全国高等教育自学考试指定教材配套辅导丛书

# 高等数学（二）概率统计

# 应试指导及模拟试题

全国高等教育自学考试命题研究组 编

教材依据 唐国兴 主编

JI SUAN JI ZHUAN YE

计算机类 权威辅导

- 重点难点精讲
- 解题技巧分析
- 教材同步训练
- 考前实战演习

中国大地出版社

全国高等教育自学考试指定教材辅导

# 高等数学(二)概率统计

## 应试指导及模拟试题

全国高等教育自学考试命题研究组 编

教材依据 唐国兴 主编

中国大地出版社

# 内容简介

本书是由全国高等教育自学考试命题研究组专家编写的应试指导与题库，依据的是国家教育部考试中心于2002年开始，正式执行自学考试新计划下的新大纲、新教材。本书的试题经过精心设计，题型标准，应试导向准确，针对性强。考生只需用少量时间，通过实战练习，就能在较短时间内巩固所学知识，掌握要点，突破难点，把握重点，熟练掌握答题方法及技巧，适应考场氛围，顺利通过考试。

**图书在版编目(CIP)数据**

高等数学(二)概率统计应试指导及模拟试题/全国高等教育自学考试命题研究组编.—北京:中国大地出版社,2002.6
(全国高等教育自学考试辅导丛书)
ISBN 7-80097-498-7

Ⅰ.高… Ⅱ.全… Ⅲ.高等数学—高等教育—自学考试—自学参考资料 Ⅳ.C932.6

中国版本图书馆 CIP 数据核字(2002)第 012477 号

---

**责任编辑:**王慧军
**出版发行:**中国大地出版社
**社址邮编:**北京市海淀区大柳树路 19 号 100081
**电　　话:**(010)-62183493(发行部)
**传　　真:**(010)-62183493
**印　　刷:**北京市顺义康华福利印刷厂
**开　　本:**787×1092 1/16
**印　　张:**220
**字　　数:**4000 千字
**版　　次:**2002 年 6 月第 1 版
**印　　次:**2002 年 6 月第 1 次印刷
**印　　数:**1-3600 册
**书　　号:**ISBN 7-80097-498-7/TP·7
**总 定 价:**300.00 元

---

# 前　　言

国家教育部考试中心于2002年开始,正式执行自学考试新计划,同时使用新编的大纲和教材。

参加自考的学生渴求在考前能通过应试指导的帮助及模拟试题的演练,全面检查自己所学的知识是否扎实,考试大纲所要求的内容是否掌握,已经理解的知识能否完整、确切、简明地进行书面表述,并借此增强考生分析和解决实际问题的能力,帮助考生顺利通过考试。因此,为配合广大考生参加考试,并能顺利过关,我们利用多年积累的自考教学辅导资源和经验,全面系统地剖析了各门专业课程新大纲和教材的内容体系,组织编写了一套"全国高等教育自学考试应试指导及模拟试题"丛书,推向全国,以满足考生之急需,适应社会之需要。

本书在编写过程中,严格按照考试大纲的要求,以指定教材为基础,包括了所有考试的知识点,并着重突出重点和难点,充分体现了"在考察课程主体知识的同时,注重考查能力尤其是应用能力"的新的命题指导思想。

本书以习题为主,完全按照指定教材的结构,以章为单位。每章设"考试要求"、"知识重点"、"反馈测试题解"三部分。"考试要求"主要是考试大纲所规定的本章考核要求。"知识重点"主要是对该章的重点、要点内容的总结归纳;"反馈测试题解"则根据考试大纲对各知识点不同能力层次的要求,将知识及知识点下的细目以各种主要考试题型的形式编写,覆盖全部考核内容,适当突出重点章节,并且加大重点内容的覆盖密度,所有试题均附详细解答;书后附有模拟试卷及2001年度试题,供考生检验自己学习情况,建议在规定时间内完成。本书由马兰、耿玉菅主编。

