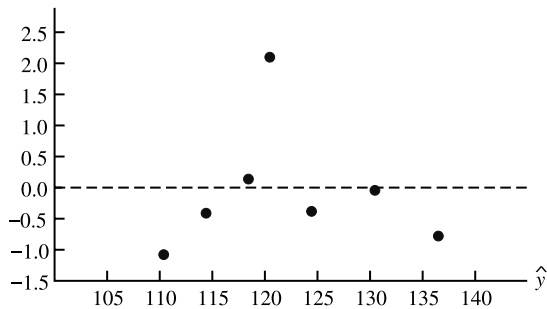
x_i	y_i	$\hat{y}_i = 29.4 + 1.55x$	$y_i - \hat{y}_i$
1	19	30.95	-11.95
2	32	32.50	-0.50
4	44	35.60	8.40
6	40	38.70	1.30
10	52	44.90	7.10
14	53	51.10	1.90
20	54	60.40	-6.40



d. 残差图使得我们对 x 和 y 之间存在线性关系的假定产生疑问。在 $\alpha=0.05$ 的显著性水平下，即使 x 和 y 之间的线性关系是显著的，但将该线性关系外推到数据范围以外是非常危险的做法。

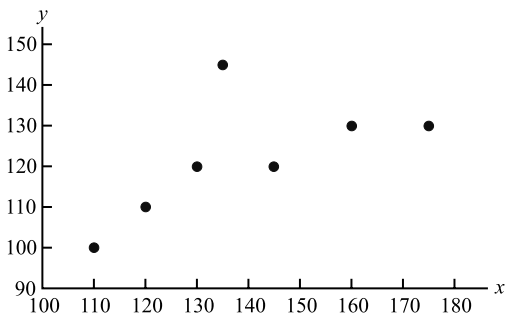
The regression equation is						
Y = 66.1 + 0.402 X						
Predictor	Coef	SE Coef	T	p		
Constant	66.10	32.06	2.06	0.094		
X	0.4023	0.2276	1.77	0.137		
S = 12.62		R-sq = 38.5%	R-sq(adj) = 26.1%			
Analysis of Variance						
SOURCE	DF	SS	MS	F	p	
Regression	1	497.2	497.2	3.12	0.137	
Residual Error	5	795.7	159.1			
Total	6	1292.9				
Unusual Observations						
Obs	X	Y	Fit	SE Fit	Residual	St Resid
1	135	145.00	120.42	4.87	24.58	2.11R
R denotes an observation with a large standardized residual						

b. 标准化残差



标准化残差图表明观测值 $x=135, y=145$ 可能是一个异常值；注意，这个观测值的标准化残差是 2.11。

c. 散点图如下所示：



48. b. 合理

50. a. 利用 Minitab，我们得到估计的回归方程 $\hat{y} = 66.1 + 0.402x$ ；Minitab 的部分输出如图 D14-50 所示；拟合值和标准化残差如下表所示：

x_i	y_i	\hat{y}_i	标准化残差
135	145	120.41	2.11
110	100	110.35	-1.08
130	120	118.40	0.14
145	120	124.43	-0.38
175	130	136.50	-0.78
160	130	130.47	-0.04
120	110	114.38	-0.41

散点图也表明观测值 $x=135, y=145$ 可能是一个异常值；这就意味着简单线性回归的异常值能够根据散点图识别出来。

52. b. $\hat{y} = 91.0 - 0.917x$
- d. Smithsonian Institution 是异常值
American Cancer Society 是有影响的观测值
54. a. 散点图确实表明潜在的异常值和（或）有影响的观测值。比如，纽约洋基队的收入和价值均为最高，似乎是一个有影响的观测值。洛杉矶道奇队的价值位于第二，似乎是一个异常值。
- b. $\hat{y} = -601.4814 + 5.9271 \times \text{收入}$
- c. 洛杉矶道奇队的标准化残差值是 4.7，应该被认为是一个异常值。
58. b. $\hat{y} = -669 + 0.157 \times \text{DJIA}$
- c. 显著； p -值 = 0.001
- d. $r^2 = 0.949$ ；非常好的拟合
60. b. 毕业率 (%) = $25.4 + 0.285 \times \text{保留率}(\%)$
- c. 显著； p -值 = 0.000
- d. 对观测数据的拟合不好； $r^2 = 0.449$
- e. 是的，关注所在学校的业绩

- f. 是的, 关注所在学校的业绩
62. a. $\hat{y} = 22.2 - 0.148x$
 b. 显著关系; p -值 = 0.028
 c. 好的拟合; $r^2 = 0.739$
 d. (12.294, 17.271)
64. a. $\hat{y} = 220 + 132x$
 b. 显著; p -值 = 0.000
 c. $r^2 = 0.873$; 非常好的拟合
 d. (559.50 美元, 933.90 美元)
66. a. 贝塔系数 = 0.95
 b. 显著; p -值 = 0.029
 c. $r^2 = 0.470$; 不是一个好的拟合
 d. 施乐公司具有较高的风险
68. b. 在这两个变量之间似乎存在一个负的线性关系
 c. $\hat{y} = 16.5 - 0.0588 \times \text{行驶里程}$
 d. 显著; p -值 = 0.000

- e. $r^2 = 0.539$; 相当不错的拟合
 g. 大约为 13 000 美元; 不是

第 15 章

2. a. 估计的回归方程是 $\hat{y} = 45.06 + 1.94x_1$
 当 $x_1 = 45$ 时, y 的估计值是
 $\hat{y} = 45.06 + 1.94 \times 45 = 132.26$
 b. 估计的回归方程是 $\hat{y} = 85.22 + 4.32x_2$
 当 $x_2 = 15$ 时, y 的估计值是
 $\hat{y} = 85.22 + 4.32 \times 15 = 150.02$
 c. 估计的回归方程是 $\hat{y} = -18.37 + 2.01x_1 + 4.74x_2$
 当 $x_1 = 45$, $x_2 = 15$ 时, y 的估计值是
 $\hat{y} = -18.37 + 2.01 \times 45 + 4.74 \times 15 = 143.18$
4. a. 255 000 美元
5. a. Minitab 的输出如图 D15-5a 所示。

The regression equation is					
Revenue = 88.6 + 1.60 TVAdv					
Predictor	Coef	SE Coef	T	p	
Constant	88.638	1.582	56.02	0.000	
TVAdv	1.6039	0.4778	3.36	0.015	
S = 1.215 R-sq = 65.3% R-sq(adj) = 59.5%					
Analysis of Variance					
SOURCE	DF	SS	MS	F	p
Regression	1	16.640	16.640	11.27	0.015
Residual Error	6	8.860	1.477		
Total	7	25.500			

- b. Minitab 输出如图 D15-5b 所示。

The regression equation is					
Revenue = 83.2 + 2.29 TVAdv + 1.30 NewsAdv					
Predictor	Coef	SE Coef	T	p	
Constant	83.230	1.574	52.88	0.000	
TVAdv	2.2902	0.3041	7.53	0.001	
NewsAdv	1.3010	0.3207	4.06	0.010	
S = 0.6426 R-sq = 91.9% R-sq(adj) = 88.7%					
Analysis of Variance					
SOURCE	DF	SS	MS	F	p
Regression	2	23.435	11.718	28.38	0.002
Residual Error	5	2.065	0.413		
Total	7	25.500			

- c. 在 (a) 中电视广告费用的系数是 1.60; 在 (b) 中电视广告费用的系数是 2.29。在 (a) 中, 系数是当电视广告费用改变一个单位时, 周的总收入改变的一个估计; 在 (b) 中, 系数是当报纸广告费用保持不变, 电视广告费用改变一个单位时, 周的总收入改变的一个

估计。

- d. 收入 = $83.2 + 2.95 \times 3.5 + 1.30 \times 1.8 = 93.56$ 或 93 560 美元
6. a. $\text{Win}\% = -58.8 + 16.4\text{Yds}/\text{Att}$
 b. $\text{Win}\% = 97.5 - 1.600\text{Int}/\text{Att}$

- c. $\text{Win}\% = -5.8 + 12.9 \text{ Yds/Att} - 1.084 \text{ Int/Att}$
 d. 35%
8. a. 总得分 = 69.3 + 0.235 岸上观光
 b. 总得分 = 45.2 + 0.253 岸上观光 + 0.248 食品/餐饮
 c. 87.76 或大约 88
10. a. $R/\text{IP} = 0.676 - 0.284\text{SO}/\text{IP}$
 b. $R/\text{IP} = 0.308 + 1.35\text{HR}/\text{IP}$
 c. $R/\text{IP} = 0.537 - 0.248\text{SO}/\text{IP} + 1.03\text{HR}/\text{IP}$
 d. 0.48
 e. 建议没有意义
12. a. $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{14052.2}{15182.9} = 0.926$
 b. $R_a^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1} = 1 - (1 - 0.926) \frac{10-1}{10-2-1} = 0.905$
 c. 是；用模型中自变量的个数修正后，我们看到 y 的 90.5% 的变异性能被模型解释。
14. a. 0.75
 b. 0.68
15. a. $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{23.435}{25.5} = 0.919$
 $R_a^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1} = 1 - (1 - 0.919) \frac{8-1}{8-2-1} = 0.887$
 b. 因为当利用两个自变量时， R^2 和 R_a^2 都表示 y 的变异性能被模型解释的比例增加了，所以我们更愿意使用多元回归分析方法。
16. a. 不能给出一个好的拟合，因为 $R^2 = 0.577$
 b. 使用多元回归方法得到更好的拟合
18. a. $R^2 = 0.563$, $R_a^2 = 0.512$
 b. 对观测数据的拟合不是很好
19. a. $\text{MSR} = \frac{\text{SSR}}{p} = \frac{6216.375}{2} = 3108.188$
 $\text{MSE} = \frac{\text{SSE}}{n-p-1} = \frac{507.75}{10-2-1} = 72.536$
 b. $F = \frac{\text{MSR}}{\text{MSE}} = \frac{3108.188}{72.536} = 42.85$
 根据分子自由度为 2，分母自由度为 7 的 F 分布表可知， p -值小于 0.01
 利用 Excel 或 Minitab，对应于 $F = 42.85$ 的 p -值是 0.0001
 因为 p -值 $\leq \alpha$ ，所以模型在总体上是显著的
- c. $t = \frac{b_1}{s_{b_1}} = \frac{0.5906}{0.0813} = 7.26$
 p -值 = 0.0002
 因为 p -值 $\leq \alpha$ ，所以 β_1 是显著的

