



# 1、初等概率模型



## 问题1：水果店的合理进货模型

某时令水果店每售出一百千克水果，可以获得利润250元，若当天进货不能出售出去，则每一百千克将损失325元。该水果店根据预测分析，每天的需求量和对应的概率值如下表：

水果需求量/百千克	0	1	2	3	4	5	6	7	8
相应的概率值	0.05	0.1	0.1	0.25	0.2	0.15	0.05	0.05	0.05

在这样的需求结构下，水果店主希望知道，他应该每天进多少水果才能够获得最大的利润？



# 1、初等概率模型



## 问题1：水果店的合理进货模型

### 问题的分析：

该问题为一个随机存储问题，要研究这类问题，主要是按平均进货量（即数学期望）准则来讨论。

### 问题的假设：

(1) 当不满足需求，即缺货时，店主没有任何损失，即不考虑缺货所带来的损失。

(2) 水果店的纯利润为卖出水果后所获利润与因未卖出的水果所带来的损失部分之差。



# 1、初等概率模型



## 问题1：水果店的合理进货模型

**模型的建立与求解：**利用概率知识及经济学中边际分析的方法，综合分析讨论这个问题。

不妨记需求量为随机变量  $\xi$ ，则需求量的期望值为  $E(\xi) = 3.65$ 。

人们对水果的需求量大致在 3.65 百千克左右变化。这样基本可以确定该水果店的进货量应在 2 百千克至 4 百千克之间比较合适。

这只是直观的推测，对 2 百千克，3 百千克和 4 百千克这三个数字进行计算分析，如果得不到最优结果，则可以再进一步考虑其他情况。



# 1、初等概率模型



## 问题1：水果店的合理进货模型

**(1) 水果店每天进货量为2百千克情况：**  
**由于该水果店每售出一百千克水果，能够获得利润250元；若不能出售时每百千克损失325元。**  
**进货 2 百千克时的需求量与纯利润表**

需求量	0	1	2	3	4	5	6	7	8
纯利润	-650	-75	500	500	500	500	500	500	500

**水果店纯利润的期望值为**

$$E(\eta) = 0.05 \times (-650) + 0.1 \times (-75) + 0.1 \times 500 + 0.25 \times 500 + 0.2 \times 500 \\ 0.15 \times 500 + 0.05 \times 500 + 0.05 \times 500 + 0.05 \times 500 = 385$$



# 1、初等概率模型



## 问题1：水果店的合理进货模型

**(2) 水果店进货量为3百千克情况：相应的需求量与对应的纯利润计算结果如下表所示。**

**进货 3 百千克时的需求量与纯利润表**

需求量	0	1	2	3	4	5	6	7	8
纯利润	-975	-400	175	750	750	750	750	750	750

**水果店纯利润的期望值为**

$$E(\eta) = 0.05 \times (-975) + 0.1 \times (-400) + 0.1 \times 175 + 0.25 \times 750 + 0.2 \times 750 \\ 0.15 \times 750 + 0.05 \times 750 + 0.05 \times 750 + 0.05 \times 750 = 491.25$$



# 1、初等概率模型



## 问题1：水果店的合理进货模型

**(3) 水果店进货量为4百千克情况：相应的需求量与对应的纯利润计算结果如下表所示。**

**进货 4 百千克时的需求量与纯利润表**

需求量	0	1	2	3	4	5	6	7	8
纯利润	-1300	-725	-150	425	1000	1000	1000	1000	1000

**水果店纯利润的期望值为**

$$E(\eta) = 0.05 \times (-1300) + 0.1 \times (-725) + 0.1 \times (-150) + 0.25 \times 425 + 0.2 \times 1000 \\ 0.15 \times 1000 + 0.05 \times 1000 + 0.05 \times 1000 + 0.05 \times 1000 = 453.75$$



# 1、初等概率模型



## 问题1：水果店的合理进货模型

该水果店每天的水果进货量为3百千克相对获得利润较大。那么问题是否是3百千克的进货量一定就是最好的呢？

引入边际分析方法，边际分析方法是西方经济学中最基本的分析方法之一。

通过已知信息，判定水果店每增加一百千克的进货量，所带来的利润或损失，进而判断进货量的合理性。

如果水果店现已有 $n$ 百千克水果，那么再进1百千克水果，从而就存有 $n+1$ 百千克水果。



# 1、初等概率模型



## 问题1：水果店的合理进货模型

首先给出以下两个概念：

**边际利润 (Marginal Profit)：** 由所增加的1个单位水果带来的纯利润，记为MP。

**边际损失 (Marginal Loss)：** 由所增加的1个单位水果所导致的损失，记为ML。

显然，当  $MP \cdot P(\xi \geq n+1) \geq ML \cdot (1 - P(\xi \geq n+1))$  时，增加1单位（即1百千克）水果是合算的，即相应的概率为  $P(\xi \geq n+1) \geq \frac{ML}{ML + MP}$ 。由  $ML=325$ ， $MP=250$ ，从而有

