## 基础课52 统计案例

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **考点考向** | **课标要求** | **真题印证** | **考频热度** | **核心素养** |
| 回归分析 | 掌握 | 2023年天津卷  2022年全国乙卷  2020年全国Ⅱ卷（理） | ★★★ | 数据分析  数学运算  逻辑推理 |
| 独立性检验 | 理解 | 2023年全国甲卷（理）  2023年全国甲卷（文）  2022年新高考Ⅰ卷  2022年全国甲卷（文） | ★★☆ | 数据分析  数学运算  逻辑推理 |
| 命题分析预测 | 从近几年高考的情况来看，对回归分析和独立性检验的考查比较频繁，一般以解答题的形式出现，预测2025年高考还会延续近几年高考的命题特点，特别是以社会现实问题为背景并与概率相结合的统计、统计案例的综合题是今后命题的重点与难点.本基础课易设置多选题，所以在复习备考中，要注意对多选题的训练，做到科学全面备考 | | | |

### 基础知识·诊断

#### 夯实基础

##### 一、散点图与曲线拟合

1*.*对于给定的两个变量*X*和*Y*(如身高和体重),可以把其成对的观测值(*x*1,*y*1),(*x*2,*y*2),…,(*xn*,*yn*)表示为平面直角坐标系中的*n*个点,这些点构成的图称为 ①散点图,每个点对应的一对数据(*xi*,*yi*)(*i=*1,2,3,…,*n*)称为②成对数据*.*

2*.*从散点图上可以看出,如果变量之间存在着某种关系,这些点会有一个大致趋势,这种趋势通常可以用一条光滑的曲线来近似地描述,这样近似描述的过程称为③曲线拟合*.*

3*.*若在两个变量*X*和*Y*的散点图中,所有点看上去都在一条直线附近波动,此时就可以用一条直线来近似地描述这两个量之间的关系,称之为④直线拟合*.*

##### 二、一元线性回归方程

直线方程*Y=+X*称作*Y*关于*X*的线性回归方程,相应的直线称作*Y*关于*X*的回归直线,,是这个线性回归方程的系数,*===*,*=-.*

##### 三、相关系数

1*.r=.*

2*.*当*r>*0时,称成对样本数据⑤正相关;当*r<*0时,称成对样本数据⑥负相关*.*

3*.|r|*≤1;当*|r|*越接近1时,成对样本数据的线性相关程度越⑦强;当*|r|*越接近0时,成对样本数据的线性相关程度越⑧弱*.*

##### 四、2×2列联表与独立性检验

1*.*2*×*2列联表:设*A*,*B*为两个变量,每一个变量都可以取两个值,其2*×*2列联表为

|  |  |  |  |
| --- | --- | --- | --- |
| *B*  *A* | *B*1 | *B*2 | 总计 |
| *A*1 | *a* | *b* | *a+b* |
| *A*2 | *c* | *d* | *c+d* |
| 总计 | *a+c* | *b+d* | *n=a+b+c+d* |

2*.*独立性检验

*χ*2的计算公式:*χ*2*=*,在统计中,用以下结果对变量的独立性进行判断:

(1)当*χ*2≤2*.*706时,没有充分的证据判断变量*A*,*B*有关联,可以认为变量*A*,*B*是没有关联的;

(2)当*χ*2*>*2*.*706时,有⑨90%的把握判断变量*A*,*B*有关联;

(3)当*χ*2*>*3*.*841时,有⑩95%的把握判断变量*A*,*B*有关联;

(4)当*χ*2*>*6*.*635时,有99%的把握判断变量*A*,*B*有关联*.*

#### 诊断自测

##### 题组1 走出误区

1. 判一判.（对的打“√”,错的打“×”）

（1） “名师出高徒”可以解释为教师的教学水平与学生的水平呈正相关.( √ )

（2） 通过线性回归方程可以估计预报变量的取值和变化趋势.( √ )

（3） 只有两个变量有相关关系，所得到的回归模型才有预测价值.( √ )

（4） 若事件，关系越密切，则由观测数据计算得到的的值越大.( √ )

2. （易错题）从某大学中随机选取8名女大学生,其身高（单位： ）与体重（单位： ）的数据如表所示：

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 165 | 165 | 157 | 170 | 175 | 165 | 155 | 170 |
|  | 48 | 57 | 50 | 54 |  | 61 | 43 | 59 |

已知表中有一个数据看不清，记作，由最小二乘法求得关于的线性回归方程，则当选取的女大学生身高为时,64.

