

第四章 统计数据的概括性度量

4. 1 一家汽车零售店的 10 名销售人员 5 月份销售的汽车数量(单位: 台)排序后如下:

2 4 7 10 10 10 12 12 14 15

要求:

(1) 计算汽车销售量的众数、中位数和平均数。

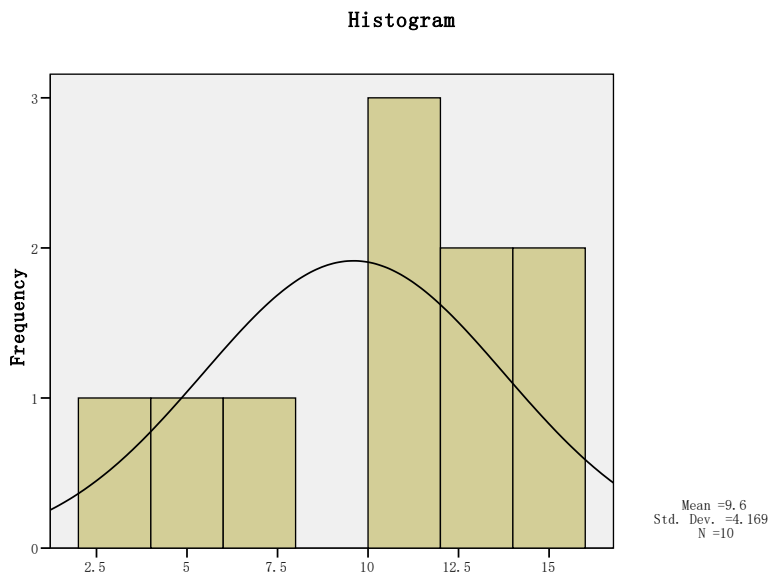
(2)根据定义公式计算四分位数。

(3)计算销售量的标准差。

(4)说明汽车销售量分布的特征。

解:

Statistics		
汽车销售数量		
N	Valid	10
	Missing	0
Mean		
Median		
Mode		10
Std. Deviation		
Percentiles	25	
	50	
	75	



4. 2 随机抽取 25 个网络用户, 得到他们的年龄数据如下:

单位: 周岁

19	15	29	25	24
23	21	38	22	18
30	20	19	19	16
23	27	22	34	24
41	20	31	17	23

要求:

(1)计算众数、中位数:

排序形成单变量分值的频数分布和累计频数分布:

网络用户的年龄

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
--	-----------	---------	----------------------	--------------------

Valid	15	1		1	
	16	1		2	
	17	1		3	
	18	1		4	
	19	3		7	
	20	2		9	
	21	1		10	
	22	2		12	
	23	3		15	
	24	2		17	
	25	1		18	
	27	1		19	
	29	1		20	
	30	1		21	
	31	1		22	
	34	1		23	
	38	1		24	
	41	1		25	
	Total	25			

从频数看出，众数 M_o 有两个：19、23；从累计频数看，中位数 $Me=23$ 。

(2)根据定义公式计算四分位数。

Q_1 位置 $=25/4=6.25$ ，因此 $Q_1=19$ ， Q_3 位置 $=3 \times \times 2=26.5$ 。

(3)计算平均数和标准差：

Mean=； Std. Deviation=

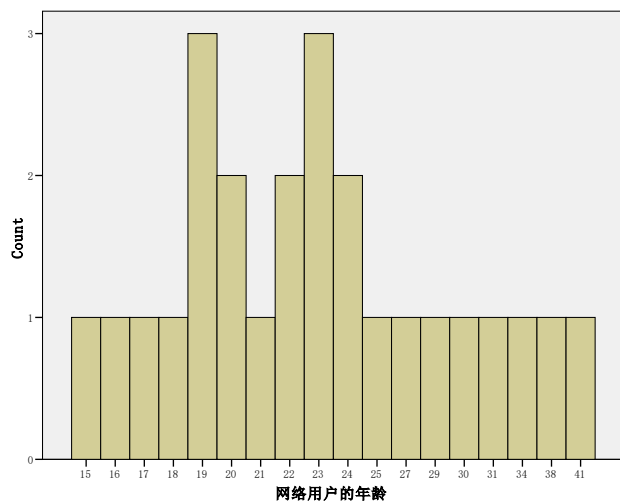
(4)计算偏态系数和峰态系数：

Skewness=； Kurtosis=

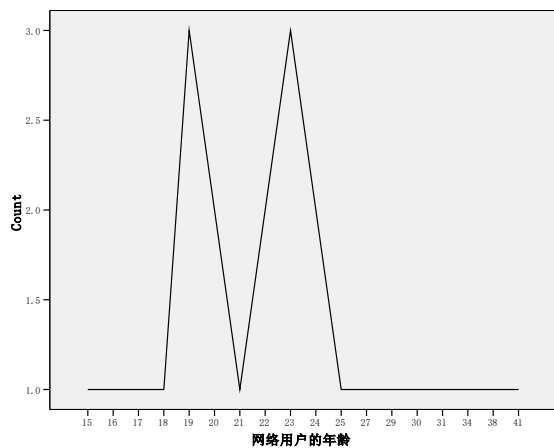
(5)对网民年龄的分布特征进行综合分析：

分布，均值 $=24$ 、标准差 $=6.652$ 、呈右偏分布。如需看清楚分布形态，需要进行分组。

为分组情况下的直方图：



为分组情况下的概率密度曲线：



分组：

1、确定组数：

$$K = 1 + \frac{\lg(n)}{\lg(2)} = 1 + \frac{\lg(25)}{\lg(2)} = 1 + \frac{1.398}{0.30103} = 5.64, \text{ 取 } k=6$$

2、确定组距：组距= (最大值 - 最小值) ÷ 组数= (41-15) ÷ 6=4.3，取 5

3、分组频数表

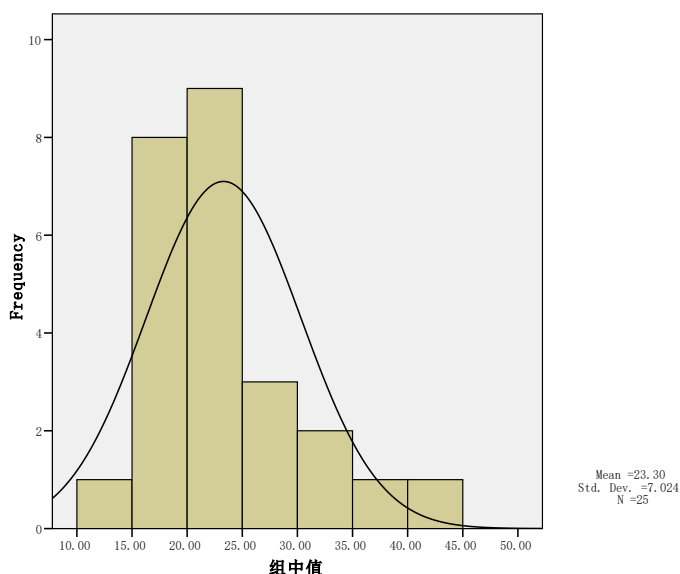
网络用户的年龄 (Binned)

		Frequency	Percent	Cumulative Frequency	Cumulative Percent
Valid	<= 15	1		1	
	16 - 20	8		9	
	21 - 25	9		18	
	26 - 30	3		21	
	31 - 35	2		23	
	36 - 40	1		24	
	41+	1		25	
	Total	25			

分组后的均值与方差：

Mean	
Std. Deviation	
Variance	
Skewness	
Kurtosis	

分组后的直方图：



4. 3 某银行为缩短顾客到银行办理业务等待的时间。准备采用两种排队方式进行试验：一种是所有顾客都进入一个等待队列；另一种是顾客在三千业务窗口处列队 3 排等待。为比拟哪种排队方式使顾客等待的时间更短。两种排队方式各随机抽取 9 名顾客。得到第一种排队方式的平均等待时间为 7. 2 分钟，标准差为 1. 97 分钟。第二种排队方式的等待时间(单位：分钟)如下：

5. 5 6. 6 6. 7 6. 8 7. 1 7. 3 7. 4 7. 8 7. 8

要求：

(1)画出第二种排队方式等待时间的茎叶图。

第二种排队方式的等待时间(单位：分钟) Stem-and-Leaf Plot

Frequency	Stem & Leaf
1.00	Extremes (<=5.5)
3.00	6 . 678
3.00	7 . 134
2.00	7 . 88

Each leaf: 1 case(s)

(2)计算第二种排队时间的平均数和标准差。

Mean	7
Std. Deviation	
Variance	

(3)比拟两种排队方式等待时间的离散程度。

第二种排队方式的离散程度小。

(4)如果让你选择一种排队方式，你会选择哪一种？试说明理由。

选择第二种，均值小，离散程度小。

4. 4 某百货公司 6 月份各天的销售额数据如下：

单位：万元

257	276	297	252	238	310	240	236	265	278
271	292	261	281	301	274	267	280	291	258
272	284	268	303	273	263	322	249	269	295

