**回归分析的基本思想及其初步应用题库及答案-高中数学选修2-3第三章**

****

一、选择题

1．(2014·重庆理，3)已知变量*x*与*y*正相关，且由观测数据算得样本平均数＝3，＝3.5，则由该观测数据算得线性回归方程可能为(　　)

A．＝0.4*x*＋2.3　　　 B．＝2*x*－2.4

C．＝－2*x*＋9.5 D．＝－0.3*x*＋4.4

[答案]　A

[解析]　因为变量*x*和*y*正相关，所以回归直线的斜率为正，排除C、D；又将点(3,3.5)代入选项A和B的方程中检验排除B，所以选A.

2．(2014·枣阳一中、襄州一中、宣城一中、曾都一中高三期中联考)由变量*x*与*y*相对应的一组数据(1，*y*1)、(5，*y*2)、(7，*y*3)、(13，*y*4)、(19，*y*5)得到的线性回归方程为＝2*x*＋45，则＝(　　)

A．135　　 B．90

C．67　　 D．63

[答案]　D

[解析]　∵＝(1＋5＋7＋13＋19)＝9，＝2＋45，

∴＝2×9＋45＝63，故选D.

3．(2014·淄博市、临淄区学分认定考试)观测两个相关变量，得到如下数据：

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *x* | －1 | －2 | －3 | －4 | －5 | 5 | 4 | 3 | 2 | 1 |
| *y* | －0.9 | －2 | －3.1 | －3.9 | －5.1 | 5 | 4.1 | 2.9 | 2.1 | 0.9 |

则两变量之间的线性回归方程为(　　)

A．＝0.5*x*－1 B．＝*x*

C．＝2*x*＋0.3 D．＝*x*＋1

[答案]　B

[解析]　因为＝0，

＝＝0，根据回归直线方程必经过样本中心点(，)可知，回归直线方程过点(0,0)，所以选B.

4．(2013·济宁梁山一中高二期中)一位母亲记录了儿子3～9岁的身高，数据(略)，由此建立的身高与年龄的回归模型为＝7.19*x*＋73.93，用这个模型预测这个孩子10岁时的身高，则正确的叙述是(　　)

A．身高一定是145.83cm B．身高在145.83cm以上

C．身高在145.83左右 D．身高在145.83cm以下

[答案]　C

[解析]　将*x*的值代入回归方程＝7.19*x*＋73.93时，得到的值是年龄为*x*时，身高的估计值，故选C.

5．对于回归分析，下列说法错误的是(　　)

A．在回归分析中，变量间的关系是非确定性关系，因此因变量不能由自变量唯一确定

B．线性相关系数可以是正的或负的

C．回归分析中，如果*r*＝±1，说明*x*与*y*之间完全线性相关

D．样本相关系数*r*∈(－1,1)

[答案]　D

[解析]　∵相关系数|*r*|≤1，∴D错．

6．甲、乙、丙、丁四位同学各自对*A*、*B*两变量的线性相关性做试验，并用回归分析方法分别求得相关系数*r*与残差平方和*m*如下表：

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

则哪位同学的试验结果体现*A*、*B*两变量有更强的线性相关性(　　)

A．甲　 　 B．乙

C．丙　 　 D．丁

[答案]　D

[解析]　*r*越接近1，相关性越强，残差平方和*m*越小，相关性越强，故选D.

二、填空题

7．已知回归直线方程为＝0.50*x*－0.81，则*x*＝25时，*y*的估计值为\_\_\_\_\_\_\_\_．

[答案]　11.69

8．在7块并排、形状大小相同的试验田上进行施化肥量对水稻产量影响的试验，得到如下表所示的一组数据(单位：kg)．由散点图初步判定其具有线性相关关系，则由此得到的回归方程的斜率是\_\_\_\_\_\_\_\_.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 施化肥量*x* | 15 | 20 | 25 | 30 | 35 | 40 | 45 |
| 水稻产量*y* | 330 | 345 | 365 | 405 | 445 | 450 | 455 |

[答案]　4.75

[解析]　列表如下，

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *i* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| *xi* | 15 | 20 | 25 | 30 | 35 | 40 | 45 |
| *yi* | 330 | 345 | 365 | 405 | 445 | 450 | 455 |
| *xiyi* | 4950 | 6900 | 9125 | 12150 | 15575 | 18000 | 20475 |
| ＝30，≈399.3，＝7000，*iyi*＝87175 | | | | | | | |

则≈≈4.75.

回归方程的斜率即回归系数.

