FRM一级核心知识点

数量分析

101% Contribution Breeds Professionalism

Conditional independence

Conditional independence

- **Independent events(or unconditional independent events):**
  - If the event (B) **is not influenced by** whether the other event (A) occurs, then we say those events are independent, otherwise they are dependent.

$$P(A \cap B) = P(A) \times P(B)$$

- **Conditional independence**
  - Like probability , independence can be redefined to hold conditional on another event(C) , two events A and B are **conditionally independent** if :

$$P(A \cap B|C) = P(A|C) \times P(B|C)$$

  - Note that two types of independence—unconditional and conditional—do not imply each other.
    - ✓ Events can be both unconditionally dependent and conditionally independent.
    - ✓ Events can be independent, yet conditional on another event they may be dependent.

Bayes’ rule

Bayes’ rule

- Specifically, we are often interested in the probability of an event happening only if another event happens first.
- **Bayes’ rule** provides a method to construct conditional probabilities using other probability measures.
  - It is both a simple application of the definition of conditional probability and an extremely important tool.

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

posterior probability **prior probability**

Bayes' rule

- **Example**
  - 一个人患心脏病的概率是3%，没有得病的概率是97%。在患病的情况下机器诊断出患病的概率是99%；在没有得病的情况下机器诊断出患病的概率是5%。若机器诊断出患病的情况下人真的得病的概率是多少?

	机器说有病	机器说没病
如果人真有病	0.99	0.01
如果人真没病	0.05	0.95

Bayes' rule

- **Solution 1:**
  - Assume
    - ✓ Event A=人患病
    - ✓ Event A^C=人不患病
    - ✓ Event B=机器诊断出人患病
    - ✓ Event B^C=机器诊断出人不患病

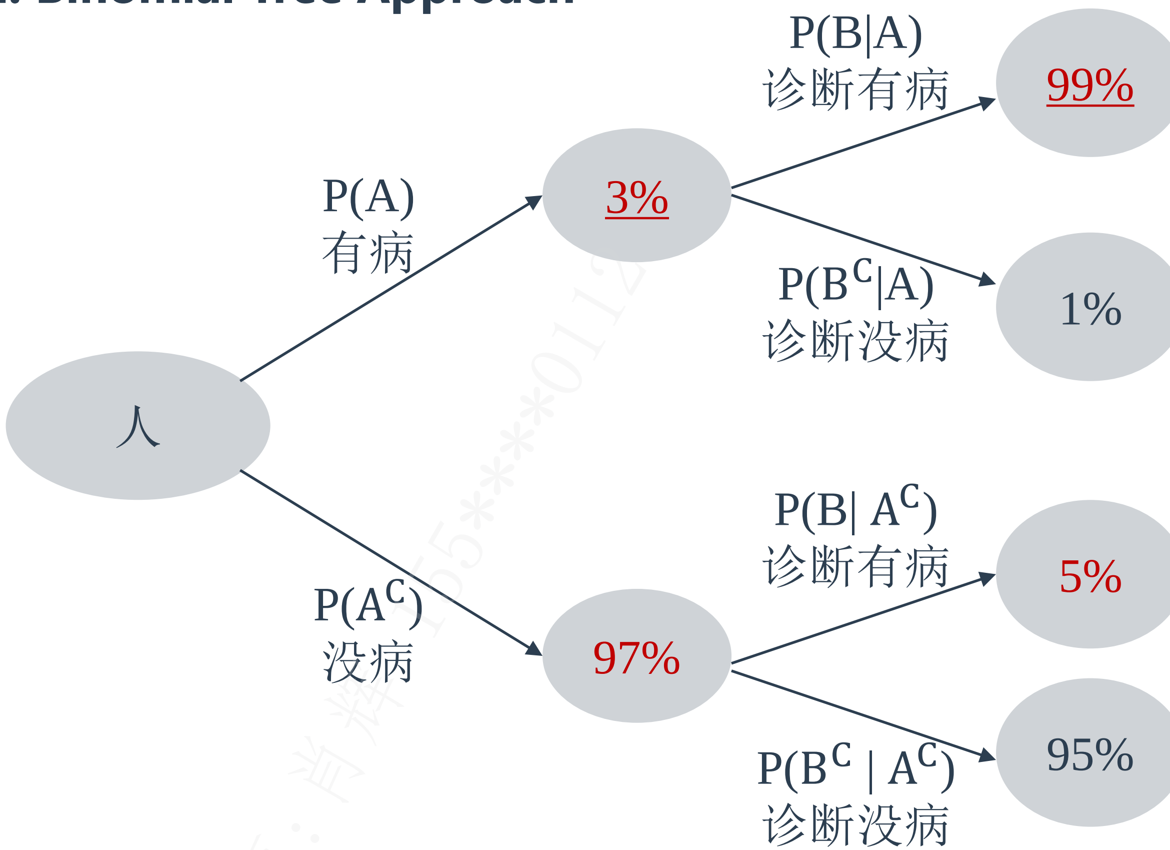
$$P(A \cap B) = P(B|A) \times P(A) = 99\% \times 3\%$$

$$P(B) = P(B|A) \times P(A) + P(B|A^C) \times P(A^C) = 99\% \times 3\% + 5\% \times 97\%$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 37.98\%$$

4.Bayes’ rule

- **Solution 2: Binomial Tree Approach**



$$P(A|B) = \frac{99\% \times 3\%}{99\% \times 3\% + 97\% \times 5\%} = 37.98\%$$

PMF & CDF

PMF & CDF

- **Probability mass function (PMF)**
  - The function returns the probability that a **random variables take a certain value**.

$$f_X(x) = \mathrm{P}(\mathrm{X} = x)$$

  - The value returned from PMF must be non-negative.
  - The sum of across all values in the support of a random variable must be one.

PMF & CDF

- **Cumulative distribution function (CDF)**
  - Measures the total probability of observing a value less than or equal to the input x.
  - $F_X(x) = \mathrm{P}(\mathrm{X} \leq x)$
    - ✓ F(X) is a non-decreasing function such that if $x_2 > x_1$, then $F(x_2) \geq F(x_1)$.
    - ✓ $P(X > k) = 1 - F(k)$
    - ✓ $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$
- **The relationship between PMF and CDF**
  - Can always be expressed as the sum of PMF for all values in support that are less than or equal to x: $\mathrm{F_X}(x) = \sum_{\mathrm{t} \leq \mathrm{x}} \mathrm{f_X}(\mathrm{t})$

PMF & CDF

- Suppose X is a random variable defined by the roll of a fair die and x is the result of a single roll. Please express the probability mass function and Cumulative distribution function for X.

