【统计简单学】

第八单元

回归分析

授课教师：唐丽英 教授

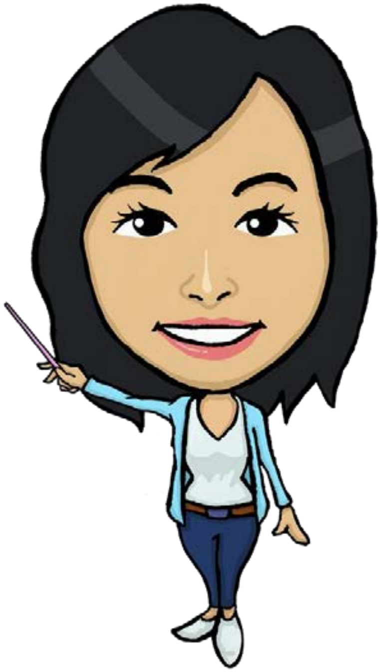
新竹交通大学
工业工程与管理学系

第八单元 内容大纲

- **第一部份：回归分析介绍**
  - 回归分析之意义
  - 回归分析之用途
  - 回归模式之类型
  - 回归分析之资料
  - 相关分析简介
  - 回归分析流程

- **第二部份：简单回归分析**

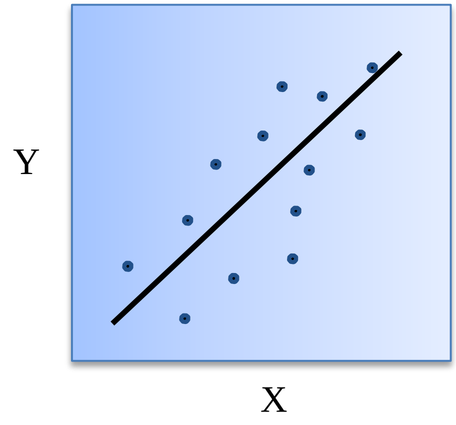
- **第三部份：复回归分析**



第一部份：回归分析介绍

回归分析介绍

- 回归分析之意义
  - 回归分析主要是探讨一个或数个「自变数」(independent variable) 和一个「依变数」(dependent variable) 间的关系，进而建构自变数与依变数之关系式(或数学方程式)，称为回归模式（regression model），此过程即为回归分析。
  - 在回归分析中，自变数通常以 X 表之，依变数则以 Y 表之。



回归分析介绍

- 回归分析之用途
  - **描述资料**：自变数与依变数间的关系。
  - **估计或预测**：利用回归模式及特定之自变数值估计或预测依变数值。
  - **控制**：给定依变数值或特性，求解自变数值。

  - 例如：某工程师欲研究某种化学合成之反应物含量（Y）是否受到合成时间（X）的影响，并希望建立反应物含量与合成时间之数学方程式，以预测在不同的合成时间下之反应物含量。

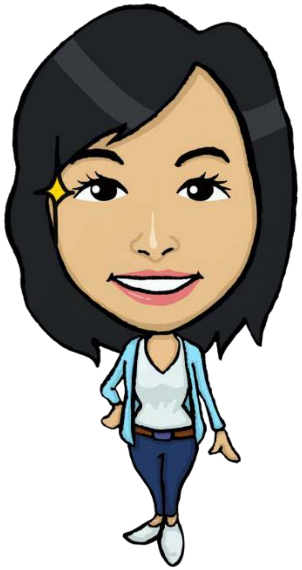
回归分析介绍

- **回归模式之类型**
  - 有时影响依变量之自变数的个数不只一个，例如：反应物含量（Y）可能与时间（X_1）、温度（X_2）、催化剂的种类（X_3）及其他自变数均有关；运用回归分析可分析各自变数是否显着，并进而建构回归模式以预测反应物含量。
  - 回归模式有以下两种：
    - **简单回归（Simple Regression）模式**：回归模式中只考虑一个自变数。
    - **复回归模式或多元回归（Multiple Regression）模式**：回归模式中考虑的自变数超过一个。

回归分析介绍

- **例1：请指出下列两例各适合适配何种回归模式？**
  1) 由房子大小(X)来预测房价(Y)?

  2) 由房子大小(X_1)、房龄(X_2)、离市区距离(X_3)、有无空调(X_4) 等因素来预测房价(Y)?

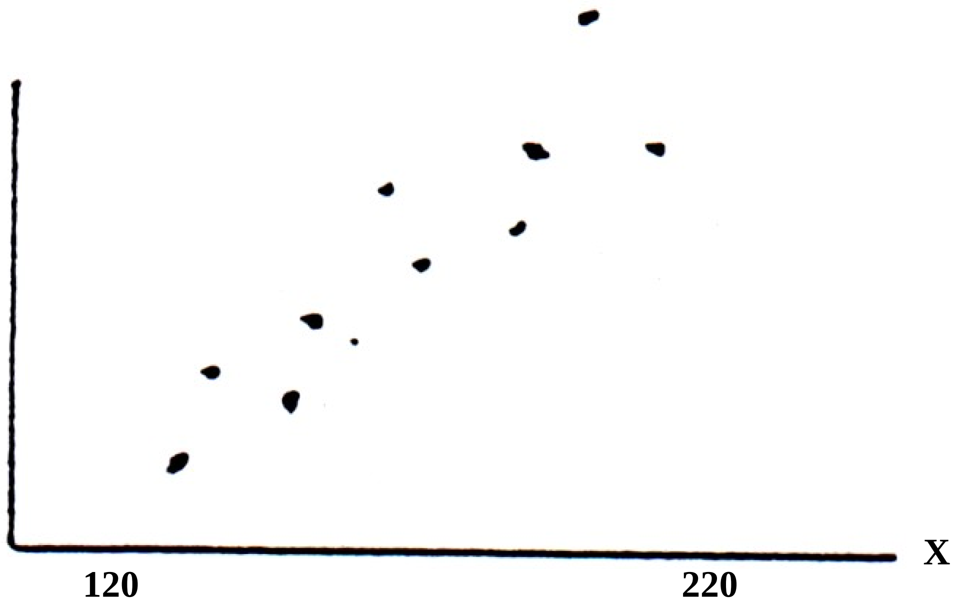


回归分析介绍

- **回归分析之资料**
  - 回归分析之资料常见的来源有以下二种：
    - A. 观察资料（observational data）
      - **观察资料来自非实验（non-experiment design）的研究**，此类资料通常无法说明、提供适当的自变数与依变数间因果关系之回归分析结果。因此，以观察的方式搜集资料进行回归分析时，须注意是否尚有其他未考虑到的自变数比回归式中的自变数更能解释自变数和依变数间的关系。
    - B. 实验资料（experimental data）
      - 实验资料是先规划实验，再以随机的方式执行实验以得到实验资料，因此，以**实验资料所得到之回归分析结果通常可说明自变数与依变数间的因果关系**。
  - 注意事项
    - 在回归分析中**数据必须能代表所欲研究的变量范围**。
    - 在作回归分析之前须先检查数据有无离群值。

