

Quantitative Methods

目录

Quantitative Methods.....	1
Time Value Calculation.....	2
Reading 6: The Value of Money	2
Reading 7: Discounted Cash Flow Applications.....	5
Probability & Descriptive Statistics	7
Reading 8: Statistical Concepts and Market Returns	7
Reading 9: Probability Concepts	10
Reading 10: Common Probability Distributions	13
Inferential Statistics.....	16
Reading 11: Sampling and Estimation.....	16
Reading 12: Hypothesis Testing.....	19
Technical Analysis.....	21
Reading 13: Technical Analysis	21

Time Value Calculation

Reading 6: The Value of Money

各种利率

- 1 Required rate of return: 要求回报率
- 2 Discount rate: 折现率
- 3 Opportunity of cost: 机会成本
- 4 Nominal risk-free rate = real risk-free rate + expected inflation rate
- 5 Required interest rate on a security = nominal risk-free rate + default risk premium + liquidity risk premium + maturity risk premium(期限风险)
- 6 单利一年按照 360 天算, 复利按照一年 365 天算

EAR Effective annually Rate

- 1 $EAR = (1 + \text{period rate})^m - 1 = EAR = (1 + \frac{r}{m})^m - 1$
 - 1.1 Periodic rate 值得是一年内的年化利率, 如果年复利, 一年计息一次, $M=1$, 如果是半年复利, 一年计息两次, $M=2$, 以此类推
 - 1.2 如果是连续复利: $EAR = e^r - 1$
- 2 FV & PV
 - 2.1 FV: future value 终值 PV: present value 现值
 - 2.1.1 投资一年: $FV = PV \left(1 + \frac{r}{m}\right)^m$
 - 2.1.2 投资 n 年: $FV = PV \left(1 + \frac{r}{m}\right)^{mn}$
 - 2.1.3 投资 n 年, 连续复利: $FV = PV \times e^{n \cdot r}$

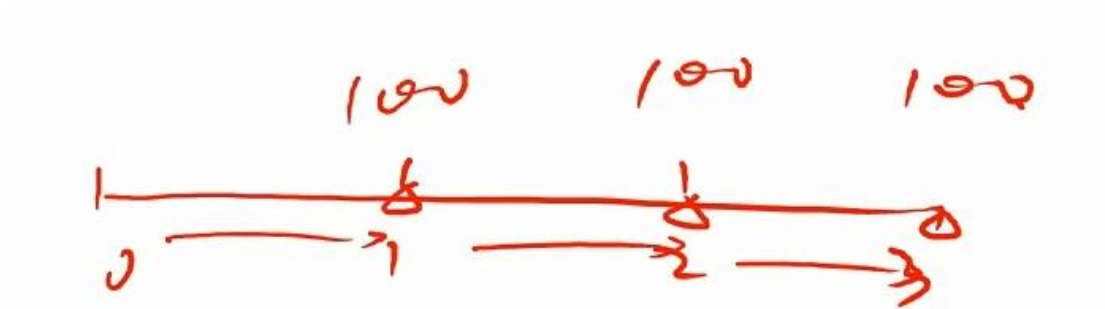
Annuities

- 1 定义: 一组特殊的现金流, 满足一下三个条件
 - 1.1 Equal intervals 等时间间隔
 - 1.2 Equal amount of cash flow
 - 1.3 Same direction
- 2 年金的构成:
 - 2.1 N=number of periods
 - 2.2 $1/Y$ =interest rate per period 注意, 在计算器输入时省略百分号
 - 2.3 PV=present values
 - 2.4 PMT=amount of each periodic payment

2.5 FV=future value

3 现付年金和后付年金

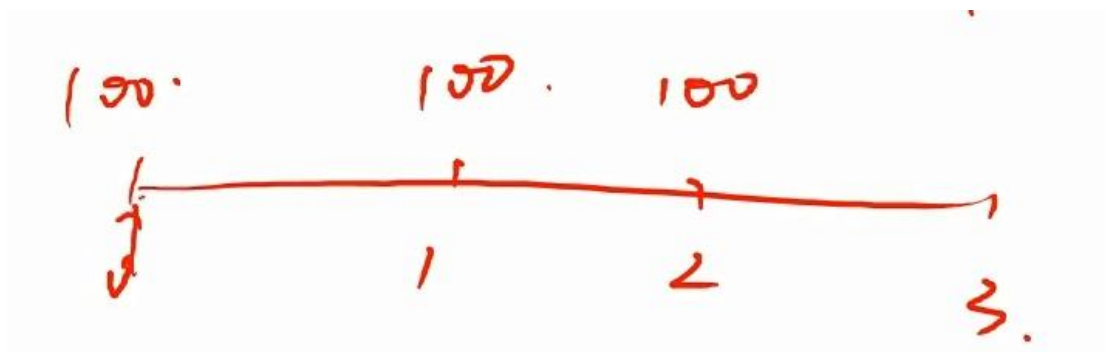
3.1 Ordinary annuities: 后付年金



$$FV = 100 \times (1 + 10\%)^2 + 100 \times (1 + 10\%) + 100$$

$$PV = \frac{100}{1 + 10\%} + \frac{100}{(1 + 10\%)^2} + \frac{100}{(1 + 10\%)^3}$$

