## 回归分析

回归分析可以从三种现实意义来解释：

**因果性：观测项 = 机制项 + 干扰项**

**预测性：预测项 = 预测项 + 误差项**

**描述性：观测项 = 概括项+残差项**

第一种理解方式接近古典计量经济学的视角；第二种方式适用于工程学领域；第三种反映社会定量性结果。三种理解方式分别从不同角度看数据。

下面我们从实际生活中的身高为例子来解释三种视角。假如我们要得到某人的身高，可能大家会说直接用身高测量仪不就可以的了，当然的这里我们如果要求你不能用仪器的方法。这样我们如何问许多人，可能会从种得到许许多多的方法，每个人的方法或有理或逗比。

这里我也说三条方法：第一种方式，大家都知道身高是由于基因(内在因素)和生活环境(外在因素)决定， 假设我们能得到这个人的所有信息，那么我们就可以通过第一视角来对这个人的身高作解释：身高=基因and环境 + 基因检测时仪器的干扰等等，这样我们就可以把基因and环境⬄机制项，基因检测时仪器的干扰等等⬄干扰项；第三方式，告诉你这个人的身边周围的人身高，；第二种方式， ……

如果上述的方法，都能被实现，我们真的到预测到这个人的身高，也就是我们的观测值。各种方法，有的有些人认为其中一些方法可靠有行，另一些方法不能被接受。评论的人多了，可能会所有的方法都会被人否定。那么问题来了，是不是我们真的不能得到这个人的”身高”，实现情况是我们是可以得到身高的，其实每种方式都有其自高的合理和不合理的地方，机制项，预测项，概括项是合理的，干扰项，干扰项，残差项是不合理的，………，(这里不谈论这个问题，见个人学习思考章节)。

一般的，一元线性回归模型可以表示为：



称为因变量，是自变量，，是随机误差项。

和为模型参数，通常是未知的，需要通过数据样本进行估计。我们定义为结构项。

下面我们把身高测量问题放到模型，y 是身高，x是基因，为出生身高，是头发，穿鞋什么的高度。