432-统计学

一、考查目标

全国硕士研究生入学统一考试应用统计硕士专业学位《统计学》考试是为高等院校和科研院所招收应用统计硕士生儿设置的具有选拔性质的考试科目。其目的是科学、公平、有效地测试考生是否具备攻读应用统计专业硕士所必须的基本素质、一般能力和培养潜能，以利用选拔具有发展潜力的优秀人才入学，为国家的经济建设培养具有良好职业道德、法制观念和国际视野、具有较强分析与解决实际问题能力的高层次、应用型、复合型的统计专业人才。考试要求是测试考生掌握数据处收集、处理和分析的一些基本统计方法。

具体来说。要求考生：

1．掌握数据收集和处理的基本分方法。

2．掌握数据分析的金发原理和方法。

3．掌握了基本的概率论知识。

4．具有运用统计方法分析数据和解释数据的基本能力。

二、考试形式和试卷结构

1．试卷满分及考试时间

试卷满分为 150 分，考试时间 180 分钟。

2．答题方式

答题方式为闭卷、笔试。允许使用计算器 (仅仅具备四则运算和开方运算功能的计算器)，但不得使用带有公式和文本存储功能的计算器。

3．试卷内容与题型结构

统计学 120 分，有以下三种题型：

单项选择题	25 题，每小题 2 分，共 50 分
简答题	3 题，每小题 10 分，共 30 分
计算与分析题	2 题，每小题 20 分，共 40 分

概率论 30 分，有以下三种题型：

单项选择题	5 题，每小题 2 分，共 10 分
简答题	1 题，每小题 10 分，共 10 分
计算与分析题	1 题，每小题 10 分，共 10 分

三、考查内容

1．统计学

调查的组织和实施。

概率抽样与非概率抽样。

数据的预处理。

用图表展示定性数据。

用图表展示定量数据。

用统计量描述数据的水平：平均数、中位数、分位数和众数。

用统计量描述数据的差异：极差、标准差、样本方差。

参数估计的基本原理。

一个总体和两个总体参数的区间估计。

样本量的确定。

假设检验的基本原理。

一个总体和两个总体参数的检验。

方差分析的基本原理。

单因子和双因子方差分析的实现和结果解释。

变量间的关系；相关关系和函数关系的差别。

一元线性回归的估计和检验。

用残差检验模型的假定。

多元线性回归模型。

多元线性回归的拟合优度和显著性检验；

多重共线性现象。

时间序列的组成要素。

时间序列的预测方法。

2．概率论

事件及关系和运算；

事件的概率；

条件概率和全概公式；

随机变量的定义；

离散型随机变量的分布列和分布函数；离散型均匀分布、二项分布和泊松分布；

连续型随机变量的概率密度函数和分布函数；均匀分布、正态分布和指数分布；

随机变量的期望与方差；

随机变量函数的期望与方差。

四、题型示例及参考答案

全国硕士研究生入学统一考试
应用统计硕士专业学位统计学试题

一、单项选择题（本题包括 1—30 题共 30 个小题，每小题 2 分，共 60 分。在每小题给出的四个选项中，只有一个符合题目要求，把所选项前的字母填在答题卡相应的序号内）。

选择题答题卡：

题号	1	2	3	4	5	6	7	8	9	10
答案										
题号	11	12	13	14	15	16	17	18	19	20
答案										
题号	21	22	23	24	25	26	27	28	29	30
答案										

1．为了调查某校学生的购书费用支出，从男生中抽取 60 名学生调查，从女生中抽取 40 名学生调查，这种抽样方法属于（　　）。

A．简单随机抽样　　B．整群抽样

C．系统抽样　　D．分层抽样

2．某班学生的平均成绩是 80 分，标准差是 10 分。如果已知该班学生的考试分数为对称分布，可以判断考试分数在 70 到 90 分之间的学生大约占（　　）。

A．95%　　B．89％　　C．68%　　D．99％

3．已知总体的均值为 50，标准差为 8，从该总体中随机抽取样本量为 64 的样本，则样本均值的数学期望和抽样分布的标准误差分别为（　　）。

A．50，8　　B．50，1　　C．50，4　　D．8，8

4．根据一个具体的样本求出的总体均值 95%的置信区间（　　）。

A．以 95%的概率包含总体均值

B．有 5%的可能性包含总体均值

C．绝对包含总体均值

D．绝对包含总体均值或绝对不包含总体均值

5．一项研究发现，2000 年新购买小汽车的人中有 40%是女性，在 2005 年所作的一项调查中，随机抽取 120 个新车主中有 57 人为女性，在 $\alpha = 0.05$ 的显著性水平下，检验 2005 年新车主中女性的比例是否有显著增加，建立的原假设和备择假设为（　　）。

