

# 数学建模

彭胜蓝

# 数学建模课程简介

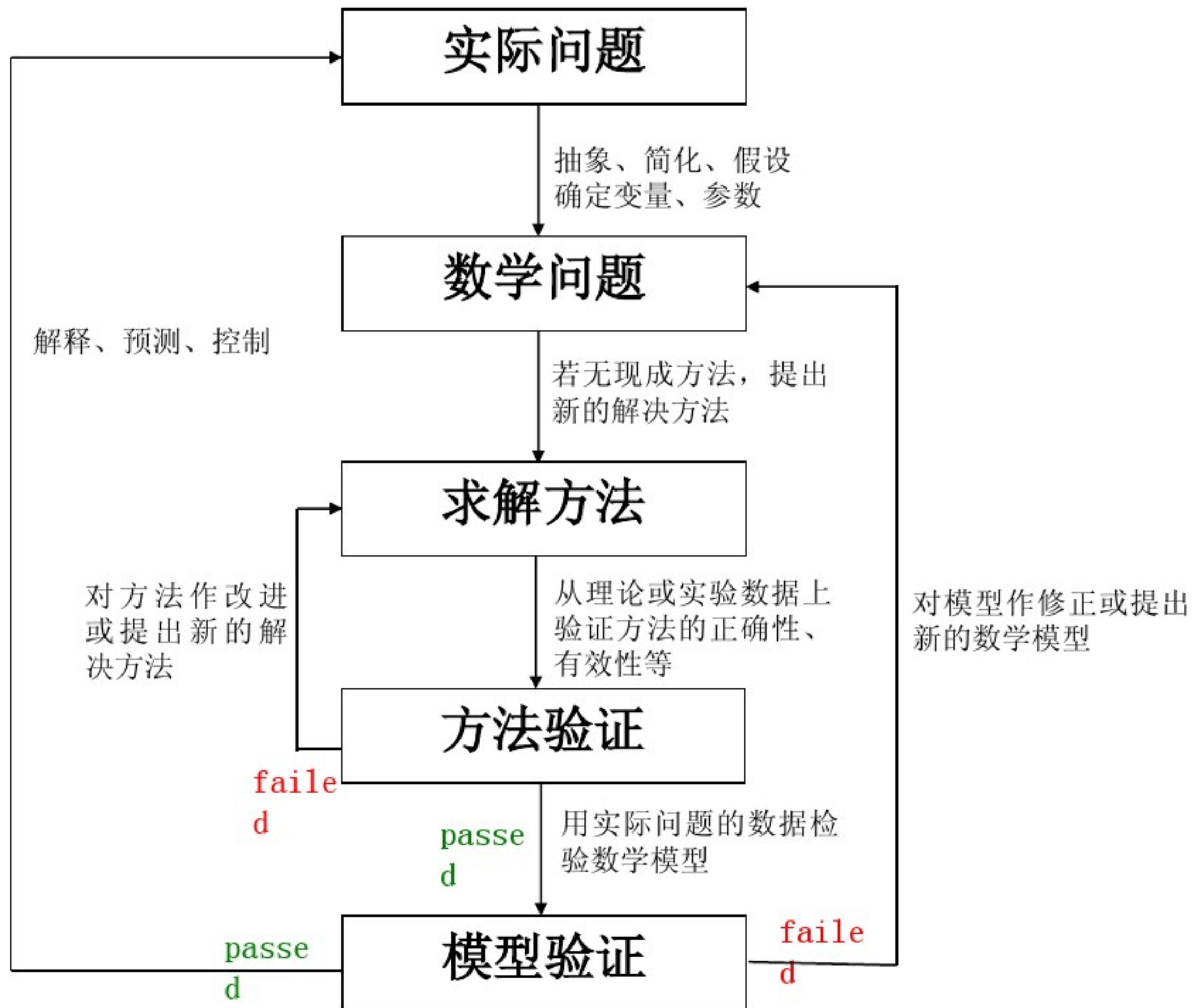
- 关于数学建模
- 数学建模的方法和工具
- 数学建模实例（线性回归）
- 论文的撰写方法

