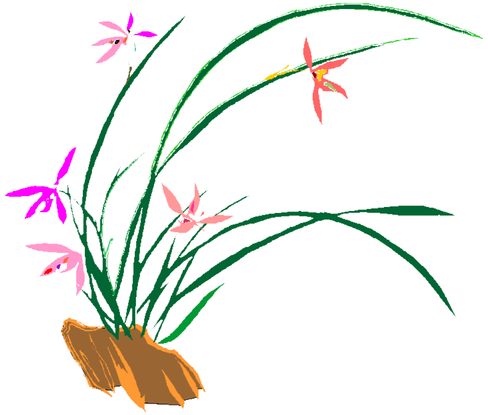
北师大版同步教材精品课件

《一元线性回归》



教学内容

在现实生活中，反映量与量之间的函数关系非常普遍，但也存在一些量与量之间不满足函数关系，如人的身高与体重.一般说来，人的身高越高，体重就越重，二者确实有关系.但是身高相同的人，体重却不一定相同.那么身高和体重究竟存在怎样的关系呢？

师生互动

教师提出问题，引发学生思考.

设计意图

通过问题引入，激发学生的学习欲望，引入本节内容.

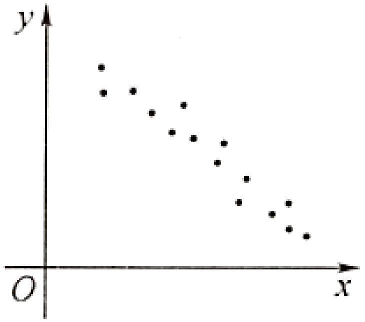
教学内容

问题1为了了解人的身高与体重的关系，我们随机抽取9名15岁的男生，测得他们的身高（单位：cm）、体重（单位：kg）如表：

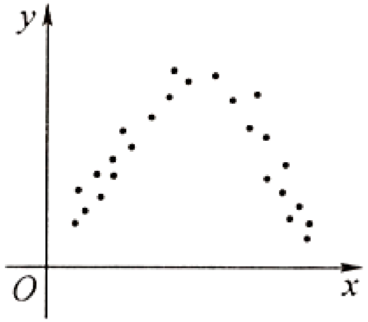
编号	1	2	3	4	5	6	7	8	9
身高 /cm	165	157	155	175	168	157	178	160	163
体重 /kg	52	44	45	55	54	47	62	50	63

如果把身高看作横坐标、体重看作纵坐标，在平面直角坐标系中画出对应的点.根据画出的图观察，随着身高的增长，体重的变化趋势是怎样的？

结论：每个点对应的一对数据(x_i, y_i)，称为成对数据.这些点构成的图称为散点图.



(1)



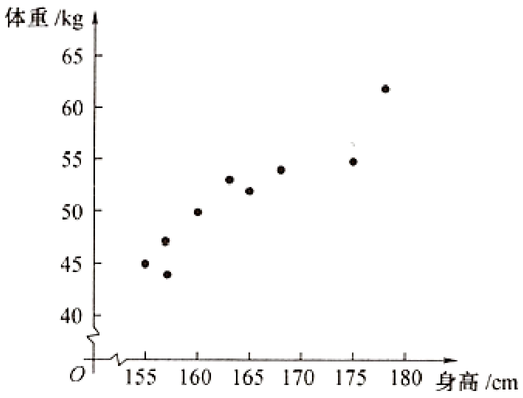
(2)

问题2观察散点图（1）和（2），试说明对相应的两变量的描述是曲线拟合，还是直线拟合，并说明理由.

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的趋势，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曲线拟合.若在两个变量X和Y的散点图中，所有的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，称之为直线拟合.

师生互动

教师引导学生以身高作为横坐标、体重作为纵坐标画出对应的点图（如图），学生由图得出随着身高的增长，体重基本上呈现直线增长的趋势.



生：图（1）中是直线拟合，图（2）中是曲线拟合.

教师引导学生理解散点图、曲线拟合和直线拟合的含义.

形成概念1

设计意图

使学生理解处理数据的一种手段是作出图，由图来反映规律.

由散点图认识直线拟合和曲线拟合的概念.

教学内容

例1　对于问题1中抽取的9名15岁男生，测得他们的身高和体重的数据后，通过散点图知两者可利用直线拟合，试求这条直线的方程，并估算一个身高166*cm*的15岁男生的体重.

方法1　选取散点图中的两个点，使得其余的点在这两个点所连直线两侧分布得尽可能一样多，如有人选取了（165，52）和（168，54）这两个成对数据，得到直线方程为$2x-3y-174=0$.因此，一个身高166*cm*的15岁男生，他的体重大致为52.667*kg*.

方法2 将所有的点分成两部分，一部分是身高在165cm以下的，一部分是身高在165cm以上（含165cm）的；然后每部分的点求一个平均点：165cm以下的身高、体重的平均数（取整近似）作为一个平均点，即（158，48），165cm以上（含165cm）的身高、体重的平均数（取整近似）作为另一个平均点，即（172，56）；最后将这两点连接成一条直线，得到直线方程为$4x-7y-296=0$.因此，一个身高166cm的15岁男生，他的体重大致为52.571kg.

