**回归分析的基本思想及其初步应用知识点总结-高中数学选修1-2第一章**

**一．知识要点，学习目标**

1.如果一组具有相关关系的数据　作出散点图大致分布在一条直线附近,那么我们称这样的变量之间的关系为线性相关关系(也称一元线性相关),这条直线就是回归直线,记为．

2.在所求回归直线方程中,当取时，与实际收集到的数据之间的偏差为,偏差的平方为

即以 来刻画出个点与回归直线在整体上**偏差的平方和**,显然Q取最小值时的的值就是我们所求的。

应注意,这个最小距离不是通常所指的各数据的点到直线的距离,而是各数据点沿平行y轴方向到直线的距离.









y





















图1

















y









图2





　　　　　这就是我们所要求的公式（无特殊要求时以此公式求回归方程中的、）．

其中为样本数据,为样本平均数，称为样本点中心，且所求线性回归直线经过样本点中心点（如图2所示）．

当回归直线斜率时,为线性正相关, 时为线性负相关.

**线性回归分析：**

**回归分析**是对具有相关关系的两个变量进行统计分析的一种常用方法．前面我们给出了线性回归方程，这里我们主要结合教材分析一元线性回归问题．

**1．以散点图分析线性相关关系，**散点图是较粗略地分析和判断两个具有相关关系的变量是否线性相关的问题，如果是线性相关的，我们可以求其线性回归方程，如果不是线性向相关的，即使求得线性回归方程，也是无效的；也就是说不能对一些数据进行分析判断，不能应用它解决和解释一些实际问题．

**2．以相关系数分析线性相关关系的强弱**

两个变量之间的相关关系的样本相关系数：

可衡量是否线性相关，以及线性相性关系的强弱．由于分子与线性回归方程中的斜率的分子一样（这也给出了公式的内在联系以及公式的记法），因此，当时，两个变量正相关；当时两个变量负相关．当的绝对值接近１，表明两个变量的线性相关性很强；当的绝对值接近０，表明两个变量之间几乎不存在线性相关关系．规定当时，我们认为两个变量有很强的线性相关关系．

**3．解释变量与随机误差对预报精度的影响以及残差分析**

**（1）有关概念**

由于样本数据点与一元线性回归方程上的点还有一定的差距，这说明了另外的一个因素随机误差的影响．

于是有线性回归模型

其中和为模型的未知参数；

图3













y

















称为解释变量，称为预报变量；

是与之间的误差，

叫随机误差。

随机误差的估计值为：



称为相应于样本点的残差（如图３）．

（2）通过残差分析判断模型拟合效果

　由计算出残差，，…，，然后选取横坐标为编号、或解释变量或预报变量，纵坐标为残差作出残差图．通过图形分析，如果样本点的残差较大，就要分析样本数据的采集是否有错误；另一方面，可以通过残差点分布的水平带状区域的宽窄（如教材图1.1－5），说明模型拟合效果，反映回归方程的预报精度．

**4．相关指数反应模型的拟合效果**

＝

**（1）变量剖析理解：**

**总偏差平方和** ，表示解释变量和随机误差产生的总的效应；

**残差平方和**，表示了随机误差效应；

**其差**　，表示了解释变量效应．

**（２）模型拟合效果**

①残差平方和占总偏差平方和的百分比，反映了随机误差对预报变量（总效应）的贡献率，显然这个数值越大，模型的拟合效果越差．

②解释变量效应占总偏差平方和的百分比反映了解释变量对预报变量（总效应）的贡献率； 因此，越接近１（即越接近0），表示回归的效果越好，即解释变量和预报变量的线性相关性越强．

**四．非线性回归的问题转化为线性回归问题**

　　　　　（1）作散点图确定曲线模型



图4













根据收集的数据作散点图（如图４），

　　　　可见两个变量不呈线性相关关系．而是

分布在某一条指数函数曲线的

周围，也可以认为样本点集中在某二次曲线的附近．

　（2）非线性转化为线性

这时通过对数变换把指数关系

变为线性关系；通过换元把二次函数关系变换为

线性关系．　在这两种情况下就可以利用线性回归模型，建立和之间的非线性回归方程了．

（3）比较两种模型的拟合效果

对于给定的样本点

　　ⅰ可以通过转换后的对应数表作散点图来确定线性回归的拟合情况，判断选用哪一种曲线模型较为合适；

ⅱ可以通过原始数据及和之间的非线性回归方程列出残差对比分析表，一

般通过残差平方和比较两种模型的拟合效果，显然残差平方和较小的拟合效果较好；

ⅲ还可以用来比较两个模型的拟合效果，越大（越接近１），拟合效果越好。