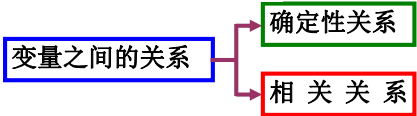


回 归 分 析

4.1 相关关系与回归



$S = \pi r^2$ 确定性关系
身高和体重 相关关系

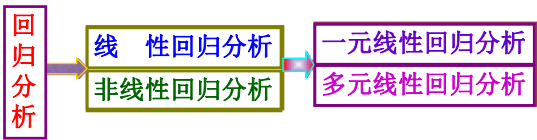
相关关系的特征是:变量之间的关系很难用一种精确的方法表示出来.

确定性关系和 相关关系的联系

由于存在测量误差等原因,确定性关系在实际问题中往往通过相关关系表示出来;另一方面,当对事物内部规律了解得更加深刻时,相关关系也有可能转化为确定性关系.

处理变量之间的相关关系的数学方法:

- 1. **相关分析**——变量均作为随机变量，变量间彼此平等。
- 2. **回归分析**——某些变量是可以观测和控制的非随机变量，另一个变量与之有关，但它是随机变量。把这个随机变量作为因变量（相应变量，Response Variable），可控变量作为自变量(预报变量，Predictor Variable)。变量的地位不可交换。



- 回归分析目前是所有统计分支中应用最广泛的一门学科之一，它被用于几乎所有的研究领域及工农业生产，包括产品的统计质量管理，市场预测，自动控制中数学模型的建立、气象预报、地质勘探、医学卫生等等。

- 回归分析，随着自变量的增加，计算会变得相当复杂，所以在计算机出现以前，它的应用受到了一定的限制。随着计算机的不断发展，速度成倍增加，回归分析的复杂的计算问题已经基本解决。目前世界上有许多能进行回归分析的统计软件：如SAS、SPSS、S-PLUS、R等等。

回归分析的基本思想和方法以及“回归(Regression)”名称的由来归功于英国统计学家 F·Galton（1822—1911 年）。

F·Galton 和他的学生、现代统计学的奠基者之一 K·Pearson (1856——1936 年)在研究父母身高与其子女身高的遗传问题时，观察了1078 对夫妇，以每对夫妇的平均身高作为解释变量X，而取他们的一个成年儿子的身高作为被解释变量Y，将结果在平面直角坐标系上绘成散点图，发现趋势近乎一条直线。计算出的回归直线方程为

$$\hat{Y} = 33.73 + 0.516X$$

| | |
|--|--|
| <p>这个结果表明，虽然高个子父辈有生高个子儿子的趋势，但父辈身高增加一个单位，儿子身高仅增加半个单位左右。低个子父辈的儿子们虽然仍为低个子，平均身高却比他们的父辈增加了。</p> <p>正是因为子代的身高有回到父辈平均身高的这种趋势，才使人类的身高在一定时间内相对稳定，没有出现父辈个子高其子女更高，父辈个子矮其子女更矮的两极分化现象。</p> <p>7</p> | <p>F·Galton 用他最有说服力且最风趣的一段话来概括了这个结论：</p> <p>孩子的遗传一部分来自父母，一部分来自祖先。家谱向前推得越远，其祖先越多样越不同，直到他们成为从一个大种族随机抽取的多样性的样本为止。这个规律解决了为何天才无法全部遗传给其后代的问题□ □ 这个规律是公正的；无论好的方面还是坏的方面的遗传都会打相同的折扣。如果它使一些有天赋的父母期待其子女也很有天赋的愿望化为泡影，那么它同样也会使另一些父母减少担心，因为他们的子女同样也不会全部继承他们的缺陷和疾病。</p> <p>8</p> |
| <p>这生动地说明了生物学中“种”的概念的稳定性。正是为了描述这种有趣的现象，F·Galton 引进了“回归(regression)”这个词来描述父辈身高X与子代身高Y的关系。尽管“回归”这个名称的由来具有其特定的含义，人们在研究大量的问题中变量X与Y之间的关系并不具有这种“回归”的含义，但借用这个词把研究变量X与Y之间的统计关系的数学方法称为“回归分析”，也算是对F·Galton 这个伟大的统计学家的一种纪念。</p> <p>9</p> | |