# 2003-2023同等学历，统计学试题

## 2003

1. 数据组(4786654536)的众数为( )。

A.6

B.5.5

C.5.4

D.5

【答案】 A

【考点】数据集中趋势的度量

【解析】数据中出现次数最多的变量值称为众数。对于一组数据，众数可以不止一个。数据组(478 6654536)出现次数最多的数值为6。

2. 用剩余法测定循环波动因素时，得到的结果是( )。

A. 只消除了长期趋势的影响

B. 只消除了长期趋势和季节因素的影响

C. 只消除了季节因素的影响

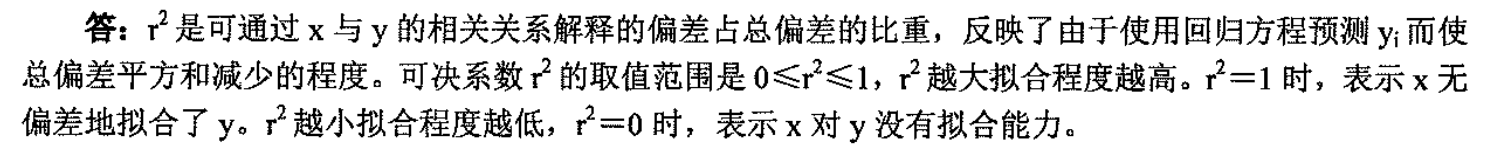
D. 只消除了长期趋势，季节因素以及不规则波动的影响

【答案】D

【考点】循环波动因素的测定

【解析】用剩余法测定循环波动的具体做法如下：①求出季节变动指数 S;②在原时序数列中消除季节因素之影响，计算公式为：T×S×C×/S=T×C×I;③计算长期趋势值T, 并在无季节影响之时序数列中消除长期趋势的影响，计算公式为：T×C×I/T=C×I;④ 用滑动平均法对时序数列C×1进行滑动平均，消除不规则波动的影响，得到循环波动值，通常用百分数表示。

1. 说明一元线性回归分析中可决系数r²的取值范围并说明r²=1 的含义。



1.(8分)某人每天乘公共汽车上班，随机选取10天，记录所花费的时间如下(分钟):

28 29 32 37 33 25 29 32 41 34

(1)对上述数据做茎叶图(以十位数为茎，个位数为叶)。

(2)估计每天乘公共汽车上班所花费的平均时间。

(3)计算乘公共汽车上班所花费时间的样本标准差。

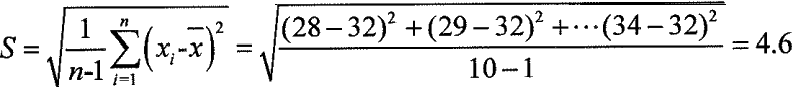
(4)该人也可以以乘另一线路公共汽车上班，根据随机选取的10天，计算乘该线路汽车上班所花费的样本平均时间为28分钟，样本标准差为5.8分钟。从时间角度看，哪一条线路比较好?请说明理由。(运算过程保留小数点后一位。)

解：(1)以十位数为茎，个位数为叶，画出的茎叶图为：



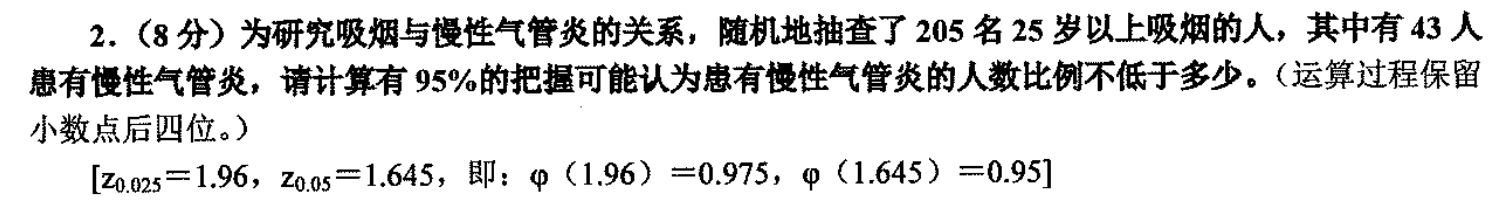
(2)每天乘公共汽车上班所花费的平均时间=(28+29+25+29+32+37+33+32+34+41)/10= 320/10=32(分钟)。

(3)乘公共汽车上班所花费时间的样本标准差为：



(4)从平均时间来看，第二条路线的平均时间更短，选第二条路线更好。但从样本标准差来看，第 一条路线的样本标准差更小，稳定性好，选第一条路线更好。

2.置信下限





## 2004

1. 在下列数据集的综合度量指标中，最容易受到极端值影响的是( )。

A. 极差

B. 上四分数

C. 中位数

D. 众数

【答案】A

【考点】数据的描述性指标与极端值

【解析】极差，又称全距，是数据的最大值与最小值之差，表示数据的跨度， 易受极端数值的影响； B 项，四分位差不像极差那么容易受极端值的影响，但是仍然存在没有充分利用数据所有信息的缺陷；C项，中数对极端值不太敏感；D项，众数是指数据中出现次数最多的变量值。因此众数对于极端值不太敏 感。

2. 设总体X～N(U,c²), 其中σ已知，若样本值容量N 和置信度1-a 均不变，则对于不同的样本观察值，总体均值U 的置信区间的长度( )。

A. 变长

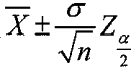
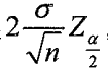
B. 变短

C. 不变

D. 不能确定

【答案】C

【考点】置信区间

【解析】U 的置信度为1-a 的置信区间是 置信区间的长度是 置信区间长度由σ与α决定， 面与X;无关，因此，总体均值U 的置信区闻的长度不变。

1. 对于某股票30天内的每日收盘价数据组成的数据集，简要说明其标准差大小的实际意义。

**答：**标准差是统计中常用的度量数据离散程度的指标，它是用数据自身与平均数之差的大小加权，因而区别对待了大小不同的数据，距离平均数远的数据权重比较大，距离平均数近的数据权重比较小，比较 合理地反映了不同数据对离散度量的作用。

股票30天内的每日收盘价数据集的标准差大小反映了每日收盘价的离散程度。标准差越大，离散程 度越大，即该股票近30天收盘股价波动越大；标准差越小，离散程度越小，即该股票近30天收盘股价波 动越小。

**1.(8分)某高校18个本科毕业生毕业后的起薪见表2(单位：元)。**



(1)以15000为起点，取区间长度为5000元，将上述数据用频数分布表表示；(1分)

(2)按(1)的结果画频率或频数直方图；(1分)

(3)计算他们的平均年薪；(2分)

(4)计算他们的年薪中位数；(2分)

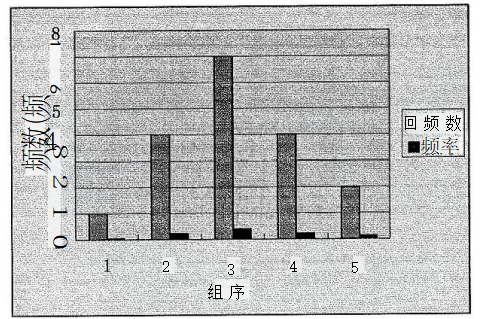
(5)计算年薪的极差。(2分)

解：(1)以15000为起点，取区间长度为5000元，决定各组界限值，确定分点。各组区间依次为：(15000, 20000)、(20000,25000)、(25000,30000)、(30000,35000)、(35000,40000)。数出各组频数，计算频率，作出频数分布表，见表3:

表3 频数分布表

|  |  |  |  |
| --- | --- | --- | --- |
| 组序 | 分组界限 | 频数 | 频率 |
| 1 | (15000,20000) | 1 | 0.056 |
| 2 | (20000,25000) | 4 | 0.222 |
| 3 | [25000,30000 | 7 | 0.389 |
| 4 | [30000,35000] | 4 | 0.222 |
| 5 | [35000,40000 | 2 | 0.111 |
| 合计 |  | 18 | 1 |

(2)按(1)的结果画出频率或频数直方图，见图4:



(3)平均年薪=(26496+23318+23447+28040+35084+31212+31817+32166+30865+35448+ 28905+28226+26686+29530+23659+22179+29730+18955)/18=505763/18=28098(元)。

(4)将数据从小到大排列，第九个和第十个数据的平均值为年薪的中位数，中位数=(28226+28905) /2=28565.5。

(5)年薪的极差=35448-18955=16493(元)。

2.

2. **(8分)许多人认为适量饮用葡萄酒可以预防心脏病。为了了解该种说法是否有道理，某研究机构**通**过调查得到以下数据(见表4),其中包括19个发达国家一年的葡萄酒消耗量(用平均每人饮用葡萄酒摄取酒精的升数表示)以及一年中因心脏病死亡的人数(每10万人死亡人数)**。

**表** **4** **葡萄酒消耗量及心脏病死亡人数**



(1)做上述数据的散点图，用X轴表示从葡萄酒中得到的酒精量，用Y轴表示每10万心脏病死亡 人数。(2分)

(2)上述散点图说明葡萄酒消耗量对心脏病死亡人数有什么影响?(2分)

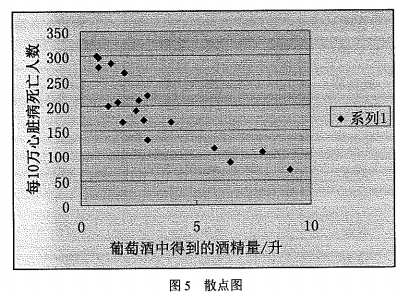
**(3)以从葡萄酒中得到的酒精量为自变量X, 以每10万人心脏病死亡人数为因变量Y, 得到的线性** **回归结果为：**



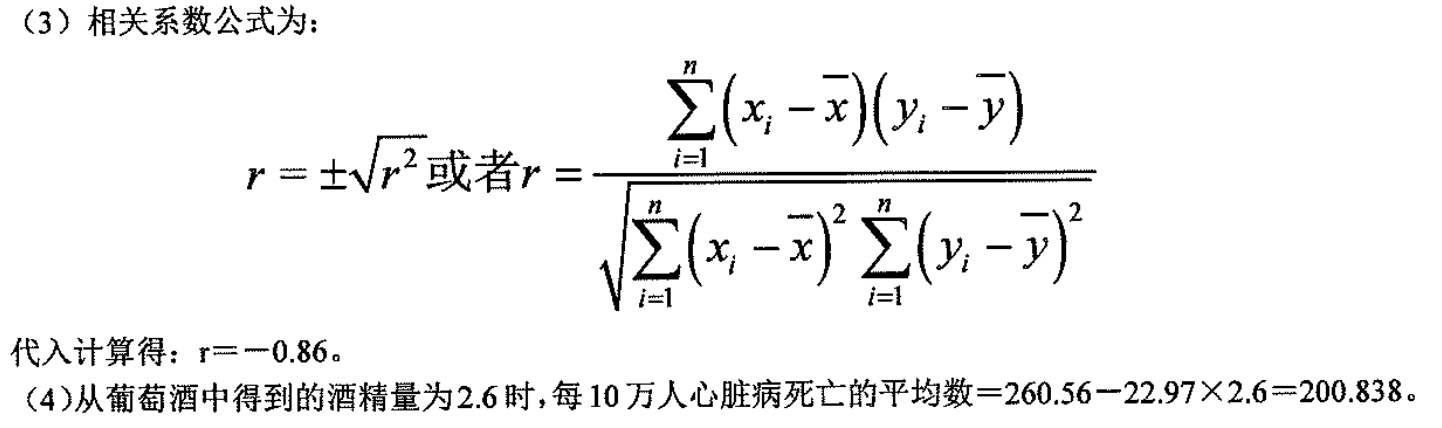
**求从葡萄酒中得到的酒精量X** **与每10万人心脏病亡人数** **Y** **之间的相关系数。(2分**)