$$\text{d. } t = \frac{b_2}{s_{b_2}} = \frac{0.4980}{0.0567} = 8.78$$

$$p\text{-值} = 0.0001$$

因为 p -值 $\leq \alpha$ ，所以 β_2 是显著的

20. a. 显著； p -值 = 0.000
 b. 显著； p -值 = 0.000
 c. 显著； p -值 = 0.002
22. a. $\text{SSE} = 4000$, $s^2 = 571.43$, $\text{MSR} = 6000$
 b. 显著； p -值 = 0.008
23. a. $F = 28.38$
 p -值 = 0.002
 因为 p -值 $\leq \alpha$ ，所以存在显著的关系
- b. $t = 7.53$
 p -值 = 0.001
 因为 p -值 $\leq \alpha$ ，所以 β_1 是显著的， x_1 不应从模型中删除
- c. $t = 4.06$
 p -值 = 0.010
 因为 p -值 $\leq \alpha$ ，所以 β_2 是显著的， x_2 不应从模型中删除
24. a. $\hat{y} = 60.5 + 0.3186\text{OffPassYds}/\text{G} - 0.2413\text{DefYds}/\text{G}$
 b. 显著； p -值 = 0.000
 c. $\text{OffPassYds}/\text{G}$ 是显著的自变量； p -值 = 0.000
 DefYds/G 是显著的自变量； p -值 = 0.011
26. a. 显著； p -值 = 0.001
 b. 所有的自变量全是显著的；所有的 p -值全都 $< \alpha = 0.05$
28. a. 利用 Minitab，95% 的置信区间是 (132.16, 154.16)
 b. 利用 Minitab，95% 的预测区间是 (111.13, 175.18)
29. a. 见图 D15-5b 的 Minitab 输出
 $\hat{y} = 83.23 + 2.29 \times 3.5 + 1.30 \times 1.8 = 93.555$ 或 93.555 美元
 b. Minitab 的输出结果：(92.840, 94.335) 或 (92.840 美元, 94.335 美元)
 c. Minitab 的输出结果：(91.774, 95.401) 或 (91.744 美元, 95.401 美元)
30. a. $\hat{y} = 60.5 + 0.3186 \times 223 - 0.2413 \times 300 = 59.1578$
 b. 部分 Minitab 输出结果如下表所示：

回归方程			
$\text{Win}\% = 60.5 + 0.3186\text{OffPassYds}/\text{G} - 0.2413\text{DefYds}/\text{G}$			
变量设		定值	
OffPassYds/G		225	
DefYds/G		300	
拟合值	标准差 拟合值	95% 置信 区间	95% 预测 区间
59.8270	4.88676	(49.8325, 69.8126)	(26.9589, 92.6952)

95% 的预测区间是 (26.958 9, 92.695 2)

32. a. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

式中, $x_2 = \begin{cases} 0 & \text{如果是水平 1} \\ 1 & \text{如果是水平 2} \end{cases}$

b. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 \times 0 = \beta_0 + \beta_1 x_1$

c. $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 \times 1 = \beta_0 + \beta_1 x_1 + \beta_2$

d. $\beta_2 = E(y | \text{水平 2}) - E(y | \text{水平 1})$

在 x_2 保持不变时, 若 x_1 变动一个单位, 则 $E(y)$ 变动 β_1 个单位

34. a. 15 300 美元

b. $\hat{y} = 10.1 - 4.2 \times 2 + 6.8 \times 8 + 15.3 \times 0 = 56.1$

预测的销售收入: 56 100 美元

c. $\hat{y} = 10.1 - 4.2 \times 1 + 6.8 \times 3 + 15.3 \times 1 = 41.6$

预测的销售收入: 41 600 美元

36. a. $\hat{y} = 1.86 + 0.291 \times \text{月数} + 1.10 \times \text{故障类型} - 0.609 \times \text{维修工}$

b. 显著; p -值 = 0.002

c. 维修工不显著; p -值 = 0.167

38. a. $\hat{y} = -91.8 + 1.08 \times \text{年龄} + 0.252 \times \text{血压} + 8.74 \times \text{吸烟者}$

b. 显著; p -值 = 0.01

c. 95% 的预测区间是 (21.35, 47.18) 或者发生中风的概率是 (0.213 5, 0.471 8); 戒烟并且开始进行某些治疗来降低病人的血压。

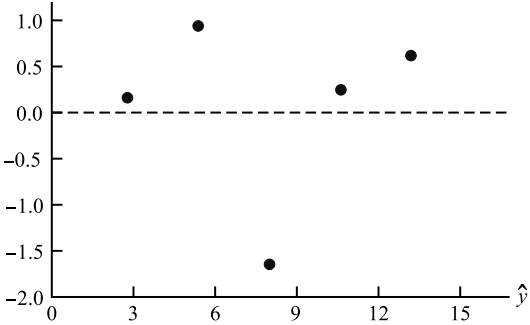
39. a. Minitab 的输出如图 D15-39 所示。

The regression equation is					
Y = 0.20 + 2.60 X					
Predictor	Coef	SE Coef	T	p	
Constant	0.200	2.132	0.09	0.931	
X	2.6000	0.6429	4.04	0.027	
S = 2.033		R-sq = 84.5%		R-sq(adj) = 79.3%	
Analysis of Variance					
SOURCE	DF	SS	MS	F	p
Regression	1	67.600	67.600	16.35	0.027
Residual Error	3	12.400	4.133		
Total	4	80.000			

b. Minitab 的输出提供了如下的数值:

x_i	y_i	\hat{y}_i	标准化残差
1	3	2.8	0.16
2	7	5.4	0.94
3	5	8.0	-1.65
4	11	10.6	0.24
5	14	13.2	0.62

标准化残差



点 (3, 5) 没有显现出其他点的趋势; 尽管这个点的标准化残差值为 -1.65, 但还没有大到使我们认为点 (3, 5) 是一个异常值的程度。

c. Minitab 的输出提供了如下的数值:

x_i	y_i	学生化删除残差
1	3	0.13
2	7	0.91
3	5	-4.42
4	11	0.19
5	14	0.54

$t_{0.025} = 4.30$ (自由度 $n - p - 2 = 5 - 1 - 2 = 2$)

因为点 (3, 5) 的学生化删除残差是 $-4.42 < -4.303$, 我们的结论是第三个观测值是一个异常值。

40. a. $\hat{y} = -53.3 + 3.11x$

b. -1.94, -0.12, 1.79, 0.40, -1.90; 无异常值

c. 0.38, 0.28, 0.22, 0.20, 0.92; 不存在有影响的观测值

d. 0.60, 0.00, 0.26, 0.03, 11.09; 有, 第 5 个观测值是有影响的观测值

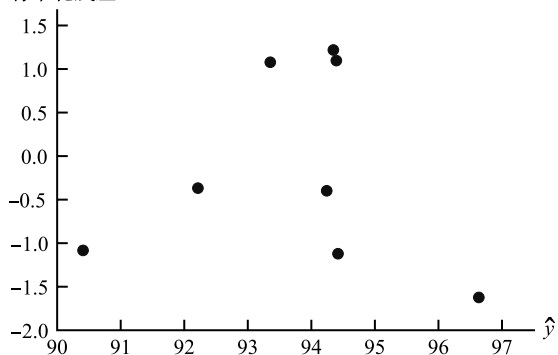
41. a. Minitab 的输出如图 D15-5b 所示。

估计的回归方程是: 周总收入 = $83.2 + 2.29 \times \text{电视广告费用} + 1.30 \times \text{报纸广告费用}$

b. Minitab 的输出提供了如下的数值:

\hat{y}_i	标准化残差	\hat{y}_i	标准化残差
96.63	-1.62	94.39	1.10
90.41	-1.08	94.24	-0.40
94.34	1.22	94.42	-1.12
92.21	-0.37	93.35	1.08

标准化残差



利用相对很少的观测值,来证实有关误差项 ε 的假定是否被违反是一项困难的工作;例如,一种观点认为在标准化残差图中似乎没有显现出任何形式的图形;而另一种观点则认为在标准化残差图中有一条曲线图形。

- c. 因为标准化残差的数值介于 -2 和 $+2$ 之间;于是,利用这一检验可知,在数据中不存在异常值。

为了进一步检测异常值,我们利用 Minitab 计算学生化删除残差如下:

观测值	学生化删除残差	观测值	学生化删除残差
1	-2.11	5	1.13
2	-1.10	6	-0.36
3	1.31	7	-1.16
4	-0.33	8	1.10

$$t_{0.025} = 2.776 \quad (\text{自由度 } n - p - 2 = 8 - 2 - 2 = 4)$$

因为没有学生化删除残差小于 -2.776 或大于 2.776 , 所以我们的结论是在数据中不存在异常值。

- d. Minitab 的输出提供了如下的数值:

观测值	h_i	D_i
1	0.63	1.52
2	0.65	0.70
3	0.30	0.22
4	0.23	0.01
5	0.26	0.14
6	0.14	0.01
7	0.66	0.81
8	0.13	0.06