$$P(\xi \geq n+1) \geq \frac{325}{325 + 250} \approx 0.5652$$





# 1、初等概率模型



## 问题1：水果店的合理进货模型

当销售概率大于0.5652时，水果店应再增加1百千克水果的进货量才是合算的。从已知的需求量与对应概率值的关系：

$$P(\xi \geq 8) \geq 0.05,$$

$$P(\xi \geq 7) = 0.05 + 0.05 = 0.1,$$

$$P(\xi \geq 6) = 0.05 + 0.05 + 0.05 = 0.15,$$

$$P(\xi \geq 5) = 0.15 + P(\xi \geq 6) = 0.15 + 0.15 = 0.3,$$

$$P(\xi \geq 4) = P(\xi \geq 5) + P(\xi = 4) = 0.3 + 0.2 = 0.5,$$

$$P(\xi \geq 3) = P(\xi \geq 4) + P(\xi = 3) = 0.5 + 0.25 = 0.75.$$

该水果店的需求量大于等于4百千克的概率小于0.5652，而需求量大于等于3百千克的概率大于0.5652。从而进货量应为3百千克为好。



## 2、简单统计模型



### 问题1：大学生平均月生活费的测算模型

大学生的日常生活水平随着整个时代的变迁发生着巨大的变化。我们了解一下，目前在校大学生的日常生活费支出与来源状况。

根据随机抽样的理论，2002年对北京某高校本科生的月生活费支出状况进行了抽样调查。

本次问卷调查对在校男女本科生共发放问卷300份，回收问卷291份，其中有效问卷共265份。

调查数据经整理后，得到全部265名学生和按性别划分的男女学生的生活费支出数据。



## 2、简单统计模型



### 问题1：大学生平均月生活费的测算模型

#### 模型假设

(1) 抽样是相互独立的，所抽到的样本都是简单随机样本。

(2) 总体即大学生日常生活费支出服从正态分布。  
用  $X_i$  表示第  $i$  个样本，即生活费支出额；  
 $\bar{X}$  表示样本均值，即所抽到学生的日常生活费支出的平均值；

$S$  表示样本标准差，即样本值与样本均值的偏离程度的度量；

$n$  是样本容量，即共抽到的有效问卷数。



## 2、简单统计模型



### 问题1：大学生平均月生活费的测算模型

#### 模型建立与求解

根据抽样结果，使用95%的置信水平，相应置信区间：

$$\left( \bar{X} - t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} \right)$$

**结论：**全校本科生的月生活费平均水平在520.70 ~ 554.40元之间；男生的月生活费平均水平在505.15 ~ 552.43元之间；女生的月生活费平均水平在545.83 ~ 596.65元之间。



## 2、简单统计模型



### 问题1：大学生平均月生活费的测算模型

#### 模型评价与应用

**模型用到了估计精度为95%的参数的区间估计，并且按性别不同，给出了不同的区间估计。**

**模型也可应用到很多实际问题的估计上，比如：一个普通家庭日常收入与支出状况、一个城市人均住房情况等问题统计分析。**



## 2、简单统计模型



### 问题2：吸烟对血压的影响模型

**“吸烟有害健康”，请你建立一个数学模型，分析说明吸烟对人体有害的影响，这里可以只就吸烟对高血压病的影响作用。**

#### **(1) 问题分析**

**为了研究吸烟对人体血压的影响，对吸烟的66人  
和不吸烟的62人两类人群进行24小时动态监测，分别  
测量24小时的收缩压(24hSBP)和舒张压(24hDBP)，  
白天(6:00~22:00)收缩压(dSBP)和舒张压(dDBP)，  
夜间(22:00~次日6:00)收缩压(nSBP) 和舒张压  
(nDBP)。**



