第3课时　线性回归方程(1)



知识技能

理解线性回归的基本思想和方法，体会变量之间的相关关系，学会求线性回归方程．

思想方法

经历线性回归方程的探求过程，体会拟合函数的方法，进一步感受直观化(散点图)与最小二乘法的思想．

核心素养

在建立具有相关关系的两个变量之间的线性回归方程的过程中，经历探求这样的拟合函数的过程，体验辩证思维的过程，发展逻辑推理和数学运算素养．



重点：用线性函数近似地刻画两个具有相关关系的变量之间的关系．

难点：最小二乘法的思维过程．



问题导引

预习教材P146～153，思考下面的问题：

1. 如何求线性回归方程？

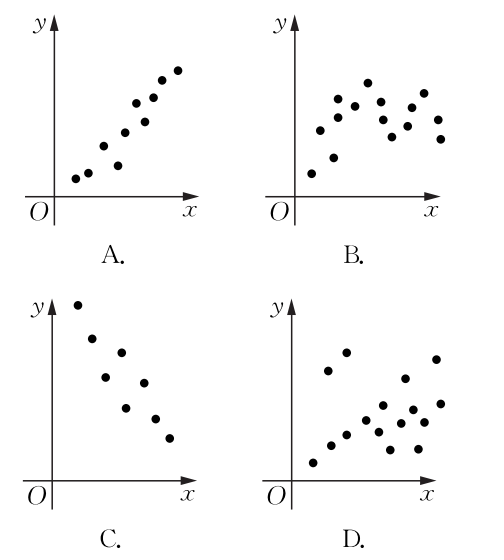
提示　利用公式法求出， ，代入直线＝ *x*＋求出线性回归方程．

2. 产生随机误差的原因是什么？

提示　①所用的确定性函数不恰当引起的误差；②忽略了某些因素的影响；③存在观测误差．

即时体验

1. (多选)下列散点图中，两个变量的关系适合用线性回归模型刻画的是(AC)



2. 从某大学随机选取8名女大学生，其身高*x*(单位：cm)与体重*y*(单位：kg)之间的线性回归方程为 ＝0.849*x*－85.712，则由方程可以预测身高为172cm的女大学生的体重为60.316kg.

3. 设变量*x*与*y*的回归方程为＝5*x*＋3，变量*x*增加1个单位时，*y*平均增加5个单位．



一、 问题情境

通过前两课时的学习，我们了解到，根据成对样本数据的散点图和相关系数，可以推断两个变量是否存在相关关系，是正相关还是负相关，以及相关程度的强弱等．

那么是否可以像两个变量之间的确定性关系那样，通过建立适当的函数模型刻画两个随机变量的相关关系呢？

下表提供了某厂节能降耗技术改造后生产甲产品过程中记录的产量*x*(单位：t)与相应的生产能耗*y*(单位：t标准煤)之间的几组对照数据．

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

已知该厂技术改造前生产100t甲产品的生产能耗为90t标准煤，你能预测生产100t甲产品的生产能耗比技术改造前降低了多少吨标准煤吗？该如何求呢？[1]

根据上表提供的数据画出散点图→求出*y*关于*x*的线性回归方程＝ *x*＋→根据求出的线性回归方程估计技术改造后生产100t甲产品的生产能耗→预测比技术改造前降低了多少吨标准煤．

二、 数学建构

(一) 生成概念

问题1　变量之间的常见关系有哪些？

(变量之间的常见关系有两类：一类是确定性函数关系，变量之间的关系可以用函数表示；另一类是相关关系，变量之间有一定的联系，但不能完全用函数来表达．)

问题2　用怎样的数学模型刻画问题情境中两个变量之间的相关关系？

画散点图(图1)．

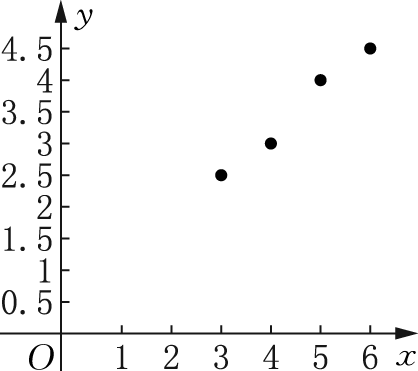


图1

在散点图中，这些点在一条直线附近，但并不都在这条直线上．也就是说，上述直线并不能精确地反映*x*与*y*之间的关系，*y*的值不能由*x*确定，在此，我们将两者之间的关系表示为

*y*＝*a*＋*bx*＋*ε*，

其中*a*＋*bx*是确定性函数，*ε*称为随机误差．

我们将*y*＝*a*＋*bx*＋*ε*称为线性回归模型．[2]