3.2 Annuities due: 先付年金



$$FV = 100 \times (1 + 10\%)^3 + 100 \times (1 + 10\%)^2 + 100 \times (1 + 10\%)$$

$$PV = \frac{100}{1 + 10\%} + \frac{100}{(1 + 10\%)^2} + 100$$

3.3 计算器默认是后付年金，换成先付年金可设置为 BGN 模式

3.4 两种典型计算问题

3.4.1 Amortization (还房贷和摊销问题)

3.4.1.1 用计算器计算摊销表格

3.4.1.2 计算出 PMT 之后，按 2nd，按 pv (AMORT)，P1 表示开始时期，P2 表示结束时期，P1=1，P2=1 表示查看第一期，P1=2，P2=2 表示查看第二期，P1=1，P2=2 表示查看第一期加第二期。

3.4.1.3 INT: 本地归还利息

3.4.1.4 PRN: 本地归还本金

3.4.1.5 BAL: 期末剩余应还

3.4.1.6 Demo

- To see how a lump sum can generate an annuity, assume that we loan \$3,170 from the bank today at 10 percent interest. Construct an **amortization** table to show the annuity payments over the next four years.

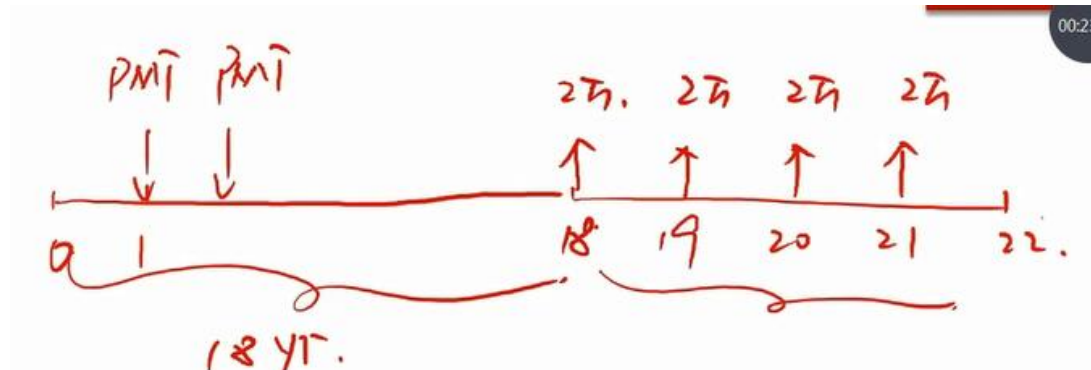
➤ **Correct Answer:**

The amount of the annuity payments : $N=4$; $I/Y=10$; $PV=-\$3,170$; $FV=0$;
CPT: $PMT=\$1,000$

Amortization Table					
Time Period	Beginning Balance (1)	Payment (2)	Interest Component (3)=(1)*10%	Principal Component (4)=(2)-(3)	Ending Balance (5)=(1)-(4)
1	3,170	1000	317	683	2,487
2	2,487	1000	248.7	751.3	1,735.7
3	1,735.7	1000	173.57	826.43	909.27
4	909.27	1000	90.93	909.27	0

3.4.2 养老金，教育金问题

3.4.2.1 18年后上大学，每年学费 20000，存 18 年，利息 5%，求 PMT



3.4.2.2 关键：第一阶段的 FV_1 =第二阶段的 PV_2

3.4.2.3 求 PV_2 : BGN-> $N=4$, $I/Y=5$, $PMT=20000$, $FV=0$, CPT: $PV=74465$,
END -> $N=18$, $I/Y=5$, $PV=0$, $FV=74465$, CPT: $PMT=2647$

3.5 Perpetuity 永续年金

3.5.1 定义：a perpetuity is a set of level never-ending sequential cash flows, with the first cash flow occurring one period from now

3.5.2 永续年金是后付的 无 FV ，只有 PV

3.5.3 $PMT=A$, interest rate= r . $PV = \frac{A}{1+r} + \frac{A}{(1+r)^2} + \frac{A}{(1+r)^3} + \dots$

3.5.4 $PV = \frac{A}{r}$

Reading 7: Discounted Cash Flow Applications

NPV & IRR

- 1 NPV: net present value 净现值 把未来所有现金流折算到现在

$$1.1 \quad NPV = CF_0 + \frac{CF_1}{(1+r)^1} + \frac{CF_2}{(1+r)^2} + \dots + \frac{CF_n}{(1+r)^n} = \sum_{t=0}^N \frac{CF_t}{(1+r)^t}$$

- 2 IRR: internal rate of return 内部收益率, NPV 为 0 时刻的利率 r

- 3 项目决策:

- 3.1 单一项目:

3.1.1 用 NPV 判断, NPV 大于 0 可投

3.1.2 用 IRR 判断, IRR 大于 r 可投, r 是要求回报率

- 3.2 多个项目:

3.2.1 Independent Projects 独立项目, 满足的都投

3.2.2 Mutually Exclusive Projects

3.2.2.1 用 NPV 判断, 选 NPV 高的

3.2.2.2 用 IRR 判断, 选 IRR 高的

3.2.2.3 一般两个指标同时符合的, 如果违背了, 使用 NPV 判断

Yield calculation

- 1 HPR: holding period return 持有期收益率, 注意不是年化收益率

$$1.1 \quad HPR = \frac{P_1 - P_0 + CF_1}{P_0}$$

1.2 P_1 : 期末价格 | P_0 : 期初价格 | CF_1 : income 例如股票分红

1.3 如果算 t-bill 的 hpr 的话 $HPR = \frac{F - P}{P}$

- 2 rBD: bank discount rate/bank discount yield 折扣率

2.1 T-bill 零息债券 face value 面值

$$2.2 \quad rBD = \frac{F - P}{F} \times \frac{360}{t} \quad \text{单利年化 (一年 360 天)}$$

