2-1
相关分析与回归分析

主讲人：范国斌

例如：为了促进经济发展，经济学家可能有各种主张

（1）认为需要增加工资，因为这将提高消费者的需求，而刺激生产；

（2）认为需要削减工资，因为降低成本将增加企业的利润，并因而刺激生产；

（3）需要提高利息率，因为可增加银行存款，而增加银行贷款的能力。

（4）需要削减利息率，因为将刺激投资，开设更多新企业

“增工资”与“减工资”、“削减利息率”与“提高利息率”，究竟应如何选择？其数量关系和数量界线究竟是什么?

这说明经济概念的定量化非常必要。正是在这类“不能解决的问题的吸引力”的影响下产生了计量经济学。

相关分析与回归分析

（对统计学的回顾）

经济变量之间的相互关系

性质上可能有三种情况:

- 确定性的函数关系　　$Y=f(X)$　　可用数学方法计算
- 不确定的统计关系—相关关系

  $Y=f(X)+\varepsilon$　　(ε为随机变量)　　可用统计方法分析
- 没有关系　　不用分析

相关程度的度量—相关系数

如果 X 和 Y 总体的全部数据 都已知，X 和 Y 的方差和协方差也已知，则

X和Y的总体线性相关系数: $\rho = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

其中：$Var(X)$ -----X 的方差　　$Var(Y)$ -----Y的方差

$Cov(X,Y)$ -----X 和 Y 的协方差

特点：

- 总体相关系数只反映总体两个变量 X 和 Y 的线性相关程度
- 对于特定的总体来说， X 和 Y 的数值是既定的，总体相关系数 ρ 是客观存在的特定数值。
- 总体的两个变量 X 和 Y 的全部数值通常不可能直接观测，所以总体相关系数一般是未知的。

X和Y的样本线性相关系数：

如果只知道 X 和 Y 的样本观测值，则 X 和 Y 的样本线性相关系数为：

$$r_{XY}=\frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum(X_i-\overline{X})^2\sum(Y_i-\overline{Y})^2}}$$

其中：X_i和 Y_i 分别是变量 X 和 Y 的样本观测值

$\overline{X}$ 和 $\overline{Y}$ 分别是变量 X 和 Y 样本值的平均值

注意：r_{XY} 是随抽样而变动的随机变量。

相关系数较为简单，也可以在一定程度上测定变量间的数量关系，但是对于具体研究变量间的数量规律性还有局限性。

对相关系数的正确理解和使用

- X和Y 都是相互**对称**的随机变量，$r_{XY} = r_{YX}$
- 样本相关系数是总体相关系数的样本估计值，由于抽样波动，样本相关系数是随抽样而变动的随机变量，其统计显著性还有待检验

只是相关分析还不能达到经济计量分析的目的

相关分析的局限:

相关系数只能说明两个变量线性相关的方向和程度,不能说明相关关系具体接近哪条直线,也就不能说明一个变量的变动会导致另一个变量变动的具体数量规律。

计量经济学关心的问题：

- 是经济变量间的因果关系以及隐藏在随机性后面的具体统计规律性。
- 在这方面回归分析方法可以发挥更为重要的作用。

回归分析

回归的古典意义：

高尔顿遗传学的回归概念

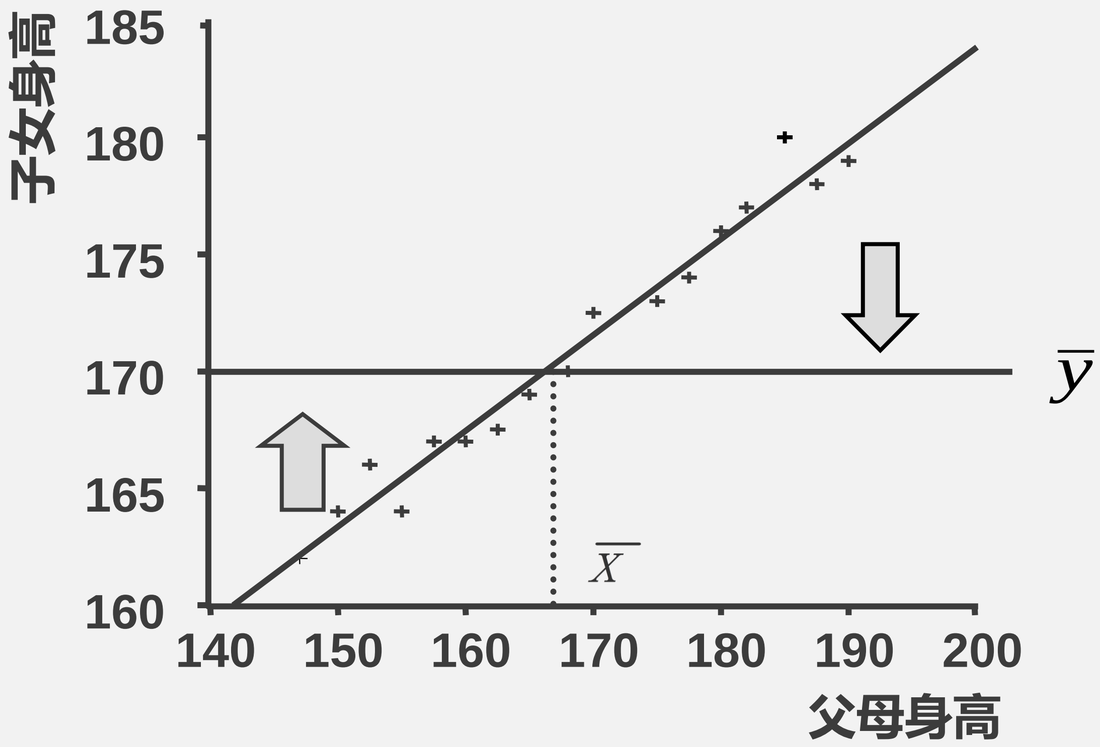
（父母身高与子女身高的关系）

子女的身高有向人的平均身高“回归”的趋势

“回归”(regression)一词最早由英国生物学家**Francis Galton**提出。