**(4)根据(3)中的回归结果，计算：从葡萄酒中得到的酒精量为2.6时，每10万人心脏病死亡的平均数。**(2分)。

**解：** (1)根据已知数据得出散点图，如图5所示：



(2)从散点图可以看出，随着葡萄酒消耗量的增加，心脏病死亡人数减少。



## 2005

1. 计算一组数据的标准差时，如果每一个数值加上一个常数a, 与原数据组标准差相比，新数据组的标准差( )。

A. 大

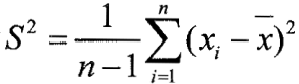
B. 小

C. 不变

D. 不一定

【答案】C

【考点】标准差

【解析】样本标准差的计算公式为，而

若令每一数值加上一个常数a 后的数值为x;’, 则x;′=x+a, 那么每一数值加上一个常数后的样本均值x'=x+a, 所以根据公式可知，新数据组的样本方差S²的大小不变，则新数据组的标准差S 也不变。

2. 设总体x 服从均值为μ、标准差为σ的正态分布，其中σ已知。在95%的置信度下，由样本容量 为25的一个样本得到的置信区间长度为24厘米。如果样本容量为400,则置信区间的长度为( )。

A.6 厘米

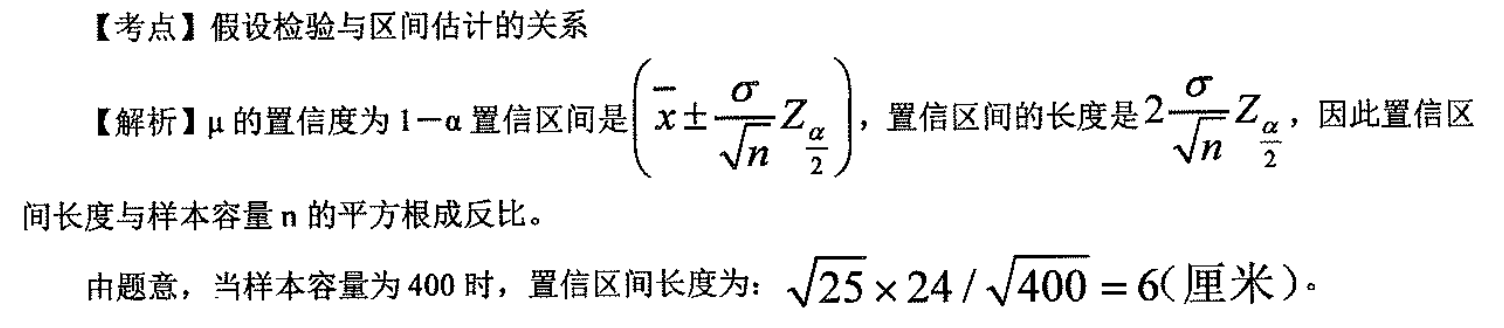
B.1.2 厘米

C.96 厘米

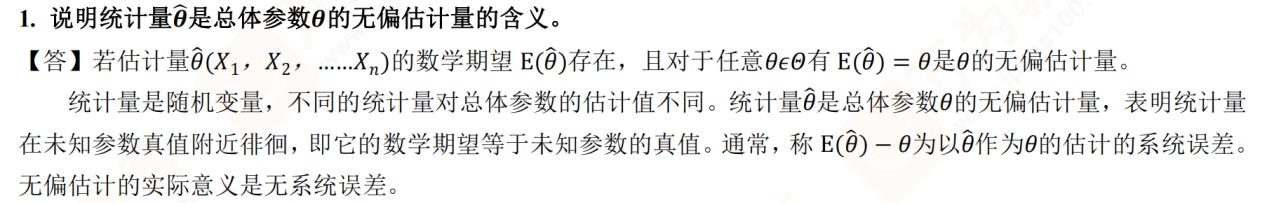
D.9.6 厘米

【答案】A

【考点】假设检验与区间估计的关系



1.说明统计量分是总体参数θ的无偏估计量的含义。



1.(8分)某杂志对其订阅者进行的调查中，有一个问题是：“在最近的4期杂志中，您读过或浏览过其中的几期?”下面的频数分布表显示了500人的回答情况。



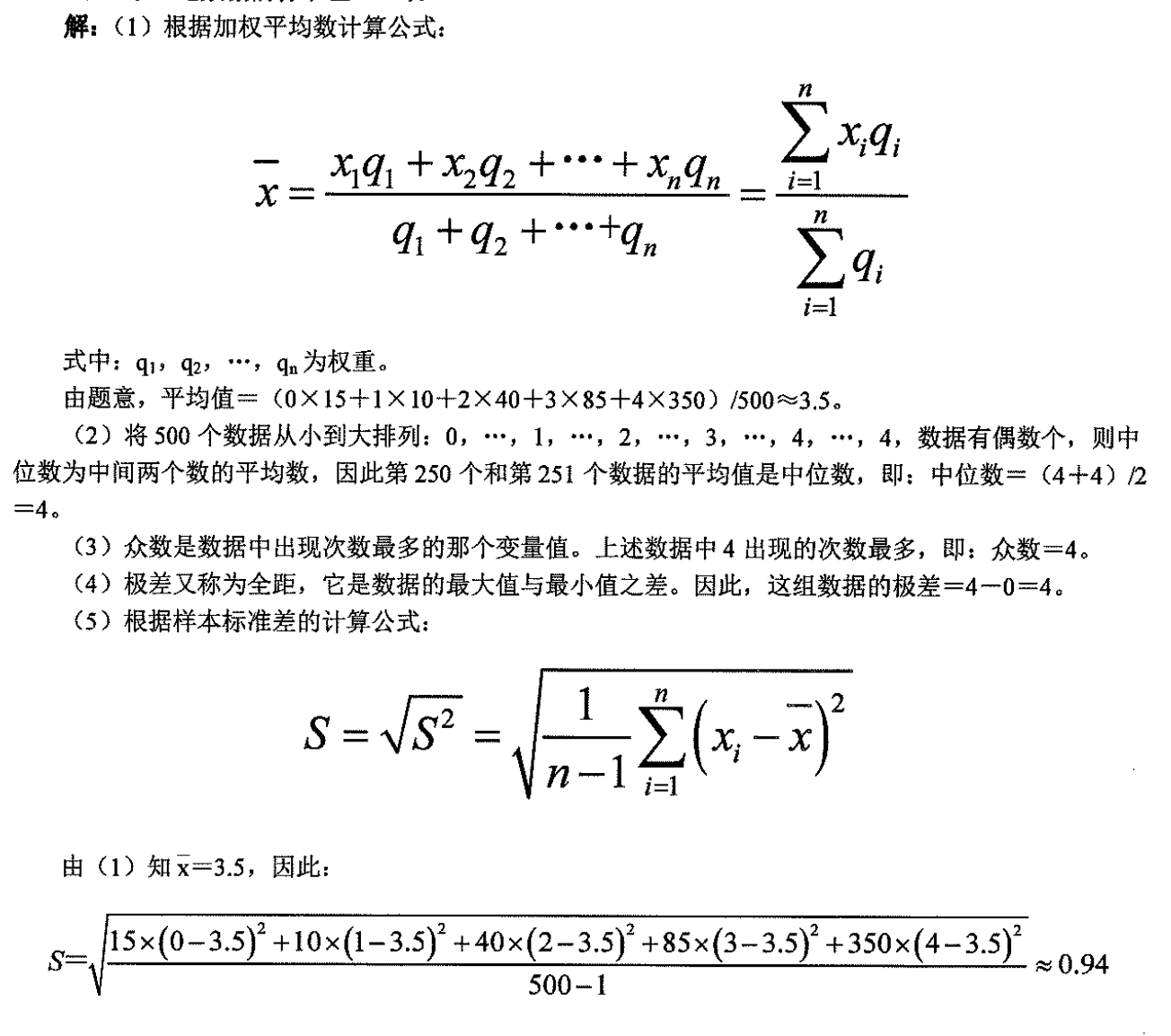
(1)求订阅者在最近4期中所读过的期数的平均值；(2分

(2)求上述数据的中位数；(1分)

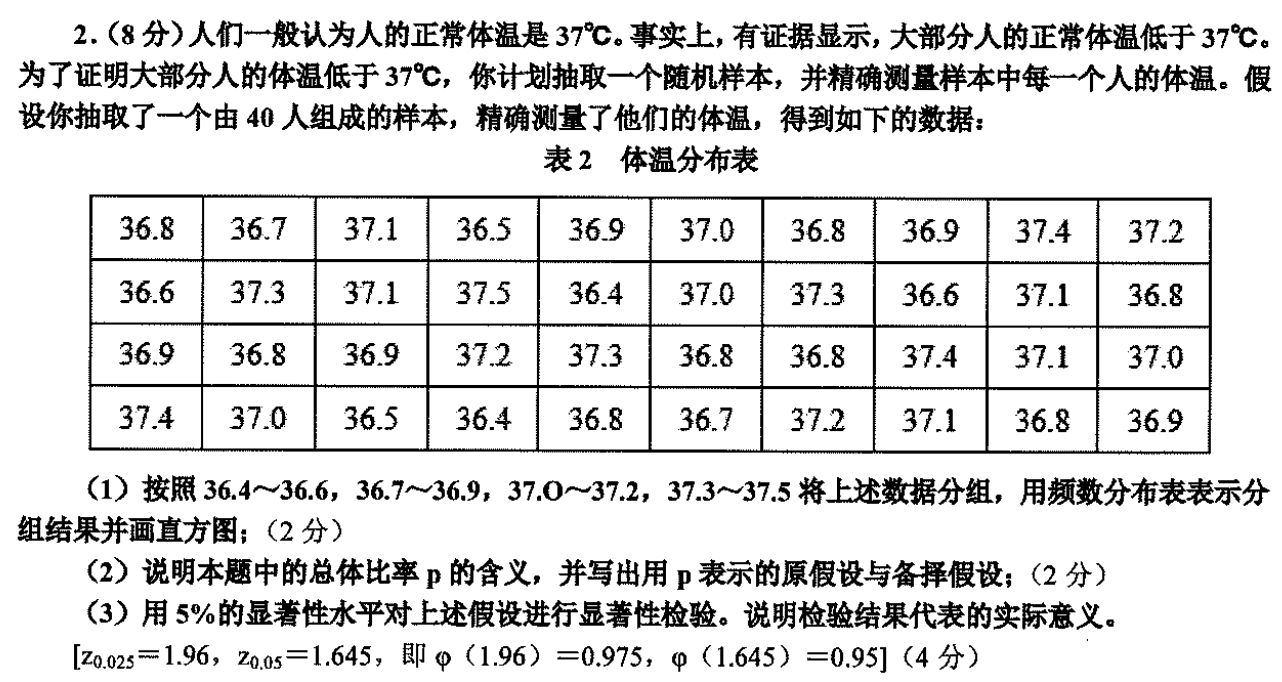
(3)求上述数据的众数；(1分)

