# CFA
# LEVEL II
## Equity Investments

泽稷网校梁老师


一级权益重难点——抽样与估计

直播时间--北京时间19:00-20:30
视频链接：http://live.zejicert.cn/item?targetId=74
详细板书请见课程视频

# Sampling and estimation

- Population & sample
- Sampling
- The central limit theorem
- Standard error of the sample mean
- Point estimates
- Confidence intervals for population mean

# Population & sample

- 用来测量总体（population）特征的统计量被称为parameter
- 用来测量样本（sample）特征的统计量被称为sample statistic

# Population & sample

- Population Mean： $\mu = \frac{\sum_{i=1}^{N} X_i}{N}$
- Sample Mean： $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
- Population Variance： $\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$
- Sample Variance： $s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$

# Sampling and estimation

- Population & sample
- Sampling
- The central limit theorem
- Standard error of the sample mean
- Point estimates
- Confidence intervals for population mean

# Sampling

- Simple random sampling
- Stratified random sampling
  - ✓ 基于一个或多个特征把总体分成不同的层（smaller groups），再从每一层中随机抽样
- Systematic sampling
  - ✓ 选择所有第k个成员，直到抽够所需的样本数

# Sampling

- 样本统计量（sample statistic）是一个随机变量，服从某一概率分布
- Sampling error
  - ✓ 样本统计量和总体参数之间的difference
  - ✓ 如：$\bar{X}-\mu$

# Sampling

- Sampling： Bias
  - ✓ Data-mining bias（数据挖掘偏差）：统计上显著的模型并不一定有经济意义，因为有可能这种统计上的显著性是通过数据挖掘（data mining）的方法得到的
  - ✓ Sample selection bias（样本选择偏差）：在抽样时，样本中排除了某些特定的数据
    - Survivorship bias（生存偏差）：只包含了现存活的样本，那些现在已不存在的样本数据（如，已清盘的对冲基金数据）并没有包括在内

# Sampling

- Look-ahead bias（前视偏差）：需要使用的数据现在暂时无法得到
- Time-period bias（时间周期偏差）：选择数据的时间周期可能太短也可能太长

# Sampling and estimation

- Population & sample
- Sampling
- The central limit theorem
- Standard error of the sample mean
- Point estimates
- Confidence intervals for population mean

# The central limit theorem

- 对于服从任意分布的总体，总体均值为μ，方差为σ2，每次从总体中抽出一个样本容量为n的样本，只要样本容量n≥30，样本均值（sample mean）近似服从均值为μ，方差为$\sigma^2/n$的正态分布N（μ，$\sigma^2/n$）

# The central limit theorem

- 每一个随机抽出的样本都有它自己的样本均值，而且这个样本均值是一个随机变量
- 总体均值μ和所有样本均值的均值（The mean of all possible sample means）是相等的
- 样本均值的方差（The variance of sample means ）等于$\sigma^2/n$，是用总体方差（The population variance）除以样本容量（The sample size） n

# Sampling and estimation

- Population & sample
- Sampling
- The central limit theorem
- Standard error of the sample mean
- Point estimates
- Confidence intervals for population mean

# Standard error of the sample mean

- The standard error of the sample mean（标准误）是样本均值的标准差（the standard deviation of the sample means）
  - ✓ Known population variance： $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
  - ✓ Unknown population variance： $s_{\bar{X}} = \frac{s}{\sqrt{n}}$
- The larger the sample size， the greater precision with which we can estimate the population parameter.（样本容量n越大，对总体参数的估计越准确）

# Sampling and estimation

- Population & sample
- Sampling
- The central limit theorem
- Standard error of the sample mean
- Point estimates
- Confidence intervals for population mean

# Point estimates & Confidence interval estimate

- Point estimates： 用一个样本统计量来估计总体参数（parameter）
  - ✓ The sample mean， x ， is an estimator of the population mean μ

# Sampling and estimation

- Population & sample
- Sampling
- The central limit theorem
- Standard error of the sample mean
- Point estimates
- Confidence intervals for population mean