要求：

(1)计算该百货公司日销售额的平均数和中位数。

(2)按定义公式计算四分位数。

(3)计算日销售额的标准差。

解：

Statistics

百货公司每天的销售额（万元）

N	Valid	30
	Missing	0
Mean		
Median		
Std. Deviation		
Percentiles	25	
	50	
	75	

4. 5 甲乙两个企业生产三种产品的单位本钱和总本钱资料如下：

产品 名称	单位本钱 (元)	总本钱(元)	
		甲企业	乙企业
A	15	2 100	3 255
B	20	3 000	1 500
C	30	1 500	1 500

要求：比拟两个企业的总平均本钱，哪个高，并分析其原因。

产品名称	单位本钱(元)	甲企业		乙企业	
		总本钱(元)	产品数	总本钱(元)	产品数
A	15	2100	140	3255	217
B	20	3000	150	1500	75
C	30	1500	50	1500	50
平均本钱 (元)					

调和平均数计算，得到甲的平均本钱为 19.41；乙的平均本钱为 18.29。甲的中间本钱的产品多，乙的低本钱的产品多。

4. 6 在某地区抽取 120 家企业，按利润额进行分组，结果如下：

按利润额分组(万元)	企业数(个)
200~300	19
300~400	30
400~500	42
500~600	18
600 以上	11
合 计	120

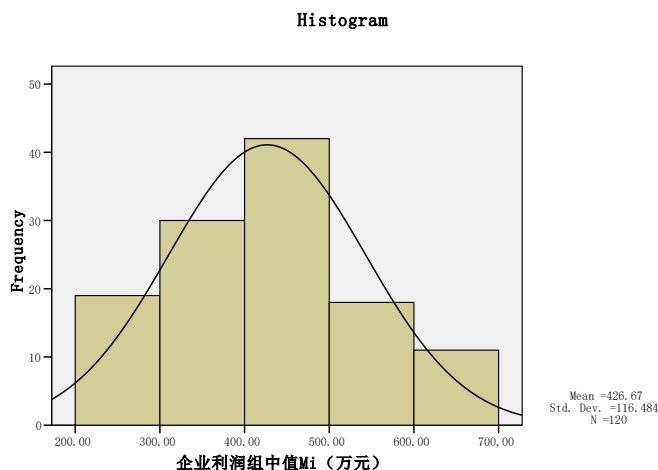
要求：

(1)计算 120 家企业利润额的平均数和标准差。

(2)计算分布的偏态系数和峰态系数。

解：

Statistics		
企业利润组中值 Mi (万元)		
N	Valid	120
	Missing	0
Mean		
Std. Deviation		
Skewness		
Std. Error of Skewness		
Kurtosis		
Std. Error of Kurtosis		



4. 7 为研究少年儿童的生长发育状况，某研究所的一位调查人员在某城市抽取 100 名 7~17 岁的少年儿童作为样本，另一位调查人员那么抽取了 1 000 名 7~17 岁的少年儿童作为样本。请答复下面的问题，并解释其原因。

- (1)两位调查人员所得到的样本的平均身高是否相同?如果不同，哪组样本的平均身高较大?
- (2)两位调查人员所得到的样本的标准差是否相同?如果不同，哪组样本的标准差较大?
- (3)两位调查人员得到这 1 100 名少年儿童身高的最高者或最低者的时机是否相同?如果不同，哪位调查研究人员的时机较大?

解：(1) 不一定相同，无法判断哪一个更高，但可以判断，样本量大的更接近于总体平均身高。

(2) 不一定相同，样本量少的标准差大的可能性大。

(3) 时机不相同，样本量大的得到最高者和最低者的身高的时机大。

4. 8 一项关于大学生体重状况的研究发现，男生的平均体重为 60kg，标准差为 5kg；女生的平均体重为 50kg，标准差为 5kg。请答复下面的问题：

- (1)是男生的体重差异大还是女生的体重差异大?为什么?

女生，因为标准差一样，而均值男生大，所以，离散系数是男生的小，离散程度是男生的小。

- (2)以磅为单位(1ks=2. 2lb)，求体重的平均数和标准差。

都是各乘以 2.21，男生的平均体重为 $60\text{kg} \times 2.21=$ ，标准差为 $5\text{kg} \times 2.21=$ ；女生的平均体重为 $50\text{kg} \times 2.21=$ ，标准差为 $5\text{kg} \times 2.21=$ 。

- (3)粗略地估计一下，男生中有百分之几的人体重在 55kg — 65kg 之间?

计算标准分数：

$$Z_1 = \frac{x - \bar{x}}{s} = \frac{55 - 60}{5} = -1; Z_2 = \frac{x - \bar{x}}{s} = \frac{65 - 60}{5} = 1, \text{ 根据经验规那么, 男生大约有 } 68\% \text{ 的人体重在 } 55\text{kg}$$

— 65kg 之间。

- (4)粗略地估计一下，女生中有百分之几的人体重在 40kg~60kg 之间?

计算标准分数：

$$Z_1 = \frac{x - \bar{x}}{s} = \frac{40 - 50}{5} = -2; Z_2 = \frac{x - \bar{x}}{s} = \frac{60 - 50}{5} = 2, \text{ 根据经验规那么, 女生大约有 } 95\% \text{ 的人体重在 } 40\text{kg}$$

— 60kg 之间。

4. 9 一家公司在招收职员时，首先要通过两项能力测试。在 A 项测试中，其平均分数是 100 分，标准差是 15 分；在 B 项测试中，其平均分数是 400 分，标准差是 50 分。一位应试者在 A 项测试中得了 115 分，在 B 项测试中得了 425 分。与平均分数相比，该应试者哪一项测试更为理想?

解：应用标准分数来考虑问题，该应试者标准分数高的测试理想。

$$Z_A = \frac{x - \bar{x}}{s} = \frac{115 - 100}{15} = 1; Z_B = \frac{x - \bar{x}}{s} = \frac{425 - 400}{50}$$

因此，A 项测试结果理想。

4. 10 一条产品生产线平均每天的产量为 3 700 件，标准差为 50 件。如果某一天的产量低于或高于平均产量，并落人士 2 个标准差的范围之外，就认为该生产线“失去控制”。下面是一周各天的产量，该生产线哪几天失去了控制？

时间	周一	周二	周三	周四	周五	周六	周日
产量(件)	3 850	3 670	3 690	3 720	3 610	3 590	3 700

时间	周一	周二	周三	周四	周五	周六	周日
产量(件)	3850	3670	3690	3720	3610	3590	3700
日平均产量	3700						
日产量标准差	50						
标准分数 Z	3						0
标准分数界限	-2	-2	-2	-2	-2	-2	-2
	2	2	2	2	2	2	2

周六超出界限，失去控制。

4. 11 对 10 名成年人和 10 名幼儿的身高进行抽样调查，结果如下：

成年组	166	169	172	177	180	170	172	174	168	173
幼儿组	68	69	68	70	71	73	72	73	74	75

要求：

(1)如果比拟成年组和幼儿组的身高差异，你会采用什么样的统计量?为什么？

均值不相等，用离散系数衡量身高差异。

(2)比拟分析哪一组的身高差异大？

成年组		幼儿组	
平均		平均	
标准差		标准差	
离散系数		离散系数	

幼儿组的身高差异大。

4. 12 一种产品需要人工组装，现有三种可供选择的组装方法。为检验哪种方法更好，随机抽取 15 个工人，让他们分别用三种方法组装。下面是 15 个工人分别用三种方法在相同的时间内组装的产品数量：
单位：个

方法 A	方法 B	方法 C
164	129	125
167	130	126
168	129	126
165	130	127
170	131	126
165	]30	128
164	129	127
168	127	126
164	128	127
162	128	127
163	127	125
166	128	126
167	128	116
166	125	126
165	132	125

要求:

(1)你准备采用什么方法来评价组装方法的优劣?