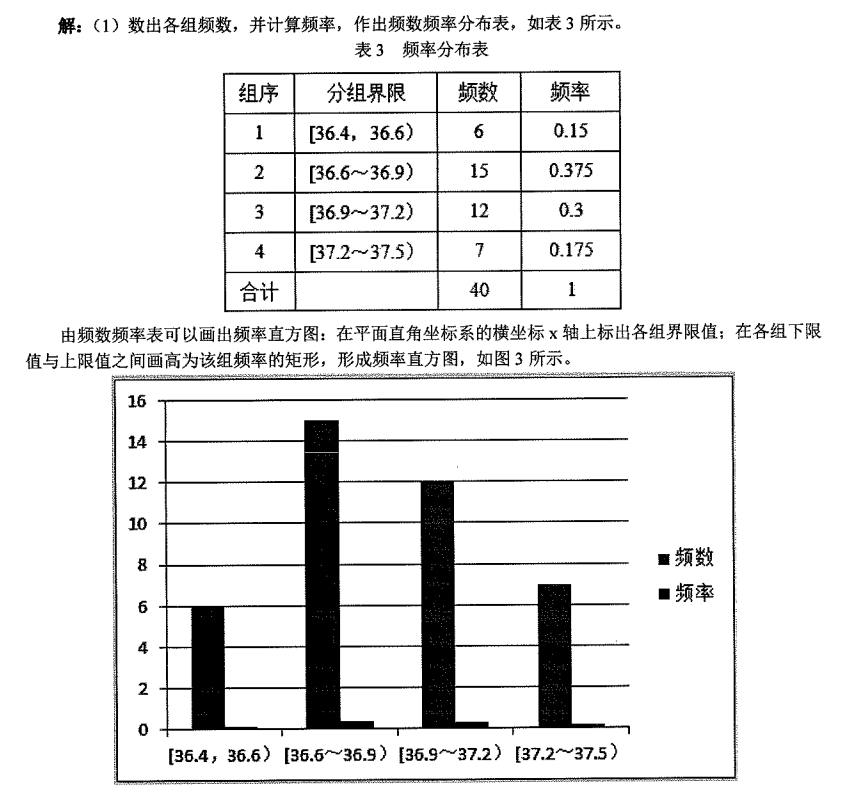
(4)求上述数据的极差； (2分)

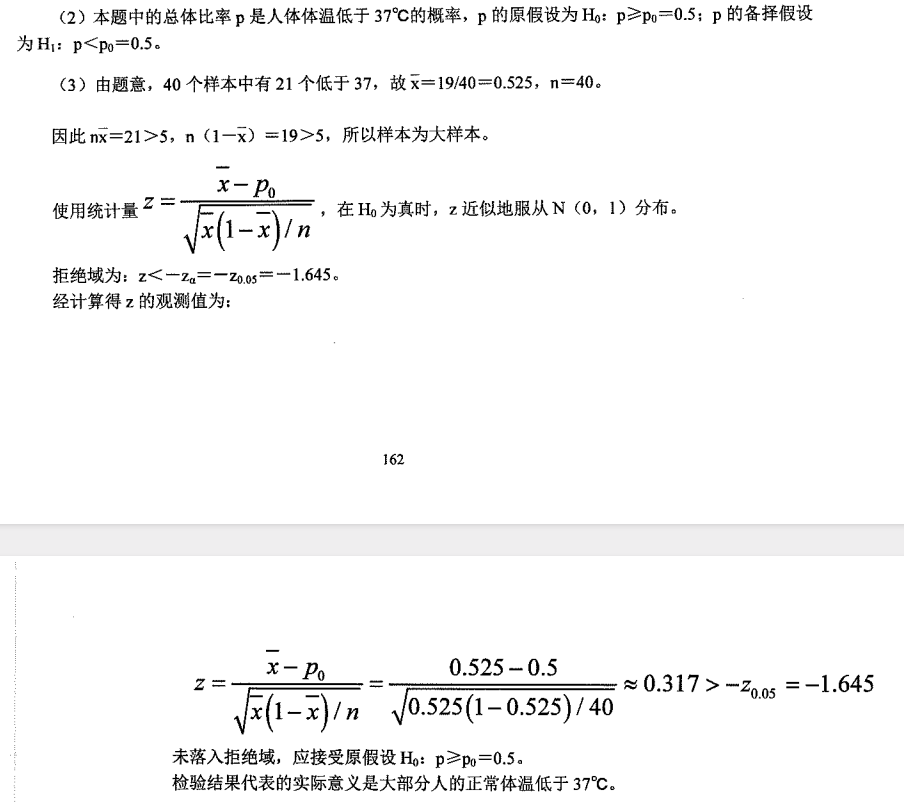
(5)求上述数据的标准差。(2分)



2.







## 2006

1. 下列数据中肯定不是某个随机事件的概率的是( )。

A.0.8

B.1/2

c.√212

D.1.2

【答案】D

【考点】随机变量概念和随机变量的数字特征

【解析】随机事件的发生是偶然性的，随机事件的概率是随机事件发生的可能性，随机事件的概率是 一个介于0到1之间的数。

2. 一个离散型随机变量的概率分布如表1所示，则这个随机变量的数学期望等于( )。

表1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | -1 | 0 | 1 | 2 |
| 概率 | 0.2 | 0.1 | 0.2 | 0.5 |