2.3 看的是在面值的基础上折了多少, 分母是面值

2.4 对于 t-bill, 已知 t-bill 的 hpr, 可以算出他的 rBD

- 3 HPR 的年化计算

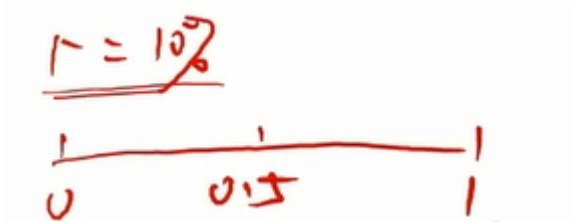
3.1 单利下的年化: $HPR \text{ 年化} = rMM = HPR \times \frac{360}{t}$ rMM : money market yield

3.2 复利下的年化: $EAR = (1 + \frac{r}{m})^m - 1$, 其中 r/m 就是他的期间收益率, 也就是等于 HPR, m 是他计息的次数, 复利下按照 365 天计算, 所以 $m=365/t$ 。最后得出:

$$EAR = (1 + HPR)^{\frac{365}{t}} - 1$$

- 4 BEY: bond equivalent yield 债券等价收益率

4.1 含义：一年计息两次的名义收益率 nominal interest rate, $m=2$



4.2 图中的 r 就是 BEY

4.3
$$EAR = (1 + \frac{BEY}{2})^2 - 1$$

4.4 中国债券一年计息一次，名义利率 10.1%，美国债券一年计息两次，名义利率 10% (BEY)

5 例题：175 天的 T-bill 的 EAR 是 3.8%，求 rBD

5.1 求 $rBD = \frac{F-P}{F} \times \frac{365}{t}$

5.2 已知 $EAR = (1 + HPR)^{\frac{365}{175}} - 1 = 3.8\%$ $HPR = \frac{F-P}{F}$

5.3 可求出 F 和 P 的线性关系

5.4 题型总结：

5.4.1 已知 rMM 或者 EAR，求 rBD，或者反过来

5.4.2 以 HPR 为跳板，因为 rMM 合格 EAR 都是 HPR 的年化形式

5.4.3 对于 t-bill 的资产，没有 income 现金流，所以 rBD 和 HPR 都是 F 和 P 的关系式，可互相求解

Money-weighted return & Time-weighted return

1 Time-weighted rate of return

1.1 年平均收益率，集合收益率

1.2 先求每期的 HPR，然后相乘开根号

1.3
$$TWRR = \sqrt[4]{(1 + 10\%)(1 - 2\%)(1 + 5\%)(1 + 8\%)} - 1$$

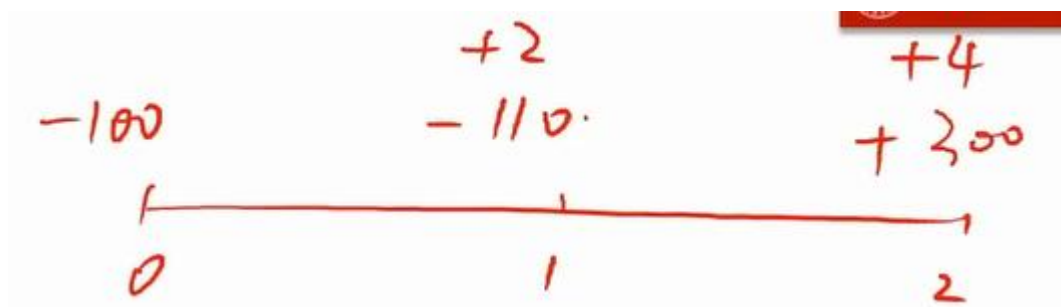
1.4 投资期限大于 1 才开根号，正好等于一年不开根号

2 Money-weighted return

2.1 本质是 IRR

3 注意 MWRR 收 CF 的影响，TWRR 不收 CF 的影响，业绩判断一般用 TWRR

4 例题：2015 年初买入股票，价格 100, 2016 年初买入股票，价格 110, 2017 年初卖出股票，价格 150，每年年末分红 2.



计算净现金流: (-100)-----(-108)-----(-304)

MWRR = IRR,使用金融计算器计算: CF0=-100,C01=-108,C02=304 CPT:IRR=28.52%

TWRR:先计算每一期的 HPR

$$HPR1=(110-100+2)/100$$

$$HPR2=(150*2-110*2+2)/110*2$$

Probability & Descriptive Statistics

Reading 8:Statistical Concepts and Market Returns

Types of measurement scales

1 四中分类

1.1 Nominal scales 类别

1.2 Ordinal scales 排序, 有大小, 不能加减 <>

1.3 Interval scales 排序+间距 eg: 温度 可加减 <>+ -

1.4 Ratio scales 什么都可以 ><+ - * /

2 四种均值计算

The Arithmetic Mean:

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{n}$$

The Weighted Mean:

$$\bar{X}_w = \sum_{i=1}^n w_i X_i = (w_1 X_1 + w_2 X_2 + \dots + w_n X_n)$$

The Geometric Mean:

$$G = \sqrt[n]{X_1 X_2 X_3 \dots X_N} = \left(\prod_{i=1}^N X_i \right)^{1/N}$$

The Harmonic Mean:

$$\bar{X}_H = \frac{n}{\sum_{i=1}^n (1/X_i)}$$

Harmonic Mean <= Geometric Mean <= Arithmetic Mean

3 如何选择计算方法:

3.1 如果关注的是 ending value 终值: 使用几何收益率

3.2 预测未来时用算术或者加权平均数收益率

3.3 如果是衡量历史业绩, 使用几何收益率

Quantile 分位数

1. Quartile 四分位数 Quintile 五分位数
2. 先排序，再均分
3. 11 个数，求第三个四分位数：(11+1) * 75%，注意要加一，如果求出来不是整数，怎么办呢：eg: 求出来是 8.25 第八个数+第九个数*0.25

Measures of dispersion (absolute dispersion)

1. Range: 范围 max-min
2. MAD: 各值到均值的距离的平均值 $MAD = \frac{\sum_{i=1}^N |X_i - \bar{X}|}{n}$

A handwritten table comparing sample and population statistics. The columns are labeled '样本' (Sample) and '总体' (Population). The rows are labeled 'Mean', 'Var.', and 'S.D.'. The sample statistics are \bar{X} , S^2 , and S . The population statistics are μ , σ^2 , and σ .

	样本	总体
Mean	\bar{X}	μ
Var.	S^2	σ^2
S.D.	S	σ

- 3.
4. $\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$ $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$
5. 为什么是 n-1, 因为这样估计出来的方差最准确, 是无偏的, 从自由度角度解释: degree of freedom,
6. 以上这些衡量的是绝对离散程度

Chebyshev's inequality 切比雪夫不等式

1. 对于任何一组观测值, 个体落在均值周围 k 个标准差的概率不小于 $1 - \frac{1}{k^2}$, 对于任意 $k > 1$
2. $P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$

3. 题型：已知 k，求 P

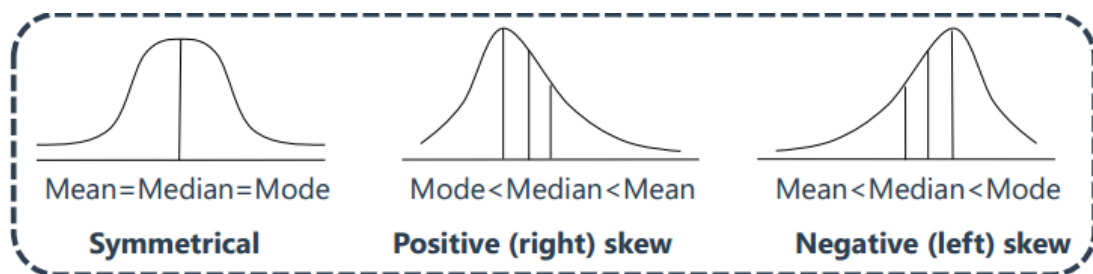
CV coefficient of variance 变异系数

1. $CV = \frac{Sx}{\bar{x}} \times 100\%$
2. $\frac{\text{标准差}}{\text{均值}}$ 提出了规模因素，修理：1,2,3 1001,1002,1003
3. 描述：没单位均值的标准差
4. 性质：
 - 4.1. 无规模影响
 - 4.2. Relative dispersion 相对离散程度

Sharp ratio

$$\text{SharpRatio} = \frac{Rp - Rf}{\sigma}$$

Skewness



- 1 右偏偏度>0 positive
- 2 左偏偏度<0 negative
- 3 右偏: mode<median<mean
- 4 左偏: mode>median>mean

Kurtosis

- 1 Lepokurtic 高峰 & Platykurtic 低峰
 - 1.1 标准正态分布，峰度是 3，大于 3 是高峰，小于 3 是低峰

Reading 9: Probability Concepts

Probability concepts

- 1 基本概念
 - 1.1 随机变量
 - 1.2 Mutually exclusive events: 互斥事件, 不可能同时发生
 - 1.3 Exhaustive events: 遍历事件, 包含了所有可能结果

Two defining properties of probability

- 1 $0 \leq P(E) \leq 1$
- 2 $P(E_1) + P(E_2) + \dots + P(E_n) = 1$

Empirical, subjective and priori probability

- 1 Objective probability 客观概率
 - 1.1 Empirical probability 经验概率, 分析过去, 得到将来
 - 1.2 Priori probability 先验概率, 分析过去, 得到过去
- 2 Subjective probability 主观概率 基于直觉或者主管假设

Odds for or against

- 1 Odds for 事件发生的赔率: $\frac{P(E)}{1-P(E)}$ 事件发生的概率除以不发生的概率
- 2 Odds against: $\frac{1-P(E)}{P(E)}$

Multiplication rule and addition rule

- 1 Conditional probability 条件概率
 - 1.1 $P(A|B)$ B 发生的情况下 A 发生的概率
- 2 Unconditional prob 非条件概率 边际概率
- 3 乘法法则
 - 3.1 $P(AB)$ AB 同时发生的概率 joint prob 联合概率
 - 3.2 $P(AB) = P(A) * P(B|A) = P(B) * P(A|B)$
 - 3.3 如果 AB 互斥, $P(AB) = 0$
- 4 加法法则
 - 4.1 $P(A \text{ or } B) = P(A) + P(B) - P(AB)$

