46.2-非线性回归模型与回归效果分

举 题 说 法

非线性经验回归方程

视角1　指数型经验回归方程

例 1-1 云计算是信息技术发展的集中体现，近年来，我国云计算市场规模持续增长．从中国信息通信研究院发布的《云计算白皮书(2022年)》可知，我国2017年至2021年云计算市场规模数据统计表如下：

年份	2017年	2018年	2019年	2020年	2021年
年份代码x	1	2	3	4	5
云计算市场规模y/亿元	692	962	1 334	2 091	3 229

经计算得：$\sum\limits_{i=1}^{5}\ln y_i=36.33$，$\sum\limits_{i=1}^{5}(x_i\ln y_i)=112.85$.根据以上数据，建立$y$关于$x$的回归方程$\hat{y}=e^{\hat{b}x+\hat{a}}$(e为自然对数的底数).

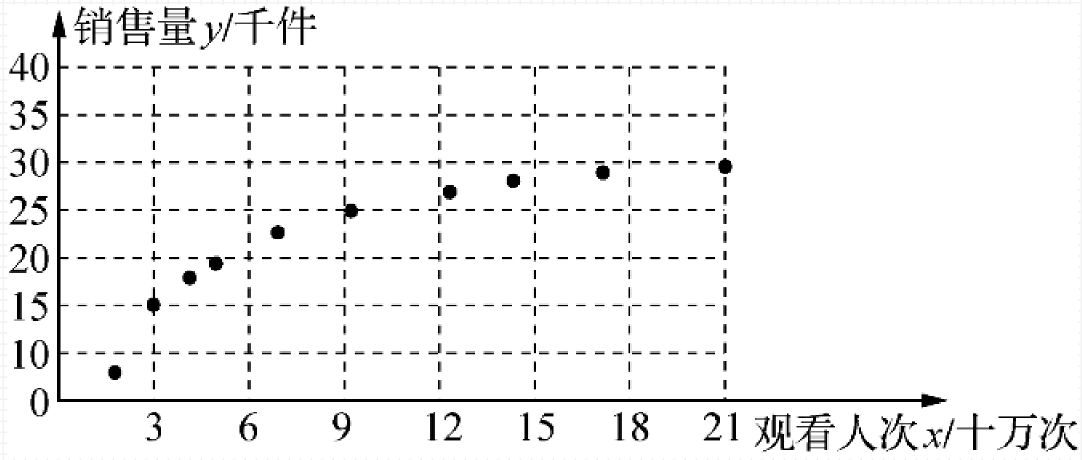
【解答】因为$\hat{y}=e^{\hat{b}x+\hat{a}}$，所以 $\ln\hat{y}=\hat{b}x+\hat{a}$，$\bar{x}=\dfrac{1+2+3+4+5}{5}=3$，所以$\hat{b}=$

$$\frac{\sum\limits_{i=1}^{5}(x_i\ln y_i)-\bar{x}\sum\limits_{i=1}^{5}\ln y_i}{\sum\limits_{i=1}^{5}x_i^2-n\bar{x}^2}=\frac{112.85-3\times36.33}{1+4+9+16+25-5\times3^2}=\frac{3.86}{10}=0.386$$，所以$\hat{a}=\frac{1}{5}\sum\limits_{i=1}^{5}\ln y_i-\hat{b}$

$\bar{x}=\frac{1}{5}\times36.33-0.386\times3=6.108$，所以$\hat{y}=e^{\hat{b}x+\hat{a}}=e^{0.386x+6.108}$.

