

【强化】2024 年复旦大学 025200 应用统计《432 统计学》考研强化模考 5 套卷

主编：掌心博阅电子书

特别说明

本书严格按照该科目今年考研专业课真题题型、试题数量和考试难度出题，结合本专业考研大纲整理编写，由考研学长严格审核校对。其内容涵盖了本科目考研常考试题及重点试题，针对性强，是报考本校该科目考研专业课复习的重要资料。

版权声明

青岛华研教育旗下掌心博阅电子书依法对本书享有专有著作权，同时我们尊重知识产权，对本电子书部分内容参考和引用的市面上已出版或发行图书及来自互联网等资料的文字、图片、表格数据等资料，均要求注明作者和来源。但由于各种原因，如资料引用时未能联系上作者或者无法确认内容来源等，因而有部分未注明作者或来源，在此对原作者或权利人表示感谢。若使用过程中对本书有任何异议请直接联系我们，我们会在第一时间与您沟通处理。

因编撰此电子书属于首次，加之作者水平和时间所限，书中错漏之处在所难免，恳切希望广大考生读者批评指正。

目录

【强化】2024 年复旦大学 025200 应用统计《432 统计学》考研强化模考 5 套卷（一）	4
【强化】2024 年复旦大学 025200 应用统计《432 统计学》考研强化模考 5 套卷（二）	11
【强化】2024 年复旦大学 025200 应用统计《432 统计学》考研强化模考 5 套卷（三）	19
【强化】2024 年复旦大学 025200 应用统计《432 统计学》考研强化模考 5 套卷（四）	28
【强化】2024 年复旦大学 025200 应用统计《432 统计学》考研强化模考 5 套卷（五）	35

【强化】2024 年复旦大学 025200 应用统计《432 统计学》考研强化模考 5 套卷（一）

说明：本书按照考试大纲、历年真题、指定参考书等公开信息潜心整理编写，由学长严格审核校对，仅供考研备考使用，与目标学校及研究生院官方无关，如有侵权请联系我们立即处理。

一、单项选择题

1. 说明统计表名称的词句，在统计表中称为_____。

- A. 横行标题
- B. 纵栏标题
- C. 总标题
- D. 主体栏

【答案】C

2. 在对总现象进行分析的基础上，有意识地选取若干有代表性的调查单位进行调查，这种调查方式是_____。

- A. 普查
- B. 典型调查
- C. 抽样调查
- D. 重点调查

【答案】B

3. 属于水资源的是_____。

- A. 工农业用水
- B. 用于产生电能或动能的水资源
- C. 农田灌溉用水
- D. 发展航运的水资源

【答案】B

4. 根据组（群）间方差的资料计算抽样平均误差的抽样组织方式是_____。

- A. 纯随机抽样
- B. 机械抽样
- C. 类型抽样
- D. 整群抽样

【答案】D

5. 调查时间是指_____。

- A. 进行调查工作的期限
- B. 调查资料所属的时间
- C. 调查工作登记的时间
- D. 调查资料的报送时间

【答案】B

6. 能够事先加以计算和控制的误差是_____。

- A. 登记性误差
- B. 代表性误差
- C. 系统性误差
- D. 抽样误差

【答案】D

7. 在一般情况下_____。

- A. 劳动力资源总数 > 社会劳动者总数
- B. 劳动力资源总数 < 社会劳动者总数
- C. 劳动力资源总数 = 社会劳动者总数
- D. B 和 C

【答案】D

8. 机床厂某日从两台机器所加工的同一种零件中，分别抽取两个样本，检验两台机床的加工精度是否相同，则提出假设_____。

- A. $H_0: \mu_1 = \mu_2; H_1: \mu_1 \neq \mu_2$
- B. $H_0: \sigma_1^2 = \sigma_2^2; H_1: \sigma_1^2 \neq \sigma_2^2$
- C. $H_0: \mu_1 \leq \mu_2; H_1: \mu_1 > \mu_2$
- D. $H_0: \sigma_1^2 \leq \sigma_2^2; H_1: \sigma_1^2 > \sigma_2^2$

【答案】B

二、简答题

9. 什么是统计综合分析？其特点如何？有哪几种？

【答案】统计综合分析，就是指根据分析研究的目的，在正确的实质性科学指导下运用统计方法，以统计资料为依据，结合具体情况，从定性与定量结合上，对客观事物进行科学的分析和综合的研究，揭示其本质和规律性，提出解决矛盾的办法的一种逻辑思维活动。它是整个统计工作的一个重要阶段，是统计工作的最终环节，其好坏直接影响统计的质量。在统计实践中，只有开展统计综合分析，才能更好地发挥统计的作用，为各级领导和有关方面的公众提供有数据、有情况、有分析的资料，为制订计划和规划，实行宏观调控，决定有关方针、政策提供科学依据。