# 什么是数学建模

- 数学建模是对于客观世界的某一特定系统或特定问题，为了某个特定的目的作出必要的简化与假设，应用适当的数学工具得到的一个数学结构，它或者可以解释特定的现实状态，或能预测对象的未来状况，或者能提供处理对象的最优决策或控制。
- 数学建模就是用数学知识和方法建立数学模型解决实际问题的一个过程。

# 数学建模的方法和步骤

- 利用数学模型解决实际问题，由于实际问题的多样性，解决方法的千差万别，很难有一个统一的、具体的步骤描述整个建模过程，但从广义的方法论而言，建模过程一般步骤可描述如下
  1. 模型准备(提出实际问题);
  2. 模型假设(简化实际问题);
  3. 模型建立(用数学语言来描述问题);
  4. 模型求解(选择适当的方法求解数学问题);
  5. 模型分析与检验.





# 数学建模的方法和工具

- 数学建模所涉及的数学知识也是非常广泛的，如微分方程、线性代数、概率统计、图与网络、回归分析、层次分析、量纲分析、机理分析、规划论、排队论、对策论、决策论、插值方法、差分方法、样条方法、变分方法，人工智能，神经网络等优化方法，
- 计算机软件的操作及编程

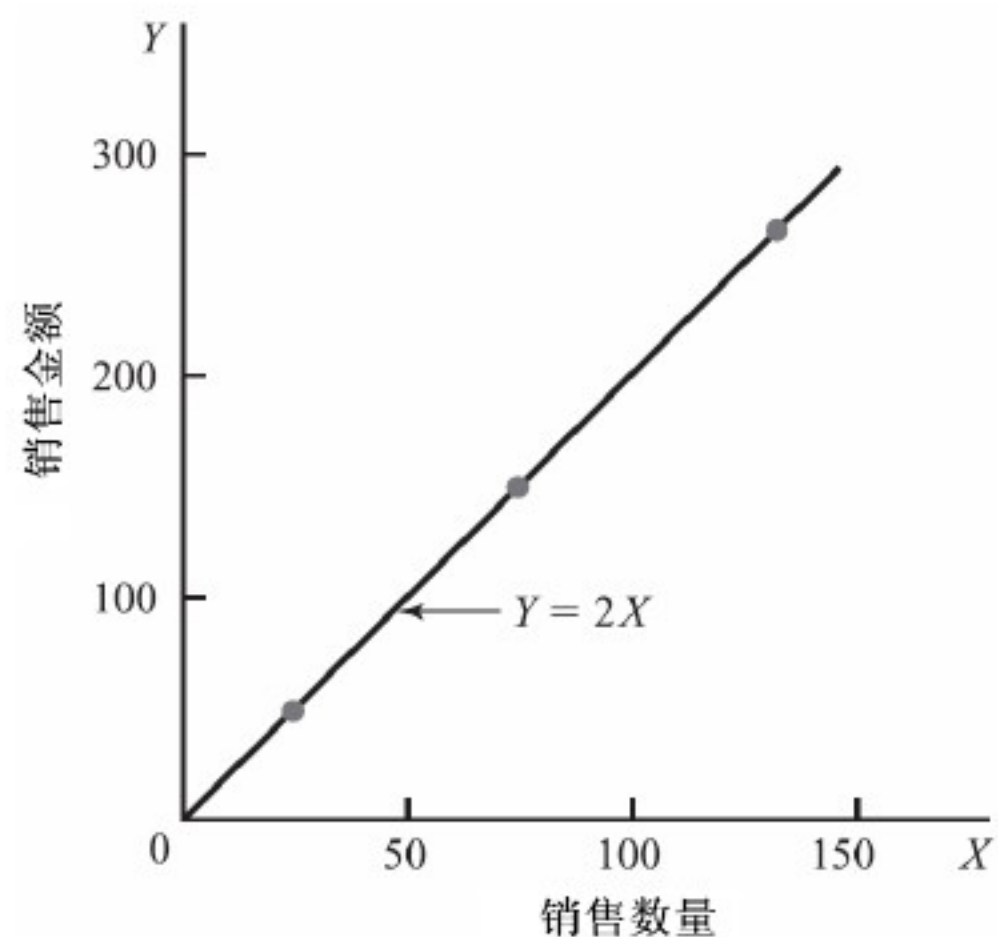
# 案例 电视广告投入影响力

FIRM	SPEND	MILIMP	FIRM	SPEND	MILIMP
MILLER LITE	50.1	32.1	BUD LITE	45.6	10.4
PEPSI	74.1	99.6	ATT/BELL	154.9	88.9
STROH'S	19.3	11.7	CALVIN KLEIN	5	12
FED'L EXPRESS	22.9	21.9	WENDY'S	49.7	29.2
BURGER KING	82.4	60.8	POLAROID	26.9	38
COCO-COLA	40.1	78.6	SHASTA	5.7	10
MC DONALD'S	185.9	92.4	MEOW MIX	7.6	12.3
MCI	26.9	50.7	OSCAR MEYER	9.2	23.4
DIET COLA	20.4	21.4	CREST	32.4	71.1
FORD	166.2	40.1	KIBBLES 'N BITS	6.1	4.4
LEVI'S	27	40.8			