师生互动

教师引导学生分析题目，得到两种方法.

教师给予总结.

教师最后给出思考问题：实际上，还有很多直线拟合的方法，你能想到哪些方法呢？

设计意图

探寻求拟合直线方程的方法，让学生认识到求回归直线的必要性.

教学内容

问题3 对于给定的两个变量X和Y，可以把其成对的观测值$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$表示为平面直角坐标系中的$n$个点.若找到一条直线$Y=a+bX$，实际观测值$y_i$与计算值$a+bx_i$的差值为$y_i-(a+bx_i)$，当偏差的平方和为何值时，该直线$Y=a+bX$拟合的效果达到理想状态？

相关概念：

对于给定的两个变量X和Y，可以把其成对的观测值$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$表示为平面直角坐标系中的$n$个点.现在希望找到一条直线$Y=a+bX$，使得对每一个$x_i\ (i=1,2,\cdots,n)$，由这个直线方程计算出来的值$a+bx_i$与实际观测值$y_i$的差异尽可能小.为此，希望$[y_1-(a+bx_1)]^2+[y_2-(a+bx_2)]^2+\cdots+[y_n-(a+bx_n)]^2$达到最小.换句话说，我们希望$a$，$b$的取值能使上式达到最小.这个方法称为最小二乘法.

问题4 先研究简单的情形，考虑3对数据$(x_1,y_1),(x_2,y_2),(x_3,y_3)$，即：求$a$，$b$的值，使得偏差$y_i-(a+bx_i)(i=1,2,3)$的平方和最小，即$[y_1-(a+bx_1)]^2+[y_2-(a+bx_2)]^2+[y_3-(a+bx_3)]^2$ 达到最小.

（1）如何用向量的语言描述偏差$y_i-(a+bx_i)(i=1,2,3)$？此时，“求a，b的值，使得偏差$y_i-(a+bx_i)(i=1,2,3)$的平方和最小”就转化成了什么问题？如何进一步利用向量描述$(y_1-(a+bx_1),y_2-(a+bx_2),y_3-(a+bx_3))$ ？

（2）如何求a，b的值，使得向量$(y_1-(a+bx_1),y_2-(a+bx_2),y_3-(a+bx_3))$ 的长度最小？

在（1）中，我们得到该向量可以表示为$\overrightarrow{OY}-(a\overrightarrow{OI}+b\overrightarrow{OX})$，该问题转化为：求$a$，$b$的值，使$\overrightarrow{OY}-(a\overrightarrow{OI}+b\overrightarrow{OX})$的长度最小.你能用向量的方法思考该问题吗？

推导过程如下：

求 $|\overrightarrow{OY}-(a\overrightarrow{OI}+b\overrightarrow{OX})|$最小时的$a$，$b$的值，就是求$\overrightarrow{OY}-(a\overrightarrow{OI}+b\overrightarrow{OX})$与$\overrightarrow{OI}$ 和$\overrightarrow{OX}$的数量积分别为0时的a，b的值，即

$$\begin{cases}[\overrightarrow{OY}-(a\overrightarrow{OI}+b\overrightarrow{OX})]\cdot\overrightarrow{OI}=0,\\ [\overrightarrow{OY}-(a\overrightarrow{OI}+b\overrightarrow{OX})]\cdot\overrightarrow{OX}=0,\end{cases}$$

概念形成2

用向量的坐标表示，即 $\begin{cases}[(y_1,y_2,y_3)-a(1,1,1)-b(x_1,x_2,x_3)]\cdot(1,1,1)=0,\\ [(y_1,y_2,y_3)-a(1,1,1)-b(x_1,x_2,x_3)]\cdot(x_1,x_2,x_3)=0,\end{cases}$

化简，得 $\begin{cases}\sum\limits_{i=1}^{3}y_i-3a-b\sum\limits_{i=1}^{3}x_i=0,\\ \sum\limits_{i=1}^{3}x_iy_i-a\sum\limits_{i=1}^{3}x_i-b\sum\limits_{i=1}^{3}x_i^2=0.\end{cases}$

若记 $\sum\limits_{i=1}^{3}y_i=3\overline{y},\sum\limits_{i=1}^{3}x_i=3\overline{x}$，则 $\begin{cases}\overline{y}-a-b\overline{x}=0,\\ \sum\limits_{i=1}^{3}x_iy_i-3a\overline{x}-b\sum\limits_{i=1}^{3}x_i^2=0.\end{cases}$

如果把它的解记作 $\hat{a},\hat{b}$，得到：

$$\hat{b}=\frac{\sum\limits_{i=1}^{3}x_iy_i-3\overline{x}\,\overline{y}}{\sum\limits_{i=1}^{3}x_i^2-3\overline{x}^2}=\frac{x_1y_1+x_2y_2+x_3y_3-3\overline{x}\,\overline{y}}{x_1^2+x_2^2+x_3^2-3\overline{x}^2}$$，① $\hat{a}=\overline{y}-\hat{b}\overline{x}$.②

①，②两式推广到n对数据 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$仍然成立，即：使 $[y_1-(a+bx_1)]^2+[y_2-(a+bx_2)]^2+\cdots+[y_n-(a+bx_n)]^2$达到最小的$a$，$b$取值为