- **Correct Answer:**
  - The PMF of X can be equivalently expressed using a list of values

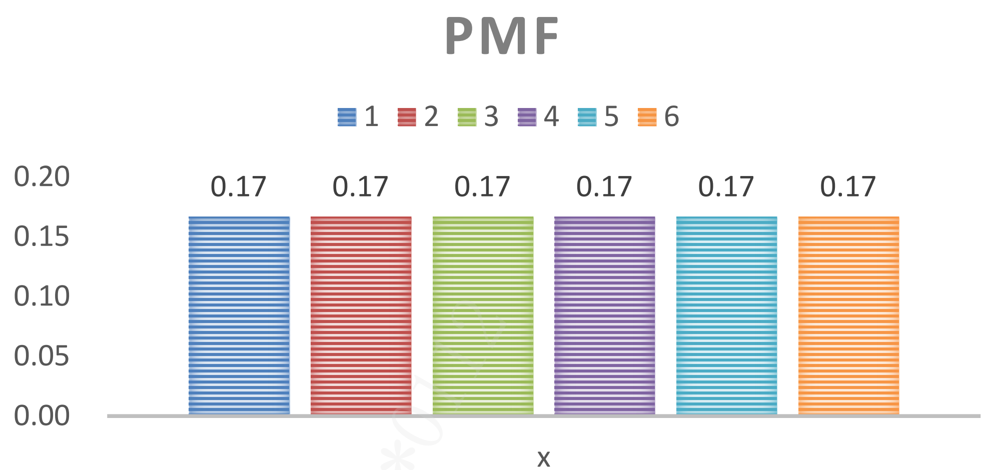
$$f_X(x) = f(x) = \begin{cases} 1/6, & x \in [1,2,\dots,6] \\ 0, & x \notin [1,2,\dots,6] \end{cases}$$

  - The counterparty of PMF is the cumulative distribution function

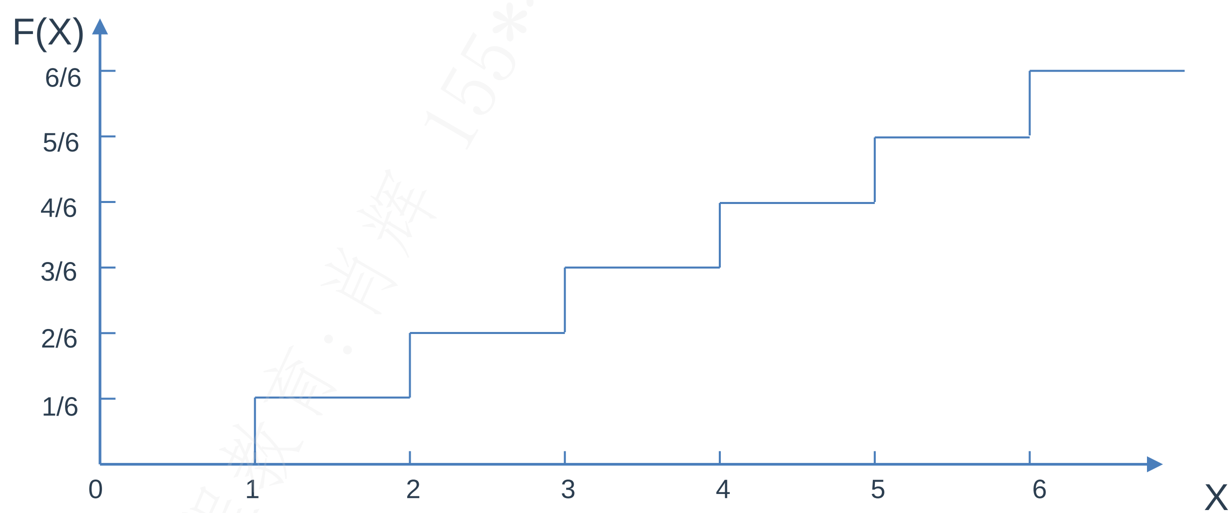
$$F_X(x) = \begin{cases} 0, & x < 1 \\ \dfrac{x}{6}, & x \in [1,2,\dots,6] \\ 1, & x > 6 \end{cases}$$

PMF & CDF

- PMF



- CDF

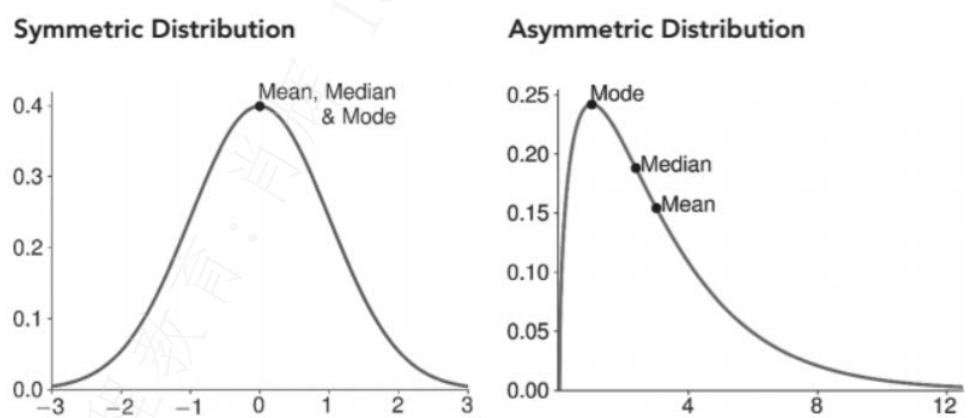


Quantiles and modes

Quantiles and modes

➢ **Mean, Median and Mode**

- **Mean**: the average value of random variable X and is also referred to as the location of distribution.
  - ✓ Very sensitive to large outliers
- **Median**: the middle number or 50%-quantile of an random variable
- **Mode**: the random variables that occur most frequently
  - ✓ random variables may have one or more modes



Binomial Distribution

Binomial Distribution

- **Binomial Distribution:** The probability of x successes in n trails
  - A binomial random variable measures the total number of successes from n independent Bernoulli random variables
  - PMF

$$f_X(x) = C_n^x p^x (1-p)^{n-x} = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