均值不相等,用离散系数衡量身高差异。

(2)如果让你选择一种方法,你会作出怎样的选择?试说明理由。

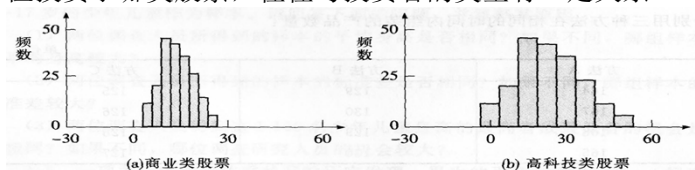
解: 比照均值和离散系数的方法, 选择均值大, 离散程度小的。

方法 A	方法 B	方法 C
平均	平均	平均
标准	标准	标准
差	差	差

离散系数: $V_A =$, $V_B =$, $V_C =$

均值 A 方法最大, 同时 A 的离散系数也最小, 因此选择 A 方法。

4. 13 在金融证券领域, 一项投资的预期收益率的变化通常用该项投资的风险来衡量。预期收益率的变化越小, 投资风险越低; 预期收益率的变化越大, 投资风险就越高。下面的两个直方图, 分别反映了 200 种商业类股票和 200 种高科技类股票的收益率分布。在股票市场上, 高收益率往往伴随着高风险。但投资于哪类股票, 往往与投资者的类型有一定关系。



(1)你认为该用什么样的统计量来反映投资的风险?

标准差或者离散系数。

(2)如果选择风险小的股票进行投资, 应该选择商业类股票还是高科技类股票?

选择离散系数小的股票, 那么选择商业股票。

(3)如果进行股票投资, 你会选择商业类股票还是高科技类股票?

考虑高收益, 那么选择高科技股票; 考虑风险, 那么选择商业股票。

第五章 概率与概率分布

5.1 略

5.2 $P(AB) = P(A) + P(B) - P(A+B) = 50\% + 60\% - 85\% = 35\%$

5.3 因为 $P(AB) + P(\bar{A}\bar{B}) + P(\bar{A}B) = 1/3$; $P(B) = P(A(B+\bar{B})) = P(AB) + P(\bar{A}B) = 1/3$

$$P(A) = P(A(B+\bar{B})) = P(AB) + P(\bar{A}B) = 1/3 - 1/9 = 2/9$$

$$Q P(AB) + P(\bar{A}\bar{B}) + P(\bar{A}B) + P(\bar{A}\bar{B}) = 1;$$

$$Q P(A|B) = P(AB) / P(B) = 1/6;$$

$$\therefore P(AB) = 1/6 * 1/3 = 1/18$$

$$P(A) = P(A(B+\bar{B})) = P(AB) + P(\bar{A}B); P(\bar{A}B) = 1/3 - 1/18 = 5/18$$

$$\text{同理 } P(B) = P(B(A+\bar{A})) = P(AB) + P(\bar{A}B); P(\bar{A}B) = 5/18$$

$$Q P(\bar{A}|\bar{B}) = P(\bar{A}\bar{B}) / P(\bar{B}) = \frac{1 - 1/18 - 5/18 - 5/18}{1 - 1/3} = 7/12$$

5.5 (1) $P(A)P(B) = 0.8 * 0.7 = 0.56;$ (2)

$$P(A+B) = P(A) + P(B) - P(AB) = 0.8 + 0.7 - 0.8 * 0.7 = 0.94$$

$$(3) P(A+B) = P(A) + P(B) - 2P(AB) = 0.8 + 0.7 - 2 * 0.8 * 0.7 = 0.38$$

5.6 $P(B) = P(A) P(B|A) = 96\% * 75\% = 0.72$

$$5.7 P(A|B) = P(AB) / P(B) = \frac{1/2}{3/4} = 2/3$$

5.8 贝叶斯公式:

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum P(A)P(B|A)} = \frac{10\% * 20\%}{10\% * 20\% + 50\% * 50\% + 40\% * 70\%} = 3.63\%$$

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum P(A)P(B|A)} = \frac{50\% * 50\%}{10\% * 20\% + 50\% * 50\% + 40\% * 70\%} = 45.45\%$$

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum P(A)P(B|A)} = \frac{40\% * 70\%}{10\% * 20\% + 50\% * 50\% + 40\% * 70\%} = 50.9\%$$

5.9 贝叶斯公式:

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum P(A)P(B|A)} = \frac{30\% * 0.1}{30\% * 0.1 + 27\% * 0.05 + 25\% * 0.2 + 18\% * 0.15} = 0.249$$

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum P(A)P(B|A)} = \frac{27\% * 0.05}{30\% * 0.1 + 27\% * 0.05 + 25\% * 0.2 + 18\% * 0.15} = 0.112$$

$$5.12 \quad (1) \int_1^{\theta} \frac{3x^2}{\theta^3} dx = \frac{7}{8}, \therefore \theta = 2 \quad (2) \quad Ex = \int_1^2 \frac{3x^3}{8} dx = 1.5; \quad Dx = \int_1^2 \frac{3x^4}{8} dx = 0.15$$

5.13 $x: B(5, 0.25)$, 学生凭猜想至少答对 4 道的概率为:

$$P(x=4) + P(x=5) = C_5^4 0.25^4 0.75^1 + C_5^5 0.25^5 0.75^0 = \frac{1}{64}$$

5.14 $P(x=k) = \lambda^k \times e^{-\lambda} / k!$ ①

$$P(x=k+1) = \lambda^{k+1} \times e^{-\lambda} / (k+1)! \quad ②$$

$$②/① \text{ 得 } P(x=k+1) / P(x=k) = \lambda / (k+1)$$

令 $P(x=k+1) / P(x=k) > 1$, 那么 $\lambda > k+1$, $k < \lambda - 1$

令 $P(x=k+1) / P(x=k) < 1$, 那么 $\lambda < k+1$, $k > \lambda - 1$

假设 $\lambda < 2$, 那么 $P(x=k)$ 随着 k 增大而减小, $\therefore k=1$ 时最大

假设 $\lambda > 2$, 那么 $P(x=1) < \dots < P(x=[\lambda-1]) < P(x=[\lambda-1]+1) > P(x=[\lambda-1]+2) > \dots$,

$\therefore k=[\lambda-1]+1=[\lambda]$ 是最大

综上, $\lambda < 2$ 时, $k=1$; $\lambda > 2$ 时, $k=[\lambda]$ (写成分段的形式, $[]$ 是取整符号)

第六章 统计量及其抽样分布

6.1 调节一个装瓶机使其对每个瓶子的灌装量均值为 μ 盎司, 通过观察这台装瓶机对每个瓶子的灌装量服从标准差 $\sigma = 1.0$ 盎司的正态分布。随机抽取由这台机器灌装的 9 个瓶子形成一个样本, 并测定每个瓶子的灌装量。试确定样本均值偏离总体均值不超过 0.3 盎司的概率。

解: 总体方差知道的情况下, 均值的抽样分布服从 $N\left(\mu, \sigma^2/n\right)$ 的正态分布, 由正态分布, 标准化得到

标准正态分布: $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$, 因此, 样本均值不超过总体均值的概率 P 为:

$$P(|\bar{x} - \mu| \leq 0.3) = P\left(\frac{|\bar{x} - \mu|}{\sigma/\sqrt{n}} \leq \frac{0.3}{\sigma/\sqrt{n}}\right) = P\left(\frac{-0.3}{1/\sqrt{9}} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.3}{1/\sqrt{9}}\right)$$

$$= P(-0.9 \leq z \leq 0.9) = 2\phi(0.9) - 1, \text{ 查标准正态分布表得 } \phi(0.9)$$

因此, $P(|\bar{x} - \mu| \leq 0.3)$

$$\begin{aligned} 6.2 \quad P(|\bar{Y} - \mu| \leq 0.3) &= P\left(\frac{|\bar{Y} - \mu|}{\sigma/\sqrt{n}} \leq \frac{0.3}{\sigma/\sqrt{n}}\right) = P\left(\frac{-0.3}{1/\sqrt{n}} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.3}{1/\sqrt{n}}\right) \\ &= P(|z| \leq 0.3\sqrt{n}) = 2\phi(0.3\sqrt{n}) - 1 = 0.95 \end{aligned}$$

查表得: $0.3\sqrt{n} = 1.96$ 因此 $n=43$

6.3 $Z_1, Z_2, \dots, Z_6$ 表示从标准正态总体中随机抽取的容量, $n=6$ 的一个样本, 试确定常数 b , 使

$$\text{得 } P\left(\sum_{i=1}^6 Z_i^2 \leq b\right) = 0.95$$

解: 由于卡方分布是由标准正态分布的平方和构成的:

设 $Z_1, Z_2, \dots, Z_n$ 是来自总体 $N(0,1)$ 的样本, 那么统计量

$$\chi^2 = Z_1^2 + Z_2^2 + \dots + Z_n^2$$

服从自由度为 n 的 χ^2 分布, 记为 $\chi^2 \sim \chi^2(n)$

因此, 令 $\chi^2 = \sum_{i=1}^6 Z_i^2$, 那么 $\chi^2 = \sum_{i=1}^6 Z_i^2 : \chi^2(6)$, 那么由概率 $P\left(\sum_{i=1}^6 Z_i^2 \leq b\right) = 0.95$, 可知:

$$b = \chi_{1-0.95}^2(6)$$

6.4 在习题 6.1 中, 假定装瓶机对瓶子的灌装量服从方差 $\sigma^2 = 1$ 的标准正态分布。假定我们方案随机抽取 10 个瓶子组成样本, 观测每个瓶子的灌装量, 得到 10 个观测值, 用这 10 个观测值我们可以求出样本方

