Quantitative Analysis

FRM一级培训讲义-强化班

讲师：周琪

Framework

1. Probability
2. Basic Statistics
3. Distributions
4. Hypothesis Tests and Confidence Intervals
5. Linear Regression
6. Elements Of Forecasting(New)
7. Simulation Modeling
8. Estimating Volatilities and Correlations

Probability

Probability

Properties of Probabilities

- If A, B, C, … are mutually exclusive events,
  - ✓ $P(A + B + C + \dots) = P(A) + P(B) + P(C) + \dots$
- If A, B, C, … are mutually exclusive and collectively exhaustive set of events,
  - ✓ $P(A + B + C + \dots) = P(A) + P(B) + P(C) + \dots = 1$
- Generally,
  - ✓ $P(A + B) = P(A) + P(B) - P(AB)$

Probability

- **Unconditional probability: P(A), P(B)**
- **Conditional probability: P(A|B)**

$$\boxed{P(A|B) = \frac{P(AB)}{P(B)};\ P(B) > 0}$$

$$\boxed{P(B|A) = \frac{P(AB)}{P(A)};\ P(A) > 0}$$

- **Joint probability: P(AB) = P(A)P(B|A) = P(B)P(A|B)**

Probability and Probability Distributions

- **Independence: The occurrence of A has <u>no influence</u> of on the occurrence of B.**
  - $P(A|B) = P(A)$ or $P(B|A) = P(B)$
  - $P(AB) = P(A) \times P(B)$
  - $P(A \text{ or } B) = P(A) + P(B) - P(AB)$
- **Independence and Mutually Exclusive are quite different.**
  - If exclusive, must not independence;
    - ✓Cause exclusive means if A occur, B can not occur, A influents B.

Probability Distributions

- **Probability density function (p.d.f): f(x)**
  - For continuous random variable commonly
- **Cumulative probability function (c.p.f): F(x)**
  - $F(x) = P(X \leq x)$
- **Properties of CDF**
  - $F(-\infty) = 0$ and $F(+\infty) = 1$
  - $F(X)$ is a non-decreasing function such that if $x_2 > x_1$ then $F(x_2) \geq F(x_1)$.
  - $P(X \geq k) = 1 - F(k)$
  - $P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$

Probability and Probability Distributions

Probability Matrix

- Summarize joint probabilities in a probability matrix.
- Unconditional/marginal probabilities can be seen by adding across a row or down a column.

Probability and Probability Distributions

Total Probability Formula

- If an event A must result in one of the mutually exclusive events A_1, A_2, A_3, ……, An, then

$$\boxed{P(A)=P(A_1)P(A|A_1)+P(A_2)P(A|A_2)+\ldots+P(A_n)P(A|A_n)}$$

Bayes' Theorem

$$\boxed{P(A|B)=\frac{P(B|A)}{P(B)}\times P(A)=\frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\bar{A})P(\bar{A})}}$$

Exercise 1

- The following is a probability matrix for X = {1, 2, 3} and Y = {1, 2, 3}.

		X		
		1	2	3
Y	1	6%	15%	9%
	2	12%	30%	18%
	3	2%	5%	3%

Each of the following is true except:

A. X and Y are independent.

B. The Covariance (X,Y) is non-zero.

C. The probability Y = 3 conditional on X = 1 is 10%.

D. The unconditional probability that X = 2 is 50%.

- Correct Answer : B

Exercise 2

- There is a unconditional probability of 20% that the Fed will initiate QE4. If the Fed announces QE4, then ABC hedge fund will outperform the market with a 70% probability. If the fed does not announce QE 4, there is only a 40% probability that ABC will outperform. If we observe that ABC outperforms the market, which is nearest to the probability that the Fed announced QE4?

  A. 20%

  B. 28%

  C. 30%

  D. 42%

- Correct Answer : C

Basic Statistics

Expected Value

Expected Value

- A measure of central tendency – the first moment

$$\boxed{E(X)=\sum P(x_i)x_i=P(x_1)x_1+P(x_2)x_2+\ldots+P(x_n)x_n}$$

Properties of Expected Value

1. If b is a constant, E(b) = b.
2. If a is a constant, E(aX) = aE(X).
3. If a and b are constants, then E(aX + b) = aE(X) + E(b) = aE(X) + b.
4. E(X + Y) = E(X) + E(Y)

Variance

Variance

- A measure of dispersion – the second moment

$$\boxed{\sigma^2 = E(X-\mu)^2}$$

- Above formula is the definition of variance. To compute the variance, we use the following formula:

$$\boxed{\sigma^2 = E\left(X^2\right) - \left[E(X)\right]^2 = E\left(X^2\right) - \mu^2}$$

- Measures how noisy or unpredictable that random variable is.
- The positive square root of σ^2, σ, is known as the standard deviation, also called volatility.

Variance

Properties of Variance

1. The variance of a constant is zero.
2. If a is constant, then: $\sigma^2(aX) = a^2\sigma^2(X)$.
3. If b is a constant, then: $\sigma^2(X + b) = \sigma^2(X)$.
4. If a and b are constant, then: $\sigma^2(aX + b) = a^2\sigma^2(X)$.
5. If X and Y are two independent random variables, then:

   $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$ $\quad$ $\sigma^2(X - Y) = \sigma^2(X) + \sigma^2(Y)$
6. If X and Y are independent random variables and a and b are constants, then $\sigma^2(aX + bY) = a^2\sigma^2(X) + b^2\sigma^2(Y)$.
7. For computational convenience, we can get: $\sigma^2(X) = E(X^2) - [E(X)]^2$.

Covariance

Covariance

$$Cov(X,Y) = E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right] = E(XY) - E(X)E(Y)$$

$$S_{XY} = \frac{1}{n-1}\sum_{i-1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$$

- Covariance measures how one random variable moves with another random variable.
- Covariance ranges from negative infinity to positive infinity.

Covariance

Properties of Covariance

- If X and Y are independent random variables, their covariance is zero.

$$Cov(X,\ X) = E\left[\left(X - E(X)\right)\left(X - E(X)\right)\right] = \sigma^2(X)$$

- Cov(X, Y+Z)=Cov(X, Y)+Cov(X, Z), or

$$Cov(a + bX,\ c + dY) = b \times d \times Cov(X, Y)$$

- If X and Y are NOT independent, then:

$$\sigma^2(X \pm Y) = \sigma^2(X) + \sigma^2(Y) \pm 2Cov(X,Y)$$

Correlation Coefficient

Correlation coefficient

$$\rho_{XY} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \sigma_{xy} = Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

Properties of Correlation coefficient

- Correlation has no units, ranges from -1 to +1.
- Correlation measures the linear relationship between two random variables.
- If two variables are independent, their covariance is zero, therefore, the correlation coefficient will be zero. The converse, however, is not true. For example, $Y = X^2$.
- Variances of correlated variables:

$$\sigma^2(X \pm Y) = \sigma^2(X) + \sigma^2(Y) \pm 2\rho\sigma(X)\sigma(Y)$$

Skewness

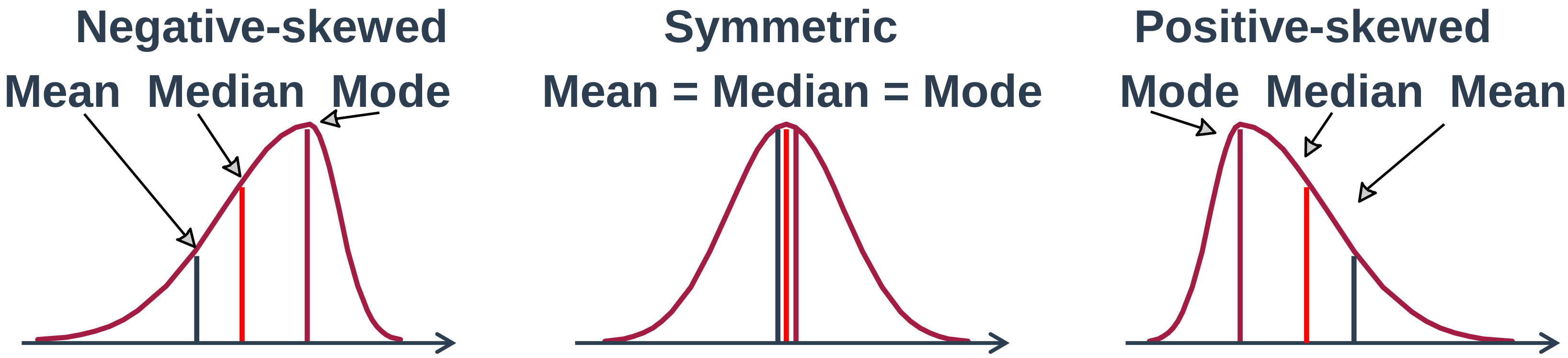
Skewness

- A measure of asymmetry of a PDF – the third moment.

$$S = \frac{E(X-\mu_x)^3}{\sigma_x^3} = \frac{\text{third moment about mean}}{\text{cube of standard deviation}}$$

Symmetrical and nonsymmetrical distributions

- Positively skewed (right skewed) and negatively skewed(left skewed)



- Positive skewed: Mode < median < mean, having a right fat tail
- Negative skewed: Mode > media > mean, having a left fat tail

Kurtosis

Kurtosis

- A measure of tallness or flatness of a PDF – the fourth moment.

$$K = \frac{E(X-\mu_x)^4}{\left[E(X-\mu_x)^2\right]^2} = \frac{\text{fourth moment}}{\text{square of second moment}}$$

- For a normal distribution, the K value is 3.
- Excess kurtosis = kurtosis – 3

	leptokurtic	mesokurtic	platykurtic
Kurtosis	> 3	= 3	< 3
Excess kurtosis	> 0	= 0	< 0
Tails (assuming same variance)	fat tail	normal	thin tail

Chebyshev's Inequality

Chebyshev's Inequality

- For any set of observations (samples or population), the proportion of the values that lie within k standard deviations of the mean is at least 1 – $1/k^2$, k > 1.

$$\boxed{P\left(\left|X-\mu\right|\le k\sigma\right)\ge 1-\frac{1}{k^2},\ k>1}$$

- This relationship applies regardless of the shape of the distribution.

Lin within		Standard deviations of the mean	
	2		$\ge 1-\frac{1}{2^2}=1-\frac{1}{4}=\frac{3}{4}=75\%$
	3		$\ge 1-\frac{1}{3^2}=1-\frac{1}{9}=\frac{8}{9}=89\%$
	4		$\ge 1-\frac{1}{4^2}=1-\frac{1}{16}=\frac{15}{16}=94\%$

Exercise 1

- A model of the frequency of losses (L) per day assumes the following discrete distribution: zero loss with probability of 20%; one loss with probability of 30%; two losses with probability of 30%; three losses with probability of 10%; and four losses with probability of 10%. What are, respectively, the expected number of loss events and the standard deviation of the number of loss events?

  A. E(L) = 1.2 and σ = 1.44

  B. E(L) = 1.6 and σ = 1.20

  C. E(L) = 1.8 and σ = 2.33

  D. E(L) = 2.2 and σ = 9.60

- Correct Answer : B

Exercise 2

- An analyst is concerned with the symmetry and peakedness of a distribution of returns over a period of time for a company she is examining. She does some calculations and finds that the median return is 4.2%, the mean return is 3.7%, and the mode return is 4.8%. She also finds that the measure of kurtosis is 2. Based on this information, the correct characterization of the distribution of return over time is:

	Skewness	Kurtosis
A.	Positive	Leptokurtic
B.	Positive	Platykurtic
C.	Negative	Platykurtic
D.	Negative	Leptokurtic

- Correct Answer : C

Exercise 3

- Using Chebyshev's inequality, what is the proportion of observations from a population of 250 that must lie within three standard deviations of the mean, regardless of the shape of the distribution?

