1. PVAD =预付年金现值，PVAo =普通年金现值

FVAD =预付年金终值， FVAo =普通年金终值

在计算present value of an **annuity due**时候，在得出PV之后，需要再算一次终值，因为要去的是annuity due：FV=PV\*(1+I/Y)

手册Page 8

1. Harmonic mean调和平均数：是倒数的平均数的倒数，用于计算不同时期以同样金额购买股票的平均成本

三种平均数之间的关系：算数平均数（A）≥几何平均数 (G)≥ 调和平均数（H）

前提条件是一组数据不相等—标准差不等于0。

否则，三者相等.

众数是一组数据中出现最频繁的数据。是唯一可以应用于计量名义数据集中趋势的指标

分位数quantile的计算方法种关于Y值的逻辑：

Quartile: 分成四份；

Quintile: 分成五份；

Decile: 分成十份；

Percentile：分成100份；

例如the third quaitile, 相当于75%，按照公式理解Y=75，因为之后还会除以100。Ly = (n+1)\* (y/100)

在算出Ly 之后（第Y个百分数位所在的位置），比如Ly =6.6，落在6.2和9.5之间，则分位数=6.2+（9.5-6.6）\*0.6=8.2 ，其中0.6是两个数之差的倍数（因为6.6小数是6，所以为0.6倍），不是用6.6-6.2减出来的。

详见手册page 33

1. Coefficient of variation (CV)是用来衡量相对于平均回报率来说，存在多大风险，是一种相对计量，可以对不同的数据集进行离差的比较。

公式： CV = standard deviation of X/ average value of X 因为是计算有多大风险，所以把象征着波动的标准差放在了分子.

比如CV= 1.44，则表示has 1.44 risk per unit of return 详见手册page 44

1. **Sharpe Ratio夏普比率：用来衡量相对于风险来说，有多少平均超额回报率**。**因为是计算有多少回报，所以把‾rp – rf 放在了分子**。

**比如sharpe ratio= 0.144，则表示earn 0.144% of excess return per unit of risk** 详见手册page 46

**夏普比率两个限制条件：**

1. **当两个投资组合的夏普比率都是负值时，不一定更高的夏普比率有更好的风险调整能力；**
2. **当标准差适合计算风险时，夏普比率才有用。如果回报曲线是非对称分布，例如大概率收益偏小，小概率损失严重，此时标准差会低估风险，从而使得夏普比率更高；**

**对比Sharpe ratio & SF Ratio:**

**Sharpe ratio是计算回报率的，常用%表示；**

**SF Ratio是计算有多少个标准差在平均值以下的，是要确保短缺风险发生的概率最小化；**

**选择最优的投资组合往往采用最高的SF Ratio，因为能够最大概率保证投资组合的收益率在能够接受的最低风险水平之上；**

**Sharpe = [E(Rp)- Rf]/σp;**

**SF Ratio= [E(Rp)-RL]/ σp; SF Ratio算出的值是Z值，采用和Z值一样的看表来算出具体数值的方法；比如计算the probability that portfolio fall short of the target amount时，算出的SF Ratio=0.5，则要看表计算-0.5 的SF Ratio或者Z值所对应的概率，因为题目要求的是fall short of the target amount;**

**当Rf = RL时，sharpe ratio 就是SF Ratio了；很多时候题目没给Rf，则用RL**

安全第一法则：计算最优投资组合也就意味着发生安全比率最高或者发生短缺风险概率最低（=1-安全比率），详见手册Page80例题

1. Kurtosis峰度：衡量的是与正态分布相比，该分布是更加尖锐还是更加扁平。正态分布的峰度等于3，而超额峰度等于0.如果峰度大于3，称为尖峰；如果小于3，称为扁峰。
2. Positive skewed & negative skewed 区别：详见手册48
3. 正态分布可正可负，因此常描述回报率；标准正态分布可正不可负，因此常描述价格
4. Time-series: 用于同一事物在不同时点的比较；

Cross-sectional data截面数据：用于同一时点上不同事物的比较；

1. 使用样本对总体参数进行试算，估算值应该具有一下3点特征：

无偏性；

效率性；

一致性；

1. Z-test: normal distribution+ known variance 用来和给定的critical z-value做对比来决定是否reject H0 ; 如果要证明大于或者小于一个值，则用one tailed test；

