## 基础课52 统计案例

### 课时评价·提能

#### 基础巩固练

1. 对两个变量，进行线性相关检验，得线性相关系数，对两个变量，进行线性相关检验，得线性相关系数，则下列判断正确的是（ C ）.

A. 变量与正相关，变量与负相关，变量与的线性相关性较强

B. 变量与负相关，变量与正相关，变量与的线性相关性较强

C. 变量与正相关，变量与负相关，变量与的线性相关性较强

D. 变量与负相关，变量与正相关，变量与的线性相关性较强

[解析]由线性相关系数 知 与 正相关，

由线性相关系数 知 与 负相关，

又，所以变量 与 的线性相关性比 与 的线性相关性强，故选.

2. 甲、乙、丙、丁四位同学各自对，两个变量的线性相关性做试验，并用回归分析方法分别求得样本相关系数与残差平方和，如表所示，

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 甲 | 乙 | 丙 | 丁 |
|  | 0.82 | 0.78 | 0.69 | 0.85 |
|  | 106 | 115 | 124 | 103 |

则（ D ）同学的试验结果体现，两个变量有更强的线性相关性.

A. 甲 B. 乙 C. 丙 D. 丁

[解析] 越大，越小，线性相关性越强.故选.

3. 给出下列说法：①回归直线恒过样本点的中心，且至少过一个样本点；②两个变量的相关性越强，则样本相关系数就越接近1；③将一组数据的每个数据都加一个相同的常数后，方差不变；④在线性回归方程中，当解释变量增加1个单位时，预报变量平均减少0.5个单位.其中说法正确的是（ B ）.

A. ①②④ B. ②③④ C. ①③④ D. ②④

[解析]对于①，回归直线 恒过样本点的中心，但不一定过一个样本点，所以①不正确；

对于②，根据样本相关系数的意义，可得两个变量相关性越强，则样本相关系数 就越接近1，所以②正确；

对于③，根据方差的计算公式，可得将一组数据的每个数据都加一个相同的常数后，方差是不变的，所以③正确；

对于④，根据回归系数的含义，可得在线性回归方程 中，当解释变量 增加1个单位时，预报变量 平均减少0.5个单位，所以④正确.故选.

4. （改编）下列说法不正确的是（ D ）.

A. 在回归分析中，残差图的带状区域越窄，模型的拟合效果越好

B. 若一组观测值，， ，满足，且，则残差恒为0

C. 回归分析是对具有相关关系的两个变量进行统计分析的一种常用方法

D. 在研究成年人的身高和体重的关系画残差图时，纵坐标为残差，横坐标一定是编号

[解析]对于，在回归分析中，残差图的带状区域越窄，残差平方和越小，则 的值越大，说明模型的拟合效果越好，正确；

对于，若一组观测值，， ，满足，且，则残差平方和为0，所以残差 恒为0，正确；

对于，回归分析是对具有相关关系的两个变量进行统计分析的一种常用方法，正确；

对于，残差图中横坐标可以是样本编号，也可以是身高数据，还可以是体重的估计值等，错误.故选.

5. “拃”是我国古代的一种长度单位，最早见于金文时代，“一拃”指张开大拇指和中指两端间的距离.某数学兴趣小组为了研究右手一拃长（单位：厘米）和身高（单位：厘米）的关系，从所在班级随机抽取了15名学生，根据测量数据的散点图发现和具有线性相关关系，其线性回 归方程为，且，.已知小明的右手一拃长为20厘米，据此估计小明的身高为（ B ）.

A. 187厘米 B. 183厘米 C. 179厘米 D. 175厘米

[解析]由题意知，，，

又，即，解得，

故线性回归方程为，

当 时，，即当小明的右手一拃长为20厘米时，可估计小明的身高为183厘米.故选.

6. （改编）某种热带观赏鱼的生产周期大约为一个月，某热带鱼爱好者购买了两条该种热带鱼（雌雄各一条），一个月后开始生产，月份与热带鱼的条数如表所示.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 月数 | 1 | 2 | 3 | 4 | 5 |
| 热带鱼的条数 | 2 | 17 | 42 | 103 | 136 |

由表格可得关于的线性回归方程为，则此回归模型第5个月的残差（实际值与预报值之差）为（ C ）.

A. 13 B. C. D. 8

[解析]设，则，，

，所以.

令，得.故选.

7. 某种细胞的存活率单位：与存放温度单位：之间具有线性相关关系，其样本数据如表所示.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 存放温度 | 20 | 15 | 10 | 5 | 0 |  |  |
| 存活率 | 6 | 14 | 26 | 33 | 43 | 60 | 63 |

计算得，，，，并求得线性回归方程，但实验人员发现表中数据的对应值录入有误，更正为,则更正后的线性回归方程为（ A ）.

A. B. C. D.

[解析]依题意，设更正后的线性回归方程为，更正后，，，，，

，所以更正后的线性回归方程为.故选.