**【易错点】**忽视回归直线过，从而直接代入175求的值而致误.

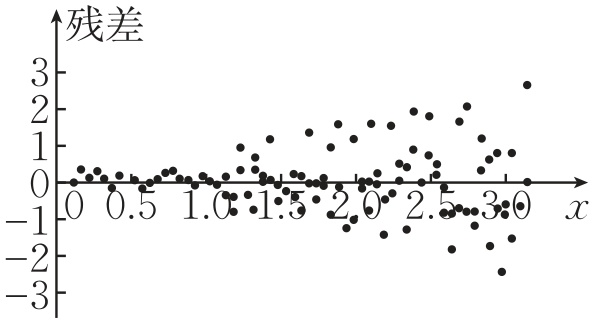
[解析]由表格中数据知，点在回归直线上，代入回归方程得，，解得.

##### 题组2 走进教材

3. （双空题）（人教A版选修③P103·练习T2改编）已知变量和变量的3对随机观测数据为，，，两个变量的样本相关系数，且不能（填“能”或“不能”）据此推出这两个变量线性相关参考公式：

[解析]，，则样本相关系数，虽然样本相关系数为，三个样本点在一条直线上，但是由于样本量太小，据此推断两个变量完全线性相关并不可靠.

4. （人教A版选修③P138·T2改编）根据变量和变量的成对样本观测数据，由一元线性回归模型得到经验回归模型，对应的残差如图所示，模型误差( C ).



A. 满足一元线性回归模型的所有假设

B. 不满足一元线性回归模型的的假设

C. 不满足一元线性回归模型的的假设

D. 不满足一元线性回归模型的和的假设

[解析]由一元线性回归模型得到经验回归模型，根据对应的残差图，残差可能成立，但明显残差在轴上方的数据更分散，所以不满足一元线性回归模型.故选.

##### 题组3 走向高考

5. [2023·全国甲卷节选]一项试验旨在研究臭氧效应，试验方案如下：选40只小白鼠，随机地将其中20只分配到实验组，另外20只分配到对照组，实验组的小白鼠饲养在高浓度臭氧环境，对照组的小白鼠饲养在正常环境，一段时间后统计每只小白鼠体重的增加量（单位：）.经计算，40只小白鼠体重的增加量的中位数为，且两样本中小于23.4与不小于23.4的个数的列联表如下：

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | 合计 |
| 对照组 | 6 | 14 | 20 |
| 实验组 | 14 | 6 | 20 |
| 合计 | 20 | 20 | 40 |

问：能否有的把握认为小白鼠在高浓度臭氧环境中与正常环境中体重的增加量有差异?

附：.

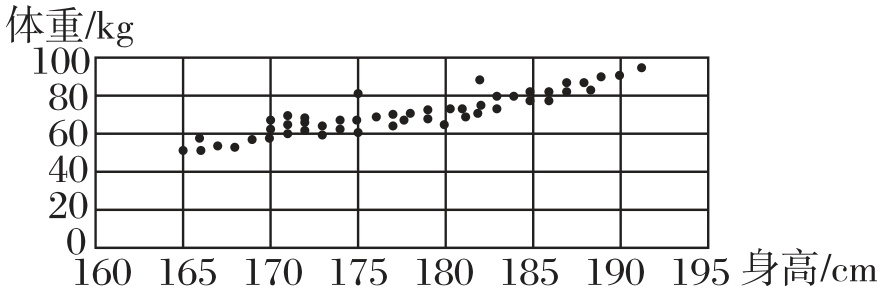
|  |  |  |  |
| --- | --- | --- | --- |
|  | 0.100 | 0.050 | 0.010 |
|  | 2.706 | 3.841 | 6.635 |

[解析]由公式代入数据计算可得，所以能有的把握认为小白鼠在高浓度臭氧环境中与正常环境中体重的增加量有差异.