统计综合分析主要有以下特点：

- (1) 以统计数据为基础，定量分析与定性分析相结合；
- (2) 统计综合分析的目的在于提出办法解决问题；
- (3) 综合运用多种分析方法。

根据统计综合分析的任务和研究重点不同，其形式综合归纳起来，主要有以下四种：

- (1) 专题性的分析；
- (2) 总结性的分析；
- (3) 进度性的分析；
- (4) 预测决策性的分析。

10. 什么是概率？它的三种定义分别是哪三个角度给出的？

【答案】概率是指随机事件发生的可能性的，其数值介于 0 到 1 之间。概率有三种不同的定义，即古典概率、统计概率和主观概率。古典概率是通过演绎方法计算得到的概率；统计概率是通过统计试验的大量数据计算而得到的频率，是概率的近似值；主观概率是通过经验和智慧主观估计的概率。

11. 统计的三种含义是什么？它们之间有何联系？

【答案】统计的三种含义是：（1）统计工作即统计实践，它是对社会自然现象客观存在的现实数量方面进行搜集、整理和分析的过程。（2）统计资料是指统计实践活动过程所取得的各项数字资料以及与之相关的其他实际资料的总称。（3）统计学是一门系统地论述统计理论和方法的科学，是对统计实践活动的经验总结和理论概括，是研究社会经济和自然现象的数量方面的方法论科学。

三者的联系：（1）统计工作与统计资料是过程与结果的关系；（2）统计工作与统计学是实践与理论的关系。

12. 贝叶斯决策有哪些优点？

【答案】贝叶斯决策的优点是：（1）一般的决策方法大多用的是不完备的信息或主观概率，而贝叶斯决策能对信息的价值或是否需要采集新的信息作出科学的判断；（2）它能对调查结果的可能性加以数量化的评价，而不是像一般的决策方法那样，对调查结果或者是完全相信，或者是完全不相信；（3）如果说任何调查结果都不可能完全准确，先验知识或主观概率也不是完全可以相信的，那么贝叶斯决策则巧妙地将这两种信息有机的结合起来了；（4）它可以在决策过程中根据具体情况不断使用，使决策逐步完善和更加科学。

13. 怎样理解统计的三种含义？

【答案】统计包含统计工作、统计资料和统计学三个方面的含义。一是统计工作，即统计实践，是对社会经济现象客观存在的现实数量方面进行搜集、整理和分析预测等活动的总称。二是统计资料（统计信息），是统计工作过程中所取得的各项数字资料以及与之相关的其他实际资料的总称。三是统计学，是系统论述统计理论和方法的科学，是长期统计工作实践的经验总结和理论概括。

14. 点估计与区间估计有什么区别？

【答案】点估计是以抽样得到的样本指标作为总体指标的估计量，并以样本指标的实际值直接作为总体未知参数的估计值的一种推断方法；区间估计则是根据抽样指标和抽样平均误差推断总体指标的可能范围，它既说明推断的准确程度，同时也表明了推断结果的可靠程度。可见，点估计所推断的总体指标是一个确定的数值，而区间估计所推断的总体指标是一个数值域，这个值域受样本指标、极限误差和样本单位数等因素的影响。

三、计算题

15. 甲、乙两个企业生产共种产品的单位成本和总成本资料如表所示。

产品名称	单位成本/元	总成本/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

表-甲、乙两个企业生产三种产品成本和总成本

比较哪个企业的总平均成本高, 并分析其原因。

【答案】设 x 为单位成本, f 为产量, xf 为总成本, 因此, 各企业的平均成本为

$$\text{甲企业平均成本} = \frac{\sum_{i=1}^3 x_i f_i}{\sum_{i=1}^3 \frac{1}{x_i} x_i f_i} = \frac{\frac{2100}{15} + \frac{3000}{20} + \frac{1500}{30}}{\frac{2100}{15} + \frac{3000}{20} + \frac{1500}{30}} = 19.41 \text{ (元)}$$

$$\text{乙企业平均成本} = \frac{\sum_{i=1}^3 x_i f_i}{\sum_{i=1}^3 \frac{1}{x_i} x_i f_i} = \frac{\frac{3255}{15} + \frac{1500}{20} + \frac{1500}{30}}{\frac{3255}{15} + \frac{1500}{20} + \frac{1500}{30}} = 18.29 \text{ (元)}$$

对比可见, 甲企业的总平均成本较高。原因是: 在甲、乙企业的产品结构中, 单位成本较低的 A 产品在乙企业中占的比例为 52.04%, 而甲企业仅占 31.82%, 所以使乙企业的总平均成本降低了。

16. 为了确定灯泡的使用寿命 (单位: 小时), 在一批灯泡中随机抽取 100 只进行测试, 所得结果如下:

700 716 728 719 685 709 691 684 705 718
706 715 712 722 691 708 690 692 707 701
708 729 694 681 695 685 706 661 735 665
668 710 693 697 674 658 698 666 696 698
706 692 691 747 699 682 698 700 710 722
694 690 736 689 696 651 673 749 708 727
688 689 683 685 702 741 698 713 676 702
701 671 718 707 683 717 733 712 683 692
693 697 664 681 721 720 677 679 695 691
713 699 725 726 704 729 703 696 717 688

(1) 利用计算机对上述数据进行排序;

(2) 以组距为 10 进行等距分组, 整理成频数分布表, 并绘制直方图;

(3) 制作茎叶图, 并与直方图作比较。

【答案】(1) 把原始数据输入 Excel 中的 A2:A101, 再把原始数据复制到 B2:B101 中, 选中 B2:B101, 点击 Excel 中“排序和筛选”, 选择“升序”, 即可把数据进行排序, 可看出数据最小值是 651, 最大值是 749。

