文科概统第四节　回归分析与独立检验

一、回归方程

1、为了解某社区住户的年收入和年饮食支出的关系，抽取了其中 5 户家庭的调查数据如下表

年收入 x(万元)	3	**4**	5	6	7
年饮食支出 y(万元)	1	**1.3**	1.5	2	2.2

(Ⅰ)根据表中数据用最小二乘法求得回归直线方程,

(Ⅱ)请预测年收入为 9 万元家庭的年饮食支出

(III)从 5 户家庭中任选 2 户，求恰有一户家庭饮食支出小于 1.6 万元的概率

解：(Ⅰ) $\overline{x}=5,\overline{y}=1.6$，$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=3.1,\sum_{i=1}^{n}(x_i-\overline{x})^2=10, b=\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}=0.31$

代入 $\hat{y}=bx+a$,得 $a=0.05$　故 $\hat{y}=0.31x+0.05$，

(Ⅱ)当 $x=9$ 时,解得 $\hat{y}=2.84$ 万元

所以年收入为 9 万元家庭的年饮食支出约为 2.84 万元

(Ⅱ) 年饮食支出小于 1.6 万元的家庭为 a,b,c; 年饮食支出不小于 1.6 万元的家庭为 M,N;

从 5 户家庭中任选 2 户，求恰有一户家庭饮食支出小于 1.6 万元为事件 A

所有基本事件为 $\{a,b\},\{a,c\},\{a,M\},\{a,N\},\{b,c\},\{b,M\},\{b,N\},\{c,M\},\{c,N\},\{M,N\}$

共 10 个

事件 A 包含的基本事件有 $\{a,M\},\{a,N\},\{b,M\},\{b,N\},\{c,M\},\{c,N\}$ 共 6 个

所以 $P(A)=\dfrac{6}{10}=0.6$

答: 从 5 户家庭中任选 2 户，求恰有一户家庭饮食支出小于 1.6 万元的概率为 0.6

2、某同学在研究性学习中，收集到某制药厂今年前 5 个月甲胶囊生产产量(万盒)的数据如下表所示

月份 x	1	**2**	3	4	5
y(万盒)	4	**4**	5	6	6

该同学为了求出 y 关于 x 的线性回归方程 $\hat{y}=\hat{b}x+\hat{a}$，根据表中数据已经正确计算出 $\hat{b}$=0.6，试求出 $\hat{a}$ 的值，并估计该厂 6 月份的甲胶囊产量数

解：(1) $\overline{x}=\dfrac{1}{5}(1+2+3+4+5)=3,\overline{y}=\dfrac{1}{5}(4+4+5+6+6)=5$

因回归直线 $\hat{y}=\hat{b}x+\hat{a}$ 过点 $(\overline{x},\overline{y})$，于是 $a=\overline{y}-b\overline{x}=5-0.6\times3=3.2$

于是 $\hat{y}=0.6x+3.2$，所以 $\hat{y}=0.6\times6+3.2=6.8$

二、独立检验

1、某学校为调查高三年级学生的身高情况，按随机抽样的方法抽取 80 名学生，得到男生身高情况的频率分布直方图 1 和女生身高情况的频率分布直方图 2,已知图 1 中身高在 170 到 175cm 的男生人数 16 人

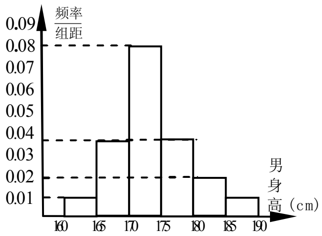


图1

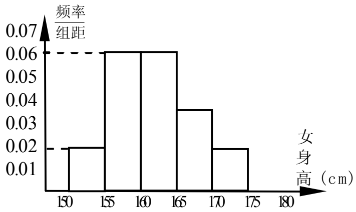


图2

(Ⅰ)试问在抽取学生中，男、女生各有多少人

(11)根据频率分布直方图，完成下列 2×2 列联表，并判断能有多大的把握认为身高与性别有关？

(III)在上述 80 名学生中，从身高在 170 到 175 之间的学生中按男女性别分层抽样的方法抽出 5 人，从这 5 人中选派 3 人当旗手，求 3 人中恰好有一名女生的概率。

$$K^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$P(K^2\geq k_0)$	0. 025	**0.010**	0. 005	0.001
k_0	5.024	**6.635**	7.879	10.828

解：(Ⅰ) 设男生总数 n，在 170 到 175cm，男生人数 16 人

频率=$0.08\times5=0.4$,$0.4n$=16,$n=40$，于是男、女生 40 人

	≥170cm	＜170cm	总计
男生身高	30	10	40
女生身高	4	36	40
总计	34	46	80

K^2的观测值$k_0=\dfrac{80\times(30\times36-4\times10)^2}{40\times40\times34\times46}=34.578>10.828$

有 99.9%的把握认为身高与性别有关系。

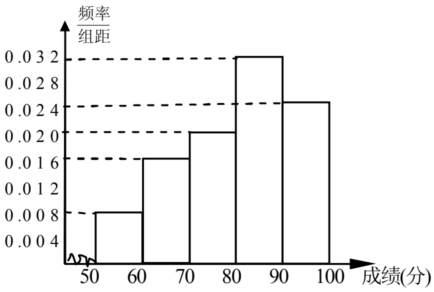
(III) 身高在 170 到 175 之间的学生中按男 16 人，女 4 人

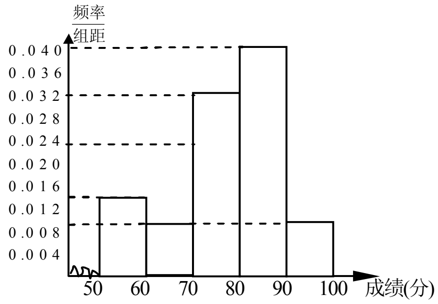
分层抽样的方法抽出 5 人，抽得男 4 人，女 1 人

从这 5 人中选派 3 人当旗手，则共有 10 种选法，其中选到女生的方法有 6 种

于是所求的概率＝0.6

2． 某中学将 100 名高一新生分成水平相同的甲乙两个平行班，每班 50 人，陈老师采用 A、B 两种不同的教学方式分别在甲乙两个班级进行教改实验。为了解教学效果，期末考试后，陈老师对甲乙两个班级的学生成绩进行统计分析，画出频率分布直方图如图。记成绩不低于 90 分者为成绩优秀





(Ⅰ)从乙班随机抽取 2 名学生的成绩，记成绩优秀的个数为ξ，求ξ的分布列和数学期望

(Ⅱ)根据频率分布直方图填写下面 2×2 列联表，并判断是否有 95%的把握认为：成绩优秀与教学方式有关

$$k^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$P(K^2\geqslant k_0)$	0.10	0. 05	0. 025	0.010
k_0	2.706	3.841	5.024	6.635

	甲班	乙班	总计
成绩优秀	12	4	16
成绩不优秀	38	46	84
总计	50	50	100

解:（Ⅰ) 2×2 列联表如图

(Ⅱ) $k^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}=\dfrac{100(12\times46-4\times38)^2}{50^2\times16\times84}=4.762$

由于 4.762>3.841, 所以有 95%的把握认为成绩优秀与教学方式有关