4.2 如果 AB 互斥, $P(A \text{ or } B) = P(A) + P(B)$

Dependent and independent events

- 1 独立事件: AB 的发生互相不产生影响
- 2 $P(A|B) = P(A)$ $P(B|A) = P(B)$ $P(AB) = P(A) * P(B)$
- 3 如果 AB 是互斥的, 那么 AB 一定不是独立事件
- 4 全概率公式 $P(A) = P(A|W1)P(W1) + P(A|W2)P(W2) + \dots + P(A|Wn)P(Wn)$
- 5 $E(X) = \sum_{i=1}^N P_i \times X_i$ $\sigma^2 = \sum_{i=1}^N P_i (X_i - E(X))^2$

Covariance and correlation

- 1 含义: 协方差反映的是两个随机变量变化的方向性 协方差大于 0, 同向变化, 协方差小于 0, 反向变化
- 2 $COV(X,Y) = E[(X-E(X))(Y-E(Y))]$
- 3 自己和自己的协方差等于方差
- 4 例题: 求联合概率

	Y1	Y2
X1	P1	0
X2	0	P2

X1Y1 同时发生的概率是 P1, X2Y2 同时发生的概率是 P2

$$E(X) = X1 * P1 + X2 * P2 \quad E(Y) = Y1 * P1 + Y2 * P2$$

$$\begin{aligned} COV_{X,Y} &= E[(X_i - \bar{X})(Y_i - \bar{Y})] \\ &= P_1 \times (X_1 - \bar{X})(Y_1 - \bar{Y}) \\ &\quad + P_2 \times (X_2 - \bar{X})(Y_2 - \bar{Y}) \end{aligned}$$

5 相关性:

$$5.1 \quad \rho_{XY} = \frac{COV(X,Y)}{Std(X) \times Std(Y)}$$

5.2 相关性反映的是线性关系, 取值范围是 -1 到 1

5.2.1 相关性=1, XY 完全正的线性关系

5.2.2 相关性=-1, XY 完全负的线性关系

5.2.3 相关性=0, XY 没有线性关系

Expected value, variance, and standard deviation of a random variable and of returns on a portfolio

1 公式

$$E(r_p) = \sum_{i=1}^n w_i E(R_i)$$

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \text{cov}(R_i, R_j)$$

$$2 \quad \sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \sigma_1 \sigma_2 \rho_{12} = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 COV(1,2)$$

Bayes formula

- $P(AB) = P(A|B) \times P(B) = P(B|A) \times P(A)$

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

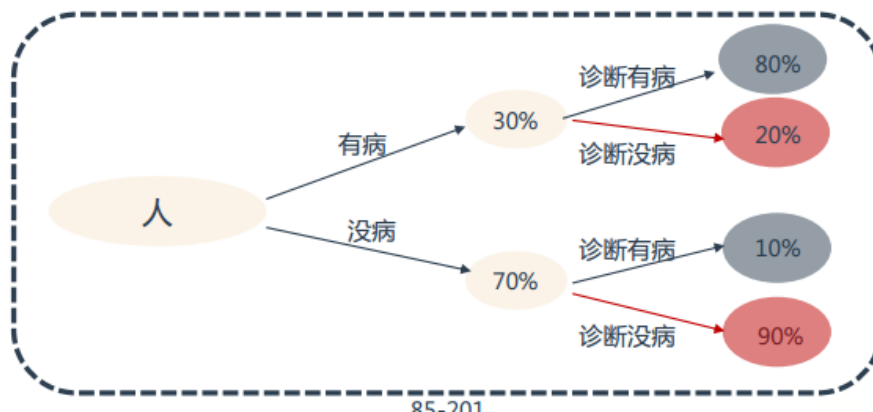
- Where $P(B)$ can be solved using total probability formula:

- ✓ $P(B) = P(B|W_1) \times P(W_1) + P(B|W_2) \times P(W_2) + \dots + P(B|W_n) \times P(W_n)$

- ✓ W_i is a set of **mutually exclusive and exhaustive** events

- The probability that a man has got sick is 30%, and if it does, a medical machine will have 80% chance to diagnose the disease. The probability that a man is healthy is 70%, and if it does, a medical machine will have 10% chance to diagnose the disease, and 90% chance not to diagnose disease. What's the probability that the man has actually got sick when the machine diagnoses disease ?

$$P(A/B) = \frac{30\% \times 80\%}{30\% \times 80\% + 70\% \times 10\%} = 77.42\%$$



Reading 10: Common Probability Distributions

Properties of discrete distribution and continuous distribution

- 1 随机变量都是有概率的，都可以画出概率分布
- 2 离散随机变量（变量数可数）和连续随机变量（变量数不可数）
- 3 连续随机变量再一定范围内考虑

Discrete uniform distribution 均匀离散分布

- 1 基本特征：离散（可数） + 均匀（概率相等）

Binomial distribution 二项式分布

- 1 Bernoulli random variable 贝努力随机变量
 - 1.1 $P(Y=1)=p$ $P(Y=0)=1-p$
 - 1.2 只做一次实验，成功概率 p ，不成功概率 $1-p$ （抛硬币）
- 2 Binomial random variable 二项式，在贝努力基础上做 N 次实验
 - 2.1 做了 n 次，其中 x 次成功的概率： $P(x) = nC_x \times p^x(1-p)^{n-x}$

Continuous uniform distribution 连续均匀分布

1 性质：连续（变量不可数）+均匀

- For all $a \leq x_1 < x_2 \leq b$:

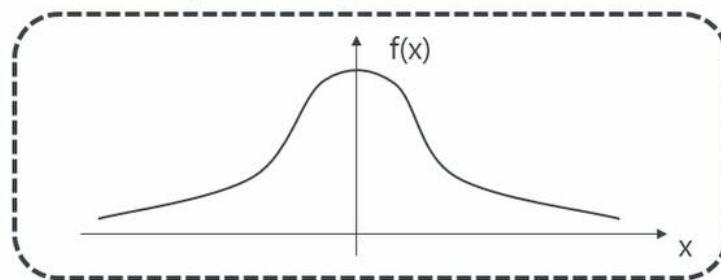
$$P(x_1 \leq X \leq x_2) = (x_2 - x_1) / (b - a)$$

- $P(X < a \text{ or } X > b) = 0$

2

Normal distribution 正态分布

➤ The shape of the density function

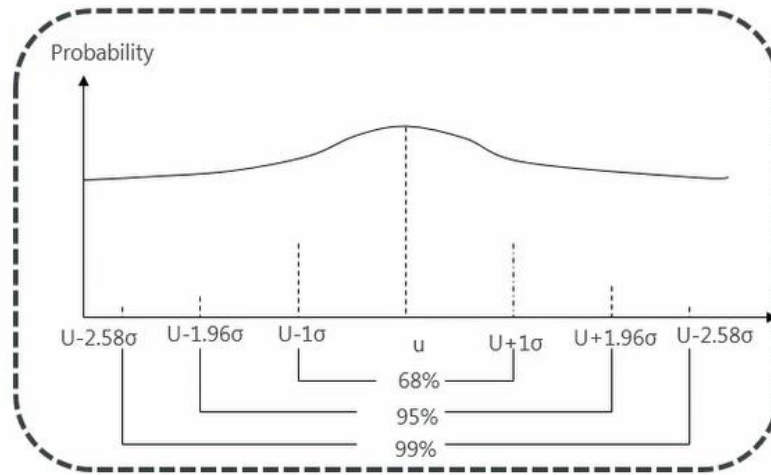


➤ Properties:

- $X \sim N(\mu, \sigma^2)$
- Symmetrical distribution: skewness=0; kurtosis=3
- A linear combination of normally distributed random variables is also normally distributed.
- As the values of x gets farther from the mean, the probability density get smaller and smaller but are always positive.

➤ The confidence intervals

- 68% confidence interval is $[\mu - \sigma, \mu + \sigma]$
- 90% confidence interval is $[\mu - 1.65\sigma, \mu + 1.65\sigma]$
- 95% confidence interval is $[\mu - 1.96\sigma, \mu + 1.96\sigma]$
- 99% confidence interval is $[\mu - 2.58\sigma, \mu + 2.58\sigma]$



103-201

标准正态分布：均值=0，方差=1

一般正态分布的标准化：

如果 X 服从 $N(\mu, \sigma^2)$ ，那么 $Z = \frac{X - \mu}{\sigma}$ 服从 0,1 正态分布

Safety-first ratio 第一安全比例

1 RI=最小要求回报率

2
$$SFR = \frac{E(R_p - R_f)}{\sigma_p}$$

3 当最小要求回报率等于无风险收益时，第一安全比例等于夏普比例

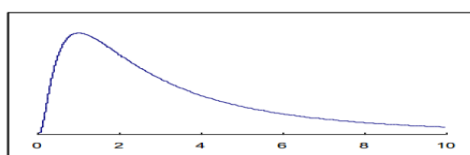
Lognormal distribution 对数正态分布

1 如果 $\ln X$ 服从正太分布，则 X 服从对数正态分布

2 对数正态分布是右偏的，且 X 全部大于 0

3 假设：价格服从对数正态分布，return 服从正态分布

4



Monte carlo simulation

Inferential Statistics

Reading 11: Sampling and Estimation

Simple random and stratified random sampling

- 1 Simple random sampling:简单随机抽样, 概率一样
- 2 Stratified random sampling:分层随机抽样
- 3 Sampling error 抽样误差 样本结果-实际结果
- 4 样本统计量是随机变量, 本身有概率分布

Time-series and cross-sectional data

时间序列数据和横截面数据

Central limit theorem 中心极限定理

- 1 研究样本均值服从什么分布的问题, 不是样本
- 2 前提条件:
 - 2.1 样本容量足够大, $N \geq 30$
 - 2.2 对于 population, 总体的均值, 方差已知
- 3 结论: 样本的均值服从正态分布
- 4 $\mu_{sample} = \mu_{population}$ 样本均值的均值等于总体均值
- 5 $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$ 样本均值的方差等于总体方差除以 n
- 6 $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ 样本均值的标准差叫标准误
- 7 有时候总体的标准差得不到, 可以抽样取得样本的标准差, 来代替总体的标准差 $S_{\bar{X}} = \frac{S}{\sqrt{n}}$

Standard error of the sample mean

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad \text{样本均值的标准差叫标准误}$$

The desirable properties of an estimator

- 1 如何评估估计量好不好
 - 1.1 Unbiasedness 无偏性 样本均值的期望=总体均值
 - 1.2 Efficiency 有效性 无偏样本选方差最小的
 - 1.3 Consistency 一致性
 - 1.3.1 n 增加时, 准确性增加
 - 1.3.2 n 增加时, 标准误减小