杠杆率的临界值是:

$$\frac{3(p+1)}{n} = \frac{3 \times (2+1)}{8} = 1.125$$

因为杠杆率的值没有一个超过 1.125 , 所以我们的结论是在数据中不存在有影响的观测值;然而,利用库克距离测度,我们看到 $D_1 > 1$ (经验准则的临界值);于是,我们的结论是第一个观测值是一个有影响的观测值。

最后的结论是第一个观测值是一个有影响的观测值。

42. b. 有不寻常的趋势
c. 没有异常值

- d. 观测值 2 是一个有影响的观测值

44. a. $E(y) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$

- b. 没有 Simmons 信用卡的顾客购买商品概率的估计

c. $\hat{g}(x) = -0.9445 + 1.0245x$

- d. 没有 Simmons 信用卡的顾客使用优惠券购买商品概率的估计值是 0.28 。

拥有 Simmons 信用卡的顾客使用优惠券购买商品概率的估计值是 0.52 。

- e. 估计的机会比率 = 2.79

46. a. $E(y) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$

b. $E(y) = \frac{e^{-2.6355 + 0.22018x}}{1 + e^{-2.6355 + 0.22018x}}$

- c. 显著; p -值 = 0.0002

- d. 0.39

- e. 1200 美元

- f. 估计的机会比率 = 1.25

48. a. $E(y) = \frac{e^{\beta_0 + \beta_1 x + \beta_2 x}}{1 + e^{\beta_0 + \beta_1 x + \beta_2 x}}$

b. $\hat{g}(x) = -39.4982 + 3.37449 \text{ Wet} + 1.81628 \text{ Noise}$

- c. 0.88

- d. 概率是 0.04

50. b. 155.14

52. a. $\hat{y} = -1.41 + 0.0235x_1 + 0.00486x_2$

- b. 显著; p -值 = 0.0001

- c. 两个都显著

d. $R^2 = 0.937$; $R_a^2 = 0.919$; 好的拟合

54. a. 再次购买 = $-7.522 + 1.8151 \times \text{转向}$

- b. 对观测数据是一个好的拟合

c. 再次购买 = $-5.388 + 0.6899 \times \text{转向} + 0.9113 \times \text{胎面磨损}$

- d. 显著; p -值 = 0.001

56. a. $\hat{y} = 4.9090 + 10.4658 \times \text{国内股票} + 21.6823 \times \text{国际股票}$

- b. $R^2 = 0.6144$; 相当好的拟合

c. $\hat{y} = 1.1899 + 6.8969 \times \text{国内股票} + 17.6800 \times \text{国际股票} + 0.0265 \times \text{资产净值 (美元)} + 6.4564 \times \text{费用比率 (\%)}$

资产净值 (美元) 是不显著的自变量, 能从估计的回归方程中删除。

d. $\hat{y} = -4.6074 + 8.1713 \times \text{国内股票} + 19.5194 \times \text{国际股票} + 5.5197 \times \text{费用比率 (\%)} + 5.9237 \times 3 \text{ 星级基金} + 8.2367 \times 4 \text{ 星级基金} + 6.6241 \times 5 \text{ 星级基金}$

- e. 15.28%

第 16 章

1. a. Minitab 的输出如图 D16-1a 所示。

The regression equation is
 $Y = - 6.8 + 1.23 X$

Predictor	Coef	SE Coef	T	p
Constant	-6.77	14.17	-0.48	0.658
X	1.2296	0.4697	2.62	0.059

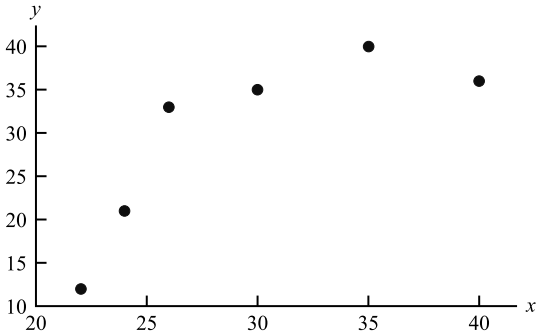
S = 7.269 R-sq = 63.1% R-sq(adj) = 53.9%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	362.13	362.13	6.85	0.059
Residual Error	4	211.37	52.84		
Total	5	573.50			

b. 因为对应于 $F=6.85$ 的 p -值 $=0.059 > \alpha=0.05$ ，所以关系是不显著的。

c.



散点图暗示：一曲线关系可能是合适的

d. Minitab 的输出如图 D16-1d 所示。

The regression equation is
 $Y = - 169 + 12.2 X - 0.177 X^2$

Predictor	Coef	SE Coef	T	p
Constant	-168.88	39.79	-4.74	0.024
X	12.187	2.663	4.58	0.020
XSQ	-0.17704	0.04290	-4.13	0.026

S = 3.248 R-sq = 94.5% R-sq(adj) = 90.8%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	2	541.85	270.92	25.68	0.013
Residual Error	3	31.65	10.55		
Total	5	573.50			

e. 因为对应于 $F=25.68$ 的 p -值 $=0.013 < \alpha=0.05$ ，所以关系是显著的。

f. $\hat{y} = -168.88 + 12.187 \times 25 - 0.17704 \times 25^2 = 25.145$

2. a. $\hat{y} = 9.32 + 0.424x$; p -值 $=0.117$ 表示在 x 和 y 之间存在一种弱相关关系。

b. $\hat{y} = -8.10 + 2.41x - 0.0480x^2$
 $R_a^2 = 0.932$ ；一个好的拟合

c. 20.965

4. a. $\hat{y} = 943 + 8.71x$

b. 显著; p -值 $=0.005 < \alpha=0.01$

5. a. Minitab 的输出如图 D16-5a 所示。

b. Minitab 的输出如图 D16-12b 所示。

Analysis of Variance					
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	229.677	76.559	641.25	0.000
Putting Ave.	1	122.801	122.801	1028.58	0.000
Greens in Reg.	1	149.336	149.336	1250.83	0.000
Drive Accuracy	1	0.000	0.000	0.00	0.964
Error	130	15.521	0.119		
Total	133	245.197			

S	R-sq	R-sq(adj)	R-sq(pred)
0.345528	93.67%	93.52%	93.08%

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	57.16	1.06	53.95	0.000	
Putting Ave.	1.0319	0.0322	32.07	0.000	1.04
Greens in Reg.	-23.102	0.653	-35.37	0.000	1.05
Drive Accuracy	-0.022	0.481	-0.05	0.964	1.02

Regression Equation
 Scoring Ave. = 57.16 + 1.0319 Putting Ave. - 23.102
 Greens in Reg. - 0.022 Drive Accuracy

$$c. F = \frac{\frac{\text{SSE(简化)} - \text{SSE(完全)}}{\text{增加的项数}}}{\frac{\text{MSE(完全)}}{0.119}} = \frac{166.95 - 15.521}{0.119} = 634.19$$

对应于 $F = 634.19$ (分子自由度为 2, 分母自由度为 130) 的 p -值是 0.00; 因为 p -值 $< \alpha = 0.05$, 所以增加的两个自变量在统计上是显著的。

14. a. $\hat{y} = -111 + 1.32 \times \text{年龄} + 0.296 \times \text{血压}$
 b. $\hat{y} = -123 + 1.51 \times \text{年龄} + 0.448 \times \text{血压} + 8.87 \times \text{吸烟者} - 0.00276 \times \text{年龄} \times \text{血压}$
 c. 显著; p -值 = 0.000
16. a. 周数 = $-8.9 + 1.51 \times \text{年龄}$
 b. 周数 = $-0.07 + 1.73 \times \text{年龄} - 2.7 \times \text{管理人员} - 15.1 \times \text{户主} - 17.4 \times \text{销售人员}$
 c. 与 (b) 中得到的最佳的估计的回归方程相同
 d. 与 (b) 中得到的最佳的估计的回归方程相同
 e. 周数 = $13.1 + 1.64 \times \text{年龄} - 9.76 \times \text{已婚} - 19.4 \times \text{户主} - 29.0 \times \text{管理人员} - 19.0 \times \text{销售人员}$
18. a. $\text{RPG} = -4.05 + 27.60\text{BP}$
 b. 各种各样的模型将给出一个好的拟合; 利用 Minitab 的逐步回归方法, 选取 Alpha-to-Enter = 0.10 和 Alpha-to-Remove = 0.10, 得到了 5 个变量的模型如下:
 $\text{RPG} = -0.0909 + 32.2 \text{ OBP} + 0.109 \text{ HR} - 21.5 \text{ AVG} + 0.2443 \text{ B} - 0.0223 \text{ BB}$

20.

x_1	x_2	x_3	处理
0	0	0	A
1	0	0	B
0	1	0	C
0	0	1	D

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

22. 因素 A: 如果水平 1, 则 $x_1 = 0$; 如果水平 2, 则 $x_1 = 1$
 因素 B:

x_1	x_2	水平
0	0	1
1	0	2
0	1	3

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1 x_3$$

23. a. 定义两个虚拟变量如下表所示。

D1	D2	制造商
0	0	1
1	0	2
0	1	3

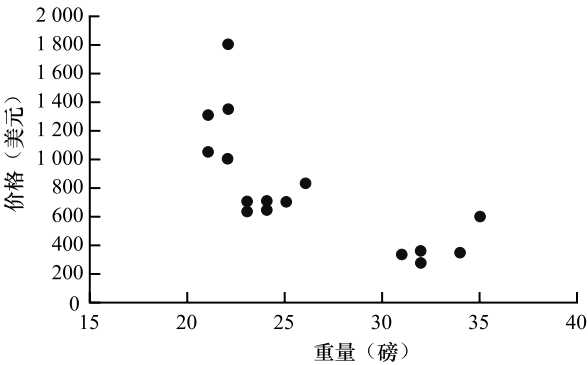
$$E(y) = \beta_0 + \beta_1 D1 + \beta_2 D2$$

b. Minitab 的输出如下图所示。

The regression equation is				
TIME = 23.0 + 5.00 D1 - 2.00 D2				
Predictor	Coef	SE Coef	T	p
Constant	23.000	1.106	20.80	0.000
D1	5.000	1.563	3.20	0.011
D2	-2.000	1.563	-1.28	0.233

S = 2.211 R-Sq = 70.3% R-Sq(adj) = 63.7%					
Analysis of Variance					
SOURCE	DF	SS	MS	F	p
Regression	2	104.000	52.000	1064	0.004
Residual Error	9	44.000	4.889		
Total	11	148.000			

- c. $H_0: \beta_1 = \beta_2 = 0$
- d. 因为 p -值 $= 0.004 < \alpha = 0.05$ ，所以我们拒绝 H_0 并且得出结论对不同制造商生产的搅拌机，搅拌一批原料需要的平均时间是不一样的。
24. a. 在 $\alpha = 0.05$ 的显著性水平下，不存在显著差异； p -值 $= 0.093$
- b. 139
26. 在总体上是显著的； p -值 $= 0.029$
- 单独地考虑，在 $\alpha = 0.05$ 的显著性水平下，没有一个变量是显著的；一个较大的样本将是有益的。
28. $d = 1.60$ ；杜宾 - 瓦特森检验不能确定是否存在自相关性。
30. a.