问题3　如何探究线性回归方程？[3]

用方程为＝ *x*＋的直线拟合散点图中的点，应使得该直线与散点图中的点最接近．

考虑用直线＝ *x*＋与各散点在垂直方向(纵轴方向)上的距离的平方和的大小，来衡量直线＝ *x*＋与给出的*n*个散点的接近程度，即用最小二乘法来求， ，它的原理和方法如下：

设有(*x, y*)的*n*对观测数据，具体数据如下表所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *x* | *x*1 | *x*2 | *x*3 | … | *xn* |
| *y* | *y*1 | *y*2 | *y*3 | … | *yn* |

于是，只要求出使Q(α， β)＝ (yi－βxi－α)2取最小值时的*α*， *β*的值，分别将它们作为， ，称＝ *x*＋为拟合这*n*对数据的线性回归方程，将该方程所表示的直线称为回归直线．其中称为回归系数，称为回归截距，称为回归值．

问题4　如何确定*a, b*的值使得*Q*最小？[4]

方法1：

Q(α， β)＝ (*y*i－β*x*i－α)2

＝[(*y*i－β*x*i)2－2α(*y*i－β*x*i)＋α2]

＝*n*α2－2α (*y*i－β*x*i)＋ (*y*i－β*x*i)2

＝*n*2＋ (*y*i－β*x*i)2－

≥ (*y*i－β*x*i)2－

＝ (*y*i－β*x*i)2－

＝β2*x*－2β*x*i*y*i＋*y*

－*n*(－β)2

＝β2－2β

＋*y*－*n*2

＝2＋*y*－*n*2－

≥*y*－*n*2－.

当且仅当α－＝0且β－＝0时，*Q*(*α*， *β*)取得最小值，

Q(α， β)*m*i*n*＝*y*－*n*2－，

此时的*α*， *β*作为*a, b*的估计值， ，

即时，*Q*(*α*， *β*)取得最小值．

方法2：

Q(α， β)＝ (*y*i－β*x*i－α)2

＝[*y*i－β*x*i－(－β)＋(－β)－α]2

＝[(*y*i－)－β(*x*i－)＋(－β)－α]2

＝[(*y*i－)－β(*x*i－)]2＋2[(*y*i－)－β(*x*i－)]×[(－β)－α]＋*n*[(－β)－α]2，

而[(*y*i－)－β(*x*i－)](－β－α)

＝(－β－α)[(*y*i－)－β(*x*i－)]

＝(－β－α)

＝(－β－α)[(*n*－*n*)－β(*n*－*n*)]

＝0，

所以Q(α， β)＝[(*y*i－)－β(*x*i－)]2＋*n*(－β－α)2.

上式右边各项均为非负数，且前*n*项与*α*无关．所以，要使*Q*取得最小值，后一项的值应为0，即*α*＝－*β*.此时

Q(α， β)＝[(*y*i－)－β(*x*i－)]2

＝β2 (*x*i－)2－2β (*x*i－)(*y*i－)

＋ (*y*i－)2.

上式是关于*β*的二次函数，因此要使*Q*取得最小值时，当且仅当*β*的取值为

＝.

此时的*α*， *β*的值作为*a, b*的估计值， ，

即

方法3:

Q(α， β)＝ (*y*i－β*x*i－α)2

＝{(*y*i－)＋[－(β＋α)]

－β(*x*i－)}2

＝ (*y*i－)2＋*n*[－(β＋α)]2＋β2(*x*i－)2＋2[－(β＋α)]· (*y*i－)－2β[－(β＋α)] (*x*i－)－2β (*x*i－)(*y*i－)

＝ (*y*i－)2＋*n*[－(β＋α)]2＋β2(*x*i－)2－2β (*x*i－)(*y*i－)

＝ (*y*i－)2＋*n*[－(β＋α)]2＋ (*x*i－)2

＝ (*y*i－)2＋*n*[－(β＋α)]2＋ (*x*i－)22－

＝*n*[－(β＋α)]2＋ (*x*i－

)22＋

(*y*i－)2－.

上式中后两项与*α*， *β*无关，前两项为非负数，因此当且仅当前两项的值都为0时，*Q*(*α*， *β*)取得最小值，即有

此时的*α*， *β*的值作为*a, b*的估计值， ，

即

综上，， 的计算公式为

其中＝*x*i, ＝*y*i.