计算步骤：

1. Specify the null and alternative hypotheses:

Two tails: H0 : μ0 =0 VS Ha : μ0 ≠ 0

1. 确定level of significance
2. **对比的本质就是计算出来的Z-statistic值和可靠因子Z-values的对比：比如5% level of significance, Z-values for two tails are ±1.96， 如果算出的Z-statistic值不在这个区间，则reject the null hypothesis;**
3. 计算standard error of the mean
4. 计算出Z-statistic值

**零假设和备择假设：**

**零假设：**It is considered to be true unless the sample provides evidence showing it is false.

上面是背景讲述。为了假设检验的原理是反证法。这种方法的逻辑，不是搜集大量的证据来正面支持自己的论点，这样做太累、更主要是因为百密难免一疏；反证法走的是捷径，给自己的观点树立一个非此即彼的反面论断，然后找到证据来推翻它，从而间接证明自己的论点。研究者把自己支持的论点放在备择假设（H1），而把他想推翻的、与备选假设对立的非此即彼的观点放在原假设（H0）

**另一种判断是否reject hypothesis 是计算population mean μ 的范围是否涵盖 μ=0。相对应的公式 sample mean- Zα/2 \* standard error ≤ μ ≤ sample mean + Zα/2 \* standard error 详见note page 273例题**

**P-value: the smallest level of significance for which the null hypothesis can be rejected**

**比如P-value = 7%, 则level of significance = 10% can be rejected, but level of significance = 5% 则不可以reject;**

**Level of significance is the probability of making type I error which reject null hypothesis when it is true**

Level of significance 显著性水平α：是指当假设为正确时人们却把它拒绝了的概率或风险，必须在每一次统计检验之前确定，就是TYPE I error

Degree of confidence 置信度 1-α

**Power of test: is the probability of correctly rejecting the null hypothesis when it is false; 就是该拒的给拒了= 1- Type II error**

T-test: normal distribution + unknown variance (sample variance known) 注意n-1 和df值两者来找出t-value. 用来和critical t-value做对比来决定是否reject H0 ;

Chi-square-test: 一个样本的方差与population的方差的比较

F-test: 两组正态分布数据的方差是否相等

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 检验 | One tail or Two tails | Variance known? | Population required normal distribution? | Sample independent? |
| t-test | Two tails | N | Y | sample VS population |
| z-test | Two tails | Y | Y | sample VS population |
| Chi-square | Two tails | Y | Y | sample VS population |
| f-test | One tail | Y and 两个population 方差假设相等 | Y | Population VS population |
| Paired difference | Two tails | Y and mean of differences between 2 samples are different | Y | Sample VS sample |

公式： μ0是假设的population mean

1. 一个population的时候：tn-1 = (‾x- μ0) / (s/√n)；
2. 两个population且variances都相等的时候：t= ‾(x1 -‾x2)/ (sp1/n1 + sp2/n2)1/2

两个population且variances不相等的时候：t= ‾(x1 -‾x2)/ (s12/n1 + s22/n2)1/2

1. z-statistic = (‾x- μ0) / (σ/√n)
2. Chi-square: Xn-12 = (n-1)\*s2 / σ02 ,其中σ02 是hypothesized value of population variance 使用情况是在一个sample和一个population都有方差或标准差时候；
3. F-test: F=S12/S22 , 两个变量是sample的方差；使用情况是在有两个samples都知道方差或标准差时候；但是必须把较大的方差放在分子；
4. Paired comparisons配对比较检验: 条件是two depend sample, H0: μd =μdz, VS μd ≠ μdz, μd is mean of the population of paired difference

μdz is hypothesized mean of paired difference, 通常是零

t-statistic = (‾d- μ0) / (Sσ¯/√n) d¯: two portfolios’ mean difference, μ0: 零假设 mean difference, Sσ¯: two portfolios’ standard deviation difference

详见易错题真题-71

**当samples are independent, 可以计算differences in means，比如测试5个人的平均身高；**

**当samples are dependent, 可以计算mean of differences, 也就是 paired comparisons test。比如测试饭前饭后的体重；**

**计算t-test 中的平均值的差异differences in means 还是差异的平均值 mean of differences,还是计算z-test，思路都是如下：**

1. **确定One tail or two tails；**
2. **根据level of significance 和df, 确定可以reject的区间，也就是1- confidence interval;**
3. **根据题目中给出的条件计算出对应的t-test值或z-test值；**
4. **判断t或z值是否符合reject 条件；**