回归分析介绍

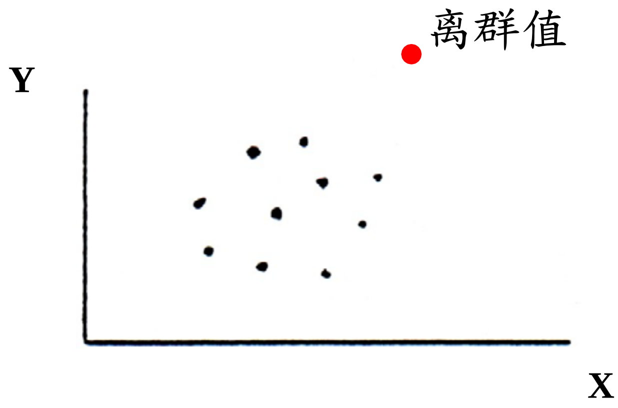
- 在回归分析中资料必须能代表所欲研究的变数范围



120 < X < 220

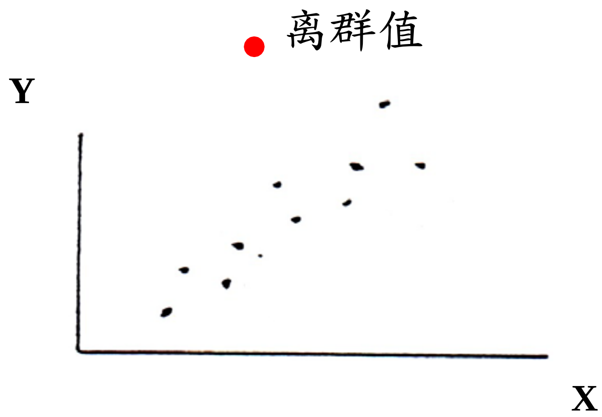
回归分析介绍

- 在作回归分析之前须先检查资料有无离群值



含离群值之r^2：0.67

不含离群值之r^2：0.001



含离群值之r^2：0.023

不含离群值之r^2：0.88

回归分析介绍

- **相关分析简介**

在作回归分析之前，可以先绘制自变数与依变数之散布图，以瞭解两变数间呈以下何种关系：

1） **正相关(Positive Relationship)**

- 假如X增加，则Y增加；或X减少，而Y减少，称为X与Y有正相关。

2） **负相关(Negative Relationship)**

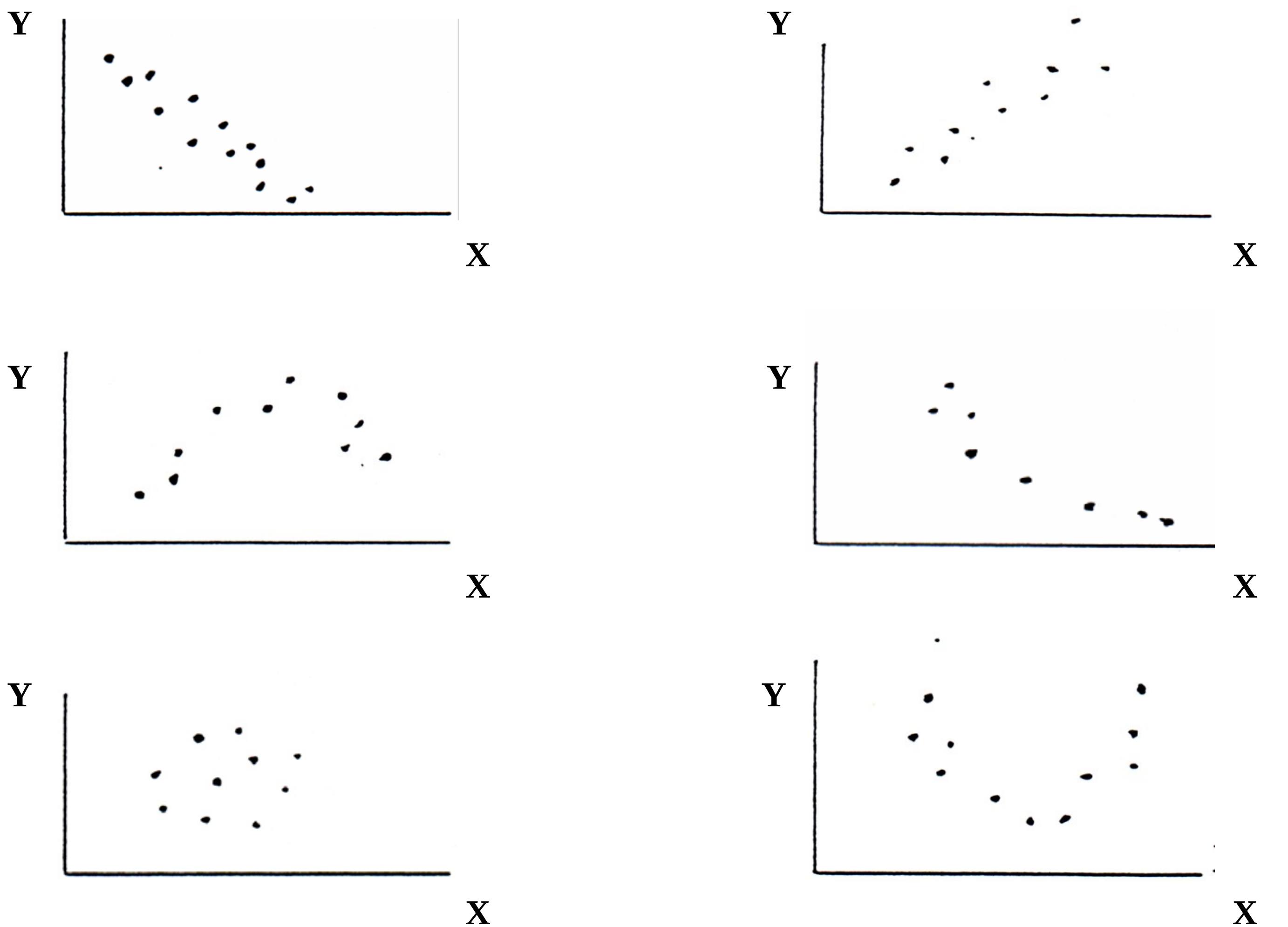
- 假如X增加，则Y减少；或X减少，而Y增加，称为X与Y有负相关。

3） **不相关(No Relationship)**

- 在散布图中之点大部份与水平轴平行，看不出任何特殊图形。

回归分析介绍

- 例2：请指出下列六张图中，X与Y变数间各具有何种关系？



回归分析介绍

- 回归分析**流程**

步骤1：搜集n组自变数与依变数之样本数据，如：(x, y) = (高中成绩,大学成绩)或(x, y) = (温度,良率)等，建立回归模式。

步骤2：评估回归模式适配资料之好坏：

方法(1) **观察(X, Y)散布图**➔数据越接近回归模式则表示回归模式与数据适配越佳。

方法(2) **利用r�指标**➔ $0 \le r^2 \le 1 \rightarrow r^2$ 越接近 1 越佳

。 方法(3) **利用ANOVA F test及 t test** ：

➔ a) ANOVA F test 之 p值 < 0.05，表整体而言，回归模式与资料适配佳。

➔ b) t test 之 p-值 < 0.05，表各别 之预测变量 X 会显著地影响 Y 。

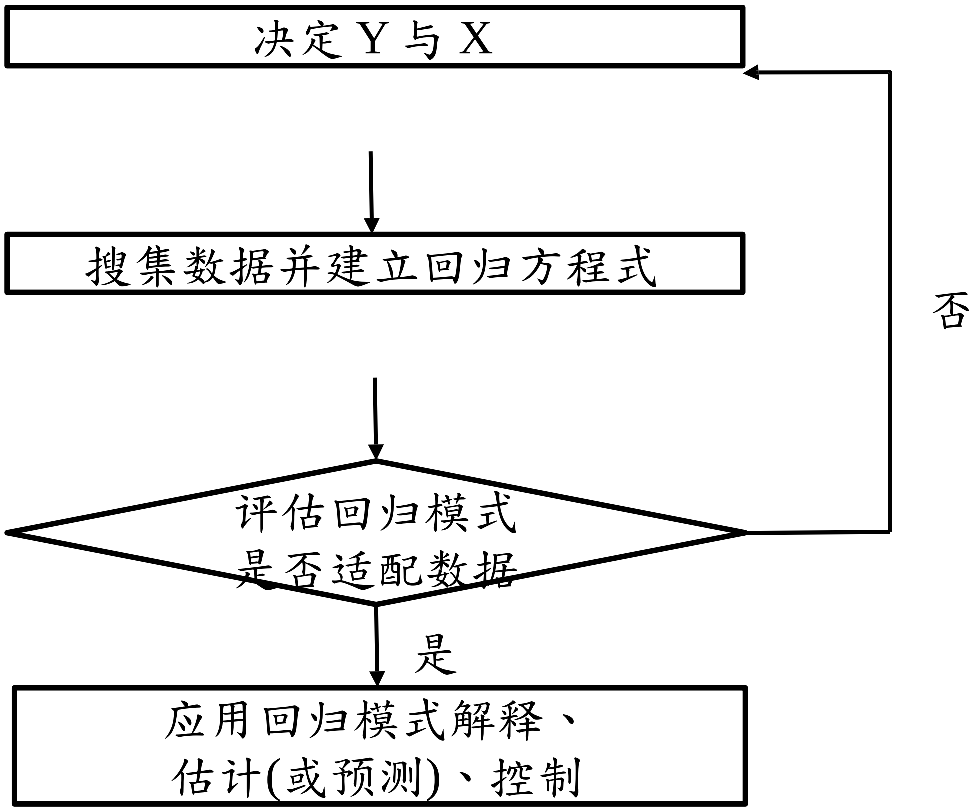
步骤3：应用回归模式

(1)解释X, Y间之关系。

(2)给定X值以估计Y值或给定Y值以找出最适之X值。

回归分析介绍

- 回归分析**流程**



第二部份：简单回归分析

简单回归分析

- 一个自变数与依变数间最简单的关系即为直线关系，简单直线回归模式可以表示如下：

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