### 考点聚焦·突破

#### 考点一 成对数据的相关性［自主练透］

1. 已知某校50名学生的身高和体重的散点图如图所示，则下列说法正确的是( C ).

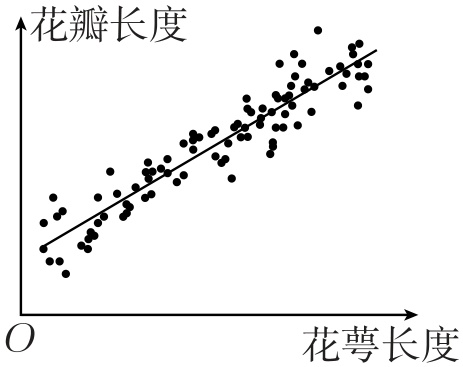


A. 身高越高，体重越重 B. 身高越高，体重越轻

C. 身高与体重成正相关 D. 身高与体重成负相关

[解析]由图可知，身高与体重成正相关，故正确，错误.但身高与体重不成函数关系，故,错误.故选.

2. [2023·天津卷]调查某种群花萼长度和花瓣长度，所得数据如图所示，其中样本相关系数，则下列说法正确的是( C ).



A. 花瓣长度和花萼长度没有相关性

B. 花瓣长度和花萼长度呈现负相关

C. 花瓣长度和花萼长度呈现正相关

D. 若从样本中抽取一部分，则这部分的样本相关系数一定是0.8245

[解析]根据散点的集中程度可知，花瓣长度和花萼长度有相关性，错误；散点的分布是从左下到右上，所以花瓣长度和花萼长度呈现正相关性，错误，正确；由于是全部数据的样本相关系数，取出来一部分数据，相关性可能变强，可能变弱，即取出的数据的样本相关系数不一定是，错误.故选.

3. （多选题）技术已经进入高速发展的阶段，手机的销量也逐渐上升，某手机商城统计了近5个月来手机的实际销量如表所示：

|  |  |  |
| --- | --- | --- |
| 月份 | 月份编号 | 销量千部 |
| 2023年2月 | 1 | 37 |
| 2023年3月 | 2 | 104 |
| 2023年4月 | 3 |  |
| 2023年5月 | 4 | 196 |
| 2023年6月 | 5 | 216 |

若与线性相关，且求得线性回归方程为，则下列说法正确的是( AB ).

A.

B. 与正相关

C. 与的相关系数为负数

D. 8月份该手机商城的手机销量约为36.5万部

[解析]由表中数据，计算得，所以，

于是，解得，故正确；

由线性回归方程中的的系数为正可知，与正相关，且其样本相关系数，故正确，错误；8月份时，，（千部），即销量约为32万部，故错误.故选.



**判断两个变量相关性的三种方法**

|  |  |
| --- | --- |
| 画散点图 | 点的分布从左下角到右上角，两个变量正相关；点的分布从左上角到右下角，两个变量负相关 |
| 相关系数 | 当时，正相关；当时，负相关；越趋近于1，线性相关性越强 |
| 线性回归方程 | 当时，正相关；当时，负相关 |

#### 考点二 回归分析［多维探究］

##### 线性回归分析角度1

典例1 某高三实验班共有60名学生参加某次考试，从中随机挑选出5名学生，他们的数学成绩与物理成绩的统计数据如表所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 数学成绩分 | 145 | 130 | 120 | 105 | 100 |
| 物理成绩分 | 110 | 90 | 102 | 78 | 70 |

数据表明与之间有较强的线性相关关系.

（1）求关于的线性回归方程；

（2）该班一名学生的数学成绩为110分，利用（1）中的线性回归方程，估计该学生的物理成绩.

参考公式:,.

