**回归分析的基本思想及其初步应用解题方法与技巧-高中数学选修2-3第三章**

　有下列说法：

①线性回归分析就是由样本点去寻找一条直线，使之贴近这些样本点的数学方法；②利用样本点的散点图可以直观判断两个变量的关系是否可以用线性关系表示；③通过回归方程＝*x*＋，可以估计和观测变量的取值和变化趋势；④因为由任何一组观测值都可以求得一个线性回归方程，所以没有必要进行相关性检验．

其中正确命题的个数是(　　)

A．1　　　　B．2　　　　C．3　　　　D．4

【思路探究】　可借助于线性相关概念及性质逐一作出判断．

【自主解答】　①反映的正是最小二乘法思想，故正确．②反映的是画散点图的作用，也正确．③解释的是回归方程＝*x*＋的作用，故也正确．④是不正确的，在求回归方程之前必须进行相关性检验，以体现两变量的关系．

**【答案】**C



1．解答例1中④时，必须明确具有线性相关关系的两个变量间才能求得一个线性回归方程，否则求得的方程无实际意义．因此必须先进行线性相关性判断，后求线性回归方程．

2．回归分析的过程：

(1)随机抽取样本，确定数据，形成样本点；

(2)由样本点形成散点图，判断是否具有线性相关关系；

(3)由最小二乘法确定线性回归方程；

(4)由回归方程观察变量的取值及变化趋势．



关于变量*y*与*x*之间的回归直线方程叙述正确的是(　　)

A．表示*y*与*x*之间的一种确定性关系

B．表示*y*与*x*之间的相关关系

C．表示*y*与*x*之间的最真实的关系

D．表示*y*与*x*之间真实关系的一种效果最好的拟合

**【解析】**　回归直线方程能最大可能地反映*y*与*x*之间的真实关系，故选项D正确．

**【答案】**D

　已知某种商品的价格*x*(元)与需求量*y*(件)之间的关系有如下一组数据：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *x* | 14 | 16 | 18 | 20 | 22 |
| *y* | 12 | 10 | 7 | 5 | 3 |

求*y*关于*x*的回归直线方程，并说明回归模型拟合效果的好坏．

【思路探究】　回归模型拟合效果的好坏可以通过计算*R*2来判断，其值越大，说明模型的拟合效果越好．

【自主解答】　＝(14＋16＋18＋20＋22)＝18，

＝(12＋10＋7＋5＋3)＝7.4，

＝142＋162＋182＋202＋222＝1 660，

*iyi*＝14×12＋16×10＋18×7＋20×5＋22×3＝620，

所以＝＝＝－1.15，

＝7.4＋1.15×18＝28.1，

所以所求回归直线方程是＝－1.15*x*＋28.1.列出残差表：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *yi*－*i* | 0 | 0.3 | －0.4 | －0.1 | 0.2 |
| *yi*－ | 4.6 | 2.6 | －0.4 | －2.4 | －4.4 |

所以(*yi*－*i*)2＝0.3，(*yi*－)2＝53.2，

*R*2＝1－≈0.994，

所以回归模型的拟合效果很好．



1．回归直线方程能定量地描述两个变量的关系，系数，刻画了两个变量之间的变化趋势，其中表示*x*变化一个单位时，*y*的平均变化量．利用回归直线可以对问题进行预测，由一个变量的变化去推测另一个变量的变化．

2．线性回归分析中：

(1)残差平方和越小，预报精确度越高．

(2)相关指数*R*2取值越大，说明模型的拟合效果越好．



某运动员训练次数与运动成绩之间的数据关系如下：

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 次数(*x*) | 30 | 33 | 35 | 37 | 39 | 44 | 46 | 50 |
| 成绩(*y*) | 30 | 34 | 37 | 39 | 42 | 46 | 48 | 51 |