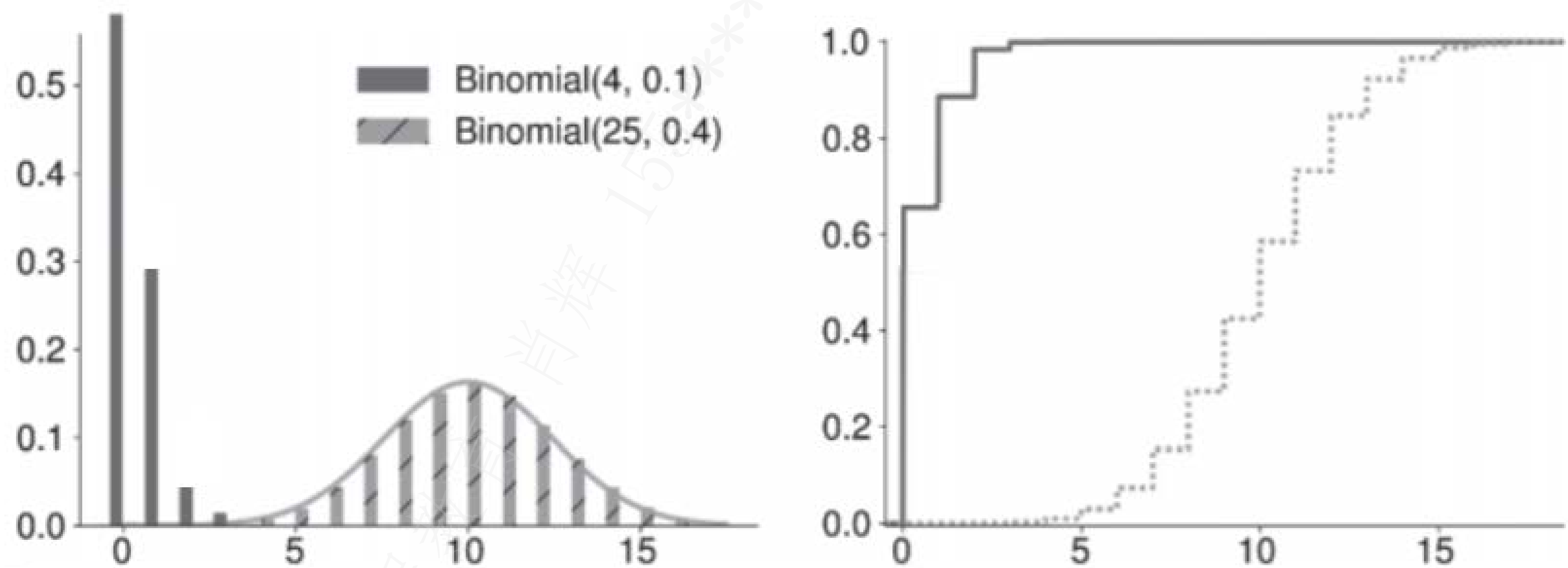
  - Expectation and variance

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

Binomial Distribution

➢ PMF Graph and CDF Graph for binomial Distribution

- The left panel shows the PMF of two binomial distributions. The solid line is the PDF of a normal random variable, N(np, np(1- p)) = N(10, 6), that approximates the PMF of the B(25, 0.4). The right panel shows the CDFs of the two binomials.



Poisson Distribution

Poisson Distribution

- **Poisson Distribution:** used to measure counts of events over fixed time spans. Such as the number of hurricanes in a century, or the number of phone calls in a day.
  - PMF

$$f_X(X = k) = \frac{(\lambda)^k}{k!} e^{-\lambda}$$

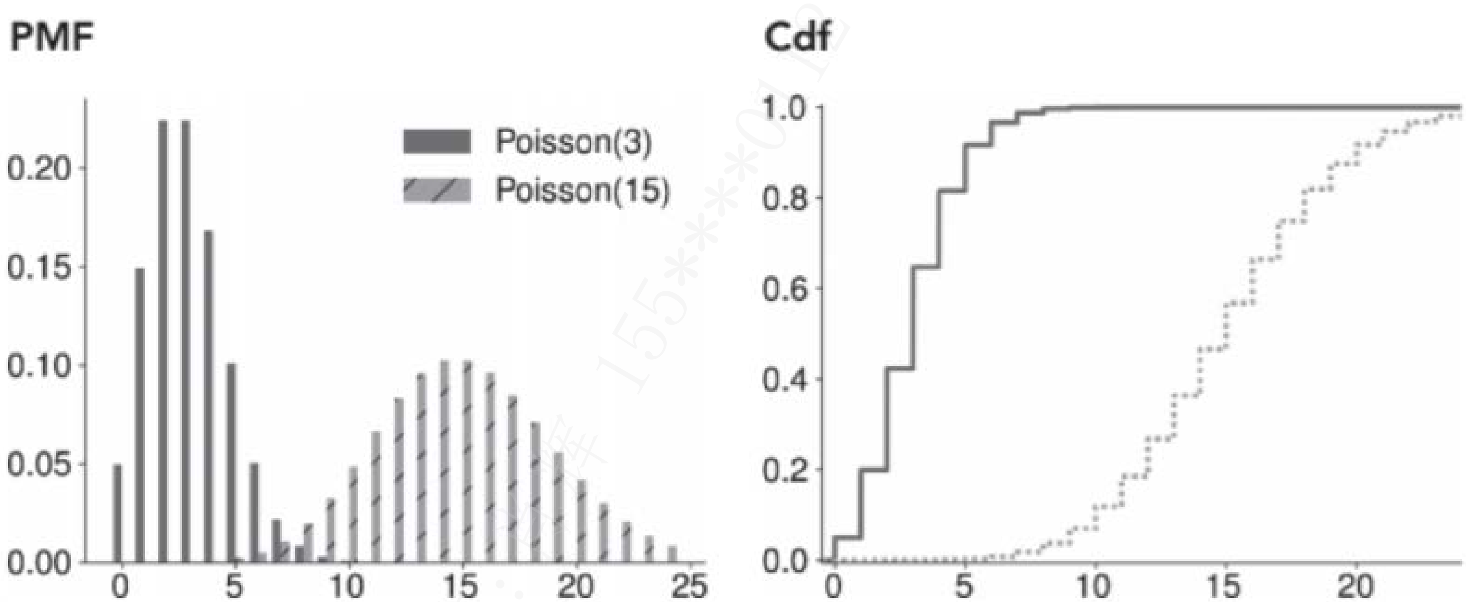
    - ✓ λ indicates average number of the occurrence per interval, also called **hazard rate**
    - ✓ If $X_1 \sim \text{Poisson}(\lambda_1)$, and $X_2 \sim \text{Poisson}(\lambda_2)$ are **independent**, and $Y = X_1 + X_2$, then $Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$.
  - Expectation and variance

	Expectation	Variance
Poisson random variable	λ	λ

Poisson Distribution

➢ PMF Graph and CDF Graph for Poisson Distribution

- The left panel shows the PMF for two values of λ, 3 and 15. The right panel shows the corresponding CDFs.



Poisson Distribution

1. A call center receives an average of two phone calls per hour. The probability that they will receive 20 calls in an 8-hour day is closest to:
   A. 5.59%
   B. 16.56%
   C. 3.66%
   D. 6.40%

➢ **Correct Answer： A**

- To solve this question, we first need to realize that the expected number of phone calls in an 8-hour day is 16. Using the Poisson distribution, we solve for the probability that X will be 20.

$$P(X = 20) = \frac{16^{20}e^{-16}}{20!} = 5.59\%$$

Normal Distribution

Normal Distribution

➢ **Normal (Gaussian, bell shaped）Distribution:** the most commonly used distribution in risk management.