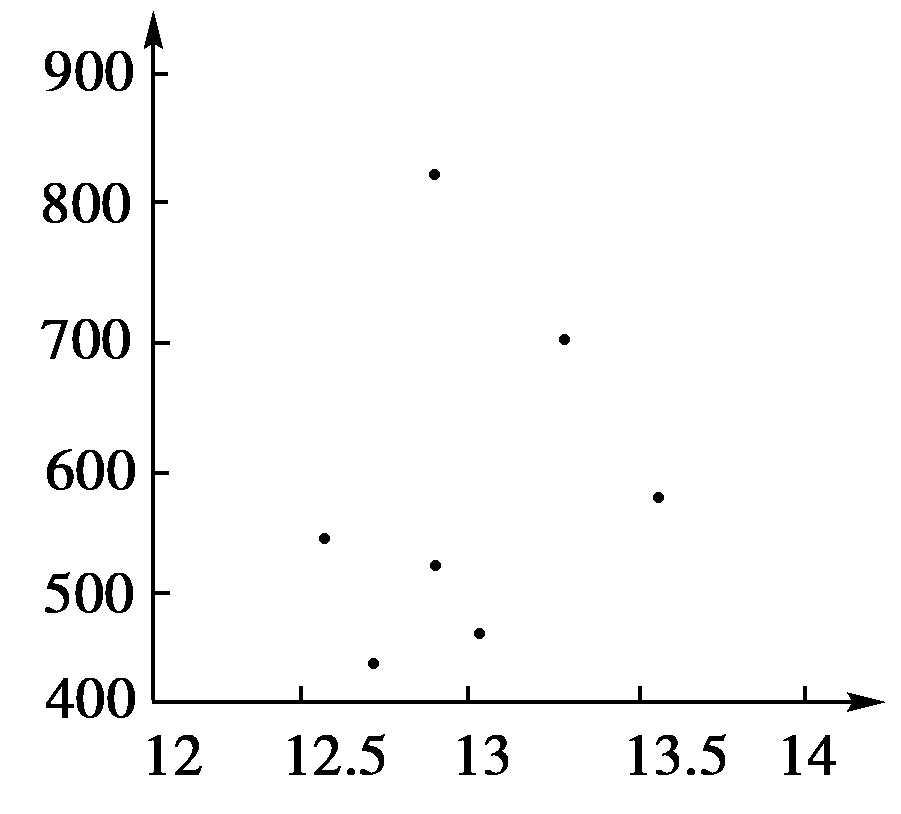
9．以下是某地区的降雨量与年平均气温的一组数据：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 年平均气温(℃) | 12.51 | 12.84 | 12.84 | 13.69 | 13.33 | 12.74 | 13.05 |
| 年降雨量(mm) | 542 | 507 | 813 | 574 | 701 | 432 | 464 |

根据这组数据可以推断，该地区的降雨量与年平均气温\_\_\_\_\_\_\_\_相关关系．(填“具有”或“不具有”)

[答案]　不具有

[解析]　画出散点图，观察可知，降雨量与年平均气温没有相关关系．



三、解答题

10．某工厂的产品产量与单位成本的资料如下表所示，请进行线性回归分析.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 月份 | 产量*x*(千件) | 单位成本*y*(元/件) | *x*2 | *xy* |
| 1 | 2 | 73 | 4 | 146 |
| 2 | 3 | 72 | 9 | 216 |
| 3 | 4 | 71 | 16 | 284 |
| 4 | 3 | 73 | 9 | 219 |
| 5 | 4 | 69 | 16 | 276 |
| 6 | 5 | 68 | 25 | 340 |
| 合计 | 21 | 426 | 79 | 1 484 |

[解析]　设回归直线方程为＝*x*＋，

＝，＝＝71，＝79，i*y*i＝1 481，

∴＝＝≈－1.818 2，

＝71－(－1.818 2)×≈77.36.

回归直线方程为＝77.36－1.818 2*x*.

由回归系数为－1.818 2知，产量每增加1 000件，单位成本下降约1.82元．

****

一、选择题

11．(2014·哈师大附中高二期中)下列说法正确的有几个(　　)

(1)回归直线过样本点的中心(，)；

(2)线性回归方程对应的直线＝*x*＋至少经过其样本数据点(*x*1，*y*1)、(*x*2，*y*2)、…、(*xn*，*yn*)中的一个点；

(3)在残差图中，残差点分布的带状区域的宽度越宽，其模型拟合的精度越高；

(4)在回归分析中，*R*2为0.98的模型比*R*2为0.80的模型拟合的效果好．

A．1　　　 B．2

C．3　　　 D．4

[答案]　B

[解析]　由回归分析的概念知①④正确，②③错误．

12．下表提供了某厂节能降耗技术改造后在生产*A*产品过程中记录的产量*x*(吨)与相应的生产能耗*y*(吨)的几组对应数据：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *x* | 3 | 4 | 5 | 6 |
| *y* | 2.5 | *t* | 4 | 4.5 |

根据上表提供的数据，求出*y*关于*x*的线性回归方程为＝0.7*x*＋0.35，那么表中*t*的值为(　　)

A．3 B．3.15

C．3.5 D．4.5

[答案]　A

[解析]　样本中心点是(，)，即(4.5，)．因为回归直线过该点，所以＝0.7×4.5＋0.35，解得*t*＝3.