  A. 75%

  B. 99%

  C. 89%

  D. 54%

- Correct Answer : C

Distributions

Binomial Distribution

- Bernoulli Distribution
- $P(X = 1) = p \quad P(X = 0) = 1 - p$

➢ Binomial Distribution

- The probability of x successes in n trails

$$\boxed{p(x) = P(X = x) = C_n^x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}}$$

➢ E

	Expectation	Variance
Bernoulli random variable	p	p(1 – p)
Binomial random variable	np	np(1 – p)

Poisson Distribution

➢ Poisson Distribution

- When there are a large number of trials but a small probability of success, Binomial calculations become impractical.
- If we substitute λ/n for p, and let n very large, the Binomial Distribution becomes the Poisson Distribution.

$$\boxed{p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (\lambda = np)}$$

✓ X refers to the number of success per unit.

✓ λ indicates the rate of occurrence of the random events; i.e., it tells us how many events occur on average per unit of time.

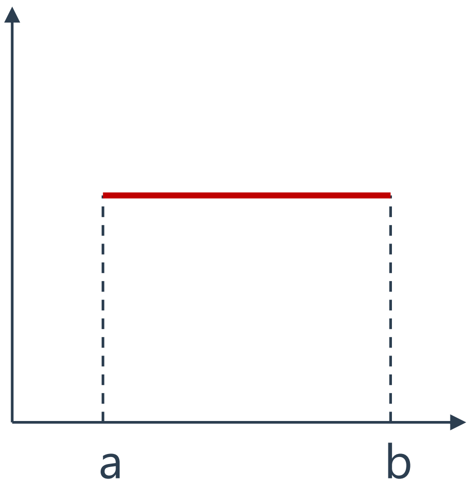
✓

Mean	Variance
λ	λ

Continuous Uniform Distribution

- **Probability density function**

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a\le x\le b\\ 0 & \text{otherwise}\end{cases}$$



- **Cumulative distribution function**

$$F(x)=\begin{cases}0 & \text{for } x\le a\\ \dfrac{x-a}{b-a} & \text{for } a<x<b\\ 1 & \text{for } x\ge b\end{cases}$$

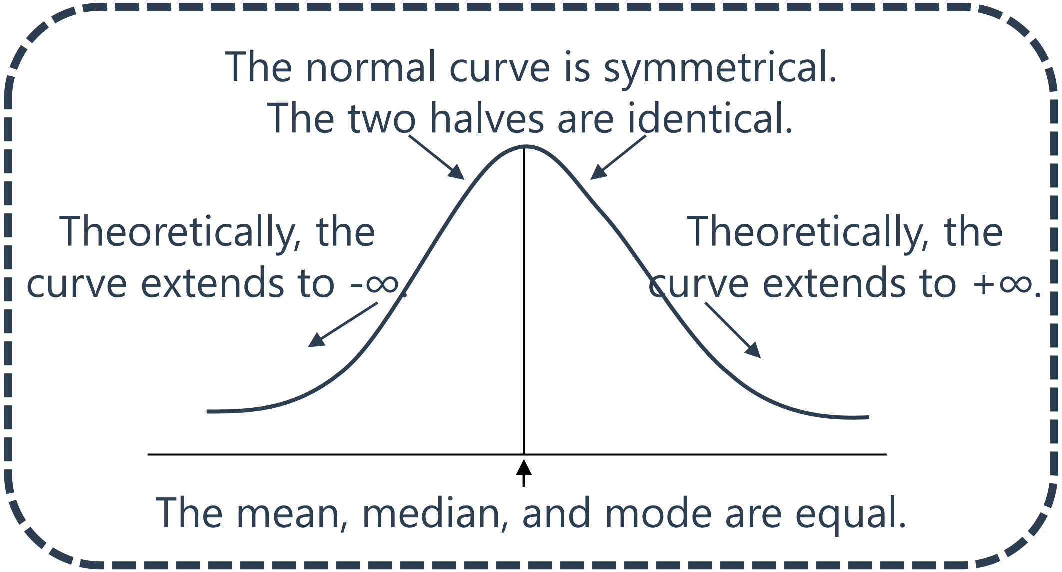
- **Properties**
  - $E(X) = (a + b)/2$, $D(X) = (b - a)^2/12$
  - For all $a \le x_1 < x_2 \le b$, we have: $P(x_1 \le X \le x_2) = \dfrac{x_2 - x_1}{b - a}$

Normal Distribution

Normal Distribution

- As n increases, the binomial distribution approaches Normal Distribution.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2\sigma^2}(x-u)^2}$$



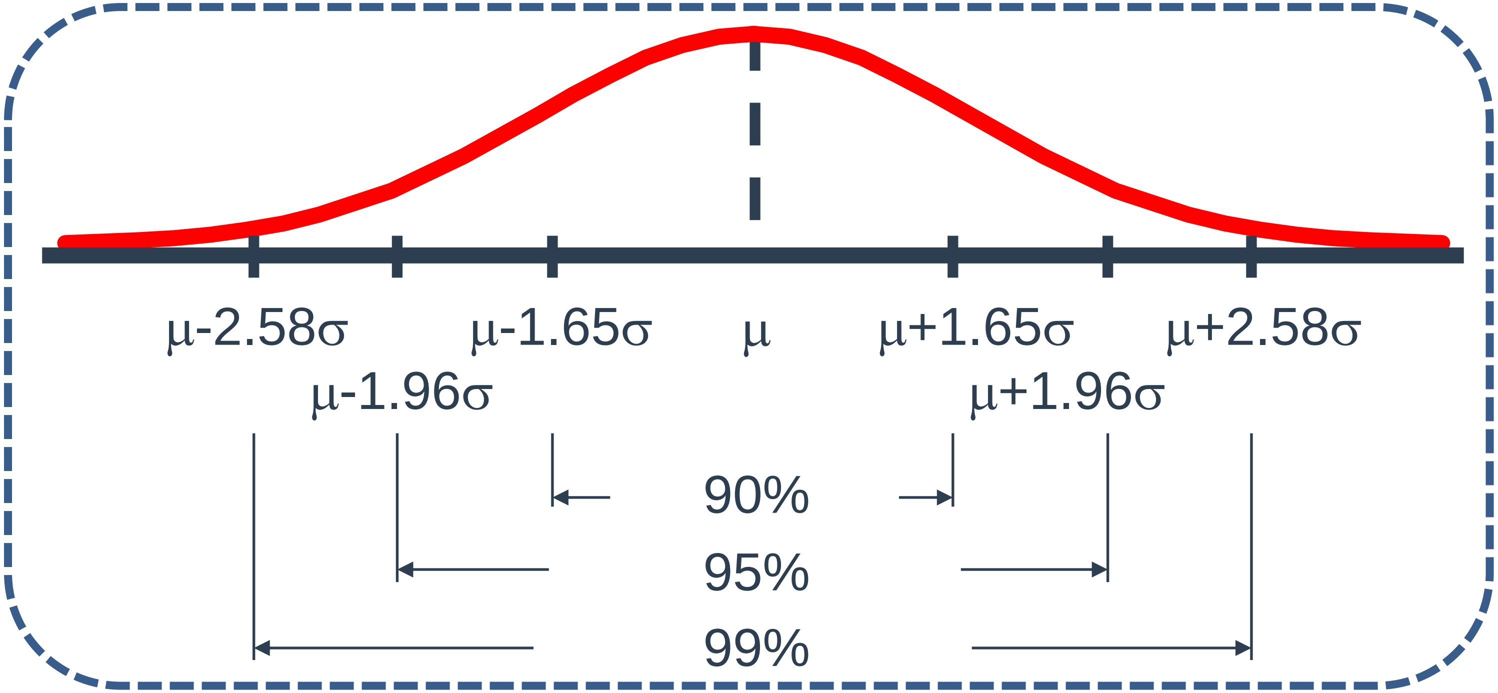
Properties

- $X \sim N(\mu, \sigma^2)$, fully described by its two parameters μ and σ^2.
- Bell-shaped, symmetrical distribution: skewness = 0; kurtosis = 3.
- A linear combination (function) of two (or more) normally distribution random variables is itself normally distributed.
- The tails get thin and go to zero but extend infinitely, asymptotic.

Normal Distribution

The confidence intervals

- Approximately 68% of all observations fall in the interval μ±σ
- Approximately 90% of all observations fall in the interval μ±1.65σ
- Approximately 95% of all observations fall in the interval μ±1.96σ
- Approximately 99% of all observations fall in the interval μ±2.58σ



The Standard Normal Distribution

- **The standard normal distribution**
  - N(0,1) or Z
  - Standardization: if $X \sim N(\mu, \sigma^2)$, then $\boxed{Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)}$
  - Z-table
- **How we use the standard normal distribution to compute various probabilities?**
  - Example: $X \sim N(70, 9)$, compute the probability of $X \le 64.12$.
    - ✓ $Z = \dfrac{X-\mu}{\sigma} = \dfrac{64.12-70}{3} \approx -1.96$
    - ✓ $P(Z \le -1.96) = 0.0250$
  - Question 1: compute the probability of $X \ge 75.9$.
  - Question 2: compute the probability of $64.12 \le X \le 75.9$.

Continuous Probability Distribution

- **Student's t distribution**
  - It is similar to the normal, except it exhibits slightly heavier tails.

$$\boxed{t=\frac{\overline{X}-\mu_X}{S_x/\sqrt{n}}\sim t_{n-1}}$$

✓ It is symmetrical.

✓ It has mean of zero.

  - As the d.f. increase, the t-distribution converges with the standard normal distribution.
  - Both the normal and student's t distribution characterize the sampling distribution of the sample mean. The difference is that the normal is used when we know the population variance. The student's t is used when we must rely on the sample variance. In practice, we don't know the population variance, so the student's t distribution is typically appropriate.

Continuous Probability Distribution

- **Lognormal Distribution**
  - If lnX is normal, then X is lognormal; if a variable is lognormal, its natural log is normal.
  - It is useful for modeling asset prices which never take negative values.
  - Right skewed.
- **Chi-Square (χ^2) Distribution**
- ***F*-Distribution**

Exercise 1

- On a multiple choice exam with four choices for each of six questions, what is the probability that a student gets less than two questions correct simply by guessing?

  A. 0.46%

  B. 23.73%

  C. 35.60%

  D. 53.39%

- Correct Answer : D

Exercise 2

- Assume random variance is normally distributed with mean of 10 and a variance of 25. Without using a calculator, what is the probability that X falls within 1.75 and 21.65?

  A. 68%

  B. 94%

  C. 95%

  D. 99%

- Correct Answer : B

Exercise 3

- The frequency of an operational risk event type – damage to physical assets – is characterized by a Poisson distribution. Over an average year, a company expects 36 of these particular loss events. During the next month, which is nearest to the probability the company will experience exactly zero of these events?

  A. 3.4%

  B. 5.0%

  C. 7.5%

  D. 9.1%

- Correct Answer : B

Exercise 4

- Which of the following statements best characterizes the relationship between the normal and lognormal distributions?
  - A. The lognormal distribution is the logarithm of the normal distribution.
  - B. If the natural log of the random variable X is lognormally distributed, then X is normally distributed.
  - C. If X is lognormally distributed, then the natural log of X is normally distributed.
  - D. The two distributions have nothing to do with one another.

- Correct Answer : C

Exercise 5

- Assume the population of hedge fund returns has an unknown distribution with mean of 8% and volatility of 10%. From a sample of 40 funds, what is the probability the sample mean return will exceed 10.6%?

