2020 FRM Part I

百题巅峰班

定量分析

2020 年 2 月

2. Quantitative Analysis

2.1. Key Point: Conditional Probability

2.1.1. 重要知识点

2.1.1.1. Conditional Probability:

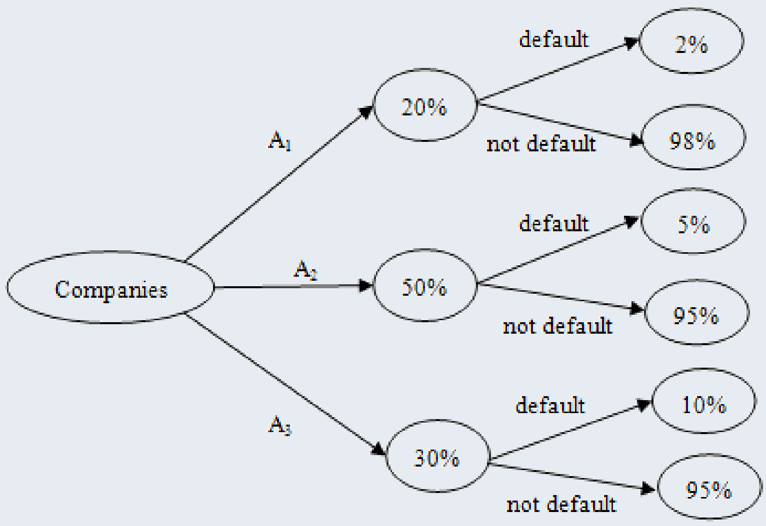
$$P(A|B) = \frac{P(AB)}{P(B)} \quad ; \quad P(B) > 0$$

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

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)$$

$$P(A_1|B) = \frac{P(B|A_1)}{P(B)} \times P(A_1)$$

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



2.1.2. 基础题

Q-1. Suppose there are two events A and B. The probability of A occurrence equals that of B. P(AB) = 4%, If event A occurred, the probability of B occurs is 80%. What is the probability of neither occurs?

A. 86%

B. 90%

C. 94%

D. 96%

Q-2. An analyst develops the following probability distribution about the state of the economy and the market.

Initial Probability P(A)	Conditional Probability P(B |A)

Good economy 60%	Bull market 50%
	Normal market 30%
	Bear market 20%
Poor economy 40%	Bull market 20%
	Normal market 30%
	Bear market 50%

Which of the following statements about this probability distribution is least likely accurate?

A. The probability of a normal market is 0.30.

B. The probability of having a good economy and a bear market is 0.12.

C. Given that the economy is good, the chance of a poor economy and a bull market is 0.15.

D. Given that the economy is poor, the combined probability of a normal or a bull market is 0.50.

Q-3. In country X, the probability that a letter sent through the postal system reaches its destination is 2/3. Assume that each postal delivery is independent of every other postal delivery, and assume that if a wife receives a letter from her husband, she will certainly mail a response to her husband. Suppose a man in country X mails a letter to his wife (also in country X) through the postal system. If the man does not receive a response letter from his wife, what is the probability that his wife received his letter?

A. 1/3

B. 3/5

C. 2/3

D. 2/5

Q-4. An insurance company estimates that 40% of policyholders who have only an auto policy will renew next year, and 60% of policyholders who have only a homeowner policy will renew next year. The company estimates that 80% of policyholders who have both an auto and a homeowner policy will renew at least one of those policies next year. Company records show that 65% of policyholders have an auto policy, 50% of policyholders have a homeowner policy, and 15% of policyholders have both an auto and a homeowner policy. Using the company's estimates, what is the percentage of policyholders that will renew at least one policy next year?

A. 20%

B. 29%

C. 41%

D. 53%

Q-5. An analyst is examining a portfolio that consists of 1,000 subprime mortgages and 600 prime mortgages. Of the subprime mortgages, 200 are late on their payments. Of the prime mortgages, 48 are late on their payments. If the analyst randomly selects a mortgage from the portfolio and it is currently late on its payments, what is the probability that it is a subprime mortgage?

A. 60%

B. 67%

C. 75%

D. 81%

Q-6. Let f(x) represent a probability function (which is called a probability mass function, p.m.f., for discrete random variables and a probability density function, p.d.f., for continuous variables) and let F(x) represent the corresponding cumulative distribution function (CDF); in the case of the continuous variable, F(X) is the integral (aka, anti-derivative) of the pdf. Each of the following is true about these probability functions EXCEPT which is false?

A. The limits of a cumulative distribution function (CDF) must be zero and one; i.e., $\mathrm{F}(-\infty) = 0 \text{ and } \mathrm{F}(+\infty) = 1.0$

B. For both discrete and continuous random variables, the cumulative distribution function (CDF) is necessarily a non-decreasing function.

