心理统计
第十四讲：回归初步

严超赣

Chao-Gan Yan, Ph.D.

yancg@psych.ac.cn
http://rfmri.org/yan

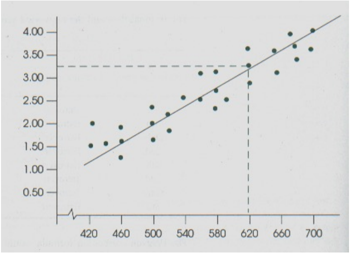
Institute of Psychology, Chinese Academy of Sciences

1

1

研究情境

- GRE与研究生GPA的关系



2

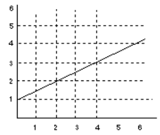
这条线有如下功用：

1. 它使GRE和GPA的关系容易被看到
2. 这条线确认了关系的“集中趋势”，对关系的简单化描述
3. 这条线可以用于预测。它建立了X与Y的精确关系。如GRE620分预测GPA3.25分

3

线性方程 （linear equations）

X = 0, Y = 1
X = 1, Y = 1.5
X = 2, Y = 2.0
X = 3, Y = 2.5
X = 4, Y = 3.0



每当 X 增加 1，Y就增加0.5. 称为**斜率（slope）**（b）.
斜率是一个常数.

截距（intercept）（a）是 X = 0时，Y 的值。截距也是一个常数.

这条线可以描述为以下**线性方程**:

Y = bX + a ---> Y = (.5)X + 1.0

用此线性方程，已知 X，b，和 a，即可确定 Y 的值

- 根据 X 预测 Y 是回归的基本目标

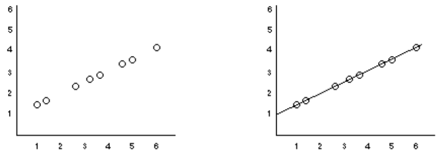
4

回归方程

$$\hat{Y} = .22(X) + .68$$

5

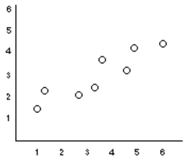
散点图: r = 1.0的情况.

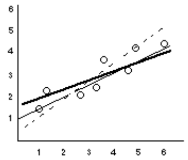


作回归分析就是试图找到一条直线(以及线性方程)以最佳地拟合数据点. 上例中是显而易见的. 只有一条可能最佳拟合线.

6

散点图:不完全相关的情况





此例中是最佳拟合线不是显而易见的。可能的拟合线不止一条。我们的目标是寻找最佳拟合线.

7

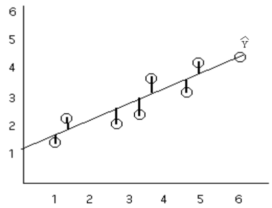
最佳拟合线

- 目标是使误差最小。即, 这条线与所有的数据点最近，是最佳拟合线.
- 回归线是给定X，a 和 b，用公式 (线性方程)来预测Y 的值。 我们的目标是找出一条线，以对Y作最佳估计. 即, 这条线使得所有Y 值的估计误差最小.

8

最小平方法（least-squares solution）

- 考察每个点, 比较Y的观测值与Y的预测值 $\hat{Y}$称为 "Y-hat")



距离 $= Y - \hat{Y}$

SS$_{误差}$ = 误差平方和 = 公式

$$\sum (Y - \hat{Y})^2$$

9

回归表达形式

- 简单线性回归（Simple Linear Regression)

$$\underset{\text{Data}}{Y} = \underset{\text{Model}}{\boxed{a + bX}} + \underset{\text{Error}}{\epsilon}$$

Y – 因变量, Dependent variable

X – 自变量, Independent (explanatory) variable

a – 截距, Intercept

b – 斜率, Slope

ϵ – 残差, Residual (error)

10

回归表达形式

- 线性代数表达

$$y_i = a + bx_i + \epsilon$$

$$\begin{cases} Y_1 = \alpha + \beta_1 X_1 + \epsilon_1 \\ Y_2 = \alpha + \beta_1 X_2 + \epsilon_2 \\ \cdots \\ Y_n = \alpha + \beta_1 X_n + \epsilon_n \end{cases}$$

11

回归表达形式

- 矩阵表达

$$Y_{n\times1} = X_{n\times2}\,\beta_{2\times1} + \epsilon_{n\times1}$$

$$Y = X\beta + \epsilon$$

$$Y_{n\times1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X_{n\times2} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{bmatrix} \quad \beta_{2\times1} = \begin{bmatrix} \alpha \\ \beta_1 \end{bmatrix} \quad \epsilon_{n\times1} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

12

示例

收入与问题解决能力是什么关系?