差 $S^2 (S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2)$, 确定一个适宜的范围使得有较大的概率保证 S^2 落入其中是有用的, 试求

b_1, b_2 , 使得

$$P(b_1 \leq S^2 \leq b_2) = 0.90$$

解: 更加样本方差的抽样分布知识可知, 样本统计量:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

此处, $n=10, \sigma^2=1$, 所以统计量

$$\frac{(n-1)S^2}{\sigma^2} = \frac{(10-1)S^2}{1} = 9S^2 \sim \chi^2(n-1)$$

根据卡方分布的可知:

$$P(b_1 \leq S^2 \leq b_2) = P(9b_1 \leq 9S^2 \leq 9b_2) = 0.90$$

又因为:

$$P(\chi_{1-\alpha/2}^2(n-1) \leq 9S^2 \leq \chi_{\alpha/2}^2(n-1)) = 1 - \alpha$$

因此:

$$P(9b_1 \leq 9S^2 \leq 9b_2) = P(\chi_{1-\alpha/2}^2(n-1) \leq 9S^2 \leq \chi_{\alpha/2}^2(n-1)) = 1 - \alpha = 0.90$$

$$\Rightarrow P(9b_1 \leq 9S^2 \leq 9b_2) = P(\chi_{1-\alpha/2}^2(n-1) \leq 9S^2 \leq \chi_{\alpha/2}^2(n-1))$$

$$= P(\chi_{0.95}^2(9) \leq 9S^2 \leq \chi_{0.05}^2(9)) = 0.90$$

那么:

$$\Rightarrow 9b_1 = \chi_{0.95}^2(9), 9b_2 = \chi_{0.05}^2(9) \Rightarrow b_1 = \frac{\chi_{0.95}^2(9)}{9}, b_2 = \frac{\chi_{0.05}^2(9)}{9}$$

查概率表: $\chi_{0.95}^2(9)=3.325, \chi_{0.05}^2(9)=19.919$, 那么

$$b_1 = \frac{\chi_{0.95}^2(9)}{9} = 0.369, b_2 = \frac{\chi_{0.05}^2(9)}{9}$$

$$7.1 \quad (1) \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{40}}$$

$$(2) \quad \Delta_{\bar{x}} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{5}{\sqrt{40}}$$

7.2 某快餐店想要估计每位顾客午餐的平均花费金额。在为期 3 周的时间里选取 49 名顾客组成了一个简单随机样本。

(1) 假定总体标准差为 15 元，求样本均值的抽样标准误差。 $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{49}}$

(2) 在 95% 的置信水平下，求估计误差。

$\Delta_{\bar{x}} = t \cdot \sigma_{\bar{x}}$ ，由于是大样本抽样，因此样本均值服从正态分布，因此概率度 $t = z_{\alpha/2}$

因此， $\Delta_{\bar{x}} = t \cdot \sigma_{\bar{x}} = z_{\alpha/2} \cdot \sigma_{\bar{x}} = z_{0.025} \cdot \sigma_{\bar{x}} \times$

(3) 如果样本均值为 120 元，求总体均值的 95% 的置信区间。

置信区间为： $\left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = (120 - 4.2, 120 + 4.2) = (115.8, 124.2)$

$$7.3 \quad \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 104560 \pm 1.96 \frac{85414}{\sqrt{100}} = (87818.856, 121301.144)$$

7.4 从总体中抽取一个 $n=100$ 的简单随机样本，得到 $\bar{x}=81$ ， $s=12$ 。

要求：

大样本，样本均值服从正态分布： $\bar{x} : N\left(\mu, \frac{\sigma^2}{n}\right)$ 或 $\bar{x} : N\left(\mu, \frac{s^2}{n}\right)$

置信区间为： $\left(\bar{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \right)$ ， $\frac{s}{\sqrt{n}} = \frac{12}{\sqrt{100}}$

(1) 构建 μ 的 90% 的置信区间。

$z_{\alpha/2} = z_{0.05} = 1.645$ ，置信区间为： $(81 - 1.645 \times 1.2, 81 + 1.645 \times 1.2) = (79.03, 82.97)$

(2) 构建 μ 的 95% 的置信区间。

$z_{\alpha/2} = z_{0.025} = 1.96$ ，置信区间为： $(81 - 1.96 \times 1.2, 81 + 1.96 \times 1.2) = (78.65, 83.35)$

(3) 构建 μ 的 99% 的置信区间。

$z_{\alpha/2} = z_{0.005} = 2.576$ ，置信区间为： $(81 - 2.576 \times 1.2, 81 + 2.576 \times 1.2) = (77.91, 84.09)$

$$7.5 \quad (1) \quad \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 25 \pm 1.96 \cdot \frac{3.5}{\sqrt{60}} = (24.114, 25.886)$$

$$(2) \quad \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 119.6 \pm 2.326 \cdot \frac{23.89}{\sqrt{75}} = (113.184, 126.016)$$

$$(3) \quad \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 3.419 \pm 1.645 \cdot \frac{0.974}{\sqrt{32}} = (3.136, 3.702)$$

$$7.6 \quad (1) \quad \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 8900 \pm 1.96 \cdot \frac{500}{\sqrt{15}} = (8646.965, 9153.035)$$

$$(2) \quad \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 8900 \pm 1.96 \cdot \frac{500}{\sqrt{35}} = (8734.35, 9065.65)$$

$$(3) \quad \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 8900 \pm 1.645 \cdot \frac{500}{\sqrt{35}} = (8761.395, 9038.605)$$

$$(4) \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 8900 \pm 2.58 \cdot \frac{500}{\sqrt{35}} = (8681.95, 9118.05)$$

7.7 某大学为了解学生每天上网的时间，在全校 7 500 名学生中采取重复抽样方法随机抽取 36 人，调查他们每天上网的时间，得到下面的数据(单位：小时)：

求该校大学生平均上网时间的置信区间，置信水平分别为 90%，95%和 99%。

解：

(1) 样本均值 $\bar{x}$

$$1-\alpha=0.9, t=z_{\alpha/2}=z_{0.05}=1.645, \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 3.32 \pm 1.645 \cdot \frac{1.61}{\sqrt{36}} = (2.88, 3.76)$$

$$1-\alpha=0.95, t=z_{\alpha/2}=z_{0.025}=1.96, \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 3.32 \pm 1.96 \cdot \frac{1.61}{\sqrt{36}} = (2.79, 3.85)$$

$$1-\alpha=0.99, t=z_{\alpha/2}=z_{0.005}=2.576, \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 3.32 \pm 2.76 \cdot \frac{1.61}{\sqrt{36}} = (2.63, 4.01)$$

$$7.8 \bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 10 \pm 2.365 \cdot \frac{3.464}{\sqrt{8}} = (7.104, 12.896)$$

7.9 某居民小区为研究职工上班从家里到单位的距离，抽取了由 16 个人组成的一个随机样本，他们到单位的距离(单位：km)分别是：

10 3 14 8 6 9 12 11 7 5 10 15 9 16 13 2

假定总体服从正态分布，求职工上班从家里到单位平均距离的 95%的置信区间。

解：小样本，总体方差未知，用 t 统计量 $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} : t(n-1)$

均值=9.375，样本标准差 s=4.11， $1-\alpha=0.95$ ， $n=16$ ， $t_{\alpha/2}(n-1)=t_{0.025}(15)$

$$\begin{aligned} \text{置信区间: } & \left(\bar{x} - t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}} \right) \\ & = \left(9.375 - 2.13 \times \frac{4.11}{\sqrt{16}}, 9.375 + 2.13 \times \frac{4.11}{\sqrt{16}} \right) = (7.18, 11.57) \end{aligned}$$

$$7.10 (1) \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 149.5 \pm 1.96 \cdot \frac{1.93}{\sqrt{36}} = (148.8695, 150.1305)$$

(2)中心极限定理

7. 11 某企业生产的袋装食品采用自动打包机包装，每袋标准重量为 100g。现从某天生产的一批产品中按重复抽样随机抽取 50 包进行检查，测得每包重量(单位：g)如下：

每包重量 (g)	包数
96~98	2
98~100	3
100~102	34
102~104	7
104~106	4
合计	50

食品包重量服从正态分布，要求：

(1)确定该种食品平均重量的 95%的置信区间。

解：大样本，总体方差未知，用 z 统计量： $z = \frac{\bar{x} - \mu}{s/\sqrt{n}} : N(0,1)$

样本均值=101.4, 样本标准差 s=1.829, $1-\alpha=0.95$, $z_{\alpha/2}=z_{0.025}$

$$\begin{aligned} \text{置信区间: } & \left(\bar{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \right) \\ & = \left(101.4 - 1.96 \times \frac{1.829}{\sqrt{50}}, 101.4 + 1.96 \times \frac{1.829}{\sqrt{50}} \right) = (100.89, 101.91) \end{aligned}$$