C. In the case of a continuous random variable, we cannot talk about the probability of a specific value occurring; e.g., $\Pr[\mathrm{R} = +3.00\%]$ is meaningless

D. Bayes Theorem can only be applied to discrete random variables, such that continuous random variables must be transformed into their discrete equivalents

Q-7. For a certain operational process, the frequency of major loss events during a one period year varies form zero to 5.0 and is characterized by the following discrete probability mass function (pmf) which is the exhaustive probability distribution and where (b) is a constant

Loss event (X)	0	1	2	3	4	5
PMF(f(x))	12b	7b	5b	3b	2b	1b

Which is nearest to the probability that next year LESS THAN two major loss events will happen?

A. 5.3%

B. 22.6%

C. 63.3%

D. 75.0%

Q-8. Assume the probability density function (pdf) of a zero-coupon bond with a notional value of \$10.00 is given by $\mathbf{f(x) = x/8 - 0.75}$ on the domain [6,10] where x is the price of the bond:

$$f(x) = \frac{x}{8} - 0.75 \text{ s.t } 6 \leq x \leq 10 \text{ where } x = \text{bond price}$$

What is the probability that the price of the bond is between \$8.00 and \$9.00?

A. 25.750%

B. 28.300%

C. 31.250%

D. 44.667%

Q-9. X and Y are discrete random variables with the following joint distribution; e.g., $\mathbf{P(X = 4, Y = 30) = 0.07}$.

		Value of Y			
		10	20	30	40
Value of X	1	0.04	0.06	0.13	0.04
	4	0.12	0.17	0.07	0.05
	7	0.05	0.03	0.13	0.11

What is the conditional standard deviation of Y given $X = 7$; i.e, Standard deviation(Y|X=7)?

A. 10.3

B. 14.7

C. 21.2

D. 29.4

2.2. Variance and Covariance

2.2.1. 重要知识点

2.2.1.1. Variance and Covariance

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

$\mathbf{Var(X) = E(X^2) - [E(X)]^2}$

$\mathbf{Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y)}$

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

$\mathbf{Cov(ax + by, cx + dy) = ac\sigma_x^2 + bd\sigma_y^2 + (ad + bc)Cov(x, y)}$

2.2.2. 基础题

Q-10. The following probability matrix contains the joint probabilities for random variables $\mathbf{X = \{2, 7, or\ 12\}}$ and $\mathbf{Y = \{1, 3, or\ 5\}}$:

		Y		
		1	3	5
X	2	5%	15%	5%
	7	10%	30%	10%
	12	5%	15%	5%

We are informed that (X) and (Y) are independent. What is the expected value of the product of X and Y, $\mathbf{E(XY)}$?

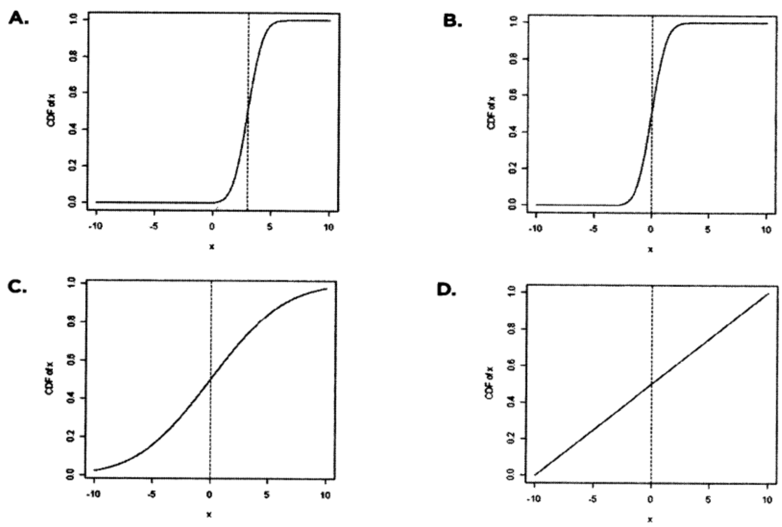
A. 15.0

B. 21.0

C. 30.5

D. 35.0

Q-11. The following graphs show the cumulative distribution function (CDF) of four different random variables. The dotted vertical line indicates the mean of the distribution. Assuming each random variable can only be values between -10 and 10, which distribution has the highest variance?



Q-12. Roy Thomson, a global investment risk manager of FBN Bank, is assessing Markets A and B using a two-factor model:

$$\mathbf{R_i = \alpha_i + \beta_{i,1}F_1 + \beta_{i,2}F_2 + \varepsilon_i}$$

where $\mathbf{R_i}$ is the return for asset i; $\boldsymbol{\beta}$ is the factor sensitivity; And $\mathbf{F}$ is the factor. The random error $\boldsymbol{\varepsilon_i}$, has a mean of zero and is uncorrelated with the factors and with the random error of the other asset returns. In order to determine the covariance between Markets A and B, Thomson developed the following factor covariance matrix for global assets:

Factor Covariance Matrix for Global Assets

	Global Equity Factor	Global Bond Factor
Global Equity Factor	0.3424	0.0122
Global Bond Factor	0.0122	0.0079

Suppose the factor sensitivities to the global equity factor are 0.70 for market A and 0.85 for Market B, and the factor sensitivities to the global bond factors are 0.30 for market A and 0.55 for Market B. The covariance between Market A and Market B is closest to:

A. 0.213

B. 0.461

C. 0.205

D. 0.453

Q-13. Let X and Y be two random variables representing the annual returns of two different portfolios. If E[X] = 3, E[Y] = 4 and E[XY] = 11, then what is Cov[X, Y]?