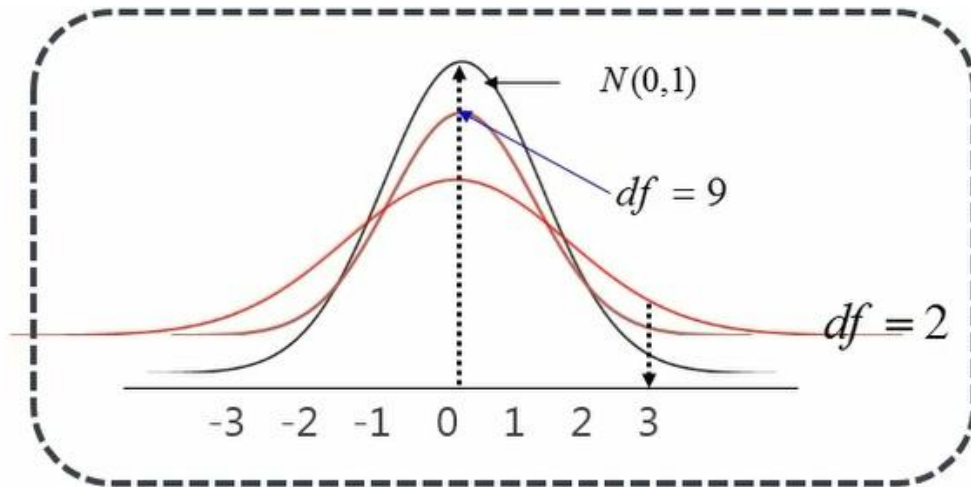
Confidence interval estimate

- 1 Point estimate 点估计, 抽一次样计算得到的均值代表样本均值
- 2 Confidence interval estimate 置信区间估计
 - 2.1 根据中心极限定理, 样本均值服从正态分布 $N(\mu = \mu_x, \sigma_{\bar{X}}^2 = \frac{\sigma_x^2}{n})$, 构造样本均值 \bar{X} 的置信区间 $CI_x = \mu_x \pm k\sigma_{\bar{X}}$
 - 2.2 样本均值 \bar{X} 的均值不知道, 用点估计方法求得, 样本均值 \bar{X} 的标准差即为标准误, 用总体标准差除以根号下 N

$$CI_{\bar{X}} : \underbrace{[\mu_{\bar{X}}] \pm k}_{\downarrow \mu_x \leftarrow \bar{X}} [\sigma_{\bar{X}}] = \frac{\sigma_x}{\sqrt{n}}$$

- 2.3 显著性水平, 双尾巴面积之和
- 2.4 置信度=1-显著性水平

Student's t-distribution



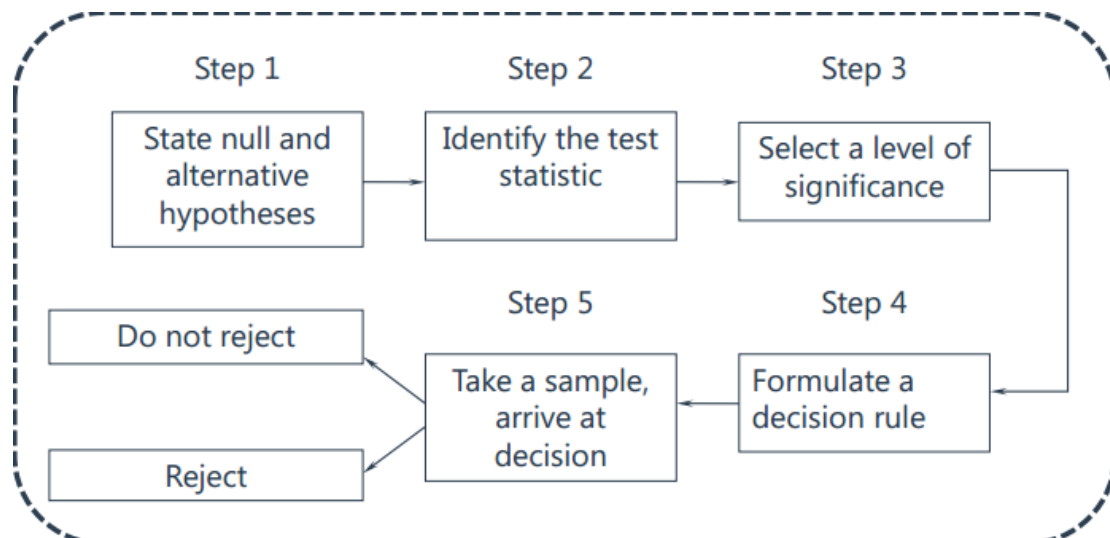
- 1 T 分布是对称的，偏度等于 0
- 2 Degree of freedom: 自由度水平 $n-1$
- 3 低峰，肥尾，峰度小于 3，方差大
- 4 N 变大的时候，T 分布接近 $N(0,1)$ 正态分布
- 5 相同的显著性水平下，T 分布的置信区间宽度更宽
- 6 如何查表，决定用 T 分布还是 Z 分布
 - 6.1 方差已知用 Z 分布，方差未知用 T 分布，非正态总体小样本不可估计
 - 6.2 $N \geq 30$ ，任何情况下都用 Z 分布

Five kinds of biases

- 1 Data-mining bias 把偶然当必然
- 2 Sample selection bias 样本选择性偏差，排除了某些样本
- 3 Survivorship bias 生存性偏差
- 4 Look-ahead bias 前视性偏差
- 5 Time-period bias 时间周期不能太长，也不能太短

