**一 线性回归**

利用最小平方误差进行建模的一种回归分析

误差经过化简后成是关于y=mx+b的m和b的函数，三维碗形，求其最低点即误差最小值

分别对m和b求偏导令等式为0可求出m和b





决定系数：

**二 协方差**

协方差 





因此斜率

**三 分布**



3.1皮尔逊****检验

以商店人流量为例周日放假：百分比为店家给出数据，观察为实际，共观察200人

周 ： 一 二 三 四 五 六

期望百分比： 10 10 15 20 30 15

观察 ： 30 14 34 45 57 20

期望 ： 20 20 30 40 60 30

解答 零假设：店家正确

备择假设：店家错误

假设店家正确

****

显著性水平设为5%，自由度为6-1=5，

查阅****分布表格得到临界值11.07<11.4

显著性水平低于5%，拒绝零假设

3.2列联表****检验（注意自由度）

药物对病的影响：

药物1 药物2 安慰剂 总（期望生病比例为总生病人数除以总人数）

有病 20 30 30 80

期望 25.3 29.4 25.3 21%

没病 100 110 90 300

期望 94.7 110.6 94.7 79%

总共 120 140 120 380

零假设： 药物无影响

备择假设：药物有影响

显著性水平设为10%

假设零假设成立



自由度（2-1）\*（3-1）=2

查阅****分布表格得到临界值4.60>2.53

无法拒绝零假设

3.3方差分析

设三组数据（SST总平方和 ）

第一组 第二组 第三组

1. 5 5

2 3 6

1. 4 7

总均值

总平方和

总平方和自由度9-1=8 即(mn-1)

组内平均值

组内平方和

组内平方和自由度：每组为2，三组共为6即m(n-1)

组间平方和

组间平方和自由度：三个组，3-1=2 即(m-1)

8=6+2 30=6+24

方差分析结果显示：组内平方和+组间平方和=总平方和

组内自由度+组间自由度=总自由度

即总波动程度=组内波动程度+组间波动程度

**3.4检验（组内检验除以全体检验）**

假设上述三组数据是三种不同食物对体重的影响，设显著性水平10%

零假设 ：均值相同 食物不会产生差别

备择假设：均值不同 食物会产生影响

假设零假设成立，服从F分布

****

查阅F分布临界值3.46<12

拒绝零假设

3.5相关性和因果性（具有显著差异）

**相关性**指两个或者多个事物同时发生具有关联

**因果性**指因为 A 所以 B

**四：归纳推理与演绎推理**

**归纳（Inductive Reasoning）：寻找规律或趋势，然后推广也就是利用已有信息进行趋势外推**

**演绎（Deductive Reasoning）:由一些正确的数据或事实出发，演绎得到其他正确事实。**