  A. 5%

  B. 8%

  C. 10%

  D. 12%

- Correct Answer : A

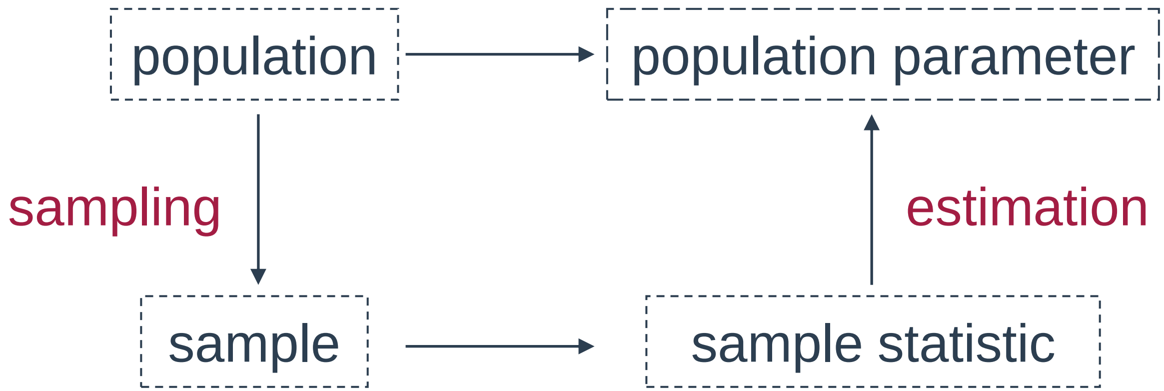
Hypothesis Tests and Confidence Intervals

Statistical Inference: Estimation and Hypothesis Testing

What is Statistical Inference?

- Concerned with drawing conclusions about the nature or some population (e.g., the normal) on the basis of a random sample that supposedly been drawn from that population.
- Loosely speaking, is the study of the relationship between a population and a sample drawn from that population.

Sampling and Estimation



Statistical Inference: Estimation and Hypothesis Testing

Central Limit Theorem (CLT)

- Properties
  - ✓ If $n \geq 30$, the sampling distribution of the sample means will be approximately normal.
  - ✓ The mean of the population μ = the mean of the distribution of all possible sample means.
  - ✓ The variance of the distribution of sample means is the population variance divided by the sample size.
- Standard Error of the Mean

$$\boxed{SE(\overline{X}) = \frac{s}{\sqrt{n}}}$$

  - ✓ The standard deviation of the sampling distribution of the sample means is called the standard error of the sample means.

Statistical Inference: Estimation and Hypothesis Testing

- Regardless of whether we are concerned with point estimates or confidence intervals, there are certain statistical properties that make some estimates more desirable than others. These desirable properties of an estimator are unbiasedness, efficiency, and consistency.
  - **Unbiasedness**: expected value of the estimator is equal to the parameter that are trying to estimate.
  - **Efficiency:** for all unbiased estimators, if the sampling dispersion is smaller than any other unbiased estimators, then this unbiased estimator is called efficient.
  - **Consistency**: the accuracy of the parameter estimate increases as the sample size increases.

Statistical Inference: Estimation and Hypothesis Testing

- Another property of a point estimate is linearity. A point estimate should be a linear estimator (i.e., it can be used as a linear function of the sample data). If the estimator is the best available (i.e., has the minimum variance), exhibits linearity, and is unbiased, it is said to be the best linear unbiased estimator (BLUE).

Point Estimation and Confidence Interval Estimate

Point Estimation

- Using a single numerical value to estimate the parameter of the population.

parameter

$$\boxed{\bar{X} \rightarrow \mu_X \qquad S^2 \rightarrow \sigma_X^2}$$

Estimator

Confidence Interval Estimate

- Level of significance (alpha)
- Degree of confidence (1 – alpha)
- Confidence Interval = [Point Estimate +/- (reliability factor) × standard error]

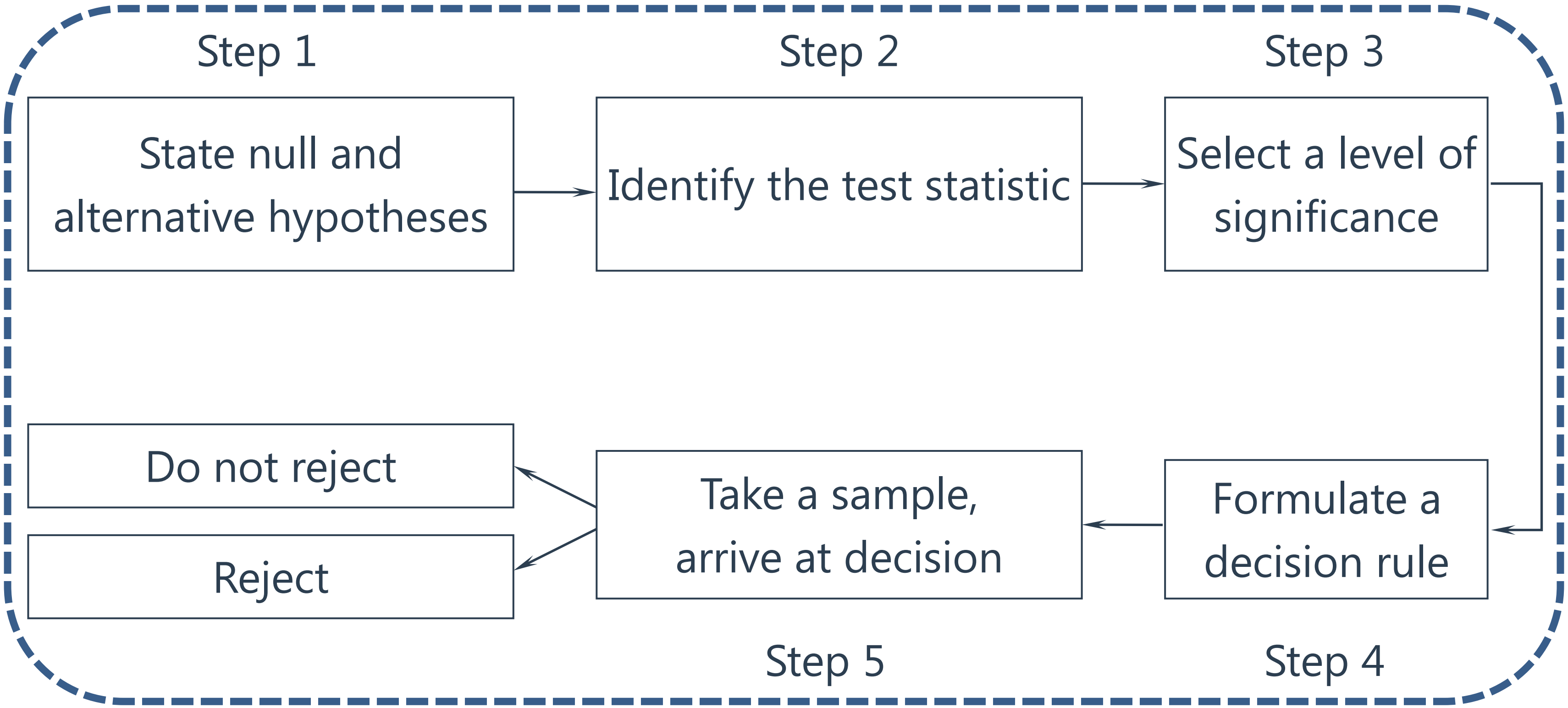
Statistical Inference: Estimation and Hypothesis Testing

$$\boxed{\overline{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}}$$

$$\boxed{\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

When sampling form a:	small sample (n < 30)	larger sample (n ≥ 30)
Normal distribution with known variance	z-Statistic	z-Statistic
Normal distribution with unknown variance	t-Statistic	t-Statistic or z-Statistic
Nonnormal distribution with known variance	not available	z-Statistic
Nonnormal distribution with unknown variance	not available	t-Statistic or z-Statistic

Hypothesis Testing



Hypothesis Testing

- **Hypothesis**
  - Statistical assessment of a statement or idea regarding a population parameter.
- **The null hypothesis (H_0) and alternative hypothesis (H_a)**
- **One-tailed test vs. Two-tailed test**
  - One-tailed test

H_0: μ ≥ 0 H_a: μ < 0	H_0: μ ≤ 0 H_a: μ > 0

  - Two-tailed test

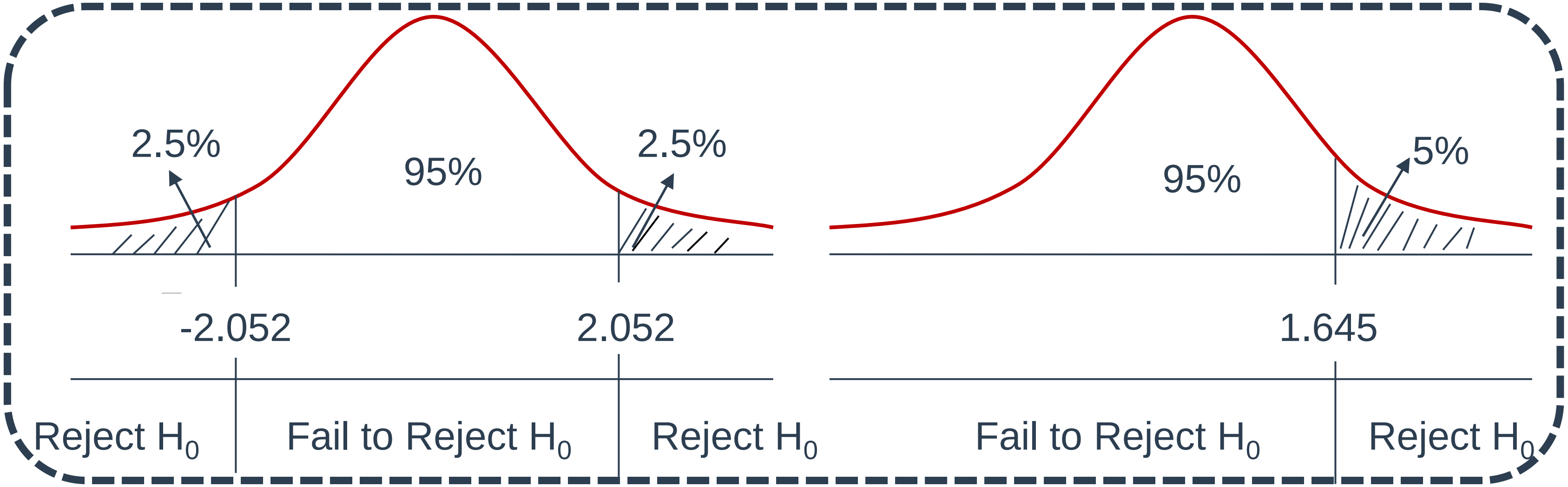
H_0: μ = 0 H_a: μ≠0

- **Critical Value**
  - The distribution of test statistic (z, t, χ^2, F)
  - Significance level (α)
  - One-tailed or two-tailed test

Decision Rule

Decision Rule

- Reject H_0 if |test statistic| > critical value.
- Fail to reject H_0 if |test statistic| < critical value.



- We can never say "accept" H_0.
- State the conclusion: μ is (not) significantly different from μ_0.