(2)如果规定食品重量低于 100g 属于不合格, 确定该批食品合格率的 95%的置信区间。

解: 总体比率的估计。大样本, 总体方差未知, 用 z 统计量: $z = \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}}: N(0,1)$

样本比率= (50-5) /50=0.9, $1-\alpha=0.95$, $z_{\alpha/2}=z_{0.025}$

$$\begin{aligned} \text{置信区间: } & \left(p - z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}, p + z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \right) \\ & = \left(0.9 - 1.96 \times \sqrt{\frac{0.9(1-0.9)}{50}}, 0.9 + 1.96 \times \sqrt{\frac{0.9(1-0.9)}{50}} \right) = (0.8168, 0.9832) \end{aligned}$$

7.12 正态分布, 大样本, 方差未知

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 16.128 \pm 2.576 \cdot \frac{0.8706}{\sqrt{25}} = (15.679, 16.576)$$

7. 13 一家研究机设想估计在网络公司工作的员工每周加班的平均时间, 为此随机抽取了 18 个员工。得到他们每周加班的时间数据如下(单位: 小时):

6	21	17	20	7	0	8	16	29
3	8	12	11	9	21	25	15	16

假定员工每周加班的时间服从正态分布。估计网络公司员工平均每周加班时间的 90%的置信区间。

解: 小样本, 总体方差未知, 用 t 统计量: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}: t(n-1)$

均值=13.56, 样本标准差 s=7.801, $1-\alpha=0.90$, n=18, $t_{\alpha/2}(n-1)=t_{0.05}(17)$

置信区间:

$$\begin{aligned} & \left(\bar{x} - t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}} \right) \\ & = \left(13.56 - 1.7369 \times \frac{7.801}{\sqrt{18}}, 13.56 + 1.7369 \times \frac{7.801}{\sqrt{18}} \right) = (10.36, 16.75) \end{aligned}$$

$$7.14 \quad (1) \quad p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} = 0.51 \pm 2.576 \cdot \sqrt{\frac{0.51(1-0.51)}{44}} = (0.33159, 0.7041)$$

$$(2) \quad p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} = 0.82 \pm 1.96 \cdot \sqrt{\frac{0.82(1-0.82)}{300}} = (0.7765, 0.8635)$$

$$(3) \quad p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} = 0.48 \pm 1.645 \cdot \sqrt{\frac{0.48(1-0.48)}{1150}} = (0.4558, 0.5042)$$

7. 15 在一家电市场调查。随机抽取了 200 个居民户, 调查他们是否拥有某一品牌的电视机。其中拥有该品牌电视机的家庭占 23%。求总体比例的置信区间, 置信水平分别为 90%和 95%。

解: 总体比率的估计

大样本，总体方差未知，用 z 统计量： $z = \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}} : N(0,1)$

样本比率=0.23, $1 - \alpha = 0.90$, $z_{\alpha/2} = z_{0.025}$

$$\text{置信区间: } \left(p - z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}, p + z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \right) \\ = \left(0.23 - 1.645 \times \sqrt{\frac{0.23(1-0.23)}{200}}, 0.23 + 1.645 \times \sqrt{\frac{0.23(1-0.23)}{200}} \right) = (0.1811, 0.2789)$$

$1 - \alpha = 0.95$, $z_{\alpha/2} = z_{0.025}$

$$\left(p - z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}, p + z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \right) \\ = \left(0.23 - 1.96 \times \sqrt{\frac{0.23(1-0.23)}{200}}, 0.23 + 1.96 \times \sqrt{\frac{0.23(1-0.23)}{200}} \right) = (0.1717, 0.2883)$$

$$7.16 \quad n = \frac{(z_{\alpha/2})^2 s^2}{E^2} = \frac{2.576^2 1000^2}{200^2} = 166$$

$$7.17 \quad (1) \quad n = \frac{(z_{\alpha/2})^2 \pi(1-\pi)}{E^2} = \frac{2.05^2 0.4(1-0.4)}{0.02^2} = 2522$$

$$(2) \quad n = \frac{(z_{\alpha/2})^2 \pi(1-\pi)}{E^2} = \frac{1.96^2 0.5(1-0.5)}{0.04^2} = 601 \quad (\text{当 } \pi \text{ 未知是, 取 } 0.5)$$

$$(3) \quad n = \frac{(z_{\alpha/2})^2 \pi(1-\pi)}{E^2} = \frac{1.645^2 0.55(1-0.55)}{0.05^2} = 328$$

$$7.18 \quad (1) \quad p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} = 0.64 \pm 1.96 \cdot \sqrt{\frac{0.64(1-0.64)}{50}} = (0.5070, 0.7731)$$

$$(2) \quad n = \frac{(z_{\alpha/2})^2 \pi(1-\pi)}{E^2} = \frac{1.96^2 0.8(1-0.8)}{0.1^2} = 62$$

7. 20 顾客到银行办理业务时往往需要等待一段时间，而等待时间的长短与许多因素有关，比方，银行业务员办理业务的速度，顾客等待排队的方式等。为此，某银行准备采取两种排队方式进行试验，第一种排队方式是：所有顾客都进入一个等待队列；第二种排队方式是：顾客在三个业务窗口处列队三排等待。为比拟哪种排队方式使顾客等待的时间更短，银行各随机抽取 10 名顾客，他们在办理业务时所等待的时间(单位：分钟)如下：

方式 1										
方式 2										10

要求：

(1)构建第一种排队方式等待时间标准差的 95%的置信区间。

$$\text{解: 估计统计量: } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

经计算得样本标准差 $s_2^2 = 3.318$, $1 - \alpha = 0.95$, $n=10$,

$$\chi_{\alpha/2}^2(n-1) = \chi_{0.025}^2(9) = 19.02, \quad \chi_{1-\alpha/2}^2(n-1) = \chi_{0.975}^2(9)$$

$$\text{置信区间: } \frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)} = \left(\frac{9 \times 0.2272}{19.02}, \frac{9 \times 0.2272}{2.7} \right) = (0.1075, 0.7574)$$

因此, 标准差的置信区间为 (0.3279, 0.8703)

(2) 构建第二种排队方式等待时间标准差的 95% 的置信区间。

$$\text{解: 估计统计量: } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

经计算得样本标准差 $s_1^2 = 0.2272$, $1 - \alpha = 0.95$, $n = 10$,

$$\chi_{\alpha/2}^2(n-1) = \chi_{0.025}^2(9) = 19.02, \quad \chi_{1-\alpha/2}^2(n-1) = \chi_{0.975}^2(9)$$

$$\text{置信区间: } \frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)} = \left(\frac{9 \times 3.318}{19.02}, \frac{9 \times 3.318}{2.7} \right) = (1.57, 11.06)$$

因此, 标准差的置信区间为 (1.25, 3.33)

(3) 根据(1)和(2)的结果, 你认为哪种排队方式更好?

第一种方式好, 标准差小。

7.21 正态总体, 独立小样本, 方差未知但相等:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \quad \left(\text{其中 } s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}, \quad df: (n_1 + n_2 - 2) \right)$$

$$(1) \quad t_{\alpha/2}(n_1 + n_2 - 1) = t_{0.05}(14 + 7 - 2) = 1.7291, \text{ 代入略}$$

$$(2) \quad t_{\alpha/2}(n_1 + n_2 - 1) = t_{0.025}(14 + 7 - 2) = 2.0930, \text{ 代入略}$$

$$(3) \quad t_{\alpha/2}(n_1 + n_2 - 1) = t_{0.05}(14 + 7 - 2) = 2.8609, \text{ 代入略}$$

7.22

(1) 正态或非正态总体, 独立大样本, 方差未知

$$(\bar{x}_1 - \bar{x}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(2) 正态总体, 独立小样本, 方差未知但 $\sigma_1 = \sigma_2$:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \quad \left(\text{其中 } s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}, \quad df: (n_1 + n_2 - 2) \right)$$

(3) 正态总体, 独立小样本, 方差未知 $\sigma_1 \neq \sigma_2$ 但 $n_1 = n_2$, $df = n_1 + n_2 - 2$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(4) 正态总体, 独立小样本, 方差未知但 $\sigma_1 = \sigma_2$, $n_1 \neq n_2$:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \quad \left(\text{其中 } s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}, \quad df: (n_1 + n_2 - 2) \right)$$

(5) 正态总体, 独立小样本, 方差未知但 $\sigma_1 \neq \sigma_2$, $n_1 \neq n_2$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \left(\text{其中 } df = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} \right)$$