Reading 12: Hypothesis Testing

The steps of hypothesis testing



定标准，均值->定测试方法，检验统计量->定标准，显著性水平->计算 K 和置信区间->看是否在区间内

The null hypothesis and alternative hypothesis, one-tailed and two-tailed test

- 1 Null hypothesis 原假设 $H_0: \text{mean}=170$
- 2 Alternative hypothesis 备择假设 $H_a: \text{mean} \neq 170$
- 3 反正法，一般把想要拒绝的假设给 H_0 原假设，想接受的放入备择假设中
- 4 假设是基于总体的 population 而不是样本的
- 5 单尾和双尾

Two-tailed	$H_0 : \mu = \mu_0$	$H_a : \mu \neq \mu_0$	$>$ $<$ 双.
One-tailed	$H_0 : \mu \leq \mu_0$ or, $H_0 : \mu \geq \mu_0$	$H_a : \mu > \mu_0$ $H_a : \mu < \mu_0$	单

- 6 等于一般放在原假设

Test statistics 检验统计量

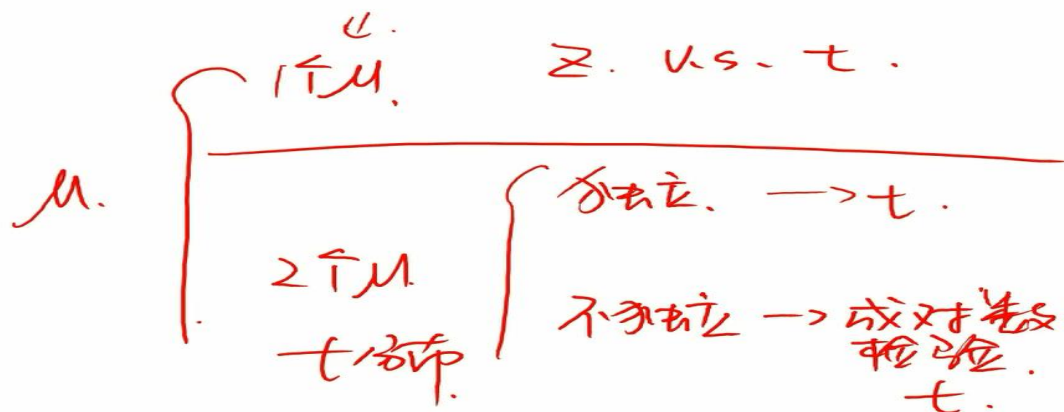
- 1 对一个总体均值检验的检验统计量: $\text{Test Statistic} = \frac{\bar{X} - \mu_x}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - \mu_x}{\frac{S}{\sqrt{n}}}$
- 2 公式中的 μ_x 就是原假设中的设定的均值 μ
- 3 公式中的标准差为样本或者总体的标准差, 用总体标准差是 z 检验统计量, 用样本标准差是 T 检验统计量

Decision rule

- 1 Critical value 关键之, 分位数 其实就是 K 值
- 2 影响 K 值得因素:
 - 2.1 和查哪张表有关, Z 还是 T
 - 2.2 和显著性水平有关系
 - 2.3 是单尾还是双尾检验有关
- 3 判断准则:
 - 3.1 落在区间内, fail to reject H0, 接受原假设
 - 3.2 落在区间外, reject H0, 接受备择假设
- 4 总体流程:
 - 4.1 设假设, 原假设以及备择假设
 - 4.2 画分布, 本质就是对于总体的什么参数进行检验, 是均值还是方差
 - 4.3 找拒绝域, 双尾还是单尾
 - 4.4 找拒绝域的面积, 确定 α , 自己确定
 - 4.5 计算 K 值
 - 4.6 计算检验统计量
 - 4.7 比较 Z 和 K 的大小, 看检验统计量落在哪里
 - 4.8 总结陈词

The chi-square test and f-test

如何判断用什么分布



P-value method

- 1 定义：拒绝原假设的最小显著性水平 α ，他的本质是概率
- 2 判断准则：
 - 2.1 $P < \alpha$ 拒绝原假设
 - 2.2 $P > \alpha$ fail to reject 原假设
 - 2.3 P 越小越拒绝

Type I and type II errors

- 1 Type I error: 据真，错杀好人
- 2 Type II error: 取伪，放走坏人
- 3 一类错误发生的概率： α 小概率事件发生，落在拒绝域
- 4 Power of test: 原假设是错误的切决绝了 $P(\text{拒绝} | H_0 \text{ 错误}) = 1 - P(\text{type II error})$
- 5 $P(\text{type I error})$ 上升 $P(\text{type II error})$ 下降
- 6 样本容量上升， $P(\text{type I error})$ 和 $P(\text{type II error})$ 都下降

Parameter tests and non-parameter-tests

非参数检验的情况

- 1 不满足分布假设
- 2 叙述排列
- 3 检验的不是参数

Technical Analysis

Reading 13: Technical Analysis

- 1 Put/call ratio $p/c r$ 变大，市场看空
- 2 Volatility index IVX 市场恐慌指数，增加，市场看空
- 3 Margin debt 借钱炒股 增加 市场看多
- 4 Short interest rate 等于做空 变大 市场看空
- 5 arm index = $\frac{V_{\downarrow}/N_{\downarrow}}{V_{\uparrow}/N_{\uparrow}}$ 大于 1 看空 小于 1 看多
- 6 mutual fund cash position 公募基金仓位增加 看多
- 7 new equity issuance IPO 增加 市场看多