A. -1

B. 0

C. 11

D. 12

Use the following data to answer Questions 9 and 10.

Probability Matrix

Returns	R_B = 50%	R_B = 20%	R_B = -30%
R_A = -10%	40%	0%	0%
R_A = 10%	0%	30%	0%
R_A = 30%	0%	0%	30%

Q-14. Given the probability matrix above, the standard deviation of Stock B is closest to?

A. 0.11

B. 0.22

C. 0.33

D. 0.15

Q-15. Given the probability matrix above, the covariance between Stock A and B is closest to?Let X and Y be two random variables representing the annual returns of two different portfolios.

A. -0.160

B. -0.055

C. 0.004

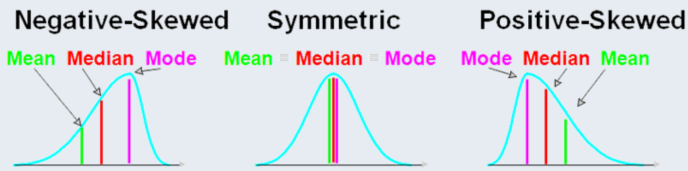
D. 0.020

2.3. Skewness & Kurtosis

2.3.1. 重要知识点

2.3.1.1. Skewness:

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

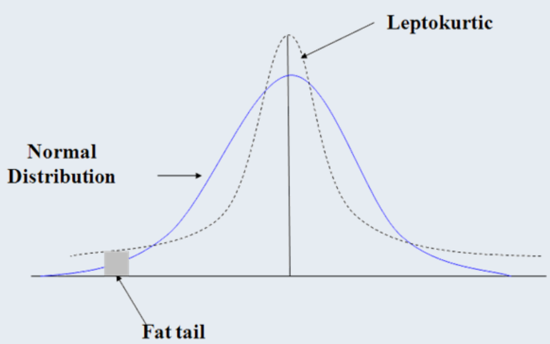


2.3.1.2. Kurtosis:

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

	leptokurtic	Mesokurtic (normal distribution)	platykurtic
Sample kurtosis	>3	=3	<3
Excess kurtosis	>0	=0	<0

Excess Kurtosis = Kurtosis – 3



2.3.2. 基础题

Q-16. An analyst gathered the following information about the return distributions for two portfolios during the same time period:

Portfolio	Skewness	Kurtosis
A	-1.6	1.9
B	0.8	3.2

The analyst states that the distribution for Portfolio A is more peaked than a normal distribution and that the distribution for Portfolio B has a long tail on the left side of the distribution. Which of the following is correct?

A. The analyst’s assessment is correct.

B. The analyst’s assessment is correct for Portfolio A and incorrect for portfolio B.

C. The analyst’s assessment is incorrect for Portfolio A but is correct for portfolio B.

D. The analyst is incorrect in his assessment for both portfolios.

Q-17. Which of the following exhibit positively skewed distributions?

I. Normal Distribution

II. Lognormal Distribution

III. The Returns of Being Short a Put Option

IV. The Returns of Being Long a Call Option

A. II only

B. III only

C. II and IV only

D. I, III, and IV only

Q-18. Which type of distribution produces the lowest probability for a variable to exceed a specified extreme value X which is greater than the mean assuming the distributions all have the same mean and variance?

A. A leptokurtic distribution with a kurtosis of 4

B. A leptokurtic distribution with a kurtosis of 8

C. A normal distribution

D. A platykurtic distribution

Q-19. 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 returns over time is:

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

Q-20. In looking at the frequency distribution of weekly crude oil price changes between 1984 and 2008, an analyst notices that the frequency distribution has a surprisingly large number of observations for extremely large positive price changes and a smaller number, but still a surprising one of observations for extremely large negative price changes. The analyst provides you with the following statistical measures. Which measures would help you identify these characteristics of the frequency distribution?

I. Serial correlation of weekly price changes

II. Variance of weekly price changes

III. Skewness of weekly price changes

IV. Kurtosis of weekly price changes

A. I, II, III and IV

B. II only

C. III and IV only

D. I, III and IV only

2.4. Distribution

2.4.1. 重要知识点

2.4.1.1. Discrete Distribution;

2.4.1.1.1. Bernoulli Distribution

$$p(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

2.4.1.1.2. Binomial distribution

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

	Expectation	Variance
Bernoulli random variable (Y)	p	p(1-p)
Binomial random variable (X)	np	np(1-p)