Parametric tests: 关于参数的假设检验，例如t-test, z-test, f-test等；

Non parametric tests: 不是关于参数的假设检验；

Spearman rank correlation test 斯皮尔曼排序相关系数：比如20个共同基金的两年排名。如果spearman rank correlation为large positive, 则表明第一年排名靠前的，第二年大概率也排名靠前，相反，如果correlation为large negative, 则表明第一年排名靠前的第二年排名靠后；

**2nd round 复习：**

1. Nominal risk-free rate = real risk-free rate + expected inflation rate **note page 57**
2. EAR (effective annual rate) = (1+ periodic rate)m -1; EAR代表实际年利率，是在将计息频率不是一年一次的利率转化成一年计息一次的利率

所以，EAR 8% compounded annually is not equal to EAR 8% compounded semiannually;

结论是the greater the compounding frequency, the greater the EAR will be.

注意算daily compounding的简便公式 e 0.06-1 = 6.1837%， 其中e = 0.06 **note page 58**

1. Ordinary annuities 普通年金: cash flow occurs at the end of each compounding period; **the first payment is at t= 1;**

Annuities due 即付年金: cash flow occurs at the beginning of each period**; the first payment is today at t=0;**

1. 计算普通年金从第三年开始，每年付款100，那么现值是多少？

思路：先画出现金流线图，由于是第三年末开始付款，首先确定第二年末的PV，之后算出起始的PV note page 63

1. 计算即付年金FV的方法有两种：一是把计算器调到【BGN】通过按【2nd】【BGN】【2nd】【SET】来实现；另一种方法是算出普通年金的FVO，再乘以 1+ I/Y，从而算出FVD , 因为即付年金比普通年金多算了一次利息，所以普通年金需要乘以 1+ I/Y来和即付年金相等 note page 65
2. 永续年金的PV的计算：注意永续年金的PV值是下一次付款之前的值，比如第四年付dividend，则对应的PV是第三年末的PV 参考例题note page 70
3. 对于不规则/不相等cash flow，计算FV只需要把各个时间点的FV都加起来即可；

计算PV有两种方式：一是可以使用计算器【CF】【2nd】【CLR WORK】来计算详见note page 71--72，二是用每次的现金流除以对应期的 (1+I/Y)m----这个方法适用于当cash flow不是一开始就有的时候，比如 note page 91 第一题

1. **NPV decision rule:**

* Accept projects with a positive NPV. Positive NPV will increase shareholder wealth;
* Reject projects with a negative NPV. Negative NPV will decrease shareholder wealth;
* When two projects are mutually exclusive (only one can be accepted), the project with the higher positive NPV should be accepted;

**IRR decision rule:**

* If IRR > firm’s required rate of return, accept projects;
* If IRR < firm’s required rate of return, reject projects;

1. **Money-weighted rate of return试用于IRR**，是指内部收益率。

**核心的逻辑是PV inflow = PV outflow , 之后确定每一期的Cash Flow，IRR具体算法也是通过【CF】【2nd】【CLR WORK】来求得；**

**或者通过下面等式算出：**

**Outlay amount A + Outlay amount B / (1 + IRR) = Income amount C / (1+ IRR)2**

**Time-weighted rate of return适用于HPR**, Time-weighted rate of return计算步骤：

1. 列出holding period beginning value/dividends paid/ ending value
2. **利用 计算出不同period 的HPR**
3. 利用公式 （1+Time-weighted rate of return）2  =HPR1 \* HPR2算出return

衡量基金经理的performance 通常用时间加权；

时间加权不受现金流影响

1. Bank discount yield: 银行贴现收益率，是指假设360天，且以单利按票面价值计算而不是按某一特定价格计算的每1美元投资的年利息。

**注意两个特征： 1. 分母是face value，而不是Market value;**

**2. 计算基础是360天 （假设为单利，而不是复利，所以用乘以360/t而不是次方）**

公式： rBD= (D/F)\*(360/t) D: difference--- face value-purchase price

HPY-holding period yield

**HPY= (P1-P0+D1)/P0 = (P1+D1)/P0 -1** 其中P0是初始价格，P1是到期价格，D1是dividend；

EAY (effective annual yield) ：实际年收益率，是复利收益率，所以计算基础是365天

公式： EAY=(1+HPY)365/t -1，**注意也可以通过EAY反推导出HPY，HPY=(1+EAY)t/365 -1**

**也就是说EAY是365天年化的HPY；因为是复利，所以用365/t次方；**

Money market yield货币市场收益，可以理解为是360天年化的HPY；

rMM = HPY \* (360/t) = (360 \* rBD)/ [360-(t\*rBD)]