# 函数关系

- 两变量间之函数关系可以通过数学方程式来表达，如果利用  $X$  表示自变量， $Y$  表示因变量，则两变量间之函数关系形式为

$$Y = f(X)$$



季度	销售数量	销售金额
1	75	\$150
2	25	50
3	130	260



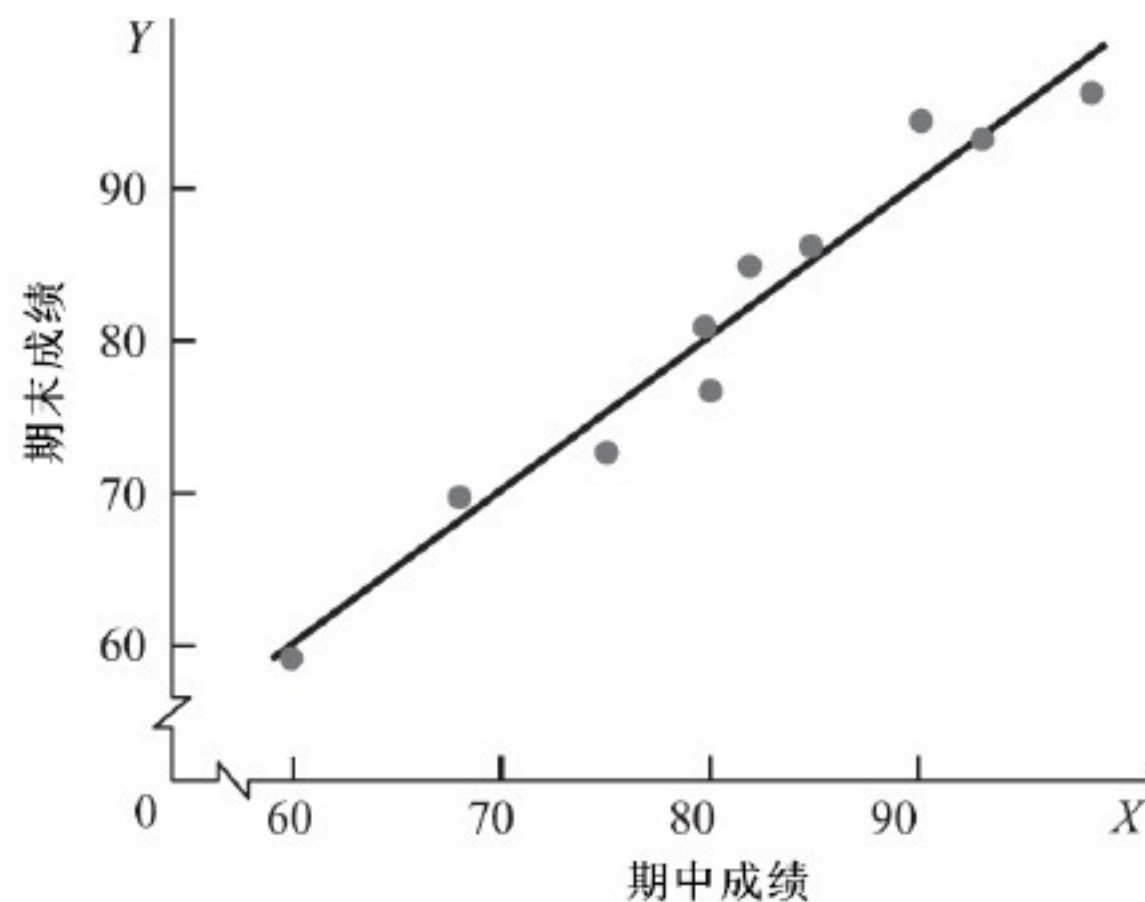
# 统计关系

- 一般而言，具备统计关系的观测值的散点并不会完美地出现在关系线上，两个变量之间的统计关系表示为

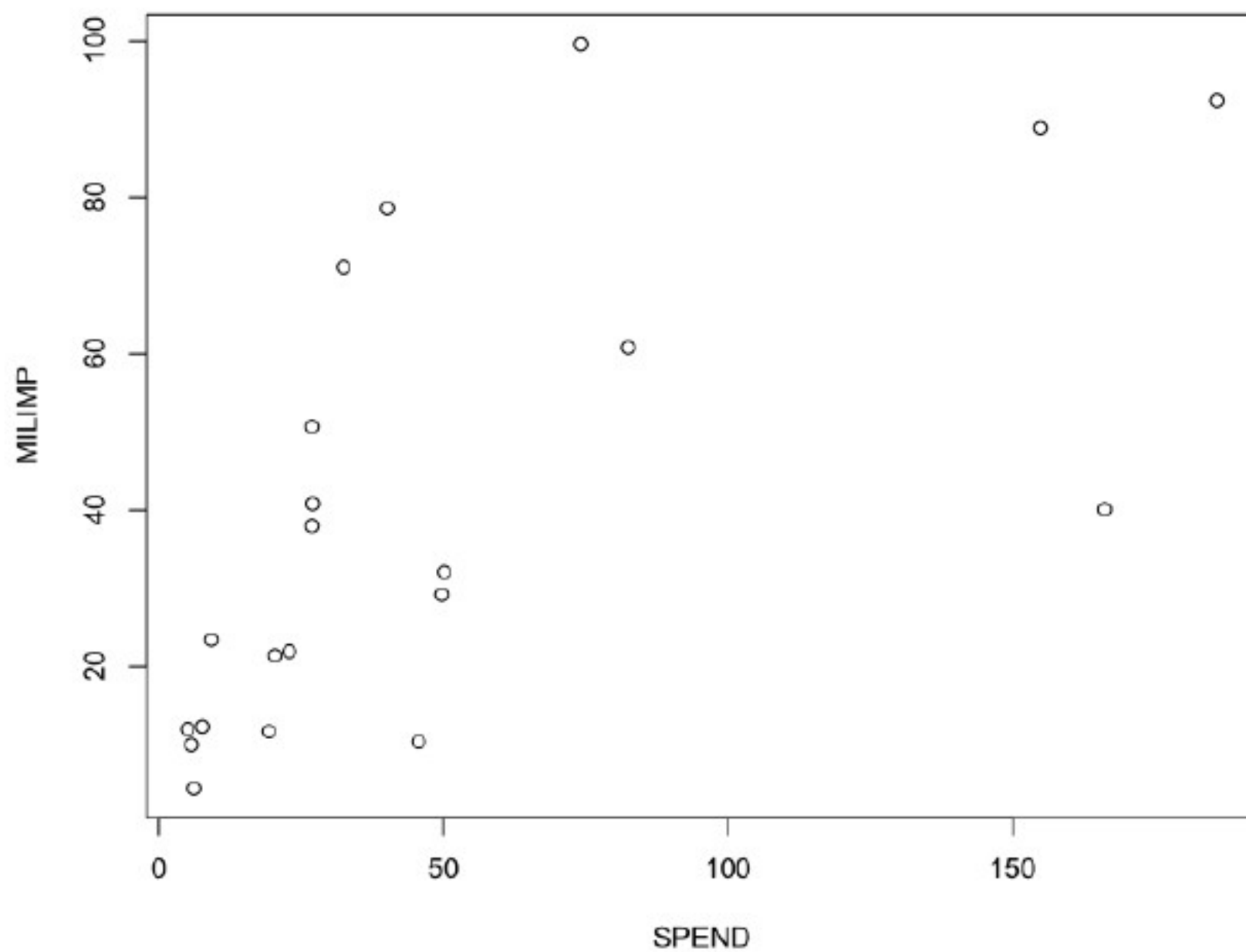
$$Y = f(X) + \varepsilon$$

$\varepsilon$ 称为随机误差

散点图与统计关系直线



# 电视广告投入金额 v.s. 受影响的人数



# 模型假设

- 的变量 $X$ 表示电视广告的投入（SPEND），变量 $Y$ 表示受广告影响的人数（MILIMP）；
- 设 $Y$ 与 $X$ 存在简单线性统计关系