（1886年**F.Gallton**的论文《**Family Likeness in Stature**》）

Galton的普遍回归定律（**law of univesal regression**）被他的朋友**Karl Pearson**(1903)证实。



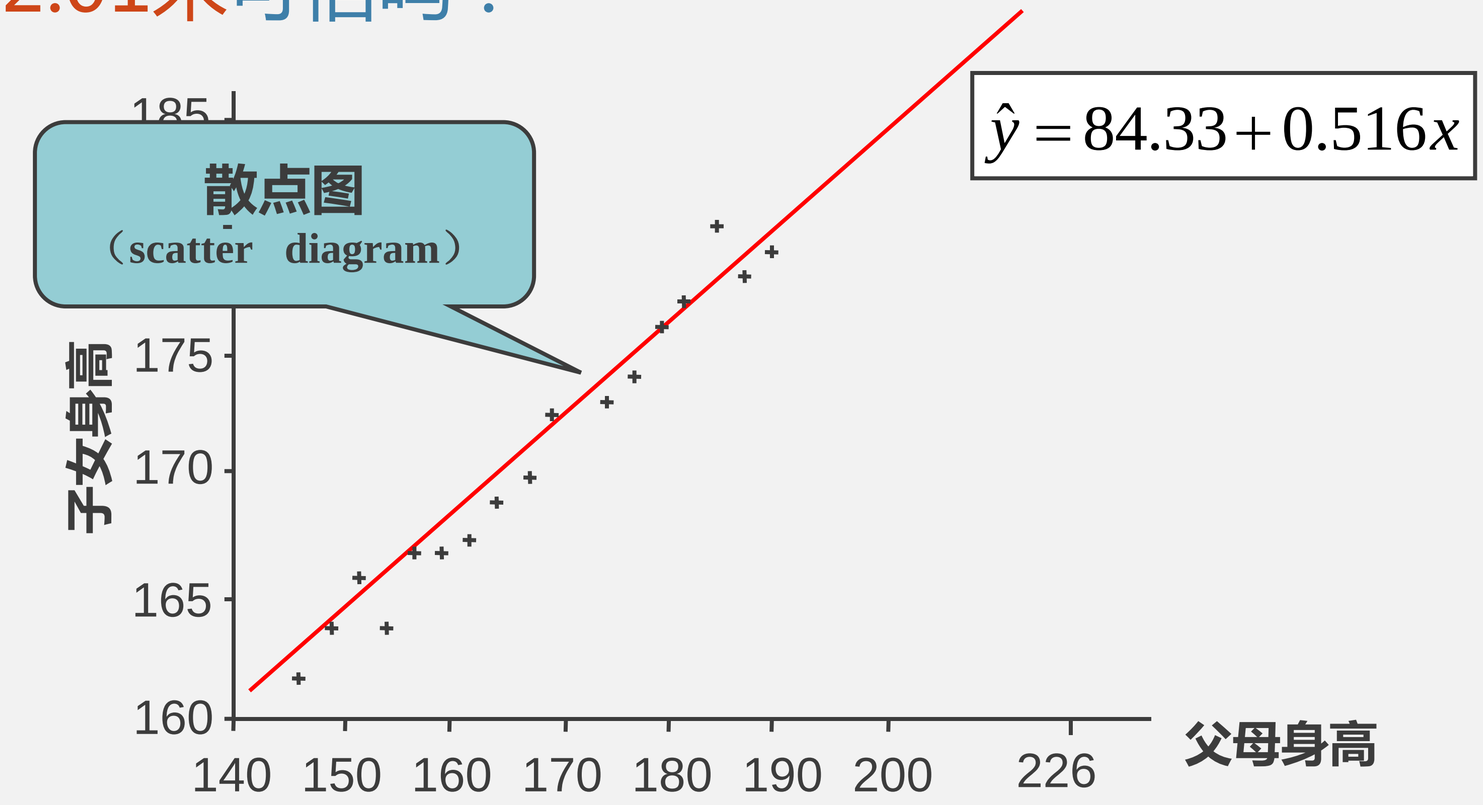
回归分析

回归的现代含义：

- 回归分析是关于研究一个叫做因变量的变量（Y）对另一个或多个叫做自变量的变量（X）的依赖关系;
- 其用意在于通过自变量在重复抽样中的已知或设定值，去估计或预测因变量的总体均值。
- 回归（Regression）是计量经济学的主要工具。