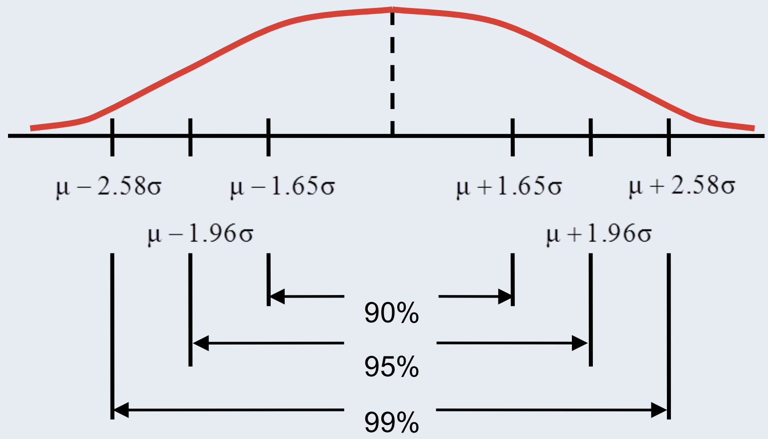
2.4.1.1.3. Poisson Distribution

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

2.4.1.2. Continuous Distribution.

2.4.1.2.1. Normal Distribution

$$\bar{X} \sim N(\mu, \sigma^2) \rightarrow Z = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} \sim N(0,1)$$



2.4.1.2.2. Lognormal Distribution

$$\ln X \sim N(\mu, \sigma^2)$$

- If lnX is normal, then X is lognormal; if a variable is lognormal, its natural log is normal.

2.4.1.2.3. *t*- Distribution

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

Fatter tail than normal distribution.

2.4.1.2.4. Chi-Square Distribution

$$(n-1)\left(\frac{S^2}{\sigma^2}\right) \sim \chi^2_{(n-1)}$$

2.4.1.2.5. F-Distribution

$$F = \frac{S_1^2}{S_2^2} \sim F_{(n_1-1),\ (n_2-1)},\ (S_1 > S_2)$$

2.4.2. 基础题

Q-21. A portfolio manager holds five bonds in a portfolio and each bond has a 1-year default probability of 17%. The event of default for each of the bonds is independent.What is the probability of exactly two bonds defaulting over the next year?

A. 1.9%

B. 5.7%

C. 16.5%

D. 32.5%

Q-22. Consider a binomial distribution with a probability of each success, p = 0.050, and that total number of trials, n = 30 trials. What is the inverse cumulative distribution function associated with a probability of 25.0%?

A. Zero successes

B. One successes

C. Two successes

D. Three successes

Q-23. A portfolio manager holds five bonds in a portfolio and each bond has a 1-year default probability of 17%. The event of default for each of the bonds is independent. What is the mean and standard deviation of the number of bonds defaulting over the next year?

A. Mean = 0.15, standard deviation = 0.71

B. Mean = 0.85, standard deviation = 0.84

C. Mean = 0.85, standard deviation = 0.71

D. Mean = 0.15, standard deviation = 0.84

Q-24. A fixed income portfolio manager currently holds a portfolio of bonds of various companies. Assuming all these bonds have the same annualized probability of default and that the defaults are independent, the number of defaults in this portfolio over the next year follows which type of distribution?

A. Bernoulli

B. Normal

C. Binomial

D. Exponential

Q-25. A multiple choice exam has ten questions, with five choices per question. If you need at least three correct answers to pass the exam, what is the probability that you will pass simply by guessing?

A. 0.8%

B. 20.1%

C. 67.8%

D. 32.2%

Q-26. 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%

Q-27. A certain low-severity administrative (operational) process tends to produce an average of eight errors per week (where each week is five workdays). If this loss frequency process can be characterized by a Poisson distribution, which is NEAREST to the probability that more than one error will be produced tomorrow; i.e., $\mathbf{P}(\mathbf{K}>\mathbf{1}|\boldsymbol{\lambda}=\mathbf{8}/\mathbf{5})$?

A. 20.19%

B. 32.30%

C. 47.51%

D. 66.49%

Q-28. Consider the following five random variables:

- A standard normal random variable; no parameters needed.
- A student's t distribution with 10 degrees of freedom; df = 10.
- A Bernoulli variable that characterizes the probability of default (PD), where PD = 4%; p= 0.040
- A Poisson distribution that characterizes the frequency of operational losses during the day, where lambda = 5.0
- A binomial variable that characterizes the number of defaults in a basket credit default swap (CDS) of 50 bonds, each with PD = 2%; n = 50, p = 2%

Which of the above has, respectively, the lowest value and highest value as its variance among the set?

A. Standard normal (lowest) and Bernoulli (highest)

B. Binomial (lowest) and Student's t (highest)

C. Bernoulli (lowest) and Poisson (highest)

D. Poisson (lowest) and Binomial (highest)

Q-29. Which of the following statements are TRUE?

I. The sum of two independent random normal variables is also an independent random normal variable.

II. The product of two random normal variables is also a random normal variable.

III. The sum of two random lognormal variables is also a random lognormal variable.

IV. The product of two random lognormal variables is also a random lognormal variable.

A. I and II only

B. II and III only

C. III and IV only

D. I and IV only

Q-30. Suppose that a quiz consists of 20 true-false questions. A student has not studied for the exam and just randomly guesses the answers. How would you find the probability that the student will get 8 or fewer answers correct?