$$Y = a + bX + \varepsilon \quad (a > 0)$$

即变量 $Y$ 直接受 $X$ 的线性函数的影响之外，还受一个不确定的随机因素 $\varepsilon$ 的影响。

# 模型建立：简单线性回归模型

■  $Y_i = a + bX_i + \varepsilon_i \quad i=1,2,\dots,21$

上式中，

$Y_i$  表示因变量  $Y$  第  $i$  次试验（第  $i$  个公司）的观测值

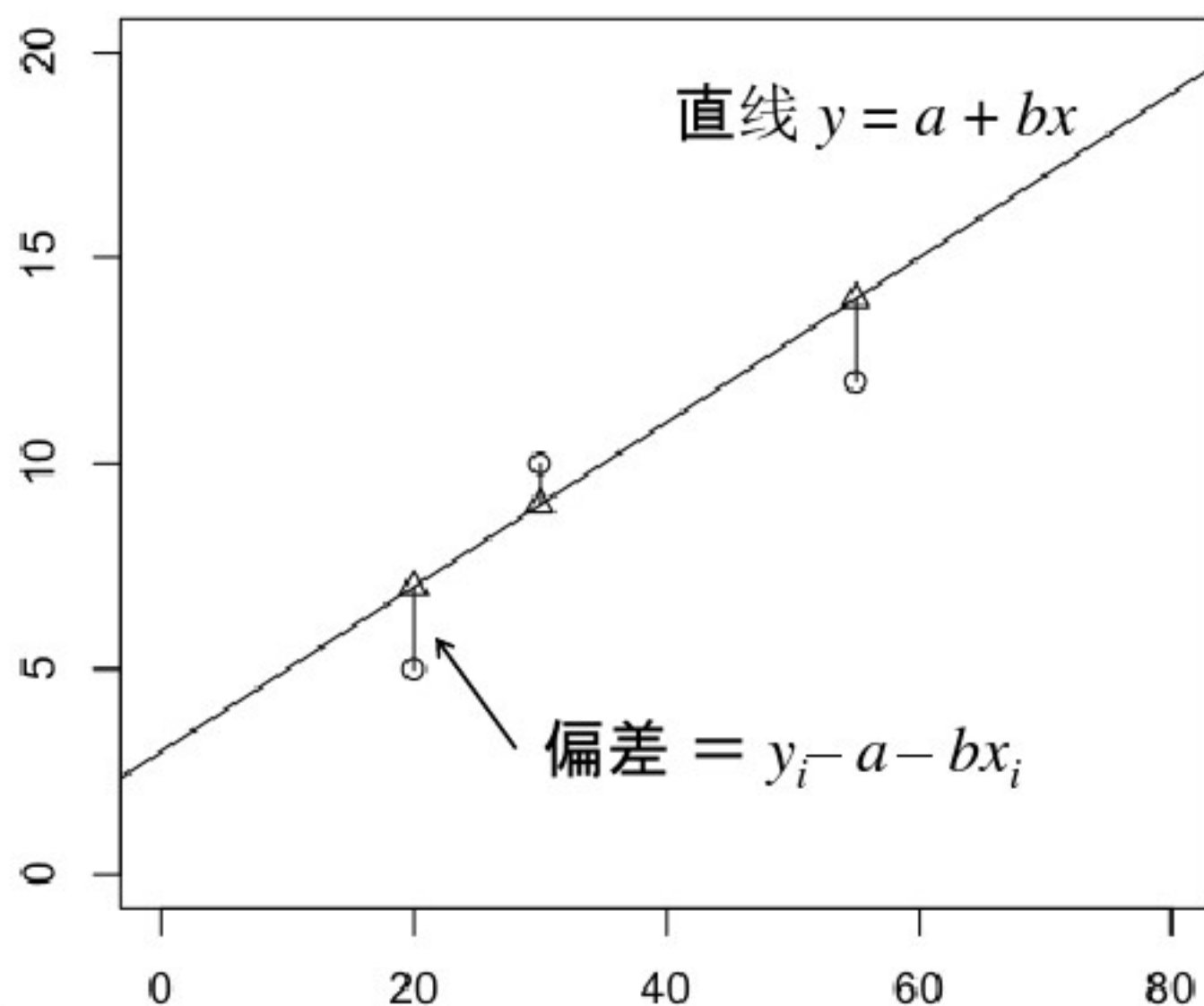
$a$  与  $b$  均为参数，称为回归系数

$X_i$  为常数，表示因变量  $X$  第  $i$  次试验下的观测值

$\varepsilon_i$  为  $i$  次试验的随机误差项，满足均值为0，不相关，方差相等

# 数学问题：模型的参数 $a$ ， $b$ 的估计

▣ 寻找一条直线，使之与实际观测数据“拟和最好”



拟和好：“总偏差”最小



# 模型求解：最小二乘法求极值问题

■ 基于（关于偏差的）最小平方准则

$$Q = \sum_{i=1}^n (Y_i - a - bX_i)^2$$

寻找“最优”的拟和直线：即寻找一对数 $(a, b)$ ，其中 $a, b$ 分别表示直线的截距与斜率，使对应的偏差平方和 $Q$ 最小，即

$$(\hat{a}, \hat{b}) = \arg \min_{a, b} Q = \sum_{i=1}^n (Y_i - a - bX_i)^2$$

# 模型求解：最小二乘法求极值问题

■ 求  $Q = \sum_{i=1}^n (Y_i - a - bX_i)^2$  对应的驻点

$$\begin{cases} \frac{\partial Q}{\partial a} = 2 \sum_{i=1}^n (Y_i - a - bX_i) = 0 \\ \frac{\partial Q}{\partial b} = 2 \sum_{i=1}^n X_i (Y_i - a - bX_i) = 0 \end{cases}$$