7. 23 下表是由 4 对观察值组成的随机样本。

配对号	来自总体 A 的样本	来自总体 B 的样本
1	2	0
2	5	7
3	10	6
4	8	5

(1) 计算 A 与 B 各对观察值之差，再利用得出的差值计算 $\bar{d}$ 和 s_d 。

$$\bar{d}=1.75, s_d$$

(2) 设 μ_1 和 μ_2 分别为总体 A 和总体 B 的均值，构造 $\mu_d = \mu_1 - \mu_2$ 的 95% 的置信区间。

解：小样本，配对样本，总体方差未知，用 t 统计量

$$t_{\bar{d}} = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} : t(n-1)$$

均值=1.75，样本标准差 s=2.62996， $1-\alpha=0.95$ ， $n=4$ ， $t_{\alpha/2}(n-1)=t_{0.025}(3)$

$$\begin{aligned} \text{置信区间: } & \left(\bar{d} - t_{\alpha/2}(n-1) \cdot \frac{s_d}{\sqrt{n}}, \bar{d} + t_{\alpha/2}(n-1) \cdot \frac{s_d}{\sqrt{n}} \right) \\ & = \left(1.75 - 3.182 \times \frac{2.62996}{\sqrt{4}}, 1.75 + 3.182 \times \frac{2.62996}{\sqrt{4}} \right) = (-2.43, 5.93) \end{aligned}$$

7.24 小样本，配对样本，总体方差未知： $t_{\alpha/2}(n-1)=t_{0.025}(10-1)$

$$\bar{d} \pm t_{\alpha/2}(n-1) \cdot \frac{s_d}{\sqrt{n}} = 11 \pm 2.2622 \cdot \frac{6.532}{\sqrt{10}} = (6.3272, 15.6728)$$

7. 25 从两个总体中各抽取一个 $n_1 = n_2 = 250$ 的独立随机样本，来自总体 1 的样本比例为 $p_1 = 40\%$ ，来自总体 2 的样本比例为 $p_2 = 30\%$ 。要求：

(1) 构造 $\pi_1 - \pi_2$ 的 90% 的置信区间。

(2) 构造 $\pi_1 - \pi_2$ 的 95% 的置信区间。

解：总体比率差的估计

$$\text{大样本，总体方差未知，用 } z \text{ 统计量： } z = \frac{p_1 - p_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} : N(0,1)$$

样本比率 $p_1=0.4$ ， $p_2=0.3$ ，

置信区间：

$$\left(p_1 - p_2 - z_{\alpha/2} \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}, p_1 - p_2 + z_{\alpha/2} \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right)$$

$1-\alpha=0.90$ ， $z_{\alpha/2}=z_{0.025}$

$$\begin{aligned} & \left(p_1 - p_2 - z_{\alpha/2} \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}, p_1 - p_2 + z_{\alpha/2} \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right) \\ & = \left(0.1 - 1.645 \times \sqrt{\frac{0.4(1-0.4)}{250} + \frac{0.3(1-0.3)}{250}}, 0.1 + 1.645 \times \sqrt{\frac{0.4(1-0.4)}{250} + \frac{0.3(1-0.3)}{250}} \right) \\ & = (3.02\%, 16.98\%) \end{aligned}$$

$1-\alpha=0.95$ ， $z_{\alpha/2}=z_{0.025}$

$$\begin{aligned} & \left(p_1 - p_2 - z_{\alpha/2} \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}, p_1 - p_2 + z_{\alpha/2} \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right) \\ & = \left(0.1 - 1.96 \times \sqrt{\frac{0.4(1-0.4)}{250} + \frac{0.3(1-0.3)}{250}}, 0.1 + 1.96 \times \sqrt{\frac{0.4(1-0.4)}{250} + \frac{0.3(1-0.3)}{250}} \right) \\ & = (1.68\%, 18.32\%) \end{aligned}$$

7.26 生产工序的方差是工序质量的一个重要度量。当方差较大时，需要对序进行改良以减小方差。下面是两部机器生产的袋茶重量(单位：g)的数据：

机器 1			机器 2		

要求：构造两个总体方差比 σ_1^2/σ_2^2 的 95% 的置信区间。

解：统计量：
$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} : F(n_1-1, n_2-1)$$

置信区间：
$$\left(\frac{s_1^2/s_2^2}{F_{\alpha/2}(n_1-1, n_2-1)}, \frac{s_1^2/s_2^2}{F_{1-\alpha/2}(n_1-1, n_2-1)} \right)$$

$$s_1^2=0.058, s_2^2=0.006, n_1=n_2=21, 1-\alpha=0.95, F_{\alpha/2}(n_1-1, n_2-1)=F_{0.025}(20, 20)=2.4645,$$

$$F_{1-\alpha/2}(n_1-1, n_2-1)=\frac{1}{F_{\alpha/2}(n_2-1, n_1-1)}$$

$$F_{1-\alpha/2}(n_1-1, n_2-1)=F_{0.975}(20, 20)=\frac{1}{F_{0.025}(20, 20)}$$

$$\left(\frac{s_1^2/s_2^2}{F_{\alpha/2}(n_1-1, n_2-1)}, \frac{s_1^2/s_2^2}{F_{1-\alpha/2}(n_1-1, n_2-1)} \right) = (4.05, 24.6)$$

7. 27 根据以往的生产数据，某种产品的废品率为 2%。如果要求 95% 的置信区间，假设要求估计误差（边际误差）不超过 4%，应抽取多大的样本？

解：
$$z_{\alpha/2} = \frac{\Delta_p}{\sqrt{\frac{p(1-p)}{n}}}, n = \frac{z_{\alpha/2}^2 \cdot p \cdot (1-p)}{\Delta_p^2}, 1-\alpha=0.95, z_{\alpha/2}=z_{0.025}$$

$$n = \frac{z_{\alpha/2}^2 \cdot p \cdot (1-p)}{\Delta_p^2} = \frac{1.96^2 \times 0.02 \times 0.98}{0.04^2} = 47.06, \text{ 取 } n=48 \text{ 或者 } 50.$$

7. 28 某超市想要估计每个顾客平均每次购物花费的金额。根据过去的经验，标准差大约为 120 元，现要求以 95% 的置信水平估计每个顾客平均购物金额的置信区间，并要求边际误差不超过 20 元，应抽取多少个顾客作为样本？

解：
$$n = \frac{z_{\alpha/2}^2 \cdot \sigma^2}{\Delta_{\bar{x}}^2}, 1-\alpha=0.95, z_{\alpha/2}=z_{0.025}=1.96,$$

$$n = \frac{z_{\alpha/2}^2 \cdot \sigma^2}{\Delta_{\bar{x}}^2} = \frac{1.96^2 \times 120^2}{20^2} = 138.3, \text{ 取 } n=139 \text{ 或者 } 140, \text{ 或者 } 150.$$

第八章 假设检验

8.1 提出假设: $H_0: \mu=4.55$; $H_1: \mu \neq$

$$\text{构建统计量 (正态, 小样本, 方差): } z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{4.484 - 4.55}{0.108/\sqrt{9}}$$

求临界值: $\alpha=0.05$, $z_{\alpha/2} = z_{0.025}$

决策: 因为 $z > -z_{\alpha/2}$, 所有, 不拒绝 H_0

8.2 一种元件, 要求其使用寿命不得低于 700 小时。现从一批这种元件中随机抽取 36 件, 测得其平均寿命为 680 小时。该元件寿命服从正态分布, $\sigma=60$ 小时, 试在显著性水平 0.05 下确定这批元件是否合格。

解: 提出假设: $H_0: \mu \geq 700$; $H_1: \mu < 700$

$$\text{构建统计量 (正态, 大样本, 方差): } z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{680 - 700}{60/\sqrt{36}} = -2$$

求临界值: 当 $\alpha=0.05$, 查表得 $z_{\alpha} = 1.645$ 。

决策: 因为 $z < -z_{\alpha}$, 故拒绝原假设, 接受备择假设

结论: 说明这批产品不合格。

8.3 提出假设: $H_0: \mu \leq 250$; $H_1: \mu > 250$

$$\text{构建统计量 (正态, 小样本, 方差): } z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{270 - 250}{30/\sqrt{25}}$$

求临界值: $\alpha=0.05$, $z_{\alpha} = z_{0.05} =$

决策: 因为 $z > z_{\alpha/2}$, 所有, 拒绝 H_0

结论: 明显增产

8.4 糖厂用自动打包机打包, 每包标准重量是 100 千克。每天开工后需要检验一次打包机工作是否正常。某日开工后测得 9 包重量(单位: 千克)如下:

99.3 98.7 100.5 101.2 98.3 99.7 99.5 102.1 100.5

包重服从正态分布, 试检验该日打包机工作是否正常($\alpha=0.05$)?

