第七章！£归© •{

§7.1 ˜元，形£归

- £归© •{ 理Ú计的 ‡工具，处理多个变量 间相关关系的˜ 学•{.
- 函ê： 定性关系. $h = \frac{1}{2}gt^2$.
- f关关X：给定x ，不能确定y的 . 如，存在测量 差.
- £归© ：建立关X，判断公 的有效性，预测！控 .
- ˜元，性£归：**随机变量** Y 与 **通变量** x 间的，性关X.
- 据成对观测：

$$(x_1;y_1);(x_2;y_2);\colon\colon\colon;(x_n;y_n).$$

- -**点**：$x_1;x_2;\cdots;x_n$ 不全f等.

例1.1. 某n 的强度Y 与拉 倍 x 有关. $n = 24$ 据:

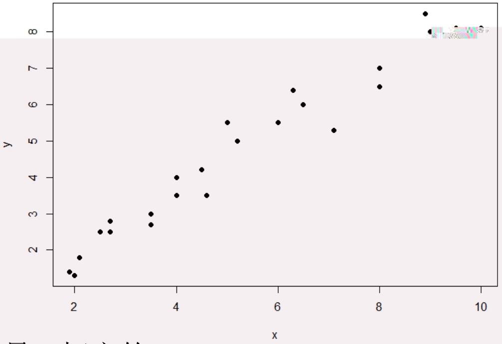
$$(1{:}9; 1{:}4), (2{:}0; 1{:}3), (2{:}1; 1{:}8), \cdots, (9{:}5; 8{:}1), (10{:}0; 8{:}1).$$

目标: 找出x 和Y 的关X .

- 散点 ã :

  散点 绕在

  ~ ^ , .



- 建立回归方程:

  $\hat{y} = a + bx,$

  £ 归X : b.

- 线性:

  x每增加1, y的变 z 量 恒定的.

- 非线性: x每增加1, y的变 z 量不 恒定的.

点估计$\hat{a}$, $\hat{b}$.

最小二乘法: 求均 • 差Q 的最小 点$\hat{a}; \hat{b}$.

$$Q = Q(a; b) = \sum_{i=1}^{n} \Big(y_i - (a + bx_i) \Big)^2 : \qquad (1.2)$$

- 方法一(微分法)! 求解 • 程组:

$$\frac{\partial Q}{\partial a} = -2 \sum_{i=1}^{n} \Big(y_i - (a + bx_i) \Big) = 0; \qquad (1.3)$$

$$\frac{\partial Q}{\partial b} = -2 \sum_{i=1}^{n} \Big(y_i - (a + bx_i) \Big) \cdot x_i = 0: \qquad (1.4)$$

- 由(1.3) 解得$a = \bar{y} - b\bar{x}$. 代入(1.4) ± 求解b.

- 由(1.3) 解得$a = \bar{y} - b\bar{x}$. 代入

$$\frac{\partial Q}{\partial b} = -2 \overset{n}{\underset{i=1}{¸}} \; y_i - (a + bx_i) \; \cdot x_i = 0: \qquad (1.4)$$

$\circ$

- 解: $\star \Rightarrow \; {}^{n}_{i=1} \; (y_i - \bar{y}) - b(x_i - \bar{x}) \; (x_i - \bar{x}) = 0.$ 记

$$`_{xx} = \overset{n}{\underset{i=1}{¸}} (x_i - \bar{x})^2; \quad `_{xy} = \overset{n}{\underset{i=1}{¸}} (x_i - \bar{x})(y_i - \bar{y}):$$

则$`_{xy} = b \cdot `_{xx}$. 解得$b = `_{xy}/`_{xx}$. (**前提**: x_i's 不全 f 等.)

- $\hat{a} = \bar{y} - \hat{b}\bar{x}$, $\hat{b} = \; \overset{\circ}{} {}^{n}_{i=1}(x_i - \bar{x})(y_i - \bar{y}) / \overset{\circ}{} {}^{n}_{i=1}(x_i - \bar{x})^2$.