- **PDF(bell shaped)**

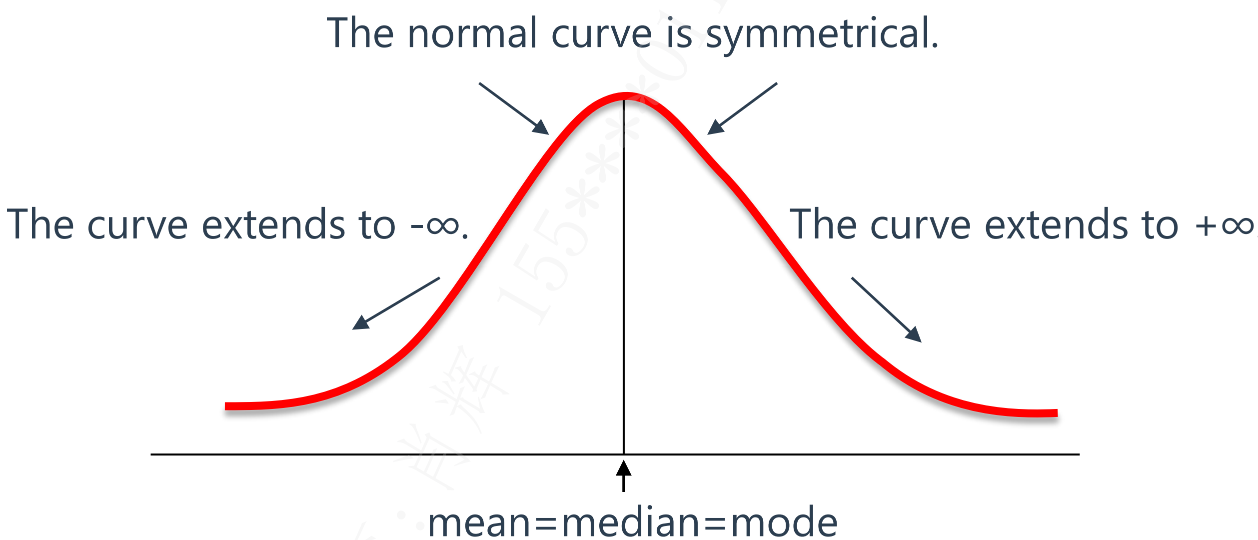
$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

- **CDF**: There is no closed form of normal distribution, fast numerical approximations are widely applied in practice.
- Expectation and variance

	Expectation	Variance
Normal random variable	μ	σ^2

Normal Distribution

- **PDF Graph for Normal Distribution**
  - Symmetrical distribution: **skewness = 0; kurtosis = 3**.
  - The tails get thin and go to zero but extend infinitely.



Normal Distribution

➢ **Normal distribution in practice**

① Sums of independent normally distributed random variables are also normally distributed.

② Standardization and Z-table Application

③ Key quantiles in normal distribution

④ Approximating discrete random variables to normal random variable

Normal Distribution

① **Sums of independent normally distributed random variables are also normally distributed**

- If $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ and they are independent, then

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

- When applied to portfolio, a and b are usually asset weights

➢ **Example:**

- A \$50 million prudent fund (PF) is merged with a \$200 million aggressive fund (AF). The return of PF~N(0.03, 0.07^2) and the return of AF~ N(0.07, 0.15^2). Assuming the returns are independent, what is the distribution of the portfolio return?
- Correct Answer: $R_p \sim N(0.062, 0.1208^2)$

Normal Distribution

1. Which of the following statement about the normal distribution is not accurate?

   A. Kurtosis equals three.

   B. Skewness equals one.

   C. The entire distribution can be characterized by two moments, mean and variance.

   D. The normal density function has the following expression:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]}$$

➢ **Correct Answer：B**

Lognormal Distribution

Lognormal Distribution

- **Lognormal Distribution:** A variable Y is said to be log-normally distributed if the natural logarithm of Y is normally distributed.
  - An important property of the log-normal distribution is that it can never be negative. E.g.
    - ✓ The **Black-Scholes Model** assumes that the **price** of the underlying asset is lognormally distributed.
  - If lnX is normal, then X is lognormal; if a variable is lognormal, its natural log is normal.
  - PDF:

$$f_Y(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} exp(-\frac{(lny-\mu)2}{2\sigma^2})$$

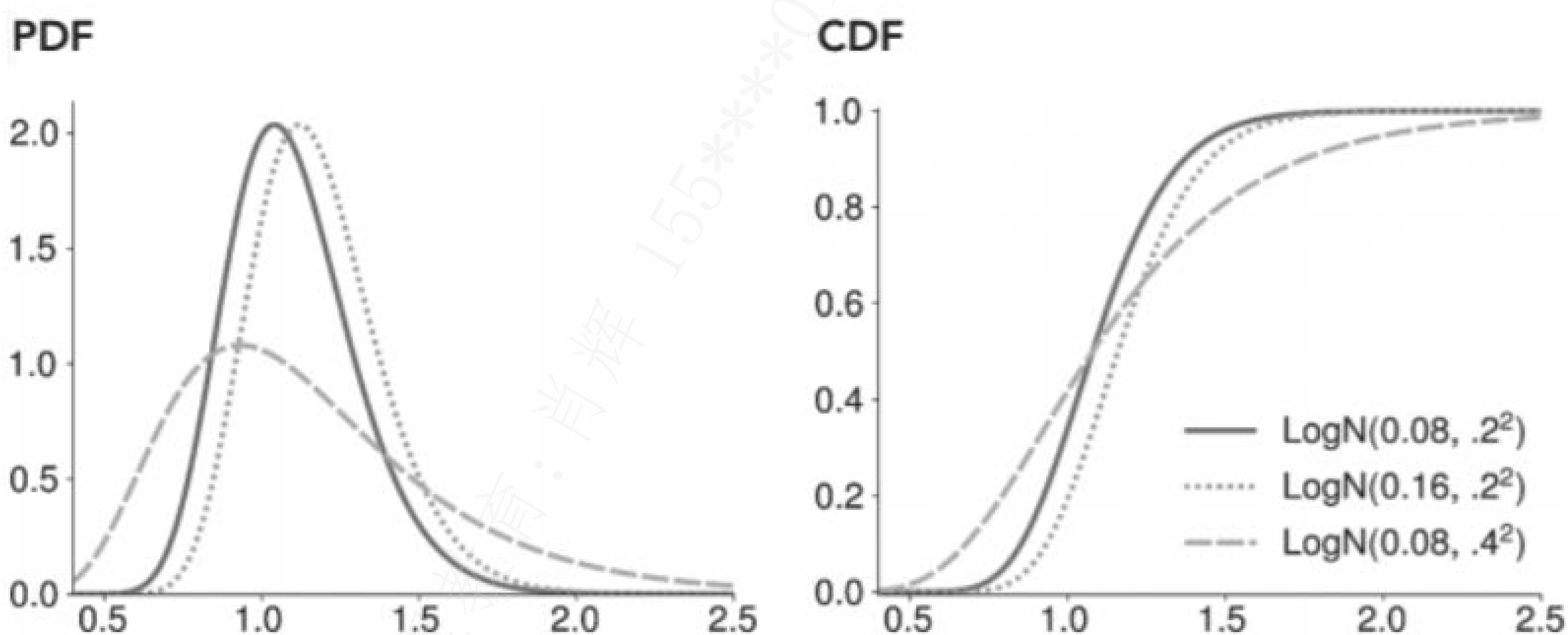
  - Expectation and variance*

	Expectation	Variance
lognormal random variable	$e^{\mu+\sigma^2/2}$	$(e^{\sigma^2}-1)\ e^{2\mu+\sigma^2}$

Lognormal Distribution

➢ **PDF Graph and CDF Graph for lognormal Distribution**

- The left panel shows the PDFs of three log-normal random variables, two of which have $\mu = 8\%$, and two with $\sigma = 20\%$, which are typical of annual equity returns. The right panel shows the corresponding CDFs.



