第二十六章　回归分析

第一节　回归模型
第二节　最小二乘法
第三节　模型的检验和预测

第一节　回归模型

回归分析的概念
一元线性回归模型

考点 1 回归分析的概念

1.含义

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，来近似的表达变量间的依赖关系。

【解释】回归分析是根据相关关系的具体形态（正相关还是负相关，是否是高度相关），选择一个合适的数学模型（也就是确定一个相关的数学方程式），根据这个数学方程式可以从已知量来推测未知量（例如从居民收入水平的数量变化情况来估算和预测食品支出的数量变化情况），来近似的表达变量间的依赖关系。

【回归分析的实质】是在相关分析的基础上，研究变量间的数量变化规律。

2、进行回归分析时，首先需要确定因变量和自变量

因变量：回归分析中，被预测或被解释变量称为因变量，一般用Y表示	自变量：用来预测或解释因变量的变量称为自变量，一般用x表示

【例 1】在研究边际消费倾向时，目的是预测在一定人均收入条件下的人均消费金额。因人均消费金额是被预测的变量，称为因变量；而用来预测人均消费的人均收入就是自变量。

【例 2】要研究质量和用户满意度之间的因果关系。

- 用户满意度是被预测的变量，称为因变量，用 Y 表示；
- 而用来预测用户满意度的质量为自变量，用 X 表示。

3、回归分析与相关分析的联系

①它们具有共同的研究对象（都是对变量间的相关关系进行研究），在具体应用时，常常必须互相补充。
②相关分析需要依靠回归分析来表明现象数量相关的具体形式。
③而回归分析则需要依靠相关分析来表明现象数量变化的相关程度。
④只有高度相关时，进行回归分析寻求其相关的具体形式才是有意义的。

4、回归分析与相关分析的区别

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：

①相关分析是研究变量之间相关的方向和相关的程度。

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

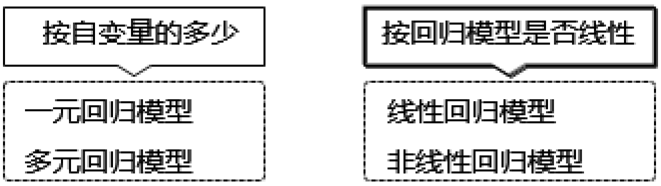
②回归分析是研究变量之间相关关系的具体形式，它对具有相关关系的变量之间的数量联系进行测定，确定相关的数学方程式，根据这个数学方程式可以从已知量来推测未知量，从而为估算和预测提供了一个重要方法。

【多选题】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.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
C.回归分析研究变量间相互关系的具体形式
D.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
E、相关分析中需要明确自变量和因变量
【答案】ACD
【解析】B，相关分析无法从一个变量的变化来推测另一个变量的变化情况；E，进行回归分析时，需要确定因变量和自变量。

考点 2 一元线性回归模型
1、 回归模型的分类



①如果在回归分析中，只包括一个自变量和一个因变量，且二者的关系可用一条直线近似表示，这种回归模型称为一元线性回归模型。
②如果回归分析中包括两个或两个以上的自变量，且因变量和自变量之间是线性关系，则称为多元线性回归模型。

2、一元线性回归模型
（1）一元线性回归是描述两个变量之间相关关系的最简单的回归模型
（2）回归模型可以用描述因变量 Y 如何依赖自变量 X 和误差项$\boldsymbol{\varepsilon}$的方程
表示为： $Y = \beta_0 + \beta_1 X + \varepsilon$
β_0、β_1为模型的参数（也叫回归系数）。
$\boldsymbol{\varepsilon}$即误差项，是一个随机变量，表示除 X 和 Y 的线性关系之外的随机因素对 Y 的影响。

【单选题】线性回归模型$Y = \beta_0 + \beta_1 X + \varepsilon$中误差$\boldsymbol{\varepsilon}$的含义是（　）。
A.回归直线的截距
B.除 X 和 Y 线性关系之外的随机因素对 Y 的影响
C.回归直线的斜率
D.观测值和估计值之间的残差
【答案】B
【解析】$\boldsymbol{\varepsilon}$ 即误差项，是一个随机变量，表示除 X 和 Y 的线性关系之外的随机因素对 Y 的影响。

2、一元线性回归模型
（3）描述因变量 Y 的期望 E（Y）如何依赖自变量 X 的方程称为回归方程。
一元线性回归方程的形式为：E（Y）= β_0 + β_1X
β_0是回归直线的截距，β_1为回归直线的斜率，表示 X 变化一个单位时，E（Y）的变动量。

第二节　最小二乘法

1、现实中，模型的参数β_0，β_1都是未知的，必须利用样本数据去估计，采用的估计方法是最小二乘法。
2、最小二乘法就是使得因变量的观测值与估计值之间的离差平方和最小来估计参数β_0和β_1的方法。