A. Find the probability that X = 8 in a binomial distribution with n = 20 and p = 0.5.

B. Find the area between 0 and 8 in a uniform distribution that goes from 0 to 20.

C. Find the probability that X = 8 for a normal distribution with mean of 10 and standard deviation of 5.

D. Find the cumulative probability for 8 in a binomial distribution with n = 20 and p = 0.5.

Q-31. A portfolio manager holds three bonds in one of his portfolios and each has a 1-year default probability of 15%. The event of default for each of the bonds is independent. What is the mean and variance of the number of bonds defaulting over the next year?

A. Mean = 0.15, variance = 0.32

B. Mean = 0.45, variance = 0.38

C. Mean = 0.45, variance = 0.32

D. Mean = 0.15, variance = 0.38

Q-32. Assume that a random variable follows a normal distribution with a mean of 80 and a standard deviation of 24. What percentage of this distribution is between 32 and 116?

A. 4.56%

B. 8.96%

C. 13.36%

D. 91.04%

Q-33. The recent performance of Prudent Fund, with USD 50 million in assets, has been weak and the institutional sales group is recommending that it be merged with Aggressive Fund, a USD 200 million fund. The returns on Prudent Fund are normally distributed with a mean of 3% and a standard deviation of 7% and the returns on Aggressive Fund are normally distributed with a mean of 7% and a standard deviation of 15%. Senior management has asked you to estimate the likelihood that returns on the combined portfolio will exceed 26%. Assuming the returns on the two funds are independent, your estimate for the probability that the returns on the combined fund will exceed 26% is closest to:

A. 1.0%

B. 2.5%

C. 5.0%

D. 10.0%

Q-34. Each of the following is true about the chi-square and F distributions EXCEPT:

A. The chi-square distribution is used to test a hypothesis about a sample variance; i.e., given an observed sample variance, is the true population variance different than a specified value?

B. As degrees of freedom increase, the chi-square approaches a lognormal distribution and the F distribution approaches a gamma distribution

C. The F distribution is used to test the joint hypothesis that the partial slope coefficients in a multiple regression are significant; i.e., is the overall multiple regression significant?

D. Given a computed F ratio, where F ratio = (ESS/df)/(SSR/df), and sample size (n), we can compute the coefficient of determination (R^2) in a multiple regression with (k) independent variables (regressors)

2.5. The Central Limit Theorem

2.5.1. 基础题

Q-35. 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

Q-36. A risk manager is calculating the VaR of a fund with a data set of 25 weekly returns. The mean weekly return is 7% and the standard deviation of the return series is 15%. Assuming that weekly returns are independent and identically distributed, what is the standard deviation of the mean weekly return?

A. 0.4%

B. 0.7%

C. 3.0%

D. 10.0%

2.5.2. 重要知识点

2.5.2.1. Nominal, ordinal, interval, ratio scales

- If $X_1, X_2 \cdots X_n$ represent n independent identically distributed random variables with mean μ and a finite variance σ^2, regardless of the distribution of these n variables, as $n \to \infty$, the distribution of the sample mean $\overline{X} = \sum X_i / n$ is close to the normal distribution with mean μ and variance σ^2/n .

2.6. Confidence Intervals

2.6.1. 重要知识点

2.6.1.1. In general, we have:

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$ **when the variance is known**

$$\bar{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}$$ **when the variance is unknown**

2.6.2. 基础题

Q-37. For a sample of the past 30 monthly stock returns for McCreary, Inc., the mean return is 4% and the sample standard deviation is 20%. Since the population variance is unknown, the standard error of the sample is estimated to be:

$$S_X = \frac{20\%}{\sqrt{30}} = 3.65\%$$

The related t-table values are ($t_{i,j}$ denotes the $(100\text{-}j)^{th}$ percentile of t-distribution value with i degrees of freedom):

$t_{29,2.5\%}$	2.045
$t_{29,5.0\%}$	1.699
$t_{30,2.5\%}$	2.042
$T_{30,5.0\%}$	1.697

What is the 95% confidence interval for the mean monthly return?

A. [-3.453%, 11.453%]

B. [-2.201%, 10.201%]

C. [-2.194%, 10.194%]

D. [-3.464%, 11.464%]

Q-38. Using the prior 12 monthly returns, an analyst estimates the mean monthly return of stock XYZ to be -0.75% with a standard error of 2.70%.

ONE-TAILED T-DISTRIBUTION TABLE

Degrees of Freedom	α		
	0.10	0.05	0.025
8	1.397	1.860	2.306
9	1.383	1.833	2.262
10	1.372	1.812	2.228
11	1.363	1.796	2.201
12	1.356	1.782	2.179

Using the t-table above, the 95% confidence interval for the mean return is between:

A. -6.69% and 5.19%

B. -6.63% and 5.15%

C. -5.60% and 4.10%

D. -5.56% and 4.06%

Q-39. A risk manager is examining a Hong Kong trader's profit and loss record for the last week, as shown in the table below:

Trading Day	Profit/Loss (HKD million)
Monday	10
Tuesday	80
Wednesday	90
Thursday	-60
Friday	30

The profits and losses are normally distributed with a mean of 4.5 million HKD and assume that transaction costs can be ignored. Part of the t-table is provided below:

Percentage Point of the t-Distribution

$$\mathbf{P(T > t) = \alpha}$$

	α		
Degrees of Freedom	0.3	0.2	0.15
4	0.569	0.941	1.19
5	0.559	0.92	1.16

According to the information provided above, what is the probability that this trader will record a profit of at least HKD 30 million on the first trading day of next week?