Test of Single Population Mean

- **H_0: $\mu = \mu_0$**
  - z-test vs. t-test

	Normal population, n < 30	n ≥ 30
Known population variance (σ^2)	z-test	z-test
Unknown population variance	t-test	t-test or z-test

- z-statistic $\boxed{z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}}$

- t-statistic $\boxed{t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}}$

Test of Single Population Variance

- **H_0: $\sigma^2 = \sigma_0^2$**
  - The Chi-Square test

$$\boxed{\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \qquad df = n-1}$$

Where:

n = sample size

s^2 = sample variance

σ_0^2 = hypothesized value for the population variance

Test of Variances Difference

- **H_0: $\sigma_1^2 = \sigma_2^2$**
  - The F-test

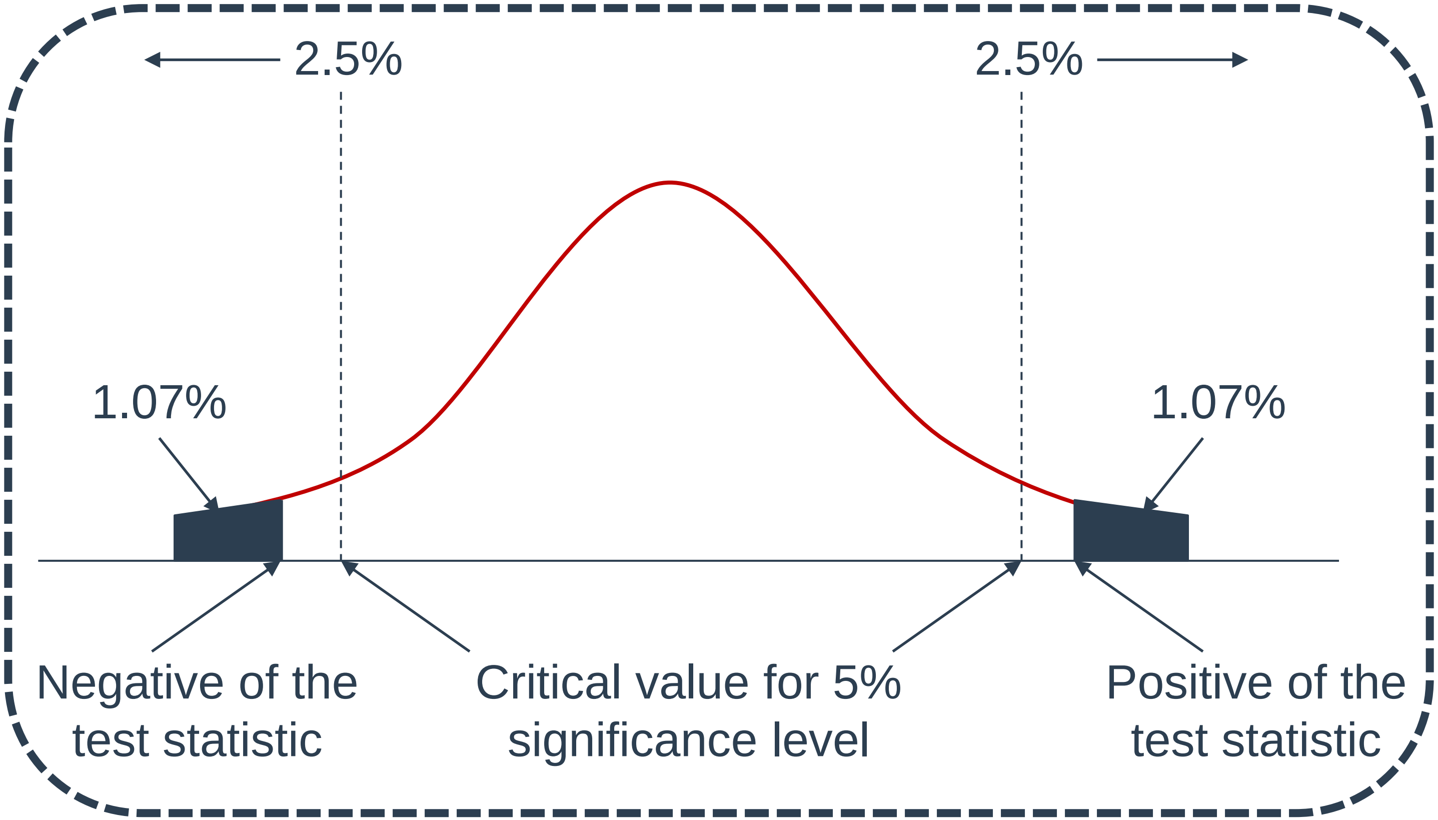
$$\boxed{F = \frac{s_1^2}{s_2^2} \quad df_1 = n_1 - 1 \quad df_2 = n_2 - 1}$$

  - Always put the larger variance in the numerator ($s_1^2 > s_2^2$).

Summary of Hypothesis Testing

Test type	Assumptions	H_0	Test-statistic	distribution
Mean hypothesis testing	Normally distributed population, known population variance	$\mu = 0$	$z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$	N(0,1)
	Normally distributed population, unknown population variance	$\mu = 0$	$t = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$	t(n-1)
Variance hypothesis testing	Normally distributed population	$\sigma^2 = \sigma_0^2$	$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$	$\chi^2(n-1)$
	Two independent normally distributed populations	$\sigma_1^2 = \sigma_2^2$	$F = s_1^2 / s_2^2$	$F(n_1 - 1,\ n_2 - 1)$

P-value testing



- If P-value < alpha, we reject null hypothesis.

Type I and Type II Errors

- **Type I error**
  - Reject the null hypothesis when it's actually true.
- **Type II error**
  - Fail to reject the null hypothesis when it's actually false.
- **Significance level (α)**
  - The probability of making a Type I error:

Significance level = P(Type I error)

- **Power of a test**
  - The probability of correctly rejecting the null hypothesis when it is false:

Power of a test = 1 – P(Type II error)

Exercise 1

- Which of the following statements regarding hypothesis testing is correct?
  - A. Type II error refers to the failure to reject the H1 when it is actually false.
  - B. Hypothesis testing is used to make inferences about the parameters of a given population on the basis of statistics computed for a sample that is drawn from another population.
  - C. All else being equal, the decrease in the chance of making a type I error comes at the cost of increasing the probability of making a type II error.
  - D. If the p-value is greater than the significance level, then the statistics falls into the reject intervals.
- Correct answer ： C

Exercise 2

- An oil industry analyst with a large international bank has constructed a sample of 1,000 individual firms on which she plans to perform statistical analyses. She considers either decreasing the level of significance used to test hypotheses from 5% to 1%, or removing 500 state-run firms from her sample. What impact will these changes have on the probability of making type I and type II errors?

	Level of Significance Decrease	Reduction in Sample Size
A.	P(type I error) increases	P(type I error) increases
B.	P(type I error) decreases	P(type II error) increases
C.	P(type II error) increases	P(type I error) decreases
D.	P(type II error) decreases	P(type II error) decreases

- Correct answer ： B

Exercise 3

- Analyst John is concerned that the average days sales outstanding has increased above its historical average of 27 days. From a large sample of 36 companies, he computes a sample mean DSO of 29 days with sample standard deviation of 7. His one-sided alternative hypothesis, stated with 95% confidence, is that DSO is greater than 27. Does she reject the null?
  - A. No, do not reject one-sided null as the t-statistic is less than 1.65.
  - B. No, do not reject one-sided null as the t-statistic is less than 1.96.
  - C. Yes, do reject one-sided null as the t-statistic is greater than 1.65.
  - D. Yes, do reject one-sided null as the t-statistic is greater than 1.96.

- Correct answer ： C

Exercise 4

- If the mean P/E of 30 stocks in a certain industrial sector is 18 and the sample standard deviation is 3.5, standard error of the mean is closest to:

  A. 0.12

  B. 0.34

  C. 0.64

  D. 1.56

- Correct answer ：C

Linear Regression

Dependent and Independent Variable

Regression analysis

- Regression analysis is concerned with the study of the relationship between one variable called the dependent or explained variable and one or more other variables called independent or explanatory variables.

Dependent variable

$$Y = b_0 + b_1X + \varepsilon$$

Independent variable

- The objectives of regression analysis: to predict or forecast dependent variable.

Ordinary Least Squares (OLS)

Ordinary least squares (OLS)

- OLS estimation is a process that estimates the population parameters bi with corresponding values for bi that minimize the squared residuals (i.e., error terms).
- The OLS sample coefficients are those that:

$$\boxed{\text{minimize} \sum \varepsilon_i^2 = \sum [Y_i - (\hat{b}_0 + \hat{b}_1 \times X_1)]^2}$$

$$\boxed{\hat{b}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{Cov(X,Y)}{Var(X)}X}$$

$$\boxed{\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}}$$

- The estimated intercept coefficient ($\hat{b}_0$): the point ($\bar{X}$, $\bar{Y}$) is on the regression line.

The Basics of Multiple Regression

- **Linear Regression with Multiple Regressors**

$$\boxed{Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki} + \varepsilon_i}$$

  - If the regressor is correlated with a variable that has been omitted from the analysis and that determines, in part, the dependent variable, then the OLS estimator will have omitted variable bias.

- The intercept term is the value of the dependent variable when the independent variables are all equal to zero.

- Each slope coefficient is the estimated change in the dependent variable for a one unit change in that independent variable, holding the other independent variables constant. That's why the slope coefficients in a multiple regression are sometimes called partial slope coefficient.

Dummy Variables

Dummy Variables

- Observations for most independent variables can take on a wide range of values. However, there are occasions when the independent variable is binary in nature - it is either "on" or "off". Independent variables that fall into this category are called **dummy variables** and are often used to quantify the impact of qualitative events.
- Dummy variables are assigned a value of "0" or "1".
- The coefficient on dummy variables indicates the **difference** in the dependent variable for the category represented by the dummy variable and the average due of the dependent variable for all classes except the dummy variable class.
  - ✓ For example, testing the slope coefficient for rise January dummy variable would indicate whether, and by how much, security returns are different in January as compared to the other months.

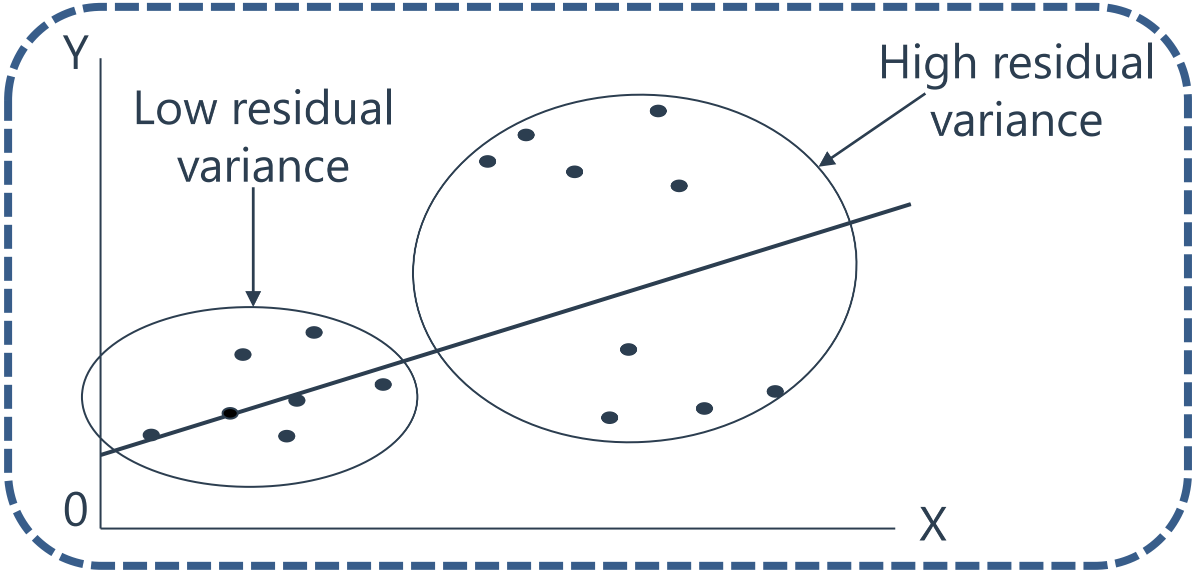
The Basics of Linear Regression

The assumptions of the linear regression

- A linear relationship exists between X and Y.
- Independent variables are not random, or independent variables are not correlated with error term, or independent variables are not correlated(only in multiple regression).
- The expected value of the error term is zero (i.e., $E(\varepsilon_i) = 0$).
- The variance of the error term is constant (i.e., the error terms are homoskedastic).
- The error term is uncorrelated across observations (i.e., $E(\varepsilon_i\varepsilon_j) = 0$ for all $i \neq j$).
- The error term is normally distributed.