欢迎广大读者对本丛书提出宝贵意见,以便我们今后工作中得以改进。

全国高等教育自学考试命题研究组

2002.6

# 目　录

# 高等数学(二)概率统计考试概述

统计是对数据进行搜集、整理、分析、推断以及从中引出结论的一门科学。统计由描述性统计和推断性统计两部分组成。本章扼要地叙述描述统计。

学习第一章,要求在收集数据之后,能熟练地将数据整理成表格、图形形式,并用一些重要的特征值如平均数、中位数、极差、标准差等来概括、显示其全貌的部分特性,从而为我们提取有用的信息。

概率论是数理统计的理论基础,第二章是概率论中的最基本的内容之一。学习本章,要求读者掌握随机事件、事件的概率、条件概率、独立性等基本概念,了解概率的 3 条基本性质和 2 种概率模型(古典概型、独立重复试验模型),并能熟悉和运用事件与概的有关运算。

学习第三章,要求读者正确理解和熟悉关于随机变量、概率密度、分布函数、数学期望与方差、随机变量的独立性等基本概念;对二项分布、普阿松分布和正态分布这三大重要分布之概率分布、期望分布与方差、有关的概率计算应能牢固掌握;对数学期望与方差的若干性质以及切比雪夫不等式的结论,了解其证明的逻辑推理,要求读者学会在各种场合运用这些结论;关于二维随机向量一节中若干定理以及推广到 $n$ 维随机向量的一些有关结论要求读者有基本或大致的了解,并能正确运用。

第四章概率论与数理统计这两部分内容起着承上启下的作用。数理统计方法的理论依据将主要在本章中给出。但是,其中绝大部分结论的证明由于超出本大纲的范围而不能给出。因此,学习本章,要求读者正确理解定理或结论的条件是什么,学会如何应用这些结论于数理统计方法中。另外,读者对总体与个体、样本、统计量等基本概念要有透彻的理解。

学习第五章,要求读者掌握参数点估计的几种常用方法,了解估计的无偏性、一致性和有效性是评价估计量是否优良的若干标准,正确理解置信区间的概念和含义,会计算一些不同情况的区间估计。

学习第六章,要求读者正确理解小概率原则、原假设与备择假设、拒绝区域与接受区域、两类错误等基本概念,掌握和熟悉运用假设检验中的几个统计量,并能熟悉地应用这几个常用的假设检验法。

学习第七章,要求读者理解产品的质量控制和抽样检验方法的基本思想和原理,它们是假设检验在实际中的具体应用;掌握工序质量控制中的平均值和极差控制方法、一次计数抽样检验方法;了解工序质量控制和计数抽样检验的其他方法。

学习第八章,要求读者理解和掌握回归与相关的概念;熟悉掌握一元线性回归、二元线性回归的参数估计及相关性检验;学会应用线性回归方法于经济管理和经济预测中去;熟悉可以化为线性形式的若干非线性类型以及了解方差分析的基本方法。

学习第九章,要求读者学会运用数理统计方法于经济预测与决策中去,掌握经济预测与决策的一些简单常用方法。

考试题设有单项选择题、简答题、计算题、证明题、综合应用题。

**(一) 单项选择题**

例 1. 反映数据组位置特征的是 ( )

A. 极差　　B. 众数　　C. 标准差　　D. 全距

答:B

例 2. 当( )时,必有 $AB = \varnothing$.

A. $P(A) = 0$ 或 $P(B) = 0$　　B. $P(AB) = 0$

C. $A$ 与 $B$ 独立　　D. $A$ 与 $B$ 对立

答:当 A 与 B 对立,即 $A = \bar{B}$ 时,必有 $AB = \varnothing$,所以,应选 D。

例 3. 当事件 $A$ 与 $B$ 满足 $B \subset A$ 时,$P(A - B) =$ ( )

A. $P(AB) - P(A)$　　B. $P(A) - P(B)$

C. $1 - P(AB)$　　D. $P(A\bar{B}) + P(AB)$

答:$P(A - B) = P(A) - P(AB) = P(A) - P(B)$,所以,应选 B。

例 4. 设 $X \sim N(3,2)$,则 $X$ 的密度函数 $p(x) =$ ( )