视角2　对数型经验回归方程

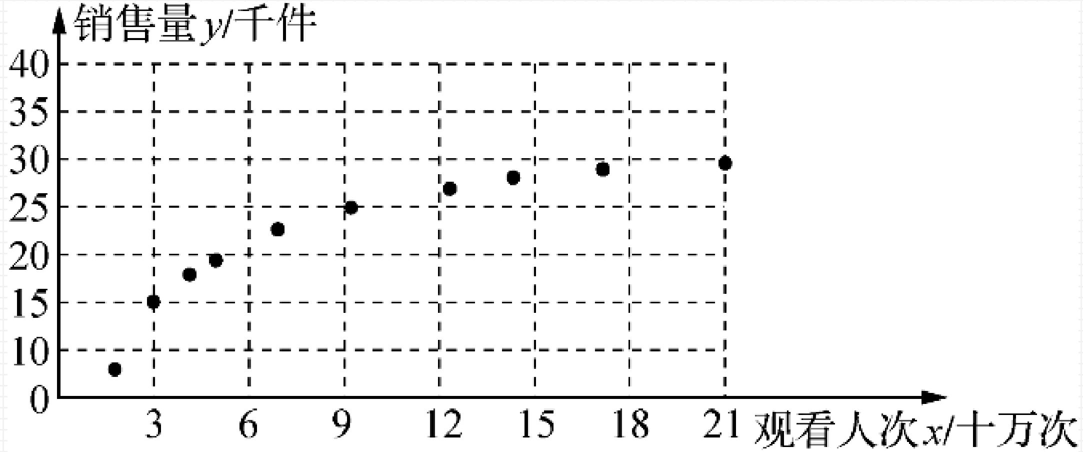
例 1-2 某乡村合作社借助互联网直播平台进行农产品销售，众多网红主播参与到直播当中，在众多网红直播中，统计了10名网红直播的观看人次x_i和农产品销售量$y_i(i=1，2，3，…，10)$的数据，得到如图所示的散点图.



(1) 利用散点图判断，$\hat{y}=\hat{a}+\hat{b}x$ 和$\hat{y}=\hat{c}+\hat{d}\ln x$ 哪一个更适合作为观看人次 x 和销售量 y 的回归方程类型(只要给出判断即可，不必说明理由).

【解答】由散点图可知，散点分布在一条对数型曲线附近，所以选择回归方程$\hat{y}=\hat{c}+\hat{d}\ln x$ 更适合.

例1-2 某乡村合作社借助互联网直播平台进行农产品销售，众多网红主播参与到直播当中，在众多网红直播中，统计了10名网红直播的观看人次x_i和农产品销售量y_i($i=1$，2，3，…，10)的数据，得到如图所示的散点图.



(2) 对数据作出如下处理，得到相关统计量的值如右表：

$\overline{x}$	$\overline{y}$	$\overline{\omega}$
9.4	30.3	2

$\sum\limits_{i=1}^{10}(x_i-\overline{x})^2$	$\sum\limits_{i=1}^{10}(\omega_i-\overline{\omega})^2$
366	6.6
$\sum\limits_{i=1}^{10}(x_i-\overline{x})(y_i-\overline{y})$	$\sum\limits_{i=1}^{10}(\omega_i-\overline{\omega})(y_i-\overline{y})$
439.2	66

其中令 $\omega_i=\ln x_i$，$\overline{\omega}=\dfrac{1}{10}\sum\limits_{i=1}^{10}\omega_i$.根据(1)的判断结果及表中数据，求 y 关于 x 的回归方程，并预测当观看人次为 280 万人时的销售量. ,参考数据：$\ln 2\approx0.69$，$\ln 7\approx1.95$.

【解答】令 $\omega=\ln x$，则$\hat{y}=\hat{c}+\hat{d}\omega$.因为$\sum\limits_{i=1}^{10}(\omega_i-\overline{\omega})(y_i-\overline{y})=66$，$\sum\limits_{i=1}^{10}(\omega_i-\overline{\omega})^2=6.6$，所以$\hat{d}=\dfrac{\sum\limits_{i=1}^{10}(\omega_i-\overline{\omega})(y_i-\overline{y})}{\sum\limits_{i=1}^{10}(\omega_i-\overline{\omega})^2}=\dfrac{66}{6.6}=10$，又$\overline{y}=30.3$，$\overline{\omega}=2$，所以$\hat{c}=\overline{y}-\hat{d}\overline{\omega}=30.3-10\times2=10.3$，所以 y 关于 ω 的经验回归方程为$\hat{y}=10.3+10\omega$，故 y 关于 x 的经验回归方程为$\hat{y}=10.3+10\ln x$.

令 $x=28$，代入经验回归方程可得$\hat{y}=10.3+10\ln 28=10.3+10\times(2\ln 2+\ln 7)\approx43.6$(千件)，所以预测观看人次为 280 万人时的销售量约为 43 600 件.