A. About 15%

B. About 20%

C. About 80%

D. About 85%

2.7. Hypothesis Testing

2.7.1. 重要知识点

2.7.1.1. The null hypothesis (H_0) and alternative hypothesis (H_a)

- One-tailed test vs. Two-tailed test
- One-tailed test: H_0: $\mu \geq 0$ H_a: $\mu < 0$ H_0: $\mu \leq 0$ H_a: $\mu > 0$
- Two-tailed test: H_0: $\mu = 0$ H_a: $\mu \neq 0$

2.7.1.2. Critical Value

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

2.7.1.3. Summary of hypothesis

Test type	Assumptions	H_0	Test-statistic	distribution
Mean hypothesis testing	Normally distributed population, known population variance	$\mu = 0$	$z = \frac{\bar{x} - u_0}{\sigma/\sqrt{n}}$	$N(0,1)$
	Normally distributed population, unknown population variance	$\mu = 0$	$t = \frac{\bar{x} - u_0}{S/\sqrt{n}}$	$t(n-1)$
	Testing the equality of two means	$\mu_1 = \mu_2$	$T = \frac{\hat{\mu}_Z}{\sqrt{\hat{\sigma}_Z^2/n}} = \frac{\hat{\mu}_X - \hat{\mu}_Y}{\sqrt{\frac{\hat{\sigma}_X^2 + \hat{\sigma}_Y^2 - 2\hat{\sigma}_{XY}}{n}}}$	$t(n-1)$
Variance hypothesis testing	Normally distributed population	$\sigma^2 = \sigma_0^2$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\chi^2(n-1)$
	Two independent normally distibuted populations	$\sigma_1^2 = \sigma_2^2$	$F = S_1^2/S_2^2$	$F(n_1-1, n_2-1)$

2.7.2. 基础题

Q-40. Hedge Fund has been in existence for two years. Its average monthly return has been 6% with a standard deviation of 5%. Hedge Fund has a stated objective of controlling volatility as measured by the standard deviation of monthly returns. You are asked to test the null hypothesis that the volatility of Hedge Fund's monthly returns is equal to 4% versus the alternative hypothesis that the volatility is not equal to 4%. Assuming that all monthly returns are independently and identically normally distributed, and using the tables below, what is the correct test to be used and what is the correct conclusion at the 5% level of significance?

t Table: Inverse of the one-tailed probability of the Student's t-distribution

Df	One-tailed Probability = 5.0%	One-tailed Probability 2.5%
22	1.717	2.074
23	1.714	2.069
24	1.711	2.064

Chi-Square Table: Inverse of the one-tailed probability of the Chi-Squared distribution

Df	One-tailed Probability = 5.0%	One-tailed Probability = 2.5%

22	33.9244	36.7807
23	35.1725	38.0757
24	36.4151	39.3641

A. t-test; reject the null hypothesis

B. Chi-square test; reject the null hypothesis

C. t-test; do not reject the null hypothesis

D. Chi-square test; do not reject the null hypothesis

Q-41. Based on 21 daily returns of an asset, a risk manager estimates the standard deviation of the asset's daily returns to be 2%. Assuming that returns are normally distributed and that there are 260 trading days in a year, what is the appropriate Chi-square test statistic if the risk manager wants to test the null hypothesis that the true annual volatility is 25% at a 5% significance level?

A. 25.80

B. 33.28

C. 34.94

D. 54.74

Q-42. Using a sample size of 61 observations, an analyst determines that the standard deviation of the returns from a stock is 21%. Using a 0.05 significance level, the analyst:
(If df = 60 and p = 0.05, the critical value is 79.08)

A. Can conclude that the standard deviation of returns is higher than 14%.

B. Cannot conclude that the standard deviation of returns is higher than 14%.

C. Can conclude that the standard deviation of returns is not higher than 14%.

D. None of the above.

Q-43. Bob tests the null hypothesis that the population mean is less than or equal to 45. From a population size of 3,000,000 people, 81 observations are randomly sampled. The corresponding sample mean is 46.3 and sample standard deviation is 4.5. What is the value of the most appropriate test statistic for the test of the population mean, and what is the correct decision at the 1 percent significance level?

A. z = 0.29, and fail to reject the null hypothesis.

B. z = 2.60, and reject the null hypothesis.

C. t = 0.29, and accept the null hypothesis.

D. t = 2.60, and neither reject nor fail to reject the null hypothesis.

Q-44. An analyst wants to test whether the standard deviation of return from pharmaceutical stocks is lower than 0.2. For this purpose, he obtains the following data from a sample of 30 pharmaceutical stocks. Mean return from pharmaceutical stocks = 8%. Standard deviation of return from pharmaceutical stocks = 12%. Mean return from the market = 12%. Standard deviation of return from the market = 16%. What is the appropriate test statistic for this test?