A. $\dfrac{1}{2\sqrt{\pi}} e^{-\left(\frac{x-3}{2}\right)^2}$　　$X \sim N(3,2)$　　B. $\dfrac{1}{2\sqrt{2\pi}} e^{-\frac{(x-3)^2}{4}}$

C. $\dfrac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{4}}$　　D. $\dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

答:由 $X \sim N(3,2)$ 知 $\mu = 3, \sigma = \sqrt{2}$,$X$ 的密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{2\sqrt{\pi}} e^{\frac{-(x-3)^2}{4}}$$

所以应选 A。

例 5. 设随机变量 $X$ 的密度函数为 $p(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$,则 ( )

A. $P\{X < x\} \geqslant P\{X > -x\}$　　B. $P\{|X| > 3\} < 0.01$

C. $P\{|X| \leqslant 1\} < \dfrac{2}{3}$　　D. $P\{X \geqslant 0\} = \dfrac{1}{\sqrt{2\pi}}$

答:由 $p(x)$ 的表达式知 $X \sim N(0,1)$,曲线 $y = p(x)$ 关于 $y$ 轴对称,应有

$$P\{X < x\} = P\{X > -x\}, -\infty < x < +\infty$$

所以,A 不正确. $P\{X \geqslant 0\} = \dfrac{1}{2}$,$D$ 也不正确,由

$$P\{|X| \leqslant 1\} = 0.6827 > \frac{2}{3}$$

知 C 也不正确,由

$$P\{|X| > 3\} = 1 - 0.9973 = 0.0027 < 0.01$$

知应选 B.

例 6. 设 $X \sim B(n,p)$,已知 $E(X) = 9, D(X) = 3.6$,则 $n =$ ( )

A. 5　　B. 15　　C. 10　　D. 20

答:由 $D(X) = E(X)(1-p)$ 知 $3.6 = 9(1-p)$,$1-p=0.4$,故 $p=0.6$,再由 $E(X)=np$ 知 $9=0.6n$,$n=15$,所以应选 $B$.

例 7.若随机变量 $X$ 的方差 $D(X)$ 存在,则 $P\{\frac{|X-EX|}{a}>1\}\leqslant$ ( )

A. $D(X)$　　B. 1　　C. $a^2D(X)$　　D. $\frac{D(X)}{a^2}$

答:首先,条件知 $a>0$,再用切比雪夫不等式知

$$P\{\frac{|X-EX|}{a}>1\}=P\{|X-EX|>a\}\leqslant\frac{D(X)}{a^2}$$

所以应选 D.

例 8.设 $X\sim N(\mu,\sigma^2)$,$X_1,X_2,\cdots,X_n$ 为 $X$ 的样本.记

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\ nS_n^2=(n-1)S^2=\sum_{i=1}^{n}(X_i-\bar{X})^2$$

则 ( )

A. $T=\frac{\bar{X}-\mu}{S_n/\sqrt{n-1}}\sim t(n)$　　B. $T=\frac{\bar{X}-\mu}{S_n/\sqrt{n}}\sim t(n-1)$

C. $T=\frac{\bar{X}-\mu}{S_n/\sqrt{n-1}}\sim t(n-1)$　　D. $T=\frac{\bar{X}-\mu}{S/\sqrt{n-1}}\sim t(n-1)$

答:应选 C.即有

$$T=\frac{\bar{X}-\mu}{S_n/\sqrt{n-1}}=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$$

例 9.设 $X\sim N(\mu,\sigma^2)$,$X_1,X_2,\cdots,X_n$ 为 $X$ 的样本,在作假设检验时,若( ),则采用统计量 $T=\frac{\bar{X}-\mu}{S/\sqrt{n}}$。

