# 第一章 经济计量学的特征和研究范围

## 经济计量学：

是利用经济理论、数学、统计推断等工具对经济现象进行分析的一门社会科学。经济计量学运用数理统计学分析经济数据，对构建于数理经济学基础之上的模型进行实证分析，并得出数值结果。

## 数理经济学：

用数学形式或方程（或模型）描述经济理论。

## 经济计量分析步骤：

* + 建立一个理论假说
  + 收集数据
  + 设定数学模型
  + 设立统计或经济计量模型
  + 估计经济计量模型参数
  + 核查模型的适用性：模型设定检验
  + 检验源自模型的假设
  + 利用模型进行预测

## 数据的类型：

* + 时间序列数据：按照时间跨度收集得到的。
  + 截面数据：一个或多个变量在某一时点上的数据集合。
  + 合并数据：既包括时间序列数据又包括截面数据。
  + 面板数据：一种特殊类型的合并数据，是同一个横截面单位（比如某个家庭或公司）的跨期调查数据。

## 回归关系与因果关系：

统计关系无论有多强，有多紧密，也绝不能建立起因果关系； 因果关系的概念来自统计学之外的某个理论。

## 两个劳动经济学假说：

* + 受挫-工人假说：当经济形势恶化时（表现为较高的失业率），许多失业工人放弃寻找工作的愿望并退出劳动力市场。
  + 增加-工人假说：当经济形势恶化时，许多尚未进入劳动力市场的后背工人（比如带孩子的母亲）可能会由于养家的人失去工作而决定进入劳动力市场。

# 第二章 线性回归的基本思想：双变量模型

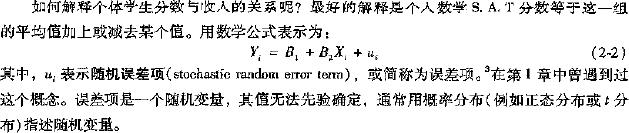
## 回归分析：

用于研究一个变量（称为被解释变量或应变量）与另一个或多个变量（称为解释变量或自变量）之间的关系。

## 总体回归函数（PRF）：

Y 的条件期望对 X 的回归函数（Y 的条件期望即给定 X 值下 Y 的期望）。

## 随机误差项：



在上述总体回归方程中，Yi 可以分解为两个部分之和：一是(B1+B2i)，这部分称为系统性成分或确定性成分；另一部分则是μi，称为非系统或非随机成分，随机项μi 也称为噪声。

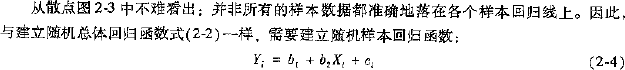
## 两种总体回归函数：

* + 随机（统计）总体回归函数：Yi
* 非随机（确定）总体回归函数：E(Y|Xi)

## 随机误差项的性质：

* + 误差项代表了未纳入模型变量的影响。
  + 即使模型中包括了决定因变量的所有变量，其内在随机性也不可避免， μ反映了人类行为的这种内在随机性。
  + μ还代表了度量误差。
  + “奥卡姆剃刀原则”，即描述应该尽可能简单，只要不遗漏重要的信息。

## 样本回归函数（SRF）：

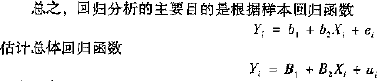
由于我们不能得到全部的总体数据，而只能得到一部分的样本数据（有时对于每个 Xi 只有一个 Yi 值），因此无法确定总体回归线，而只能用这些已知的样本数据去拟合，得到的拟合函数称为样本回归函数。对于不同的样本，所得到的样本回归函数也是不同的。

## 残差：

样本回归函数（2-4）中的 ei 是μi 的估计量，称为残差。从概念上讲， 它与μi 类似，可以作为μi 的估计量。简单地说，ei 表示了 Y 的实际值与根据样本回归得到的估计值的差。

## 回归分析的目的：

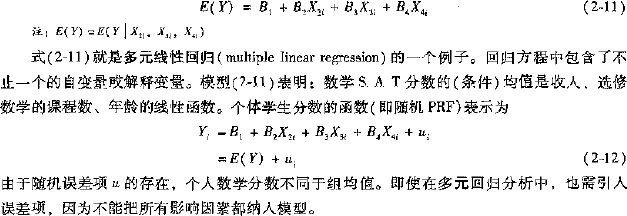
根据样本回归函数估计总体回归函数。



## 线性回归的含义：

* + 变量线性：应变量的条件均值是自变量的线性函数。（本质含义）
  + 参数线性：应变量的条件均值是参数 B 的线性函数，而变量之间并不一定是线性的。（所用含义）