可 ± 明$(\hat{a}; \hat{b})$ $Q(a; b)$ 的最小 点. (理由: 阶导 矩 /海色 H 定.)

$$H = \begin{matrix} 2n & & 2n\bar{x} \\ 2n\bar{x} & 2 & \overset{\circ}{} {}^{n}_{i=1}(x_i - \bar{x})^2 + n\bar{x}^2 \end{matrix} \quad :$$

最小二乘法: 求均 • 差Q 的最小 点$a, \hat{b}$.

$$Q = Q(a;b) = \sum_{i=1}^{n} \left(y_i - (a + bx_i) \right)^2 : \qquad (1.2)$$

- 方法二(方法)、(注: $\circ \sum_{i=1}^{n} \star \cdot \star_i = 0$, $\ell_{yy} = \circ \sum_{i=1}^{n} (y_i - \bar{y})^2$.)

$$Q(a;b) = \sum_{i=1}^{n} \left((y_i - \bar{y}) + \left(\bar{y} - (a + b\bar{x}) \right) - b(x_i - \bar{x}) \right)^2$$

$$= \ell_{yy} + n \cdot \star^2 + \underline{b^2 \ell_{xx} - 2b\ell_{xy}}$$

$$= \ell_{yy} + n \cdot \star^2 + $$ m1(xy))]-16/P66b10.9001)Tf[TJ10.907[

,909126

— •) 点

总结:

- 均 • 差Q 的最小 点 $\hat{a} = \bar{y} - \hat{b}\bar{x}$, $\hat{b} = \frac{\ell_{xy}}{\ell_{xx}}$.

$$Q = Q(a;b) = \sum_{i=1}^{n} \left(y_i - (a + bx_i) \right)^2: \qquad (1.2)$$

- 回归方程/回归†线/经验公ª:

$$y = \hat{a} + \hat{b}x:$$

- $\bar{y} = \hat{a} + \hat{b}\bar{x}$. Ï 此, (1) 点$(\bar{x};\bar{y})$ 落在£归 , 上, (2) $\hat{y}_i$'s 的平均 • $\bar{y}$.

$$\hat{y}_i := \hat{a} + \hat{b}x_i:$$

- 例1.1. 由24 个 据计算得$\hat{y} = \hat{a} + \hat{b}x = 0:15 + 0:859x$. £归X $\hat{b}$ 的含Â: 拉 倍 x 每增加1, 强度Y 平均增加0:859.

例1.2. 彩色w影 ，染料光学密度Y 与 出õ的光学密度x 关X如e，$A; B > 0$:

$$Y \approx Ae^{-B/x};$$

- 这不 ，性关X. 两边取对 得

$$Y^* \approx \ln A - Bx^*; \quad 其 \; ; Y^* = \ln Y; \; x^* = \frac{1}{x}:$$

- 据: $(x_1; y_1); \cdots; (x_n; y_n) \to (x_1^*; y_1^*); \cdots; (x_n^*; y_n^*)$.
- 建立，性£归•程: $\hat{y}^* = \hat{a} + \hat{b}x^*$.
- ‡变†: $\hat{A} = e^{\hat{a}}$, $\hat{B} = -\hat{b}$.
- £归•程: $\hat{Y} = \hat{A}e^{-\hat{B}/x}$.

例1.3. 炼钢钢包容 È Y 随 用次 x 增大. $n = 13$, 据:

$$(2; 106{:}42); (3; 108{:}20); (4; 109{:}58); \cdots; (19; 111{:}20)$$

- x 散点 ã . 用双曲 , $\frac{1}{y} \approx a + b \cdot \frac{1}{x}$.
- , 性 £ 归 • 程: $\hat{y}^* = \hat{a} + \hat{b}x^*$, $x^* = 1/x$, $y^* = 1/y$,

  其 $\hat{a} = 0{:}008967$, $\hat{b} = 0{:}0008292$.
- 经 公 :

$$\frac{1}{\hat{y}} = 0{:}008967 + 0{:}0008292 \cdot \frac{1}{x}{:}$$

- 前提: $x_1; x_2; : : : ; x_n$ 不全 f 等.