例子：姚明身高2.26米，姚明的子女会有多高呢？

2.01米可信吗？



因此，一旦知道了父母的身高，就可以按照上述关系式（回归线）来预测子女的平均身高（而不是具体身高）。

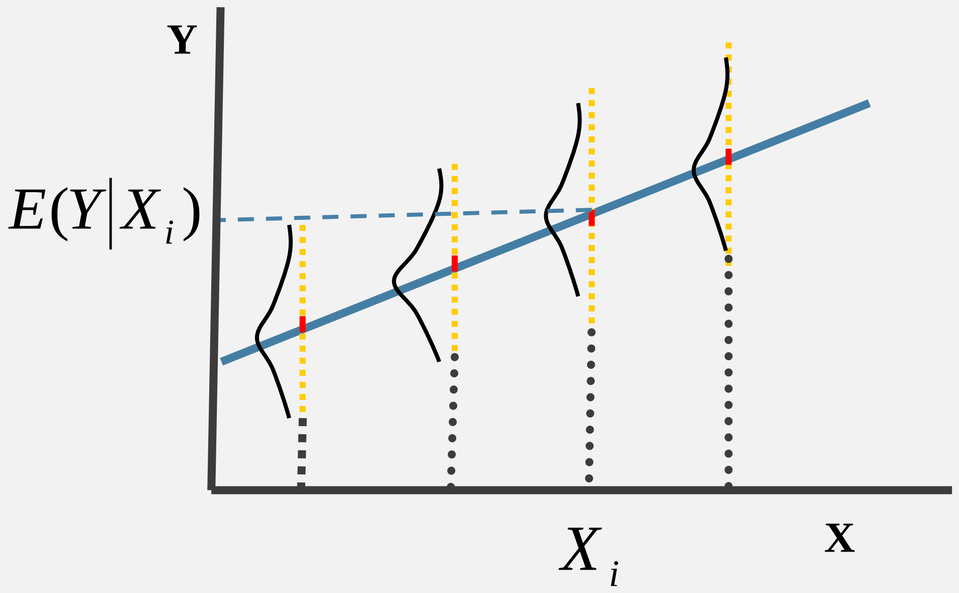
注意明确几个概念（为深刻理解“回归”）

●被解释变量 Y 的条件分布和条件概率：

当解释变量 X 取某固定值时（条件），Y 的值不确定，Y 的不同取值会形成一定的分布，这是 Y 的条件分布。 X 取某固定值时，Y 取不同值的概率称为条件概率。

●被解释变量Y的条件期望：

对于X 的每一个取值，对Y所形成的分布确定其期望或均值，称为Y的条件期望或条件均值，用 $E(Y|X_i)$ 表示。



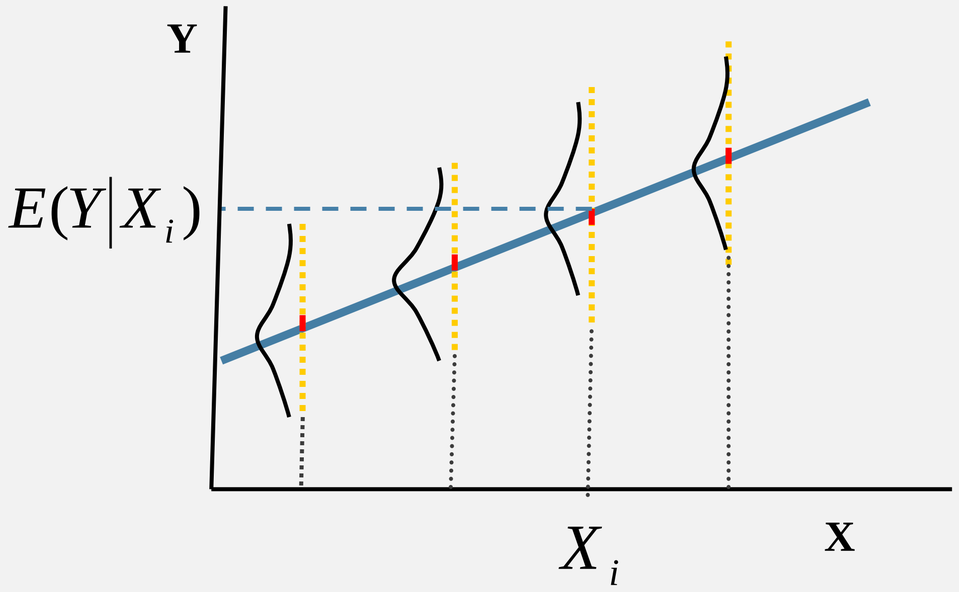
注意:Y的条件期望是随X的变动而变动的

- **回归线**：对于每一个 X 的取值 X_i ，都有 Y 的条件期望 $E(Y|X_i)$ 与之对应，代表 Y 的条件期望的点的轨迹形成的直线或曲线称为回归线。

- **回归函数**：被解释变量 Y 的条件期望 $E(Y|X_i)$ 随解释变量 X 的变化而有规律的变化，如果把Y的条件期望表现为 X 的某种函数，

$$\mathrm{E}(Y|X_i) = f(X_i)$$

这个函数称为回归函数。



回归函数分为：

总体回归函数和样本回归函数