重量和价格之间似乎是一种曲线关系。

b. Minitab 的部分输出如下图所示。

The regression equation is					
Price = 11376 - 728 Weight + 12.0 WeightSq					
Predictor	Coef	SE Coef	T	p	
Constant	11376	2565	4.43	0.000	
Weight	-728.3	193.7	-3.76	0.002	
WeightSq	11.974	3.539	3.38	0.004	
S = 242.804 R-Sq = 77.0% R-Sq(adj) = 74.1%					
Analysis of Variance					
SOURCE	DF	SS	MS	F	p
Regression	2	3161747	1580874	26.82	0.000
Residual Error	16	943263	58954		
Total	18	4105011			

Minitab 得到的结果支持结论：重量和价格之间存在一种曲线关系。

c. Minitab 的部分输出如下图所示。

The regression equation is					
Price = 1284 - 572 Type_Fitness - 907 Type_Comfort					
Predictor	Coef	SE Coef	T	p	
Constant	1283.75	95.22	13.48	0.000	
Type_Fitness	-571.8	153.5	-3.72	0.002	
Type_Comfort	-907.1	145.5	-6.24	0.000	

S = 269.328 R-Sq = 71.7% R-Sq(adj) = 68.2%					
Analysis of Variance					
SOURCE	DF	SS	MS	F	p
Regression	2	2944410	1472205	20.30	0.000
Residual Error	16	1160601	72538		
Total	18	4105011			

在预测自行车的价格时，自行车的类型似乎是一个显著的因素，但是在（b）中建立的估计的回归方程似乎给出了一个略微好一些的拟合。

d. Minitab 的部分输出如下所示。在这一输出中，WxF 表示自行车重量与虚拟变量 Type-Fitness 之间的交互作用；WxC 表示自行车重量与虚拟变量 Type-Comfort 之间的交互作用。

The regression equation is					
Price = 5924 - 215 Weight - 6343 Type_Fitness - 7232					
Type_Comfort + 261 WxF + 266 WxC					
Predictor	Coef	SE Coef	T	p	
Constant	5924	1547	3.83	0.002	
Weight	-214.56	71.42	-3.00	0.010	
Type_Fitness	-6343	2596	-2.44	0.030	
Type_Comfort	-7232	2518	-2.87	0.013	
WxF	261.3	111.8	2.34	0.036	
WxC	266.41	93.98	2.83	0.014	
S = 224.438 R-Sq = 84.0% R-Sq(adj) = 77.9%					
Analysis of Variance					
SOURCE	DF	SS	MS	F	p
Regression	5	3450170	690034	13.70	0.000
Residual Error	13	654841	50372		
Total	18	4105011			

考虑到自行车类型、重量以及这两个变量之间的相互作用，估计的回归方程给出了非常好的拟合。

32. a. $\text{Delay} = 63.0 + 11.1 \text{ Industry}$ ；不存在显著的正的自相关性
34. 三种类型的顾客在逛商店时的舒适度存在显著差异； p -值 = 0.034

第 17 章

1. 下表中给出了（a）、（b）和（c）中的计算结果。

周	时间序列值	预测值	预测误差	预测误差的绝对值	预测误差平方	百分数误差	百分数误差的绝对值
1	18						
2	13	18	-5	5	25	-38.46	38.46
3	16	13	3	3	9	18.75	18.75
4	11	16	-6	5	25	-45.45	45.45
5	17	11	6	6	36	35.29	35.29
6	14	17	-3	3	9	-21.43	21.43
合计				22	104	-51.30	159.38

- a. $\text{MAE} = \frac{22}{5} = 4.4$
b. $\text{MSE} = \frac{104}{5} = 20.8$
c. $\text{MAPE} = \frac{159.38}{5} = 31.88$
d. 第 7 周的预测值为 14

2. 下表中给出了 (a)、(b) 和 (c) 中的计算结果。

周	时间 序列值	预测值	预测 误差	预测误差的 绝对值	预测 误差平方	百分数 误差	百分数误差的 绝对值
1	18						
2	13	18.00	-5.00	5.00	25.00	-38.46	38.46
3	16	15.50	0.50	0.50	0.25	3.13	3.13
4	11	15.67	-4.67	4.67	21.81	-42.45	42.45
5	17	14.50	2.50	2.50	6.25	14.71	14.71
6	14	15.00	-1.00	1.00	1.00	-7.14	7.14
合计				13.67	54.31	-70.21	105.86

- a. $MAE = \frac{13.67}{5} = 2.73$
- b. $MSE = \frac{54.31}{5} = 10.86$
- c. $MAPE = \frac{105.89}{5} = 21.18$
- d. 第7周的预测值为 $\frac{18 + 13 + 16 + 11 + 17 + 14}{6} = 14.83$
4. a. $MSE = \frac{363}{6} = 60.5$
- 8月的预测值是15
- b. $MSE = \frac{216.72}{6} = 36.12$
- 8月的预测值是18
- c. 所有过去数值的平均值较好，因为 MSE 较小。
5. a. 数据显示有水平模式。
- b. 3周的移动平均

周	时间序列值	预测值	预测误差	预测误差的平方
1	18			
2	13			
3	16			
4	11	15.67	-4.67	21.78
5	17	13.33	3.67	13.44
6	14	14.67	-0.67	0.44
				35.67

- $MSE = \frac{35.67}{3} = 11.89$
- 第7周的预测值 = $\frac{11 + 17 + 14}{3} = 14$
- c. 平滑常数 = 0.2

周	时间序列值	预测值	预测误差	预测误差的平方
1	18			
2	13	18.00	-5.00	25.00
3	16	17.00	-1.00	1.00
4	11	16.80	-5.80	33.64
5	17	15.64	1.36	1.85
6	14	15.91	-1.91	3.66
				65.15

- $MSE = \frac{65.15}{5} = 13.03$
- 第7周的预测值 = $0.2 \times 14 + (1 - 0.2) \times 15.91 = 15.53$
- d. 3周的移动平均提供较好的预测，因为 MSE 较小。
- e. 平滑常数 = 0.4

周	时间序列值	预测值	预测误差	预测误差的平方
1	18			
2	13	18.00	-5.00	25.00
3	16	16.00	0.00	0.00
4	11	16.00	-5.00	25.00
5	17	14.00	3.00	9.00
6	14	15.20	-1.20	1.44
				60.44

- $MSE = \frac{60.44}{5} = 12.09$
- 用 $\alpha = 0.4$ 比用 $\alpha = 0.2$ 提供了较好的指数平滑预测，因为它的 MSE 较小。
6. a. 数据显示有水平模式。
- b. $MSE = \frac{110}{4} = 27.5$
- 第8周的预测值为19
- c. $MSE = \frac{252.87}{6} = 42.15$
- 第7周的预测值为19.12
- d. 3周的移动平均提供较好的预测，因为 MSE 较小。
- e. $MSE = 39.79$
- 用 $\alpha = 0.4$ 比用 $\alpha = 0.2$ 提供较好的指数平滑预测，因为它的 MSE 较小。
8. a.
- | 周 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 预测值 | 19.33 | 21.33 | 19.83 | 17.83 | 18.33 | 18.33 | 20.33 | 20.33 | 17.83 |
- b. $MSE = 11.49$
- 这里偏好未加权移动平均；它有较小的 MSE
- c. 你总能找到一个至少与未加权移动平均一样好的加权移动平均；事实上，未加权移动平均是权数都相等时的加

权平均。

10. b. 在确定预测时, 较近期的数据得到较大的权数或重要性; 移动平均法在确定预测值时, 对最近 n 个数据值赋予相同的权数。

12. a. 数据显示有水平模式。

b. $MSE(3 \text{ 个月}) = 0.12$

$MSE(4 \text{ 个月}) = 0.14$

用 3 个月的移动平均

c. 9.63

13. a. 数据显示有水平模式。

b.