其中

- y_i 表示依变量之第 i 个观察值，$i = 1, \ldots, n$；
- x_i 表示对应于第 i 个观察值之自变量值；
- α 与 β 分别为适配之回归系数；
- α 表截距
- β 表斜率，即自变量每增加一单位时，依变量的平均改变量；
- ε_i 为随机误差项。

简单回归分析

- **何谓残差(Residual) e_i？**
  - 观察值与其预测值间的差异(即 $e_i = y_i - \hat{y}_i$) 称为「残差」(residual)，残差值愈小表示回归模式的解释能力愈强，此回归模式预测依变量的效果亦会愈佳。
  - 利用**最小平方法（Least Square Method）**（即最小化残差值平方和的方法）可以得到回归模式中各回归系数的估计式，从而建构出回归模式。

- **最小误差平方法之公式**

$$\min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$$

其中 $e_i = y_i - \hat{y}_i$ 表示第 i 個觀察值下的殘差值，且 $\sum_{i=1}^{n} e_i = 0$。

简单回归分析

- 简单回归模式中各回归系数的估计式
  - 样本简单线性回归模式之公式如下：

$$\hat{y}_i = a + bx_i$$

其中 $\hat{y}_i$ 为在特定 x_i 下 y 的平均值，a 与 b 分别为 α 与 β 之估计值。

利用最小误差平方法可推得 a 与 b 的公式如下：

$$b = \frac{SS_{xy}}{SS_x}$$

$$a = \bar{y} - bx$$

其中

$$SS_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = \sum xy - (\sum x)(\sum y)/n$$

$$SS_x = \sum(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - (\sum x)^2/n$$

简单回归分析－回归式好坏之评断

- 如何定义回归分析中的三个变异量？
  - 在回归模式中，为了要评估自变数预测依变数的能力，必须要知道下列三个变异数的衡量值：

1) 总变异量（Total Variation）：

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \boxed{\sum_{i=1}^{n} {y_i}^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}} = SS_y$$