， 的计算公式也可化为

问题5　求线性回归方程的步骤是什么？

用回归直线进行数据拟合的一般步骤：

① 作出散点图，判断散点是否在一条直线附近；

② 如果散点在一条直线附近，用上面的公式求出， ，并写出线性回归方程．

问题6　你能解决问题情境中的问题，求出线性回归方程＝ *x*＋吗？

＝＝4.5, ＝＝3.5，

*x*i*y*i＝3×2.5＋4×3＋5×4＋6×4.5＝66.5,

*x*＝32＋42＋52＋62＝86，

所以＝＝＝0.7, ＝－ ＝3.5－0.7×4.5＝0.35.

所以所求的线性回归方程为＝0.7*x*＋0.35.

当*x*＝100时，＝0.7×100＋0.35＝70.35，所以降低了90－70.35＝19.65t标准煤．

(二) 理解概念

(1) 求线性回归方程，首先应注意到，只有在散点图大致呈线性时，求出的线性回归方程才有实际意义．否则，求出的线性回归方程毫无意义．因此，对一组数据作线性回归分析时，应先看其散点图是否成线性．

(2) 由＝－ 可知线性回归直线必过点(， )，点(， )通常称为样本的中心．

三、 数学运用

例1[5]　(1) 某工厂生产的某产品产量*x*(单位：千件)与单位成本*y*(单位：元)满足线性回归方程＝77.36－1.82*x*，下列说法中正确的是(A)

A. 产量每增加1000件，单位成本约下降1.82元

B. 产量每减少1000件，单位成本约下降1.82元

C. 当产量为1千件时，单位成本为75.54元

D. 当产量为2千件时，单位成本为73.72元

(2) 某企业想通过做广告来提高销售额，经预测可知产品的广告费*x*(单位：百万元)与销售额*y*(单位：百万元)之间有如下对应数据：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *x*/百万元 | 2 | 4 | 5 | 6 | 8 |
| *y*/百万元 | 30 | 40 | 60 | 50 | 70 |

由表中的数据得线性回归方程为＝ *x*＋，其中＝6.5，由此预测当广告费为7百万元时，销售额为\_\_63\_\_万元．

(见学生用书课堂本P91)

[处理建议]　(1) 令*f*(*x*)＝77.36－1.82*x,* 将(， )代入，计算*f*(*x*＋1)－*f*(*x*)即可．

(2) 求出，得到线性回归方程＝*x*＋，将*x*代入，即可求到对应的的值．

[规范板书]　解析　(1) 令*f*(*x*)＝77.36－1.82*x, f*(*x*＋1)－*f*(*x*)＝77.36－1.82(*x*＋1)－77.36＋1.82*x*＝－1.82，

所以产量每增加1000件，单位成本约下降1.82元．

(2) ＝＝5，

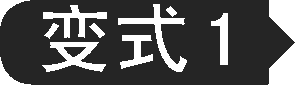
＝＝50，

所以＝－ ＝50－6.5×5＝17.5.

所以＝6.5*x*＋17.5.

当*x*＝7时，＝6.5×7＋17.5＝63.

[题后反思]　在线性回归方程＝ *x*＋中，当回归系数>0时，说明两个变量正相关，当*x*每增加一个单位时，*y*就增加个单位；当回归系数<0时，说明两个变量负相关，当*x*每增加一个单位时，*y*就减少||个单位．

　(多选)对两个变量*y*和*x*进行回归分析，下列说法中正确的有(ACD)

A. 由样本数据得到的回归方程＝ *x*＋必过样本点的中心(， )

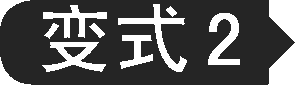
B. 设有一个回归方程＝3－5*x*，变量*x*增加1个单位时，*y*平均增加5个单位

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. 在回归分析中，变量间的关系若是非确定关系，那么因变量不能由自变量唯一确定

[处理建议]　根据回归分析的概念、线性回归方程的特征逐一判断．

[规范板书]　解析　由样本数据得到的回归方程＝ *x*＋必过样本点的中心(， )；对于回归方程＝3－5*x*，变量*x*增加1个单位时，*y*平均减少5个单位；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； 在回归分析中，变量间的关系不是函数关系，所以因变量不能由自变量唯一确定．

　(多选)5G技术在我国已经进入高速发展的阶段，5G手机的销量也逐渐上升，某手机商城统计了近5个月来5G手机的实际销量，具体数据如下表所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 月　　份 | 2021年  2月 | 2021年  3月 | 2021年  4月 | 2021年  5月 | 2021年  6月 |
| 月份编号*x* | 1 | 2 | 3 | 4 | 5 |
| 销量*y*/千部 | 37 | 104 | *a* | 196 | 216 |

若*y*与*x*线性相关，且求得的线性回归方程为＝45*x*＋5，下列说法中正确的有(AB)