月份	时间序列值	3 个月移动平均预测值	误差平方	$\alpha = 0.2$ 预测	误差平方
1	240				
2	350		240.00	12 100.00	
3	230		262.00	1 024.00	
4	260	273.33	177.69	255.60	19.36
5	280	280.00	0.00	256.48	553.19
6	320	256.67	4 010.69	261.18	3 459.79
7	220	286.67	4 444.89	272.95	2 803.70
8	310	273.33	1 344.69	262.36	2 269.57
9	240	283.33	1 877.49	271.89	1 016.97
10	310	256.67	2 844.09	265.51	1 979.36
11	240	286.67	2 178.09	274.41	1 184.05
12	230	263.33	1 110.89	267.53	1 408.50
合计			17 988.52		27 818.49

$MSE(3 \text{ 个月}) = 17\,988.52/9 = 1\,998.72$

$MSE(\alpha = 0.2) = 27\,818.49/11 = 2\,528.95$

根据上述的 MSE 数值, 3 个月的移动平均比较合适。但是, 由于指数平滑法包含第 2 个月而处于不利的地位, 这对于任何一种预测方法都是很困难的。如果仅用 4~12 月的误差平方, 指数平滑法的 MSE 为:

$MSE(\alpha = 0.2) = 14\,694.49/9 = 1\,632.72$

因此, 考虑 4~12 月的预测时指数平滑法比较合适。

- c. 使用指数平滑法

$F_{13} = \alpha Y_{12} + (1 - \alpha) F_{12}$

$= 0.20 \times 230 + 0.80 \times 267.53 = 260$

14. a. 数据显示有水平模式。

- b. 2~12 月的数字如下:

105.00	114.00	115.80	112.56	105.79	110.05
120.54	126.38	118.46	106.92	104.85	

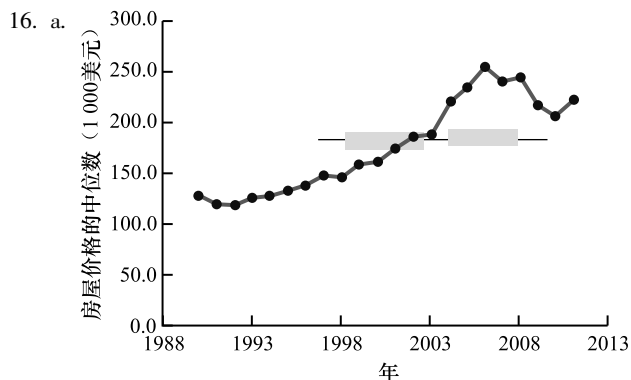
$MSE = 510.29$

- c. 2~12 月的数字如下:

105.00	120.00	120.00	112.50	101.25	110.63
127.81	133.91	116.95	98.48	99.24	

$MSE = 540.55$

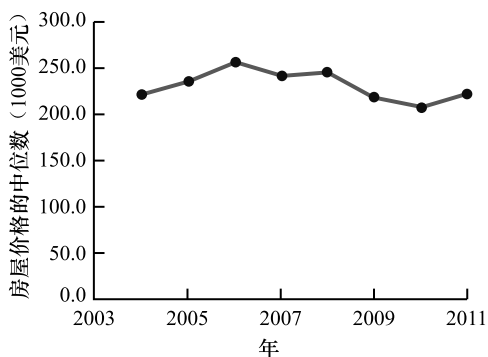
结论: 平滑常数 0.3 比平滑常数 0.5 要好, 因为 0.3 的 MSE 较小。



时间序列图表明数据存在趋势模式; 尽管 2008 年的经济衰退导致价格开始下降, 2010~2011 年价格的中位数上升。

- b. 本节讨论的方法仅适合于有水平模式的时间序列, 因此如果数据中的确存在趋势模式, 本节讨论的方法将不适合。

- c. 2003 年价格的中位数为 189 500 美元, 2004 年价格的中位数为 222 300 美元。于是, 这显示时间序列在 2004 年移动到一个新水平, 使用 2004 年及随后数据的时间序列图如下。



这个时间序列图显示有水平模式, 因此本节讨论的方法适合。

17. a. 时间序列图呈现线性趋势。

b. $\bar{t} = \frac{\sum_{i=1}^n t}{n} = \frac{15}{5} = 3, \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{55}{5} = 11$

$\sum (t - \bar{t})(Y_i - \bar{Y}) = 21, \sum (t - \bar{t})^2 = 10$

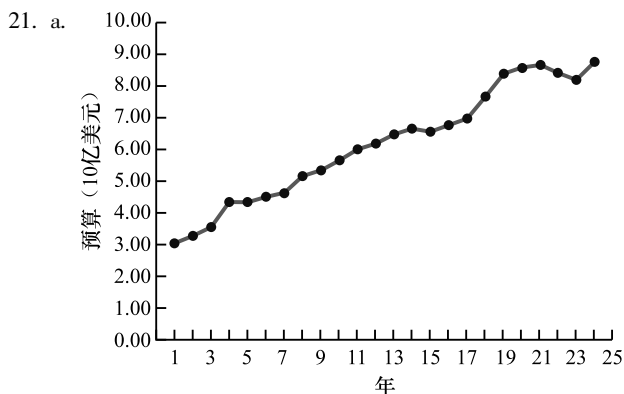
$b_1 = \frac{\sum_{i=1}^n (t - \bar{t})(Y_i - \bar{Y})}{\sum_{i=1}^n (t - \bar{t})^2} = \frac{21}{10} = 2.1$

$b_0 = \bar{Y} - b_1 \bar{t} = 11 - 2.1 \times 3 = 4.7$

$T_i = 4.7 + 2.1t$

- c. $T_t = 4.7 + 2.1 \times 6 = 17.3$
 18. a. 时间序列图呈现向下的线性趋势。
 b. $T_t = 11 - 1.0357t$
 c. $T_t = 11 - 1.0357 \times 8 = 2.7$

20. a. 时间序列图呈现曲线趋势。
 b. $T_t = 107.857 - 28.9881t + 2.6547t^2$
 c. 45.86



$$b. \bar{t} = \frac{\sum_{i=1}^n t}{n} = \frac{300}{24} = 12.5, \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{148.2}{24} = 6.175$$

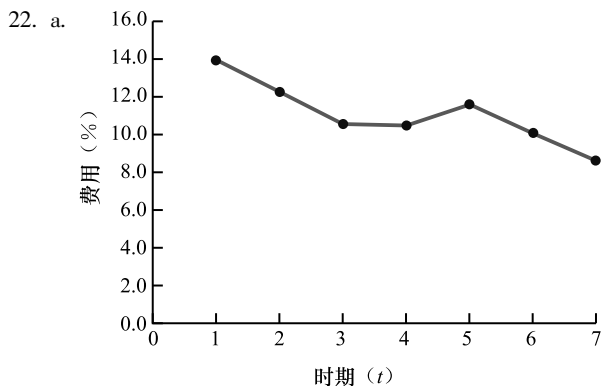
$$\sum (t - \bar{t})(Y_t - \bar{Y}) = 290.86, \sum (t - \bar{t})^2 = 1150$$

$$b_1 = \frac{\sum_{i=1}^n (t - \bar{t})(Y_t - \bar{Y})}{\sum_{i=1}^n (t - \bar{t})^2} = \frac{290.86}{1150} = 0.25292$$

$$b_0 = \bar{Y} - b_1 \bar{t} = 6.175 - 0.25292 \times 12.5 = 3.0135$$

$$c. \hat{y} = 3.0135 + 0.25292 \times 25 = 9.34$$

2012 年的预测值为 93.4 亿美元。



时间序列图呈现向下的线性趋势。

$$b. \bar{t} = \frac{\sum_{i=1}^n t}{n} = \frac{28}{7} = 4, \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{77}{7} = 11$$

$$\sum (t - \bar{t})(Y_t - \bar{Y}) = -19.6, \sum (t - \bar{t})^2 = 28$$

$$b_1 = \frac{\sum_{i=1}^n (t - \bar{t})(Y_t - \bar{Y})}{\sum_{i=1}^n (t - \bar{t})^2} = \frac{-19.6}{28} = -0.7$$

$$b_0 = \bar{Y} - b_1 \bar{t} = 11 - (-0.7) \times 4 = 13.8$$

$$T_t = 13.8 - 0.7t$$

$$c. 2013 \text{ 对应的时期为 } t=8, T_8 = 13.8 - (-0.7) \times 8 = 8.2$$

d. 如果 SCF 每年能以 0.7% 的幅度继续减少花在管理和资金筹集上的费用, 则 2018 年费用的预测值为 4.70%。

24. a. 时间序列图呈现线性趋势。

$$b. T_t = 7.5623 - 0.07541t$$

$$c. 6.7328$$

d. 在全球市场条件不确定的情况下, 仅用时间对 12 月进行预测是不可取的。

26. a. 线性趋势是不适当的。

$$b. T_t = 5.702 + 2.889t - 1.618t^2$$

$$c. 17.90$$

28. a. 时间序列图显示一种水平模式, 但数据中有季节模式。例如, 每年最低值出现在第二季度, 最高值出现在第四季度。

- b. 部分 Minitab 回归输出结果如下:

The regression equation is Value = 77.0 - 10.0 Qtr1 - 30.0 Qtr2 - 20.0 Qtr3

- c. 下一年的季节预测如下:

第一季度预测值 = $77.0 - 10.0 \times 1 - 30.0 \times 0 - 20.0 \times 0 = 67$

第二季度预测值 = $77.0 - 10.0 \times 0 - 30.0 \times 1 - 20.0 \times 0 = 47$

第三季度预测值 = $77.0 - 10.0 \times 0 - 30.0 \times 0 - 20.0 \times 1 = 57$

第四季度预测值 = $77.0 - 10.0 \times 0 - 30.0 \times 0 - 20.0 \times 0 = 77$

30. a. 这里显示数据中有季节模式, 且或许有适当的向上线性趋势。

- b. 部分 Minitab 回归输出结果如下:

The regression equation is Value = 2492 - 712 Qtr1 - 1512 Qtr2 + 327 Qtr3

- c. 下一年的季节预测如下:

第一季度预测值为 1780

第二季度预测值为 980

第三季度预测值为 2819

第四季度预测值为 2492

- d. 部分 Minitab 回归输出结果如下:

The regression equation is Value = 2307 - 642 Qtr1 - 1465 Qtr2 + 350 Qtr3 + 23.1 t
--

下一年的季节预测如下:

第一季度预测值为 2058

第二季度预测值为 1258

第三季度预测值为 3096

第四季度预测值为 2769

32. a. 时间序列图显示有线性趋势和季节影响。
b. 部分 Minitab 回归输出结果如下：

The regression equation is
Revenue = 70.0 + 10.0 Qtr1 + 105
Qtr2 + 245 Qtr3

第一季度预测值为 80
第二季度预测值为 174
第三季度预测值为 315
第四季度预测值为 70

- c. 部分 Minitab 回归输出结果如下：

The regression equation is
Revenue = -70.1 + 45.0 Qtr1 + 128
Qtr2 + 257 Qtr3 + 11.7 Period

第一季度预测值为 221
第二季度预测值为 315
第三季度预测值为 456
第四季度预测值为 211

34. a. 时间序列图显示有季节和线性趋势影响。
b. 注意：如果是 1 月，则 Jan = 1，否则 Jan = 0；如果是 2 月，则 Feb = 1，否则 Feb = 0；依此类推。
部分 Minitab 回归输出结果如下：

The regression equation is
Expense = 175 - 18.4 Jan - 3.72 Feb +
12.7 Mar + 45.7 Apr + 57.1
May + 135 Jun + 181 Jul + 105
Aug + 47.6 Sep + 50.6 Oct +
35.3 Nov + 1.96 Period

- c. 注意：时间序列下一时期是 Period = 37（第 4 年的 1 月）；1 ~ 12 月的预测值为 229, 246, 264, 299, 312, 392, 440, 366, 311, 316, 302, 269。

35. a. 时间序列图显示有线性趋势和季节模式。
b.