2) 回归模式解释的变异量（Explained Variation）：

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \boxed{bSS_{xy}} = \frac{(SS_{xy})^2}{SS_x}$$

3) 其他因素解释的变异量（Unexplained Variation）：

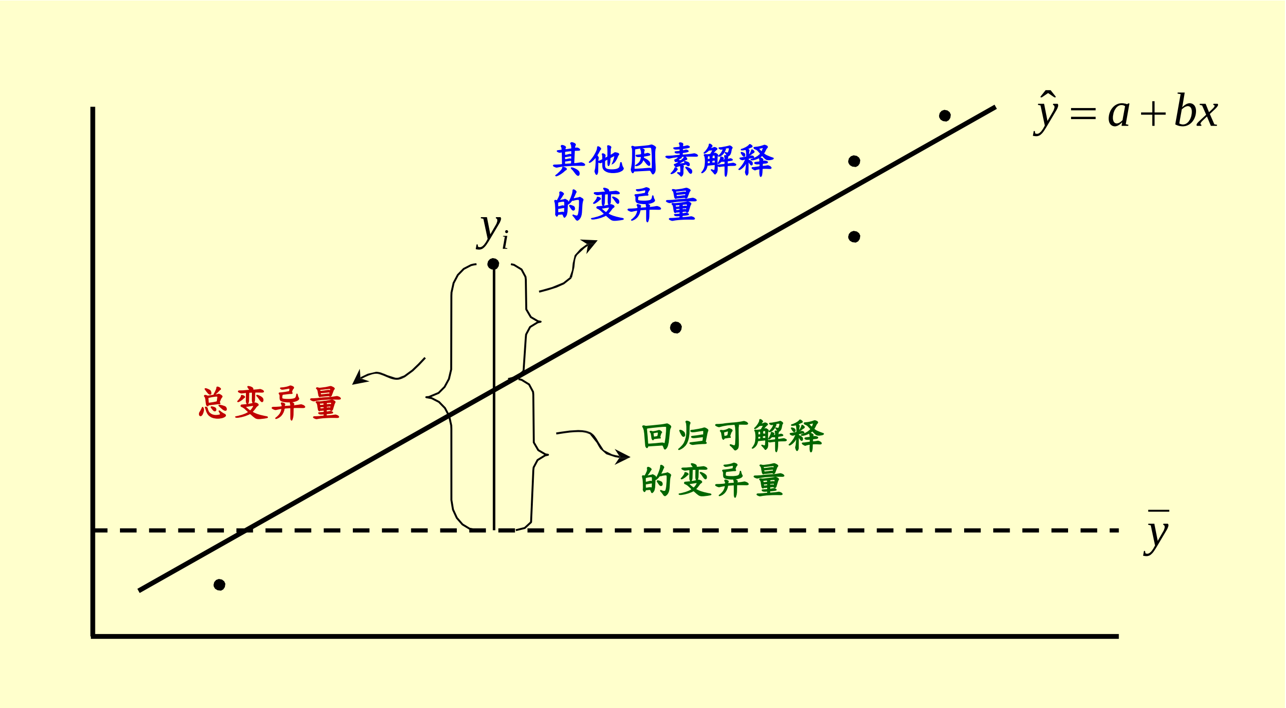
$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \boxed{SST - SSR}$$

简单回归分析 – 回归式好坏之评断

- 总变异量 = 回归模式解释的变异量 + 其他因素解释的变异量

SST = SSR + SSE

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

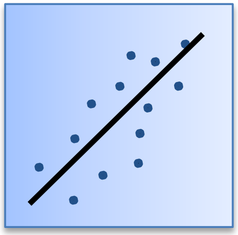


简单回归分析 – 回归式好坏之评断

- **如何判断 X 是否对预测 Y 提供了有用的资讯（或回归方程式是否有用）？**

1) 由图形判定（限简单回归）

- 数据点与回归线越接近表回归方程式越有用。



2) 判定系数（Coefficient of Determination，r^2）

- 判定系数是用来衡量自变数(x)所能解释依变数(y)之变异量占总变异量的百分比。 r^2值介于0与1之间，其值越接近1表示回归模式与资料适配越佳。

$$r^2 = \frac{\text{回归可解释的变异量}}{\text{总变异量}} = \frac{SSR}{SST} \qquad (0 \le r^2 \le 1)$$

- 相关系数 r (Correlation Coefficient) 是用来衡量两个随机变数 x 与 y 间 <u>直线</u>关系的方向与强弱。r 可由$\pm\sqrt{r^2}$ 求得，‘＋’或‘—’符号则与斜率b同。
  - ✓ $-1 \leqq r \leqq 1$
  - ✓ r=0 并不表示 y 与 x 间<u>无</u>关系，仅表 y 与 x 间无<u>线性</u>关系。

简单回归分析 – 回归式好坏之评断

3) 统计检定—F检定及t检定

- 利用统计检定可检定 X 是否对预测 Y 提供了有用之信息。假设 X 与 Y 之间完全无关，亦即在预测 Y 值时 X 几乎不提供任何有用的信息，则在线性模式：$y=\alpha+\beta x$ 中，β 值应等于零。

- 对β进行检定。若 $H_0: \beta=0$ 被推翻，则可下结论：有足够的证据显示「X 与 Y 之间有显著之线性关系」或「X 在对 Y 的预测上提供了有用的资讯」。

- 估计与检定回归系数时之统计假设：
  - 假设误差项ε_i为独立且服从平均数为0和变异数为σ^2 的 常态 分配（即 $\varepsilon_i \sim N(0,\sigma^2)$）。

- 关于 $H_0:\beta=0$ 有两种检定：

  a) **ANOVA F test**：检定整个回归模式之显着性。

  b) **t test**：检定各别 X 对 Y 之显著性。

简单回归分析－回归式好坏之评断

a) ANOVA F test

– 使用ANOVA的程序来检定所有之X与Y之间是否有显着之线性关系。

变异来源	平方和	自由度	均方	F
回归	SSR	1	SSR/1=MSR	MSR/MSE
误差	SSE	n-2	SSE/(n-2)=MSE=$\hat{\sigma}^2$	
总和	SST	n-1		