A. *a*＝147

B. *y*与*x*正相关

C. *y*与*x*的相关系数为负数

D. 8月份该手机商城的5G手机销量约为36.5万部

[处理建议]　利用回归方程的性质代入(， )求*a,* 利用分析正相关、负相关， 赋值预测．

[规范板书]　解析　＝＝3, ＝45×3＋5＝140，由37＋104＋*a*＋196＋216＝140×5，得*a*＝147，故A正确；由回归方程中的*x*的系数为正可知，*y*与*x*正相关，且其相关系数*r*>0，故B正确，C错误；8月份时，*x*＝7, ＝45×7＋5＝320(千部)＝32(万部)，故D错误.

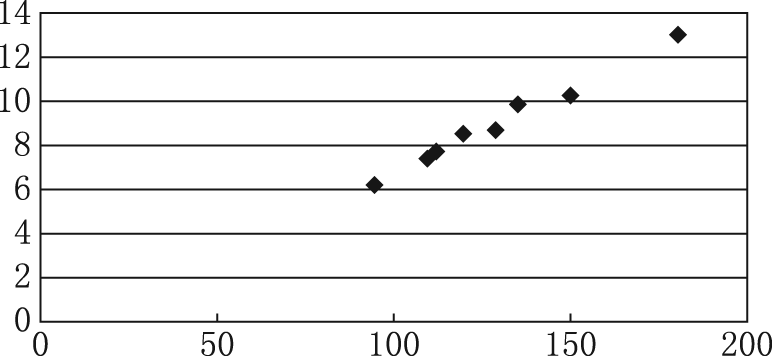
例2　(教材P151例4)下表为某地近几年机动车辆数与交通事故数的统计资料，请判断机动车辆数与交通事故数之间是否具有线性相关关系．如果具有线性相关关系，求出线性回归方程；如果不具有线性相关关系，说明理由．[6]

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 机动车辆数*x*/103辆 | 95 | 110 | 112 | 120 | 129 | 135 | 150 | 180 |
| 交通事故数*y*/103件 | 6.2 | 7.5 | 7.7 | 8.5 | 8.7 | 9.8 | 10.2 | 13 |

(见学生用书课堂本P92)

[处理建议]　作出散点图，利用图形判断出*x, y*具有线性相关关系， 利用公式求出线性回归方程．

[规范板书]　解　在平面直角坐标系中画出数据的散点图(如图)．



(例2)

直观判断散点在一条直线附近，故机动车辆数与交通事故数之间具有线性相关关系．

计算相应的数据之和：

*x*i＝1031, *y*i＝71.6，

*x*＝137835, *y*＝671, *x*i*y*i＝9611.7.

由公式得＝≈0.0774, ＝－ ≈－1.0249.

所以所求线性回归方程为＝0.0774*x*－1.0249.

[题后反思]　也可如教材那般，根据相关系数公式得*r*≈0.9927，从而说明两变量具有线性相关关系．

　某医院用光电比色计检验尿汞时，得尿汞含量(单位：mg/L)与消光系数如下表所示：

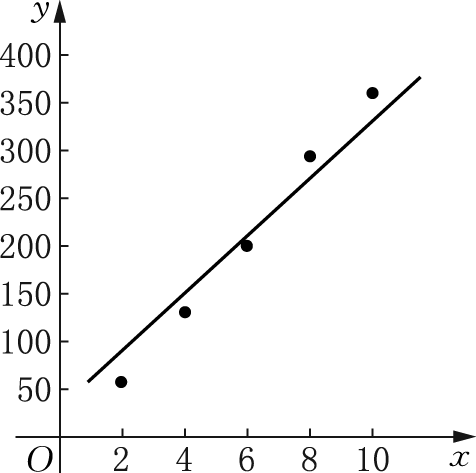
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 尿汞含量*x* | 2 | 4 | 6 | 8 | 10 |
| 消光系数*y* | 64 | 134 | 205 | 285 | 360 |

(1) 画出散点图；

(2) 若*y*与*x*之间具有线性相关关系，求线性回归方程；

(3) 估计尿汞含量为9mg/L时的消光系数．

[规范板书]　解　(1) 散点图如图所示：



(变式)

(2) 由散点图可知*y*与*x*线性相关．

计算相应的数据之和：

*x*i＝30, *y*i＝1048, *x*＝220, *x*i*y*i＝7774，从而＝6, ＝209.6.

由公式得＝＝37.15, ＝－ ＝－13.3.

所以所求线性回归方程为＝37.15*x*－13.3.