## 2、简单统计模型



### 问题2：吸烟对血压的影响模型

(2) 模型假设：将人群（样本总体）分为两类：吸烟者和不吸烟者，分别记为A类和B类，主要研究这两类人血压的分布情况。假设：

A和B两类人的血压都服从正态分布，均值分别为 $\mu_1, \mu_2$ ，而方差相同。抽样是随机的，相互独立的。

小概率事件在一次试验中是不可能发生的。

选取样本容量 $n_A = 66, n_B = 62$ ，显著性水平 0.05， $\bar{x}_A, \bar{x}_B$  表示吸烟者和不吸烟者血压的样本均值， $S_1, S_2$  表示两类样本的标准差。



## 2、简单统计模型



### 问题2：吸烟对血压的影响模型

**(3) 模型建立：**吸烟对人体血压是否有影响？从这些数据中能得到什么样的推断？

吸烟者和不吸烟者两类样本分别来自两个非常大的总体，这个问题需要从两个样本的参数（均值与标准差）来推断总体参数的性质。

分别对6项血压指标作假设检验，针对每组数据指标提出假设：

$$H_0 : \mu_1 = \mu_2, H_1 : \mu_1 \neq \mu_2$$

其中  $\mu_1, \mu_2$  分别是吸烟者和不吸烟者群体（总体）的血压指标均值。



## 2、简单统计模型



### 问题2：吸烟对血压的影响模型

#### (3) 模型建立与求解：根据抽样数据，作检验统计量

$$t = \frac{\overline{x_A} - \overline{x_B}}{s_w \sqrt{1/n_A + 1/n_B}} \quad s_w^2 = \frac{(n_A - 1)S_1^2 + (n_B - 1)S_2^2}{n_A + n_B - 2}$$

经计算，第五项指标即夜间收缩压（nSBP）没有拒绝原假设，其余五项的指标即24小时的收缩压（24hSBP）和舒张压（24hDBP）、白天收缩压（dSBP）和舒张压（dDBP）、夜间舒张压（nDBP）都拒绝了原假设。



## 2、简单统计模型



### 问题3：男大学生的身高分布模型

#### (1) 问题提出

现在考虑我国在校大学生中男性的身高分布问题，根据有关统计资料表明，在校男大学生群体的平均身高约为170cm，且该群体中约有99.7%的人身高在150cm至190cm之间。

试问该群体身高的分布情况是怎样的呢？

进一步地将 $[150, 190]$ 等分成20个区间，在每一高度区间上，研究相应人数的分布情况。

特别是中等身高（165cm至175cm之间）的人占该群体的百分比能超过60%吗？



## 2、简单统计模型



### 问题3：男大学生的身高分布模型

#### (2) 问题分析：

一个同类群体的身高分布可近似看作正态分布，根据已知数据，“在校男大学生群体的平均身高约为 170cm”，可确定该分布的均值为  $\mu=170$ 。而由“该群体中约有 99.7% 的人身高在 150cm 至 190cm 之间”和正态分布的“ $3\sigma$  规则”，

$\mu - 3\sigma = 150, \mu + 3\sigma = 190$ ，于是  $\sigma = \frac{20}{3}$ 。其分布密度函数为

$f(x) = \frac{3}{20\sqrt{2\pi}} e^{-\frac{9(x-170)^2}{800}}$ ，身高在任一区间  $[a, b]$  的人数的百分比可利

用积分  $\int_a^b \frac{3}{20\sqrt{2\pi}} e^{-\frac{9(x-170)^2}{800}} dx$  来计算。



## 2、简单统计模型



### 问题3：男大学生的身高分布模型

#### (3) 模型建立：

将[150, 190]等分成20个区间，得到高度区间：

[150,152],[152,154], $\cdots$ , [188,190]

对应的分布：

$$\int_k^{k+2} f(x)dx = \int_k^{k+2} \frac{3}{20\sqrt{2\pi}} e^{-\frac{9(x-170)^2}{800}} dx \quad (k = 150, 152, 154, \cdots, 188)$$

身高在165cm至175cm之间  
的人占该群体的百分比为

$$\int_{165}^{175} \frac{3}{20\sqrt{2\pi}} e^{-\frac{9(x-170)^2}{800}} dx$$



## 2、简单统计模型



### 问题3：男大学生的身高分布模型

#### (4) 模型求解：

虽然，通过变换再查标准正态分布的数值表，可以计算上面积分。但是，要得到各个身高区间上人数的分布情况，显然都用这种方法是很繁杂的。而采用计算机却是轻而易举的事，通过数值积分的基本方法来解决这个问题。

选用数值积分中的复合梯形公式求积方法，可以计算出误差小于0.0001的定积分值，从而可得出相应分布。



## 2、简单统计模型



### 问题3：男大学生的身高分布模型

#### (4) 模型求解：

用数值积分命令：

$$\int_{165}^{175} \frac{3}{20\sqrt{2\pi}} e^{-\frac{9(x-170)^2}{800}} dx \approx 0.5467246173$$