A．$H_0 : \pi = 40\%$，$H_1 : \pi \neq 40\%$

B．$H_0 : \pi \geq 40\%$，$H_1 : \pi < 40\%$

C．$H_0 : \pi \leq 40\%$，$H_1 : \pi > 40\%$

D．$H_0 : \pi < 40\%$，$H_1 : \pi \geq 40\%$

6．在回归分析中，因变量的预测区间估计是指（　　）。

A. 对于自变量 x 的一个给定值 x_0，求出因变量 y 的平均值的区间

B. 对于自变量 x 的一个给定值 x_0，求出因变量 y 的个别值的区间

C. 对于因变量 y 的一个给定值 y_0，求出自变量 x 的平均值的区间

D. 对于因变量 y 的一个给定值 y_0，求出自变量 x 的平均值的区间

7. 在多元线性回归分析中，如果 F 检验表明线性关系显著，则意味着（　　）。

A. 在多个自变量中至少有一个自变量与因变量之间的线性相关系著

B. 所有的自变量与因变量之间的线性关系都显著

C. 在多个自变量中至少有一个自变量与因变量之间的线性关系不显著

D. 所有的自变量与因变量之间的线性关系都不显著

8. 如果时间序列的逐期观察值按一定的增长率增长或衰减，则适合的预测模型是（　　）。

A. 移动平均模型

B. 指数平滑模型

C. 线性模型

D. 指数模型

9. 雷达图的主要用途是（　　）。

A. 反映一个样本或总体的结构

B. 比较多个总体的构成

C. 反映一组数据的分布

D. 比较多个样本的相似性

10. 如果一组数据是对称分布的，则在平均数加减 2 个标准差之内的数据大约有（　　）。

A. 68%　　B. 90%　　C. 95%　　D. 99%

11. 从均值为 200、标准差为 50 的总体中，抽出 $n=100$ 的简单随机样本，用样本均值 $\bar{x}$ 估计总体均值 μ，则 $\bar{x}$ 的期望值和标准差分别为（　　）。

A. 200，5　　B. 200，20　　C. 200，0.5　　D. 200，25

12. 95%的置信水平是指（　　）。

A. 总体参数落在一个特定的样本所构造的区间内的概率为 95%

B. 总体参数落在一个特定的样本所构造的区间内的概率为 5%

C. 在用同样方法构造的总体参数的多个区间中，包含总体参数的区间比例为 95%

D. 在用同样方法构造的总体参数的多个区间中，包含总体参数的区间比例为 5%

13. 在假设检验中，如果所计算出的 P 值越小，说明检验的结果（　　）。

A. 越显著　　B. 越不显著　　C. 越真实　　D. 越不真实

14. 在下面的假定中，哪一个不属于方差分析中的假定（　　）。

A. 每个总体都服从正态分布　　B. 各总体的方差相等

C. 观测值是独立的　　D. 各总体的方差等于 0

15．在方差分析中，数据的误差是用平方和来表示的，其中组间平方和反映的是（ ）。

A．一个样本观测值之间误差的大小

B．全部观测值误差的大小

C．各个样本均值之间误差的大小

D．各个样本方差之间误差的大小

16．在多元线性回归分析中，t检验是用来检验（ ）。

A．总体线性关系的显著性

B．各回归系数的显著性

C．样本线性关系的显著性

D．$H_0:\beta_1=\beta_2=\cdots=\beta_k=0$

17．为研究食品的包装和销售地区对其销售量是否有影响，在三个不同地区中用三种不同包装方法进行销售，根据获得的销售量数据计算得到下面的方差分析表。表中“A”单元格和“B”单元格内的结果是（ ）。