(2) 全距 = 749 - 651 = 98, 等距分组分成 10 组, 利用 Excel 中的 FREQUENCY 函数功能编制频数分布表, 见下表。

按使用寿命分组/小时	灯泡个数/只	频率/%
650~660	2	2
660~670	5	5
670~680	6	6
680~690	16	16
690~700	26	26
700~710	18	18
710~720	12	12
720~730	9	9
730~740	3	3
740~750	3	3
合计	100	100

表-100 只灯泡使用寿命频数分布表

利用 Excel 中描述统计的直方图功能绘制直方图，如下图 1 所示。

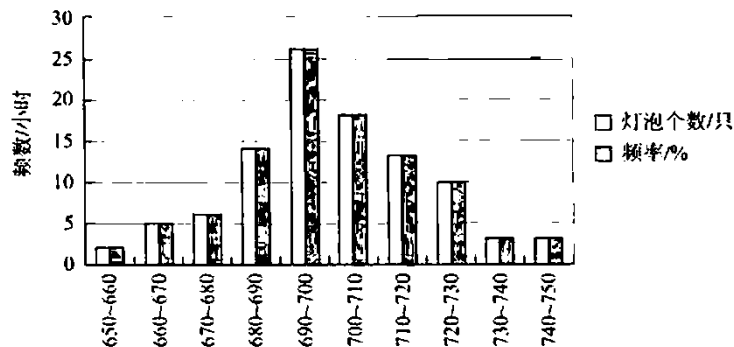


图 1-灯泡使用寿命分布直方图

(3) 利用 SPSS 制作茎叶图，见图 2。

茎	树叶	数据个数
65	18	2
66	14568	5
67	134679	6
68	11233345558899	11
69	00111122233445566677888899	26
70	001122345666778889	18
71	0022335677889	13
72	0122567899	10
73	356	3
74	179	3

图 2-灯泡使用寿命茎叶图

利用 Excel 制作的直方图中,每组的上限包含在本组之中,而茎叶图中,每组的上限不在本组内。本题中直方图和茎叶图的分布状况是相似的,茎叶图类似于横置的直方图。由此可以看出、灯泡的使用寿命服从正态分布。

17. 已知某地区 2007 年的农副产品收购总额为 360 亿元, 2008 年比上年的收购总额增长 12%, 农副产品收购价格总指数为 105%。试考虑, 2008 年与 2007 年对比:

- (1) 农民因销售农副产品共增加多少收入?
- (2) 农副产品收购量增加了百分之几? 农民因此增加了多少收入?
- (3) 由于农副产品收购价格提高 5% 农民增加了多少收入?
- (4) 验证以上三方面分析结论能否保持协调一致。

【答案】(1) 农民因销售农副产品共增加收入 $= 360 \times 12\% = 43.2$ (亿元)。

(2) 农副产品收购额指数 $= 112\%$, 而收购价格总指数 $= 105\%$ 。根据指数体系: 农副产品收购量指数 $= 112\% \div 105\% = 106.67\%$, 说明收购量增加了 6.67%。由于 2007 年农副产品的收购总额为 360 亿元, 增加的农副产品收购额为 $= 360 \times 6.67\% = 24.012$ (亿元)。

(3) 根据指数体系, 农副产品收购总额增加了 43.2 亿元, 由于产量增加而增加的收购额为 24.012 亿元, 因此由于价格上升而增加的收购额为 $43.2 - 24.012 = 19.2$ (亿元)。

(4) 根据指数体系, 有: $106.67\% \times 105\% = 112\%$; 24 亿元 + 19.2 亿元 = 43.2 亿元说明以上三方面的分析结论能协调一致。

18. 某校工商管理系学生体重资料如下表:

按体重分组(公斤)	学生人数
52 以下	24
52~55	25
55~58	38
58~61	21
61 以上	17
合计	125

表

试计算该系学生体重的算术平均数、中位数和众数。

【答案】算术平均数:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{50.5 \times 24 + 53.5 \times 25 + 56.5 \times 38 + 59.5 \times 21 + 62.5 \times 17}{125} = 56.068$$

中位数:

$$M_e = L + \frac{\frac{\sum f}{2} - S_{m-1}}{f_m} \cdot i = 55 + \frac{\frac{125}{2} - 49}{38} \times 3 = 56.07$$

众数:

$$M_o = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot i = 55 + \frac{13}{13 + 17} \times 3 = 56.3$$

19. 对成人组和幼儿组共 500 人身高资料分组如下表:

成人组		幼儿组	
按身高分组/cm	人数	按身高分组/cm	人数
150 ~ 155	30	70 ~ 75	20
155 ~ 160	120	75 ~ 80	80
160 ~ 165	90	80 ~ 85	40
165 ~ 170	40	85 ~ 90	30
170 以上	20	90 以上	30
合计	300	合计	200

表

(1) 计算成年组和幼儿组身高的平均数、标准差和标准差系数。

(2) 成年组和幼儿组平均身高的代表性哪个大? 为什么?