A.1

B.2.4

C.0.25

D.3.6

【答案】A

【考点】离散型随机变量的数学期望

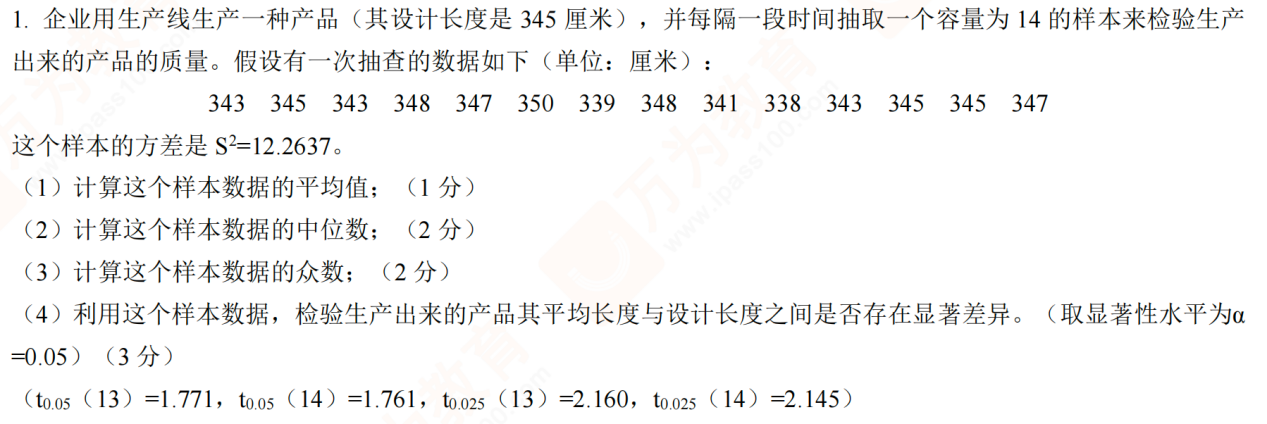
【解析】离散型随机变量的数学期望：

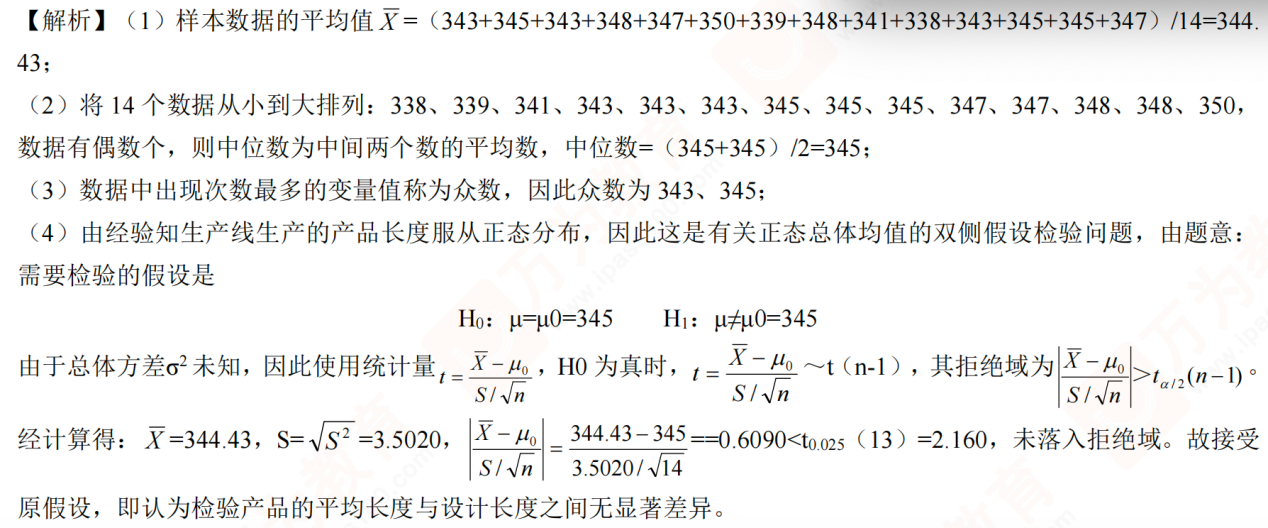


1. 有一大批产品的次品率为I%, 从中随机地抽取一个产品，令随机变量X=I代表该产品为次品， 请写出X的概率函数。

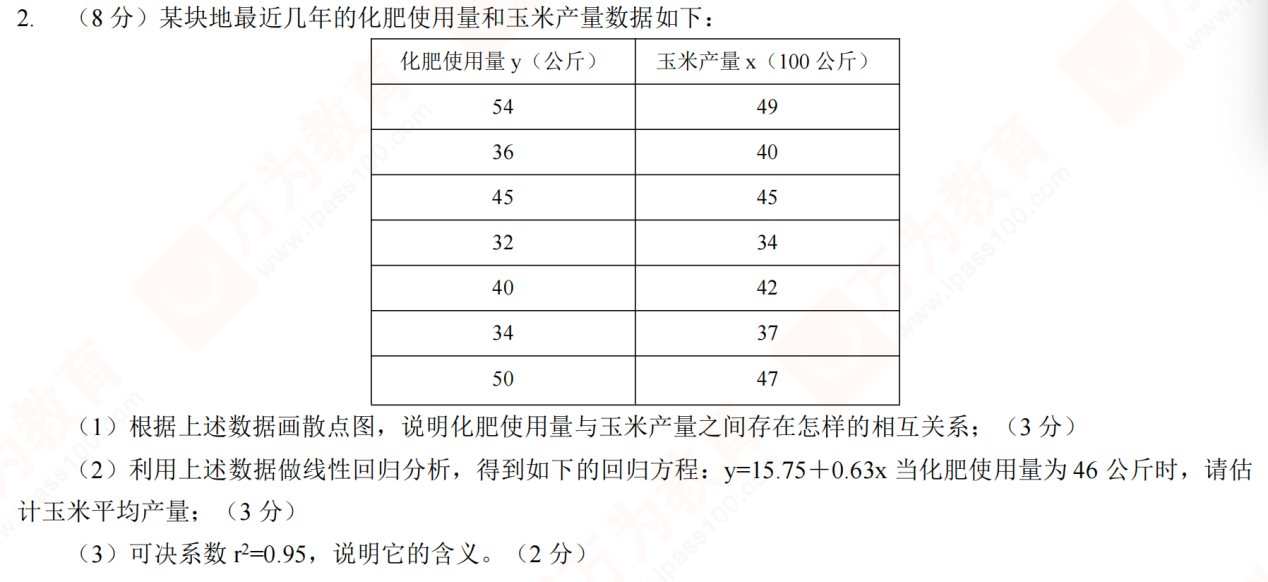
答： X的概率函数为：P{X=I}=I%。

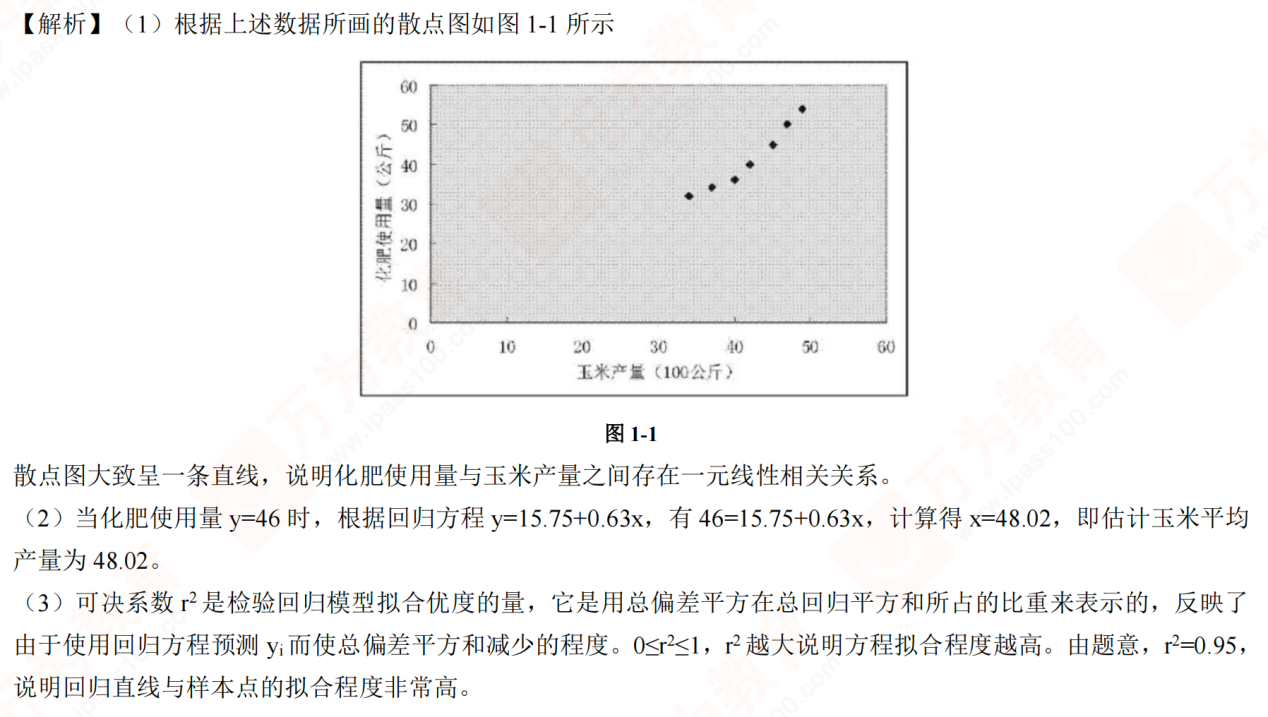
1.





2.





## 2007

1. 数据28,25,26,25,30,27的中位数是( )。

A.26

B。26.5

C.27

D.27.5

【答案】B

【考点】中位数

【解析】将数据从小到大排列：25、25、26、27、28、30,数据有偶数个，则中位数为中间两个数的 平均数，中位数=(26+27)/2=26.5。

2. 在假设检验中，记 H₀ 为待检验的假设，则第一类错误为( )。

A.H₀ 为真，接受H₀

B.H₀ 不真，拒绝H

C.H₀ 为真，拒绝Ho

D.H₀ 不真，接受H₀

【答案】C

【考点】假设检验的两类错误

【解析】H₀ 为真时仍可能作出拒绝H₀ 的错误判断，这种错误称为第一类错误，或“以真为假”的错 误。H₀ 不为真时作出接受H₀ 的错误判断，这种错误称为第二类错误，或“以假为真”的错误。

1. 正态分布概率密度函数曲线的性质有哪些?