差异源	SS	df	MS	F
行	22.22	2	11.11	A
列	955.56	2	477.78	B
误差	611.11	4	152.78	
总计	1588.89	8		

A．0.073 和 3.127　　B．0.023 和 43.005

C．13.752 和 0.320　　D．43.005 和 0.320

18．对某时间序列建立的预测方程为$\hat{Y}_t=100\times(0.8)^t$，这表明该时间序列各期的观察值（ ）。

A．每期增加 0.8　　B．每期减少 0.2

C．每期增长 80%　　D．每期减少 20%

19．进行多元线性回归时，如果回归模型中存在多重共线性，则（ ）。

A．整个回归模型的线性关系不显著

B．肯定有一个回归系数通不过显著性检验

C．肯定导致某个回归系数的符号与预期的相反

D．可能导致某些回归系数通不过显著性检验

20．如果时间序列不存在季节变动，则各期的季节指数应（ ）。

A．等于 0　　B．等于 1　　C．小于 0　　D．小于 1

21．一所中学的教务管理人员认为，中学生中吸烟的比例超过 30%，为检验这一说法是否属实，该教务管理人员抽取一个随机样本进行检验，建立的原假设和备择假设为$H_0:\pi\leq 30\%$，$H_1:\pi>30\%$。检验结果是没有拒绝原假设，这表明（ ）。

A. 有充分证据证明中学生中吸烟的比例小于 30%

B. 中学生中吸烟的比例小于等于 30%

C. 没有充分证据表明中学生中吸烟的超过 30%

D. 有充分证据证明中学生中吸烟的比例超过 30%

22. 某药品生产企业采用一种新的配方生产某种药品，并声称新配方药的疗效远好于旧的配方。为检验企业的说法是否属实，医药管理部门抽取一个样本进行检验。该检验的原假设所表达的是（　　）。

A. 新配方药的疗效有显著提高　　B. 新配方药的疗效有显著降低

C. 新配方药的疗效与旧药相比没有变化　　D. 新配方药的疗效不如旧药

23. 在回归分析中，残差平方和 SSE 反映了 y 的总变差中（　　）。

A. 由于 x 与 y 之间的线性关系引起的 y 的变化部分

B. 由于 x 与 y 之间的非线性关系引起的 y 的变化部分

C. 除了 x 对 y 的线性影响之外的其他因素对 y 变差的影响

D. 由于 y 的变化引起的 x 的误差

24. 在公务员的一次考试中，抽取 49 个应试者，得到的平均考试成绩为 81 分，标准差 $s=12$ 分。该项考试中所有应试者的平均考试成绩 95%的置信区间为（　　）。

A. 81±1.96　　B. 81±3.36　　C. 81±0.48　　D. 81±4.52

25. 某大学共有 5000 名本科学生，每月平均生活费支出是 500 元，标准差是 100 元。假定该校学生的生活费支出为对称分布，月生活费支出在 400 元至 600 元之间的学生人数大约为（　　）。

A. 4750 人　　B. 4950 人　　C. 4550 人　　D. 3400 人

26. 将一颗质地均匀的骰子（它是一种各面上分别标有点数 1，2，3，4，5，6 的正方体玩具）先后抛掷 3 次，至少出现一次 6 点向上的概率是(　　)

A. $\frac{5}{216}$　　B. $\frac{25}{216}$　　C. $\frac{31}{216}$　　D. $\frac{91}{216}$

27. 离散型随机变量 ξ 的分布列为 $\begin{pmatrix}0 & 1 & 2\\ 0.2 & a & b\end{pmatrix}$，其中 a,b 是未知数，如果已知 ξ 取 1 的概率和取 2 的概率相等，则 $a=$（ ）。

A. 0.2　　B. 0.3　　C. 0.4　　D. 0.5

28. 甲乙两人将进行一局象棋比赛，考虑事件 $A=\{\text{甲胜乙负}\}$，则 $\overline{A}$ 为（　　）。

A. 甲负乙胜　　B. 甲乙平局　　C. 甲负　　D. 甲负或平局

29. 对于随机变量 ξ，有 $D(10\xi)=10$，则 $D(\xi)=$（　　）。其中 $D(\xi)$ 表示随机变量 ξ 的方差。

A. 0.1　　B. 1　　C. 10　　D. 100

30．设函数 $f(x)$ 在区间 $[a,b]$ 上等于 0.5，在此区间之外等于 0，如果 $f(x)$ 可以作为某连续型随机变量的密度函数，则区间 $[a,b]$ 可以是（　　）。