【答案】 (1) $\bar{x}_{成} = \frac{\sum xf}{\sum f} = \frac{152.5 \times 30 + 157.5 \times 120 + 162.5 \times 90 + 17.5 \times 40 + 172.5 \times 20}{300}$

$= 160.83$

$\bar{x}_{幼} = \frac{\sum xf}{\sum f} = \frac{72.5 \times 20 + 77.5 \times 80 + 82.5 \times 40 + 87.5 \times 30 + 92.5 \times 30}{200} = 81.75$

$\sigma_{成} = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}} = 5.22$

$\sigma_{幼} = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}} = 6.18$

$V_{成} = \frac{\sigma}{\bar{x}} = 3.25\% \quad V_{幼} = \frac{\sigma}{\bar{x}} = 7.56\%$

(2) 成年组平均数代表性大。

20. 已知随机变量 X 的分布函数为

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{x}{4}, & 0 < x \leq 4 \\ 1, & x > 4 \end{cases}$$

求 E (x) 和 Var (X) 。

【答案】 X 的密度函数 $f(x)$ 为:

$f(x) = \frac{1}{4} (0 < x \leq 4)$

$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^4 \frac{x}{4}dx = 2$

$Var(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x)dx = \int_0^4 \frac{(x-2)^2}{4}dx \approx 1.33$

【强化】2024 年复旦大学 025200 应用统计《432 统计学》考研强化模考 5 套卷（二）

说明：本书按照考试大纲、历年真题、指定参考书等公开信息潜心整理编写，由学长严格审核校对，仅供考研备考使用，与目标学校及研究生院官方无关，如有侵权请联系我们立即处理。

一、单项选择题

1. 根据抽样测定 100 名 4 岁男孩身体发育情况的资料，平均身高为 95cm，标准差为 4cm。用_____概率可确信 4 岁男孩平均身高在 93.8~96.2cm 之间。

- A.68.27%
- B.90%
- C.95.45%
- D.99.74%

【答案】D

2. 对于越高越好的现象按连续型变量分组，如第一组为 75 以下，第二组为 75~85,第三组为 85~95,第四组为 95 以上，则数据_____。

- A.85 在第三组
- B.75 在第一组
- C.95 在第三组
- D.85 在第二组

【答案】A

3. 属于不同总体的不同性质指标对比的相对数是_____。

- A.动态相对数
- B.比较相对数
- C.结构相对数
- D.强度相对数

【答案】D

4. 根据平均指标的确定方法和依据资料不同主要有 5 种，其中_____。

- A.中位数和算术平均数是位置平均数
- B.众数和调和平均数是位置平均数
- C.算术平均数和几何平均数是位置平均数
- D.中位数和众数是位置平均数

【答案】D

5. 若要检验两个独立样本是否取自均值相等的总体，则可采用_____。

- A.独立性检验
- B.成对比较检验
- C.曼-惠特尼 U 检验
- D.游程检验

【答案】C

6. 用“水平法”检查长期计划的执行情况适用于_____。

- A.规定计划期初应达到的水平
- B.规定计划期内某一期应达到的水平
- C.规定计划期末应达到的水平
- D.规定整个计划期累计应达到的水平

【答案】C

7. 估计量的含义是指_____。

- A.用来估计总体参数的统计量的名称
- B.用来估计总体参数的统计量的具体数值
- C.总体参数的名称
- D.总体参数的具体数值

【答案】A

8. 算术平均数的分子和分母是_____。

- A.两个有联系的而性质不同的总体总量
- B.分子是总体单位总量，分母是总体标志总量
- C.分子是总体标志总量，分母是另一总体单位总量
- D.同一总体的标志总量和总体单位总量

【答案】D

二、简答题

9. 简述方差分析的基本步骤。

【答案】方差分析是对总体均值是否相等的检验，所以方差分析的基本步骤遵循假设检验的基本步骤提出假设、构造检验统计量、统计决策。

以单因素方差分析为例：

第一步，提出假设：

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k, H_1: \mu_1, \mu_2, \cdots, \mu_k \text{ 不全相等。}$$

第二步，构造检验统计量。

构造检验统计量的关键是总平方和的分解，其步骤为：计算总平方和、水平项误差项平方和、误差项平方和；将各平方和除以自由度，得到组间均方、组内均方；组间均方除以组内均方得到检验统计量 F_v 。

第三步，进行统计决策。

将统计的值 F 与给定的显著性水平 α 的临界值 F_{α} 进行比较，做出对原假设 H_0 的决策：若 $F > F_{\alpha}(k-1, n-k)$ ，则拒绝原假设 H_0 ，表明均值之间的差异是显著的，所检验的因素对观测值有显著影响；若 $F < F_{\alpha}(k-1, n-k)$ ，则不能拒绝原假设 H_0 ，表明所检验的因素对观察值没有显著影响。

或者计算 P 值，用 P 值决策法做出拒绝或不拒绝原假设的决策 若 $P < \alpha$ ，则拒绝原假设 H_0 ；若 $P > \alpha$ ，不能拒绝原假设。

10. 统计指数按不同的标志可以有哪些分类方法？

【答案】统计指数的分类有如下形式：

(1) 按所反映现象的特征不同, 可分为数量指数和质量指数。数量指数反映现象的总规模、水平或工作总量的变化; 质量指数反映工作质量的变化情况。

(2) 按计算指数时所用的基期不同, 可分为定基指数和环比指数。定基指数的基期是固定不变的, 环比指数的基期是随着报告期的变化而变化的。

(3) 按所反映现象的范围不同, 可分为个体指数和总指数。个体指数是说明单个事物或现象在不同时期上的变动程度, 总指数是说明多种事物或现象在不同时期上的综合变动程度。

11. 回归系数与偏回归系数有什么不同?