数据:"HouseholdIncome"(房产收入)和"Ravens.score"(瑞文推理测验)这两个变量进行研究(Eisenberg et al., 2019)

(Eisenberg et al. 2019)

13

示例

- 若我们只取前4个样本的数据:

Y	X
40000	2
19500	4
60000	3
81000	9

14

示例

- 用线性方程表示为:

$$\begin{cases} 40000 = \boldsymbol{\alpha} + 2\times\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1 \\ 19500 = \boldsymbol{\alpha} + 4\times\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_2 \\ 60000 = \boldsymbol{\alpha} + 3\times\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_2 \\ 81000 = \boldsymbol{\alpha} + 9\times\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_4 \end{cases}$$

15

示例

- 更简洁的表达方式——矩阵:

$$\begin{bmatrix} 40000 \\ 19500 \\ 60000 \\ 81000 \end{bmatrix} = \boldsymbol{\alpha} + \boldsymbol{\beta} \begin{bmatrix} 2 \\ 4 \\ 3 \\ 9 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

16

补充知识:矩阵运算

矩阵的加减法

$$X = \underbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}}_{\text{m列 m columns}} \left.\vphantom{\begin{bmatrix} 1\\1\\1\\1 \end{bmatrix}}\right\} \text{n行 n rows} \qquad Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix}$$

$$X \pm Y = \begin{bmatrix} x_{11} \pm y_{11} & x_{12} \pm y_{12} & \cdots & x_{1m} \pm y_{1m} \\ x_{21} \pm y_{21} & x_{22} \pm y_{22} & \cdots & x_{2m} \pm y_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} \pm y_{n1} & x_{n2} \pm y_{n2} & \cdots & x_{nm} \pm y_{nm} \end{bmatrix}$$

17

补充知识:矩阵运算

$$C = A + B = \begin{bmatrix} 2 & 4 & -1 \\ 1 & 8 & 7 \\ 3 & 5 & 6 \end{bmatrix} + \begin{bmatrix} 7 & 5 & 2 \\ 9 & -3 & 1 \\ 2 & 1 & 8 \end{bmatrix} = \begin{bmatrix} 9 & 9 & 1 \\ 10 & 5 & 8 \\ 5 & 6 & 14 \end{bmatrix}$$

18

补充知识：矩阵运算

矩阵的乘法

假设有维数有$n \times m$ 的矩阵X，以及维度为$l \times k$的矩阵Y，则

（1）当 $m = l$ 时，矩阵X乘以矩阵Y才是可行的，结果矩阵XY才存在；

（2）当 $k = n$ 时，矩阵Y乘以矩阵X才是可行的，结果矩阵YX才存在。

下面假设$m = l$ 成立，则矩阵Y的维数可以表示成$m \times k$。设矩阵X乘以矩阵Y得到的结果矩阵为C；矩阵C的维数为$n \times k$，其第i行第j列元素遵循如下的计算公式：

$$c_{ij} = \sum x_{ih} y_{hj}$$

19

补充知识：矩阵运算

$1 \times 3 + 9 \times 5 + 7 \times 6 = 90$

$$C = AB = \begin{bmatrix} 1 & 9 & 7 \\ 8 & 1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 2 & 1 & 5 \\ 5 & 4 & 7 & 3 \\ 6 & 9 & 6 & 8 \end{bmatrix} = \begin{bmatrix} 90 & 101 & 106 & 88 \\ 41 & 38 & 27 & 59 \end{bmatrix}$$

20

补充知识：矩阵运算

矩阵的转置

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \qquad X^T = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1m} & x_{2m} & \cdots & x_{nm} \end{bmatrix}$$

21

补充知识：矩阵运算

$$A = \begin{bmatrix} 1 & 5 \\ 4 & 8 \\ 7 & 9 \end{bmatrix} \qquad A' = A^T = \begin{bmatrix} 1 & 4 & 7 \\ 5 & 8 & 9 \end{bmatrix}$$

22

回归表达形式

- **矩阵表达**

$$\boldsymbol{Y_{n\times1} = X_{n\times2} \overset{\text{矩阵乘法}}{\downarrow} \beta_{2\times1} \overset{\text{矩阵加法}}{+} \epsilon_{n\times1}}$$

$$Y_{n\times1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad X_{n\times2} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{bmatrix} \qquad \beta_{2\times1} = \begin{bmatrix} \alpha \\ \beta_1 \end{bmatrix} \qquad \epsilon_{n\times1} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