A. $[0,0.5]$　　B. $[0.5,2.5]$　　C. $[1,1.5]$　　D. $[2,3]$

二、简要回答下列问题（本题包括 1—4 题共 4 个小题，每小题 10 分，共 40 分）。

1．简述假设检验中 P 值的含义。

2．已知甲乙两个地区的人均收入水平都是 5000 元。这个 5000 元对两个地区收入水平的代表性是否一样？请说明理由。

3．简述分解法预测的基本步骤。

4．正态分布的概率密度函数 $f(x)$ 有两个参数 μ 和 σ，请结合函数 $f(x)$ 的几何形状说明 μ 和 σ 的意义。

三、计算与分析题（本题包括 1—3 题共 3 个小题，第 1 小题和第 2 小题每题 20 分，第 3 小题 10 分，共 50 分）。

1．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 100 克。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（克）如下：

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

（1）确定该种食品平均重量 95%的置信区间。

（2）采用假设检验方法检验该批食品的重量是否符合标准要求？（$\alpha = 0.05$，写出检验的具体步骤）。

2．一家产品销售公司在 30 个地区设有销售分公司。为研究产品销售量(y)与该公司的销售价格（x1）、各地区的年人均收入(x2)、广告费用(x3)之间的关系，搜集到 30 个地区的有关数据。利用 Excel 得到下面的回归结果（$\alpha = 0.05$）：

方差分析表

变差来源	df	SS	MS	F	Significance F
回归			4008924.7		8.88341E-13
残差				—	—
总计	29	13458586.7	—	—	—

参数估计表

	Coefficients	标准误差	t Stat	P-value
Intercept	7589.1025	2445.0213	3.1039	0.00457
X Variable 1	-117.8861	31.8974	-3.6958	0.00103
X Variable 2	80.6107	14.7676	5.4586	0.00001
X Variable 3	0.5012	0.1259	3.9814	0.00049

（1）将方差分析表中的所缺数值补齐。

（2）写出销售量与销售价格、年人均收入、广告费用的多元线性回归方程，并解释各回归系数的意义。

（3）检验回归方程的线性关系是否显著?

（4）计算判定系数R^2，并解释它的实际意义。

（5）计算估计标准误差s_e，并解释它的实际意义。

3．用A,B,C三类不同元件连接成两个系统N_1和N_2。当元件A,B,C都正常工作时，系统N_1正常工作；当元件A正常工作且元件B,C中至少有一个正常工作时，系统N_2正常工作。已知元件A,B,C正常工作的概率依次为0.80,0.90,0.90，且某个元件是否正常工作与其他元件无关。分别求系统N_1和N_2正常工作的概率P_1和P_2。

参考答案

一、单项选择题

1. D；2. C；3. B；4. D；5. C；6. B；7. A；8. D；9. D；10. C；

11. A；12. C；13. A；14. D；15. C；16. B；17. A；18.D；19.D；20.B；

21.C；22.C；23.C；24.B；25.D；26.D；27.C；28.D；29.A；30.B。

二、简要回答题

1．（1）如果原假设H_0是正确的，所得到的样本结果会像实际观测结果那么极端或更极端的概率，称为P值。

（2）P值是指在总体数据中，得到该样本数据的概率。

（3）P值是假设检验中的另一个决策工具，对于给定的显著性水平α，若$P<\alpha$，则拒绝原假设。

2．这要看情况而定。如果两个地区收入的标准差接近相同时，可以认为5000元对两个地区收入水平的代表性接近相同。如果标准差有明显不同，则标准差小的，5000元对该地区收入水平的代表性就要好于标准差大的。