【答案】在多元线性回归模型中, 偏回归系数表示当控制其他自变量不变的条件下, 某个自变量的单位变动对因变量均值的影响, 是該自变量对因变量的净影响, 二回归系数并没有派出其他自变量对因变量的影响, 即回归系数中既包含要考察的某个自变量对因变量的影响, 又包含其他自变量对因变量的影响。

12. 什么是“国内总产出”、“国内生产总值”和“国内生产净值”?试比较说明三者的价值构成和实物构成有何不同, 从核算和分析的角度看, 各有何优劣。

【答案】“国内总产出”是指一定时期内用货币计量的各单位、各部门生产的社会总产品的总量。“国内生产总值”是用货币计量的由本国常住单位在一定时期内所创造的社会最终产品的总量, 同时又是全社会各常住单位所创造的增加值之和。“国内生产净值”是指本国常住单位新创造的价值总和, 它等于国内生产总值减去全社会的固定资产折旧。

国内总产出的价值构成: ①生产资料转移价值, 包括劳动手段转移价值(即固定资产折旧)和劳动对象转移价值; ②活劳动新创造的价值, 包括必要劳动价值和剩余劳动价值。国内总产出的实物构成包括货物和服务两大类。货物通常是有实物形态, 能够确定所有权并满足人们需求的实物产品。服务产品是无形的, 用于满足使用者的需要。国内总产出指标能够表明全社会的生产活动的总规模, 并能用于对国民经济各部门间的技术经济联系进行投入产出分析, 但是, 它不能说明国民经济生产活动的最终成果。

国内生产总值是一国的各常住单位所创造的增加值的总和, 是常住单位在生产过程中把中间产品加工, 成最终产品所追加的价值, 从实物角度看, 国内生产总值是本国常住单位在一定时期内所生产出来的可供居民和社会最终消费和使用的产品总量。国内生产总值指标具有显著的优越性: 一是覆盖了国民经济各行各业; 二是避免了中间消耗的重复计算, 能确切地反映社会生产活动最终成果的价值量; 三是能够完整地反映社会最终产品的实物内容, 能够同时从生产, 收入和最终使用的角度进行观察和计算; 四是它具有国际可比性, 为世界各国广泛采用。因此, 国内生产总值核算在国民经济统计中处于核心地位。

国内生产净值是本国常住单位新创造的价值总和, 是国内生产总值减去全社会的固定资产折旧。国内生产净值是一个没有任何重复计算的社会新创造成果的指标。现实中, 由于固定资产折旧的计算方法较多, 每种计算方法都存在一定的假定性, 全社会在计算周定资产折旧时难以做到客观统一, 因此, 因内生产净值没有完整的实物形态与之对应, 这在某种程度上影响到该指标的客观性和稳定性。

13. 时间序列分解常用的模型有哪些? 简述乘法模型分解的基本步骤。

【答案】时间序列可以表示为以下四个因素的函数, 即

$$y_t = f(T_t, S_t, C_t, I_t)$$

时间序列分解的方法有很多, 较常用的模型有加法模型和乘法模型。

加法模型为

$$y_t = T_t + S_t + C_t + I_t$$

乘法模型为

$$y_t = T_t * S_t * C_t * I_t$$

乘法模型分解的基本步骤是：

- (1) 运用移动平均法剔除长期趋势和周期变化，得到序列 TC。然后再用按月（季）平均法求出季节指数 S。
- (2) 作散点图，选择适合的曲线模型拟合序列的长期趋势，得到长期趋势 T。
- (3) 计算周期因素 C。用序列 TC 除以 T 即可得到周期变动因素 C。
- (4) 将时间序列的 T、S、C 分解出来后，剩余的即为不规则变动，即

$$I = \frac{y}{TSC}$$

14. 反映总体集中趋势的指标有哪几种？

【答案】常用来表达数列集中趋势的测度有算术平均数、调和平均数、几何平均数、中位数和众数。这些测度在统计学中也称为平均指标或平均数，可以用来反映标志值的典型水平和标志值分布的中心位置或集中趋势。

算术平均数的定义公式为：算术平均数=总体标志总量/总体单位总量。调和平均数又称“倒数平均数”它是根据各标志值的倒数来计算的平均数，即各个标志值倒数的算术平均数的倒数。几何平均数是计算平均比率和平均速度最适用的一种方法，是所有标志值乘积后的 N 次方。三种平均数均有简单平均数和加权平均数两种。中位数是一种按其在数列中的特殊位置而决定的平均数。把总体各单位标志值按大小顺序排列后，处在中点位次的标志值就是中位数。众数是一种位置平均数。众数是总体单位中，标志值出现次数最多的那个数值。

三、计算题

15. 下表为 130 位同学统计学成绩分组资料：

考试成绩	学生数
50—55	2
55—60	4
60—65	10
65—70	14
70—75	20
75—80	28
80—85	26
85—90	16
90—95	7
95—100	3