解: 提出假设: $H_0: \mu=100$; $H_1: \mu \neq 100$

$$\text{构建统计量 (正态, 小样本, 方差未知): } t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{99.9778 - 100}{1.21221/\sqrt{9}}$$

求临界值: 当 $\alpha=0.05$, 自由度 $n-1=8$ 时, 查表得 $t_{\alpha/2}(8) =$ 。

决策: 因为 $|t| < t_{\alpha/2}$, 样本统计量落在接受区域, 故接受原假设, 拒绝备择假设

结论: 说明打包机工作正常。

8.5 某种大量生产的袋装食品, 按规定不得少于 250 克。今从一批该食品中任意抽取 50 袋, 发现有 6 袋低于 250 克。假设规定不符合标准的比例超过 5% 就不得出厂, 问该批食品能否出厂($\alpha=0.05$)?

解: 提出假设: $H_0: \pi \leq 0.05$; $H_1: \pi$

$$\text{构建统计量: } Z = \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}} = \frac{0.12 - 0.05}{\sqrt{\frac{0.05 \times (1 - 0.05)}{50}}}$$

求临界值: 当 $\alpha=0.05$, 查表得 $z_{\alpha} = 1.645$ 。

决策: 因为 $z > z_{\alpha}$, 样本统计量落在拒绝区域, 故拒绝原假设, 接受备择假设

结论: 说明该批食品不能出厂。

8.6 提出假设: $H_0: \mu \leq 25000$; $H_1: \mu > 25000$

构建统计量 (正态, 小样本, 方差): $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{27000 - 25000}{5000/\sqrt{15}}$

求临界值: 当 $\alpha=0.05$, 查表得 $z_\alpha=1.645$ 。

决策: 因为 $z < z_\alpha$, 故不能拒绝原假设

结论: 没有充分证据证明该厂家的广告是真实的

8. 7 某种电子元件的寿命 x (单位: 小时)服从正态分布。现测得 16 只元件的寿命如下:

159 280 101 212 224 379 179 264
222 362 168 250 149 260 485 170

问是否有理由认为元件的平均寿命显著地大于 225 小时($\alpha=0.05$)?

解: 提出假设: $H_0: \mu \leq 225$; $H_1: \mu > 225$

构建统计量 (正态, 小样本, 方差): $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{241.5 - 225}{98.726/\sqrt{16}}$

求临界值: 当 $\alpha=0.05$, 自由度 $n-1=15$ 时, 查表得 $t_\alpha(15)$

决策: 因为 $t < t_\alpha$, 样本统计量落在接受区域, 故接受原假设, 拒绝备择假设

结论: 说明元件寿命没有显著大于 225 小时。

8. 10 装配一个部件时可以采用不同的方法, 所关心的是哪一个方法的效率更高。劳动效率可以用平均装配时间反映。现从不同的装配方法中各抽取 12 件产品, 记录各自的装配时间(单位: 分钟)如下:

甲方法: 31 34 29 32 35 38 34 30 29 32 31 26

乙方法: 26 24 28 29 30 29 32 26 31 29 32 28

两总体为正态总体, 且方差相同。问两种方法的装配时间有无显著不同 ($\alpha=0.05$)?

解: 提出假设: $H_0: \mu_1 - \mu_2 = 0$; $H_1: \mu_1 - \mu_2 \neq 0$

构建统计量 (总体正态, 小样本抽样, 方差未知, 方差相等): $t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

根据样本数据计算, 得 $n_1=12$, $n_2=12$, $\bar{x}_1=31.75$, $s_1=3.19446$, $\bar{x}_2=28.6667$, $s_2=2.46183$ 。

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} = \frac{(12-1) \times 0.92216^2 + (12-1) \times 0.71067^2}{12+12-2}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

求临界值: $\alpha=0.05$ 时, 临界点为 $t_{\alpha/2}(n_1 + n_2 - 2) = t_{0.025}(22)$

决策: 此题中 $|t| > t_{\alpha/2}$, 故拒绝原假设

结论: 认为两种方法的装配时间有显著差异

8. 11 调查了 339 名 50 岁以上的人, 其中 205 名吸烟者中有 43 个患慢性气管炎, 在 134 名不吸烟者中有 13 人患慢性气管炎。调查数据能否支持“吸烟者容易患慢性气管炎”这种观点($\alpha=0.05$)?

解: 提出假设: $H_0: \pi_1 \leq \pi_2$; $H_1: \pi_1 > \pi_2$

$$p_1 = 43/205 = 0.2097 \quad n_1 = 205 \quad p_2 = 13/134 = 0.097 \quad n_2 = 134$$

$$\text{构建统计量: } z = \frac{(p_1 - p_2) - d}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} = \frac{(0.2098 - 0.097) - 0}{\sqrt{\frac{0.2098(1-0.2098)}{205} + \frac{0.097(1-0.097)}{134}}} = 3$$

求临界值: 当 $\alpha=0.05$, 查表得 z_α

决策：因为 $z > z_{\alpha}$ ，拒绝原假设

结论：说明吸烟者容易患慢性气管炎

8.12 为了控制贷款规模，某商业银行有个内部要求，平均每项贷款数额不能超过 60 万元。随着经济的开展，贷款规模有增大的趋势。银行经理想了解在同样工程条件下，贷款的平均规模是否明显地超过 60 万元，故一个 $n=144$ 的随机样本被抽出，测得 $\bar{x}=68.1$ 万元， $s=45$ 。用 $\alpha=0.01$ 的显著性水平，采用 p 值进行检验。

解：提出假设： $H_0: \mu \leq 60$ ； $H_1: \mu > 60$

$$\text{构建统计量 (大样本, 方差未知): } z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{68.1 - 60}{45/\sqrt{144}}$$

求临界值：由于 $\bar{x} > \mu$ ，因此 P 值 $= P(z \geq 2.16) = 1 - \phi(2.16)$ ，查表的 $\phi(2.16)$

决策：由于 $P > \alpha = 0.01$ ，故不能拒绝原假设

结论：说明贷款的平均规模没有明显地超过 60 万元。

8.13 有一种理论认为服用阿司匹林有助于减少心脏病的发生，为了进行验证，研究人员把自愿参与实验的 22 000 人随机平均分成两组，一组人员每星期服用三次阿司匹林(样本 1)，另一组人员在相同的时间服用抚慰剂(样本 2)持续 3 年之后进行检测，样本 1 中有 104 人患心脏病，样本 2 中有 189 人患心脏病。以 $\alpha=0.05$ 的显著性水平检验服用阿司匹林是否可以降低心脏病发生率。

解：提出假设： $H_0: \pi_1 \geq \pi_2$ ； $H_1: \pi_1 < \pi_2$

$$p_1 = 104/11000 = 0.00945 \quad n_1 = 11000 \quad p_2 = 189/11000 = 0.01718 \quad n_2 = 11000$$

$$\begin{aligned} \text{构建统计量: } z &= \frac{(p_1 - p_2) - d}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \\ &= \frac{(0.00945 - 0.01718) - 0}{\sqrt{\frac{0.00945(1-0.00945)}{11000} + \frac{0.01718(1-0.01718)}{11000}}} = -5 \end{aligned}$$

求临界值：当 $\alpha=0.05$ ，查表得 z_{α}

决策：因为 $z < -z_{\alpha}$ ，拒绝原假设

结论：说明用阿司匹林可以降低心脏病发生率。

8.14

8.15 有人说在大学中男生的学习成绩比女生的学习成绩好。现从一个学校中随机抽取了 25 名男生和 16 名女生，对他们进行了同样题目的测试。测试结果说明，男生的平均成绩为 82 分，方差为 56 分，女生的平均成绩为 78 分，方差为 49 分。假设显著性水平 $\alpha=0.02$ ，从上述数据中能得到什么结论？

解：方差比检验：

$$\text{提出假设: } H_0: \sigma_1^2 = \sigma_2^2; H_1: \sigma_1^2 \neq \sigma_2^2$$

$$(: n_1=25, s_1^2=56, n_2=16, s_2^2=49)$$

$$\text{构建统计量: } F = \frac{s_1^2}{s_2^2} = \frac{56}{49}$$

求临界值：当 $\alpha=0.02$ 时， $F_{\alpha/2}(24,15)=3.294$ ， $F_{1-\alpha/2}(24,15)=0.346$ 。

决策：由于 $F_{1-\alpha/2}(24,15) < F < F_{\alpha/2}(24,15)$ ，检验统计量的值落在接受域中，所以接受原假设

结论：说明总体方差无显著差异。

检验均值差：

$$\text{提出假设: } H_0: \mu_1 - \mu_2 \leq 0; H_1: \mu_1 - \mu_2 > 0$$

$$\text{构建统计量 (总体正态, 小样本抽样, 方差未知, 方差相等): } t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