# Confidence intervals for population mean

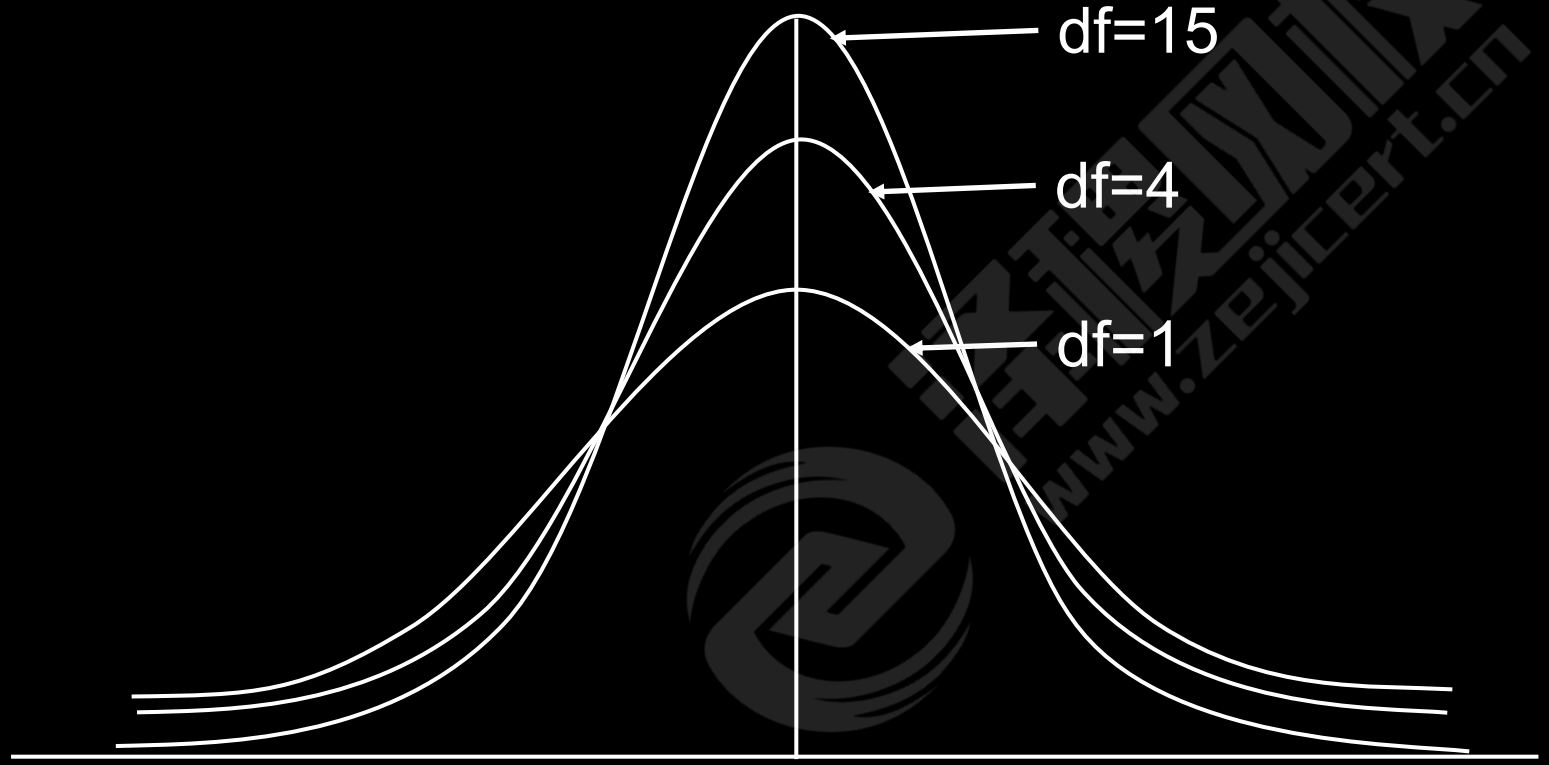
- Confidence interval estimate
  - ✓ Level of significance（alpha，显著性水平）
  - ✓ Degree of Confidence（1-alpha，置信水平）
  - ✓ Confidence Interval=[ Point Estimate+/-（reliability factor）×Standard error ]

# Confidence intervals for population mean

- Z distribution
  - ✓ 总体方差已知
  - ✓ $\bar{X} \pm z\alpha_{/2} \frac{\sigma}{\sqrt{n}}$
    - 90 percent confidence intervals： Use $Z_{0.05} = 1.65$
    - 95 percent confidence intervals： Use $Z_{0.025} = 1.96$
    - 99 percent confidence intervals： Use $Z_{0.005} = 2.58$

# T Distribution

- T Distribution



# T Distribution

- Properties of the T Distribution
  - ✓ Symmetrical
  - ✓ Degrees of freedom（df）： n-1
  - ✓ 与正态分布相比，尾巴更肥（ “fatter tails”），意味着It has more probability in the tails
  - ✓ 随着自由度（the degrees of freedom）， 也即样本容量（the sample size）n的增大， t分布的图形更趋近于标准正态分布（standard normal distribution）