A. t-statistic
B. z-statistic
C. F-statistic
D. χ^2 statistic

2.8. Best Linear Unbiased Estimator

2.8.1. 重要知识点

2.8.1.1. Unbiasedness、Efficiency、Consistency、Linearity

2.8.1.2. The OLS estimator is BLUE.

2.8.2. 基础题

Q-45. If the variance of the sampling distribution of an estimator is smaller than all other unbiased estimators of the parameter of interest, the estimator is:

A. Reliable
B. Efficient
C. Unbiased
D. Consistent

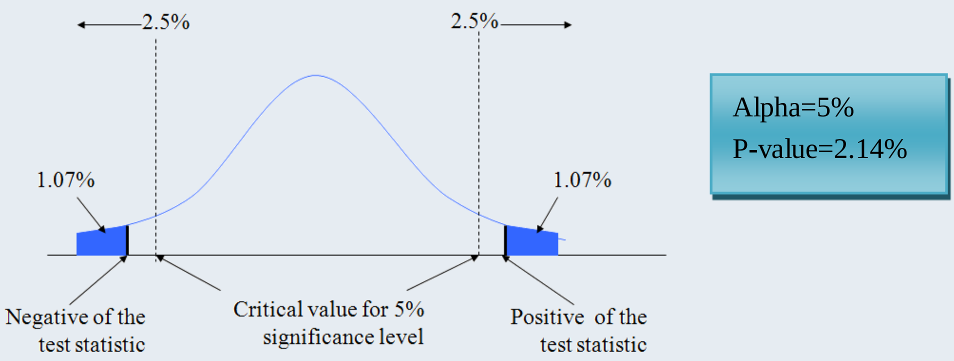
Q-46. Analyst Rob has identified an estimator, denoted T(.), which qualifies as the best linear unbiased estimator (BLUE). If T(.) is BLUE, which of the following must also necessarily be TRUE?

A. T(.) must have the minimum variance among all possible estimators.
B. T(.) must be the most efficient (the "best") among all possible estimators.
C. It is possible that T(.) is the maximum likelihood (MLE) estimator of variance; i.e., SUM([X - average (X)]^2)/(n-1).
D. Among the class of unbiased estimators that are linear, T(.) has the smallest variance.

2.9. P-value Testing

2.9.1. 重要知识点

2.9.1.1. P-value Testing:



- **Decision rule:**
  - If P-value<alpha, we will reject the null hypothesis
  - If P-value>=alpha, we will not reject the null hypothesis

2.9.2. 基础题

Q-47. Which of the following statements regarding hypothesis testing is correct?

A. Type II error refers to the failure to reject the H_1 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.

2.10. Type I & Type II errors

2.10.1. 重要知识点

2.10.1.1. The P (Type I error) equals to the significance level α.

	True Condition	
Decision	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Incorrect decision **Type II error**
Reject H_0	Incorrect decision **Type I error** Significance level, α, =P(Type I error)	Correct decision Power of the test =1- P(Type II error)

- Given the sample size, Type I and II errors cannot be reduced simultaneously.

2.10.2. 基础题

Q-48. When testing a hypothesis, which of the following statements is correct when the level of significance of the test is decreased?

A. The likelihood of rejecting the null hypothesis when it is true decreases.

B. The likelihood of making a Type I error increases.

C. The null hypothesis is rejected more frequently, even when it is actually false.

D. The likelihood of making a Type II error decreases.

Q-49. 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

Q-50. According to the Basel back-testing framework guidelines, penalties start to apply if there are five or more exceptions during the previous year. The Type I error rate of this test is 11 percent. If the true coverage is 97 percent of exceptions instead of the required 99 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.

2.11. Regression & Variance Analysis

2.11.1. 重要知识点

2.11.1.1. Linear Regression:

$$\mathbf{Y_i = b_0} + \boldsymbol{b}_\mathbf{i} \times \mathbf{X_i} + \boldsymbol{\varepsilon}_\mathbf{i}$$

2.11.1.2. Ordinary least squares (OLS):

$$\textbf{minimize} \sum \mathbf{e_i^2} = \sum [\mathbf{Y_0} - (\mathbf{b_0} + \mathbf{b_1} \times \mathbf{X_i})]^2$$

$$\textbf{minimize} \sum \mathbf{e_i^2} = \sum \left[\mathbf{Y_i} - \left(\mathbf{b_0} + \sum_{i=1}^{k} \mathbf{b_i} \times \mathbf{X_i}\right)\right]^2$$

2.11.1.3. The Assumptions of Classical Linear Regression Model:

- A linear relationship exists between X and Y;
- X is uncorrelated with the error term;
- The expected value of the error term is zero;
- The variance of the error term is constant (i.e., the error terms are homoskedastic);
- The error term is uncorrelated across observations;
- The error term is normally distributed.

2.11.1.4. There are five key assumptions for the linear regression model:

- The regression errors, εi , have a mean of zero conditional on the regressors Xi
- Constant Variance of Shocks
- Variance of independent variable is strictly greater than 0 ($\sigma_X^2 > 0$).
- The sample observations are i.i.d. random draws from the population
- Large outliers are unlikely
- The explanatory variables are not perfectly linearly dependent.