A. $\sigma^2$ 未知,检验 $H_0:\mu=\mu_0$　　B. $\sigma^2$ 已知,检验 $H_0:\mu=\mu_0$

C. $\mu$ 未知,检验 $H_0:\sigma^2=\sigma_0^2$　　D. $\mu$ 已知,检验 $H_0:\sigma^2=\sigma_0^2$

答:应选 A.当 $\sigma^2$ 未知,检验 $H_0:\mu=\mu_0$ 时,取统计量

$$T=\frac{\bar{X}-\mu}{S/\sqrt{n}}=\frac{\bar{X}-\mu}{S_n/\sqrt{n-1}}\sim t(n-1)$$

例 10.对总体 $X\sim N(\mu,\sigma^2)$ 的均值 $\mu$ 作区间估计时,得到置信概率为 95% 的置信区间,其含义是指这个区间 ( )

A. 平均含总体 95% 的值　　B. 平均含样本 95% 的值

C. 有 95% 的概率含 $\mu$ 的真值　　D. 有 95% 的概率含样本的值

答:本题中的置信区间是满足

$$P\{\underline{\theta}<\mu<\bar{\theta}\}=95\%$$

的随机区间$(\underline{\theta},\bar{\theta})$,它是样本的函数,它以 95% 的概率包含真值 $\mu$,所以,应选 C.

例 11.一元线性回归分析中,全差平方和 $l_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2$ 的自由度是 ( )

A. $n$　　B. $n-1$　　C. $n-2$　　D. 1

答:应选 B. 因为在平方和分解式 $l_{yy} = Q + U$ 中,$Q$ 和 $U$ 的自由度分别为 $n-2$ 和 1,所以 $l_{yy}$ 的自由度为 $n-1$. 或者由 $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$. 知 $l_{yy}$ 中 $n$ 个平方项满足一个约束方程,所以,减少了一个自由度.

例 12. 关于事件的独立性,结论(　　)成立.

A. 两两独立的事件组必相互独立

B. 当 $\bar{A}$ 与 $\bar{B}$ 独立时,$A$ 与 $B$ 必独立

C. $\varnothing$ 与 $\Omega$ 不独立

D. 若 $P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3)$,则 $A_1, A_2$ 和 $A_3$ 必相互独立

答:应选 B. A 与 B 独立 $\Leftrightarrow \bar{A}$ 与 $B$ 独立 $\Leftrightarrow A$ 与 $\bar{B}$ 独立 $\Leftrightarrow \bar{A}$ 与 $\bar{B}$ 独立,$\varnothing$ 与 $\Omega$ 必独立,即 $P(\varnothing\Omega) = P(\varnothing)P(\Omega) = 0$.

### (二) 简答题

例 13. 两台机床加工同一种零件,设零件直径都服从正态分布,现从两批产品中分别随机抽取6件和9件,测得直径的修正样本方差分别为 $S_1^2 = 0.537$ 和 $S_2^2 = 0.358$,试判断能否认为两者的方差无显著差异?(取 $\alpha = 0.05$,已知 $F_{0.975}(8,5) = 6.757$, $F_{0.975}(5,8) = 4.82$)

答:设 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, $X$ 与 $Y$ 独立,先求出左临界值

$$\lambda_1 = F_{0.025}(5,8) = \frac{1}{F_{0.975}(8,5)} = \frac{1}{6.757} = 0.184.$$

设 $H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2$,在 $H_0$ 成立的条件下,

$$F = \frac{S_1^2}{S_2^2} \sim F(5,8)$$

故由统计值

$$F = \frac{0.537}{0.358} = 1.5 \in (0.148, 4.82)$$

知可接受 $H_0$,即两批零件的直径方差无显著差异(显著水平为 0.05)

例 14. 设 $X \sim N(\mu, \sigma^2)$,其 $\sigma^2$ 已知而 $\mu$ 未知,在对 $\mu$ 作假设检验时,可得 $H_0: \mu = \mu_0$ 的显著水平为 $\alpha$ 的接受区间;在对 $\mu$ 作区间估计时,可得 $\mu$ 的置信水平为 $1-\alpha$ 的置信区间. 问这两个区间何时是一致的?