[解析]（1）由题意，可知，，

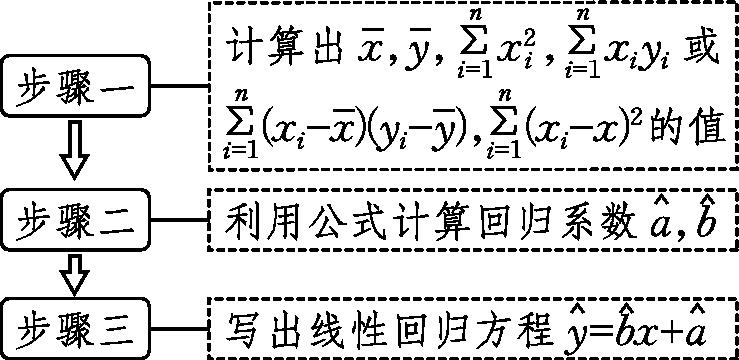
故，

，故线性回归方程为.

（2）当时，，故该学生的物理成绩约为82分.



**求线性回归方程的三个步骤**



##### 非线性回归分析角度2

典例2 [2024·深圳模拟]数独是源自18世纪瑞士的一种数学游戏，玩家需要根据盘面上的已知数字，推理出所有剩余空格的数字，并满足每一行、每一列、每一个粗线宫内的数字均含，不重复.数独爱好者小明打算报名参加“丝路杯”全国数独大赛初级组的比赛，赛前小明在某数独上进行一段时间的训练，每天的解题平均速度（单位：秒）与训练天数（单位：天）有关，经统计得到的数据如表所示：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 天 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 秒 | 990 | 990 | 450 | 320 | 300 | 240 | 210 |

（1）现用作为回归方程模型，请利用表中数据，求出该线性回归方程；

（2）请用第（1）题的结论预测，小明经过100天训练后，每天解题的平均速度约为多少秒？

参考数据其中：

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 1845 | 0.37 | 0.55 |

参考公式：对于一组数据，， ，，其回归直线的斜率和截距的最小二乘估计公式分别为：，.

[解析]（1）由题意，.令，设关于的线性回归方程为，

则有，

则，所以.

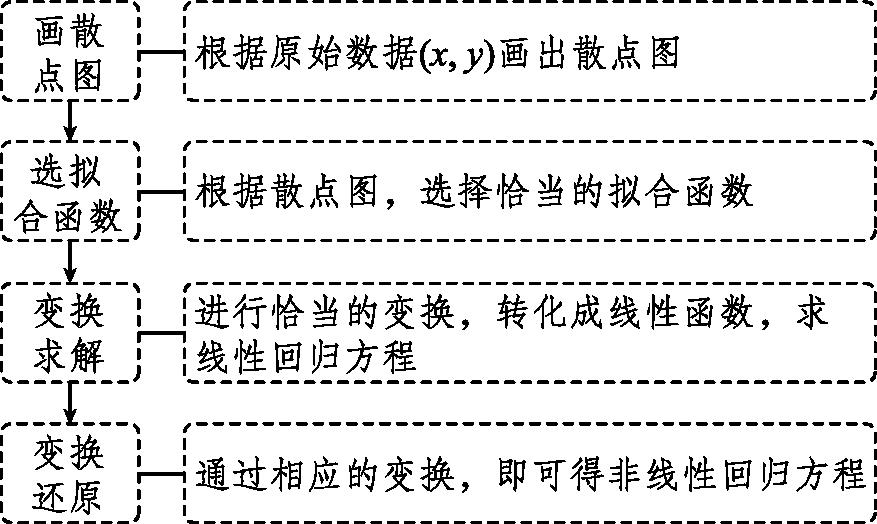
又因为，所以关于的线性回归方程为.

（2）当时，，

所以经过100天训练后，小明每天解题的平均速度约为140秒.



有些非线性回归分析问题并不给出公式，这时我们可以画出已知数据的散点图，把它与学过的各种函数（幂函数、指数函数、对数函数等）的图象进行比较，挑选一种跟散点图拟合得最好的函数模型，用适当的变量进行变换，把问题化为一元线性回归分析问题，使之得到解决.其一般步骤为：



##### 多维训练

1. [2024·安庆模拟]（多选题）某市物价部门对5家商场的某商品一天的销售量及其价格进行调查，5家商场的售价（单位：元）和销售量（单位：件）之间的一组数据如表所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 售价元 | 9 | 9.5 | 10 | 10.5 | 11 |
| 销售量件 | 11 | 10 | 8 | 6 | 5 |

根据表中数据得到关于的线性回归方程是，则下列说法正确的是( ABC ).