【单选题】对于一元线性回归方程*Y=β_0+β_1 X+ε*，确定*β*_0 和*β*_1 的方法是（　）。
A.二次平均
B.加权平均

C.斯特基方法
D.最小二乘法
【答案】D

【单选题】在回归分析中，估计回归系数的最小二乘法的原理是（　）。
A.使得因变量观测值与均值之间的离差平方和最小
B.使得因变量估计值与均值之间的离差平方和最小
C.使得观测值与估计值之间的乘积最小
D.使得因变量观测值与估计值之间的离差平方和最小
【答案】D

第三节　模型的检验和预测

回归模型的拟合效果分析
模型预测

考点 1 回归模型的拟合效果分析
（1）一般情况下，使用估计的回归方程之前，需要对模型进行检验：
①结合经济理论和经验分析回归系数的经济含义是否合理；
②分析估计的模型对数据的拟合效果如何（用决定系数来测度）；
③对模型进行假设检验。

决定系数
含义　决定系数，也称为$R2$，可以测度回归直线对样本数据的拟合程度。
取值
- 决定系数的取值在 0 到 1 之间。
- 决定系数越接近 1，回归直线的拟合效果越好。
- $R2 = 1$，说明回归直线可以解释因变量的所有变化。
- $R2 = 0$，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。

【单选题】下列关于回归分析的说法错误的是（　）。
A.描述因变量如何依赖自变量和误差项的方程称为回归模型
B.决定系数可以测度回归直线对样本数据的拟合程度
C.决定系数越接近 1，回归直线的拟合效果越好
D.$R2$=0，说明回归直线可以解释因变量的所有变化
【答案】D
【解析】 $R2$=1，说明回归直线可以解释因变量的所有变化。

【单选题】回归模型决定系数的取值范围是（　）。
A.-1 到 0 之间
B.0 到 1 之间
C.-1 到 1 之间
D.-∞到∞
【答案】B
【解析】决定系数的取值在 0 到 1 之间。决定系数越接近 1，回归直线的拟合效果越好。

考点 2 模型预测
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。

【教材例题】估计的城镇居民人均可支配收入和人均消费的一元线性直线回归方程为：
$\hat{Y}$=1292.6+0.629X，根据估计的回归方程，当城镇居民人均可支配收入 X=15000 元时，人均消费支出时多少？
【计算】将 X=15000 元代入回归方程，得：
$\hat{Y}$ =1292.6+0.629X = 1292.6+0.629×15000
=10731 元

【单选题】（2016）关于回归方程决定系数的说法，正确的有（ ）。
A.决定系数测度回归模型对样本数据的拟合程度
B.决定系数取值越大，回归模型的拟合效果越差
C.决定系数等于 1，说明回归模型可以解释因变量的所有变化
D.决定系数取值在[0，1]之间
E.如果决定系数等于 1，所有观测点都会落在回归线上
【答案】ACDE

模型预测

- 用 t 检验方法验证自变量 X 对因变量 Y 是否有显著影响
- 如果 P<0.05，则可以在 0.05 的显著性水平下拒绝原假设，认为自变量 X 对因变量 Y 有显著影响

第三节　模型的检验和预测

模型预测

	系数	标准误差	t 检验	P 值
截距	1368.674	107.0967227	12.77979	1.82×10^{-9}
城镇居民人均可支配收入	0.637271	0.004248599	149.9955	3.05×10^{-25}
最小二乘估计值分别为 1368.674 和 0.637271， “P 值”列给出自变量城镇居民人均可支配收入的 t 检验 P 值=3.05×10^{-25} 可以在 0.05 的显著性水平下拒绝β_1=0 的原假设，认为城镇居民人均可支配收入对城镇居民人均消费支出有显著影响。				

	系数	标准误差	t 检验	P 值
截距	-18716.2	16600.7096	-1.12743	0.26948
受教育年限（edu）	3669.002	1209.412447	3.033684	0.00529
职位（position）	27144.47	8197.976538	3.311118	0.00265
二元线性回归模型的调整后 R^2=0.605，表示受教育年限和职位的二元回归模型对年薪的变化可解释程度为 60.5%。 自变量受教育年限和职位的 t 检验 P 值 0.00529 和 0.00265 都小于 0.05，表明 2 个自变量都通过了 t 检验，对年薪有显著的线性影响。 在相同的职位上，受教育年限（edu）每增长 1 年，年薪平均增长 3669 元；相同受教育年限条件下，管理者年薪比一般职员平均增长 27144.47 元。 已知员工甲受教育年限为 10 年，在该公司为一般职员，可利用该模型预测其年薪大约为$\hat{Y}$=-18716.2+3669×10=17937.8 元。				