# 探究 1　一元线性回归模型（P105）

探究人：　　　　　　　　　时间：　　　　　　　　　指导老师：

## 探究目的：

1、当两个变量线性相关时，用 Geogebra 探究成对样本数据的线性回归模型；

2、体会回归直线方程与成对数据的关系。

## 器材：

电脑（或平板或手机等设备），Geogebra 软件、实验手册

## 探究步骤：

**（参照教材** P105-P110 实施以下实验**）**

**实验 1：探究儿子的身高与父亲身高的线性相关模型，并利用模型预测。**

**第一步**：打开实验资源包中的实验软件（如图 1，右下侧表格为身高数据输入区，已经输入一组数据，请参照书 105 页的表格输入剩余数据，软件自动在右上侧描点，右上侧为散点图图模拟函数呈现区，左侧为实验过程代数呈现区）。

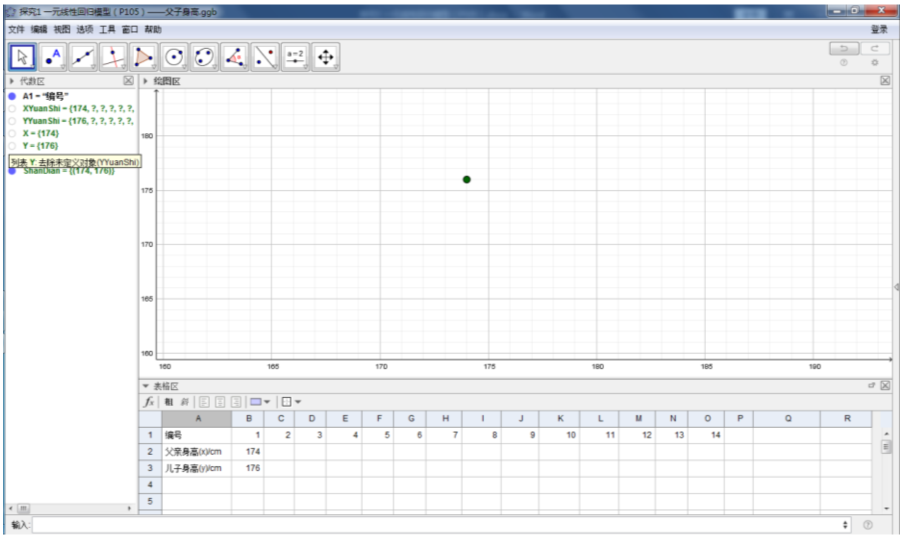

图 1

**第二步**：儿子的身高和父亲的身高之间是函数关系吗？能用函数模型刻画父子身高的关系吗？（比对第 6 和第 8 或者第 5 和第 11 等数据说明）

结论：儿子的身高和父亲的身高之间_____（是、不是）函数关系，______（能、不能）用函数关系刻画，______（能、不能）用函数模型刻画。

**第三步**：分析散点的分布，它们大致分布在___________附近，表明儿子身高和父亲身高这两个变量之间有较强的_______相关关系，因此我们可以用一次函数模型来刻画父亲身

高对儿子身高的影响，而把影响儿子身高的其它因素（如饮食习惯等）作为______误差，得到刻画两个变量之间关系的线性回归模型：______

**第四步**：参照书第 108 页、第 109 页最小二乘法的公式，计算线性回归模型中斜率参数 b、截距参数 a：

首先，计算父亲身高、儿子身高的平均数：底部输入栏输入“xJunZ=平均数[X]”、“yJunZ=平均数[Y]”，得到计算结果如：xJunZ = 173.214 yJunZ = 174.286，

其次，计算父亲身高、儿子身高与平均值的差值：通过底部输入栏输入“XChaZ=X-xJunZ”、“YChaZ=Y-yJunZ”，得到计算结果如：XChaZ = {0.786, -3.214, -0.214, -4.214,

YChaZ = {1.714, 1.714, -4.286, -4.286,

最后，计算分子、分布、b、a：通过底部输入栏输入“FenZi=sum[XChaZ* YChaZ]”、“FenMu=sum[XChaZ^2]”、“b=FenZi/FenMu”、“a=yJunZ-b*xJunZ”，得到计算结果如：