Student's t distribution

Student's t distribution

- **t Distribution (Student's t distribution):** The Student's t distribution is closely related to the normal, but it has heavier tails.
  - Developed for testing hypotheses using **small samples**.
  - If Z is a standard normal variable and U is a chi-square variable with k degrees of freedom, then the random variable X follows a t-distribution with k degrees of freedom.

$$X = \frac{Z}{\sqrt{\frac{U}{K}}}$$

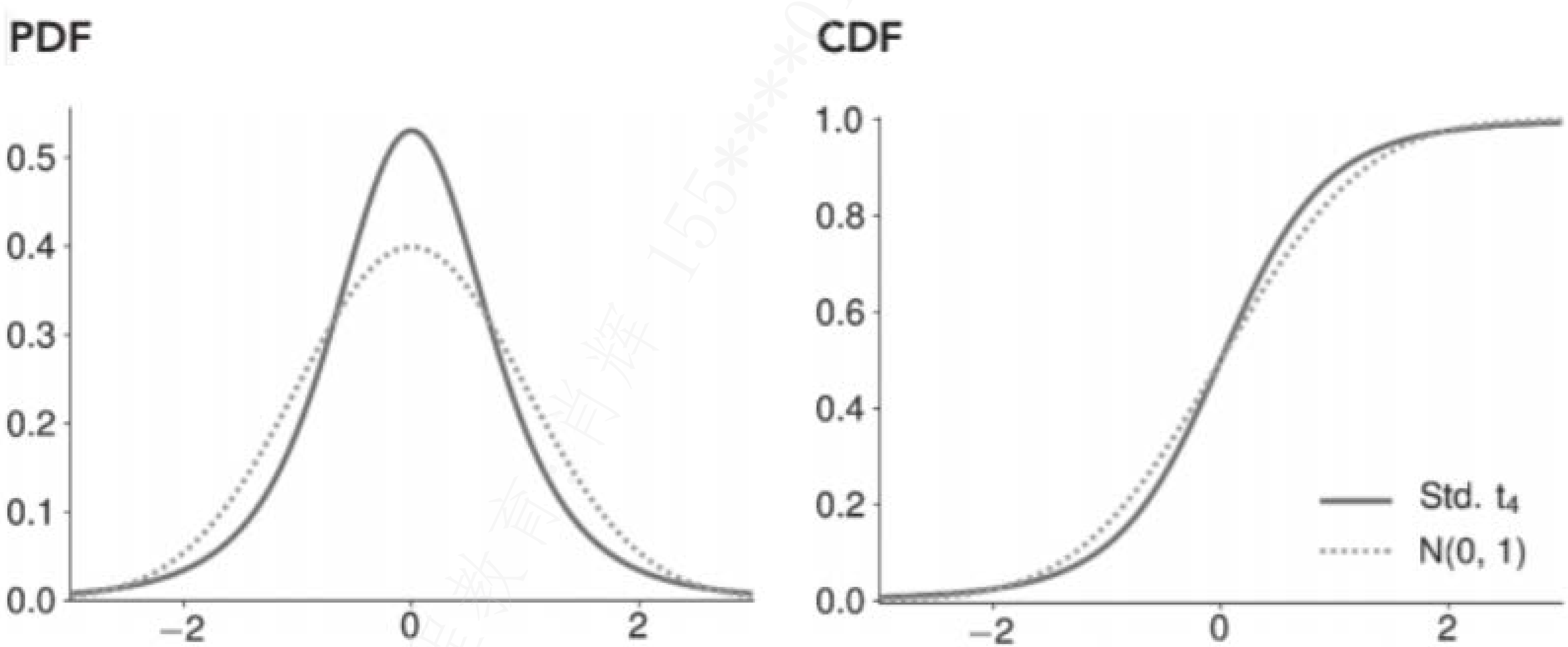
  - Expectation, variance and kurtosis(Variance>=1, Kurtosis>=3)

	Expectation	Variance	Kurtosis*
Student's t random variable	0	k/k-2	3(k-2)/(k-4)

Student's t distribution

➢ **PDF Graph and CDF Graph for Student's t Distribution**

- The left panel shows the PDF of a generalized Student's t with four degrees of freedom and the PDF of a standard normal. The right panel shows the corresponding CDFs.



Probability Matrix

Probability Matrix

- Multivariate random variables extend the concept of a single random variable to include measures of dependence between two or more random variables.
  - like their univariate counterparts, they can be **discrete or continuous**.
- **Probability matrix(probability table):** When dealing with the joint probabilities of two variables, it is often convenient to summarize the various probabilities in a probability matrix.
  - **Example**: We take X from 1 or 2 with the same probability. We take Y from [1,X] with the same probability.

Y	X	
	1	2
1	0.50	0.25
2	0.00	0.25

  - The probability in green blanks are **joint probabilities**

Probability Matrix

- **Marginal distributions:**
  - The distribution of a single component of a bivariate random variable is called a marginal distribution.
  - When a PMF is represented as a probability matrix, the two marginal distributions are computed by either summing across columns or summing down rows.
    - ✓ E.g. The probability in yellow blanks are **marginal probabilities**
    - ✓ The yellow column and row are the marginal distribution of Y and X

Y	X		
	1	2	$f_Y(y)$
1	0.50	0.25	0.75
2	0.00	0.25	0.25
$f_X(x)$	0.50	0.50	

Probability Matrix

- **Conditional Distributions:**
  - The conditional distribution summarizes the probability of the outcomes for one random variable conditional on the other taking a specific value.
    - ✓ E.g. What is the conditional distribution of Y condition on X=2

$$f_{Y|X}(y|x=2)=\frac{f_{X,Y}(2,y)}{f_X(2)}=\frac{f_{X,Y}(2,y)}{0.5}$$

    - ✓ The **conditional distribution** is then

Y	
1	2
50%	50%

Covariance and Correlation

3.Covariance and correlation

- **Covariance**
  - Covariance measures how one random variable moves with another random variable.

$$\mathrm{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

✓ Unless the zero covariance, the expectation function is non-multiplicative.

- Covariance ranges from negative infinity to positive infinity.