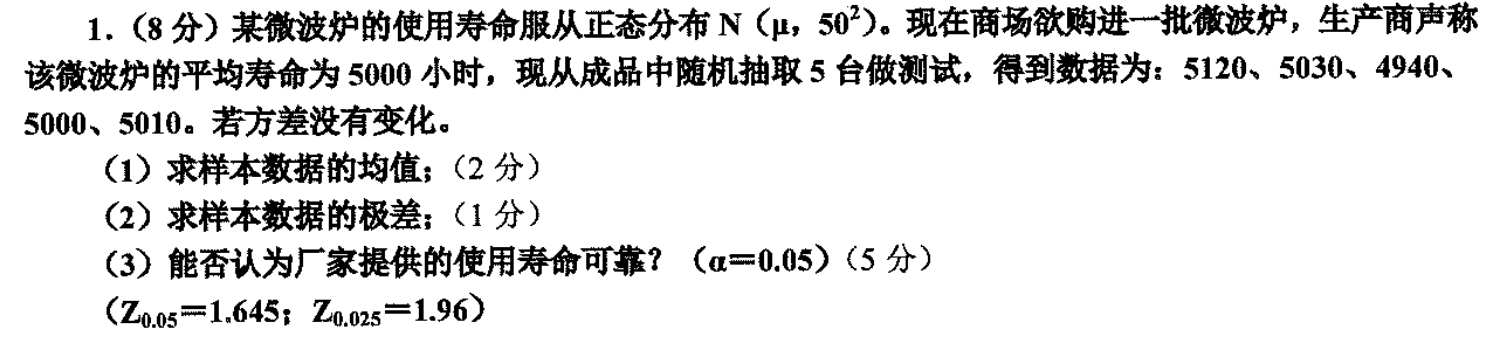
答 ：正态分布概率密度函数曲线的性质有：

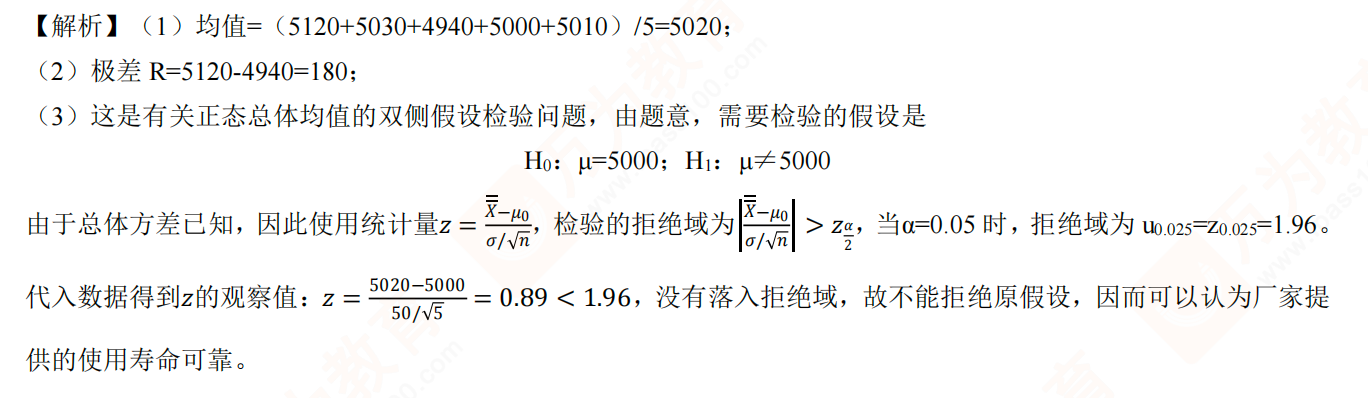
(1)对于任意的μ和Q², 密度函数p(x) 与 x 轴之间的面积是1。

(2)密度函数的曲线关于直线x=μ 对称，因此μ是正态分布的位置参数。

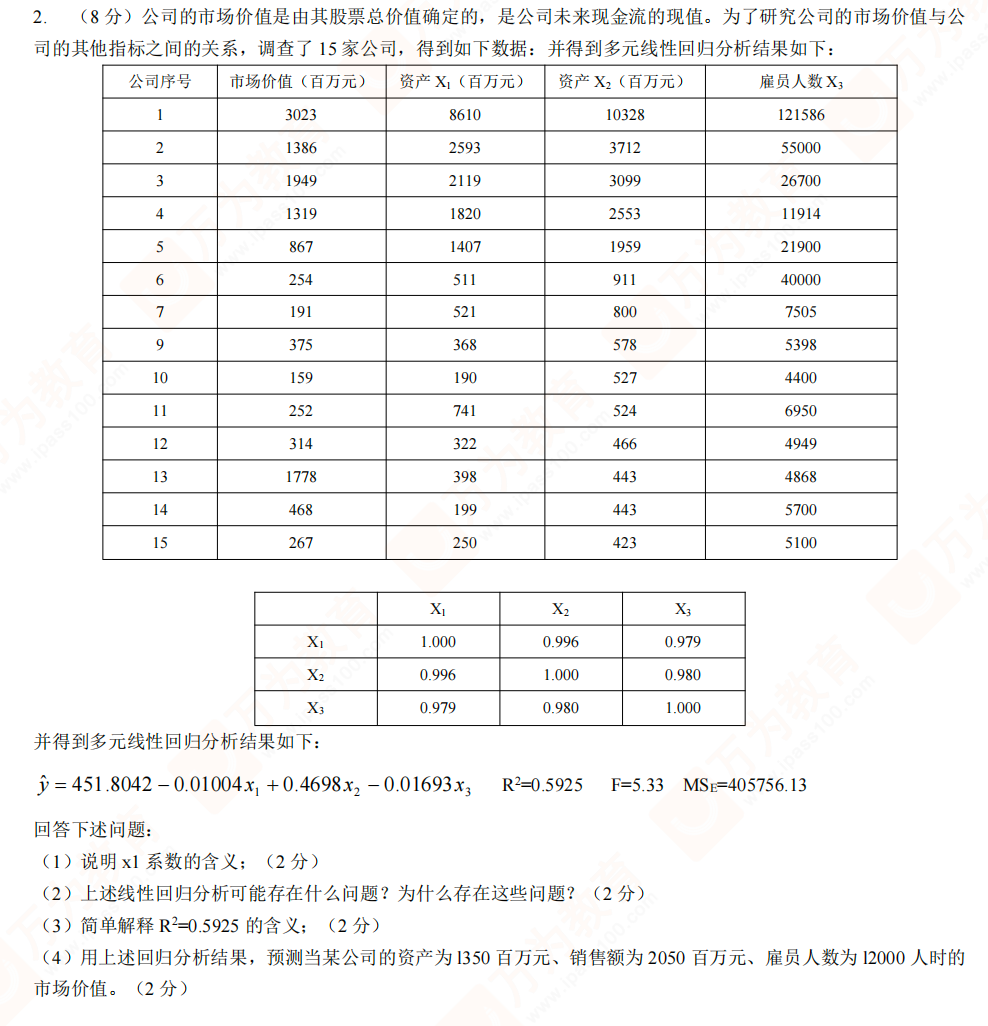
(3)方差o²的大小决定了密度曲线的高矮胖瘦：o²越大，曲线越矮越胖； o²越小，曲线越高越瘦。

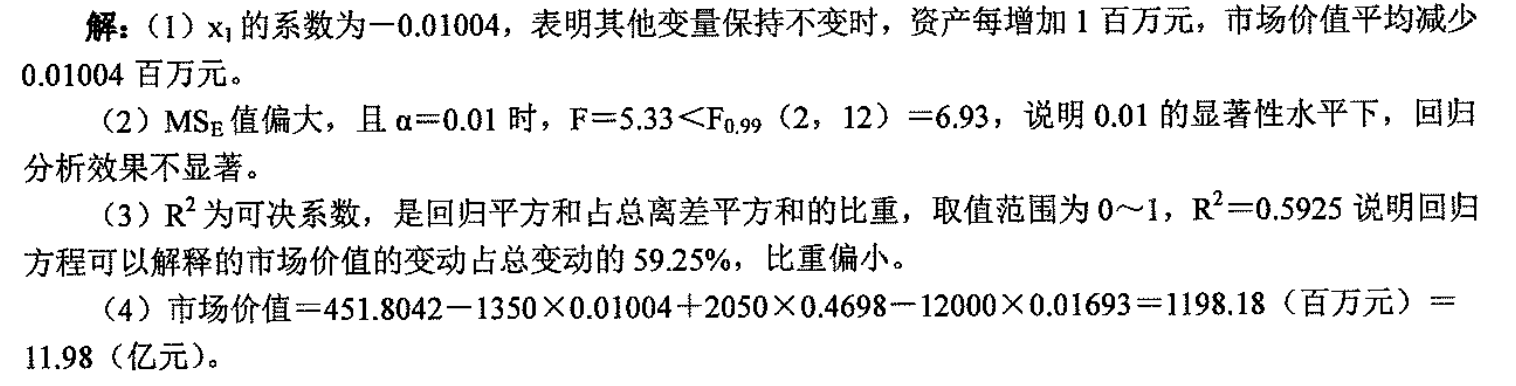
1.





2.多元线性回归





## 2008

1. 数据88,85,89,87的极差是( )。

A.1

B.2

C.3

D.4

【答案】D

【考点】数据的离散趋势度量

【解析】极差，又称全距，是数据的最大值与最小值之差，表示数据的跨度，但是它易受极端数值的 影响。题中，极差=89-85=4。

2. 下列数据集的综合度量指标中，最不容易受极端值影响的是( )。

A. 均值

B. 上四分位数

C. 中位数

D. 众数

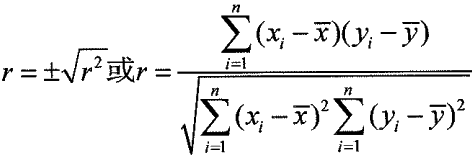
【答案】D

【考点】数据集中趋势的度量

【解析】众数是数据中出现次数最多的变量值， 不受极端值的影响。均值(平均数)的优点在于容易 理解和计算，缺点是对极端值特别敏感。中位数对极端值不像平均数那么敏感。四分位差，即上四分位数 与下四分位数的差数， 一般不受极端值的影响。

1. 两个自变量的相关系数接近一1,说明什么?

答： 为可决系数。

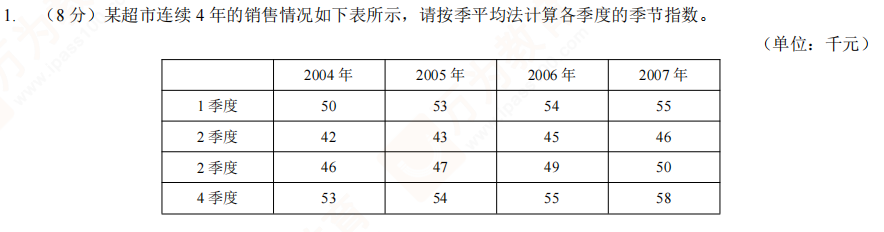
为相关系数。

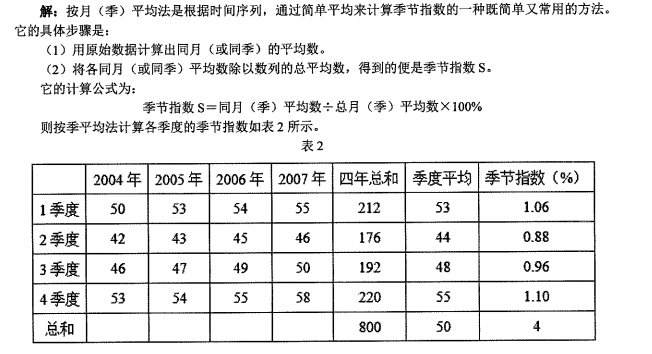
式中，相关系数r不但反映了y 和 x 之间的线性相关密切程度，而且反映了y 和x 之间的相关方向。

r>0,y 和 x之间呈一定程度的正相关关系； r<0,y和x之间呈一定程度的负相关关系，同时， r与b同号(这里的b 是 y=a+bx中的b, 即 b为边际消费倾向)。

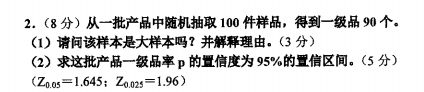
两个自变量的相关系数接近一1,说明这两个自变量之间存在高度负相关。

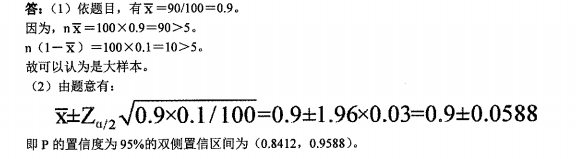
1.按季平均法





2.





## 2009

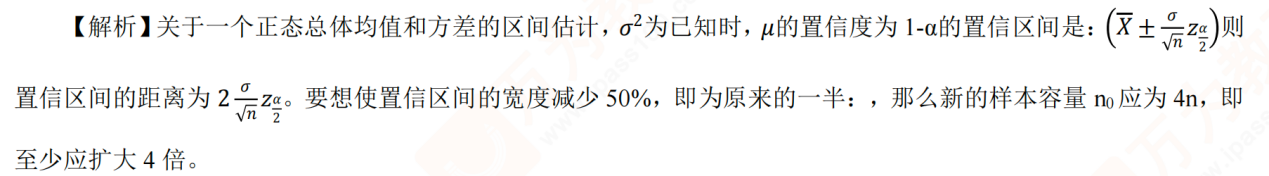
1.在对某个总体均值的区间估计中，在给定置信水平下，如果要想使给定的置信区间宽度再减少50%, 则样本容量至少应扩大( )。

A.2 倍

B.3 倍

C.4 倍

D.6 倍



2. 当计算一个时期到另一个时期的销售额的年平均增长速度时，应采用哪种平均数?( )

A. 众数

B. 中位数

C. 算术平均数

D. 几何平均数

【答案】 D

【考点】数据集中趋势的度量

【解析】n个观察值连乘积的n 次方根就是几何平均数。计算几何平均数要求各观察值之间存在连乘积关系，它的主要用途包括：①对比率、指数等进行平均；②计算平均发展速度。题中所述年平均增长速 度应采用几何平均数进行计算。

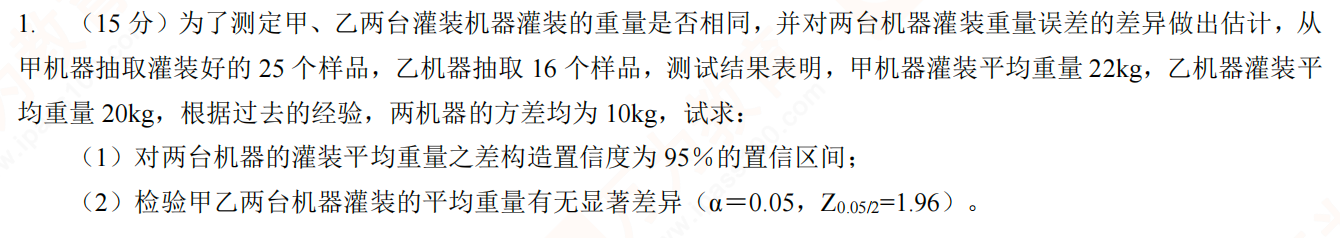
1. 有人说：“在方差分析中，组间方差和组内方差都包括系统性因素和随机性因素”。这种说法对吗? 为什么?