回归效果分析

视角1　残差分析

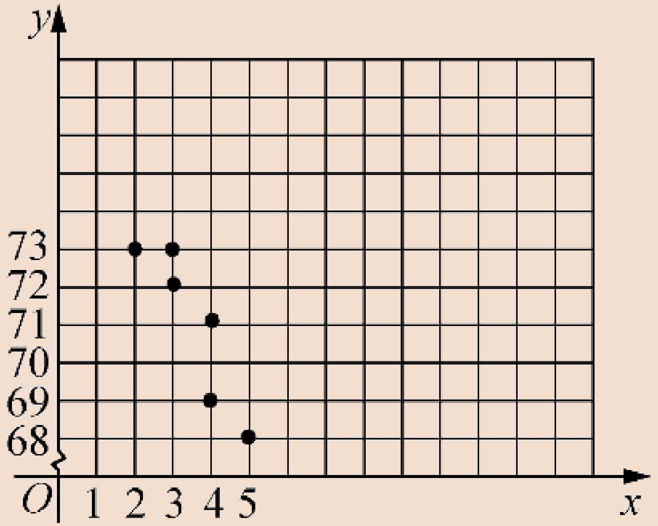
例2-1 假定产品产量x(单位：千件)与单位成本y(单位：元/件)之间存在相关关系.

数据如右：

x	2	3	4	3	4	5
y	73	72	71	73	69	68

(1) 以x为解释变量，y为预报变量，作出散点图；

【解答】散点图如图所示.



例 2-1 假定产品产量x(单位：千件)与单位成本y(单位：元/件)之间存在相关关系.数据如下：

x	2	3	4	3	4	5
y	73	72	71	73	69	68

(2) 求y与x之间的经验回归方程，当单位成本为70元/件时，预报产量为多少；

【解答】因为$\overline{x}=\frac{2+3+4+3+4+5}{6}=3.5$，$\overline{y}=\frac{73+72+71+73+69+68}{6}=71$，$\sum_{i=1}^{6}x_i^2=79$，$\sum_{i=1}^{6}x_iy_i=1\,481$，所以$\hat{b}=\frac{\sum_{i=1}^{6}x_iy_i-6\overline{x}\cdot\overline{y}}{\sum_{i=1}^{6}x_i^2-6\overline{x}^2}=\frac{1\,481-6\times3.5\times71}{79-6\times3.5^2}\approx-1.82$，$\hat{a}=\overline{y}-\hat{b}\overline{x}=71+1.82\times3.5=77.37$，

所以经验回归方程为$\hat{y}=-1.82x+77.37$.令$y=70$,则$70=-1.82x+77.37$,解得$x\approx4.05$

所以当单位成本为 70 元/件时，预报产量约为 4.05 千件.

例 **2-1** 假定产品产量x(单位：千件)与单位成本y(单位：元/件)之间存在相关关系.

数据如下：

x	2	3	4	3	4	5
y	73	72	71	73	69	68

(3) 计算各组残差，并计算残差平方和.

注：保留两位有效数字.

【解答】各组残差分别为$\hat{e}_1=y_1-\hat{y}_1=73-(-1.82\times 2+77.37)=-0.73$,

$\hat{e}_2=y_2-\hat{y}_2=72-(-1.82\times 3+77.37)=0.09$,

$\hat{e}_3=y_3-\hat{y}_3=71-(-1.82\times 4+77.37)=0.91$,

$\hat{e}_4=y_4-\hat{y}_4=73-(-1.82\times 3+77.37)=1.09$,

$\hat{e}_5=y_5-\hat{y}_5=69-(-1.82\times4+77.37)=-1.09$,

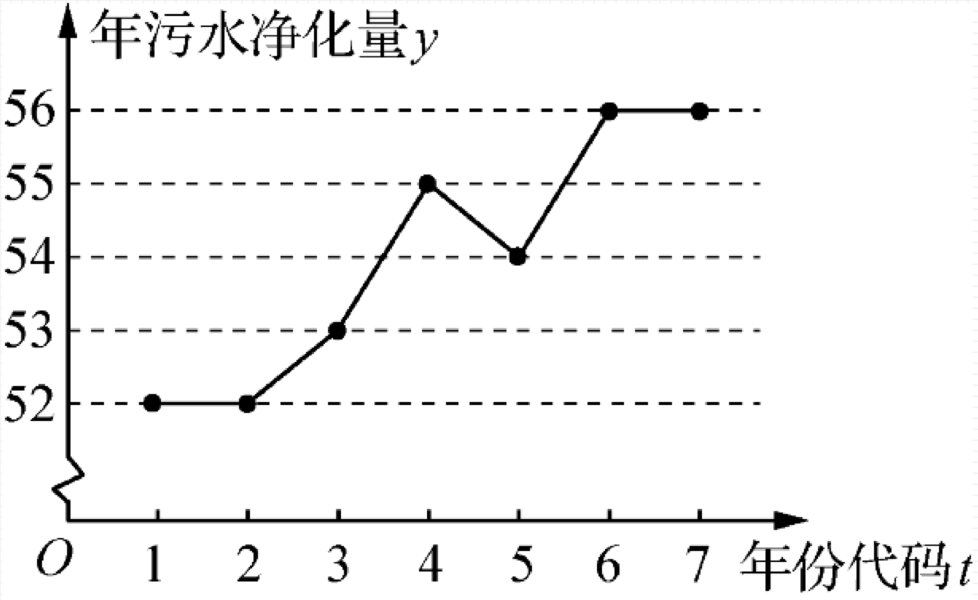
$\hat{e}_6=y_6-\hat{y}_6=68-(-1.82\times5+77.37)=-0.27$,