答:$H_0: \mu = \mu_0$ 的显著水平为 $\alpha$ 的接受区间为

$$\left(\mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$\mu$ 的置水平为 $1-\alpha$ 的置信区间为

$$\left(\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

因此,当 $H_0$ 中的 $\mu_0$ 取作样本均值 $\bar{x}$ 时,这两个区间是相同的.

### (三) 计算题

例 15. 设某批灯泡使用寿命在 2000 小时以上的概率为 0.4，试求出 5 个灯泡在使用 2000 小时后只有 2 个不坏的概率 $p$.

答：5 个灯泡在使用 2000 小时后，仍可继续使用的个数 $X \sim B(5, 0.4)$，因此，所求概率为

$$p = P\{X = 2\} = C_5^2 (0.4)^2 (0.6)^3$$
$$= 10 \times 0.16 \times 0.216 = 0.3456$$

例 16. 设某种元件的寿命 $X$ 服从指数分布 $E(\lambda)$，已知期望 $E(X) = 1000$(小时).

(1) 求 $P\{1000 \leqslant X \leqslant 1500\}$

(2) 若某元件已正常使用了 600 小时，求它还可继续正常使用 100 小时的概率.

答：$X$ 的分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & 若 x \geqslant 0 \\ 0 & 若 x < 0, \lambda = \dfrac{1}{1000} \end{cases}$$

(1) $P\{1000 \leqslant X \leqslant 1500\} = F(1500) - F(1000)$

$= (1 - e^{-1.5}) - (1 - e^{-1}) = e^{-1} - e^{-1.5} (\approx 14.5\%)$

(2) 对 $x \geqslant 0$ 有 $P\{X \geqslant x\} = 1 - P\{X < x\} = 1 - F(x) = e^{-\lambda x}$，于是所求概率为

$$P\{X \geqslant 700 \mid X \geqslant 600\} = \frac{P\{X \geqslant 700\}}{P\{X \geqslant 600\}} = \frac{e^{-0.7}}{e^{-0.6}}$$
$$= e^{-0.1} (\approx 90\%)$$

例 17. 设有 $A_1, A_2, A_3$ 三个口袋，已知 $A_1$ 中有 5 个黑球和 2 个白球；$A_2$ 有 7 个黑球和 3 个白球；$A_3$ 中有 6 个黑球和 4 个白球. 现随机地选择一个口袋，再从中取出一球，用 $B$ 表示取出的是白球. 求 $P(B)$ 和 $P(A_3 \mid B)$.

答：由全概率公式知

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2)$$
$$+ P(B \mid A_3)P(A_3)$$
$$= \frac{1}{3}\left(\frac{2}{7} + \frac{3}{10} + \frac{4}{10}\right) = \frac{20 + 49}{210} = \frac{69}{210} = \frac{23}{70}$$

$$P(A_3 \mid B) = \frac{P(B \mid A_3)P(A_3)}{P(B)} = \frac{\dfrac{4}{10} \times \dfrac{1}{3}}{\dfrac{23}{70}} = \frac{4 \times 7}{3 \times 23} = \frac{28}{69}$$

### (四) 证明题

例 18. 设 $\hat{\theta}$ 是未知参数 $\theta$ 的无偏估计，证明当 $D(\hat{\theta}) = 0$ 时，$\hat{\theta}^2$ 必是 $\theta^2$ 的无偏估计.

答：由 $E(\hat{\theta}) = \theta$ 和方差定义直接可得

$$0 = D(\hat{\theta}) = E(\hat{\theta}^2) - (E(\hat{\theta}))^2 = E(\hat{\theta}^2) - \theta^2$$

即 $E(\hat{\theta}^2) = \theta^2$，所以 $\theta^2$ 的估计 $\hat{\theta}^2$ 也是无偏的.