求学生成绩分布的偏度系数和峰度系数。

【答案】 $\nu_2 = 92.25$

$$\begin{aligned} \nu_3 &= -220.154 \\ \nu_4 &= 23195.29 \\ \alpha &= \frac{\nu_3}{\nu_2^{3/2}} = -0.24847 \\ \beta &= \frac{\nu_4}{\nu_2^2} - 3 = -0.27437 \end{aligned}$$

16. 某家电制造公司准备购进一批5#电池, 现有 A、B、C 三个电池生产企业愿意供货, 为比较它们生产的电池质量, 从每个企业各随机抽取 5 只电池, 经试验得出其寿命 (小时) 数据见下表。

试验号	电池生产企业		
	A	B	C
1	50	32	15
2	50	28	12
3	43	30	38
4	40	34	48
5	39	26	40

表

试分析三个企业生产的电池的平均寿命之间有无显著差异 ($\alpha=0.05$)。如果有差异用 LSD 方法检验哪些企业之间有差异。

【答案】 $F=17.0684 > F_{0.05}(2, 12)=3.8853$ (或 $P\text{-value}=0.0003 < \alpha=0.05$) , 拒绝原假设, 表明电池的平均寿命之间有显著差异。

$|\bar{x}_A - \bar{x}_B| = |44.4 - 30| = 14.4 > LSD = 5.85$, 拒绝原假设, 则企业 A 与企业 B 电池的平均使用寿命之间有显著差异;

$|\bar{x}_A - \bar{x}_C| = |44.4 - 42.6| = 1.8 < LSD = 5.85$, 不能拒绝原假设, 没有证据表明企业 A 与企业 C 电池的平均使用寿命之间有显著差异;

$|\bar{x}_B - \bar{x}_C| = |30 - 42.6| = 12.6 > LSD = 5.85$, 拒绝原假设, 则企业 B 与企业 C 电池的平均使用寿命之间有显著差异。

17. 为了调查某地区人口总数, 在该地区 15000 户中以不重复抽样方法随机抽选 30 户作为样本, 每户人口数据如下:

5, 6, 3, 3, 2, 3, 3, 3, 4, 4

3, 2, 6, 4, 3, 5, 4, 5, 3, 3

4, 3, 3, 1, 2, 5, 3, 4, 2, 4

(1) 估计平均每户人口数及其标准差; (2) 推断该地区人口总数。

【答案】 (1) 平均每户人口数的估计值即为样本均值 $\bar{X} = \frac{\sum X}{n} = 3.4$

$$\text{样本方差 } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = 1.441$$

$$\text{样本均值的标准差 } \sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)} \approx \sqrt{\frac{S^2}{n}} = \sqrt{\frac{1.441}{30}} = 0.219$$

(2) 该地区人口总数为 $N\bar{X} = 15000 \times 3.4 = 51000$ (人)

18. 下表是根据《2000 年国际统计年鉴》中 1998 年世界 30 多个国家若干现代化指标数据计算的相关指标, 其中 x_j^* 为高收入国家 1997 年平均值。试对中国香港和美国的现代化综合水平作出比较。

	购买力 法人均 GNP (美元)	农业占 GDP 比 重(%)	第三产 业占 GDP 比重(%)	非农业 劳动力 占劳动 力比重 (%)	城市人 口占总 人口比 重(%)	平均预 期寿命 (岁)	成人识 字率(%)	大学生 占适龄 人口比 重(%)	千人拥 有医生 数(%)
$\bar{x}_j$	12190.21	8.07	56.82	0.83	70.59	72.89	94.87	37.03	2.41
σ_j	8084.72	6.71	13	0.16	19.45	5.43	5.49	21.08	1.3
x_j^*	22930	2	63	0.96	76	77	97.5	58	2.5
中国香港	20763	0.001	84.1	0.9969	100	78.6	92.9	28	1.3
美国	29240	0.018	62.2	0.9733	76.8	76.8	99	81	2.6

【答案】第一步, 采用相对化方法进行指标的无量纲化。以高收入国家 1997 年的平均值 x_j^* 为对比基准, 则指标评价值 $x_{ij}^* = \frac{x_{ij}}{x_j^*}$, 其中农业占 GDP 比重为逆指标, 采用公式 $x_{ij}^* = \frac{x_j^*}{x_{ij}}$ 。

第二步, 采用变异系数法确定各指标权重。分别计算第 j 个指标的变异系数 $V_j = \frac{\sigma_j}{\bar{x}_j}$, 再对各指标的变异系数进行归一化处理, 即得各指标的权重系数 $w_j = \frac{V_j}{\sum_{j=1}^m V_j}$, 计算结果列于下表:

	购买力 法人均 GNP (美元)	农业占 GDP 比 重(%)	第三产 业占 GDP 比重(%)	非农业 劳动力 占劳动 力比重 (%)	城市人 口占总 人口比 重(%)	平均预 期寿命 (岁)	成人识 字率(%)	大学生 占适龄 人口比 重(%)	千人拥 有医生 数(%)
中国香港	90.5	2000	133.5	103.8	131.6	102.1	95.3	48.3	52
美国	127.52	111.1	113.97	101.6	101.05	99.7	101.5	139.66	104
V_j	0.663	0.831	0.229	0.189	0.276	0.075	0.058	0.569	0.539
w_j	0.193	0.242	0.067	0.055	0.08	0.022	0.017	0.166	0.157