（1）确定并分离季节成分。计算季节指数，以确定时间序列中的季节成分。然后将季

节成分从时间序列中分离出去，即用每一个时间序列观测值除以相应的季节指数，以消除季节成分。

（2）建立预测模型并进行预测。对消除季节成分的时间序列建立适当的预测模型，并根据这一模型进行预测。

（3）计算出最后的预测值。用预测值乘以相应的季节指数，得到最终的预测值。

3．正态分布的概率密度函数是一个左右对称的钟形曲线，参数 μ 是这个曲线的对称轴，同时也决定了曲线的位置，μ 也是正态分布的数学期望；而参数 σ 的大小决定了曲线的陡峭程度，σ 越小，则曲线的形状越陡峭，越集中在对称轴 $x=\mu$ 的附近，这和 σ^2 是正态分布的方差的直观意义一致。

三、计算与分析题

1．（1）已知：$n=50$，$M_{0.05/2}=1.96$。

样本均值为：$\bar{x}=\dfrac{\sum_{i=1}^{k}M_i f_i}{\sum_{i=1}^{k}f_i}=\dfrac{5066}{50}=101.32$ 克，

样本标准差为：$s=\sqrt{\dfrac{\sum_{i=1}^{k}(M_i-\bar{x})^2 f_i}{n-1}}=\sqrt{\dfrac{130.88}{49}}=1.634$ 克。

由于是大样本，所以食品平均重量 95%的置信区间为：

$$\bar{x}\pm z_{\alpha/2}\frac{s}{\sqrt{n}}=101.32\pm1.96\times\frac{1.634}{\sqrt{50}}=101.32\pm0.453$$

即（100.867，101.773）。

（2）提出假设：$H_0:\mu=100$，$H_1:\mu\neq100$

计算检验的统计量：$z=\dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}=\dfrac{101.32-100}{1.634/\sqrt{50}}=5.712$

由于 $z=5.712>z_{0.05/2}=1.96$，所以拒绝原假设，该批食品的重量不符合标准要求。

2．（1）

方差分析表

变差来源	df	SS	MS	F	Significance F
回归	3	12026774.1	4008924.7	72.80	8.88341E-13
残差	26	1431812.6	55069.7	—	—
总计	29	13458586.7	—	—	—

（2）多元线性回归方程为：

$\hat{y}=7589.1025-117.8861x_1+80.6107x_2+0.5012x_3$。

$\hat{\beta}_1=-117.8861$ 表示：在年人均收入和广告费用不变的情况下，销售价格每增加一个单位，销售量平均下降 117.8861 个单位；$\hat{\beta}_2=80.6107$ 表示：在销售价格和广告费用不变的情况下，年人均收入每增加一个单位，销售量平均增加 80.6107 个单位；$\hat{\beta}_3=0.5012$ 表示：在年销售价格和人均收入不变的情况下，广告费用每增加一个单位，销售量平均增加 0.5012 个单位。

（3）由于 Significance F=8.88341E-13< $\alpha=0.05$，表明回归方程的线性关系显著。

（4）$R^2=\frac{SSR}{SST}=\frac{12026774.1}{13458586.7}=89.36\%$，表明在销售量的总变差中，被估计的多元线性回归方程所解释的比例为 89.36%，说明回归方程的拟合程度较高。

（5）$s_e=\sqrt{\frac{SSE}{n-k-1}}=\sqrt{MSE}=\sqrt{55069.7}=234.67$。表明用销售价格、年人均收入和广告费用来预测销售量时，平均的预测误差为 234.67。

3. 解：分别记元件 A,B,C 正常工作为事件 A,B,C，由已知条件可得

$P(A)=0.8,P(B)=0.9,P(C)=0.9$

记系统 N_1 正常工作为事件 N_1，则有 $P_1=P(N_1)=P(ABC)$；

由于事件 A,B,C 相互独立，所以

$P_1=P(A)P(B)P(C)=0.8\times0.9\times0.9=0.648$

记系统 N_2 正常工作为事件 N_2，则有

$P_2=P(N_2)=P\left(A\cap\left(B\cup C\right)\right)$；

由于 A,B,C 相互独立，则有

$$P_2=P(A)\cdot[1-P(\overline{B})\cdot P(\overline{C})]=P(A)\left[1-\left(1-P(B)\right)\left(1-P(C)\right)\right]$$
$$=0.8\times\left[1-0.1\times0.1\right]=0.792$$