# 模型求解：回归系数 $a, b$ 的估计值

■ 解方程组的最小二乘法估计式

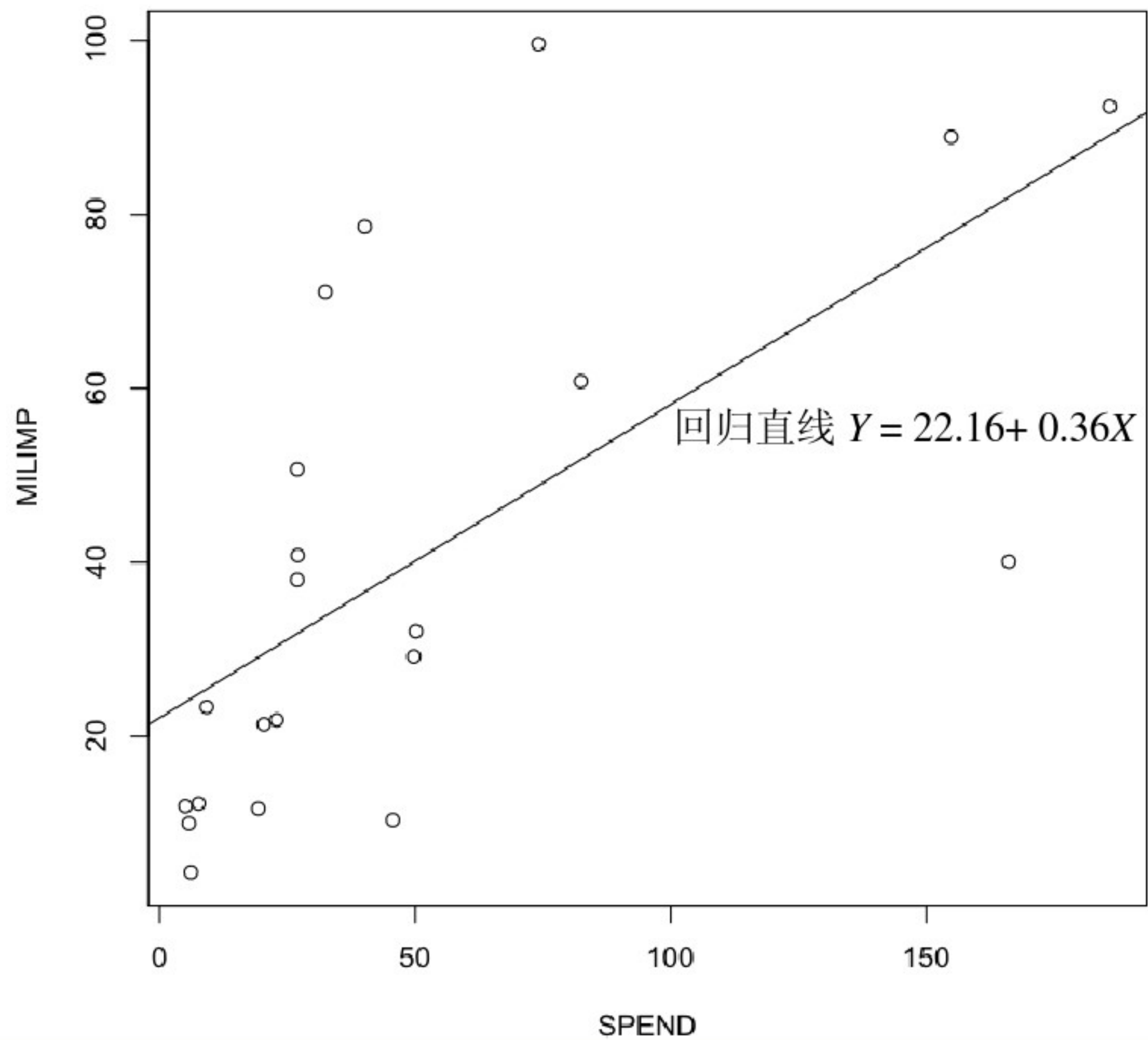
$$\hat{b} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X}$$

此处

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n), \bar{Y} = \frac{1}{n}(Y_1 + \dots + Y_n)$$

FIRM	$X$	$Y$	$X'=X-\bar{X}$	$Y'=Y-\bar{Y}$	$X'*Y'$	$X'*X'$
MILLER LITE	50.10	32.10	-0.30	-8.40	2.52	0.09
PEPSI	74.10	99.60	23.70	59.10	1400.67	561.69
STROH'S	19.30	11.70	-31.10	-28.80	895.68	967.21
FED'L EXPRESS	22.90	21.90	-27.50	-18.60	511.50	756.25
BURGER KING	82.40	60.80	32.00	20.30	649.60	1024.00
COCO-COLA	40.10	78.60	-10.30	38.10	-392.43	106.09
MC DONALD'S	185.90	92.40	135.50	51.90	7032.45	18360.25
MCI	26.90	50.70	-23.50	10.20	-239.70	552.25
DIET COLA	20.40	21.40	-30.00	-19.10	573.00	900.00
FORD	166.20	40.10	115.80	-0.40	-46.32	13409.64
LEVI'S	27.00	40.80	-23.40	0.30	-7.02	547.56
BUD LITE	45.60	10.40	-4.80	-30.10	144.48	23.04
ATT/BELL	154.90	88.90	104.50	48.40	5057.80	10920.25
CALVIN KLEIN	5.00	12.00	-45.40	-28.50	1293.90	2061.16
WENDY'S	49.70	29.20	-0.70	-11.30	7.91	0.49
POLAROID	26.90	38.00	-23.50	-2.50	58.75	552.25
SHASTA	5.70	10.00	-44.70	-30.50	1363.35	1998.09
MEOW MIX	7.60	12.30	-42.80	-28.20	1206.96	1831.84
OSCAR MEYER	9.20	23.40	-41.20	-17.10	704.52	1697.44
CREST	32.40	71.10	-18.00	30.60	-550.80	324.00
KIBBLES 'N BITS	6.10	4.40	-44.30	-36.10	1599.23	1962.49
	$\bar{X}=50.40$	$\bar{Y}=40.47$			$\hat{b}=0.36$	$\hat{a}=22.16$