– 检定程序：

① 假设：$\varepsilon_i \sim NID(0,\sigma^2)$

② $H_0:\beta=0$ (X与Y之间没有线性关系；或对预测Y而言，回归或无法提供有用之资讯)
$H_a:\beta\neq0$ (X与Y之间有线性关系，亦即斜率不为0；或在预测Y上，回归式有用)

③ 检定值：$F=\dfrac{MSR}{MSE}$

④ 弃却域：查F-表，自由度=(1,n-2) 或 p-值

⑤ 下结论

简单回归分析 – 回归式好坏之评断

b) t test

- 检定程序：
  ① 假设：$\varepsilon_i \sim NID(0,\sigma^2)$

  ② $H_0: \beta=0$
  $H_a: \beta\neq 0$

  ③ 检定值： $t = \dfrac{b-\beta}{\sqrt{MSE/SS_x}}$

  ④ 弃却域：查 t- 表，自由度= n-2 或 p-值

- **在简单回归中，ANOVA F 检定与 t 检定有何关系？**
  - 对检定相同的假设，即 $H_0: \beta=0$ ，ANOVA F 检定与 t 检定间的关系为： $F_{(1,r)} = t^2_{(r)}$

简单回归分析范例

- 例3：某产品之良率会随温度之增加而减少，经实验得结果如下表。

温 度 (F)	190	191	193	195	198	187	199	200
良率 (%)	88	88	86	85	83	89	84	83

试回答下列问题。

1) 本例中之自变数与依变数各为何?
2) 请画出x-y散布图并判断自变数与依变数之关系为何?
3) 请适配简单线性回归模式。
4) 请判断回归模式是否适配原始资料?
5) 请解释回归系数b在本例中之意义为何?
6) 请预测当温度为192时，则估计之平均良率为何?

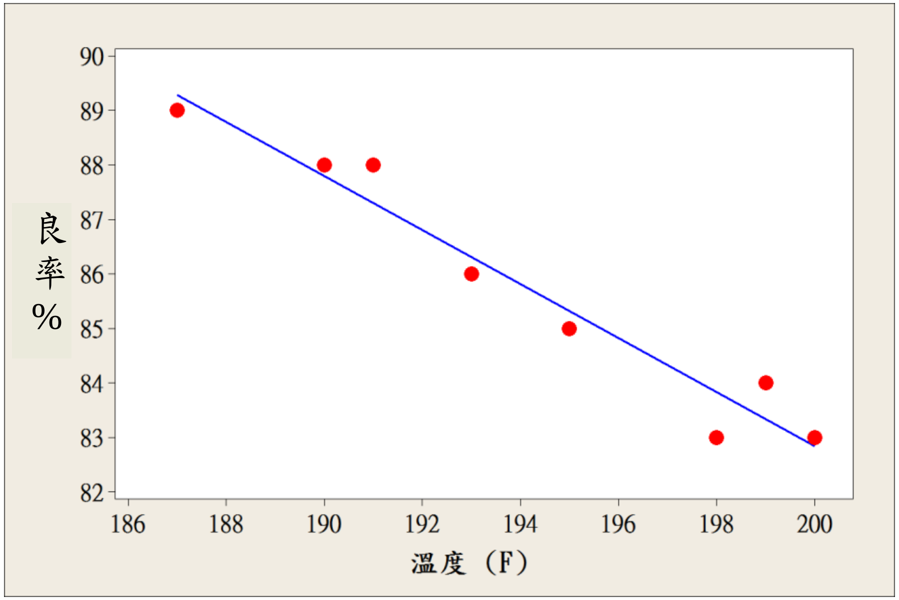
简单回归分析范例

例3：【解】

1) 本例中之自变数与依变数各为何?

- 本例「温度」为自变数；「良率」为依变数。

2) 请画出x-y散布图并判断自变数与依变数之关系为何?



自变量与依变量间有 __负向线性__ 之关系。

简单回归分析范例

3) 请适配简单线性回归模式

【将数据整理如下表】

计算表						令u=(x-194) 令v=(y-86)
x	y	u	v	uv	u^2	v^2
190	88	-4	2	-8	16	4
191	88	-3	2	-6	9	4
193	86	-1	0	0	1	0
195	85	1	-1	-1	1	1
198	83	4	-3	-12	16	9
187	89	-7	3	-21	49	9
199	84	5	-2	-10	25	4
200	83	6	-3	-18	36	9
总计		1	-2	-76	153	40

请注意 **u=x-194**， **v=y-86**

简单回归分析范例

3) 请适配简单线性回归模式

【套用公式计算回归系数】

$$b=\frac{SS_{xy}}{SS_x}=\frac{-75.75}{152.875}=-0.4955$$

$$SS_{xy}=SS_{uv}=\sum uv-\frac{\sum u\sum v}{n}=-76-\frac{(1)(-2)}{8}=-75.75$$

$$\overline{Y}=86+\frac{-2}{8}=85.75$$

$$SS_x=SS_u=\sum u^2-\frac{(\sum u)^2}{n}=153-\frac{(1)^2}{8}=152.875$$

$$\overline{X}=194+\frac{1}{8}=194.125$$

$$SS_y=SS_v=\sum v^2-\frac{(\sum v)^2}{n}=40-\frac{(-2)^2}{8}=39.5$$

$$a=\overline{Y}-b\overline{X}=85.75-(-0.4955)(194.125)=85.75-(-96.189)=181.939$$

$$\boxed{\hat{Y}=a-bX=181.939-0.4955X}$$

→ 回归模式

简单回归分析范例

3) 请适配简单线性回归模式

【Minitab 报表】

Session

Regression Analysis: 良率(%) versus 温度 (F)

The regression equation is
良率(%) = 182 – 0.496 温度 (F)

Predictor	Coef	SE Coef	T	P
Constant	181.939	8.989	20.24	0.000
温度 (F)	-0.49550	0.04629	-10.70	0.000

S = 0.572372   R-Sq = 95.0%   R-Sq(adj) = 94.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	37.534	37.534	114.57	0.000
Residual Error	6	1.966	0.328		
Total	7	39.500			

Obs	温度 (F)	良率(%)	Fit	SE Fit	Residual	St Resid
1	190	88.000	87.794	0.278	0.206	0.41
2	191	88.000	87.298	0.249	0.702	1.36
3	193	86.000	86.307	0.209	-0.307	-0.58
4	195	85.000	85.316	0.206	-0.316	-0.59
5	198	83.000	83.830	0.270	-0.830	-1.65
6	187	89.000	89.280	0.387	-0.280	-0.66
7	199	84.000	83.334	0.303	0.666	1.37
8	200	83.000	82.839	0.339	0.161	0.35

简单回归分析范例

4) 请判断回归方程式是否适配原始资料?