Homoskedasticity and Heteroskedasticity

- **Conditional Heteroskedasticity**
  - The heteroskedasticity is related to the level of (i.e., conditional on) the independent variable.
  - The residual variance with be larger if the values of the independent variable X is larger.
  - Conditional heteroskedasticity does create significant problems for statistical inference.



Homoskedasticity and Heteroskedasticity

➢ Effect of Heteroskedaticity on Regression Analysis

- The standard errors are usually unreliable estimates.
- The coefficient estimates (the b_1) aren't affected.
- If the standard errors are too small, but the coefficient estimates themselves are not affected, the t-statistics will be too large and the null hypothesis of no statistical significance is rejected too often.
  - ✓ The opposite will be true if the standard errors are too large.

Violation of the Basic Assumption

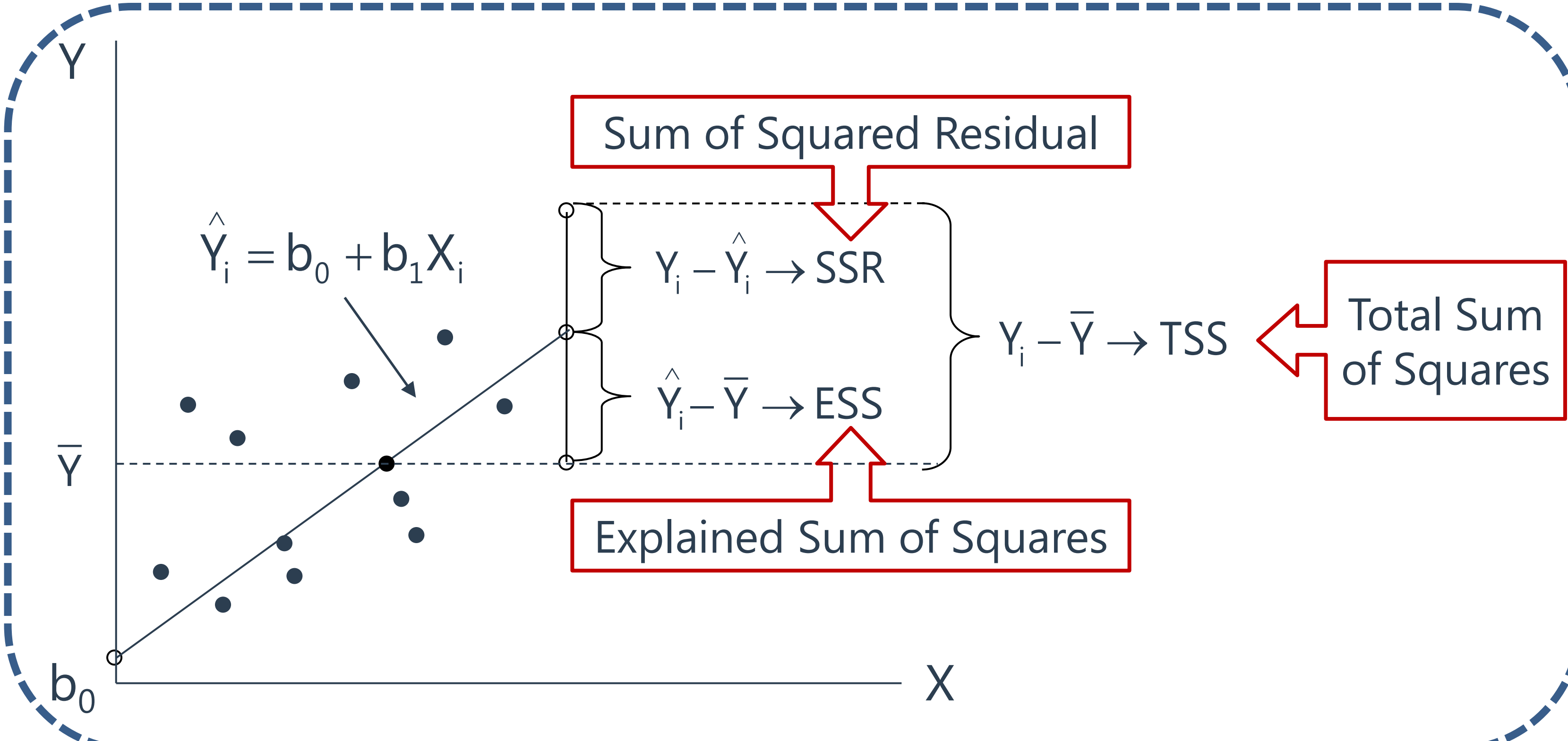
- **Autocorrelation**
  - Autocorrelation, also known as serial correlation, is the cross-correlation between the values of the same variables.
  - Impact
    - ✓ Autocorrelation violates the ordinary least squares assumption that the error terms are uncorrelated. While it does not bias the OLS coefficient estimates, <u>the standard errors tend to be underestimated (and the t-scores overestimated) when the auto correlations of the errors are positive</u>.

Violation of the Basic Assumption

- **Multicollinearity(Multiple regression only)**
  - Multicollinearity refers to the situation that two or more independent variables are highly correlated with each other.
  - Impact
    - ✓ When conducting regression analysis, we need to be cognizant of imperfect multicollinearity since OLS estimators will be computed, but the resulting coefficients may be improperly estimated.
  - Two methods to detect multicollinearity
    1. t-tests indicate that none of the individual coefficients is significantly different than zero, while the F-test indicates overall significance and the R^2 is high.
    2. The absolute value of the sample correlation between any two independent variables is greater than 0.7 (i.e., $|r| > 0.7$).
  - Methods to correct multicollinearity
    - ✓ Omit one or more of the correlated independent variables.

Analysis of Variance (ANOVA) Table

- **Components**



Total sum of squares = explained sum of squares + sum of squared residuals

$$\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y} - \bar{Y})^2 + \sum(Y_i - \hat{Y})^2$$

$$TSS = ESS + SSR$$

Analysis of Variance (ANOVA) Table

- **Analysis of Variance (ANOVA) Table**

	df	**SS**	**MSS**
Regression	k = 1	ESS	ESS/k
Residual	n – 2	SSR	RSS/(n – 2)
Total	n – 1	TSS	-

Measures of Fit

- **SER**

  - Standard error of regression (SER) estimates the standard deviation of the error term. In this way, the SER is a measure of spread of the distribution of Y around the regression line.
  - The SER gauges the "fit" of the regression line. The smaller the standard error, the better the fit.
  - SER will be low(relative to total variability) if the relationship is very strong, or will be high if the relationship is weak.

$$\boxed{SER = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}}}$$

Measures of Fit

- **R^2 (the Coefficient of Determination)**
  - A measure of the "goodness of fit" of the regression. It is interpreted as a percentage of variation in the dependent variable explained by the independent variable. Its limits are $0 \le R^2 \le 1$.

$$\boxed{R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}}$$

  - In a simple two-variable regression, the square of R^2 is the correlation coefficient(r) between X_i and Y_i:

$$\boxed{r^2 = R^2 \rightarrow r = \pm\sqrt{R^2}}$$

R^2 and Adjusted R^2

R^2 and Adjusted R^2

- In multiple regression, the R^2 increases whenever a regressor (independent variable) is added, unless the estimated efficient on the added regressor is exactly zero.
- The adjusted R^2 is a modified version of the R^2 that does not necessarily increase with a new independent variable is added. Adjusted R^2 is given by:

$$\boxed{\text{Adjusted } R^2 = 1 - \frac{SSR/n-k-1}{TSS/n-1} = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}}$$

✓ Adjusted $R^2 \leq R^2$; adjusted R^2 may be less than zero.

Confidence Interval

➢ The confidence interval for the regression coefficient, b_1, is calculated as:

$$\boxed{\hat{b}_1 \pm (t_c \times s_{\hat{b}_1}), \text{or } [\hat{b}_1 - (t_c \times s_{\hat{b}_1}) < b_1 < \hat{b}_1 + (t_c \times s_{\hat{b}_1})]}$$

- t_c: the critical two-tailed t-value for the selected confidence level with the appropriate number of degrees of freedom, which is equal to the number of sample observations minus 2 (i.e. n – 2).
- $s_{\hat{b}_1}$: the standard error of the regression coefficient.

➢ $s_{\hat{b}_1}$ is a function of the SER: as SER rises, $S_{\hat{b}_1}$ also increases, and the confidence interval widens. SER measures the variability of the data about the regression line, and the more variable the data, the less confidence interval.

Hypothesis Testing

Regression Coefficient Hypothesis Testing

- The hypothesis that is tested is whether the true slope is zero ($b_1 = 0$). The appropriate test structure for the null and alternative hypotheses is:

$$H_0: b_1 = 0 \quad H_a: b_1 \neq 0$$

- A t-test may also be used to test the hypothesis that the true slope coefficient, b_1, is equal to some hypothesized value. Letting b_1 be the point estimate for b_1, the appropriate test statistic with n – 2 degrees of freedom is:

$$\boxed{t = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} \sim t(n-2)}$$

- The decision rule for tests of significance for regression coefficients is:

Reject H_0 if $t > +t_{critical}$ or $t < -t_{critical}$

Hypothesis test for a Partial Slope Coefficient

Hypothesis test for a Partial Slope Coefficient

- H_0: $b_j = 0$ ($j = 1, 2, \ldots, k$)

$$\boxed{t = \frac{\hat{b}_j}{s_{\hat{b}_j}} \sim t(n-k-1)}$$

- Regression coefficient confidence interval

$$\boxed{\hat{b}_j \pm \left(t_c \times s_{\hat{b}_j} \right)}$$

Joint Hypothesis Testing

Joint Hypothesis Testing

- An F-test is used to test whether at least one slope coefficient is significantly different from zero.

H_0: $b_1 = b_2 = b_3 = \ldots = b_k = 0$; H_a: at least one $b_j \neq 0$ ($j = 1$ to k)

F-Statistic

$$\boxed{F = \frac{ESS/k}{SSR/(n-k-1)}}$$

- The F-test here is always a one-tailed test.
- The test assesses the effectiveness of the model as a whole in explaining the dependent variable.
- Decision rule: reject H_0, if $F_{(test\text{-}statistic)} > F_{c(critical\ value)}$.

Exercise 1

- John is trying to get some insight into the relationship between the return on stock LMD ($R_{LMD,t}$) and the return on the S&P 500 index ($R_{S\&P,t}$). Using historical data he estimates the following:

Annual mean return for LMD	11%
Annual mean return for S&P 500 index	7%
Annual volatility for S&P index	18%
Covariance between the return of LMD and S&P 500 index	6%

Assuming he uses the same data to estimate the regression model given by:

$$R_{LMD,t} = \alpha + \beta R_{S\&P,t} + \varepsilon_t$$

Using the ordinary least squares technique, which of the following models will she obtain?

A. $R_{LMD,t} = -0.02 + 0.54 R_{S\&P,t} + \varepsilon_t$

B. $R_{LMD,t} = -0.02 + 1.85 R_{S\&P,t} + \varepsilon_t$

C. $R_{LMD,t} = 0.04 + 0.54 R_{S\&P,t} + \varepsilon_t$

D. $R_{LMD,t} = 0.04 + 1.85 R_{S\&P,t} + \varepsilon_t$

- Correct Answer : B

Exercise 2

- An analyst is given the data in the following table for a regression of the annual sales for Company XYZ.

Parameters	Coefficient	Standard Error of the Coefficient
Intercept	-94.88	32.97
Slope (industry sales)	0.2796	0.0363

The correlation between company and industry sales is 0.9757. Which of the following is closest to the value and reports the most likely interpretation of R^2?