答 ：这种说法是错误的。原因如下：

方差分析用于两个及两个以上样本均数差别的显著性检验。观察数据的误差来源可以从两个方面来说明。衡量因素的同一水平(同一总体)下样本数据的方差，称为组内方差，衡量因素的不同水平(不同总 体)下样本数据的方差，称为组间方差。组内方差只包含随机误差，而组间方差既包含随机误差也包括系统误差。

方差分析中，如果不同水平对结果没有影响，那么在组间方差中只含有随机误差没有系统误差，这时，组间方差与组内方差就应该很接近，两个方差比值就会接近1;反之，如果不同水平对结果有影响，在组间方差分析中除了包含随机误差外还包括系统误差，这时的组间方差就会大于组内方差，组间方差与组内方差的比值就大于1。当这个比值大到一定程度时，就可判断不同水平之间存在显著的差异。

1.两个正态总体的均值差





## 2010

1. 产品产量与单位成本之间的简单相关系数为0.92,并通过检验，这说明二者之间存在着( )。

A. 高度相关

B. 中度相关

C. 低度相关

D. 极弱相关

【答案】A

【考点】a² 的估计、可决系数与相关系数

【解析】相关系数r 用来度量两个定量变量之间的线性相关程度。根据经验可将相关程度分为以下几种情况：①当⊥r「≥0.8时，可视为高度相关；②当0.5≤|r|<0.8 时，可视为中度相关；③当0.3≤|r <0.5时，视为低度相关；④当|r|<0.3 时，说明两个变量之间的相关程度极弱。

**2.** **在因变量的总方差中，若回归方差所占比重大，而相应剩余方差所占比重小，则自变量与因变量**( )。

A. 零相关

B. 相关程度低

**C. 完全相关**

**D. 相关程度高**

**【答案】D**

**【考点】o²的估计、可决系数与相关系数**

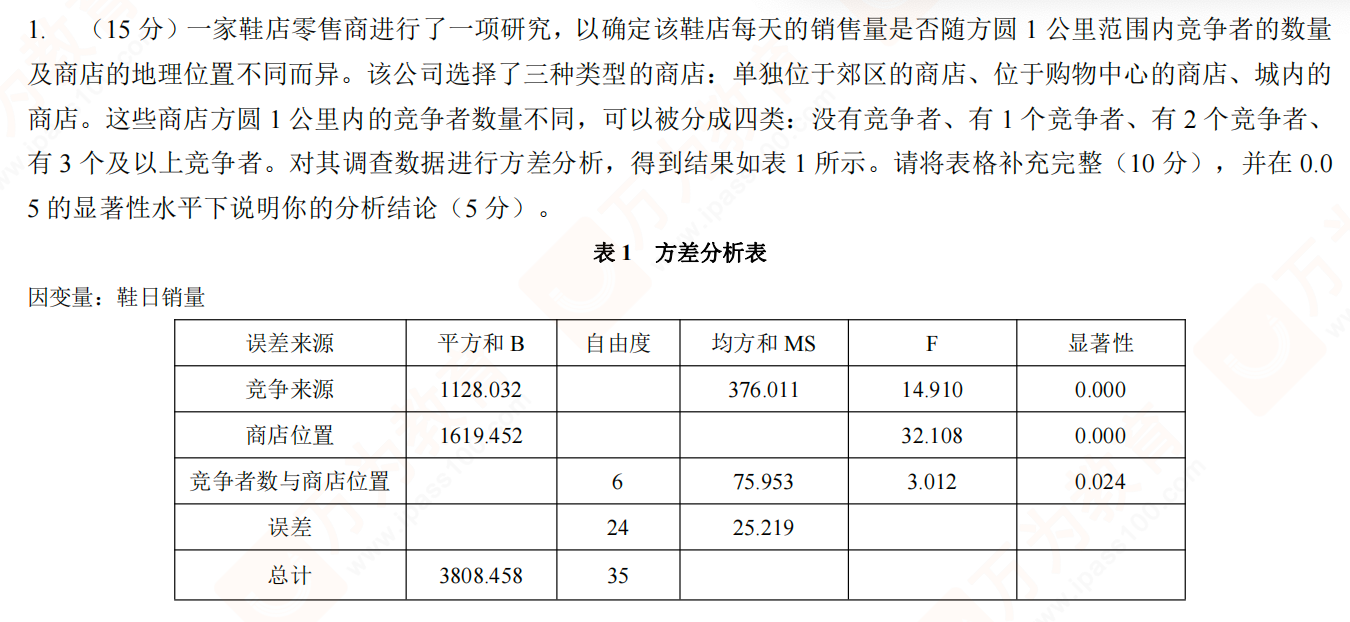
**【解析】自变量所能解释的方差在总方差中所占的比重很大，即该回归方程的拟合程度很好，进而决定系数的取值接近1。说明自变量和因变量的相关程度高**。

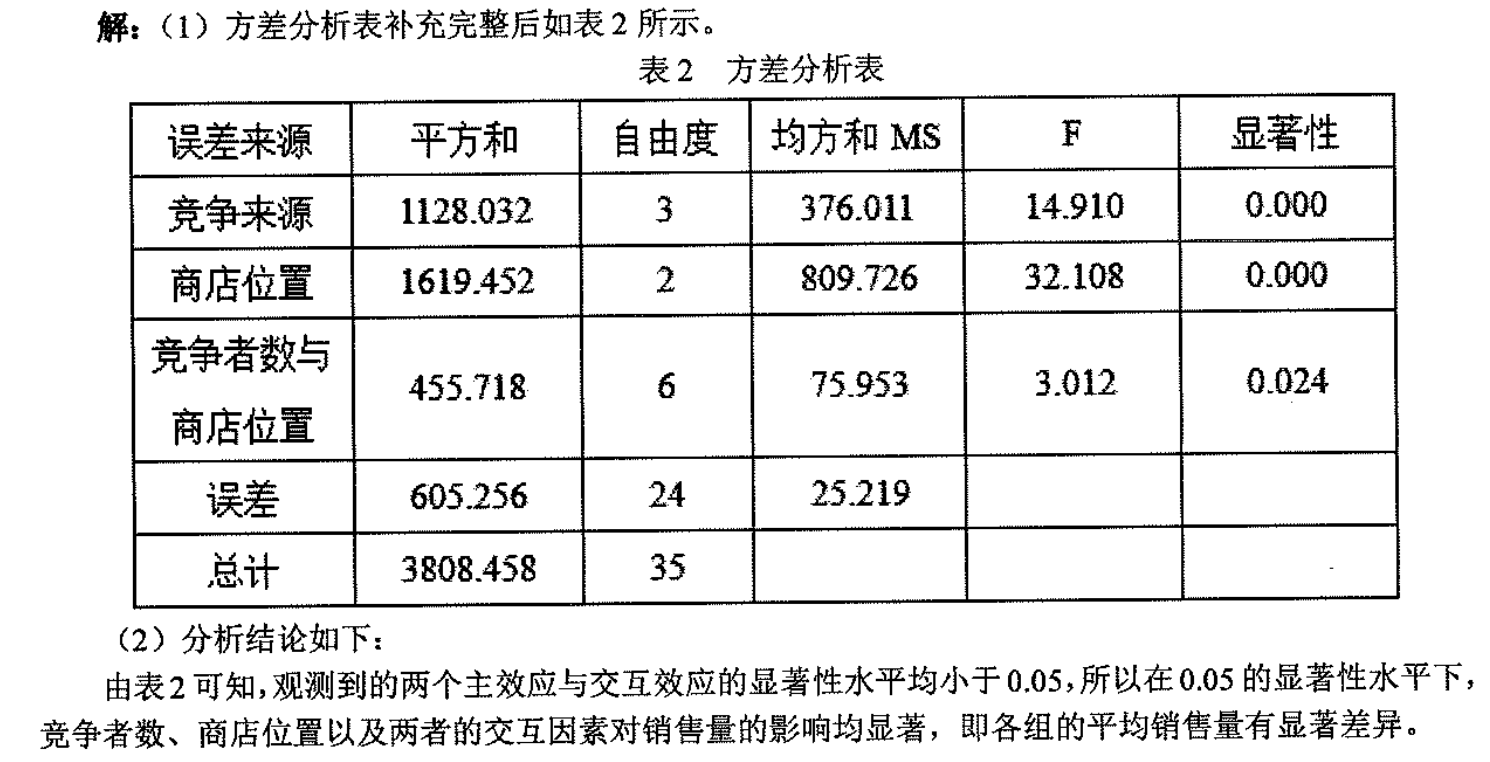
1. 有人说：“当数据偏态程度较大时，应选择众数或者中位数等代表集中趋势，而不应使用均值。” 这话对吗?为什么?

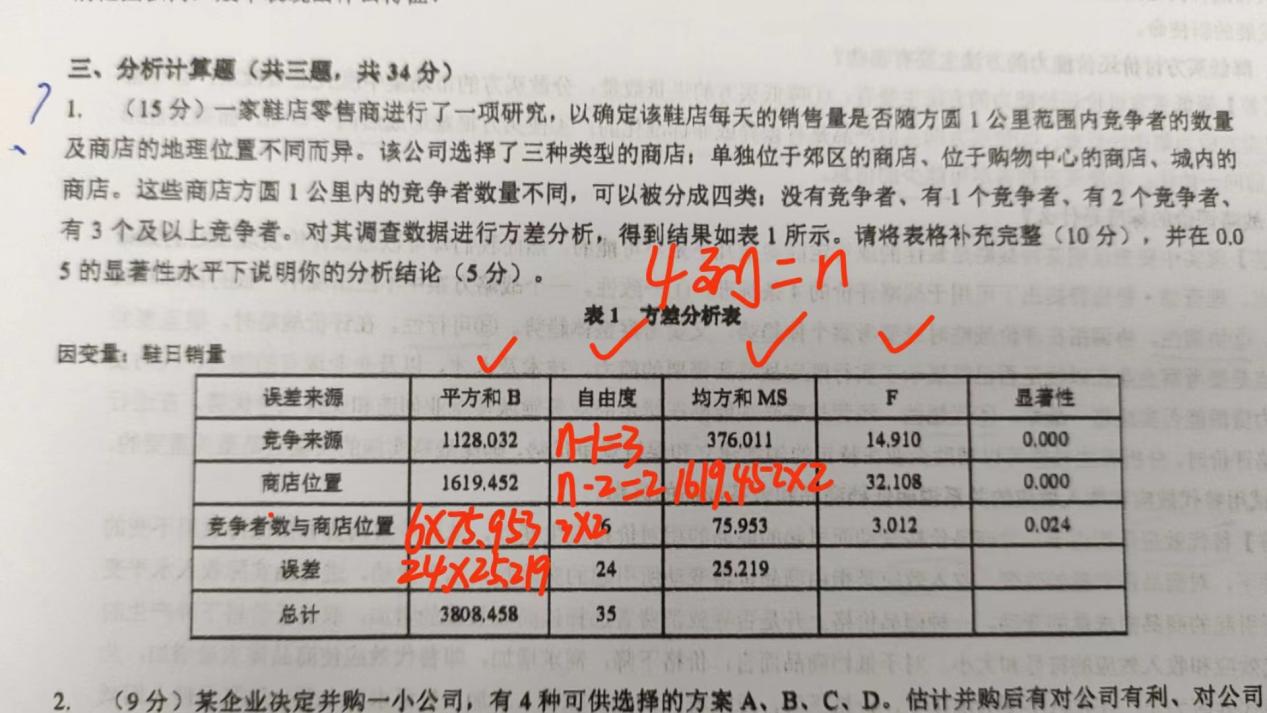
答： 正确。具体分析如下：

当数据偏态程度较大时，说明数据中存在极端值。而均值的一个主要缺点是它对极端值特别敏感(极端值就是数据集中特别大或特别小的个别数据)。但是中位数和众数对极端值不像平均数那么敏感，受极 端值的影响较小，具有统计上的稳健性，所以当数据偏态程度较大时，应选择众数或者中位数等代表集中趋势。

1.