残差的平方和为$\sum\limits_{i=1}^{6}(y_i-\hat{y}_i)^2=(-0.73)^2+0.09^2+0.91^2+1.09^2+(-1.09)^2+(-0.27)^2=$

$3.818\ 2$.

视角2　决定系数

例2-2 如图是某企业2016年至2022年的污水净化量(单位：吨)的折线图.



注：年份代码1～7分别对应年份2016～2022.

参考数据：$\overline{y}=54$，$\sum\limits_{i=1}^{7}(t_i-\overline{t})(y_i-\overline{y})=21$，$\sqrt{14}\approx 3.74$，$\sum\limits_{i=1}^{7}(y_i-\hat{y}_i)^2=\dfrac{9}{4}$.

(1) 由折线图看出，可用线性回归模型拟合y和t的关系，请建立y关于t的回归方程，并预测2025年该企业的污水净化量；

【解答】由折线图中的数据得 $\overline{t}=4$，$\overline{y}=54$，$\hat{b}=\dfrac{\sum\limits_{i=1}^{n}(t_i-\overline{t})(y_i-\overline{y})}{\sum\limits_{i=1}^{n}(t_i-\overline{t})^2}=$

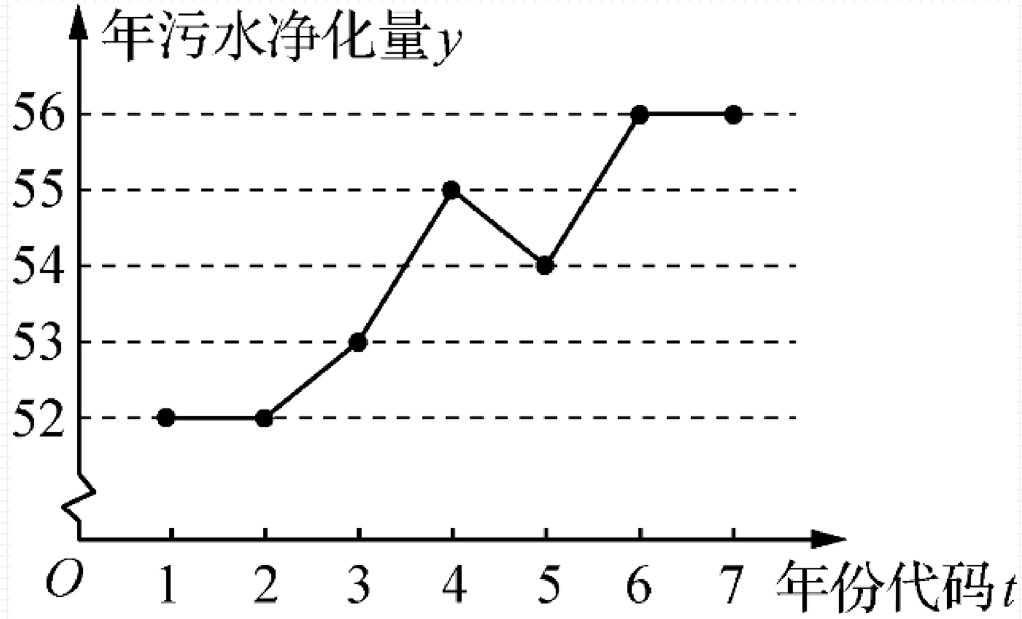
$\dfrac{21}{(1-4)^2+(2-4)^2+(3-4)^2+(4-4)^2+(5-4)^2+(6-4)^2+(7-4)^2}=\dfrac{21}{28}=\dfrac{3}{4}$，所以 $\hat{a}=\overline{y}-\hat{b}\overline{t}=54-\dfrac{3}{4}\times 4=51$，所以 y 关于 t 的经验回归方程为 $\hat{y}=\hat{b}t+\hat{a}=\dfrac{3}{4}t+51$.