✓ A positive covariance indicates positive relationship

✓ A negative covariance indicates inverse relationship

✓ A zero covariance indicates no discernible relationship

Covariance and Correlation

Properties of Covariance

- If X and Y are independent random variables, their covariance is zero.
- The covariance of X and itself is the variance of x
  - ✓ $\mathrm{Cov}(X,\ X) = E\big[\big(X - E(X)\big)\big(X - E(X)\big)\big] = \sigma^2(X)$
  - ✓ $\mathrm{Cov}(X,\ X) = \sigma_{xx}$, $\mathrm{Cov}(X,\ Y) = \sigma_{xy}$
- If a, b and c are constant, then
  - ✓ $\mathrm{Cov}(a + bX,\ cY) = \mathrm{Cov}(a, cY) + \mathrm{Cov}(bX, cY) = b \times c \times \mathrm{Cov}(X, Y)$
- The relationship between covariance and variance:
  - ✓ $\sigma^2_{X\pm Y} = \sigma^2_x + \sigma^2_y \pm\ 2\mathrm{Cov}(X, Y)$
  - ✓ $\sigma^2_{aX\pm bY} = a^2\sigma^2_x + b^2\sigma^2_y \pm\ 2ab\mathrm{Cov}(X, Y)$

Covariance and Correlation

- **Correlation coefficient**
  - The covariance depends on the scales of X and of Y. It is therefore more common to report the correlation, which is a **scale-free measure**.

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

- **Properties of Correlation coefficient**
  - Correlation has no units, ranges from -1 to +1.
  - Variances of correlated variables:

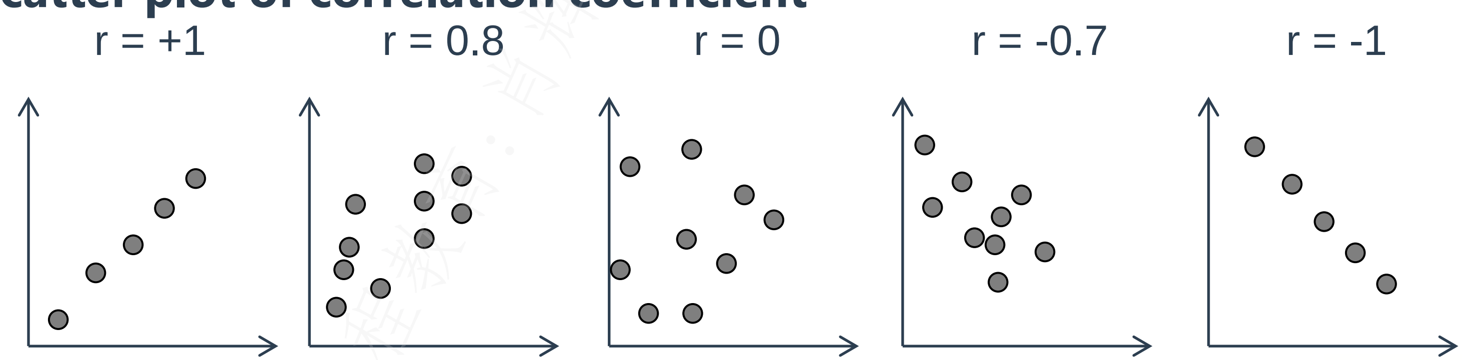
$$\sigma^2_{X \pm Y} = \sigma^2_X + \sigma^2_Y \pm 2\rho_{XY}\sigma_X\sigma_Y$$

Covariance and Correlation

- **Correlation and independence:** Correlation measures the linear relationship between two random variables.
  - If two variables are independent, the correlation coefficient will be zero. The converse, however, is not necessary. For example, $Y = X^2$.
  - If the correlation between two variables is 0, the **linear relationship** must not exist, and the expectation of the product of the two is the product of the expectations.

$$E(XY) = E(X)E(Y)$$

- **Scatter plot of correlation coefficient**



Covariance and Correlation

Which one of the following statements about the correlation coefficient is false?

A. It always ranges from -1 to +1.

B. A correlation coefficient of zero means that two random variables are independent.

C. It is a measure of linear relationship between two random variables.

D. It can be calculated by scaling the covariance between two random variables.

➢ Correct Answer: B

Four sample moments

Four sample moments

➢ **Sample mean**

- Assumed the random variables X_i are i.i.d. and $E[X_i] = \mu$ and $V[X_i] = \sigma^2$.
- When population mean μ are not observable , it is estimated using the sample mean estimator, $\hat{\mu}$, or $\overline{x_i}$ in mathematical expression.

$$\hat{\mu} = \overline{x_i} = \frac{\sum_{i=1}^{n} x_i}{n}$$

✓ In this case , $\hat{\mu}$ is an estimator of the unknown population parameter μ .

Four sample moments

- **Features of sample mean**
  - The mean estimator is a function of random variables, and so it is also a random variable.
    - ✓ The mean estimator is **unbiased** because the expected value of the mean estimator is the same as the population mean

$$E[\hat{\mu}] = \frac{1}{n} \times \sum_{i=1}^{n} E[x_i] = \mu$$

    - ✓ The variance of the mean estimator decreases as the number of observations increases, and so larger samples are better to estimate population mean.

$$V[\hat{\mu}] = \frac{1}{n^2} \times \sum_{i=1}^{n} V[x_i] = \sigma^2/n$$

Four sample moments

- **Sample Variance**
  - Similarly as $\hat{\mu}$, sample variance is estimated using the sample variance estimator, denoted by $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

    - ✓ Unlike the sample mean estimator, sample variance estimator is biased

$$E[\hat{\sigma}^2] = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

- **The unbiased variance estimator**

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = \frac{n}{n-1}\hat{\sigma}^2$$

  - The expression (n – 1) is known as the **degrees of freedom**.
  - The sample variance estimator, S^2, is unbiased , as $E[S^2] = \sigma^2$.
- **Sample standard deviation**
  - S or $\hat{\sigma}$ (the positive square root of S^2 or $\hat{\sigma}^2$), is called the **sample standard deviation estimator**.
  - Both estimator are biased because square root is a nonlinear transformation.

Four sample moments

- **Estimators for the skewness**
  - The estimators for the skewness is :

$$\widehat{S}(\mathrm{X}) = \frac{E(X-\hat{\mu})^3}{\hat{\sigma}^3}$$

  - $\mathrm{E}(\mathrm{X}-\hat{\mu})^3$ is the estimate of the **third central moment** and is defined as :

$$E(X-\hat{\mu})^3 = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{x_i}-\hat{\mu})^3$$

- **Estimators for the kurtosis**
  - The estimators for the kurtosis is :

$$\widehat{\mathrm{K}}(\mathrm{X}) = \frac{E(X-\hat{\mu})^4}{\hat{\sigma}^4}$$

  - $\mathrm{E}(\mathrm{X}-\hat{\mu})^4$ is the estimate of the **fourth central moment** and is defined as :

$$E(X-\hat{\mu})^4 = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{x_i}-\hat{\mu})^4$$

Four sample moments

- A sample estimator is used to estimate the parameter of population, the **best** estimator should be **unbiased, and linear(BLUE).**
  - **Linear**: estimator is be calculated as a linear function of the sample data.
  - **Unbiased**: the expected value of the estimator should equal to the parameter of population.
  - **Best**: If the estimator has the minimum variance(Best) among all the other linear unbiased estimators.
- The **mean estimator** is the Best Linear Unbiased Estimator (BLUE) of the population mean <u>when the data are i.i.d.</u>.