# T Distribution

| df | $p = 0.10$ | $p = 0.05$ | $p = 0.025$ | $p = 0.01$ | $p = 0.005$ | df | $p = 0.10$ | $p = 0.05$ | $p = 0.025$ | $p = 0.01$ | $p = 0.005$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 31 | 1.309 | 1.696 | 2.040 | 2.453 | 2.744 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 32 | 1.309 | 1.694 | 2.037 | 2.449 | 2.738 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 33 | 1.308 | 1.692 | 2.035 | 2.445 | 2.733 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 34 | 1.307 | 1.691 | 2.032 | 2.441 | 2.728 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 35 | 1.306 | 1.690 | 2.030 | 2.438 | 2.724 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 36 | 1.306 | 1.688 | 2.028 | 2.434 | 2.719 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 37 | 1.305 | 1.687 | 2.026 | 2.431 | 2.715 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 38 | 1.304 | 1.686 | 2.024 | 2.429 | 2.712 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 39 | 1.304 | 1.685 | 2.023 | 2.426 | 2.708 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 40 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 41 | 1.303 | 1.683 | 2.020 | 2.421 | 2.701 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 42 | 1.302 | 1.682 | 2.018 | 2.418 | 2.698 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 43 | 1.302 | 1.681 | 2.017 | 2.416 | 2.695 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 44 | 1.301 | 1.680 | 2.015 | 2.414 | 2.692 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 45 | 1.301 | 1.679 | 2.014 | 2.412 | 2.690 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 46 | 1.300 | 1.679 | 2.013 | 2.410 | 2.687 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 47 | 1.300 | 1.678 | 2.012 | 2.408 | 2.685 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 48 | 1.299 | 1.677 | 2.011 | 2.407 | 2.682 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 49 | 1.299 | 1.677 | 2.010 | 2.405 | 2.680 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 50 | 1.299 | 1.676 | 2.009 | 2.403 | 2.678 |
| 21 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 60 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 70 | 1.294 | 1.667 | 1.994 | 2.381 | 2.648 |
| 23 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 80 | 1.292 | 1.664 | 1.990 | 2.374 | 2.639 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 90 | 1.291 | 1.662 | 1.987 | 2.368 | 2.632 |
| 25 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 100 | 1.290 | 1.660 | 1.984 | 2.364 | 2.626 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 110 | 1.289 | 1.659 | 1.982 | 2.361 | 2.621 |

# Confidence intervals for population mean

- T distribution
  - ✓ 总体方差未知
  - ✓ $\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$
  - ✓ $t = \frac{\bar{X} - \mu_X}{S_x/\sqrt{n}} \sim t_{(n-1)}$

# Confidence intervals for population mean

| When sampling from a： | Test statistic | |
|---|---|---|
| | Small sample ( n<30 ) | Large sample ( n≥30 ) |
| 正态分布，总体方差已知 | z-statistic | z-statistic |
| 正态分布，总体方差未知 | t-statistic | t-statistic /z-statistic |
| 非正态分布，总体方差已知 | Not available | z-statistic |
| 非正态分布，总体方差未知 | Not available | t-statistic /z-statistic |

# Confidence intervals for population mean

## Example

The average salary for a sample of 61 CFA charter holders with 10 years experience is $200,000, and the sample standard deviation is $80,000. Assume the population is normally distributed. Which of the following is a 99% confidence interval for the population mean salary of CFA charter holders with 10 years of experience?

A. $172,514 to $227,486
B. $172,754 to $227,246
C. $160,000 to $240,000

# Confidence intervals for population mean

## Example

- The 99% confidence interval is

  $200,000±2.660($80,000/√61) or

  $200,000±2.660×$10,243, or $200,000±$27,246

# Point estimates & Confidence interval estimate

- Desirable properties of an estimator
  - ✓ Unbiasedness（无偏）： expected value of the estimator is equal to the parameter that are trying to estimate
    - The sample mean is an unbiased estimator of the population mean
  - ✓ Efficiency（有效）： if the variance of its sampling distribution is smaller than all the other unbiased estimators of the parameter you are trying to estimate
  - ✓ Consistency（一致）： the accuracy of the parameter estimate increases as the sample size increases

扫码加入全球最大的
CFA人脉圈

# CFA全球考友百日打卡集训营
# 在全球最大的CFA考友社群中寻找CP

本期主题：CFA数量重难点--抽样与估计

本期主讲人：梁老师

祝大家都能顺利通过CFA一级的考试！