**后面的公式很难记，可以通过推导来完成BDY—转化成actual discount—算出初始价格或者叫做购买价格—算出HPY—转化成360天年化的rMM 详见note page 107例题包含推导过程；**

**Money market yield: 是360天年化的HPY**

**Bond-equivalent yield 债券等值收益率：是365天年化的HPY，以每半年为标准换算利息；**

**它是单利，而EAY是复利，两者都是实际年收益率；**

**BEY= HPR \* (365/days to maturity)**

**从3个月的BEY转化为半年的BEY= (1+HPY)2 -1;**

**从一年的BEY转化为半年的BEY=(1+HPY)0.5 -1;**

**Effective annualized cost of bank acceptance = interest / net**

1. When describing investments, measures of central tendency provide an indication of an investment’s expected return;

Measures of dispersion indicate the riskiness of an investment;

1. Nominal scales名义测定：对数据进行定性分类，但不排序，比如性别；

Ordinal scales 定序测定：对数据进行分类并排序，但这种排序不能测定各等级之间间距的大小；

Internal scale区间测定：可以测定出间距的大小，比如温度；但缺点是不能比较，比如20度并不是比10度热一倍；

Ratio scales 比例测定：还可以进一步测定间距比例的大小，比如价格，4元比2元贵一倍；

1. Mean absolute deviation (MAD)是标准差的绝对值的平均数，表示偏离平均数的数值；例如一组数据的平均数是50，MAD是10%，代表整体分布的数据平均下来，每个数据偏离平均数50的偏离度是正负10%；

MAD = (∑x¯ -X) /n 算术平均数的标准差的绝对值

**Population standard deviation: 是衡量总体数据的波动性**，population variance 是为了计算总体标准差而存在的，最终会归结到总体标准差上；

同理，sample standard deviation 和Sample variance也是一个逻辑，只不过涉及到计算用sample来计算时，要除以n-1，而不是n，因为本身sample所体现的方差或者标准差都比总体的方差或者标准差要小，**所以通过除以n-1来增加sample的方差或者标准差； 无论是population还是sample的标准差都是先除以n或n-1，再开平方**

**Chebyshev’s inequality 切比雪夫不等式**：是计算落在某一标准差区间的概率；

**公式=1-1/k2, K代表标准差的数值；**

1. Sample skewness 和 sample kurtosis都是描述数据分布形态的，一个是左右偏，一个是尖扁峰；

Sample skewness =0.5时，表明significant levels of skewness;

Sample kurtosis > 1.0时，表明excess kurtosis value is large;

1. Odds赔率=发生的概率/ 不发生的概率，假设=1/3，表述为赔率是1-to-3，**发生的概率在前；**
2. 概率乘法法则multiplication rule of probability:

P(AB)= P(A I B) \* P(B)，中间竖线表示”given”，考虑到B发生的条件下，A发生的概率；

概率加法法则 addition rule of probability:

P(A or B)= P(A) + P(B) -P(AB)

全概率法则 total probability rule:

总结下：乘法法则适用于非独立事件；

加法法则适用于A 或B;

全概率法则适用于互斥并遍历事件；

Bayes’ formula是全概率法则的延伸，使用事前概率来推算事后概率: updated probability= (probability of new information for a given event/ unconditional probability of new information) \* prior probability of event

简单来说，**假设P(I) 是prior probability**—确定这步非常关键，则需要计算P(I⏐O)时的概率公式为：

P(I⏐O)= [P(O⏐I)/ P(O)] \* P(I), 或者P(I⏐O) \* P(O)= P(O⏐I) \* P(I)

假设有fail, pass, survivor, no survivor

P(fail) = P (F ⏐ S) \* P (S) + P (F ⏐ N) \* P (N)， 这个公式适用于求其中某个变量的；

1. Expected value期望值：是指随机变量取值的数学平均值或加权平均值，用E表示；

E(X)= ∑ P(xi)\*xi= P(X1)\*X1 + P(X2)\*X2 + ….+P(Xn)\*Xn

当各个概率都相等时，expected value就是数学平均数，否则是加权平均数；

σ2 = P(X1)\*(X1-Ex)2 +….+P(Xn)\*(Xn-Ex)2 ，这时，Ex就是相当于之前章节中的平均数；详见note page 169例题；

同样，**对于两个随机变量的协方差Cov(Ra, Rb)= ∑P(S)\* [Ra-E(Ra)] \* [Rb-E(Rb)]，其中P代表概率，R代表收益率，详见note page 172例题**