LLN and CLT

LLN and CLT

➢ **Law of Large Numbers (LLN)**

- The Kolmogorov proved **Strong Law of Large Numbers** which states that if $\{X_i\}$ is a sequence of i.i.d. random variables with $E[X_i] \equiv \mu$(a finite parameter, μ), then :

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu$$

✓ The symbol $\xrightarrow{a.s.}$ means **converges** almost surely.

✓ The mean in large samples $\hat{\mu}_n$ (formally , when $n \rightarrow \infty$) is approximate to the true population mean μ.

LLN and CLT

- **Consistency**: The feature of LLN applied estimator is called consistent.
  - **Consistency** requires that the **accuracy** of the parameter estimate increases and so many finite sample bias must diminish as the **sample size increases** .
  - As the sample size n grows larger , the variance of the mean estimator converges to zero .

$$V[\hat{\mu}_n] = \frac{\sigma^2}{n} \to 0$$

  - Consistency ensures the chance of a large deviation from the population value is negligible in large samples.

LLN and CLT

- **Comparisons between BLUE and Consistency**
  - An estimator is BLUE means the true parameter can be approached through infinite sampling.
    - ✓ $\hat{\mu}$ and S^2 are BLUE, while $\hat{\sigma}^2$ is not BLUE.
  - An estimator is **consistency** means the estimator will converge to population value through finite sampling.
    - ✓ $\hat{\mu}$ and S^2 are consistent, besides, $\hat{\sigma}^2$ is consistent.

LLN and CLT

- **The Central Limit Theorem (CLT)**
  - If X_1, X_2, … , X_n, a i.i.d. random sample from any population (**regardless of any distributions)** with a finite mean, μ and finite variance, σ^2 , then the mean estimator, $\hat{\mu}_n$, tends to be normally distributed with mean μ and $\frac{\sigma^2}{n}$ variance, as the sample size increases.

$$\hat{\mu}_n = \bar{X} \xrightarrow{d} N\left(\mu, \frac{\sigma^2}{n}\right)$$

or

$$\left(\frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}}\right) \xrightarrow{d} N(0,1)$$

Hypothesis Testing

Hypothesis Testing

- **A hypothesis test has <u>six distinct components</u>:**
  ① The **null hypothesis**, which specifies a parameter value that is assumed to be true; The **alternative hypothesis**, which defines the range of values where the null should be rejected;
  ② The **test statistic**, which has a known distribution when the null is true;
  ③ The **size of the test**, which captures the willingness to make a mistake and falsely reject a null hypothesis that is true;
  ④ The **critical value**, which is a value that is compared to the test statistic to determine whether to reject the null hypothesis; and
  ⑤ The **decision rule**, which combines the test statistic and critical value to determine whether to reject the null hypothesis.
  ⑥ The **Test Power**, which measures the probability that a false null is rejected.

Hypothesis Testing

① **The null and alternative hypothesis**

- The null hypothesis (H_0) is a statement about the population values of one or more parameters. Usually two types,
  - ✓ Consistent with something we want to reject.
    - ◆ E.g., when considering whether to invest in a mutual fund, the natural null hypothesis is that the fund does not generate an abnormally high return.
  - ✓ Testing the accuracy of an econometric model.
    - ◆ E.g., when testing a VaR model, the null hypothesis is that the model is correct.

Hypothesis Testing

① **The null and alternative hypothesis**

- Alternative hypothesis (H_1), which determines the values of the population parameter where the null hypothesis <u>should be rejected</u>.
  - ✓ Two-Sided Alternative Hypotheses
    - ◆ H_0: $\mu = 0$ H_1: $\mu \neq 0$
  - ✓ One-Sided Alternative Hypotheses
    - ◆ H_0: $\mu \geq 0$ H_1: $\mu < 0$
    - ◆ H_0: $\mu \leq 0$ H_1: $\mu > 0$

Hypothesis Testing

② **Test statistic**

- The test statistic is a summary of the observed data that has a known distribution when the null hypothesis is true.
- Test statistics take many forms and follow a wide range of distributions. This chapter focuses on test statistics that are normally distributed.
- E.g. Consider a test of the null hypothesis about a mean: H_0: $\mu = \mu_0$.
  - ✓ When the true value of the mean (μ) is equal to the value assumed by the null (μ_0), then the asymptotic distribution leads to the test statistic:
  - ✓ The test statistic T (also known as the t-statistic) is asymptotically standard normally distributed according to CLT.

$$T = \frac{\hat{\mu} - \mu_0}{\sqrt{\hat{\sigma}^2/n}} \sim N(0,1)$$

Hypothesis Testing

③ **The size of the test**

- In an ideal world, a false (true) null would always (never) be rejected. However, in practice there is a tradeoff between avoiding a rejection of a true null and avoiding a failure to reject a false null.
- Rejecting a true null hypothesis is called a **Type I error**. The probability of Type I error is referred to **the test size.**
  - ✓ The **test size(α)** is chosen to reflect the willingness to mistakenly reject a true null hypothesis, it is set by the tester.
  - ✓ The most common test size is **5%.** Smaller test sizes (e.g., 1% or even 0.1%) are used when it is especially important to avoid incorrectly rejecting a true null.

Hypothesis Testing

④ The critical value

- The critical value depends on the distribution of the test statistic ② and **defines a range of values** where the null hypothesis should be rejected in favor of the alternative. This range is known as the **rejection region**.
  - ✓ The test size③ and the alternative① are combined with the distribution of the test statistic② to construct the **critical value** of the test.

Hypothesis Testing

④ **The critical value**

- In most cases, test statistic follows **standard normal distribution**.
- When testing against two-sided alternatives, if
  - ✓ $\alpha = 10\%$ ($C_{10\%} = \pm 1.645$)
  - ✓ $\alpha = 5\%$ ($C_{5\%} = \pm 1.96$)
  - ✓ $\alpha = 1\%$ ($C_{1\%} = \pm 2.57$)
- When testing against one-sided alternatives, if
  - ✓ $\alpha = 5\%$ ($C_{5\%} = 1.645$ or -1.645)
  - ✓ $\alpha = 1\%$ ($C_{1\%} = 2.326$ or -2.326)
  - ✓ The sign depends on the alternative hypothesis

Hypothesis Testing

④ **The critical value**

- However, there are **two circumstances** where the Student's t should be used in place of the normal to determine the critical value of a test.
  - ✓ First, when the random variables in the average are i.i.d. normally distributed, then the t-test statistic has an exact Student's t_{n-1} distribution("t-1" refers to the degree of freedom)

$$T = \frac{\hat{\mu} - \mu_0}{\sqrt{s^2/n}} \sim t_{n-1}$$

  - ✓ Second, when n is small (i.e., less than 30), the Student's t has been documented to provide a better approximation than the normal.

Hypothesis Testing

⑤ The decision rule

- The decision rule combines the critical value, the alternative hypothesis, and the test statistic into two decisions:
  - ✓ reject the null in favor of the alternative
  - ✓ to fail to reject the null
- The conclusion of reject the null is more useful in practice
  - ✓ P.S. There is never “accepting” the null.