表一中国香港和美国的各项指标评价及各指标权重计算结果表

第三步, 求各指标的综合评价值

先用加权算术平均法对各指标评价值进行加权算术平均, 求综合评价值, 即加权算术平均综合值
$$= \frac{\sum_{j=1}^n x_j w_j}{\sum_{j=1}^n w_j} = \sum_{j=1}^n x_j w_j$$
。计算可得, 美国的现代化综合评价值为 116.24%, 略高于高收入国家的平均水平,

而香港的现代化综合评价值高达 546.7%, 远远高于高收入国家的平均水平, 其主要原因是香港的农业占 GDP 比重非常小, 该项指标的评价值非常高, 使得各指标评价值间得以线性补偿, 难以准确反映香港现代化水平的均衡发展状况。因此, 考虑用加权几何平均法, 即: 加权几何平均综合值
$$= \sqrt[n]{\prod_{j=1}^n x_j^{w_j}} = \prod_{j=1}^n x_j^{w_j}$$
。计算可得, 美国的现代化综合评价值为 115.03%, 香港的现代化综合评价值为 168.31%。加权几何平均综合评价法较之加权算术平均法更合理地反映了各项现代化指标的均衡发展状况, 比较可得, 香港的现代化综合水平高于美国, 并处于世界领先水平。

19. 对目标进行射击, 每次发一颗子弹, 直到击中 n 次为止, 设各次射击相互独立, 且每次射击时击中目标的概率为 p , 试求子弹的消耗量 X 的均值。

分析: 使用随机变量分解法, 可方便求解。

【答案】设: X_i = “第 $i-1$ 次击中后至第 i 次击中消耗子弹数” ($i=1, 2, \dots, n$);

X_1 = “第一次击中目标所消耗的子弹数”。

则有:

$$X = \sum_{i=1}^n X_i$$

$X_1, X_2, \dots, X_n$ 独立同分布, 且 $P\{X_i = k\} = (1-p)^{k-1} p = q^{k-1} p$ ($k=1, 2, \dots, n$)。

$$\begin{aligned} E(X_i) &= \sum_{k=1}^{\infty} k P(X_i = k) = p(1 + 2q + 3q^2 + \dots) * \\ &= p(q + q^2 + \dots)' = p\left(\frac{q}{1-q}\right)' = p \frac{1}{(1-q)^2} = \frac{1}{p} \end{aligned}$$

$$E(X) = \sum_{i=1}^n E(X_i) = \frac{n}{p}$$

故子弹的消耗量 X 的均值为 $\frac{n}{p}$

20. 一家建筑公司有向两个合同进行投标的机会, 第一个合同是建造一个废物处理工厂, 第二个合同是设计并建造一个城市蒸汽分配系统。这种情况下, 公司标明有五种方案。方案一是公司作为项目经理, 对外承包; 方案二是自己承担建设, 对外承包设计工作; 方案三是公司做设计, 其他对外承包; 方案四是公司包干全部工作; 方案五是联合另一家厂联合投标。损益矩阵如下:

单位: 万元

方 案	自然状态		
	订合同一	订合同二	两合同都订
方案一	-400	100	200
方案二	100	100	400
方案三	-200	150	600
方案四	0	200	500
方案五	100	300	200

【答案】这是不确定型的决策问题。首先, 观察损益矩阵表, 第一方案在三种自然状态下的收益都最小, 因而没有独立存在的价值, 应该去掉。现在对四种方案分别用五种方法决策。

(1) 用 “好中求好” 决策方法。

四种方案在各种自然状态下的最大收益分别为: 400, 600, 500, 300。选择最大的一个, 即 600, 所对应的为方案三。故方案三为最优方案。

(2) 用 “坏中求好” 决策方法。

四种方案在各种自然状态下的最小收益分别为: 100, -200, 0, 100。选择最大的一个, 即 100, 所对应的为方案二和方案五。故方案二和方案五为最优方案。

(3) 用 α 系数决策方法。

$$\begin{aligned} f(d_2) &= \alpha \max(100, 100, 400) + (1-\alpha) \min(100, 100, 400) = 300\alpha + 100 \\ f(d_3) &= \alpha \max(-200, 150, 600) + (1-\alpha) \min(-200, 150, 600) \\ &= 800\alpha - 200 \end{aligned}$$

$$f(d_4) = \alpha \max(0, 200, 500) + (1 - \alpha) \min(0, 200, 500) = 500\alpha$$

$$f(d_5) = \alpha \max(100, 300, 200) + (1 - \alpha) \min(100, 300, 200) = 200\alpha + 100$$

若方案二有最大收益, 需要 $\alpha \leq \frac{1}{2}$, 若方案三有最大收益, 需要 $\alpha \geq \frac{2}{3}$; 若方案四有最大收益, 需要 $\frac{1}{2} \leq \alpha \leq \frac{2}{3}$ 。若方案五有最大收益, 需要 $\alpha = 0$, 并且此时方案二有相同收益。