1. Correlation coefficient 相关系数：表示两个变量之间的线性关系，介于-1到+1之间。

-1表示完全负相关，0表示不相关，+1表示完全正相关；

**Corr (Ri, Rj)= Cov (Ri, Rj) / [σ(Ri) \* σ(Rj)] ，Corr (Ri, Rj)也可以用ρ(Ri, Rj)表示；**

**Cov (Ra, Ra)=Var(Ra);**

**Coefficient of variation of portfolio = portfolio return的标准差/ portfolio return的平均数**

**一个group 数组的方差variance = [∑(Xi-x )2]/(n-1)**

1. **Portfolio variance: 投资组合中两个变量A, B的方差：**

**Var(Rp)= WA2 σ2 (RA) + WB2 σ2 (RB) + 2WAWBCov (RA, RB) 或者**

**Var(Rp)= WA2 σ2 (RA) + WB2 σ2 (RB) + 2WAWBσ(RA)σ(RB)ρ(RA, RB);**

**其中W代表所占比例，σ代表标准差，ρ代表相关系数，R代表回报率；**

同样，投资组合中有三个变量A, B, C的方差：

Var(Rp)= WA2 σ2 (RA) + WB2 σ2 (RB) + WC2 σ2 (RC) +2WAWBCov (RA, RB)+ 2WAWCCov (RA, RC)+ 2WBWCCov (RB, RC);

Note page 176—178例题解题思路：

第一步：确定两个assets各自所占比例，以便为接下来算方差所用；

第二步：先抛开概率，如果只是算Asset A的平均回报率，只需要把可能各自的回报率加在一起除以N即可，所以回到这里，只不过变成各自回报率的概率，所以各自回报率乘以相对应可能发生的概率即为期望值expected value。因为最后要算出整个portfolio的expected value，所以再乘以各自所占portfolio的比例即可：E(P)=WAE(RA) + WBE(RB)；

第三步：再算整体portfolio的方差之前，需要先算出asset A 和B的各自方差和协方差。

同理，如果抛开概率，一般算方差的方法是Asset A的各自值减去asset A的平均数的结果的平方求和，这里涉及概率，所以各自再乘以对应的概率；

协方差只能根据记忆协方差公式来算了；

1. Labeling贴标签 & permutation 排序：

nCr = n! / [(n-r)! \* r!], 这个公式**不考虑**组合内排序的问题；从n 个数中选出r个作为一个组合；

nPr = n! / (n-r)!, 这个公式**考虑**组合内排序的问题；从n个数中选出r个并排序作为一个组合；

1. Binomial distribution 二项式分布：

Binominal random variable 二项式随机变量是指当试验只有成功和失败时，成功的次数；如果实验为一次，则为伯努利实验，即非成功即失败；

P(x) = n! / [(n-x)! \*x!] \* Px \* (1-P)n-x

**如果每次试验成功的概率P(x)=P , 则试验了n次时的方差：**

**Variance of x = np \* (1-p)**

1. Tracking error追踪误差：the difference between the total return on a portfolio and the total return on the benchmark;

The role of correlation in the Multivariate normal distribution多变量分布的相关性角色：

1. Number of correlations= 0.5n\*(n-1), 比如有4个变量，那么0.5\*4\*(4-1)=6，相关系数就是6；
2. 对于asset的portfolio，it is desirable to combine assets having low returns correlation，这样能有效规避风险；
3. 对数正态分布lognormal distribution:

由于对数的特点，数值是大于等于0，所以应用在asset price上，这样asset price不会出现负值甚至-100%；

**公式effective annual rate 或者Holding period return= eRcc -1，其中Rcc是continuous compounding return;**

**同样，反过来可以通过EAR或HPR来得到Rcc: Rcc= Ln (1+EAR);**

**通过相对价格来计算：Ln (S1/S0)= Ln (1+HPR)=Rcc，逻辑是S1/S0 =1+HPR：比如一只股票年初100元，年末120元，则S1/S0=1.2, HPR=20%;**

**HPRT = eRcc \* T -1，也就是说假设**

**Rcc=10%, the HPR over two years is e(0.10)\*2 -1=22.14%;**

**反过来，假设股票price from $1000 to $1221.4 over 2 years, the 2-year continuously compounded rate is Ln(1221.40/ 1000)= 20%, and the annual continuously compound rate Rcc is 20%/2=10%;**