23

示例

收入与问题解决能力是什么关系？

数据：“HouseholdIncome”(房产收入)和“Ravens.score”(瑞文推理测验)这两个变量进行研究(Eisenberg et al., 2019)

(Eisenberg et al. 2019)

24

示例

$$Y_{n\times1} = X_{n\times2}\,\beta_{2\times1} + \epsilon_{n\times1}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

25

示例

$$Y_{n\times1} = X_{n\times2}\,\beta_{2\times1} + \epsilon_{n\times1}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{\text{Ravens},1} \\ 1 & x_{\text{Ravens},2} \\ \vdots & \vdots \\ 1 & x_{\text{Ravens},n} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

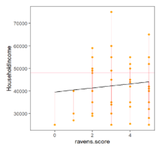
26

最小二乘法

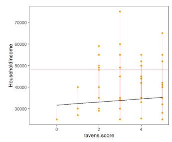
要建立一元线性回归方程，就要先计算方程中的参数a和b。根据最佳拟合原则，回归线是指散点图中每一个点沿Y轴方向到该直线的距离平方和最小的那条直线，即使使误差平方和最小，这就是常规最小二乘法(ordinary least squares, OLS)的基本思想。

27

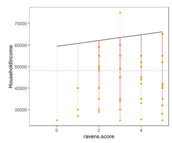
最小二乘法



$\alpha^*, \beta^* = argmin \sum \epsilon^2$



$\alpha^*, \beta^* \neq argmin \sum \epsilon^2$



$\alpha^*, \beta^* \neq argmin \sum \epsilon^2$

28

最小二乘法

一般形式

$$\alpha^*, \beta^* = argmin \sum \epsilon^2$$

$$\alpha^*, \beta^* = argmin \sum (Y - a - bX)^2$$

29

最小二乘法

$$\alpha^*, \beta^* = argmin \sum (Y - a - bX)^2$$

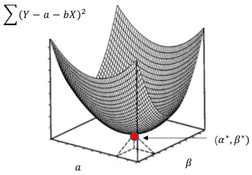
要使$\sum(Y - a - bX)^2$最小，则$\frac{\partial \sum(Y-a-bX)^2}{\partial a} = 0$，$\frac{\partial \sum(Y-a-bX)^2}{\partial b} = 0$

即$-2\sum(Y - a - bX) = 0$，$-2\sum(XY - aX - bX^2) = 0$

$$a = Y - \mathrm{b}X \qquad b = \frac{\sum(X - \bar{\bar{X}})(Y - \bar{\bar{Y}})}{\sum(X - \bar{X})^2} = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}}$$

30

最小二乘法



令 $\frac{\partial\sum(Y-a-bX)^2}{\partial a}=0$， $\frac{\partial\sum(Y-a-bX)^2}{\partial b}=0$，

(α^*,β^*)位于最低点

$\sum(Y-a-bX)^2$ 最小

31

最小二乘法

矩阵形式

$$Y = X\beta + \epsilon$$

$$\epsilon_{n\times1}=\begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \vdots\\ \epsilon_n\end{bmatrix} \quad \epsilon_{n\times1}{}^T=\begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \vdots\\ \epsilon_n\end{bmatrix}^T$$

$$\epsilon_{n\times1}{}^T\epsilon_{n\times1}=\begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \vdots\\ \epsilon_n\end{bmatrix}^T\begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \vdots\\ \epsilon_n\end{bmatrix}=\sum\epsilon_i^2$$

32

最小二乘法

矩阵形式

$$Y = X\beta + \epsilon$$

用矩阵形式表达其残差平方和

$$\begin{aligned}\mathrm{SSR} &= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\\ &= (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\\ &= \mathbf{y}'\mathbf{y}-\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}-\mathbf{y}'\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\\ &= \mathbf{y}'\mathbf{y}-2\mathbf{y}'\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\end{aligned}$$

33

最小二乘法

矩阵形式

用矩阵形式表达其残差平方和

$$\begin{aligned}\mathrm{SSR} &= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\\ &= (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\\ &= \mathbf{y}'\mathbf{y}-\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}-\mathbf{y}'\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\\ &= \mathbf{y}'\mathbf{y}-2\mathbf{y}'\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\end{aligned}$$

以β为参数，求残差平方和对β的偏导数

$$\frac{\partial(\mathrm{SSR})}{\partial(\boldsymbol{\beta})}=-2\mathbf{X}'\mathbf{y}+2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}=0$$