【由图形判定】

检视 2) XY之散布图，在此散布图中画出回归模式。

令X_1= 190 ，Y_1＝ 87.794 ；X_2= 195 ，Y_2＝ 85.317 ；

连结（X_1,Y_1）及（X_2,Y_2）可绘制回归方程式。

简单回归分析范例

4) 请判断回归方程式是否适配原始资料?

【判定系数（Coefficient of Determination，r²）】

计算判定系数 r^2

$$SST = SS_y = SS_v = 39.5$$

$$SSR = bSS_{xy} = (-0.4955)(-75.75) = 37.5341$$

$$SSE = SST - SSR = 1.966$$

$$r^2 = \frac{SSR}{SST} = \frac{37.5341}{39.5} \cong 0.95$$

温度(X)的变动会引起95%良率的变动。
即温度确实是影响良率的一个重要因素。

计算相关系数 r

$$r = \sqrt{r^2} \cong 0.975$$

温度 含有率间有非常强的负向线性关系。（r之符号同斜率b之符号）
在本例中，简单线性模式是一个很好的预测模式。

简单回归分析范例

4) 请判断回归方程式是否适配原始资料?

【统计检定—F检定及 t 检定】

利用ANOVA F test来检定温度(x)与含氧率(y)间是否有显著的直线关系

变异来源	平方和	自由度	均方	F
回归	37.5341	1	37.5341	37.5341/0.3277 = 114.54
误差	1.966	6	0.3277	
总和	39.5	7		

① 假设：$\varepsilon_i \sim NID(0,\sigma^2)$

② H_0: $\beta=0$ (x与y之间没有线性关系)
Ha: $\beta\neq0$ (x与y之间有线性关系亦即斜率不为0)

③ 检定值：F=<u>114.54</u>

④ 弃却域：临界值$F_{0.05,1,6}$ =<u> 5.99 </u>

⑤ 下结论：温度与良率间<u>有</u>显著的直线关系。

简单回归分析范例

4) 请判断回归方程式是否适配原始资料?

【统计检定—F检定及 t 检定】

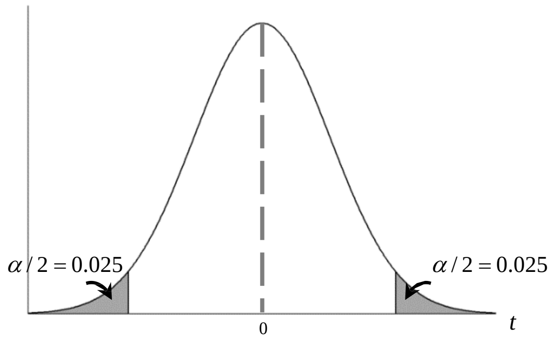
利用 t tset 来检定温度(x)是否为良率(y)的一个有用的预测变数

① 假设：$\varepsilon_i \sim NID(0,\sigma^2)$

② $H_0: \beta=0$
$H_a: \beta\neq 0$

③ 检定值：$t=\dfrac{b-\beta}{\sqrt{MSE/SS_x}}=\dfrac{-0.4955-0}{\sqrt{0.3277/152.875}}\cong -10.702$

④ 弃却域：临界值$t_{0.025,7}$=<u>2.365</u>



⑤ 下结论：温度(x)<u>　是　</u>良率(y)的一个有用的预测变量。

简单回归分析范例

5) 请解释回归系数b在本例中之意义为何?

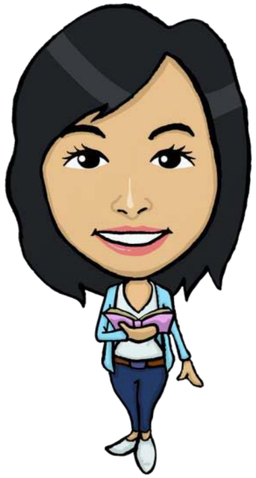
b = -0.4955：表示当温度每增加1°F时，良率平均会下降0.4955%。

6) 请预测当温度为192，则估计之平均良率为何?

$$\hat{Y} = a - bX = 182.461 - 0.4955(192) = 87.325$$

简单回归分析范例

- **作回归分析时应注意事项：**
  - 利用回归式估计 y 时，**所给定之 x 值必须在样本之 X 值范围内，Y 之估计值才会准确。**上例中，当所给定之 x 值介于 187 与 200 间，y之估计值才会准确。
  - **回归模式显着并不表示自变数与依变数间一定有 因果 关系**。其因果关系也可能是经由第三变数或其他理论依据而成立。



第三部份：复回归分析

复回归分析

- **复回归分析简介**
  - 复回归分析之主要目的在探讨多个自变数（X）与依变数（Y）之间的关系，并将其关系以数学方程式表之。
  - 复回归分析目前已广泛地应用在各个领域中，包括自然科学、社会科学、管理、工程、医药等。
  - 例如：房地产中介商希望能利用房子所在之地区（X_1）、房子的型式（透天、别墅、公寓）（X_2）、房子的坪数（X_3）、有无公共设施（X_4）等多个变数来建立一个复回归模式以预测房价（Y）
  - 例如：工程师想要利用温度（X_1）、压力（X_2）、催化剂种类（X_3）等多个变数来建立一个复回归模式以预测某化学实验之收成率（Y）。

复回归分析

- **复回归模式**
  - 复回归模式之形式如下：

$$y=\alpha+\beta_1 x_1+\cdots+\beta_i x_i+\cdots+\beta_k x_k+\varepsilon$$

其中

- y 表依变数值；
- �表截距；
- x_i, i=0,1,…,k，表自变数值；
- �$_i$, i=0,1,…,k，表对应于各自变量之回归系数；
- ε 表随机误差。

  - 复回归模式之假设
    - $\varepsilon_i \sim NID(0,\sigma^2)$。

复回归分析

- 例：两个自变数之复回归模式
  - $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

    其中
    - α 表截距；
    - β_1 表当 x_2 为固定常数时，x_1 对 y 之斜率；
    - β_2 表当 x_1 为固定常数时，x_2 对 y 之斜率。
- 样本资料建构之复回归模式
  - $\hat{y} = a + b_1 x_1 + b_2 x_2$