  结论: 均 • 差Q 的(˜ 的)最小 点 $\hat{a} = \bar{y} - \hat{b}\bar{x}$, $\hat{b} = \frac{`_{xy}}{`_{xx}}$.

$$Q = Q(a; b) = \sum_{i=1}^{n} \left(y_i - (a + bx_i) \right)^2 :$$

估计: $\hat{y} = \hat{a} + \hat{b}x$, $\bar{y} = \hat{a} + \hat{b}\bar{x}$, $\hat{y}_i = \hat{a} + \hat{b}x_i$.

- 注: 经 公 并不都能‡映 际情况. 需‡判别x与Y 间 Ä 真的具有 , 性 f 关关 $X : Y$ Ä 随着x增大 , 性地增大(½ 者 , 性地减小).

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2.$$

$\hat{a} = (1$

模型的 • 差估计与 f 关性检

- 态模型:

$$Y_i = a + bx_i + "_i; \quad i = 1; \cdots; n \qquad (1.10)$$

其 $"_1, \cdots, "_n$ 独立Ó©布, 都Ñ从$\sim \mathrm{N}(0; \ ^2)$, ^{2} .

- 注: ´ 得$(Y_1; \cdots; Y_n)$ 的联合密度.
- 可± 明: 残差平•和$Q = \mathring{}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \mathring{}\sum_{i=1}^{n} "_i^2$ 满足

$$\frac{1}{\ ^2}Q \sim \ ^2(n-2);$$

- $"_i = y_i - \hat{y}_i = y_i - \bar{y} - \hat{b}(x_i - \bar{x})$ 满足两个约 •程:

$$\sum_{i=1}^{n} "_i = 0; \qquad \sum_{i=1}^{n} "_i(x_i - \bar{x}) = 0:$$

- ^{2} 的无 估计: $\hat{}^2 = \frac{1}{n-2}Q$. (Ï $E\frac{1}{^2}Q = n - 2$.)

- 相关性检验. $H_0 : b = 0.$

(注: H_0 不成立, 则Y 与x 有，性 f 关关 X .)

- f 对 言, $U = \hat{b}^2 `_{xx}$ 越大& $Q = {}^{n}_{i=1}(y_i - \hat{y}_i)^2$ 越小, 则模型越准确描 x 和Y 间的，性 f 关关 X . ‡ ，则x 和Y 间没有，性 f 关关 X .

- 检验统计量:

$$F = F(x; Y) = \frac{U}{Q/(n-2)}: \tag{1.9}$$

若F f 当大, 则表明x 对Y 的，性影 • 强, 两者有，性 f 关性. Ä 则, 没有，性 f 关性.

- 可 明: 在H_0 e, $F \sim F(1; n-2)$.
- 否定域: $\mathcal{W} = \{(x; y) : F(x; y) > \}$, 其 $= F_{1-}(1; n-2)$.
- 结论: $F(x; y) >$ ，则拒绝H_0, 认 Y w 地，性 • 赖于x.

Ä 则, 接 H_0, 认 x 与Y 间没有 w 的，性 f 关性.

- 有些 的检 Ú计量取 本 f 关X ：

$$R = R(x;y) = `_{xy}/ \overset{a}{\overline{`_{xx}`_{yy}}}.$$

(£顾, 随Å变量 与 的(, 性) f 关X ： $= \frac{\mathrm{Cov}(\ ;\ )}{\sqrt{D(\ )D(\ )}}$.)