用矩阵形式表达估计的β

$$\mathbf{b}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

34

斜率和截距

- 最佳拟合线斜率的公式是:
  b = SP/SS_X =r S_Y/S_X
- 最佳拟合线截距的公式是:
  a =$Y-bX$

35

在前述相关的例子中

X	Y	$X-\bar{X}$	$Y-\bar{Y}$	$(dev_X)(dev_Y)$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$
0	1	−6	−1	6	36	1
10	3	+4	+1	4	16	1
4	1	−2	−1	2	4	1
8	2	+2	0	0	4	0
8	3	+2	+1	2	4	1

和 30 10 14 64 4

均值 6.0 2.0

SP = 14; SS_X = 64; SS_Y = 4

- 斜率 = b = SP/SS_X = 14/64 = .22
- 截距 = a = $Y-bX$ = 2.0 - (.22)(6.0) = .68

36

解释回归要注意

- 预测值不是百分之百准确的 (除非 $r = \pm 1.0$). 注意图中数据点并没有位于回归线上. 所以有残差（误差）. 估计的标准误描述了用来估计 Y 的典型误差。
- 回归方程不能对 X值范围之外的数据作出预测。 这一点在相关中已有说明。

37

回归估计的标准误

- 回归方程允许我们作出预测，但未给出预测准确性的信息
- 估计的标准误给出了回归线与数据点之间标准距离的量度
- 回归估计的标准误在概念上类似标准差

38

如何计算估计的标准误?

- 首先计算误差的平方和

$$SS_{误差} = \sum (Y - \hat{Y})^2$$

- 将误差的平方和除以自由度 (即误差的方差, 或误差的均方)
  - $SS_{误差} / df$
  - $df = n - 2$
- 为求得估计的标准误，将误差的方差取平方根 (类似标准差)
- 最后得到公式: $S_{误差} = \sqrt{\frac{SS_{error}}{df}} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{SS_{error}}{df}} = \sqrt{MS_{error}}$

39

在上例中

	X	Y	$\hat{Y}$	$(Y - \hat{Y})$	$(Y - \hat{Y})^2$
	0	1	0.68	.32	.102
	10	3	2.88	.12	.014
	4	1	1.56	-.56	.314
	8	2	2.44	-.44	.193
	8	3	2.44	.56	.314
和	30	10	10	0	.937
均值	6.0	2.0			

SP = 14; SSX = 64; SSY = 4; r = 0.875

$$\hat{Y} = .22(X) + .68$$

$$S_{误差} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{SS_{error}}{df}} = \sqrt{\frac{.937}{3}} = .559$$

40

标准误与相关系数之间的关系

- $SS_{误差} = (1 - r^2)SS_Y$

41

回归方程和标准误

- 回归方程描述了最佳拟合线和预测值，估计的标准误和相关系数则提供了预测的误差的信息

42

计算 S误差的简便方法：利用相关系数

- $SS_{误差} = (1 - r^2)SS_Y = (1 - (+0.875)^2)(4) = (1 - .766)(4) = .9375$

$$S_{误差} = \sqrt{\frac{SS_{误差}}{df}} = \sqrt{\frac{.937}{3}} = .559$$

43

X	Y
1	4
2	1
3	7
4	13
5	10

1. 求由X预测Y的回归方程
2. 用回归方程求Y的预测值
3. 求每个数据点Y- $\hat{Y}$ ，将其平方，并求和
4. 计算积差相关系数
5. 用该相关系数和SSy计算SS误差

44

X	Y	
1	4	SS_X=10
2	1	
3	7	SS_Y=90
4	13	
5	10	SP=24

$b=SP/SS_X=24/10=2.4$
$a=Y_{bar}-bX_{bar}=7-2.4*3=-0.2$
回归方程是$Y_{hat}=2.4X-0.2$
$Y_{hat1}= 2.4*1-0.2=2.2$
$Y_{hat2}= 2.4*2-0.2=4.6$
$Y_{hat3}= 2.4*3-0.2=7$
$Y_{hat4}= 2.4*4-0.2=9.4$
$Y_{hat5}= 2.4*5-0.2=11.8$
$SS_{error}=(4-2.2)^2+(4.6-1)^2+(7-7)^2+(9.4-13)^2+(11.8-10)^2=32.4$
$r=SP/sqrt(SS_xSS_Y)=24/sqrt(10*90)=0.8$
$SS_{error} = (1 - r^2)SS_Y = (1-0.64)*90=32.4$