13．(2012·湖南文，5)设某大学的女生体重*y*(单位：kg)与身高*x*(单位：cm)具有线性相关关系，根据一组样本数据(*xi*，*yi*)(*i*＝1,2，…，*n*)，用最小二乘法建立的回归方程为＝0.85*x*－85.71，则下列结论中不正确的是(　　)

A．*y*与*x*具有正的线性相关关系

B．回归直线过样本点的中心(，)

C．若该大学某女生身高增加1cm，则其体重约增加0.85kg

D．若该大学某女生身高为170cm，则可断定其体重必为58.79kg

[答案]　D

[解析]　本题考查线性回归方程．

D项中身高为170cm时，体重“约为”58.79，而不是“确定”，回归方程只能作出“估计”，而非确定“线性”关系．

14．(2014·哈师大附中高二期中)某咖啡厅为了了解热饮的销售量*y*(个)与气温*x*(℃)之间的关系，随机统计了某4天的销售量与气温，并制作了对照表：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 气温(℃) | 18 | 13 | 10 | －1 |
| 销售量(个) | 24 | 34 | 38 | 64 |

由表中数据，得线性回归方程*y*＝－2*x*＋*a*.当气温为－4℃时，预测销售量约为(　　)

A．68 B．66

C．72 D．70

[答案]　A

[解析]　∵＝(18＋13＋10－1)＝10，＝(24＋34＋38＋64)＝40，

∴40＝－2×10＋*a*，∴*a*＝60，

当*x*＝－4时，*y*＝－2×(－4)＋60＝68.

二、填空题

15．已知两个变量*x*和*y*之间有线性相关性，5次试验的观测数据如下表：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *x* | 100 | 120 | 140 | 160 | 180 |
| *y* | 45 | 54 | 62 | 75 | 92 |

那么变量*y*关于*x*的回归方程是\_\_\_\_\_\_\_\_．

[答案]　＝0.575*x*－14.9

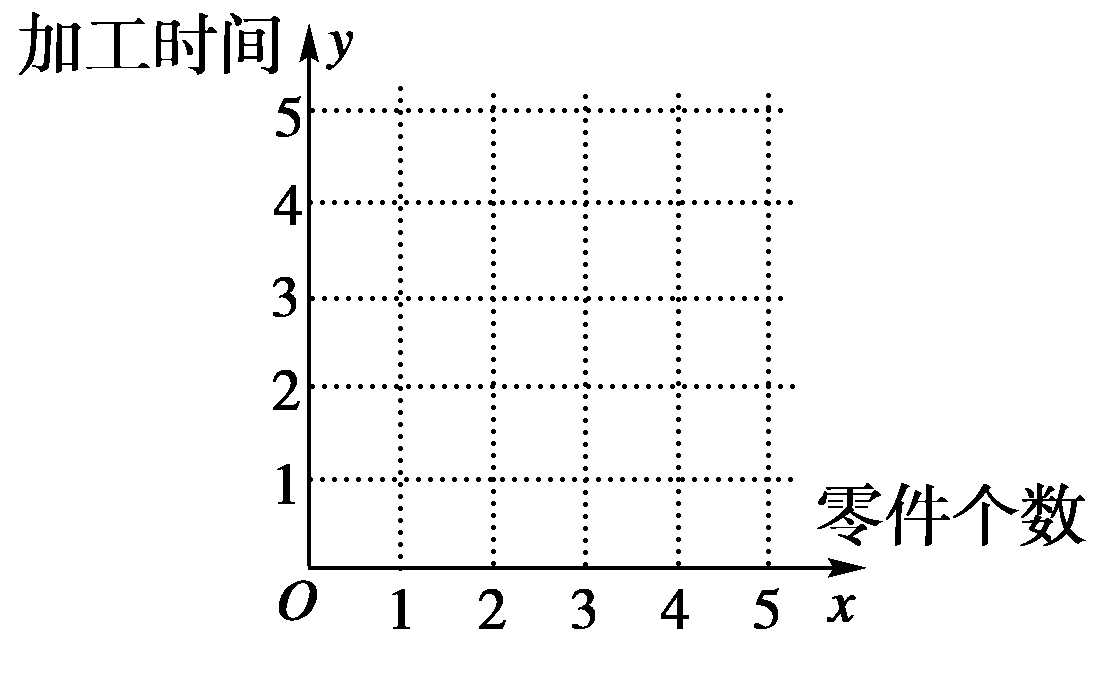
[解析]　根据公式计算可得＝0.575，＝－14.9，所以回归直线方程是＝0.575*x*－14.9.