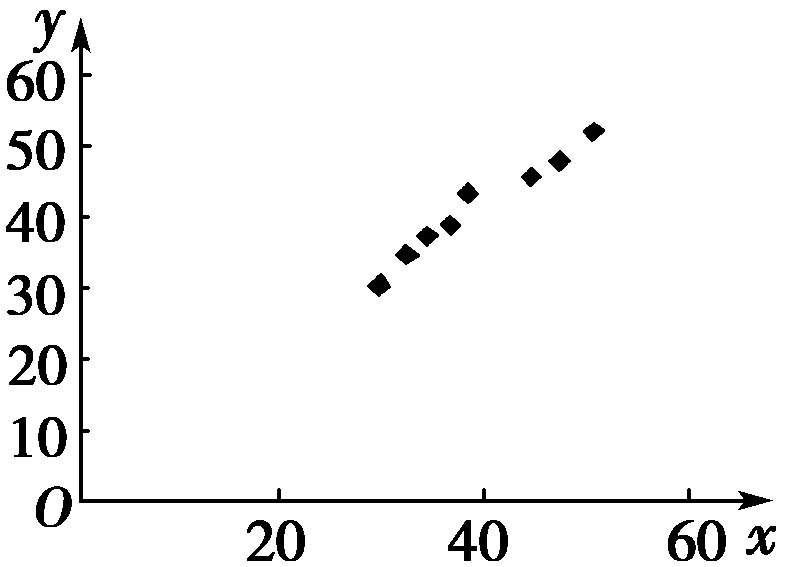
(1)作出散点图；

(2)求出线性回归方程；

(3)作出残差图，并说明模型的拟合效果；

(4)计算*R*2，并说明其含义．

**【解】**(1)作出该运动员训练次数(*x*)与成绩(*y*)之间的散点图，如图所示．



(2)可求得＝39.25，＝40.875，＝12 656，

＝13 731，*iyi*＝13 180，

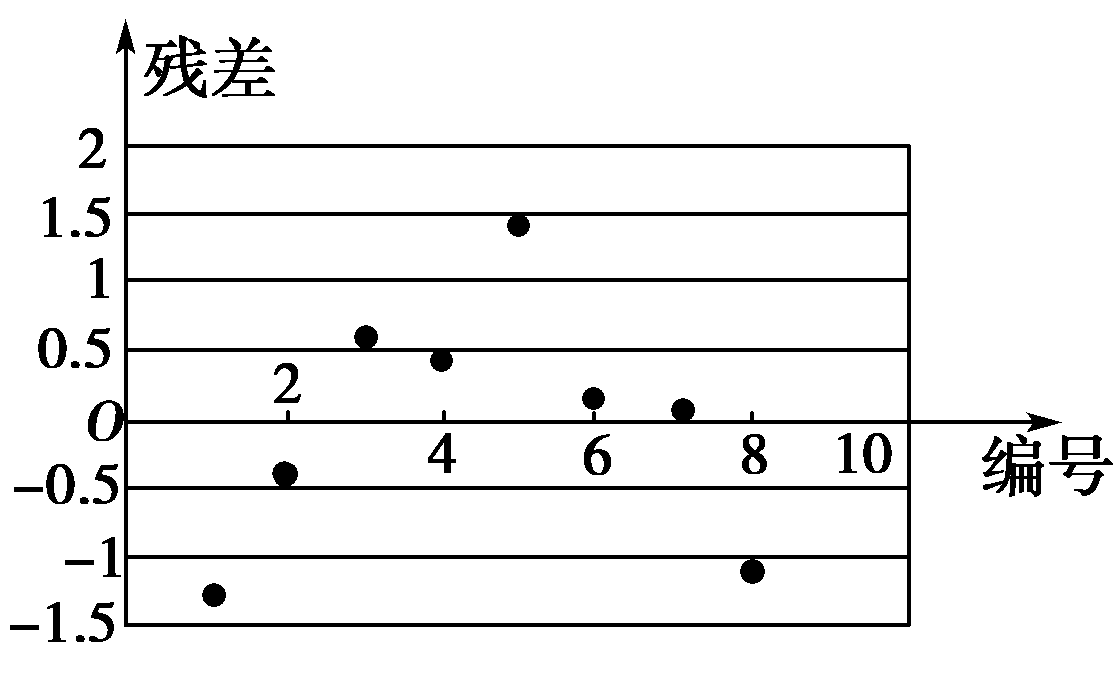
∴＝

＝≈1.041 5，

＝－＝－0.003 875，

∴线性回归方程为＝1.041 5*x*－0.003 875.

(3)作残差图如图所示，



由图可知，残差点比较均匀地分布在水平带状区域中，说明选用的模型比较合适．

(4)相关指数*R*2＝0.985 5.说明了该运动员的成绩的差异有98.55%的可能性是由训练次数引起的.