    其中
    - a 表 $\widehat{\alpha}$；
    - b_1 表 $\widehat{\beta_1}$；
    - b_2 表 $\widehat{\beta_2}$。
- 如何估计复回归模式之系数？
  - a) 利用统计软件
  - b) 解联立方程序

复回归分析范例

- 例4：某洗碗精制造商欲研究洗碗精销售量与广告费用及促销费用间的关系，以了解广告与促销费用对于销售量的影响。现收集相关变数之资料如下表所示，试进行复回归分析。

时间周期	销售量（万元）	广告费用（万元）	促销费用（万元）
1	70	16	14
2	63	15	16
3	51	8	9
4	71	12	27
5	79	20	22
6	81	16	37
7	80	23	25
8	44	5	7
9	59	10	10
10	61	17	8
11	60	7	12
12	69	13	21
13	70	14	22
14	74	11	25
15	50	8	6
16	62	11	17

复回归分析范例

例4：【解】

1) 本例中之自变数与依变数各为何?

- 自变数为「广告费用」与「促销费用」；依变数为「销售量」。

2) 解读电脑报表。

Session

Regression Analysis: 銷售量 versus 廣告費用, 促銷費用

```
The regression equation is
銷售量 = 37.9 + 1.05 廣告費用 + 0.795 促銷費用


Predictor    Coef  SE Coef      T      P
Constant   37.897    2.444  15.51  0.000
廣告費用   1.0520   0.2028   5.19  0.000
促銷費用   0.7947   0.1132   7.02  0.000


S = 3.25881   R-Sq = 92.4%   R-Sq(adj) = 91.2%


Analysis of Variance

Source          DF       SS      MS      F      P
Regression       2  1672.94  836.47  78.76  0.000
Residual Error  13   138.06   10.62
Total           15  1811.00
```

复回归分析范例

a) 适配的复回归模式为

$$\boxed{\text{销售量} = 37.9 + 1.05 \times (\text{广告费用}) + 0.795 \times (\text{促销费用})}$$

- 常数项的系数为37.9，表示当广告费用与促销费用均为零时，平均销售量约为37.9万元。
- 广告费用的系数为1.05，表示当促销费用固定时，广告费用每增加一万元，平均销售量约增加1.05万元。
- 促销费用的系数为0.795，表示当广告费用固定时，促销费用每增加一万元，平均销售量约增加0.795万元。

b) 由结果报表可知，$r^2 = 92.4\%$，$r^2\,(\text{adj}) = 91.2\%$，表示广告费用与促销费用之变异共可解释约91%的总销售量之变异。

复回归分析范例

c) 由 ANOVA 结果报表可知，整个复回归模式之 F 检定统计量为78.76，其P值＝0.000，表示此回归模式非常显著，即两个自变数中至少有一个变数对依变数有显着之影响。

d) 广告费用与促销费用两个自变数 t 检定之 P 值分别为0.000及0.000，表示广告费用与促销费用对销售量的变异均具有显着的解释能力。

复回归分析范例

3) 当广告费用与促销费用分别为10万元与20万元时，请预测平均销售量约可达多少万元？

$$销售量 = 37.9 + 1.05 \times 10 + 0.795 \times 20 = 64.3$$

预估之平均销售量约可达 64.3 万元

复回归分析范例

- 例5：某饼干师傅为了解烘焙时间及烤箱温度对苏打饼干良率之影响，他搜集12批不同条件下之烘焙时间、烤箱温度与苏打饼干良率的资料，如下表所示，试进行复回归分析。

次数	苏打饼干良率（%）	烘焙时间（10秒）	烤箱温度（℃）
1	87	92	257
2	93	87	252
3	80	93	262
4	90	89	254
5	98	85	250
6	83	92	260
7	93	88	253
8	95	86	251
9	88	91	256
10	99	84	249
11	84	92	259
12	90	90	255

复回归分析范例

例5：【解】

1) 本例中之自变数与依变数各为何?

- 自变数为「烘焙时间」与「烤箱温度」；依变数为「苏打饼干良率」 。

2) 解读电脑报表。

Session

Regression Analysis: 蘇打餅乾良率 versus 烘焙時間, 烤箱溫度

```
The regression equation is
蘇打餅乾良率 = 427 - 0.299 烘焙時間 - 1.22 烤箱溫度


Predictor      Coef  SE Coef      T      P
Constant     427.06    26.91  15.87  0.000
烘焙時間    -0.2992   0.2524  -1.19  0.266
烤箱溫度    -1.2181   0.1879  -6.48  0.000


S = 0.686829   R-Sq = 98.9%   R-Sq(adj) = 98.7%


Analysis of Variance

Source          DF      SS      MS       F      P
Regression       2  381.75  190.88  404.63  0.000
Residual Error   9    4.25    0.47
Total           11  386.00
```

复回归分析范例

a) 适配的复回归模式为

$$\boxed{\text{苏打饼干良率} = 427 - 0.299 \times (\text{烘焙时间}) - 1.22 \times (\text{烤箱温度})}$$

- 常数项的系数为427。
- 烘焙时间的系数为-0.299，表示当烤箱温度固定时，烘焙时间每增加10秒，苏打饼干的平均良率约降低 0.299 %。
- 烤箱温度的系数为-1.22，表示当烘焙时间固定时，烤箱温度每增加1℃，苏打饼干的平均良率约降低 1.22 %。

b) 由结果报表可知，R-Sq＝98.9%，R-Sq(adj)＝98.7%，表示烘焙时间与烤箱温度之变异共可解释约99%的苏打饼干良率之变异。

复回归分析范例

c) 由ANOVA结果报表可知，整个复回归模式之F检定统计量为404.63，其P值＝0.000，表示此回归模式非常显著，即两个自变数中至少有一个变数对依变数有显着之影响。

d) 烘焙时间与烤箱温度两个自变数t检定之P值分别为0.266及0.000，其中烘焙时间之P值 > 0.05，表示烘焙时间在回归模式中不显著，应将烘焙时间删除后，再进行简单回归分析。

