



主讲老师：徐娜

中级经济师  
经济基础知识  
课程精讲班



# 第二十六章

## 回归分析



# 目录

## 第二十六章回归分析

### 第一节回归模型

1.回归分析的概念

2.一元线性回归模型

### 第二节最小二乘法

### 第三节模型的检验和预测

1.回归模型的拟合效果分析

2.模型预测

3.二元回归模型案例



## 第一节 回归模型

本章所讲内容，一般均指定量变量的相关分析与回归分析

。

### 一、回归分析

1. 回归分析：就是根据相关关系的具体形态，选择一个合适的数学模型，来近似地表达变量间的依赖关系。

2. 进行回归分析时，首先需要确定因变量和自变量。

(1) 因变量：被预测或被解释的变量，一般用 $Y$ 表示；

(2) 自变量：用来预测或解释因变量的变量，一般用 $X$ 表

示。



## 第一节 回归模型

3. 回归分析的一个重要应用就是预测，即利用估计的回归模型预估因变量数值。

	回归分析	相关分析
联系	需要依靠相关分析来表明现象数量变化的相关程度，高度相关时，进行回归分析寻求其相关的具体形式才是有意义的	依靠回归分析来表明现象数量相关的具体形式
区别	研究变量之间相关关系的具体形式，它对具有相关关系的变量之间的数量联系进行测定，确定相关的数学方程式	研究变量之间相关的方向和相关的程度，不能指出变量间相互关系的具体形式，无法从一个变量的变化来推测另一个变量的变化情况



## 典型真题

【真题·2023多选】关于相关分析和回归分析的说法，正确的有（ ）。

- A. 相关分析研究变量间相关的方向和相关程度
- B. 回归分析研究变量间相互关系的具体形式
- C. 相关分析和回归分析在研究方法和研究目的上有明显区别
- D. 相关分析可以从一个变量的变化来推测另一个变量的变化
- E. 相关分析中需要明确自变量和因变量



## 典型真题

答案：ABC

解析：相关分析不能指出变量间相互关系的具体形式，也无法从一个变量的变化来推测另一个变量的变化情况；A项错误。进行回归分析时，首先需要确定因变量和自变量；B项错误。只有当变量之间存在着高度相关时，进行回归分析寻求其相关的具体形式才有意义；C项错误。



## 第一节 回归模型

### 一、回归模型

#### 1. 回归模型分类

(1) 根据自变量的多少：一元回归模型和多元回归模型。

(2) 根据回归模型是否线性：线性回归模型和非线性回归模型。





## 第一节 回归模型

### 2. 一元线性回归模型

(1) 一元线性回归模型：描述两个变量之间相关关系的最简单的回归模型，只涉及一个自变量的回归问题。

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

$\beta_0$ 、 $\beta_1$ —模型的参数

因变量Y是自变量X的线性函数（ $\beta_0 + \beta_1 X$ ）加上误差项

$\varepsilon$

(1)  $\beta_0 + \beta_1 X$ ：反映了由于自变量X的变化而引起的因变量Y的线性变化；

(2) 误差项  $\varepsilon$  是随机变量，表示除X和Y的线性关系之外的随机因素对Y的影响；是不能由X和Y的线性关系所解释的Y的变异



## 第一节 回归模型

### (2) 回归方程

回归方程：描述因变量 $Y$ 的期望 $E(Y)$ 如何依赖自变量 $X$ 的方程。

一元线性回归方程的形式：

$$E(Y) = \beta_0 + \beta_1 X \quad (\text{方程的图示是一条直线})$$

**【注意1】**  $\beta_0$ 是回归直线的截距；

**【注意2】**  $\beta_1$ 是回归直线的斜率：表示 $X$ 每变动一个单位时， $E(Y)$ 的变动量。



## 典型真题

【真题·2023多选】以 $Y$ 为因变量、 $X$ 为自变量，建立一元线性回归模型 $Y = \beta_0 + \beta_1 X + \varepsilon$ 。关于该回归模型的说法（ ）。