**（五）综合应用题**

例 19. 设 $X$ 和 $Y$ 是两个相互独立的随机变量，若

$$E(X)=\mu_1,D(X)=\sigma_1^2;E(Y)=\mu_2,D(Y)=\sigma_2^2$$

均为未知，试据如下样本数据

$n_1=500,\bar{x}=3000,S_1=400$

$n_2=1000,\bar{y}=4200,S_2=500$

其中，$S_1$ 和 $S_2$ 是修正样本标准差，求出

(1) $\mu_1-\mu_2$ 的置信区间（取 $\alpha=0.05$）；

(2) $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信区间（$\alpha=0.10$）.

（已知 $z=z_{0.975}=1.96;f_1=F_{0.05}(499,999)=\dfrac{1}{1.11},f_2=F_{0.95}(499,999)=1.13$）

答：(1) 因为是大样本，所以可用

$$Z=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}\sim N(0,1)$$

求出置信区间. 先求出

$$\bar{x}-\bar{y}=3000-4200=-1200$$

$$\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}=\sqrt{\frac{400^2}{500}+\frac{500^2}{1000}}=\sqrt{570}\approx 23.87$$

所以，$\mu_1-\mu_2$ 的置信水平为 0.95 的置信区间为

$$-1200\pm 1.96\times 23.87=-1200\pm 46.79$$

即 $(-1246.79,-1153.21)$

(2) 可用

$$F=\frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2}\sim F(n_1-1,n_2-1)$$

求出置信区间，先计算

$$\frac{S_1^2}{S_2^2}=(\frac{400}{500})^2=0.64$$

所以 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信水平为 0.90 的置信区间为

$$(\frac{0.64}{1.13},0.64\times 1.11)=(0.566,0.710)$$

# 第一章　描述统计

## ◎ 考试要求

1. 能在收集数据后熟练地将数据整理成表格、图形形式。

2. 能用一些重要特征值如平均数、中位数、极差等显示问题的部分特性。

3. 主要题目类型

① 基本知识的填空、选择、简答。

② 图形描述。

③ 求位置特征与变异特征。

## ◎ 知识重点

(一) 资料的整理与分组

由观察或其他方法所收集到的一批资料记为

$$x_1, x_2, \cdots, x_n$$

1. 一般,按其数值从小到大的顺序重新排列成 $x_1^* \leqslant x_2^* \leqslant \cdots \leqslant x_n^*$,由此找出最小值 $a = x_1^*$ 和最大值 $b = x_n^*$,并比原始数据获得了较多的认识.

2. 分组:在区间 $[a, b]$ 或比它略为放大一些的区间插入分点:$a_0$(取区间的左端点值),$a_1, a_2, \cdots, a_l$(取区间的右端点值). $a_0 < a_1 < a_2 < \cdots < a_{l-1} < a_l$. 作半开区间:$[a_0, a_1), [a_1, a_2) \cdots [a_{l-1}, a_l)$. 采用画记法计算落在每个区间中的数据频数. 这样就把全部数据分成了 $l$ 组,通常采用组距相等的方法,分多少组可以是任意的,$l$ 一般取 7 ~ 15.

(二) 直方图

直方图是一种条形图,它是描述分组资料最普遍的一种图形. 横轴 $x$ 标出取值的区间 $[a_0, a_1), [a_1, a_2), \cdots, [a_{l-1}, a_l)$,纵轴 $y$ 表示与频数成正比的各区间的条形高.

如果数据是一个连续型随机变量的观察值,则直方图顶部的阶梯图形是它的分布密度曲线的近似。

(三) 累积频率函数图

横轴 $x$ 表示取值的区间的中点值:

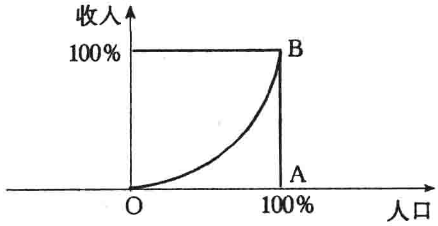
$$x_1 = \frac{a_1 + a_0}{2}, x_2 = \frac{a_2 + a_1}{2}, \cdots, x_l = \frac{a_l + a_{l-1}}{2}$$

纵轴 $y$ 表示区间的累加频率.