年	季度	时间序列值	4 个季度 移动平均	中心化的 移动平均
	1	4		
	2	2		
			3.50	
1	3	3		3.750
			4.00	
	4	5		4.125
			4.25	
	1	6		4.75
			4.75	
	2	3		5.000
			5.25	
2	3	5		5.375
			5.50	
	4	7		5.587
			6.25	

(续)

年	季度	时间序列值	4 个季度 移动平均	中心化的 移动平均
	1	7		6.375
			6.50	
3	2	6		6.625
			6.75	
	3	6		
	4	8		

c.

年	季度	时间序列值	中心化的 移动平均	季节不规则 成分
	1	4		
	2	2		
1	3	3	3.750	0.800
	4	5	4.125	1.212
	1	6	4.500	1.333
2	2	3	5.000	0.600
	3	5	5.375	0.930
	4	7	5.875	1.191
	1	7	6.375	1.098
3	2	6	6.625	0.906
	3	6		
	4	8		

季度	季节 不规则	成分值	季节指数	调整后的 季节指数
1	1.333	1.098	1.216	1.205
2	0.600	0.906	0.752	0.746
3	0.800	0.930	0.865	0.857
4	1.212	1.191	1.201	1.191
合计			4.036	

季节指数的调整系数 = $\frac{4}{4.036} = 0.991$

36. a.

年	季度	消除季节影响的数值
	1	3.320
	2	2.681
1	3	3.501
	4	4.198
	1	4.979
2	2	4.021
	3	5.834
	4	5.877
	1	5.809
3	2	8.043
	3	7.001
	4	6.717

- b. 令 $\text{Period} = 1$ 表示 1 年第一季度的时间序列值; $\text{Period} = 2$ 表示 1 年第二季度的时间序列值, 以此类推; 将 Period 作为自变量, 将 $\text{Deseasonalized Values}$ 作为因变量的值, 部分 Minitab 回归输出结果如下:

The regression equation is
Deseasonalized Value = 2.42 + 0.422
Period

- c. 第 4 年 (Period 为 13, 14, 15 和 16) 的消除季节影响的趋势预测值如下:
第一季度预测值为 7.906
第二季度预测值为 8.328
第三季度预测值为 8.750
第四季度预测值为 9.172
- d. 调整消除季节影响的趋势预测值得到如下季度估计值:
第一季度预测值为 9.527
第二季度预测值为 6.213
第三季度预测值为 7.449
第四季度预测值为 10.924
38. a. 时间序列图显示有线性趋势和季节影响。
b. 0.71, 0.78, 0.83, 0.97, 1.02, 1.30, 1.50, 1.23, 0.98, 0.99, 0.93, 0.79

c.

月	消除季节影响的支出
1	239.44
2	230.77
3	246.99
4	237.11
5	235.29
6	242.31
7	240.00
8	235.77
9	244.90
10	242.42
11	247.31
12	246.84
13	253.52
14	262.82
15	259.04
16	252.58
17	259.80
18	253.85
19	266.67
20	272.36
21	265.31
22	272.73
23	274.19

(续)

月	消除季节影响的支出
24	278.48
25	274.65
26	269.23
27	277.11
28	288.66
29	284.31
30	300.00
31	280.00
32	268.29
33	295.92
34	297.98
35	301.08
36	316.46

- b. 令 $\text{Period} = 1$ 表示 1 年 1 月的时间序列值; $\text{Period} = 2$ 表示 1 年 2 月的时间序列值, 以此类推; 将 Period 作为自变量, 将 $\text{Deseasonalized Values}$ 作为因变量的值, 部分 Minitab 回归输出结果如下:

The regression equation is
Deseasonalized Expense = 228 + 1.96
Period

c.

月	月份预测值
1	213.37
2	235.93
3	252.69
4	297.21
5	314.53
6	403.42
7	486.42
8	386.52
9	309.88
10	314.98
11	297.71
12	254.44

40. a. 时间序列图表示存在季节影响; 用电量在上午 0 ~ 4 点最低, 随后平稳增加, 到下午 12 ~ 4 点达到最高值, 然后再次减少。数据也可能存在线性趋势。

b.

时间段	调整后的季节指数
上午 0 ~ 4 点	0.325 6
上午 4 ~ 8 点	0.447 6
中午 8 ~ 12 点	1.362 2
下午 12 ~ 4 点	1.695 9
下午 4 ~ 8 点	1.457 8
下午 8 ~ 0 点	0.710 9

c. 下面的 Minitab 输出结果是对消除季节影响的时间序列拟合的线性趋势方程:

The regression equation is
Deseasonalized Power = 63108 + 1854 t

消除季节影响后的用电量 ($t = 19$) = $63\,108 + 1\,854 \times 19 = 98\,334$

下午 12 ~ 4 点的预测值 = $1.695\,9 \times 98\,334 = 166\,764.63$
或近似 166 765 千瓦时

消除季节影响后的用电量 ($t = 20$) = $63\,108 + 1\,854 \times 20 = 100\,188$

下午 4 ~ 8 点的预测值 = $1.457\,8 \times 100\,188 = 146\,054.07$
或近似 146 054 千瓦时

因此,从中午 12 点到下午 8 点用电量的预测值为
 $166\,765 + 146\,054 = 312\,819$ 千瓦时。

42. a. 时间序列图显示有水平模式。
b. $MSE(\alpha = 0.2) = 1.40$
 $MSE(\alpha = 0.3) = 1.27$
 $MSE(\alpha = 0.4) = 1.23$
平滑常数 $\alpha = 0.4$ 提供最佳的预测,因为它的 MSE 较小。
c. 31.00
44. a. 数据显示有增长的趋势,2011 年 4 月后是降低成本和增加成本时期。
b. 成本 = $81.29 + 0.58t$ $t = 1$ 对应于 2010 年 1 月, $t = 2$ 对应于 2010 年 2 月。依此类推。2014 年 1 月 ($t = 49$) 预测值 = $81.29 + 0.58 \times 49 = 109.71$ (美元)
c. 成本 = $68.82 + 2.08t - 0.03t^2$ 2014 年 1 月 ($t = 49$) 的预测值 = $68.82 + 2.08 \times 49 - 0.03 \times 49^2 = 98.71$ (美元)
d. 线性趋势方程: $MSE = 69.6$ 。二次趋势方程: $MSE = 41.9$
46. a. 7 月的预测值为 236.97
用 7 月的预测值作为 7 月的实际销售额,则 8 月的预测值是 236.97。
指数平滑法对未来每一期提供相同的预测,这也是对长期预测通常不推荐它的原因。
b. 用 Minitab 回归程序我们得到如下线性趋势方程:
 $T_t = 149.72 + 18.451t$
7 月的预测值为 278.88
8 月的预测值为 297.33
c. 理赔方案不合理,因为它没有考虑销售额的向上趋势;基于趋势推测法,理赔总额应以 7 月和 8 月预测的损失销售额 278 880 美元和 297 330 美元为依据。
48. a. 时间序列图显示有线性趋势。
b. $T_t = -5 + 15t$
斜率 15 表明每年平均增加 15 架钢琴。
c. 85, 100

50. a.

季度	调整后的季节指数
1	1.271 7
2	0.612 0
3	0.497 8
4	1.618 5

注意: 季节指数的调整系数 = $\frac{4}{3.898\,5} = 1.026\,0$

- b. 最大的影响在第四季度;这似乎是合理的,因为零售额通常在 10 月、11 月和 12 月较高。
52. a. 是,显示线性趋势模式
b. 部分 Minitab 回归输出结果如下:

The regression equation is
Number Sold = 22.9 + 15.5 Year

- c. 第 8 年的预测值近似为 147 个单位
54. b. 中心化移动平均数消除了时间序列中的季节影响和一些随机波动;中心化移动平均时间序列显示数据中存在趋势。

c.