## 2011

1. 下列计量属于集中趋势的是( )。

A. 总体单位数

B. 样本单位数

C. 平均数

D. 方差

【答案】C

【考点】数据集中趋势的度量

【解析】数据集中趋势的度量表示数据的集中点或中心值。根据不同的实际需要，可以采用众数、中 数(或中位数)、 平均数(或均值)、加权平均数、上四分位数、下四分位数来对数据的集中趋势进行度量

2. 回归分析中的一个重要假定是( )。

A. 自变量X 是随机变量，且无观测误差

B. 自变量X 是非随机变量，且无观测误差

C. 因变量y 是随机变量，且无观测误差

D. 因变量y 是非随机变量，且无观测误差

【答案】C

【考点】简单线性回归模型及其基本理论假设

【解析】回归分析中的一个重要假定是：因变量y是随机变量，且无观测误差。而自变量x 是非随机变量，且其有观测误差。

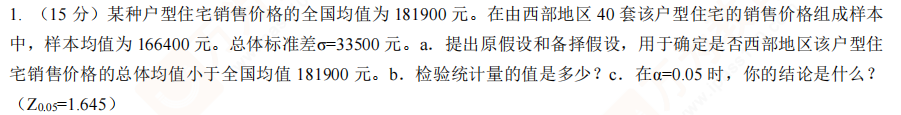
1. 在连续的变量中，如果分布没有明显的最高点或者数据个数不多，则众数不存在。这种判断正确 吗?为什么?

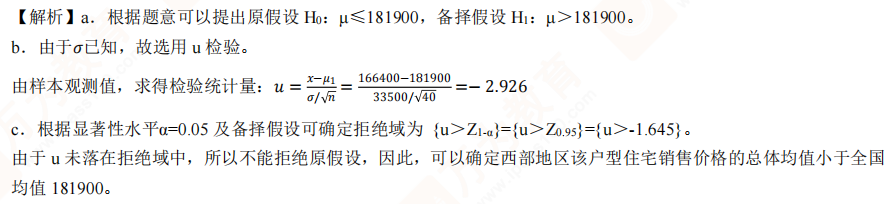
答： 不正确，具体分析如下：

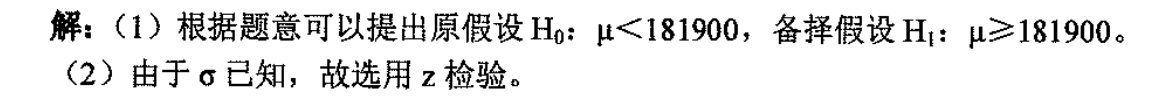
(1)众数是指数据中出现次数最多的变量值。对于一组数据，众数可以不止一个。

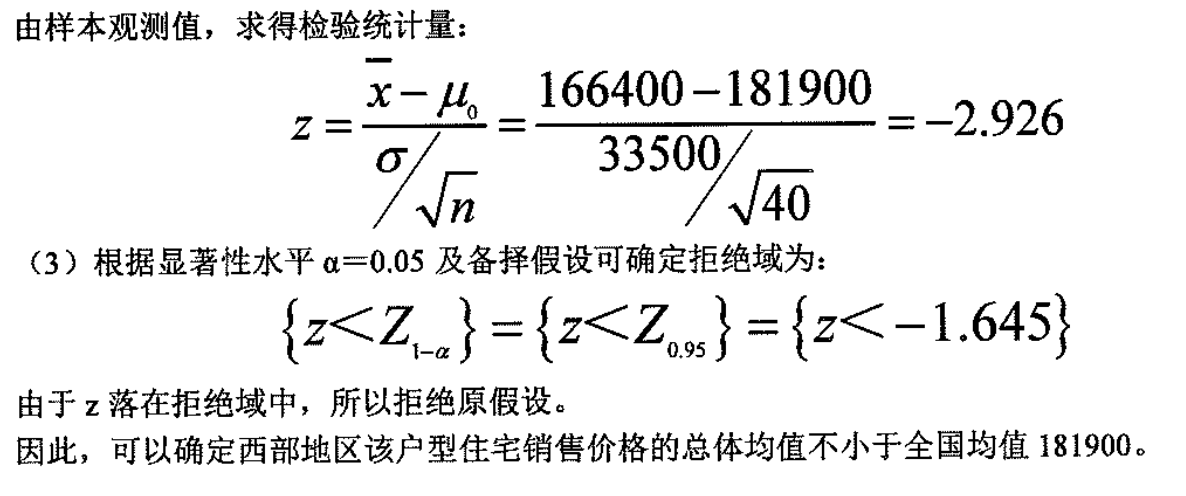
(2)众数的主要缺点是一个数据集可能没有众数，或众数不唯一，而数据集的平均数和中数都是存在并且唯一的。它的优点是它反映了数据集中最常见的数值，即最普遍的数值，并且它不仅对数量型数据 集有意义，对分类型数据集也有意义。

1.假设校验









## 2012

1. 数据88,85,89,87的极差是( )。

A.1

B.2

C.3

D.4

【答案】D

【考点】数据的离散趋势度量

【解析】 极 差，又称为全距，它是数据的最大值与最小值之差，表示数据的跨度，但是它易受极端数 值的影响。题中，该组数据的极差=89-85=4。

2. 按地理区域划片所进行的区域抽样，其抽样方法属于( )。

A. 简单随机抽样

B. 等距抽样

C. 分层抽样

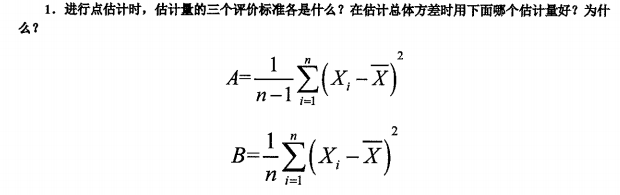
D. 整群抽样

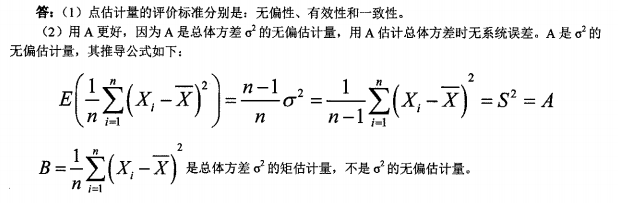
【答案】D

【考点】整群抽样

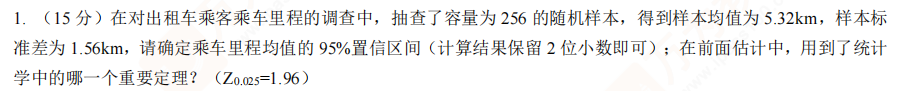
【解析】整群抽样是指将总体中的成员分为若干群(或组),从这些群(或组)中抽取部分群(或组), 调查对象是被抽中的这些群(或组)中的所有成员。A项，简单随机抽样，又称为完全随机抽样，它是在 无限总体中进行的无放回独立抽样或在有限总体中进行的有放回随机抽样。B 项，系统抽样，又称为等距抽样，是指将总体中的调查单位按某种次序排列，随机地选定初始单位，然后按相等的间距抽取其他样本单位。C项，分层抽样是指将总体中的成员按某种原则划分成若干个子总体，每个子总体称为一个层，在每层中独立进行简单随机抽样或其他抽样。

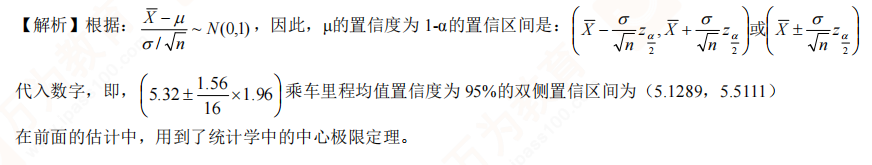
1.点估计





1.中心极限定理





## 2013

1. 相关性和因果关系都可以描述两个变量之间的关系。下面说法正确的是( )。

A. 两个强正相关的变量必然存在因果关系

B. 两个强负相关的变量必然存在因果关系

C. 只要两个变量统计相关就必然存在因果关系

D. 两个统计相关的变量不一定存在因果关系

【答案】D

【考点】多元线性回归分析

【解析】存在因果关系的两个变量肯定是相关的，但是存在相关性的两个变量不一定存在因果关系 可能是并列关系等。

【说明】根据最新版考试教材，该知识点已删除。

2. 关于置信区间，下面描述正确的是( )。

A.99% 的置信区间与95%的置信区间一样宽

B. 在置信水平相同的情况下，减少样本数量，则可以获得更窄的置信区间

C. 在置信水平相同的情况下，增加样本数量，则可以获得更窄的置信区间

D. 置信区间和样本数量无关

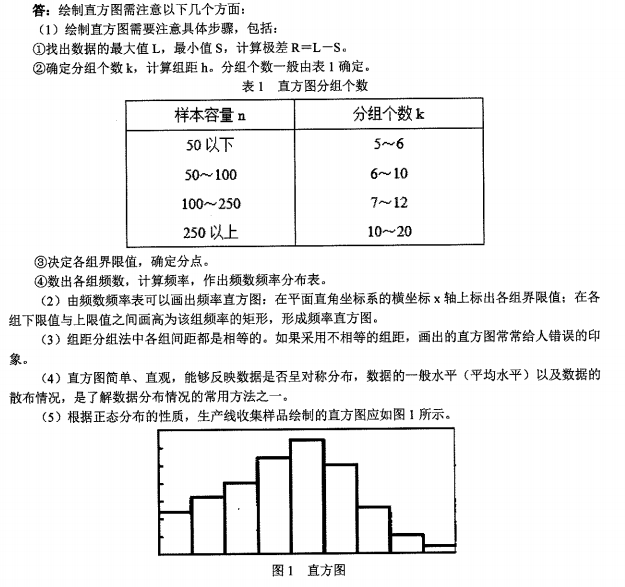
【答案】C

【考点】参数区间估计的一般提法

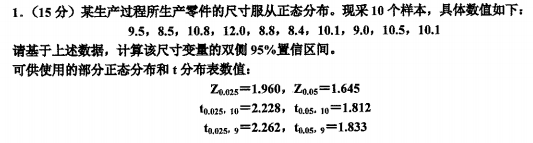
【解析】置信区间是指由样本统计量所构造的总体参数的估计区间。在置信水平相同的情况下，样本量越多，置信区间越窄。