A. 0.048, indicating that the variability of industry sales explains about 4.8% of the variability of company sales.

B. 0.048, indicating that the variability of company sales explains about 4.8% of the variability of industry sales.

C. 0.952, indicating that the variability of industry sales explains about 95.2% of the variability of company sales.

D. 0.952, indicating that the variability of company sales explains about 95.2% of the variability of industry sales.

- Correct Answer : C

Exercise 3

- A regression of a stock's return (in percent) on an industry index's return (in percent) provides the following results:

	Coefficient	Standard Error
Intercept	2.1	2.01
Industry Index	1.9	0.31

	Degrees of Freedom	SS
Explained	1	92.648
Residual	3	24.512
Total	4	117.160

Which of the following statements regarding the regression is incorrect?

A. The correlation coefficient between the X and Y variables is 0.889.

B. The industry index coefficient is significant at the 99% confidence interval.

C. If the return on the industry index is 4%, the stock's expected return is 9.7%.

D. The variability of industry returns explains 21% of the variation of company returns.

- Correct Answer : D

Exercise 4

- You built a linear regression model to analyze annual salaries for a developed country. You incorporated two independent variables, age and experience, into your model. Upon reading the regression results, you notice that the coefficient of experience is negative, which appears to be counterintuitive. In addition, you discover that the coefficients have low t-statistics but the regression model has a high R square. What is the most likely cause of these results?

  A. Incorrect standard errors

  B. Heteroskedasticity

  C. Serial correlation

  D. Multicollinearity

- Correct Answer : D

Forecasting Trends

Modeling and Forecasting Trend

- **Linear trend**: The trend in which appears roughly linear, meaning that it increases or decreases like **a straight line**.
- **Non-linear Trend**
  - Quadratic trend models

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t^{\ 2}$$

  - Polynomial trend:

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t^{\ 2} + \beta_3 TIME_t^{\ 3} + \ldots + \beta_n TIME_t^{\ n}$$

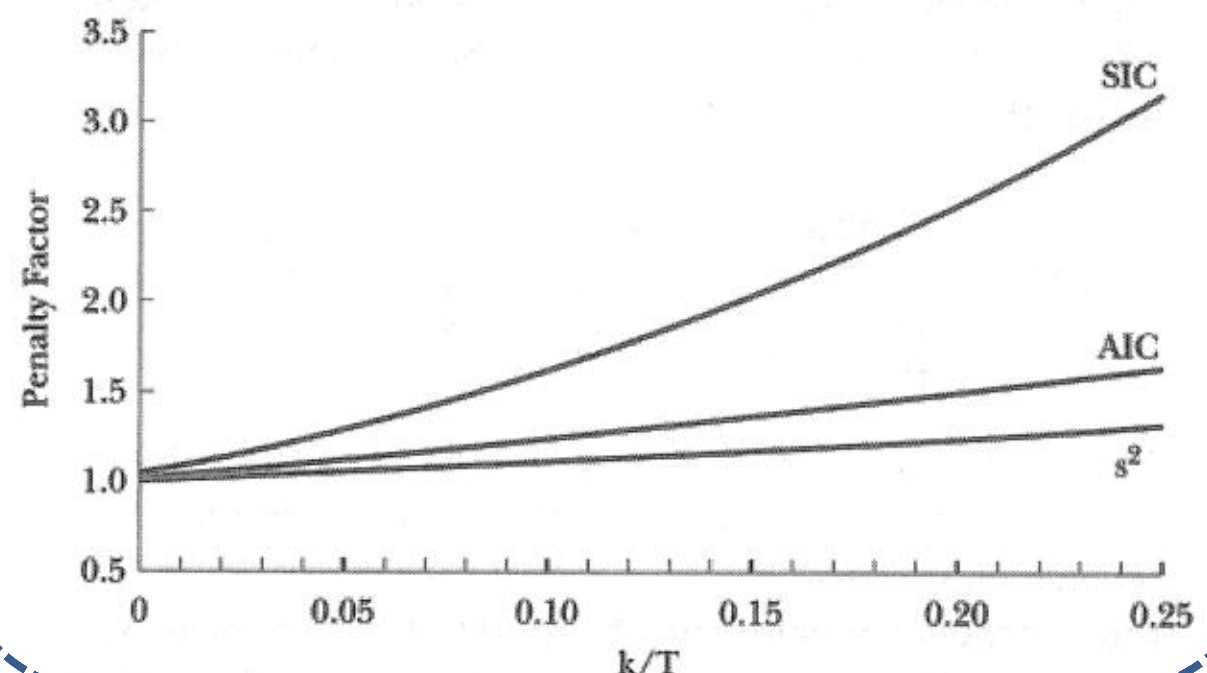
  - Exponential trend, If trend is characterized by constant growth at rate β_1, then we can write

$$T_t = \beta_0 e^{\beta_1 TIME_t}$$

Modeling and Forecasting Trend

- The SIC （ the most consistency ceritria ） generally penalizes free parameters more strongly than does the Akaike information criterion, though it depends on the size of T and relative magnitude of T and k.

Degrees-of-freedom penalties, various model selection criteria.



Modeling and Forecasting Trend

- We evaluate model selection criteria in terms of a key property called **consistency.** A model selection criterion is consistent if the following conditions are met:
  - as the sample size gets large, the chosen measure will choose the true model correctly or with the biggest probability.
  - SIC is consistency while others are not.
- If the no model discussed has the property of consistency. We're then led to a different optimality property, called **asymptotic efficiency.**
  - as the sample size gets large, an asymptotically efficient model selection criterion chooses the model has the fastest speed to approach to the true error variance.
  - The AIC, although inconsistent, is asymptotically efficient, whereas the SIC is not.

Modeling and Forecasting Seasonality

- **The Sources Of Seasonality**
  - Any **technology** that involves the weather, such as production of agricultural commodities, is likely to be seasonal as well.
  - **Preferences** may also be linked to the calendar. People want to do more vacation travel in the summer, which tends to increase both the price and quantity of summertime gasoline sales.
  - **Social institutions** that are linked to the calendar, such as holidays.
- **A key technique for modeling seasonality is regression on seasonal dummies.**

Modeling cycle

- **What is cycle?**
  - As we mentioned the cycles, we generally considered the all-encompassing notion of cyclicality into two parts:
    - ✓Can be captured by trend or seasonal
    - ✓Cannot be captured by trend or seasonal
- **Charactering cycle**
- **Modeling cycle**

Characterizing Cycles

- **Why we need covariance stationary?**
  - If the underlying probabilistic structure of the series were changing over time, there would be no way to predict the future accurately on the basis of the past, **because the laws governing the future would differ from those governing the past**.
  - If we want to forecast a series, at minimum we require the mean and covariance to be stable and finite over time, which we call it **covariance stationary**.
  - In this chapter all we mentioned covariance stationarity is called **second-order stationarity** or **weak stationarity** .It means that a series whose mean and variance and covariance are stable and finite all the time but the skewness and kurtosis are not necessary.

Characterizing Cycles

- The autocovariance is just the covariance between y_t and $y_{t-\tau}$, as the series is covariance stationary, so the autocorrelation will only depend on τ, and have no relationship with t , so the function can be written as follow:

$$\gamma = (\tau) = \operatorname{cov}(y_t, y_{t-\tau}) = E(y_t - u)(y_{t-\tau} - u)$$

- The autocorrelations are just the “simple” or “regular” correlations between y_t and $y_{t-\tau}$ so the function can be written as follow:

$$\rho(\tau) = \frac{\operatorname{cov}(y_t, y_{t-\tau})}{\sqrt{\operatorname{var}(y_t)}\sqrt{\operatorname{var}(y_{t-\tau})}} = \frac{\gamma(\tau)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} = \frac{\gamma(\tau)}{\gamma(0)}$$

Characterizing Cycles

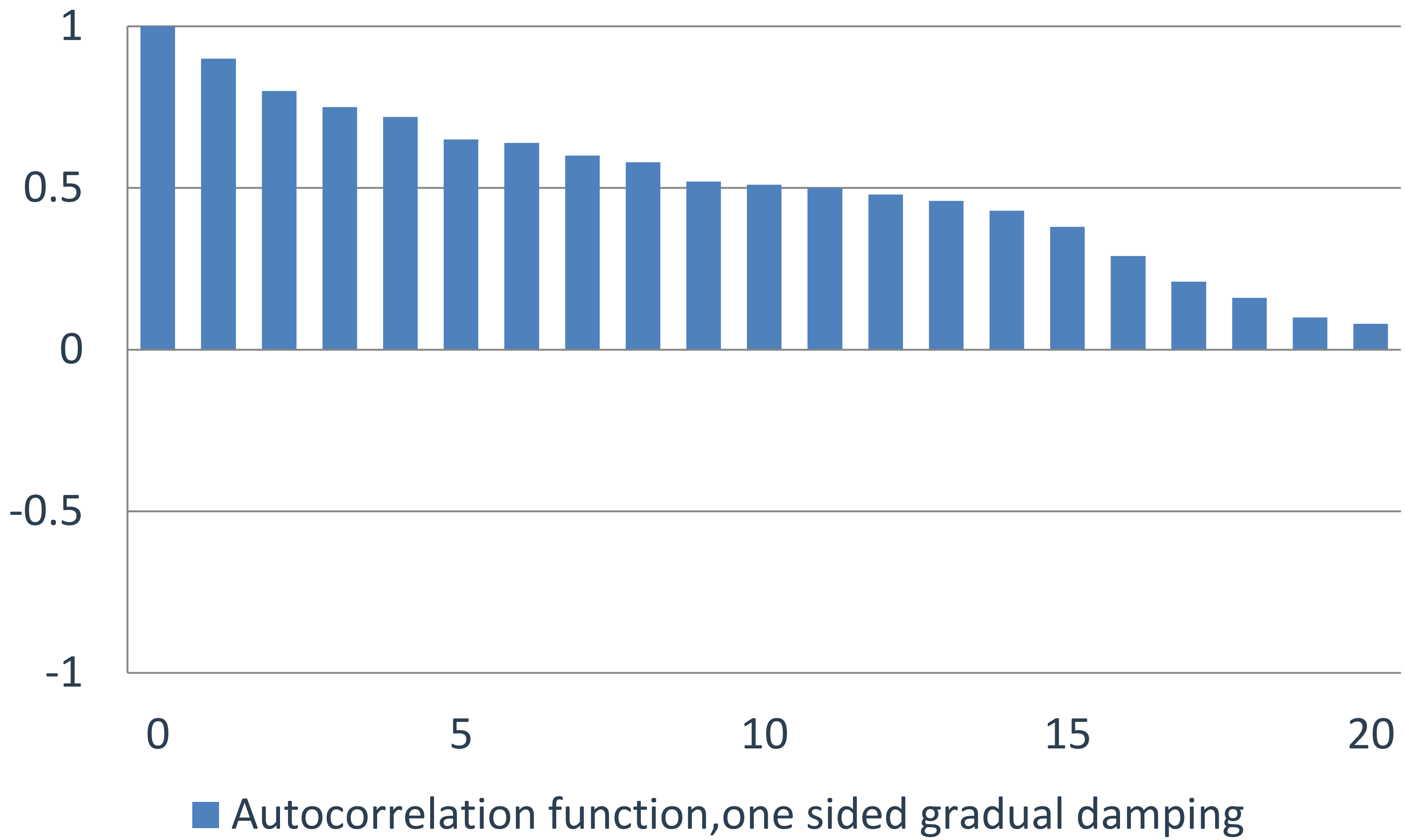
- The partial autocorrelations function measures the relationship between y_t and $y_{t-\tau}$ removed the effects of $y_{t-1}\cdots\cdots y_{t-\tau}$,in other words, the partial autocorrelations function measures the relationship only between y_t and $y_{t-\tau}$.