根据样本数据计算，得 $n_1=25$, $n_2=16$, $\bar{x}_1=82$, $s_1^2=56$, $\bar{x}_2=78$, $s_2^2=49$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = 53.308; \quad t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

求临界值： $\alpha=0.02$ 时，临界点为 $t_\alpha(n_1+n_2-2) = t_{0.02}(39) = 2.125$, $t < t_\alpha$ ，故不能拒绝原假设
结论：不能认为大学中男生的学习成绩比女生的学习成绩好。

第9章 分类数据分析

9.1

	低收入组	偏低收入组	偏高收入组	高收入组	
经常购置	25 ()	40 ()	47 ()	46 ()	158
不购置	69 ()	51 ()	74 ()	57 ()	251
有时购置	36 ()	26 ()	19 ()	37 (3)	118
	130	117	140	140	527

第一步：提出假设：

H_0 : 收入与购置习惯无关 (相互独立); H_1 : 收入与购置习惯有关 (相互不独立)

或者 $H_0: P_{ij} = P_{ig} \times P_{gj}$; $H_1: P_{ij} \neq P_{ig} \times P_{gj}$

第二步：构建统计量：

利用公式： $E_{ij} = \frac{N_{ig} \times N_{gj}}{N}$ 先计算期望频次分布，如上表括号中数据。

由 $\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$: $\chi^2((r-1)(c-1))$ 得

$$\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(n_{ij} - E_{ij})^2}{E_{ij}} = \frac{(25 - 38.98)^2}{38.98} + \frac{(40 - 35.08)^2}{35.08} + \dots = 17.63 \quad (p=0.007227)$$

第三步：求临界值： $\chi_{0.1}^2(6) = 10.6446$

(注意：①对于 $r \times c$ 列联表的自由度是： $df=(r-1) \times (c-1)$ ；②按右侧检验方法)

第四步：决策：

因为 $\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(n_{ij} - E_{ij})^2}{E_{ij}} = 17.63 > \chi_{0.1}^2(6) = 10.6446$ ，所以拒绝 H_0

第五步：结论：所以收入与购置习惯有关

假设： $H_0: \pi_1 = 0.1, \pi_2 = 0.2, \pi_3 = 0.3, \pi_4 = 0.2, \pi_5 = 0.2$; H_1 : 至少有一个不成立

$$\text{统计量: } \chi^2 = \sum_{i=1}^n \frac{(f_o - f_e)^2}{f_e} =$$

$$\frac{(0.14 - 0.1)^2}{0.1} + \frac{(0.28 - 0.2)^2}{0.2} + \frac{(0.24 - 0.3)^2}{0.3} + \frac{(0.18 - 0.2)^2}{0.2} + \frac{(0.16 - 0.2)^2}{0.2}$$

$$\text{临界值: } \chi_{0.1}^2(5-1) = 7.7794 \quad (P=0.9994)$$

决策：因为 $\chi^2 = \sum_{i=1}^n \frac{(f_o - f_e)^2}{f_e} = 7.7794 < \chi_{0.1}^2(4) = 7.7794$ ，所以不能拒绝原假设。

结论：没有发生变化。

9.3

阅读习惯	大学以上	大学和大专	高中	高中以下	
早上看	6 ()	13 ()	14 ()	17 ()	50
中午看	12 ()	16 ()	8 ()	8 ()	44
晚上看	38 ()	40 ()	11 ()	6 ()	95
有空看	21 ()	22 ()	9 ()	13 ()	65
	77	91	42	44	254

第一步：提出假设：

H_0 ：阅读习惯与其文化程度无关（相互独立）； H_1 ：（相互不独立）

第二步：构建统计量：

利用公式： $E_{ij} = \frac{N_{i\cdot} \times N_{\cdot j}}{N}$ 先计算期望频次分布，如上表括号中数据。

由 $\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$: $\chi^2((r-1)(c-1))$ 得

$$\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(n_{ij} - E_{ij})^2}{E_{ij}} = \frac{(6-15.16)^2}{15.16} + \frac{(13-17.91)^2}{17.91} + \dots = 36.8611 \quad (p=0.00021)$$

第三步：求临界值： $\chi_{0.05}^2(9) = 16.919$

第四步：决策：因为 $\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$ 大于 $\chi_{0.05}^2(9)$ ，所以拒绝 H_0

第五步：结论：所以阅读习惯与其文化程度有关

9.4

本科专业	MBA 所选课程			
	会计	统计	市场营销	
专业一	31 ()	13 ()	16 ()	60
专业二	8 ()	16 ()	7 ()	31
专业三	12 ()	10 ()	17 ()	39
其他专业	10 ()	5 ()	7 ()	22
	61	44	47	152

(1) 第一步：提出假设： H_0 ：本科专业不影响读 MBA 期间所选课程（相互独立）； H_1 ：（相互不独立）

第二步：构建统计量：

利用公式： $E_{ij} = \frac{N_{i\cdot} \times N_{\cdot j}}{N}$ 先计算期望频次分布，如上表括号中数据。

由 $\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$: $\chi^2((r-1)(c-1))$ 得

$$\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(n_{ij} - E_{ij})^2}{E_{ij}} = \frac{(31-24.08)^2}{24.08} + \frac{(13-17.91)^2}{17.91} + \dots = 14.7019 \quad ((2) p=0.0227)$$

第三步：求临界值： $\chi_{0.05}^2(6) = 12.5916$

第四步：决策：因为 $\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$ 大于 $\chi_{0.05}^2(6)$ ，所以拒绝 H_0

第五步：结论：本科专业影响读 MBA 期间所选课程

$$9.5 \quad \varphi = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{17.63}{527}}; \quad c = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{17.63}{17.63 + 527}};$$

$$v = \sqrt{\frac{\chi^2}{n \times \min[(r-1)(c-1)]}} = \sqrt{\frac{17.63}{527 \times 2}}$$

第 10 章 方差分析

10.1

方差分析

差异源	SS	df	MS	F	P-value	F crit
组间		2				
组内	598	9				
总计		11				

因为 $F=4.6574 < \text{临界值}$ ，所以不能拒绝 H_0 ，有显著差异

方差分析

差异源	SS	df	MS	F	P-value	F crit
组间		4				
组内		18				
总计		22				

因为 $F=15.52337 > \text{临界值}$ ，所以拒绝 H_0 ，不相等

10. 3 一家牛奶公司有 4 台机器装填牛奶，每桶的容量为 4L。下面是从 4 台机器中抽取的样本数据：

机器 1	机器 2	机器 3	机器 4
4.02			
			4.01

取显著性水平 $\alpha =$ ，检验 4 台机器的装填量是否相同？

解：ANOVA

每桶容量 (L)

	平方和	df	均方	F	显著性
组间	0.007	3			
组内		15			
总数		18			

不相同。

方差分析

差异源	SS	df	MS	F	P-value	F crit
组间		2				
组内		15				
总计		17				

有显著差异

方差分析

差异源	SS	df	MS	F	P-value	F crit
组间		2				
组内		12				
总计	832	14				

有显著差异
LSD 方法略

方差分析

差异源	SS	df	MS	F	P-value	F crit
组间		2				
组内		23				
总计		25				

有显著差异

10. 7 某企业准备用三种方法组装一种新的产品，为确定哪种方法每小时生产的产品数量最多，随机抽取了 30 名工人，并指定每个人使用其中的一种方法。通过对每个工人生产的产品数进行方差分析得到下面的结果

方差分析表

差异源	SS	df	MS	F	P-value	F crit
组间	420	2	210			
组内	3836	27		—	—	—
总计	4256	29	—	—	—	—

要求：

(1)完成上面的方差分析表。

(2)假设显著性水平，检验三种方法组装的产品数量之间是否有显著差异？

解：（2）>，没有显著差异。

方差分析

差异源	SS	df	MS	F	P-value	F crit
行		4				
列		2				
误差		8				
总计		14				

(1) 有显著影响

(2) 有显著影响

10.9 有 5 种不同品种的种子和 4 种不同的施肥方案，在 20 块同样面积的土地上，分别采用 5 种种子和 4 种施肥方案搭配进行试验，取得的收获量数据如下表：

品种	施肥方案			
	1	2	3	4
1	12. 0	9. 5	10. 4	9. 7
2	13. 7	11. 5	12. 4	9. 6
3	14. 3	12. 3	11. 4	11. 1
4	14. 2	14. 0	12. 5	12. 0
5		14. 0	13. 1	

检验种子的不同品种对收获量的影响是否有显著差异?不同的施肥方案对收获量的影响是否有显著差异 (a=0.05)?

方差分析

差异源	SS	df	MS	F	P-value	F crit
行		4				
列		3				
误差		12				

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/606004011042010220>