Hypothesis Testing

⑥ The Test Power

- The **power of a test** measures the probability that a false null is rejected and alternative conclusion is drawn.
  - ✓ In the opposite, a **type II error** occurs when the alternative is true, but the null is not rejected. The probability of a Type II error is denoted by the Greek letter β.
  - ✓ In practice, β should be small so that the power of the test, defined as 1 – β, is high.

Type I and type II error

Type I and type II error

➢ **The summary of type I and type II error**

		Null Hypothesis	
		Ture	False
Decision	Fail to reject	Correct $(1-\alpha)$	Type II error (β)
	Reject	Type I (α)	Correct $(1-\beta)$

- The power of test is high when
  - ✓ The sample size is high
  - ✓ The size of the test is high.

Type I and type II error

1. According to the Basel back-testing framework guidelines, penalties start to apply is there are five or more exceptions during the previous year. The Type I error rate of this test is 11 percent, the power of the test is 87 percent. This implies that there is a(an):
   A. 89% probability regulators will reject the correct model.
   B. 11% probability regulators will reject the incorrect model.
   C. 87% probability regulators will not reject the correct model.
   D. 13% probability regulators will not reject the incorrect model.

- Correct Answer：D

P-Value Approach

P-Value Approach

① P Value Approach

- P value is the **smallest level of significance** for which the null hypothesis can be rejected.
- Return to our P/E example,

$$t=\frac{(\bar{X}-\mu_X)}{S_X/\sqrt{n}}=\frac{(23.25-18.5)}{9.49/\sqrt{28}}=2.65\sim t_{\frac{p}{2}}(27)\rightarrow P\approx 0.015$$

➢ The Rule of P Value

- Reject H_0 if the p-value is less than the significance level of the hypothesis test.
- Do not reject H_0 if the p-value is greater than the significance level.

Inference in linear regression

Inference in linear regression

- The **confidence interval** for the regression coefficient, b_1 and b_0

$$\hat{b}_1 \pm (\text{Critical Value} \times s_{\hat{b}_1})$$

$$\hat{b}_0 \pm (Critical\ Value \times s_{\hat{b}_0})$$

- The standard error of the regression coefficient.

$$S_{\hat{b}_1} = \frac{1}{\sqrt{n}}\frac{s}{\hat{\sigma}_x}$$

$$S_{\hat{b}_0} = \sqrt{\frac{s^2(\hat{\mu}_X^2 + \hat{s}_X^2)}{n\hat{\sigma}_X^2}}$$

Inference in linear regression

- **Regression Coefficient Hypothesis Testing**
  - A t-test may also be used to test the hypothesis that the true slope coefficient b_1, is equal to some hypothesized value, k (usually k=0), the appropriate test procedure is:
    - ✓ State null hypothesis and alternative hypothesis

$$H_0\colon b_1 = 0 \quad H_1\colon b_1 \neq 0$$

    - ✓ Calculate test statistic

$$\mathrm{T} = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}}$$

    - ✓ The test statistic can also be transformed into a p-value,

$$p - value = 2(1 - \Phi|T|)$$

    - ✓ Reject H_0 if T falls into rejection region, or p-value is less than test size.

ANOVA Table

ANOVA Table

Summary output

Regression Statistics	
Multiple R	0.8148004
R Square	0.6638996
Adjusted R Square	0.6390033
Standard Error	5.5120076
Observations	30

ANOVA

	df	SS	MS	F	Significance F
Regression	2	1620.3799	810.18993	26.666575	4.04884E-07
Residual	27	820.32014	30.382227		
Total	29	2440.7			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-140.7652	29.755964	-4.730656	6.282E-05	-201.8194407	-79.71104917
Age	-0.048528	1.1637722	-0.041699	0.9670455	-2.436391407	2.33933526
Height	1.1972821	0.1800553	6.6495241	3.896E-07	0.827839137	1.566725091

ANOVA Table

1. 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.

ANOVA Table

- **Answer: D**
  - $r^2 = R^2 = 92.648/117.160 = 79\%$, the variability of industry returns explains 79% of the variation of company.
  - t-stat (industry index) = $1.9/0.31 = 6.13$, so the coefficient of industry index is significant.
  - $R = 2.1\% + 1.9 \times R(\text{industry index}) = 2.1\% + 1.9 \times 4\,\% = 9.7\%$

Covariance Stationary

Covariance Stationary

- **Covariance Stationary:** A time series is covariance stationary if its first two moments satisfy three key properties**.**
  - The mean is constant and does not change over time.
    - ✓ $E[Y_t] = \mu$ for all t
  - The autocovariance is finite, does not change over time, and <u>only depends on the distance between observation h</u>
    - ✓ $Cov[Y_t, Y_{t-h}] = \gamma_h$ for all t
  - The variance is finite and does not change over time(**specially when h=0**)
    - ✓ $V[Y_t] = \gamma_0 < \infty$ for all t

Covariance Stationary

- **Autocovariance**
  - Autocovariance is defined as the covariance between a stationary time series at different points in time.
    - ✓ The h^{th} autocovariance is defined as:

$$\gamma_{t,h} = E[(Y_t - E(Y_t))(Y_{t-h} - E(Y_{t-h}))]$$

    - ✓ When h=0 , then $\gamma_{t,0}$ is the variance of Y_t

$$\gamma_{t,0} = E[(Y_t - E(Y_t))^2]$$

- The **autocovariance function** is a function of h that returns the autocovariance between Y_t and Y_{t-h}

$$\gamma(h) = \gamma_{|h|}$$

  - The **symmetry** follows from the third property of covariance stationary , because the autocovariance depends only on h :

$$\mathrm{Cov}[Y_t, Y_{t-h}] = \mathrm{Cov}[Y_{t+h}, Y_t]$$

White Noise

White Noise

- **White noise**
  - White noise is the **fundamental building block** of any time-series model to model shocks, which means that all the stationary time series models are established on white noise process.

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

  - White noise processes is covariance-stationary because of three properties
    - ✓ Mean zero
    - ✓ Constant and finite variance, σ^2
    - ✓ No autocorrelation or autocovariance, $\gamma(h) = \rho(h) = 0 \ for \ h \neq 0$
  - The **lack of correlation** is the essential characteristic of a white noise process.

White Noise

- **Differences** between error term in regression model and white noise process in time series model
  - In regression model, error term follows a normal distribution
  - In time series model, the white noise process follows any distribution
- **Similarities** between error term in regression model and white noise process in time series model, they both has
  - A mean of zero
  - Constant variance
  - No autocorrelation

White Noise

➢ **Independent white noise**

- If ε is serially independent, then we say that ε is **independent white noise.**

$$\varepsilon_t \overset{i.i.d.}{\sim} (0, \sigma^2)$$

- Note that white noise itself has no autocorrelation, but not necessarily to be independent.

➢ **Gaussian white noise(normal white noise)**

- A special case of i.i.d. noise , where $\varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2)$