$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_n y_{t-n} + \varepsilon_t$$

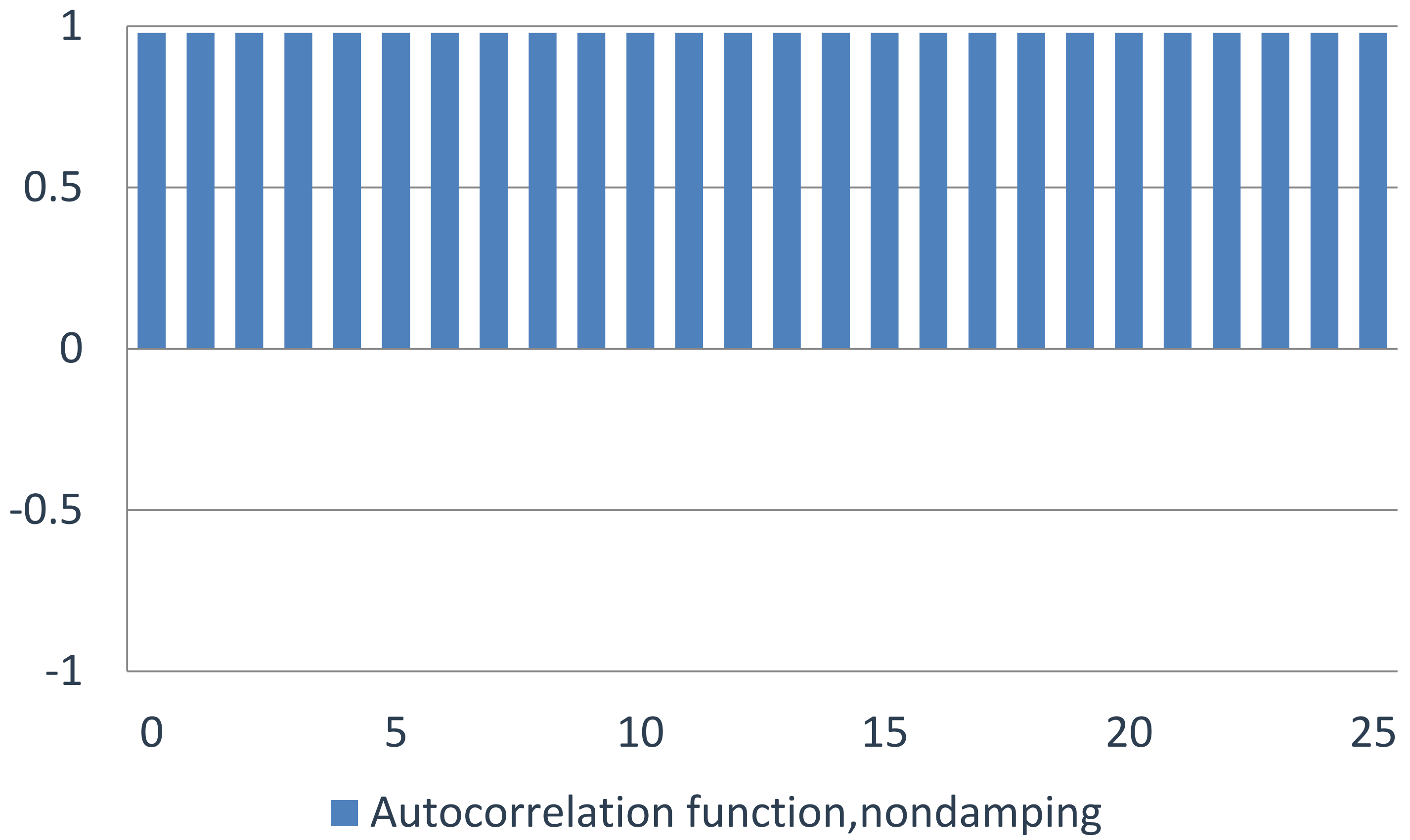
$$p(\tau) = b_\tau$$

Characterizing Cycles

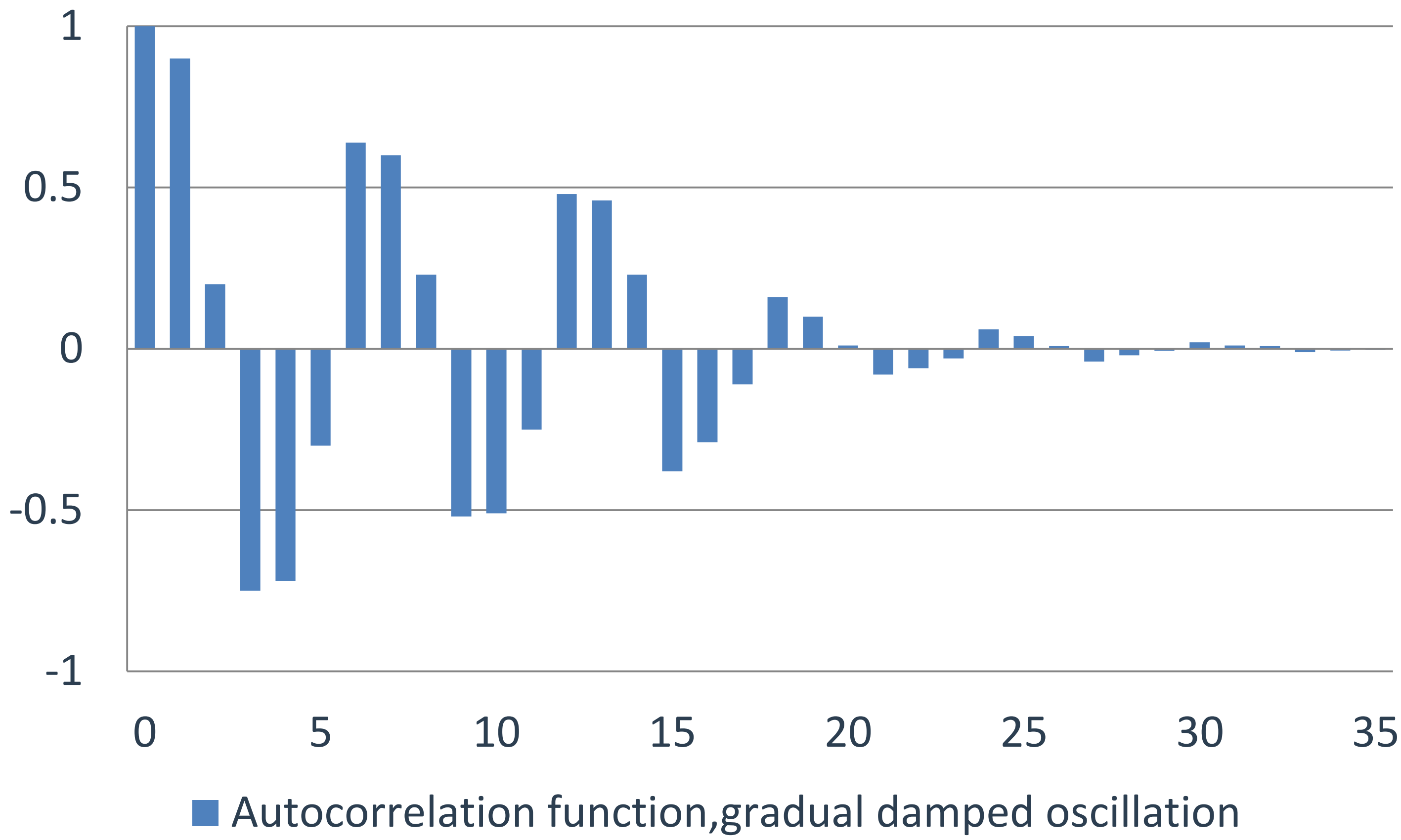
- The four characteristics of autocorrelation function(Graphs)



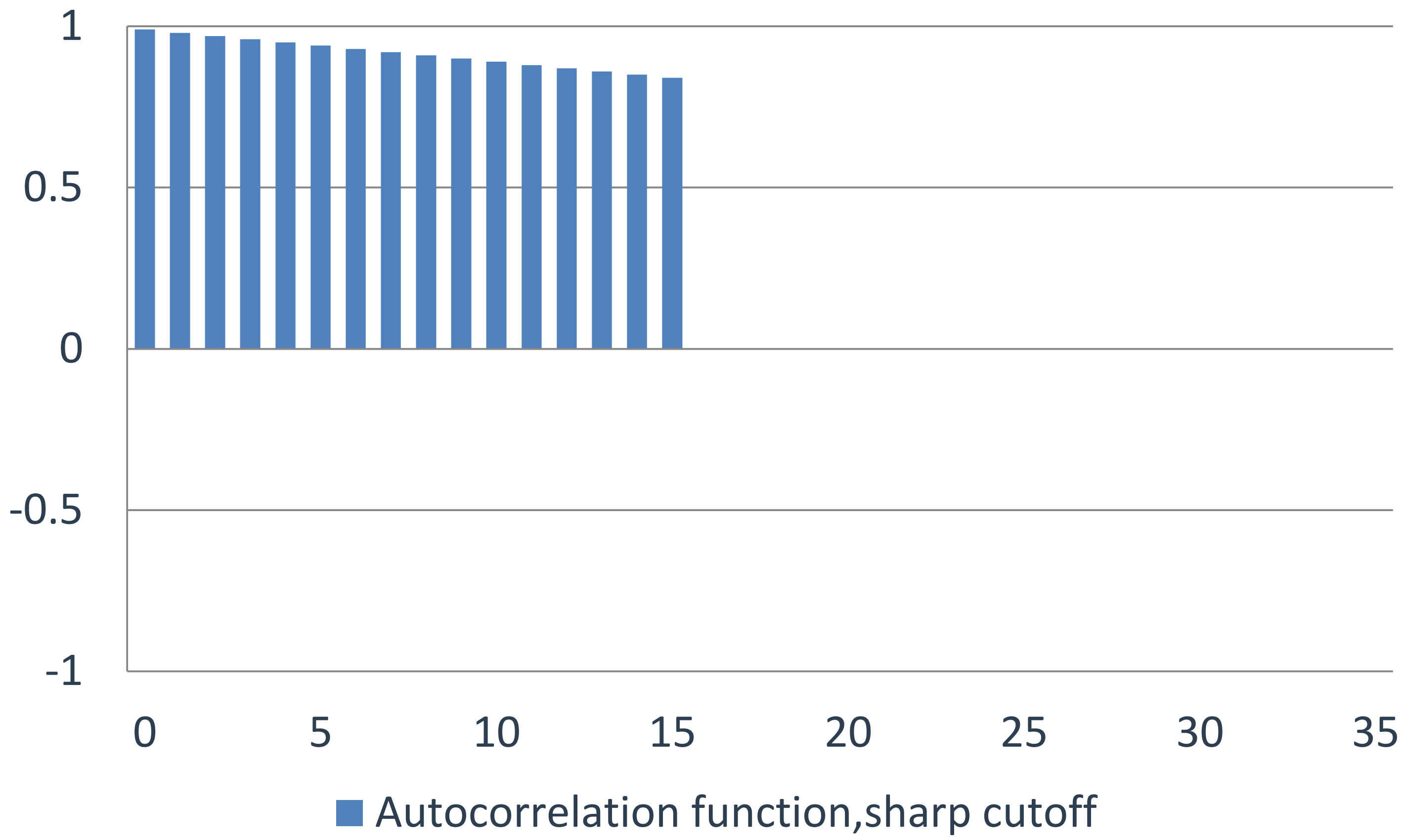
Characterizing Cycles



Characterizing Cycles



Characterizing Cycles



Characterizing Cycles

- **White noise**
  - In general, if there is a process that have zero mean, constant variance, and no serial correlation, the process is called **zero-mean white noise, or simply white noise**. Sometimes for short we write:

$$\varepsilon_t \sim WN(0, \sigma^2)$$

$$y_t = \varepsilon_t$$

hence:

$$y_t \sim WN(0, \sigma^2)$$

Characterizing Cycles

- **Wold's theorem(Wold's representation)**
  - Let $\{y_t\}$ be any zero-mean covariance-stationary process.