- A.  $\beta_0$ 是因变量 $Y$ 的均值
- B.  $\beta_1$ 是回归直线的斜率
- C.  $\varepsilon$ 是不能由 $X$ 和 $Y$ 的线性关系所解释的 $Y$ 的变异性
- D.  $\beta_1$ 表示 $X$ 每变动一个单位时 $Y$ 的平均变动量
- E.  $\varepsilon$ 是随机变量



## 典型真题

答案：BCDE

解析： $\beta_0$ 是回归直线的截距。



## 第二节 最小二乘法和拟合效果分析

### 一、最小二乘估计

最小二乘法就是使因变量的观测值与估计值之间的离差（残差）平方和最小来求得  $\beta_0$  和  $\beta_1$  的方法。

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n y_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x}$$



## 第二节 最小二乘法 and 拟合效果分析

### 二、拟合效果分析

1. **决定系数** $R^2$ ：也称为拟合优度或判定系数，可以测度回归直线对样本数据的**拟合程度**。

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



## 第二节 最小二乘法和拟合效果分析

2. 决定系数的取值：在0到1之间。决定系数越高，模型的拟合效果就越好，即模型解释因变量的能力越强。

(1)  $R^2 = 1$ ，回归直线可以解释因变量的所有变化；所有观测点都落在回归直线上

(2)  $R^2 = 0$ ，回归直线无法解释因变量的变化，因变量的变化与自变量无关。

现实中的大都落在0和1之间，越接近于1，回归直线的拟合效果越好；越接近于0，回归直线的拟合效果越差



## 典型真题

【真题·2021多选】关于决定系数，下列说法正确的有（ ）。

- A. 决定系数的取值在0到1之间
- B. 决定系数越接近于0，说明模型的拟合效果越好
- C. 决定系数为1，说明回归直线可以解释因变量的所有变化
- D. 决定系数为0，说明回归直线无法解释因变量的变化
- E. 决定系数的取值，大体上说明了回归模型所能解释的因变量

变化占因变量总变化的比例





## 典型真题

答案：ACDE

解析：决定系数，也称为 $R^2$ ，可以测度回归直线对样本数据的拟合程度。决定系数的取值在0到1之间，大体上说明了回归模型所能解释的因变量变化占因变量总变化的比例。决定系数越高，模型的拟合效果就越好，即模型解释因变量的能力越强。如果所有观测点都落在回归直线上， $R^2=1$ ，说明回归直线可以解释因变量的所有变化。 $R^2=0$ ，说明回归直线无法解释因变量的变化，因变量的变化与自变量无关。现实中的 $R^2$ 大都落在0和1之间， $R^2$ 越接近于1，回归直线的拟合效果越好； $R^2$ 越接近于0，回归直线的拟合效果越差。



# 本章内容总结

## 第26章 回归分析

### 相关和回归的关系

#### 相关分析

研究变量之间相关的方向和相关的程度，不能指出变量间相互关系的具体形式，不能预测

#### 回归分析

研究变量之间相关关系的具体形式，确定相关的数学方程式，可以预测

### 回归模型分类

#### 根据自变量的多少

一元回归模型和多元回归模型

#### 根据回归模型是否线性

线性回归模型和非线性回归模型

### 一元线性回归模型

#### 模型

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

#### $\beta_0 + \beta_1 X$

反映了由于自变量X的变化而引起的因变量Y的线性变化

#### 误差项 $\varepsilon$

随机变量，表示除X和Y的线性关系之外的随机因素对Y的影响  
是不能由X和Y的线性关系所解释的Y的变异性

### 回归方程

一元线性回归方程： $E(Y) = \beta_0 + \beta_1 X$

### 最小二乘法

对回归模型进行参数估计的方法使得因变量的观测值与估计值之间的离差(残差)平方和最小来估计参数  
 $\beta_0$ 和 $\beta_1$ 的方法

### 决定系数 $R^2$

取值：在0到1之间。决定系数越高，模型的拟合效果就越好

$R^2 = 1$ ，回归直线可以解释因变量的所有变化；所有观测点都落在回归直线上

$R^2 = 0$ ，回归直线无法解释因变量的变化，因变量的变化与自变量无关

谢谢 观看  
THANK YOU