- 当$|R|$ 很大 , 拒绝$H_0 : b = 0$.
- 复 f 关X 平•：

$$R^2 = \frac{`^2_{xy}}{`^2_{xx}} \cdot \frac{`_{xx}}{`_{yy}} = \frac{\hat{b}^2 `_{xx}}{`_{yy}} = \frac{U}{`_{yy}} = 1 - \frac{Q}{`_{yy}}.$$

注: $0 \leqslant R^2 \leqslant 1$. R^2 越接近1, £归模型对 据拟合得越好.

- F 与R^2 的关X: ˜ ˜ 对应, 严格增,

$$F = \frac{U}{Q/(n-2)} = (n-2)\frac{U}{`_{yy} - U}$$

$$= (n-2)\frac{R^2}{1-R^2} = \frac{n-2}{1/R^2 - 1}.$$

例1.4. 炼钢，% 含量x 越高，Ž 炼　间Y 越长. $n = 34$,　据:

$$(180; 200); (104; 100); \cdots ; (143; 160)$$

- x 散点 ã，观察　据，建，形 £ 归模型.
- 计算结果: $\hat{a} = -23.20$, $\hat{b} = 1.270$, $F = 145.0$, $R^2 = 0.8192$.
- 查$F(1; n-2) = F(1; 32)$ 表: $\quad = 0.01$, 得 $\quad = 4.15$.
- e 结论: $F >$ ，Ä 定H_0, 认　x, Y 存在，性 f 关关 X. ½:
  ，£ 归　w　的.

应用1. 预报.

- 对新的自变量 x_0, 预报

$$Y_0 = a + bx_0 + {''}_0 :$$

- 点估计: 用$\hat{y}_0 = \hat{a} + \hat{b}x_0$ 预报Y_0.
- „ 需 ‡ 衡量预报精度.

  区间估计: 用 $\hat{\ }^2 = \frac{1}{n-2}Q$ 代 o ^{2}.

$$T := \frac{Y_0 - \hat{y}_0}{s\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\grave{\ }_{xx}}}} \sim t(n-2):$$

- 查$t(n-2)$ 表: $= t_{1- \ /2}(n-2)$.
- 区间两端点: $\hat{y}_{\pm} = \hat{y}_0 \pm s\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\grave{\ }_{xx}}}$.

- 区间估计: $= t_{1-\ /2}(n-2)$,

$$\hat{y}_0 - \ s\star;\ \hat{y}_0 + \ s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\grave{}_{xx}}}\ :$$

- x_0 离$\bar{x}$ 越远，预报区间长度越长.
  差标准差的估计s 越小, 预报区间越短, 预报越精确.
- 注: x_0 不能超出原　据的‰　.
- 当n 较大且$x_0 - \bar{x}$ 较小　, $\star \approx 1$. 于　预测区间近似

$$[\hat{y}_0 - \ s; \hat{y}_0 + \ s]:$$

- 当n 较大　,　可用标准　©布的临界　$z_{1-\ /2}$.

- ‡求控 Y 在区间$[A;B]$ 内, 如何选取x?
- 办{: 让$\hat{y}_{\pm} \in [A;B]$, ‡解出x 的区间.

£归　断和残差©

:d [( )]TJ/F53 Tf 9.04.000036 Td [(:d [( )]TJ/.Tf 9.0480.9

o

- 即　Ä定了H_0：　()　/　59 10.1 131 (　)　/　9091　7.0　(8)

- 残差: $\varepsilon_i = y_i - \hat{y}_i$.
- 令

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\ell_{xx}}, \quad s = \sqrt{\frac{Q}{n-2}}, \quad r_i = \frac{\varepsilon_i}{s\sqrt{1-h_i}}.$$

- 近似地, $r_1, \ldots, r_n$ f p 独立, 且都 Ñ 从N$(0,1)$. 故

$$P(|r_i| > 2) \approx 0.05.$$

- 当n 比较大 , r_i's 应该 有约$[0.05n]$ 个的绝对 大于2.
注: 可用来检 模型关于 差 ' 的假 Ä 成立, ± 及 u y É 常 点.