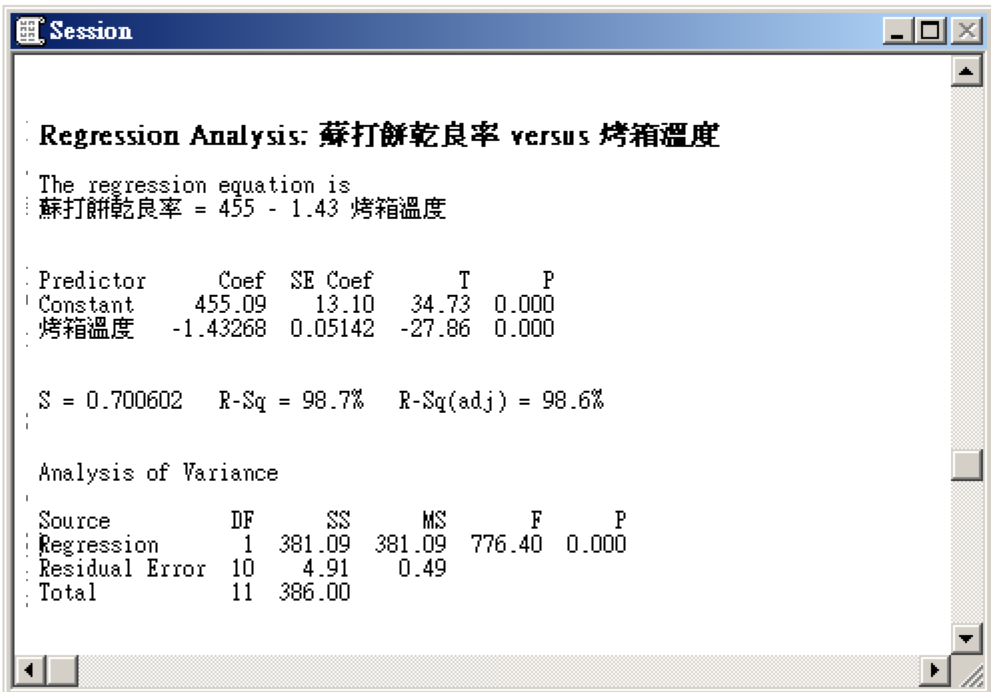
复回归分析范例

例5：【解】

3)　重新进行简单回归分析

- 自变数为「烤箱温度」；依变数为「苏打饼干良率」。

4)　解读电脑报表。


Session

Regression Analysis: 蘇打餅乾良率 versus 烤箱溫度

```
The regression equation is
蘇打餅乾良率 = 455 - 1.43 烤箱溫度


Predictor        Coef   SE Coef       T      P
Constant       455.09     13.10   34.73  0.000
烤箱溫度     -1.43268   0.05142  -27.86  0.000


S = 0.700602   R-Sq = 98.7%   R-Sq(adj) = 98.6%


Analysis of Variance

Source          DF      SS      MS       F      P
Regression       1  381.09  381.09  776.40  0.000
Residual Error  10    4.91    0.49
Total           11  386.00
```

复回归分析范例

a) 适配的简单回归模式为

$$\boxed{\text{苏打饼干良率} = 455.09 - 1.43 \times (\text{烤箱温度})}$$

b) 由结果报表可知，R-Sq＝98.7%，表示苏打饼干良率变异的98.7%可被烤箱温度之变异解释，显示此模式的解释能力非常良好。

c) 由 ANOVA 结果报表可知， F 检定统计量为776.40，其P值＝0.000，表示此回归模式非常显着。

d) 烤箱温度 t 检定之 P 值为0.000，表示烤箱温度可有效地用来预测苏打饼干的良率，且苏打饼干的良率与烤箱温度间存有非常强的负向线性关系。

本单元结束

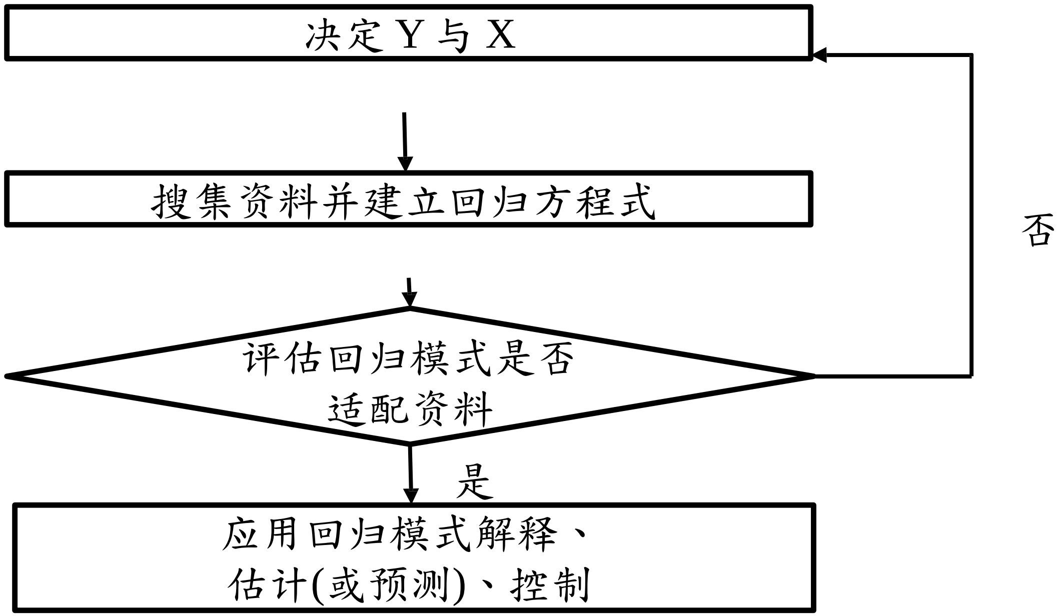
第八单元 简单回顾

简单回顾

- 回归分析：
  - 回归分析之意义
    - 自变数
    - 依变数
    - 回归模式
  - 回归分析之用途
  - 回归模式之类型
    - 简单回归模式
    - 复回归模式或多元回归模式
  - 回归分析之资料
    - 观察资料
    - 实验资料

简单回顾

- 回归分析介绍：
  - 相关分析简介
    - 正相关
    - 负相关
    - 不相关
  - 回归分析流程



简单回顾

- 简单线性回归分析：
  - 残差
  - 最小平方法
  - 回归式好坏之判断
    - 由图形判定（限简单回归）
    - 由判定系数判定
    - 由统计检定－F检定及t检定判定
- 复回归分析：
  - 如何估计复回归模式之系数
    - 解联立方程式
    - 利用统计软件