1. Monte carlo simulation用于金融风险管理，复杂证券估值，它提供统计的估计，而不是具体结果和分析方法。
2. 连续型均匀分布的概率函数f(x)= 1/(b-a), 指a到b区间占整体的概率；
3. T-distribution: 利用小样本来估计呈正态分布且方差未知的总体的均值；

* 样本数量N<30；
* 自由度越大，越靠近呈现正态分布，置信区间越小；
* 自由度是可以自由变化的变量数量；
* df= n -1，n是sample 数量;
* 自由度的数量是基于观察值的数量，也就是说观察值的数量越多，越接近分布的中心（粗暴的理解为事实的真相）

1. **Z-statistic & t-distribution:** Z-value表示距离平均值有几个标准差的距离：

* 当总体呈正态分布且**总体方差已知**情况下，用Z-statistic; n>30;
* 当总体呈正态分布且**总体方差未知**情况下，用t-distribution; n<30;
* **当总体呈非正态分布且总体方差已知的情况下**，只要sample数量n>30，就可以用**z-statistic;**
* **当总体呈非正态分布且总体方差未知的情况下**，只要sample数量n>30，就可以用**t-statistic;**

**所以，当总体呈非正态分布时，如果n<30, 我们不能创建confidence interval;**

* 当方差未知，还想用z-test，条件是sample 足够大，而且适用于中心极限理论-central limit theorem

1. 置信区间 confidence of interval

当不知道总体的标准差和均值时，可以通过sample的均值‾X 和s来计算confidence interval:

* 90% confidence interval for X is ‾X- 1.65s to ‾X+ 1.65s;
* 95% confidence interval for X is ‾X- 1.96s to ‾X+ 1.96s;
* 99% confidence interval for X is ‾X- 2.58s to ‾X+ 2.58s;

**1.65, 1.96, 2.58 可以理解为t值，但是还会有更多的t值需要结合自由度在书后面的表里去查；**

当知道总体标准差时，confidence interval= ‾x ± Zα/2 \*σ/√n , ‾x是sample mean, Zα/2 是可靠因子，是在α/2的概率所对应的Z值；σ/√n 是标准误差；

重点：计算T值和Z值都用上面的公式，区别在于：

对于Z值，比如95% 置信区间， Zα/2 =1.96，以此类推其他情形；

对于T值，比如95% 置信区间，就要看有多少个sample了，df =n-1，通过给的置信区间和df来锁定可靠因子Zα/2 ；

* Zα/2 =1.645 for 90% confidence intervals (the significance level is 10%, 5% in each tail);
* Zα/2 =1.960 for 95% confidence intervals (the significance level is 5%, 2.5% in each tail);
* Zα/2 =2.575 for 99% confidence intervals (the significance level is 1%, 0.5% in each tail);

总结计算总体均值population mean的步骤：

1. 确定自由度df=n-1；
2. 根据One tail (use α) or two tails (use α/2)，确定α；
3. 确定Zα/2；
4. Line charts: show closing prices for each period;

Bar charts: displayed as a vertical line with the closing price indicated as a point of dash on the right side of the line;

Candlestick charts 蜡烛图: can make patterns easier to recognize;

Point and figure charts: helpful in identifying changes in the direction of price movements;

1. 绝对频率 VS 相对频率 VS 累计相对频率

绝对频率：是在一个区间出现的次数；

相对频率：是绝对频率所占所有数值出现次数的百分比；

累计绝对频率：是在区间上限以下所有出现的概率相加；

1. Combination formula: 组合公式

Multinomial formula：多项式公式

Permutation formula：排列公式

1. consistent estimator VS unbiased estimator VS efficient estimator

无偏性说明某种方法求出的估计量在真实值的周围波动，对于两个估计量来说，方差小的那个说明此种方法的估计量更集中于真实值附近，说明有效性较大，一致性原理说明样本容量越大，求出的估计量越靠近实际值。

1. Df 自由度越高，尾巴越瘦，反之尾巴越胖；
2. **Data-mining bias:** use out-of-sample data

Sample selection bias: occurs when some data is systematically excluded from the analysis, usually because of the lack of availability;

**Survivor-ship bias: is the most common form of sample selection bias**. **最典型的例子是mutual funds**, which normally only include the “survivors” but not include funds that have exit due to closure or merger;

**Look-ahead bias:** data is not available during a specific period, like 30 to 60 days after fiscal year ends;

**Time-period bias**: if the time period over which the data is gathered is either too short or too long;

1. 风险偏好的类型，详见手册153