如果数据是一个连续随机变量的观察值,那么累积频率函数图是它的分布函数曲线的近似。

(四) 洛伦茨曲线

横轴 $x$ 表示人口的百分比,纵轴 $y$ 表示收入的百分比(收入最低的 10% 的人的总收入与全部人的总收入之比表示收入百分比).




要求读者了解洛伦茨曲线在经济学中的应用.

(五) 平均数

1. $N$ 个数值 $x_1, x_2, \cdots, x_N$,其算求平均数 $\bar{x}$ 为

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

2. 假如 $N$ 个数据分为 $l$ 个组(区间),各组的中点是 $x_1, x_2, \cdots, x_l$ 相应的频数是 $f_1, f_2, \cdots, f_l$,那么平均数 $x$ 为

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i f_i.$$

(六) 中位数

$N$ 个数按递增大小为 $x_1^*, x_2^*, \cdots, x_N^*$,其中位数 $M_d$ 为

$$M_d = \begin{cases} x^*_{\frac{N+1}{2}} & \text{当 } N \text{ 为奇数时} \\ \frac{1}{2}\left(x^*_{\frac{N}{2}} + x^*_{\frac{N}{2}+1}\right) & \text{当 } N \text{ 为偶数时} \end{cases}$$

(七) 众数

频数最大的那个数称为众数,记为 $M_0$.

(八) 极差

数据的最大数与最小数之差叫极差,记为 $R$,

$$R = x_N^* - x_1^*$$

(九) 平均偏差

$N$ 个数 $x_1, x_2, \cdots, x_N$ 其平均偏差 $M_D$ 为

$$M_D = \frac{1}{N}\sum_{i=1}^{N} |x_i - \bar{x}|$$

其中 $\bar{x}$ 为算术平均数.

(十) 方差和标准差

1. 偏差平方和的平均数称为方差:

$$S_N{}^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2$$

2. 方差的平方根称为标准差:

$$S_N=\sqrt{S_N{}^2}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

3. 一个计算公式:

$$\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-c)^2-(\bar{x}-c)^2,$$

其中 $c$ 为任一常数.

要求读者理解和掌握将位置特征和变异特征结合起来,它能画龙点睛地将数据变化的情况显示出来.

## ◎ 反馈测试题解

**一、单项选择题**

1. 一组数据 $x_1,x_2,\cdots,x_n$ 的标准差是指 ( )

A. $S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$　　B. $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

C. $S^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$　　D. $S=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}^2)}$

答:D

2. 描述数据组变异特征的量,一般选用 ( )

A. 平均数　B. 众数　C. 中位数　D. 极差

答:D

3. 若某种商品的一张调查问卷表有 3 个指标,各自的权重及得分如下:

| | $A$ | $B$ | $C$ |
|---|---|---|---|
| 权重(%) | 50 | 30 | 20 |
| 得分 | 9 | 8 | 10 |

则这张调查问卷表对该种商品的最后打分是 ( )

A. 9 分　B. 40 分　C. 8.9 分　D. 890 分

答:C

4. 设有 23,25,22,35,20,24 一组数据,那么这组数据的中位数是 ( )

A. 22　B. 23　C. 24　D. 23.5

答:D

5. 如果一组数据如下:1,2,3,4,4,4,4,4,5,6,7,8,10000,为了具有代表性和反映一般水平,应选取 ( )

A. 4　B. 5　C. 3　D. 773.23

答:A

6. 实际问题中,测量一物体的长度,反复测量 6 次,所得数据如下:

| 数据 | 4.8 | 4.9 | 5.0 |
| --- | --- | --- | --- |
| 次数 | 3 | 2 | 1 |

则该物体的长度计算公式应选用　　　　　　　　　　　　　　　　　　　　　　　　　　(　　)