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i} \qquad \varepsilon_t \sim WN(0, \sigma^2)$$

✓ where the b_i are coefficients with $b_0 = 1$ and $\sum_{i=0}^{\infty} b_i^2 < \infty$

✓ In short, the correct "model" for any covariance stationary series is some infinite distributed lag of white noise, called the **Wold's representation.** The ε_t are often called **innovations.**

Characterizing Cycles

- Hypothesis test for white noise
  - Hypothesis testing for one autocorrelation coefficient is not enough to conclude covariance stationary, instead, a joint hypothesis test for all the autocorrelation coefficients to be zero is needed.
  - If the autocorrelation coefficients follow a normal distribution, then the sum of squared autocorrelation coefficients follow a **chi-squared distribution.**
    - ✓ **Box-Pierce Q-Statistic & Ljung-Box Q-Statistic**

Modeling Cycle

MA(1) Model

$$\boxed{y_t = \varepsilon_t + \theta\varepsilon_{t-1} = (1+\theta L)\varepsilon_t;\ \varepsilon_t \sim WN(0,\sigma^2)}$$

- The current value of the observed series is expressed as a function of current and lagged unobservable shocks.
- It is the very special case of Wold's representation.

MA(q) Model

$$\boxed{\begin{array}{l} y_t = \varepsilon_t + \theta\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q} = \Theta(L)\varepsilon_t;\ \varepsilon_t \sim WN(0,\sigma^2) \\ \Theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q \end{array}}$$

Modeling Cycle

- Characteristic of MA(1)
  - If the coefficient $|\theta|<1$, then MA(1) process is **a convergent process**.
  - The structure of the MA(1) process, in which only the first lag of the shock appears on the right, forces it to **have a very short memory**, and hence weak dynamics, regardless of the parameter value.
  - MA(1) process with parameter θ = 0.95 almost share the same persistence with the process with a parameter of θ = 0.4.
  - Autocorrelation graph appears to **have sharp cutoff**.
  - Partial Autocorrelation graph appears to have **gradual damped oscillation.**

Modeling Cycle

- **AR(1) Model**

$$\boxed{\begin{aligned} & y_t = \varphi y_{t-1} + \varepsilon_t;\ \varepsilon_t \sim WN\left(0,\sigma^2\right) \\ & (1-\varphi L) y_t = \varepsilon_t \end{aligned}}$$

- **The relationship between AR and MA model**
  - The AR model described a relationship between y_t and y_{t-i}
  - The MA model described a relationship between y_t and ε_t ,which ε_t is a white noise process.
- AR(p) Model

$$\boxed{\begin{aligned} & y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t;\ \varepsilon_t \sim WN\left(0,\sigma^2\right) \\ & \Phi(L) y_t = \left(1 - \varphi_1 L - \varphi_2 L^2 + \ldots - \varphi_p L^p\right) y_t = \varepsilon_t \end{aligned}}$$

Modeling Cycle

- **Characteristic of AR(1)**
  - The AR(1) model is capable of capturing much more persistent dynamics than is the MA(1).
  - Autocorrelation graph appears to have **gradual damped oscillation or one sided gradual damping.**
    - ✓ $|\varphi| < 1$ is the condition for covariance stationary in the AR(1). If φ is positive, the autocorrelation decay is one-sided. If φ is negative, the decay involves back-and-forth oscillations.
  - Partial Autocorrelation appears to **have sharp cutoff.**

Modeling Cycle

ARMA(p, q) Model

- ARMA models are often both highly accurate and highly parsimonious.

$$\boxed{y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \theta\varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t ; \varepsilon_t \sim WN(0, \sigma^2)}$$

- Regardless of autocorrelation or partial autocorrelation, their graphs all appear to be gradual damped.

Estimating Volatilities and Correlations

Estimating Volatilities and Correlations

- **ARCH: Add Long-Run Average Variance Rate**

$$\boxed{\sigma_n^2 = \gamma V_L + \sum_{i=1}^{m} \alpha_i u_{n-i}^2}$$

- **EWMA**
  - In EWMA, the weights decline (in constant proportion, given by lambda).

$$\boxed{\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1-\lambda) u_{n-1}^2}$$

$$\boxed{\sigma_n^2 = (1-\lambda) \sum_{i=1}^{m} \lambda^{i-1} u_{n-i}^2 + \lambda^m \sigma_{n-m}^2}$$

Estimating Volatilities and Correlations

➢ GARCH

- GARCH(1,1) is the weighted sum of a long run-variance, the most recent squared-return, and the most recent variance.

$$\boxed{\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2}$$

$$\boxed{\begin{aligned} &\omega = \gamma V_L \\ &\gamma + \alpha + \beta = 1 \\ &V_L = \frac{\omega}{1-\alpha-\beta} \end{aligned}}$$

- EWMA is a special case of GARCH (1,1).
- Persistence (α + β) defines the speed at which shocks to the variance revert to their long-run values. The higher the persistence, the longer it will take to revert to the mean.

Estimating Volatilities and Correlations

Estimating Correlations

$$\hat{\rho}_{XY} = \frac{Cov_n}{\sigma_{x,n}\sigma_{y,n}}$$

- For EWMA model

$$Cov_n = \lambda Cov_{n-1} + (1-\lambda) x_{n-1} y_{n-1}$$

- For GARCH(1,1) model

$$Cov_n = \omega + \alpha x_{n-1} y_{n-1} + \beta Cov_{n-1}$$

Exercise 1

- Assume an exponentially weighted moving average (EWMA) model with a lambda parameter of 0.94. Yesterday's daily volatility was 1.9%. The price of the stock closed at $10 yesterday and closed up today at 11.275. What is the updated EWMA volatility estimate?

  A. 0.1202%

  B. 2.3750%

  C. 3.4676%

  D. 5.2255%

- Correct Answer : C

Exercise 2

- Given lambda of 0.94, under an infinite series, what is the weight assigned to the seventh prior daily squared return?

  A. 4.68%

  B. 4.40%

  C. 4.14%

  D. 3.89%

- Correct Answer : C

Exercise 3

- A risk manager is using the EWMA approach to estimate the volatilities of asset X and Y and their correlation. With λ = 0.92, the estimate of the covariance between assets X and Y on day t – 1 is 0.000243, some of the estimated returns and volatilities of the two assets are given in the following table:

	Asset X		Asset Y	
Day	Return	Volatility	Return	Volatility
t-1	0.3%		2.1%	
T	0.2%	1.44%	2.5%	2.94%

What is the estimated correlation coefficient on day t?

A. 0.18

B. 0.50

C. 0.54

D. 0.57

- Correct Answer : C

Exercise 4

- Consider the historical series of stock prices below, including the daily returns and the squared daily return. Although only the previous nine days are displayed, the actual horizon includes 55 trading days. The volatility estimates computed for the previous day, t – 1, are given for each of the GARCH(1, 1), EWMA. Also given are the lambda for the EWMA model and the GARCH parameters:

EWMA Parameter	
λ	0.82

GARCH(1,1) Parameters	
ω	0.00009
α	0.08
β	0.82

			Volatility Estimate	
Period	Price	Return	GARCH(1,1)	EWMA
t	$84.37	-10%		
t - 1	$93.24	-0.06%	4.3667%	6.1528%

Exercise 4

- Today's price drop contributes a dramatic -10% return to the series. Which are nearest to the updated volatility estimates given by, respectively, GARCH(1,1) and EWMA?
  A. GARCH = 4.73% and EWMA = 4.9%
  B. GARCH = 4.95% and EWMA = 7%
  C. GARCH = 5.25% and EWMA = 5.25%
  D. GARCH = 5.25% and EWMA = 5.74%

- Correct Answer : B

Exercise 5

- The parameters of a GARCH(1,1) model are $\omega = 0.00003$, $\alpha = 0.04$, and $\beta = 0.92$. These figures imply a long-run daily standard deviation of:
  - A. 1.68%
  - B. 1.55%
  - C. 1.45%
  - D. 2.74%

- Correct Answer : D

Correlations and Copulas

Copula

- **Drawbacks of using correlation to measure dependence.**
  - Correlation is a good measure of dependence when random variables are distributed as multivariate elliptical (e.g., normal, student's).
  - However, correlation is only defined if variance is finite. There might be a problem for **non-elliptical** distributions. e.g. Levy distribution can have infinite variance.
  - Correlation measures the linear relationship of two variables. If risks are independent → zero correlation, however zero correlation does not imply independence.

Copula

Introduction of Copula

- Mathematically, a "copula" is a function which allows us to combine univariate distributions to obtain a joint distribution with a particular dependence structure.
- Formally, the copula is a function of the marginal distributions F (x), plus some parameters, θ, that are specific to this function (and not to the marginals).

$$f_{12}(x_1, x_2) = f_1(x_1) \times f_2(x_2) \times c_{12}[F_1(x_1), F_2(x_2); \theta]$$

Tail Dependence

- **Tail dependence:**
  - Tail dependence is an important issue because extreme events are often related (i.e., disasters often come in pairs or more)
  - If marginal distributions are continuous, we can define a coefficient of (upper) tail dependence of X and Y as the limit, as $\alpha \to 1$ from below, of

$$\Pr\left[Y > F_y^{-1}(\alpha) \mid Y > F_x^{-1}(\alpha)\right] = \lambda$$

Simulation Modeling

Disadvantages of the Simulation Approach

- It might be computationally expensive.
- The results might not be precise.
- The results are often hard to replicate.
- Simulation results are experiment-specific.

Variance Reduction Techniques

- **Sampling Variation**
  - The sampling variation in a Monte Carlo study is measured by the standard error estimate, denoted S_x:

$$\boxed{S_x = \sqrt{\frac{\text{var}(x)}{N}}}$$

    ✓ where var(x) is the variance of the estimates of the quantity of interest over the N replications.

- **Antithetic Variates**
- **Control Variates**
- **Random Number Re-Usage across Experiments**

Bootstrapping Method

- Bootstrapping is used to obtain a description of the properties of empirical estimators by using the sample data points themselves, and it involves sampling **repeatedly with replacement from the actual data.**
- The advantage of bootstrapping over the use of analytical results is that it allows the researcher to make inferences without making strong distributional assumptions, since the distribution employed will be that of the actual data.
- Situations where the bootstrap will be ineffective:
  - Outliers in the Data
  - Non-Independent Data: Use of the bootstrap implicitly assumes that the data are independent of one another.

Random Number Generation

- Most econometrics computer packages include a random number generator.
  - The simplest class of numbers to generate are from a uniform (0,1) distribution. Computers generate continuous uniform random number draws. The initial value is called the seed.
- Computer-generated random number draws are known as **pseudo-random numbers**, since they are in fact not random at all, but entirely deterministic, since they have been derived from an exact formula.

Exercise

- The 95% confidence interval for the output of ending capital is calculated to be ($117.03, $122.97) for a simulation run with 100 scenarios. In addition, the simulation resulted in a mean ending capital amount of $120 with a standard deviation of $15. Suppose we want to improve the accuracy of this confidence interval by running a simulation of 400 scenarios. What is the new 95% confidence interval with a simulation of 400 scenarios using the same mean and standard deviations from the model with 100 scenarios?

  A. ($117.23, $122.95)

  B. ($118.52, $121.48)

  C. ($119.02, $121.99)

  D. ($119.71, $122.27)

- Correct Answer : B

It's not the end but just beginning.

By training your thoughts to concentrate on the bright side of things, you are more likely to have the incentive to follow through on your goals. You are less likely to be held back by negative ideas that might limit your performance.

试着训练自己的思想朝好的一面看，这样你就会汲取实现目标的动力，而不会因为消极沉沦停滞不前。