**结论：身高中等（165cm至175cm之间）的大学生约占54.67%，不足60%。如果放宽些，如164cm至176cm之间，则大约有63.2%以上。**



### 3、一元线性回归模型



#### 问题1：一元线性回归方法

如果已知实际检测数据  $(x_i, y_i) (i = 1, 2, \dots, n)$  大致成为一条直线，则变量  $y$  与  $x$  之间的关系大致可以看作是近似的线性关系。

要是研究  $y$  与  $x$  之间的关系，则假定有如下函数关系：

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

其中  $\beta_0$  和  $\beta_1$  是未知待定常数， $\varepsilon$  表示其他随机因素对  $y$  的影响，并且服从于  $N(0, \sigma^2)$  分布。

$y = \beta_0 + \beta_1 x + \varepsilon$  称为一元线性回归模型， $x$  称为回归变量，

$y$  称为响应变量， $\beta_0$  和  $\beta_1$  称为回归系数。



### 3、一元线性回归模型



#### 问题1：一元线性回归方法

要确定一元线性回归模型，就要确定回归系数  $\beta_0$  和  $\beta_1$ 。

参数  $\beta_0$  和  $\beta_1$  的最小二乘法估计：要确定一组  $\beta_0$  和  $\beta_1$  的估计值，使得回归模型与直线方程  $y = \beta_0 + \beta_1 x$  在所有数据点  $(x_i, y_i)(i = 1, 2, \dots, n)$  都比较“接近”。

只要使  $y$  的观察值与估计值偏差的平方和最小，即只需求函数

$$Q = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

的最小值，这种方法叫最小二乘法。





### 3、一元线性回归模型



#### 问题1：一元线性回归方法

对上式分别求对  $\beta_0$  和  $\beta_1$  的偏导数，并令它们等于零，解方程组

求得  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{l_{xy}}{l_{xx}},$

其中  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$

$$l_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), l_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

于是，所求的线性回归方程为  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x .$



### 3、一元线性回归模型



#### 问题2：确定身高与体重关系模型

##### 问题提出：

身体肥胖现已成为人们关注的一个社会问题，过于肥胖一是影响美观，二是可能导致很多影响健康的疾病。肥胖的主要特征是体重过重，那么体重与什么有关呢？当然这与每个人的身高有关。

一般来说，人越高，体重也就要越重，但同样高度的人，体重往往不相同。试通过身高和体重之间的关系，确定人的身体是否肥胖的问题。



### 3、一元线性回归模型



#### 问题2：确定身高与体重关系模型

##### 问题分析与假设：

人的身高与体重之间存在着关系，这种关系是非确定性的关系，即所谓相关关系，因为涉及的变量体重是随机变量，回归分析是研究这种相关关系的一种数学方法。

假设：

- (1) 用  $x$  表示一个人的身高 (cm)，即是一个确定的变量。
- (2) 用  $y$  表示人的体重 (kg)，即是随机变量。
- (3) 对不同的人而言，随机误差  $\varepsilon_i \sim N(0, \sigma^2) (i = 1, 2, \dots, n)$  之间相互独立。
- (4) 收集尽量多的样本，即样本容量尽量的大。



### 3、一元线性回归模型



#### 问题2：确定身高与体重关系模型

模型建立与求解：

假设  $y$  与  $x$  之间具有线性回归关系：

$$y = \beta_0 + \beta_1 x + \varepsilon \quad .$$

为了估计其中的参数  $\beta_0, \beta_1$ ，通过抽样测量，得到了一些人身高和体重相关数据，做最小二乘估计得  $\hat{\beta}_0 = -40, \hat{\beta}_1 = 0.6$ ，将其代入上式得到一元线性回归方程： $\hat{y} = -40 + 0.6x$  .