A. B. 回归直线过点

C. 当时，的估计值为12.8 D. 点处的随机误差为0.4

[解析]由表中数据得，，故回归直线过点，且，解得，故，正确.当时，，故正确.点处的随机误差为，故错误.故选.

2. [2024·珠海模拟]中国茶文化博大精深，已知茶水的口感与茶叶类型以及水温有关.经验表明，某种绿茶用的水泡制，再等到茶水温度降至时饮用，可以达到最佳口感.某学习研究小组通过测量，得到了下面表格中的数据（室温是）.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 泡制时间 | 0 | 1 | 2 | 3 | 4 |
| 水温 | 85 | 79 | 74 | 71 | 65 |
|  | 4.2 | 4.1 | 4.0 | 3.9 | 3.8 |

（1）小组成员根据上面表格中的数据绘制散点图，并根据散点图分布情况，考虑到茶水温度降到室温（即）就不能再降的事实，决定选择函数模型来刻画.

①令，求出关于的线性回归方程；

②利用①的结论，求出中的与.

（2）你认为该品种绿茶用的水大约泡制多久后饮用，可以达到最佳口感？

参考数据：,,,.参考公式：,,.

[解析]（1）①设线性回归方程为，

由题意，得，，

，

，

则，

，

则关于的线性回归方程为.

②由，得，

两边取对数得，

利用①的结论得,，

，.

（2）由（1）得，

令，得，

该品种绿茶用的水泡制后饮用，口感最佳.

#### 考点三 独立性检验［师生共研］

典例3 某企业生产的产品按质量分为一等品和二等品，该企业计划对现有的生产设备进行改造，为了分析设备改造前后的效果，现从设备改造前后生产的大量产品中各抽取200件产品作为样本，产品的质量情况统计结果如表所示：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 一等品 | 二等品 | 合计 |
| 设备改造前 | 120 | 80 | 200 |
| 设备改造后 | 150 | 50 | 200 |
| 合计 | 270 | 130 | 400 |

附：,

|  |  |  |  |
| --- | --- | --- | --- |
|  | 0.050 | 0.010 | 0.001 |
|  | 3.841 | 6.635 | 10.828 |

依据的独立性检验，能否认为该企业生产的这种产品的质量与设备改造有关?

[解析]零假设为质量指标值与设备改造无关.

，

不成立，依据的独立性检验，能认为该企业生产的这种产品的质量指标值与设备改造有关.



**独立性检验的一般步骤**

1.根据样本数据制成列联表.

2.根据公式计算.

3.比较与临界值的大小关系，作统计推断.

##### 针对训练

[2022·全国甲卷]甲、乙两城之间的长途客车均由和两家公司运营，为了解这两家公司长途客车的运行情况，随机调查了甲、乙两城之间的500个班次，得到下面列联表：

|  |  |  |
| --- | --- | --- |
|  | 准点班次数 | 未准点班次数 |
| A | 240 | 20 |
| B | 210 | 30 |

（1）根据上表，分别估计这两家公司甲、乙两城之间的长途客车准点的概率；

（2）能否有的把握认为甲、乙两城之间的长途客车是否准点与客车所属公司有关？

附：，

|  |  |  |  |
| --- | --- | --- | --- |
|  | 0.100 | 0.050 | 0.010 |
|  | 2.706 | 3.841 | 6.635 |

[解析]（1）根据表中数据可知，公司共有班次260次，准点班次有240次，

设“公司长途客车准点”为事件，则.

公司共有班次240次，准点班次有210次，

设“公司长途客车准点”为事件，则.

故公司长途客车准点的概率为，公司长途客车准点的概率为.

（2）列联表：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 准点班次数 | 未准点班次数 | 合计 |
|  | 240 | 20 | 260 |
|  | 210 | 30 | 240 |
| 合计 | 450 | 50 | 500 |

，根据临界值表可知，有的把握认为甲、乙两城之间的长途客车是否准点与客车所属公司有关.