# 模型检验

- 原假设  $H_0: b = 0$  （不存在线性统计关系）

备选假设  $H_a: b \neq 0$  （存在线性统计关系）

- 计算检验统计量  $t^*$  及对应的  $p$ -值

$$t^* = \frac{\hat{b}}{\sqrt{Q(\hat{a}, \hat{b}) / (n - 2)}} = 3.74 \quad p\text{-value} = \Pr\{|t| > t^*\} = 0.0014$$

- 检验法则：给定一个风险水平  $\alpha$  （一般取0.05或0.01），如果  $p$ -值大于或等于  $\alpha$ ，接受原假设  $H_0$ ；否则，如果  $p$ -值小于  $\alpha$ ，拒绝原假设  $H_0$ ，接受备选假设  $H_a$ 。本例中  $p$ -值非常小，所有拒绝原假设，即广告投入和受影响人数存在显著的线性统计关系。

# 数据分析工具：Excel

- ▣ 下载数据：<ftp://172.22.23.251/math/Ad.csv>，并用Excel打开；
- ▣ 画散点图：点击菜单“插入”→“图表”→“XY散点图”→“下一步”→数据区域填入“B1:C21”，并选择列→“下一步”→X轴填入“SPEND”，Y轴填入“MILIMP”→“下一步”→作为新工作表插入→完成
- ▣ 在散点图中添加趋势线：右击图中任意一个数据点→“添加趋势线”

# 利用数据分析工具作回归分析

▣ 加载数据分析工具：点击菜单“工具” —> “加载宏” —> 选择“分析工具库”，点击“确定”。

▣ 如果没有安装数据分析包，需要用MS Office安装光盘安装

<ftp://172.22.23.251/Windows/>

▣ 加载成功后，在“工具”子菜单栏中会出现“数据分析”的子菜单，打开它 —> 选择“回归” —> 在回归对话框中，“Y值输入区域”输入“C1:C21”，“X值输入区域”输入“B1:B21”，点击“确定”，观察输出结果。

# 学术论文的撰写

- ▣ 研究性的学术论文按结构通常分为三个部分：提出问题、分析并解决问题及结论部分。
  - ✓ 提出问题：主要说明论文中针对的问题的实际背景和意义，其中主要包括解决什么问题？为什么要解决该问题；
  - ✓ 分析并解决问题：对所提出的问题进行适当的简化，并作合理的假设，以构造一个适合问题需要的模型，并在此基础上提出解决问题的方法，同时对模型和方法作严格的理论分析；
  - ✓ 结论：分析、讨论关于真实数据的试验结果，说明论文提出的模型及方法对解决实际问题的效果。



# 论文的基本格式

## 1. 论文开头

- 题名 (Title)
- 作者, 单位 (Authors, Affiliations)
- 摘要 (Abstract)
- 关键词 (Keywords)

## 2. 正文

- 引论 (Introduction)
- 方法 (Methods)
- 结果与讨论 (Results and Discussions)

## 3. 论文结尾

- 结论 (Conclusions)
- 参考文献 (References)



# 论文范本