FenMu = 442.357
FenZi = 371.143
b = 0.839
a = 28.957

；

**第五步**：输入回归直线方程并画图：通过底部输入栏输入“y=b*x+a”，得到结果如：f: y = 0.839x + 28.957 和回归直线图像；

**第六步**：观察散点和回归直线的关系，思考：散点图中的全部点或者部分点一定会在回归直线上吗？通过底部输入栏输入点坐标“（xJunZ，yJunZ）”，观察生成的点与回归直线图像有什么关系？

（结论：散点图中的点一般______在回归直线上；由平均值生成的点______回归直线上。）

**第七步**：改变一组父子身高的数据对，或者在第 14 组后面输入更多父子身高的数据对（最多支持 100 组），观察散点、身高的均值点和回归直线方程，思考：散点图中的点、身高的均值点、回归直线方程有变化吗？前两者和回归直线图像有什么关系？

（结论：改变（增加）父子身高值数据对，均值点、回归直线方程都有变化，说明即使经验回归直线方程不过身高数据对任意一点，却与每一个身高数据对______关系；散点图中的点一般______在回归直线上；但平均值生成的点______回归直线上。）

**第八步**：经验回归直线方程与过散点图中某两点的直线方程比较（教材第 107 页），谁在整体上更接近散点图（谁的误差平方和更小）：

首先，用上部工具栏中的“直线”工具，先后单击你认为能较好体现散点图走势的两点（如第 1 和第 8 对），得直线方程如：g: y = 1.1x - 15.4，

其次，计算散点图与两条直线的误差平方和：通过底部输入栏输入 “FWuChaHe=误差平方和(ShanDianTu, f)” 、 “GWuChaHe=误差平方和(ShanDianTu, g)” ，得散点图与两条直线的误差平方和，如：FWuChaHe = 85.464 GWuChaHe = 125.71，

最后，比较误差值的大小，判断哪条直线在整体上更接近散点图：直线 g 的误差平方和______，所以直线 f 在整体上更______散点图（还可以继续尝试经过其它两点的直线，与经验回归方程比较，谁的误差平方和更小）；

**第九步**：用经验回归方程，输入父亲的身高（如第 5、11 组值中，父亲的身高 182），预测儿子的升高：通过底部输入栏输入 “YuChe1=（182，f（182））” 得倒结果如：YuChe1 = (182, 181.657)，其中 181.657 即为儿子身高的预测值。

请思考：预测值一定等于样本值吗？

结论：__________，因为还有其他影响儿子身高的因素，可以把预测值视为该父亲身高总体中儿子身高_________的______。

**第十步**：观察回归直线图像，思考：经验回归方程 y=bx+a 中斜率 b 如何理解？

结论：斜率 b 可以理解父亲身高每增加 1cm，其儿子身高______增加 _____cm（如_____cm）

## 探究结论：

当由两个变量的数据对（如数据对（父亲身高，儿子身高））生成的散点图大致分布在一条直线附近时，可用一元线性回归模型拟合两个变量间隐藏的关系。

## 交流与反思：

1、经验回归直线方程中的字母参数 b、截距参数分别表示什么意义？

2、是否存在其他回归方程，在整体上比经验回归直线方程更接近散点图？

## 探究练习：

1． “绿水青山就是金山银山” 的生态文明发展理念已经深入人心，推动着新能源汽车产业的迅速发展.下表是 2020 年我国某地区新能源乘用车的前 5 个月销售量与月份的统计表：

| 月份代码$x$ | 1 | 2 | 3 | 4 | 5 |
| --- | --- | --- | --- | --- | --- |
| 销售量$y$(万辆) | 0.5 | 0.6 | 1 | 1.4 | 1.5 |

由上表可知其线性回归方程为：$\hat{y}=0.28x+a$，则$a$的值为（　）

A．0.16　　　　B．1.6　　　　C．0.06　　　　D．0.8

1．设一个回归方程为$\hat{y}=3+1.2x$，则变量$x$增加一个单位时（　）．

A．$y$平均增加 12 个单位　　　　B．$y$平均增加 3 个单位

C．$y$平均减少 1.2 个单位　　　　D．$y$平均减少 3 个单位

3．已知$x$与$y$之间的一组数据如下表：

| $x$ | 3 | 4 | 5 | 6 |
| --- | --- | --- | --- | --- |
| $y$ | 30 | 40 | 60 | 50 |

若$y$与$x$线性相关，根据上表求得$y$与$x$的线性回归方程，$\hat{y}=\hat{b}x+\hat{a}$中的$\hat{b}$为 8，据此模型预报$x=7$时$y$的值为（　）

A．70　　　　B．63　　　　C．65　　　　D．66