由此得出如下结论:

- ① $\alpha \leq \frac{1}{2}$ 时, 方案二为最优方案;
- ② $\alpha \geq \frac{2}{3}$ 时, 方案三为最优方案;
- ③ $\frac{1}{2} \leq \alpha \leq \frac{2}{3}$ 时, 方案四为最优方案;
- ④ $\alpha = 0$ 时, 方案五和方案二为最优方案。

(4) 按“最小的最大后悔值”决策方法决策。

三种自然状态下的最大收益值分别为:

$$\max(100, -200, 0, 100) = 100$$

$$\max(100, 150, 200, 300) = 300$$

$$\max(400, 600, 500, 200) = 600$$

各方案的最大后悔值分别为:

$$G(d_2) = \max(100 - 100, 300 - 100, 600 - 400) = 200$$

$$G(d_3) = \max(100 + 200, 300 - 150, 600 - 600) = 300$$

$$G(d_4) = \max(100 - 0, 300 - 200, 600 - 500) = 100$$

$$G(d_5) = \max(100 - 100, 300 - 300, 600 - 200) = 400$$

取最小的最大后悔值, 得到 100, 对应的为方案四。因此, “最小的最大后悔值”决策方法结果是方案四为最优方案。

(5) 等概率决策方法。

假设三种自然状态出现的概率相等, 均为 $1/3$, 计算四种方案的期望收益。

$$\text{方案二: } (-400 + 100 + 200) \times 1/3 = -100/3$$

$$\text{方案三: } (-200 + 150 + 600) \times 1/3 = 550/3$$

$$\text{方案四: } (0 + 200 + 500) \times 1/3 = 700/3$$

$$\text{方案五: } (100 + 300 + 200) \times 1/3 = 200$$

比较可知, 方案四的期望收益最大, 因此, 方案四为最优方案。

【强化】2024 年复旦大学 025200 应用统计《432 统计学》考研强化模考 5 套卷 (三)

说明：本书按照考试大纲、历年真题、指定参考书等公开信息潜心整理编写，由学长严格审核校对，仅供考研备考使用，与目标学校及研究生院官方无关，如有侵权请联系我们立即处理。

一、单项选择题

1. 是非标志的标准差是_____。

A. $\sqrt{p(1-p)}$

B. $p(1-p)$

C. $\sqrt{(p-1)}$

D. $1-p$

【答案】A

2. 某商场第二季度商品零售额资料如下：

月 份	4 月	5 月	6 月
完成商品零售额(万元)	50	62	78
完成计划(%)	100	124	104

该商场第二季度平均完成计划为_____

A. $\frac{100\% + 124\% + 104\%}{3} = 108.6\%$

B. $\frac{50 + 62 + 78}{\frac{50}{100\%} + \frac{62}{124\%} + \frac{78}{104\%}} = 108.6\%$

C. $\frac{50}{100\%} + \frac{62}{124\%} + \frac{78}{104\%} = 92.1\%$

D. $\frac{50 \times 100\% + 62 \times 124\% + 78 \times 104\%}{50 + 62 + 78} = 109.5\%$

【答案】B

3. 下列_____可以作为离散型随机变量的概率分布。

A.

X	1	2	3
P	$-\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

B.

X	1	2	3
P	$\frac{1}{2}$	$-\frac{1}{3}$	$\frac{5}{6}$

C.

X	1	2	3
P	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

D.

X	1	2	3
P	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{5}{6}$

【答案】C

4. 抛一枚均匀硬币 100 次, 根据切贝谢夫不等式可知, 出现正面的次数在 40 ~ 60 次之间的概率 p 为_____。

- A. ≤ 0.25
- B. ≤ 0.75
- C. ≥ 0.75
- D. ≥ 0.25

【答案】C

5. 无偏估计是指_____。

- A. 样本统计量的值恰好等于待估的总体参数
- B. 所有可能样本估计值的数学期望等于待估总体参数
- C. 样本估计值围绕待估总体参数使其误差最小
- D. 样本量扩大到和总体单元相等时与总体参数一致

【答案】B

6. 中心极限定理可保证在大量观察下_____。

- A. 样本平均数趋近于总体平均数的趋势
- B. 样本方差趋近于总体方差的趋势
- C. 样本平均数分布趋近于正态分布的趋势
- D. 样本比例趋近于总体比例的趋势

【答案】C

7. 按调查登记的时间是否连续, 统计调查可分为连续调查和不连续调查。下述调查中属于连续调查的是_____。

- A. 每隔 10 年进行一次人口普查
- B. 对 2006 年大学毕业生分配状况的调查
- C. 对近年来物价变动情况进行一次摸底调查
- D. 按旬上报钢铁生产量

【答案】D

8. 评价估计量是否优良的标准一般有三个, 下面哪一个不是? _____。

- A. 无偏性
- B. 一致性
- C. 有效性
- D. 真实性

【答案】D

二、简答题

9. 统计学是怎样产生和发展的?

【答案】统计学的产生距今约有 300 多年的历史, 经过统计实践活动和统计理论的发展, 加以总结和概括后统计学成为一门科学。从统计学发展的历史来看, 曾经产生过较大影响的主要有以下几个流派 (1)

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/788104031126006100>