　下表为收集到的一组数据：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *x* | 21 | 23 | 25 | 27 | 29 | 32 | 35 |
| *y* | 7 | 11 | 21 | 24 | 66 | 115 | 325 |

　(1)作出*x*与*y*的散点图，并猜测*x*与*y*之间的关系；

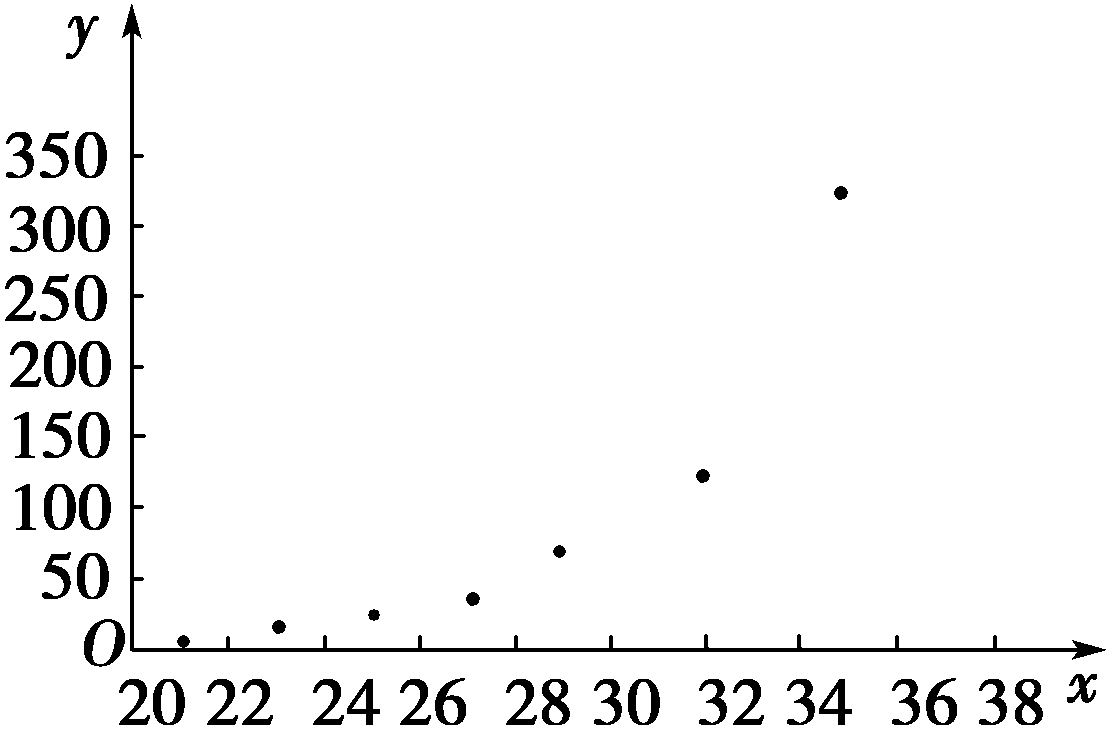
(2)建立*x*与*y*的关系，预报回归模型并计算残差；

(3)利用所得模型，预报*x*＝40时*y*的值．

【思路探究】　(1)画出散点图或进行相关性检验，确定两变量*x*、*y*是否线性相关．由散点图得*x*、*y*之间的回归模型．

(2)进行拟合，预报回归模型，求回归方程．

【自主解答】　(1)作出散点图如图，从散点图可以看出*x*与*y*不具有线性相关关系，根据已有知识可以发现样本点分布在某一条指数函数曲线*y*＝*c*1e*c*2*x*的周围，其中*c*1、*c*2为待定的参数．



(2)对两边取对数把指数关系变为线性关系，令*z*＝ln *y*，则有变换后的样本点应分布在直线*z*＝*bx*＋*a*，*a*＝ln *c*1，*b*＝*c*2的周围，这样就可以利用线性回归模型来建立*y*与*x*之间的非线性回归方程了，数据可以转化为：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *x* | 21 | 23 | 25 | 27 | 29 | 32 | 35 |
| *z* | 1.946 | 2.398 | 3.045 | 3.178 | 4.190 | 4.745 | 5.784 |

　求得回归直线方程为＝0.272*x*－3.849，

∴＝e0.272*x*－3.849.

残差如下表：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *yi* | 7 | 11 | 21 | 24 | 66 | 115 | 325 |
| *i* | 6.443 | 11.101 | 19.125 | 32.950 | 56.770 | 128.381 | 290.325 |
| *i* | 0.557 | －0.101 | 1.875 | －8.950 | 9.23 | －13.381 | 34.675 |

　(3)当*x*＝40时，*y*＝e0.272*x*－3.849≈1 131.