8. 足球运动是深受学生喜爱的一项体育运动，为了研究是否喜爱足球运动与学生性别的关系，从某高校男、女生中各随机抽取80名学生进行调查问卷，得到数据如表所示：

|  |  |  |
| --- | --- | --- |
|  | 喜爱 | 不喜爱 |
| 男生 |  |  |
| 女生 |  |  |

若根据小概率值的独立性检验，可认为是否喜爱足球运动与学生性别有关，则的最小值为（ B ）.

附：，其中.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 0.25 | 0.10 | 0.05 | 0.01 |
|  | 2.072 | 2.706 | 3.841 | 6.635 |

A. 17 B. 15 C. 13 D. 11

[解析]因为根据小概率值 的独立性检验，可认为是否喜爱足球运动与学生性别有关，

所以，

即，因为 在，时单调递增，且，，

所以 的最小值为15.故选.

#### 综合提升练

9. （多选题）由样本数据组成的一个样本得到线性回归方程，且，去除两个样本点和后，得到新的线性回归方程.在余下的8个样本数据和新的线性回归方程中，下列结论正确的是（ AB ）.

A. 相关变量，具有正相关关系

B. 新的回归方程为

C. 当回归方程确定时，随着自变量值的增加，因变量值增加的速度变大

D. 样本的残差为1

[解析]由题意可知，，

所以8个样本数据的 新平均数为，新平均数为，

所以，所以.

新的线性回归方程，，具有正相关关系，，正确；

由线性回归方程知，随着自变量 值增加，因变量 值增加的速度恒为3，错误；

当 时，，残差为，错误.

故选.

10. （多选题）某同学在家独自用表格分析高三前五次月考中数学的班级排名与考试次数的相关性时，忘记了第二次和第四次月考排名，但该同学记得平均排名，于是分别用和得到了线性回归方程和，对应的样本相关系数分别为，，排名对应的方差分别为，，则下列结论正确的是（ BD ）.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
|  | 10 |  | 6 |  | 2 |

附：，，.

A. B. C. D.

[解析]当 时，,，解得，

则，

，，

，

，

，

所以，

得，

，

.

同理，当 时，,,，.

所以,,,.故选.

11. 甲、乙两个班级进行数学考试，按学生考试及格与不及格统计成绩后的列联表（数据有缺失）如表所示，则0.56.（保留到小数点后两位）

|  |  |  |  |
| --- | --- | --- | --- |
|  | 不及格 | 及格 | 总计 |
| 甲班 | 12 |  | 45 |
| 乙班 |  | 36 |  |
| 总计 | 21 |  |  |

参考公式：

，.

[解析]由题意，得到 列联表,

|  |  |  |  |
| --- | --- | --- | --- |
|  | 不及格 | 及格 | 总计 |
| 甲班 | 12 | 33 | 45 |
| 乙班 | 9 | 36 | 45 |
| 总计 | 21 | 69 | 90 |

则.

12. （双空题）根据某市有关统计公报，该市2019年至2022年每年进口总额（单位：千亿元）和出口总额（单位：千亿元）之间的一组数据如表所示：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 2019年 | 2020年 | 2021年 | 2022年 |
|  | 1.8 | 2.2 | 2.6 | 3.0 |
|  | 2.0 | 2.8 | 3.2 | 4.0 |

若每年的进出口总额，满足线性回归方程，则1.6；若计划2023年出口总额达到5千亿元，预计该年进口总额为3.65千亿元.

[解析]由数表得，，

因此，回归直线 过点（2.4,3），由，解得，

此时，，当 时，，解得，

预计该年进口总额为3.65千亿元.

#### 应用情境练

13. 某奶茶店统计某月一周奶茶销售的杯数，用表示星期一至星期天，表示每天销售奶茶的杯数，统计数据如表所示.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|  | 6 | 11 | 21 | 34 | 66 | 101 | 196 |

根据散点图判断，销售奶茶的杯数关于天数的线性回归方程适合用来表示，则其线性回归方程为  .

参考数据：设，，，.

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]由，两边同时取对数可得.

设，则，

因为，，

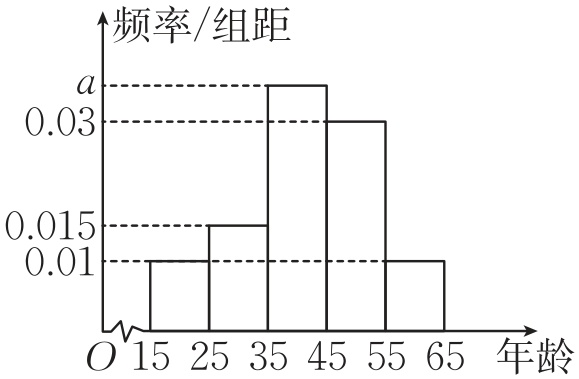
，

所以，

，

所以，所以.

14. 某调查网站从观看央视春晚的观众中随机选出200人（年龄在15岁到65岁之间），经统计，这200人中通过传统的传媒方式电视端口观看的人数与通过新型的传媒端口观看的人数之比为.将这200人按年龄分组：第1组，第2组，第3组，第4组，第5组.其中统计通过传统的传媒方式电视端口观看的观众得到的频率分布直方图如图所示.