三、解答题

16．某车间为了规定工时定额，需要确定加工零件所花费的时间，为此作了四次试验，得到的数据如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 零件的个数*x*(个) | 2 | 3 | 4 | 5 |
| 加工的时间*y*(小时) | 2.5 | 3 | 4 | 4.5 |

(1)在给定的坐标系中画出表中数据的散点图；

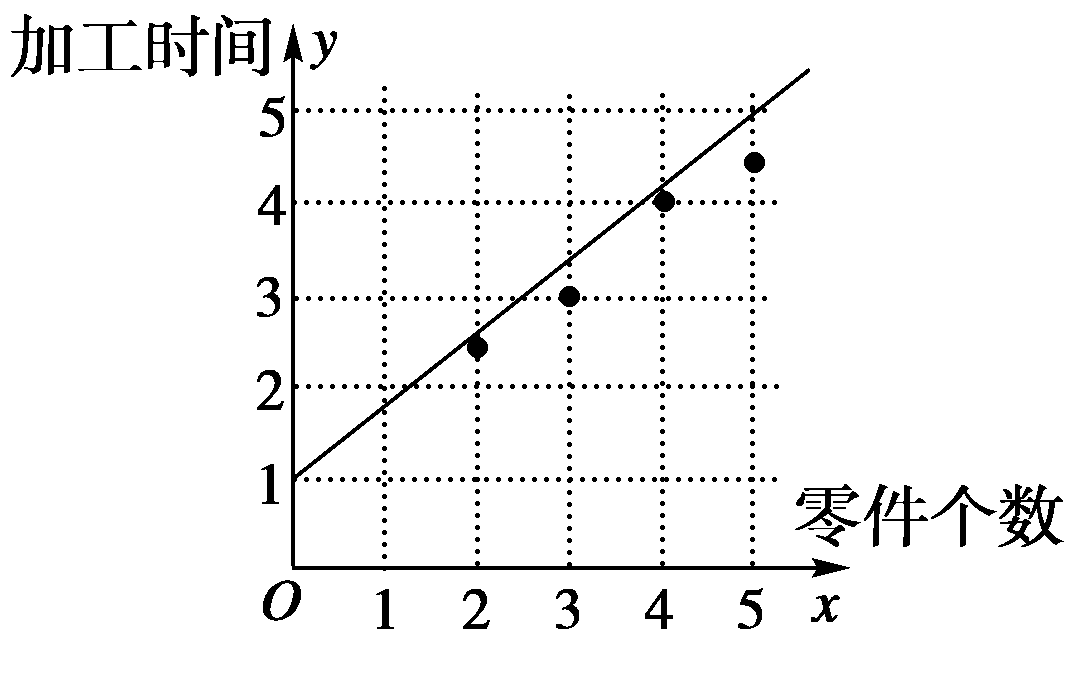


(2)求出*y*关于*x*的线性回归方程＝*x*＋，并在坐标系中画出回归直线；

(3)试预测加工10个零件需要多少时间？

(注：＝，＝－)

[解析]　(1)散点图如下图．



(2)由表中数据得*iyi*＝52.5，

＝3.5，＝3.5，＝54，

∴＝0.7，＝1.05.