(3) 当*x*＝9时，＝37.15×9－13.3＝321.05，

故估计尿汞含量为9mg/L时的消光系数为321.05.

四、 课堂练习

1. 在对两个变量*x, y*进行线性回归分析时有以下步骤：

① 利用回归方程进行预测；

② 收集数据(*x*i*, y*i), i＝1, 2, …， *n*；

③ 求线性回归方程；

④ 根据所收集的数据绘制散点图．

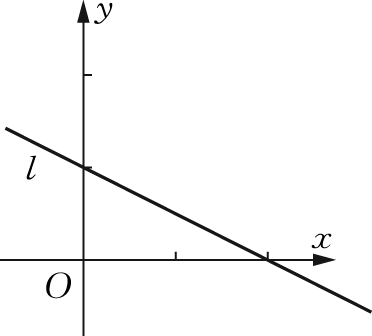
正确的操作顺序是！！！\_\_②④③①\_\_.

2. 3月15日，某市物价部门对本市5家商场的某种商品的一天销售量及其价格进行了调查，5家商场的售价*x*(单位：元)和销售量*y*(单位：件)之间的一组数据如下表所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 售价*x*/元 | 9 | 9.5 | 10 | 10.5 | 11 |
| 销售量*y*/件 | 11 | 10 | 8 | 6 | 5 |

由散点图可知销售量*y*与售价*x*之间有较好的线性相关关系，其线性回归方程为＝－3.2*x*＋，则＝\_\_40\_\_.

3. (多选)设(*x*1, *y*1), (*x*2, *y*2), …， (*xn, yn*)是变量*x*和*y*的*n*对数据，直线*l*是由这些样本点通过最小二乘法得到的线性回归直线(如图)，下列结论中正确的有(AC)



(第3题)

A. 直线*l*过点(， )

B. 回归直线必通过散点图中的多个点

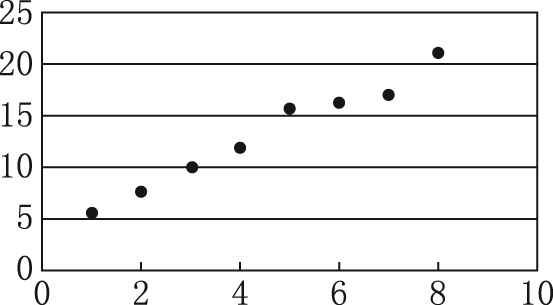
C. *x*和*y*的相关系数在－1和0之间

D. 当*n*为偶数时，分布在*l*两侧的样本点的个数一定相同

4. 对一个做直线运动的质点的运动过程观测了8次，得到的数据如下表所示，试估计当*x*＝9时*y*的值．

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 时刻*x*/s | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 位置观测值*y*/cm | 5.54 | 7.52 | 10.02 | 11.73 | 15.69 | 16.12 | 16.98 | 21.06 |

解　散点图如图所示：



(第4题))

从散点图中可以看出，时刻*x*与位置观测值*y*之间有着较好的线性相关关系，可以用线性回归方程来刻画它们之间的关系．

由表中数据计算得＝4.5, ＝13.0825, *x*＝204, *x*i*y*i＝560.07.

由公式得＝≈2.1214，

＝－ ≈3.5362，

所以线性回归方程为＝2.1214*x*＋3.5362.

当*x*＝9时，＝2.1214×9＋3.5362＝22.6288，即位置观测值为22.6288cm.

五、 课堂小结

1. 线性回归模型*y*＝*a*＋*bx*＋*ε*与确定性函数*y*＝*a*＋*bx*相比，它表示*y*与*x*之间是相关关系(非确定性关系)，其中的随机误差*ε*提供了选择模型的准则以及在模型合理的情况下探求最佳估计值， 的工具．

2. 线性回归方程＝ *x*＋中， 的意义：以 为基数，＞0时，*x*每增加1个单位，*y*相应地平均增加个单位，＜0时，*x*每增加一个单位时，*y*相应地平均减少||个单位．

3. ， 的计算公式：

或



[1] 从实际问题出发，引导学生利用已知数据寻找合理的拟合函数，利用数学的方法预测并构造解决实际问题的模型，经历将生活问题转化为数学问题的过程，体会数学的应用性．

[2] 对于线性回归模型，我们应该考虑下面两个问题：

① 模型是否合理(这个问题在上一节课已解决)；

② 在模型合理的情况下，如何估计*a, B.*

[3] 利用最小二乘法的思想推理出公式，发展逻辑推理素养．

[4] 还可参考教材P149～151的方法．

[5] 线性回归方程的性质与运算．

[6] 利用线性回归方程求解实际应用问题．