将 2025 年对应的年份代码 $t=10$ 代入得 $\hat{y}=\dfrac{3}{4}\times 10+51=58.5$，所以预测 2025 年该企业污水净化量约为 58.5 吨.

例2-2 如图是某企业2016年至2022年的污水净化量(单位：吨)的折线图.

注：年份代码1～7分别对应年份2016～2022.

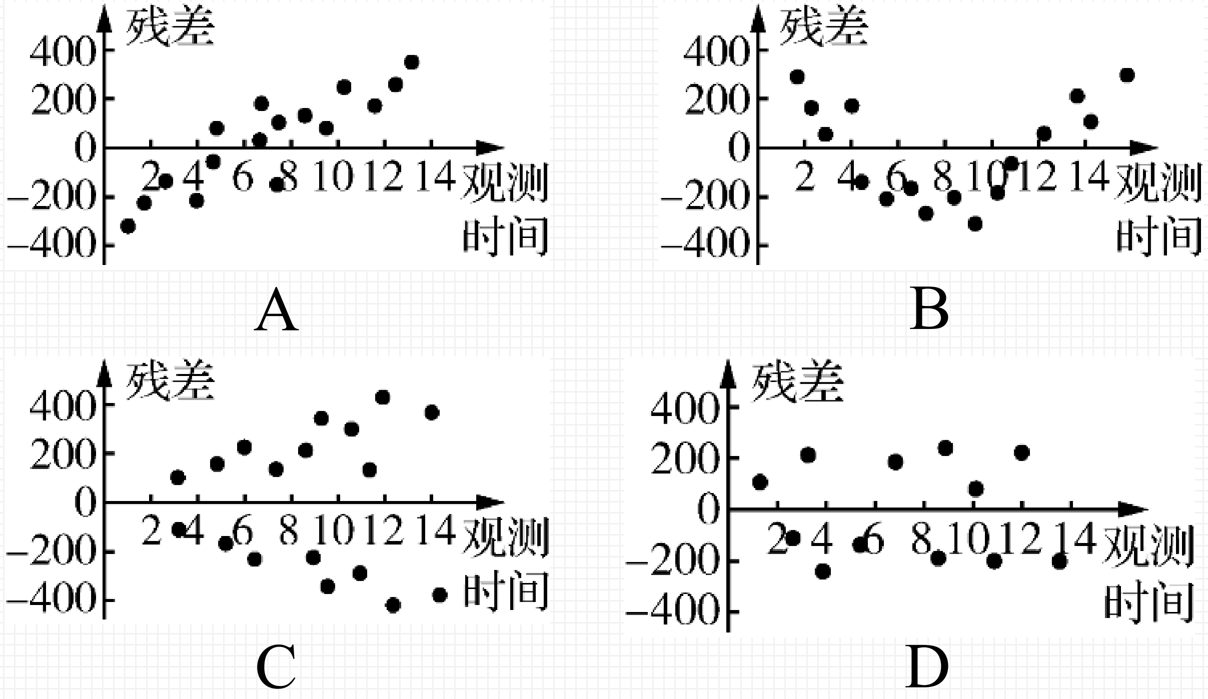
参考数据：$\overline{y}=54$，$\sum\limits_{i=1}^{7}(t_i-\overline{t})(y_i-\overline{y})=21$，$\sqrt{14}\approx3.74$，$\sum\limits_{i=1}^{7}(y_i-\hat{y}_i)^2=\frac{9}{4}$.



(2) 请用决定系数说明回归方程预报的效果.

随 堂练习

1．下列四幅残差分析与一元线性回归模型拟合精度最高的是　　　　　　(　　D　　)



【解析】由图知D中残差均匀分布在横轴附近，故D中图象与一元线性回归模型拟合精度最高.

【解答】因为 $R^2=1-\frac{\sum\limits_{i=1}^{7}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{7}(y_i-\overline{y})^2}=1-\frac{9}{4}\times\frac{1}{18}=1-\frac{1}{8}=\frac{7}{8}=0.875$，与 1 接近，说明回归方程预报的效果是良好的.