A. $\frac{1}{6}(4.8+4.9+5.0)$　　　　　　　　B. $\frac{1}{3}(4.8+4.9+5.0)$

C. $\frac{1}{6}(3\times 4.8+2\times 4.9+1\times 5.0)$　　D. $\frac{1}{3}(3\times 4.8+2\times 4.9+1\times 5.0)$

答:C

7.下列数据中既可以反映是横截面数据,又可以反映是时间序列数据的是　　(　　)

A.洗衣粉镬生产线上每隔 30 分钟抽取 5 包洗衣粉测得的平均重量;

B.某月全厂各班组的出勤率;

C.第一百货公司每月的营业额;

D.1989 年 10 月 29 日个大中城市零售物价总指数(以上年同期价格为 100)。

答:D

8.有两种股票,它们在 5 天内的收盘价(元) 如:$A$:7.8,7.9,7.8,7.7,7.8;$B$:7.8,6.6,7.4,8.2,9 下列判断(　　)比较合理。

A. $A$ 股票平均价比 $B$ 股高　　　　B. $B$ 股票风险大,但可能获利较大

C. $B$ 股票会继续攀升　　　　　　D. $A$ 股票获利大

答:B

9.已知一组数据:1,2,2,2,3,3,3,3,4,5,5,6,7。则这组数据的众数是　　(　　)

A.2　　B.3　　C.6　　D.7

答:B

10.反映数据组 $x_1,x_2,\cdots,x_n$ 的变异特征,人们常用　　(　　)

A.中位数　　B.众数　　C.方差　　D.平均数

答:C

11.设有数据 $x_1,\cdots,x_{96}$ 按其数值大小的顺序重新排列或 $x_1^{\cdot}\leqslant x_2^{\cdot}\leqslant\cdots\leqslant x_{96}^{\cdot}$,则(　　)是数据的中位数。

A. $x_{48}$　　　　B. $x_{48}^{\cdot}$

C. $\frac{1}{2}(x_{47}^{\cdot}+x_{48}^{\cdot})$　　D. $\frac{1}{2}(x_{48}^{\cdot}+x_{49}^{\cdot})$

答:D

## 二、计算题

1.有 6 个观察值 100,120,200,210,220,230。为了简便地计算其平均数,甲两人分别想出了一种方法。

甲是:先计算前两个数的平均值 110,再计算后 4 个数的平均值 215,最后把它们平均一下得 162.5。

乙是:先把前 3 个数平均一下得 140,再将后 3 个数平均一下得 220,然后再把它们平均一下得 180。

请鉴定两种方法对不对?为什么?

答:因甲的方法为

$$\frac{1}{2}\left[\frac{1}{2}(x_1+x_2)+\frac{1}{4}(x_3+x_4+x_5+x_6)\right]$$
$$=\frac{1}{4}(x_1+x_2)+\frac{1}{8}(x_3+x_4+x_5+x_6)$$
$$=162.5$$

而6个数的平均数为$\frac{1}{6}(x_1+x_2+x_3+x_4+x_5+x_6)$。

可见甲的方法不正确。

乙的方法为

$$\frac{1}{2}\left[\frac{1}{3}(x_1+x_2+x_3)+\frac{1}{3}(x_4+x_5+x_6)\right]$$
$$=\frac{1}{6}(x_1+x_2+x_3+x_4+x_5+x_6)=180$$

所以,乙的方法正确。

2. 观察得10个数据,分成3组,第一组含3个数据,第二组含5个数据,第三组含2个数据,试填下表:

| 组 | 频数 | 相对频数(%) | 累计相对频数(%) |
|---|---|---|---|
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 合计 | | | |

答:频数:3,2,5;合计:10。

相对频数:30,20,50;合计:100。

累计相对频数:30,50,100。

3. 1982年我国不同规模家庭户数的抽样调查数据如下:

| 家庭规模 | 户数(万户) |
|---|---|
| 1人户 | 1749 |
| 2人户 | 2215 |
| 3人户 | 3533 |
| 4人户 | 4305 |
| 5人户 | 4040 |
| 6人户 | 2887 |
| 7人户 | 1751 |
| 8人及以上户 | 1530 |
| 总　计 | 22010 |