　两个变量不具有线性关系，不能直接利用线性回归方程建立两个变量的关系，可以通过变换的方法转化为线性回归模型，如*y*＝*c*1e*c*2*x*，我们可以通过对数变换把指数关系变为线性关系，令*z*＝ln *y*，则变换后样本点应该分布在直线*z*＝*bx*＋*a*(*a*＝ln *c*1，*b*＝*c*2)的周围．

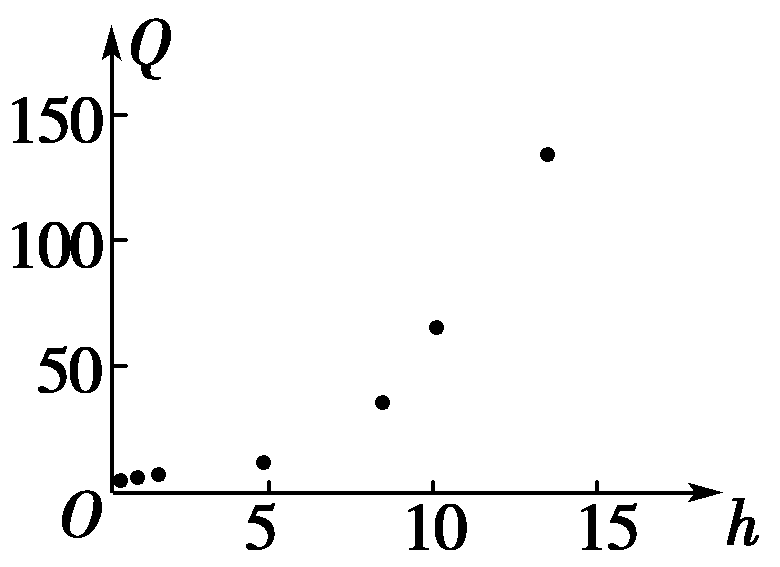


有一个测量水流量的实验装置，测得试验数据如下表：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *i* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 水高*h*(厘米) | 0.7 | 1.1 | 2.5 | 4.9 | 8.1 | 10.2 | 13.5 |
| 流量 |  |  |  |  |  |  |  |
| *Q*(升/分钟) | 0.082 | 0.25 | 1.8 | 11.2 | 37.5 | 66.5 | 134 |

根据表中数据，建立*Q*与*h*之间的回归方程．

**【解】**由表中测得的数据可以作出散点图，如图．



观察散点图中样本点的分布规律，可以判断样本点分布在某一条曲线附近，表示该曲线的函数模型是*Q*＝*m*·*hn*(*m*，*n*是正的常数)．两边取常用对数，

则lg *Q*＝lg *m*＋*n*·lg *h*.

令*y*＝lg *Q*，*x*＝lg *h*，那么*y*＝*nx*＋lg *m*，

即为线性函数模型*y*＝*bx*＋*a*的形式(其中*b*＝*n*，*a*＝lg *m*)．

由下面的数据表，用最小二乘法可求得≈2.509 7，＝－0.707 7，所以*n*≈2.51，*m*≈0.196.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *i* | *hi* | *Qi* | *xi*＝lg *hi* | *yi*＝lg *Qi* | *x* | *xiyi* |
| 1 | 0.7 | 0.082 | －0.154 9 | －1.086 2 | 0.024 | 0.168 3 |
| 2 | 1.1 | 0.25 | 0.041 4 | －0.602 1 | 0.001 7 | －0.024 9 |
| 3 | 2.5 | 1.8 | 0.397 9 | 0.255 3 | 0.158 3 | 0.101 6 |
| 4 | 4.9 | 11.2 | 0.690 2 | 1.049 2 | 0.476 4 | 0.724 2 |
| 5 | 8.1 | 37.5 | 0.908 5 | 1.574 0 | 0.825 4 | 1.430 0 |
| 6 | 10.2 | 66.5 | 1.008 6 | 1.822 8 | 1.017 3 | 1.838 5 |
| 7 | 13.5 | 134 | 1.130 3 | 2.127 1 | 1.277 6 | 2.404 3 |
| ∑ |  |  | 4.022 | 5.140 1 | 3.780 7 | 6.642 |

于是所求得的回归方程为*Q*＝0.196·*h*2.51.