季度	调整后的季节指数
1	0.899
2	1.362
3	1.118
4	0.621

- d. Hudson Marine 公司在第二季度经受最大季节性增长。因为这个季度发生在夏季水上运动高峰季之前,结果似乎是合理的,但是最大季节影响是第四季度的季节性减少。这也是合理的,因为水上运动在秋季和冬季减少。

第 18 章

1. 得到 $n = 27$ 个数值不等于 150
正态近似, $\mu = 0.5n = 0.5 \times 27 = 13.5$,
 $\sigma = \sqrt{0.25n} = \sqrt{0.25 \times 27} = 2.598\,1$
由于加号个数 22 在上侧,用连续性校正因子如下
 $P(x \geq 21.5) = P(z \geq \frac{21.5 - 13.5}{2.598}) = P(z \geq 3.08)$
 p -值 = $1.000\,0 - 0.999\,0 = 0.001\,0$
 p -值 ≤ 0.01 ; 拒绝 H_0 , 得到中位数 > 150 的结论。
2. 去掉没有偏好的数据, $n = 9$ 和 $p = 0.5$ 的二项概率为

x	概率
0	0.002 0
1	0.017 6
2	0.070 3
3	0.164 1
4	0.246 1

(续)

x	概率
5	0.246 1
6	0.164 1
7	0.070 3
8	0.017 6
9	0.002 0

加号个数 = 7
 $P(x \geq 7) = P(x = 7) + P(x = 8) + P(x = 9)$
 $= 0.070\ 3 + 0.017\ 6 + 0.002\ 0 = 0.089\ 9$
双侧 p -值 $= 2 \times 0.089\ 9 = 0.179\ 8$
 p -值 > 0.05 ; 不拒绝 H_0 ; 得出没有迹象表明存在差异的结论。

4. a. H_0 : 中位数 ≥ 15
 H_a : 中位数 < 15
b. $n = 9$; 加号个数 = 1
 p -值 $= 0.019\ 6$
拒绝 H_0 ; 债券型基金有较低的中位数。
6. $n = 48$; $z = 1.88$
 p -值 $= 0.030\ 1$
拒绝 H_0 ; 即中位数 $> 56\ 200$ 美元。
8. a. $n = 15$
 p -值 $= 0.118\ 6$
不拒绝 H_0 ; 对节奏不存在显著差异。
b. 25%, 68.8%; 建议大样本
10. $n = 600$, $z = 2.41$
 p -值 $= 0.016\ 0$
拒绝 H_0 ; 更偏好嘻哈帝国。
12. H_0 : 添加剂 1 的中位数 - 添加剂 2 的中位数 $= 0$
 H_a : 添加剂 1 的中位数 - 添加剂 2 的中位数 $\neq 0$

差	差的绝对值	秩	符号秩	
			负	正
2.07	2.07	9		9
1.79	1.79	7		7
-0.54	0.54	3	-3	
2.09	2.09	10		10
0.01	0.01	1		1
0.97	0.97	4		4
-1.04	1.04	5	-5	
3.75	3.75	12		12
1.84	1.84	8		8
3.08	3.08	11		11
0.43	0.43	2		2
1.32	1.32	6		6
正的符号秩之和			$T^+ = 70$	

$$\mu_{T^+} = \frac{n(n+1)}{4} = \frac{12 \times 13}{4} = 39$$
$$\sigma_{T^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{12 \times 13 \times 25}{24}} = 12.747\ 5$$
$$P(T^+ \geq 70) = P(z \geq \frac{69.5 - 39}{12.747\ 5}) = P(z \geq 2.39)$$
$$p\text{-值} = 2 \times (1.000\ 0 - 0.991\ 6) = 0.016\ 8$$
$$p\text{-值} \leq 0.05, \text{ 拒绝 } H_0; \text{ 添加剂之间存在显著差异。}$$

13. H_0 : 无弛缓药物的中位数时间 - 有弛缓药物的中位数时间 ≤ 0
 H_a : 无弛缓药物的中位数时间 - 有弛缓药物的中位数时间 > 0

差	差的绝对值	秩	符号秩	
			负	正
5	5	9		9
2	2	3		3
10	10	10		10
-3	3	6.5	-6.5	
1	1	1		1
2	2	3		3
-2	2	3	-3	
3	3	6.5		6.5
3	3	6.5		6.5
3	3	6.5		6.5
正的符号秩之和			$T^+ = 45.5$	

$$\mu_{T^+} = \frac{n(n+1)}{4} = \frac{10 \times 11}{4} = 27.5$$
$$\sigma_{T^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{10 \times 11 \times 21}{24}} = 9.810\ 7$$
$$P(T^+ \geq 45.5) = P(z \geq \frac{45 - 27.5}{9.810\ 7}) = P(z \geq 1.78)$$
$$p\text{-值} = 1.000\ 0 - 0.992\ 5 = 0.037\ 5$$
$$p\text{-值} \leq 0.05, \text{ 拒绝 } H_0; \text{ 得出无弛缓药物有较大的中位数时间的结论。}$$

14. $n = 11$; $T^+ = 61$; $z = 2.45$
 p -值 $= 0.014\ 2$
拒绝 H_0 ; 存在显著差异; 2006 年正点率较好

16. $n = 12$; $T^+ = 49.5$; $z = 0.788$
 p -值 $= 0.435\ 4$
不拒绝 H_0 ; 得出中位数分数之间不存在差异的结论。

18. H_0 : 两个添加剂总体相同
 H_a : 两个添加剂总体不相同

添加剂 1	秩	添加剂 2	秩
17.3	2	18.7	8.5
18.4	6	17.8	4

(续)

添加剂 1	秩	添加剂 2	秩
19.1	10	21.3	15
16.7	1	21.0	14
18.2	5	22.1	16
18.6	7	18.7	8.5
17.5	3	19.8	11
		20.7	13
		20.2	12
W = 34			

$$\mu_W = \frac{1}{2}n_1(n_1 + n_2 + 1) = \frac{1}{2} \times 7 \times (7 + 9 + 1) = 59.5$$

$$\sigma_T = \sqrt{\frac{1}{12}n_1n_2(n_1 + n_2 + 1)} = \sqrt{\frac{1}{12} \times 7 \times 9 \times (7 + 9 + 1)} = 9.4472$$

由于 $W = 34$ 在下侧, 用连续性校正

$$P(W \leq 34) = P\left(z \leq \frac{34.5 - 59.5}{9.4472}\right) = P(z \leq -2.65)$$

$$p\text{-值} = 2 \times 0.0040 = 0.0080$$

$p\text{-值} < 0.05$, 拒绝 H_0 ; 得出添加剂不相同的结论。

添加剂 2 的每加仑行驶里程数较多。

19. a. H_0 : 两个年薪总体相同

H_a : 两个年薪总体不相同

注册会计师	秩	财务分析师	秩
25.2	5	24.0	2
33.8	19	24.2	3
31.3	16	28.1	10
33.2	18	30.9	15
29.2	13	26.9	8.5
30.0	14	28.6	11
25.9	6	24.7	4
34.5	20	28.9	12
31.7	17	26.8	7
26.9	8.5	23.9	1
W = 136.5			

$$\mu_T = \frac{1}{2}n_1(n_1 + n_2 + 1) = \frac{1}{2} \times 10 \times (10 + 10 + 1) = 105$$

$$\sigma_T = \sqrt{\frac{1}{12}n_1n_2(n_1 + n_2 + 1)} = \sqrt{\frac{1}{12} \times 10 \times 10 \times (10 + 10 + 1)} = 13.2288$$

由于 $W = 136.5$ 在上侧, 用连续性校正

$$P(W \geq 136.5) = P\left(z \geq \frac{136 - 105}{13.2288}\right) = P(z \geq 2.34)$$

$$p\text{-值} = 2 \times (1.0000 - 0.9904) = 0.0192$$

$p\text{-值} < 0.05$, 拒绝 H_0 ; 得出总体不相同的结论, 注册会计师的年薪较高。

$$\text{b. 注册会计师: } \frac{55.0 + 56.3}{2} = 55.65 \text{ (1 000 美元)}$$

$$\text{财务分析师: } \frac{51.8 + 51.9}{2} = 51.85 \text{ (1 000 美元)}$$

20. a. 54 900 美元, 40 400 美元

$$\text{b. } W = 69; z = 2.04$$

$$p\text{-值} = 0.0414$$

拒绝 H_0 ; 得出收入之间存在差异的结论; 男性较高。

22. $W = 157; z = 2.74$

$$p\text{-值} = 0.0062$$

拒绝 H_0 ; 得出市盈率之间存在差异的结论; 日本较高。

24. $W = 116; z = -0.22$

$$p\text{-值} = 0.8258$$

不拒绝 H_0 ; 结论是没有价格不同的证据。

26. H_0 : 三种产品评分的总体相同

H_a : 三种产品评分的总体不全相同

A	B	C
4	11	7
8	14	2
10	15	1
3	12	6
9	13	5
秩和 34	65	21

$$H = \frac{12}{15 \times 16} \left[\frac{34^2}{5} + \frac{65^2}{5} + \frac{21^2}{5} \right] - 3 \times 16 = 10.22$$

利用 $df = 2$ 的 χ^2 表, $\chi^2 = 10.22$; $p\text{-值}$ 介于 0.005 和 0.01 之间

$p\text{-值} \leq 0.01$, 拒绝 H_0 ; 得出评分总体不同的结论。

28. H_0 : 所有消耗卡路里的总体相同

H_a : 消耗卡路里的总体不全相同

游泳	网球	自行车
8	9	5
4	14	1
11	13	3
6	10	7
12	15	2
秩和 41	61	18

$$H = \frac{12}{15 \times (15 + 1)} \left[\frac{41^2}{5} + \frac{61^2}{5} + \frac{18^2}{5} \right] - 3 \times 16 = 9.26$$

利用 $df = 2$ 的 χ^2 表, $\chi^2 = 9.26$; $p\text{-值}$ 介于 0.005 和 0.01 之间
 $p\text{-值} \leq 0.05$, 拒绝 H_0 ; 得出消耗卡路里总体不同的结论。