（1）求的值及通过传统的传媒方式电视端口观看的观众的平均年龄；

（2）把年龄在第1，2，3组的观众称青少年组，年龄在第4，5组的观众称为中老年组，若选出的200人中通过新型的传媒方式端口观看的中老年人有12人，请完成下面的列联表，则依据的独立性检验，能否认为观看央视春晚的方式与年龄有关？

附：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 通过端口观看 | 通过电视端口观看 | 合计 |
| 青少年 |  |  |  |
| 中老年 |  |  |  |
| 合计 |  |  |  |

其中.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 0.10 | 0.05 | 0.025 | 0.010 | 0.005 | 0.001 |
|  | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 | 10.828 |

[解析]（1）由频率分布直方图可得，

解得，

所以通过传统的传媒方式电视端口观看的观众的平均年龄为.

（2）通过电视端口观看的有（人），通过 端口观看的有（人），

由频率分布直方图可得，通过电视端口观看的中老年有（人），通过 端口观看的中老年人有12人，

所以通过 端口观看的青少年有（人），

可得 列联表为

|  |  |  |  |
| --- | --- | --- | --- |
|  | 通过 端口观看 | 通过电视端口观看 | 合计 |
| 青少年 | 28 | 96 | 124 |
| 中老年 | 12 | 64 | 76 |
| 合计 | 40 | 160 | 200 |

计算得，

所以依据小概率值 的独立性检验，不能认为观看央视春晚的方式与年龄有关.

#### 创新拓展练

15. 针对“中学生追星问题”，某校团委对“学生性别和中学生追星是否有关”作了一次调查，其中女生人数是男生人数的，男生追星的人数占男生人数的，女生追星的人数占女生人数的，若根据小概率值的独立性检验，认为中学生追星与性别有关，则男生至少有30人.

参考数据及公式如下：

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

，.

[解析]设男生人数为，依题意可得列联表为

|  |  |  |  |
| --- | --- | --- | --- |
|  | 喜欢追星 | 不喜欢追星 | 总计 |
| 男生 |  |  |  |
| 女生 |  |  |  |
| 总计 |  |  |  |

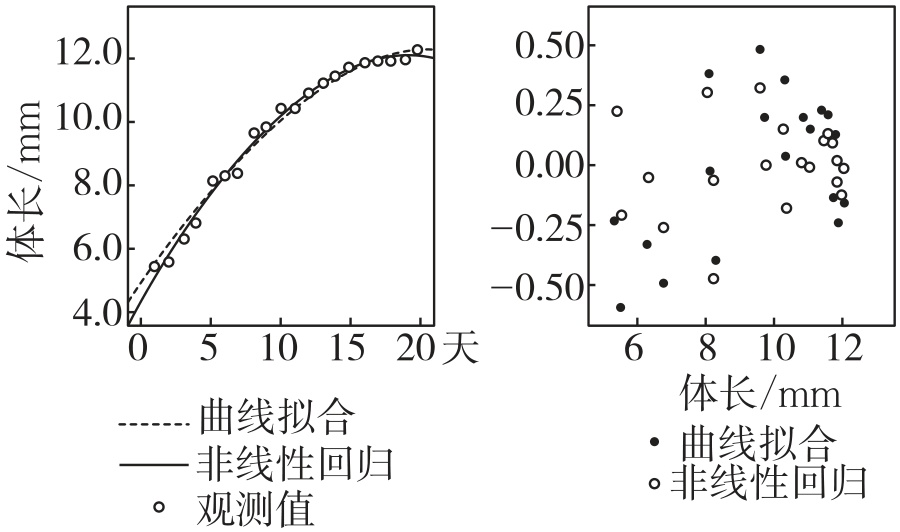
若根据小概率值 的独立性检验，认为是否喜欢追星和性别有关，则，

由，解得，

由题意知,应为6的整数倍，

若根据小概率值 的独立性检验，认为是否喜欢追星和性别有关，则男生至少有30人.

16. 某院校研究小组以当地某水产养殖基地的黄河鲤仔鱼为研究对象,从出卵开始持续观察20天,试验期间,每天固定时段从试验水体中随机取出同批次9尾黄河鲤仔鱼测量体长,取其均值作为第天的观测值单位：,其中.根据以往的统计资料,该组数据可以用曲线拟合模型或非线性回归模型进行统计分析,其中,,为参数.基于这两个模型,绘制得到如下的散点图和残差图：



（1）你认为哪个模型的拟合效果更好?分别结合散点图和残差图进行说明.

（2）假定,且黄河鲤仔鱼的体长与天数具有很强的相关关系.现对数据进行初步处理,得到如下统计量的值：,,,,,,其中,,根据（1）的判断结果及给定数据,求关于的线性回归方程,并预测第22天时黄河鲤仔鱼的体长.（结果精确到小数点后2位）

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

[解析]（1）非线性回归模型 的拟合效果更好.

从散点图看,散点更均匀地分布在该模型拟合曲线附近;从残差图看,该模型下的残差更均匀地集中在以残差为0的直线为对称轴的水平带状区域内.

（2）可化为,即,

两边取对数得,

则,

即,

.

所以 关于 的线性回归方程为.

当 时,体长.

所以估计第22天时仔鱼的体长为.