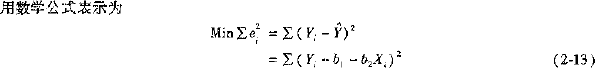
### 多元线性回归模型：



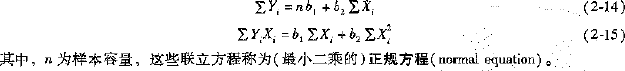
# 最小二乘原理：

选择合适的参数 b 和 b ，使得残差平方和Σei2 最小。

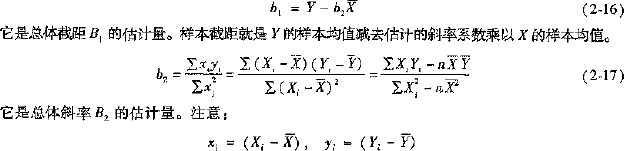
1 2



# 正规方程：

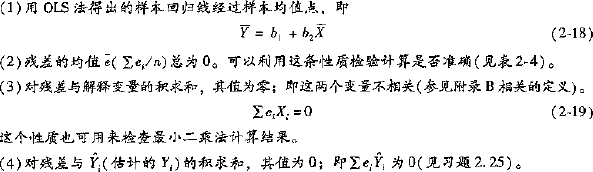


# OLS 估计量：



# OLS 估计量的重要性质：

* + 用 OLS 法得出的样本回归线经过样本均值点。
  + 残差的均值(Σei/n)总为 0。
  + 对残差与解释变量的积求和，其值为 0，即这两个变量不相关。
  + 对残差与 Yihat 的积求和，其值为 0。



# 第三章 双变量模型：假设检验：

* **古典线性回归模型（CLRM）的基本假定：** 
  + 回归模型是参数线性的，但不一定是变量线性的。



* + 解释变量(X)与扰动误差项μ不相关。但是，如果 X 是非随机的（即为固定值），则该假定自动满足。即使 X 值是随机的，如果样本容量足够大，也不会对分析产生严重影响。
  + 给定 Xi，扰动项的期望或均值为 0。



* + μi 的方差为常数，或同方差（homoscedastic）。



* + 无自相关假定，即两个误差项之间不相关。

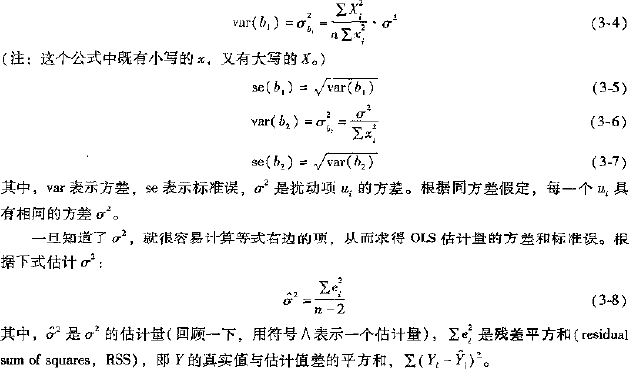


* + 回归模型是正确设定的。也就是说，实证分析的模型不存在设定偏差或设定误差。
  + （推导 b1b2 抽样分布增加假定）在总体回归函数中，误差项服从均值为

0，方差为σ2 的正态分布。

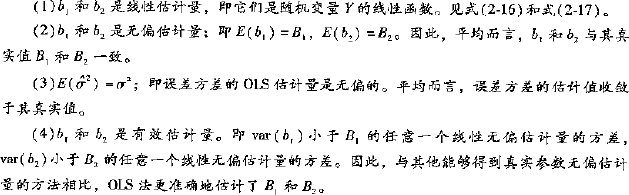


# 普通最小二乘估计的方差和标准误：



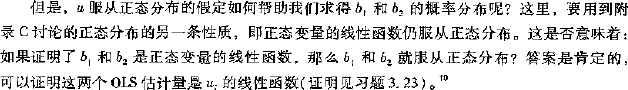
# OLS 估计量的性质：

* + b1 和 b2 是线性估计量，即它们是随机变量 Y 的线性函数。
  + b1 和 b2 是无偏估计量。
  + 误差方差的 OLS 估计量是无偏的，平均说，误差方差的估计值收敛于其真实值。
  + b1 和 b2 是有效估计量。