30. $H = 8.03; df = 3$

$p\text{-值}$ 介于 0.025 和 0.05 之间

拒绝 H_0 ；得出课程的质量之间有差异的结论。

32. a. $\sum d_i^2 = 52$

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} = 1 - \frac{6 \times 52}{10 \times 99} = 0.685$$

b. $\sigma_{r_s} = \sqrt{\frac{1}{n-1}} = \sqrt{\frac{1}{9}} = 0.333\ 3$

$$z = \frac{r_s - 0}{\sigma_{r_s}} = \frac{0.685}{0.333\ 3} = 2.05$$

$$p\text{-值} = 2 \times (0.000\ 0 - 0.979\ 8) = 0.040\ 4$$

$p\text{-值} \leq 0.05$ ，拒绝 H_0 ；得出显著的正秩相关的结论。

34. $\sum d_i^2 = 250$

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} = 1 - \frac{6 \times 250}{11 \times 120} = -0.136$$

$$\sigma_{r_s} = \sqrt{\frac{1}{n-1}} = \sqrt{\frac{1}{10}} = 0.316\ 2$$

$$z = \frac{r_s - 0}{\sigma_{r_s}} = \frac{-0.136}{0.316\ 2} = -0.43$$

$$p\text{-值} = 2 \times 0.333\ 6 = 0.667\ 2$$

$p\text{-值} > 0.05$ ，不拒绝 H_0 ；我们不能得到存在显著相关的结论。

36. $r_s = -0.709$ ， $z = -2.13$

$$p\text{-值} = 0.033\ 2$$

拒绝 H_0 ；得出显著的秩相关结论。

38. 加号个数 = 905， $z = -3.15$

$$p\text{-值} < 0.002\ 0$$

拒绝 H_0 ；得出偏好之间存在显著差异的结论。

40. $n = 12$ ； $T^+ = 6$ ； $z = -2.55$

$$p\text{-值} = 0.010\ 8$$

拒绝 H_0 ；得出价格之间存在显著差异的结论。

42. $W = 70$ ； $z = -2.93$

$$p\text{-值} = 0.003\ 4$$

拒绝 H_0 ；得出重量总体不同的结论。

44. $H = 12.61$ ； $df = 2$

$$p\text{-值} < 0.005$$

拒绝 H_0 ；得出评分总体不同的结论。

46. $r_s = 0.757$ ， $z = 2.83$

$$p\text{-值} = 0.004\ 6$$

拒绝 H_0 ；得出显著的正秩相关的结论。

第 19 章

2. a. 5.42

b. UCL = 6.09，LCL = 4.75

4. R 控制图：

$$UCL = \bar{RD}_4 = 1.6 \times 1.864 = 2.98$$

$$LCL = \bar{Rd}_3 = 1.6 \times 0.136 = 0.22$$

\bar{x} 控制图：

$$UCL = \bar{\bar{x}} + A_2 \bar{R} = 28.5 + 0.373 \times 1.6 = 29.10$$

$$LCL = \bar{\bar{x}} - A_2 \bar{R} = 28.5 - 0.373 \times 1.6 = 27.90$$

6. 20.01，0.082

8. a. 0.047 0

b. UCL = 0.098 9，LCL = -0.004 9（用 LCL = 0）

c. $\bar{p} = 0.08$ ；处于在控状态

d. UCL = 14.826，LCL = -0.726（用 LCL = 0）

如果有缺陷的数量超过 14，则过程处于失控状态。

e. 因为 12 个有缺陷，所以处于在控状态。

f. np 控制图

10. $f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

当 $p = 0.02$ 时，接收一批的概率为

$$f(0) = \frac{25!}{0!(25-0)!} 0.02^0 (1-0.02)^{25} = 0.603\ 5$$

当 $p = 0.06$ 时，接收一批的概率为

$$f(0) = \frac{25!}{0!(25-0)!} 0.06^0 (1-0.06)^{25} = 0.212\ 9$$

12. $p_0 = 0.02$ ；生产者风险 = 0.059 9

$p_0 = 0.06$ ；消费者风险 = 0.339 6

当接收数量 c 增加时，生产者风险将减少。

14. $n = 20$ ， $c = 3$

16. a. 95.4

b. UCL = 96.07，LCL = 94.73

c. 否

18.

	R 控制图	\bar{x} 控制图
UCL	4.23	6.57
LCL	0	4.27

标准差的估计值 = 0.86

20.

	R 控制图	\bar{x} 控制图
UCL	0.112 1	3.112
LCL	0	3.051

22. a. UCL = 0.081 7，LCL = -0.001 7（用 LCL = 0）

24. a. 0.03

b. $\beta = 0.080\ 2$

第 20 章

1. a.

商品项目	价比
A	$103 = 7.75/7.50 \times 100$
B	$238 = 1\ 500/630 \times 100$

$$b. I_{2011} = \frac{7.75 + 1\,500.00}{7.50 + 630.00} \times 100 = \frac{1\,507.75}{637.50} \times 100 = 237$$

$$c. I_{2011} = \frac{7.75 \times 1\,500 + 1\,500.00 \times 2}{7.50 \times 1\,500 + 630.00 \times 2} \times 100$$

$$= \frac{14\,625.00}{12\,510.00} \times 100 = 117$$

$$d. I_{2011} = \frac{7.75 \times 1\,800 + 1\,500.00 \times 1}{7.50 \times 1\,800 + 630.00 \times 1} \times 100$$

$$= \frac{15\,450.00}{14\,130.00} \times 100 = 109$$

2. a. 32%

b. 8.14 美元

3. a. A 的价比 = $6.00/5.54 \times 100 = 110$

B 的价比 = $5.95/5.60 \times 100 = 106$

C 的价比 = $6.20/5.50 \times 100 = 113$

$$b. I_{2014} = \frac{6.00 + 5.95 + 6.20}{5.45 + 5.60 + 5.50} \times 100 = 110$$

$$c. I_{2014} = \frac{6.00 \times 150 + 5.95 \times 20 + 6.20 \times 120}{5.45 \times 150 + 5.60 \times 200 + 5.50 \times 120} \times 100 = 109$$

2 年间增长 9%

$$4. I_{2014} = 114$$

6.

项目	价比	基期		权数	加权价比
		价格	使用量		
A	150	22.00	20	440	66 000
B	90	5.00	50	250	22 500
C	120	14.00	40	560	67 200
合计				1 250	155 700

$$I = \frac{155\,700}{1\,250} = 125$$

7. a. A 的价比 = $3.95/2.50 \times 100 = 158$

B 的价比 = $9.90/8.75 \times 100 = 113$

C 的价比 = $0.95/0.99 \times 100 = 96$

b.

项目	价比	基期价格	数量	权数 $P_{i0}Q_i$	加权价比
A	158	2.50	22	62.5	9 875
B	113	8.75	15	131.3	14 837
C	196	0.99	60	59.4	5 702
合计				253.2	30 414

$$I = \frac{30\,414}{253.2} = 120$$

化学用品原材料的成本上涨了 20%

8. $I = 105$, 证券投资组合上涨 5%

10. a. 减缩 2007 年工资:

$$\frac{30.04}{207.3} \times 100 = 14.49 \text{ 美元}$$

减缩 2011 年工资:

$$\frac{33.23}{224.9} \times 100 = 14.78 \text{ 美元}$$

$$b. \frac{33.23}{30.04} \times 100 = 1.11, \text{ 现实工资增长 } 11\%$$

$$c. \frac{14.78}{14.49} \times 100 = 1.02, \text{ 实际工资增长 } 2\%$$

$$12. a. \text{第一年: } \frac{29.1}{216.0} \times 100 = 13.47 \text{ 美元}$$

$$\text{第二年: } \frac{33.3}{218.4} \times 100 = 15.25 \text{ 美元}$$

$$\text{第三年: } \frac{32.9}{226.9} \times 100 = 14.50 \text{ 美元}$$

$$b. \text{第一年: } \frac{29.1}{173.4} \times 100 = 16.78 \text{ 美元}$$

$$\text{第二年: } \frac{33.3}{180.2} \times 100 = 18.47 \text{ 美元}$$

$$\text{第三年: } \frac{32.9}{192.5} \times 100 = 17.09 \text{ 美元}$$

c. 生产者价格指数比消费者价格指数更适合, 因为这些数值反映零售商 (而非消费者) 的价格。

$$14. I = \frac{300 \times 18.00 + 400 \times 4.90 + 850 \times 15.00}{350 \times 18.00 + 220 \times 4.90 + 730 \times 15.00} \times 100$$

$$= \frac{20\,110}{18\,328} \times 100 = 110$$

$$15. I = \frac{95 \times 1\,200 + 75 \times 1\,800 + 50 \times 2\,000 + 70 \times 1\,500}{120 \times 1\,200 + 86 \times 1\,800 + 35 \times 2\,000 + 60 \times 1\,500} \times 100$$

$$= 99$$

数量有所下降

$$16. I = 83$$

$$18. a. 151, 197, 143, 178$$

$$b. I = 170$$

$$20. I_{\text{Jan}} = 73.5, I_{\text{Mar}} = 70.1$$

22.

项目	价比 (P_t/P_{i0})	基期 价格 P_{i0}	数量 Q_i	权重 $W_i = P_{i0}Q_i$	加权价比 (P_t/P_{i0}) $100W_i$
剃须柄	126.9	7.46	1	7.46	946.47
刀片	153.7	1.9	17	32.30	4 964.51
合计				39.76	5 911.18

$$I_{\text{第11年}} = (5\,911.18/39.76) = 148.7$$

$$24. \text{第一年: } (84\,810/215.3) \times 100 = 39\,392 \text{ 美元}$$

$$\text{第二年: } (87\,210/214.5) \times 100 = 40\,657 \text{ 美元}$$

$$\text{第三年: } (87\,650/218.1) \times 100 = 40\,188 \text{ 美元}$$

$$\text{第四年: } (91\,060/224.9) \times 100 = 40\,489 \text{ 美元}$$

$$(40\,489/39\,392) = 1.028$$

这期间年薪中位数增加了 2.8%

$$26. I = 143, \text{ 数量上涨 } 43\%$$