2. 经验表明，树高y与胸径x具有线性关系，为了解回归方程的拟合效果，利用右面数据计算残差，用来绘制残差图.

胸径 x/cm	18.2	19.1	22.3	24.5	26.2
树高的观测值 y/m	18.9	19.4	20.8	22.8	24.8
树高的预测值$\hat{y}$/m	18.6	19.3	21.5	23.0	24.4

则残差的最大值和最小值分别是 (C)

A．0.4，－1.8　　B．1.8，－0.4

C．0.4，－0.7　　D．0.7，－0.4

【解析】由表可得，各组数据的残差为18.9－18.6＝0.3，19.4－19.3＝0.1，20.8－21.5＝－0.7，22.8－23＝－0.2，24.8－24.4＝0.4，故残差最大值为0.4，最小值为－0.7.

3. 某部门统计了某地区今年前7个月在线外卖的规模如下表：

月份代号x	1	2	3	4	5	6	7
在线外卖规模y/百万元	11	13	18	★	28	★	35

其中 4，6 两个月的在线外卖规模数据模糊，但这 7 个月的平均值为 23.若利用经验回归方程$\hat{y}=\hat{b}x+\hat{a}$来拟合预测，且 7 月相应于点(7，35)的残差为-0.6，则$\hat{a}-\hat{b}=$

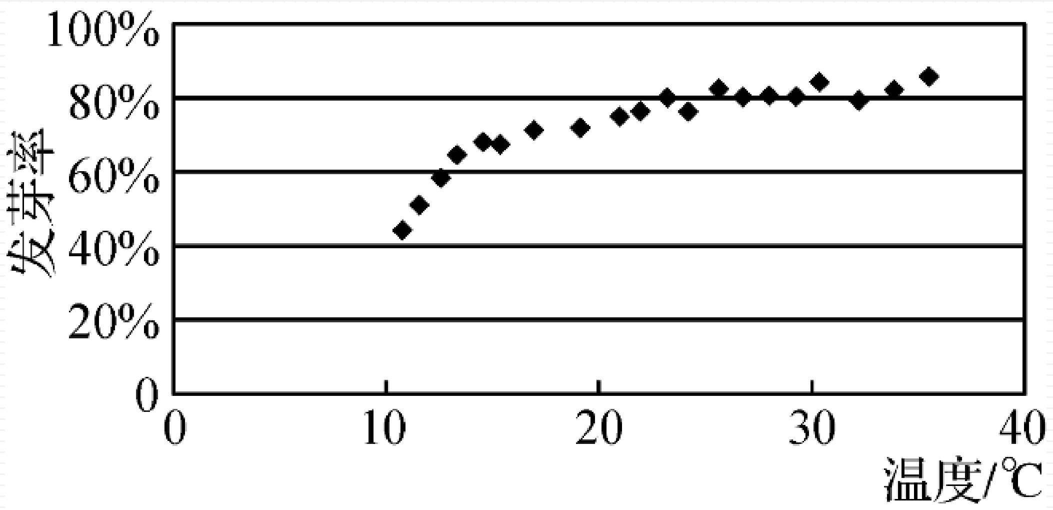
( )

A．1.0　　B．2.0　　C．3.0　　D．4.0

【解析】依题意，$\overline{x}=\frac{1}{7}\times(1+2+3+4+5+6+7)=4$，而 $\overline{y}=23$，于是得 $4\hat{b}+\hat{a}=23$，而当 $x=7$ 时，$35-(7\hat{b}+\hat{a})=-0.6$，即 $7\hat{b}+\hat{a}=35.6$，解得$\hat{a}=6.2$，$\hat{b}=4.2$，所以$\hat{a}-\hat{b}=2.0$.

【答案】B

4. 某校一个课外学习小组为研究某作物种子的发芽率y和温度x(单位：℃)的关系，在20个不同的温度条件下进行种子发芽实验，由实验数据$(x_i，y_i)(i=1，2，…，20)$得到下面的散点图：



由此散点图，在10 ℃至40 ℃之间，下面四个回归方程类型中最适宜作为发芽率y和温度x的回归方程类型的是 (　　)

A．$y=a+bx$　　B．$y=a+bx^2$

C．$y=a+be^x$　　D．$y=a+b\ln x$