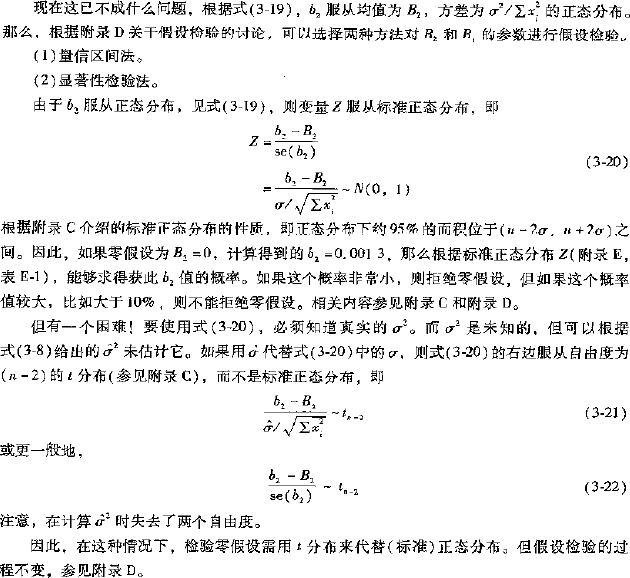


# 中心极限定理：

随着变量个数的无限增加，独立同分布随机变量近似服从正态分布。



# 假设检验：

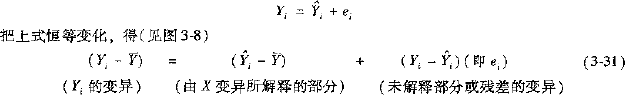


# 两种假设检验基本方法：

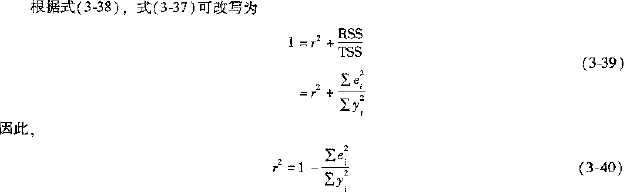
* + 置信区间法：计算参数的置信区间
  + 显著性检验法：计算参数估计的 t 值

# 判定系数 r2：

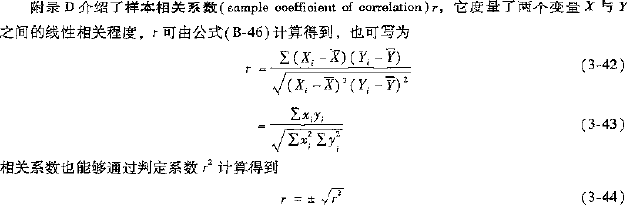




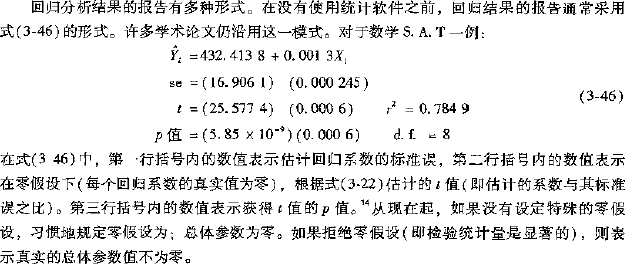
# r2 的计算公式：



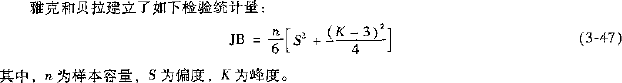
# 相关系数 r：



# 回归分析结果的报告：



* **正态性检验：用 ei 来获悉μi 的正态性** 
  + 残差直方图：
  + 正态概率图：
  + 雅克-贝拉检验（JB 检验）：



在正态性假设下，式（3-47）给出的 JB 统计量渐进服从自由度为 2

的卡方分布，用符号表示为：

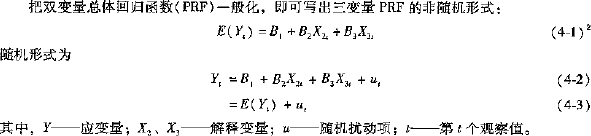


# 预测时的注意事项：

当 X0 = X 均值时，置信带的宽度最小；但是，随着 X0 逐渐远离 X 均值， 置信带将迅速变宽（即预测误差将会增加），这表明随着 X0 逐渐远离 X 均值， 历史回归的预测能力将显著减弱。因此，在用历史回归线“外推”预测 Y 的均值时，需要格外谨慎。

# 第四章 多元回归：估计与假设检验：

* **多元回归模型：**

包含多个解释变量的回归模型。多元是指有多种因素（即变量）对应变量有影响。

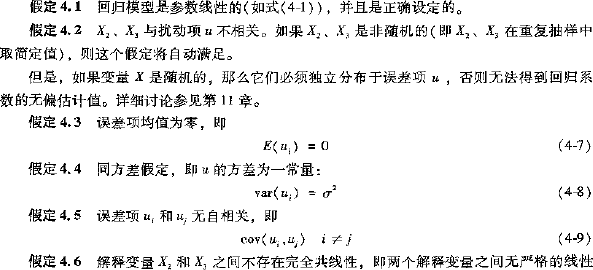
# 偏回归系数：

在多元回归模型中，B2、B3 等参数称为偏回归系数或偏斜率系数。

# 多元回归模型的基本假定：

* + 回归模型是参数线性的，并且是正确设定的。
  + X2、X3 与扰动项μ不相关。如果 X2、X3 是非随机的（即 X2、X3 在重复抽样中取简定值），则这个假定将自动满足。但如果变量 X 是随机的，那么它们必须独立分布于误差μ，否则无法得到回归系数的无偏估计值。
  + 误差项均值为 0。
  + 同方差假定。
  + 误差项无自相关。
  + 解释变量 X2 与 X3 之间不存在完全共线性，即两个解释变量之间无严格的线性关系。
  + （假设检验补充条件）为了进行假设检验，假定随机误差μ服从均值为

0，方差为σ2 的正态分布。



# 多元回归参数的估计：