这个一元回归方程在一定程度上描述了体重与身高的相关关系。给定一个人的身高  $x$  的具体值，就可以算出对应正常的体重值。例如某人的身高 160 cm，代入上式得到体重值为 56 kg，即 56 kg 为身高 160 cm 的人正常体重值。



## 4、参数估计模型



### 问题1：参数估计方法

所谓参数估计就是利用样本的信息来估计总体中的参数。  
参数估计法包括点估计法和区间估计法两种。

#### 1.点估计法

点估计就是构造一个适当的统计量，用它的观察值来估计未知参数。

常用的点估计法有两种：矩估计法和极大似然估计法。

矩估计法:用样本矩作为相应的总体矩的估计量，而以样本矩的连续函数作为相应的总体矩的连续函数的估计量。

极大似然估计法:固定样本观察值，在取值的可能范围内挑选使概率达到最大的参数值作为参数的估计值。即构造似然函数，求使似然函数达到最大的参数值，从而得到参数的估计值。



## 4、参数估计模型



### 问题1：参数估计方法

#### 2. 区间估计法

设总体  $X$  的分布函数含有未知参数  $\theta$ ，对于给定显著水平  $\alpha (0 < \alpha < 1)$ ，若由样本  $X_1, X_2, \dots, X_n$  存在两个统计量  $\hat{\theta}_1(X_1, X_2, \dots, X_n)$  和  $\hat{\theta}_2(X_1, X_2, \dots, X_n)$ ，使得

$$P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = 1 - \alpha$$

则称随机区间  $(\hat{\theta}_1, \hat{\theta}_2)$  是  $\theta$  的置信度为  $1 - \alpha$  的置信区间，

$\hat{\theta}_1, \hat{\theta}_2$  分别称为置信下限和置信上限，

$1 - \alpha$  称为置信度或置信水平。



## 4、参数估计模型



### 问题2：糖果重量的估计模型

在某食品公司生产的一大批糖果中，由于采用自动化的生产线进行糖果的包装，使生产包装出来的袋装糖果重量或多或少地都会有一定的误差。

现在从生产出的这批袋装糖果中随机抽取了 16 袋进行称重，称得重量分别为 506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509, 496（单位：g）。

一般认为自动化生产出的袋装糖果的重量近似地服从正态分布，请通过建模估计分析一下，这批袋装糖果的平均重量范围是多少？



## 4、参数估计模型



### 问题2：糖果重量的估计模型

#### 模型假设：

这是一个随机抽样参数的区间估计问题。

先给出如下假设：

(1) 给定问题的显著水平为  $\alpha = 0.05$ ，即置信度为  $1 - \alpha = 0.95$ 。

(2) 总体服从  $N(\mu, \sigma^2)$ ， $X_1, X_2, \dots, X_n$  为抽样样本。

(3)  $\bar{X}, S^2$  分别为样本的均值和方差。





## 4、参数估计模型



### 问题3：糖果重量的估计模型

模型建立：

假设样本方差  $S^2$  是总体方差  $\sigma^2$  的无偏估计，则有  $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$ ，

并且右边的  $t$ -分布  $t(n-1)$  不依赖任何参数，于是可得

$$P\left\{-t_{\alpha/2}(n-1) < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}(n-1)\right\} = 1 - \alpha$$

$$P\left\{\bar{X} - \frac{S}{\sqrt{n}}t_{\alpha/2}(n-1) < \mu < \bar{X} + \frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right\} = 1 - \alpha$$

即

故总体的均值  $\mu$  的置信度为  $1 - \alpha$  的置信区间为

$$\left(\bar{X} - \frac{S}{\sqrt{n}}t_{\alpha/2}(n-1), \bar{X} + \frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right)$$



## 4、参数估计模型



### 问题3：糖果重量的估计模型

模型求解：

$$1 - \alpha = 0.95, \alpha/2 = 0.025, n - 1 = 15, t_{0.025}(15) = 2.1315$$

$\bar{X} = 503.75, S = 6.2022$ ，总体的均值  $\mu$  的置信度为 0.95 的置信区间为  $(503.75 - \frac{6.2022}{\sqrt{16}} \times 2.1315, 503.75 + \frac{6.2022}{\sqrt{16}} \times 2.1315)$ ，即为 (500.4, 507.1)。

该公司生产的这批袋装糖果重量的均值在 500.4 克与 507.1 克之间，这个估计值的可信程度为 95%。

若以此区间内任一值作为这批袋装糖果的均值  $\mu$  的近似值，其误差不大于  $\frac{6.2022}{\sqrt{16}} \times 2.1315 \times 2 = 6.61$  (g)，其误差估计的可信程度为

95%。



## 4、参数估计模型



### 问题3：糖果重量的估计模型

#### 模型评价与应用：

本模型利用估计精度为 95% 的区间估计，解决了这批袋装糖果的平均重量范围的问题。

本模型也可应用到很多实际问题的估计上，比如一个妇产科医院对新生儿（女）的平均体重、平均身高的估计，一个农民对明年粮食亩产量的估计，一个工厂对一批灯泡的平均寿命的估计等。