45

一元线性回归的假设检验

Source	df	SS	MS	F
Model	1	$\sum(\hat{y}_i - \bar{y})^2$	SS_{Model}/df_{Model}	MS_{Model}/MS_{Error}
Error	N - 2	$\sum(y_i - \hat{y}_i)^2$	SS_{Error}/df_{Error}	
Total	N - 1	$\sum(y_i - \bar{y})^2$		

46

一元线性回归的假设检验

截距项a

$$S_a^2 = MS_{error}\left[\frac{1}{n} + \frac{\bar{X}}{\sum(X-\bar{X})^2}\right] = \frac{MS_{error}\sum X^2}{n\sum X^2 - (\sum X)^2}$$

$$t = \frac{a-0}{\sqrt{S_a^2}} = \frac{a-0}{SE}, \quad df = n-2$$

路径系数b

$$S_b^2 = \frac{MS_{error}}{\sum(X-\bar{X})^2} = \frac{MS_{error}}{SS_X}$$

$$t = \frac{b-0}{\sqrt{S_b^2}} = \frac{b-0}{SE}, \quad df = n-2$$

47

一元线性回归的效应量

- R-Squared

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

$$R^2 = 1 - \frac{SS_{error}}{SS_Y}$$

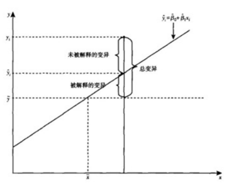


图 3－4 判定系数 R^2 的含义

48

一元线性回归的效应量

校正复相关系数

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

$$R^2 = 1 - \frac{SS_{error}}{SS_Y}$$

$$R^2_{adj} = 1 - \frac{SS_{error}/df_{error}}{SS_Y/df_Y}$$

49

一元线性回归的统计前提

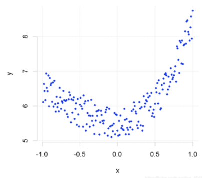
1. 模型设定假定(线性预设)
2. 正交预设
3. 残差方差齐性预设
4. 正态分布预设

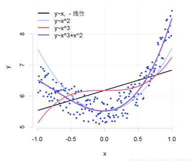
50

一元线性回归的统计前提

1. 线性预设

(1) 该预设规定Y的条件均值是自变量X的线性函数，若变量之间是非线性关系，线性回归的拟合效果不佳；



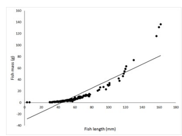


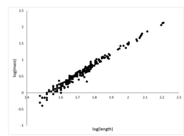
http://tecdat.cn/r

51

一元线性回归的统计前提

(2) 在某些情况下，我们会遇到非线性函数的形式。最常见的变换方式就是对数变换。





52

一元线性回归的统计前提

2. 正交预设

(1) 误差项ϵ和x不相关；

(2) 误差项ϵ的期望值为0。

注意：不管正交假定是否成立，最小二乘估计在计算中已运用了这一预设。

53

一元线性回归的统计前提

2. 正交预设

若误差项ϵ的期望值不为0，

E(y)=E(a+bx)+E(ϵ)。

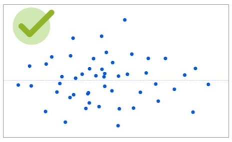
如果E(ϵ)不为零，回归系数将有无穷解，结果就是总体回归方程无法通过样本去估计。

54

一元线性回归的统计前提

3. 残差方差齐性

对X变量的每一个可能的值，Y变量有相同的总体方差。即残差的方差不受X变量取值的影响。

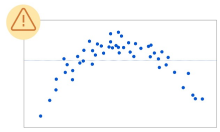


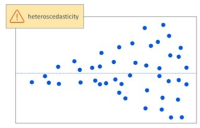
55

一元线性回归的统计前提

3. 残差方差齐性

若残差方差不齐性，误差较大的观测值将对拟合模型产生更大的影响。





56

一元线性回归的统计前提

4. 残差正态分布预设

该预设规定残差项ϵ独立且同分布。

实际中无法确定ϵ的分布。

对于大样本数据，可根据中心极限定理对参数进行统计推断。然而在小样本情况下，我们只有假定ϵ服从正态分布时才能使用t检验。

57

一元线性回归的统计前提

4. 正态分布预设

若ϵ不满足正态分布预设，OLS的估计是有偏的。

58

作业

1.对于下列数据：

X	Y
1	2
4	7
3	5
2	1
5	14
3	7

a. 找出回归方程
b. 对于数据中的每一个X，计算Y的预测值
c. 计算X与Y的相关系数
d. 计算残差的和方SS_{error}

59

59