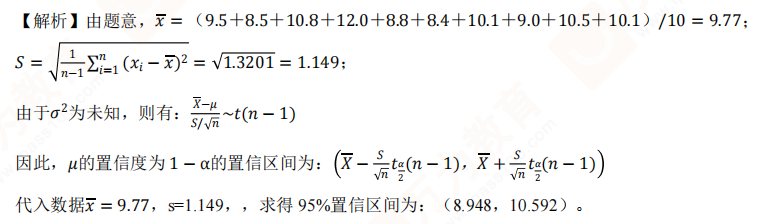
【说明】根据最新版考试教材，该知识点已删除。





1.置信区间





## 2014

1. 下列样本统计量中受奇异点的影响最小的是( )。

A. 均值

B. 方差

C. 标准差

D. 中位数

【答案】D

【考点】数据集中趋势的度量

【解析】奇异点，即样本中的极端值。将数据按大小次序排列后，若数据组包含奇数个数据，则中间 位置上的数值或变量值就是中位数；若数据有偶数个，则中位数是中间两个数的平均数。中位数是一组数据排序后处于中间位置上的变量值，是位置平均数，不受极端值的影响。

2. 下面图形工具中可以用来分析两个变量之间的相关性的是( )。

A. 直方图

B. 散点图

C. 茎叶图

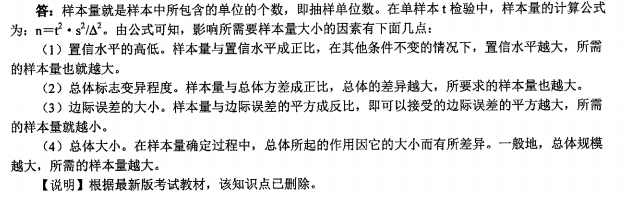
D. 饼状图

【答案】B

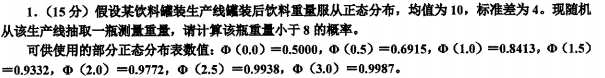
【考点】散点图

【解析】 散点图是用二维坐标展示两个变量之间关系的一种图形，可以用来分析西个变量之间的相关性。A项，直方图又称质量分布图，是由一系列高度不等的纵向条纹或线段表示数据分布的情况；C项，茎叶图是反映原始数据分布的图形；D项，饼状图显示一个数据系列，这些数据源自数据表的行或列。ACD三项都是用来描述单变量数据特征的图形工具，不可以用来分析两个变量之间的相关性。

~~1.在单样本t 检验中，影响所需要样本量大小的因素有哪些?~~



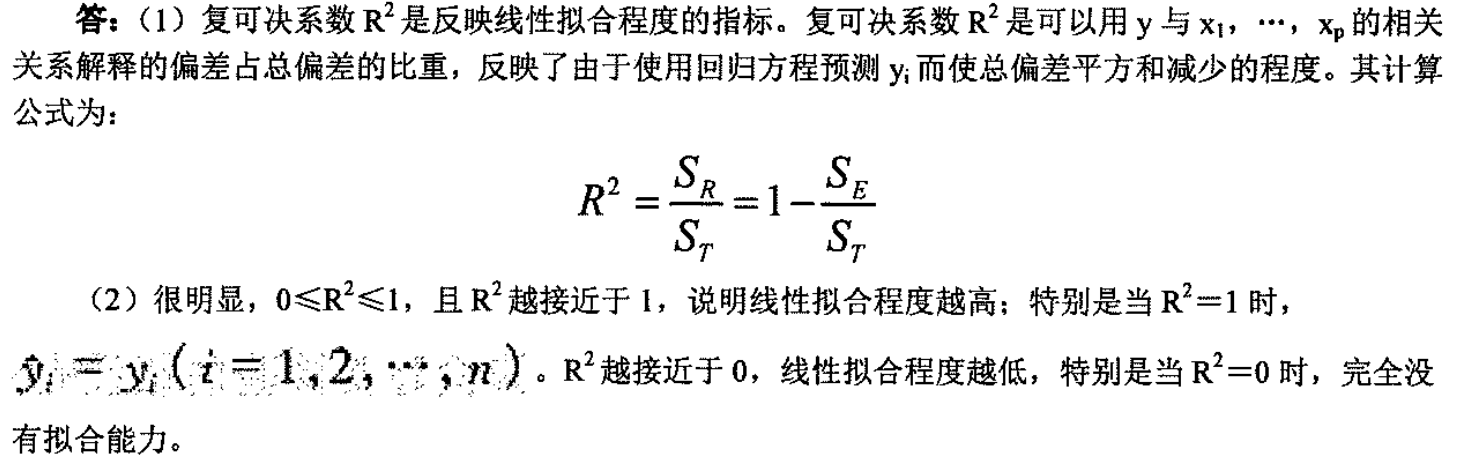
1.正态分布





## 2019

1.多元线性回归分析中，复可决系数R² 的含义是什么?其值通常在什么范围?



线性回归计算(5组身高、体重)。

## 2020

1. 简述抽样方法。

答：几种简单的抽样方法如下：

(1)简单随机抽样，又称完全随机抽样，是在无限总体中进行的无放回独立抽样或在有限总体中进 行的有放回随机抽样。这样得到的样本称为简单随机样本。这种抽样方法在理论上最易处理，但实施时较 难。简单随机抽样通常用样本平均数来估计总体均值，即简单估计。

(2)分层抽样，即将总体中的成员按某种原则划分成若干个子总体，每个子总体称为一个层，在每 层中独立进行简单随机抽样或其他抽样。分层抽样的估计是先对各层进行，然后再综合成总体参数的估计。 分层抽样适用于调查既需要对总体进行估计也需要对局部进行估计的情况。当层内成员差异较小，而层间 成员差异较大时，分层抽样可以提高估计的精度。

(3)整群抽样，即将总体中的成员分为若干群(或组),从这些群(或组)中抽取部分群(或组), 调查对象是被抽中的这些群(或组)中的所有成员。其优点是当被调查单位地理位置比较集中时，实施起 来比较方便；缺点是由于调查单位集中在若干群或组中，而不能均匀地分布在总体的各部分。因此，其准 确性较差，但可以通过适当地多抽样来得到弥补。

(4)系统抽样，又称等距抽样，即将总体中的调查单位按某种次序排列，随机地选定初始单位，然 后按相等的间距抽取其他样本单位。系统抽样的好处是总体各部分都能在一定程度上被包括到样本中，实 施方便；缺点是当初始单位决定之后，样本只有一种组合，不再具有随机性，当次序排列具有周期性时， 容易产生严重的偏差。因此，在进行等距抽样之前，对调查单位的排列次序必须进行细致研究，以避免这 种情况的发生。

## 2021

1. 在总体样本中，按照一定规律排序后，随机抽取首件样品，然后按照一定间距抽取后面的样本，是什么抽样方式?( )

A. 整群抽样

B. 简单随机抽样

C. 系统抽样

D. 分层抽样

【答案】C

【考点】抽样方法

【解析】系统抽样又称等距抽样，是指将总体中的调查单位按某种次序排列， 随机地选定初始单位 然后按相等的间距抽取其他样本单位。这里只有初始单位是随机抽取的，其他样本单位随初始单位的确定 而确定。其好处是总体各部分都能在一定程度上被包括到样本中，实施方便。缺点是当初始单位决定之后， 样本只有一种组合，不再具有随机性，当次序排列具有周期性时，容易产生严重的偏差。

2. 有一车辆检测员，每个车情况都是独立的，质检员每天检有无不合格的车，是什么分布?( )

A. 正态分布

B. 0—1分布

C. 二项分布

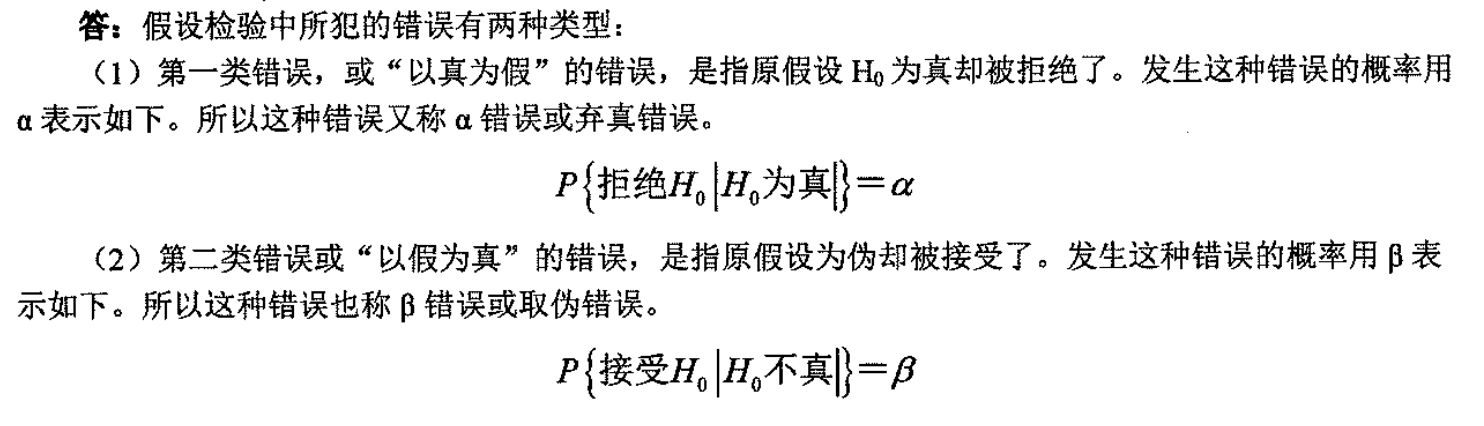
D. 泊松分布

【答案】 B

【考点】常用随机变量以及分布

【解析】只取两个不同数值的随机变量X称为0—1分布。 0—1分布常常被用来描述产品是合格品还是不合格品这类质量检验的结果，或者描述答案只有是与否两种可能的调查，

1.假设检验的两种错误。



## 2022

1.能观测数据是否对称分布、均匀的图形是( )

A 饼形图 B 直方图 C 条形图 D 柱形图

答案：B

解析：应用统计学——第一章数据的整理与图形表示（第一节数据的整理与图形

表示）。

直方图简单、直观，能够反映数据是否呈对称分布，数据的一般水平以及数据的

散步情况。

这个题目也是大纲后面的原题。

1.两个变量之间的相关关系形式有哪些?

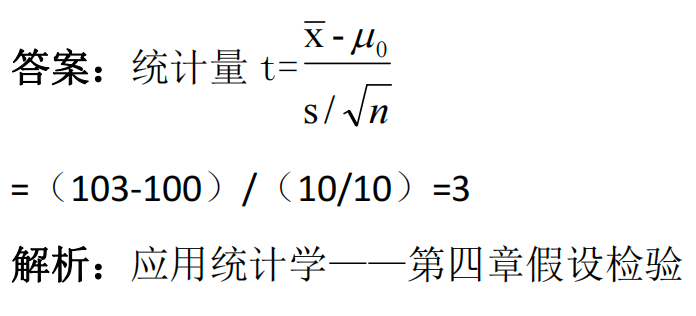
答案：简单线性回归，多元线性回归，非线性，方差序列相关非正太，多重共线。

解析：应用统计学——第五章线性回归分析

显著性题，关键内容为：

总体平均值 100，观测 10 个样本，样本均值 103，样本标准差 10，让算显著性

是否偏移(显著水平 0.05) 给的是 t 分布的值。



## 2023