∴＝0.7*x*＋1.05.回归直线如图中所示．

(3)将*x*＝10代入回归直线方程，得

*y*＝0.7×10＋1.05＝8.05(小时)，

∴预测加工10个零件需要8.05小时．

17．以下是某地搜集到的新房屋的销售价格*y*和房屋的面积*x*的数据：

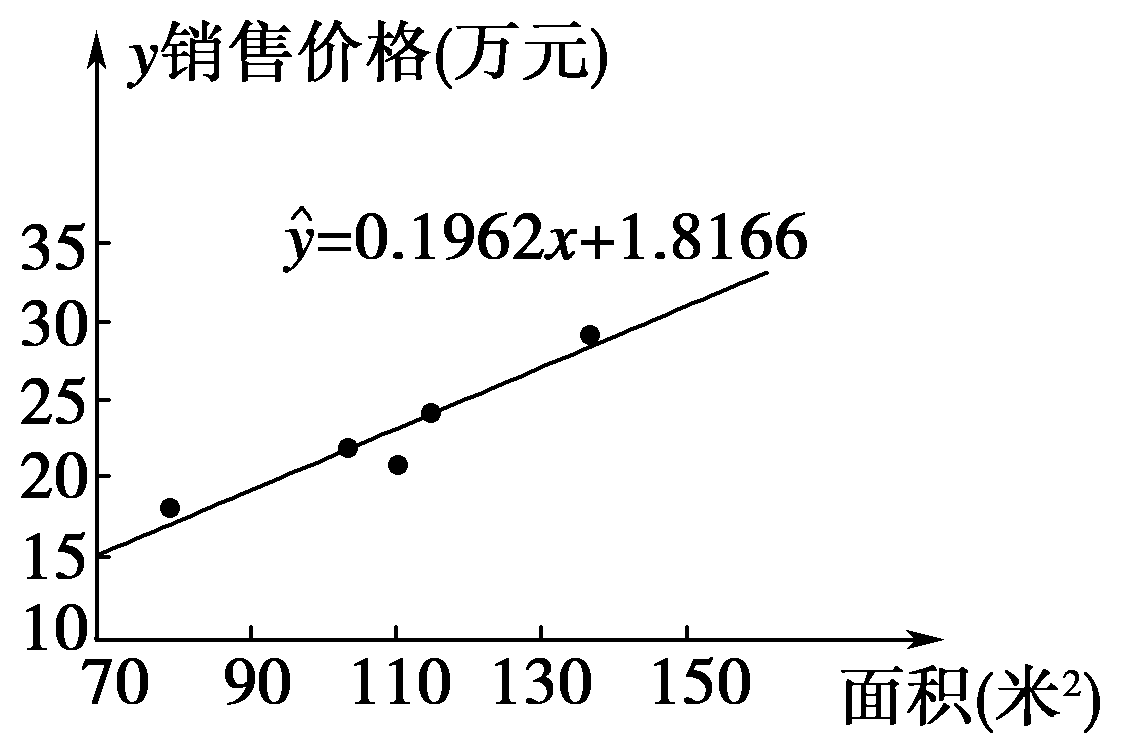
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 房屋面积(m2) | 115 | 110 | 80 | 135 | 105 |
| 销售价格(万元) | 24.8 | 21.6 | 18.4 | 29.2 | 22 |

(1)画出数据对应的散点图；

(2)求线性回归方程，并在散点图中加上回归直线；

(3)据(2)的结果估计当房屋面积为150m2时的销售价格．

[解析]　(1)数据对应的散点图如下图所示：



(2)＝*x*i＝109，*lxx*＝ (*x*i－)2＝1570，

＝23.2，*lxy*＝ (*x*i－)(*y*i－)＝308.

设所求回归直线方程为＝*x*＋，

则＝＝≈0.1962，＝－＝1.8166.

故所求回归直线方程为＝0.1962*x*＋1.8166.

(3)据(2)，当*x*＝150m2时，销售价格的估计值为

＝0.1962×150＋1.8166＝31.2466(万元)．

18．关于*x*与*y*有如下数据：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *x* | 2 | 4 | 5 | 6 | 8 |
| *y* | 30 | 40 | 60 | 50 | 70 |

有如下的两个线性模型：

①＝6.5*x*＋17.5；

②＝7*x*＋17.

试比较哪一个模型拟合效果更好．

[解析]　由①可得*yi*－*i*与*yi*－的关系如下表：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *yi*－*i* | －0.5 | －3.5 | 10 | －6.5 | 0.5 |
| *yi*－ | －20 | －10 | 10 | 0 | 20 |

所以(*yi*－*i*)2＝(－0.5)2＋(－3.5)2＋102＋(－6.5)2＋0.52＝155，

(*yi*－)2＝(－20)2＋(－10)2＋102＋02＋202＝1000.

所以*R*＝1－＝1－＝0.845.

由②可得*yi*－*i*与*yi*－的关系如下表：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *yi*－*i* | －1 | －5 | 8 | －9 | －3 |
| *yi*－ | －20 | －10 | 10 | 0 | 20 |

所以(*yi*－*i*)2＝(－1)2＋(－5)2＋82＋(－9)2＋(－3)2＝180，

(*yi*－)2＝(－20)2＋(－10)2＋102＋02＋202＝1000.

所以*R*＝1－＝1－＝0.82.

由于*R*＝0.845，*R*＝0.82,0.845>0.82，所以*R*>*R*.

故①的拟合效果好于②的拟合效果．

[点评]　*R*2的取值越大，模型的拟合效果越好．