2.11.1.5. Regression Assumption Violations:

- Heteroskedasticity occurs when the variance of the residuals is not rho same across all observations in the sample.
- Multicollinearity refers to the condition when two or more of the independent variables, or linear combinations of the independent variables, in a multiple regression are highly correlated with each other.
- Omitted variable bias occurs when the omitted variable is correlated with the included regressor and is a determinant of the dependent variable.
- Serial correlation refers to the situation in which the residual terms are correlated with one another.

2.11.1.6. Analysis of Variance (ANOVA) Table:

	df	SS	MSS
Regression	k	ESS	ESS/k
Residual	n-k-1	RSS	RSS/(n-k-1)
Total	n-1	TSS	-

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

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

$$\text{TSS} = \text{ESS} + \text{SSR}$$

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{SSR}}{\text{TSS}} \; ; \; r = \pm\sqrt{R^2}$$

2.11.1.7. Adjusted R-Squared:

$$\mathbf{Adjusted\ R^2 = 1 - (1 - R^2) \times \frac{n-1}{n-k-1}}$$

2.11.1.8. Multiple Linear Regressions:

$$\mathbf{Y_i = b_0 + \boldsymbol{b_1} \times X_{1,i} + b_2 \times X_{2,i} + \boldsymbol{\varepsilon}_i}$$

2.11.2. 基础题

Q-51. Samantha Xiao is trying to get some insight into the relationship between the return on stock LMD ($\mathbf{R_{LMD,t}}$) and the return on the S&P 500 index ($\mathbf{R_{S\&P,t}}$). Using historical data she estimates the following:

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

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

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

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

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

Q-52. For a sample of 400 firms, the relationship between corporate revenue (Y_i) and the average years of experience per employee (Xi) is modeled as follows:

$$\mathbf{Y_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 X_i + \boldsymbol{\beta}_2 X_2 + \boldsymbol{\varepsilon}}$$

You wish to test the joint null hypothesis that $\beta_1 = 0$ and $\beta_2 = 0$ at the 95% confidence level. The p-value for the t-statistic for β_1 is 0.07, and the p-value for the t-statistic for β_2 is 0 06. The p-value for the F-statistic for the regression is 0.045. Which of the following statements is correct?

A. You can reject the null hypothesis because each β is different from 0 at the 95% confidence level.

B. You cannot reject the null hypothesis because neither β is different from 0 at the 95% confidence level.

C. You can reject the null hypothesis because the F-statistic is significant at the 95% confidence level.

D. You cannot reject the null hypothesis because the F-statistic is not significant at the 95% confidence level.

Q-53. An analyst is testing a hypothesis that the beta, β, of stock CDM is 1. The analyst runs an ordinary least squares regression of the monthly returns of CDM, R_{CDM}, on the monthly returns of the S&P 500 index, Rm, and obtains the following relation:

$$R_{CDM} = 0.86\,R_m - 0.32$$

The analyst also observes that the standard error of the coefficient of R_m is 0.80. In order to test the hypothesis $H_0: \beta = 1$ against $H_1: \beta \neq 1$, what is the correct statistic to calculate?

A. t-statistic

B. Chi-square test statistic

C. F test statistic

D. Sum of squared residuals

Q-54. The proper selection of factors to include in an ordinary least squares estimation is critical to the accuracy of the result. When does omitted variable bias occur?

A. Omitted variable bias occurs when the omitted variable is correlated with the included regressor and is a determinant of the dependent variable.

B. Omitted variable bias occurs when the omitted variable is correlated with the included regressor but is not a determinant of the dependent variable.

C. Omitted variable bias occurs when the omitted variable is independent of the included regressor and is a determinant of the dependent variable.

D. Omitted variable bias occurs when the omitted variable is independent of the included regressor but is not a determinant of the dependent variable.

Q-55. Each of the following is true about the adjusted $\mathbf{R^2}$ EXCEPT which is false?

A. Adjusted R^2 = 1 - (SSR/TSS)*[(n-1)/(n-k-1)].

B. Adding a regressor (independent variable) always causes the adjusted R^2 to decrease.

C. Adjusted R^2 is always less than R^2.

D. The adjusted R^2 can be negative.

Q-56. Which of the following is assumed in the multiple least squares regression model?

A. The dependent variable is stationary.

B. The independent variables are not perfectly multicollinear.

C. The error terms are heteroskedastic.

D. The independent variables are homoskedastic.

Q-57. Which of the following statements about the ordinary least squares regression model (or simple regression model) with one independent variable are correct?

I. In the ordinary least squares (OLS) model, the random error term is assumed to have zero mean and constant variance.

II. In the OLS model, the variance of the independent variable is assumed to be positively correlated with the variance of the error term.

III. In the OLS model, it is assumed that the correlation between the dependent variable and the random error term is zero.

IV. In the OLS model, the variance of the residuals is assumed to be constant.

A. I, II, III and IV

B. II and IV only

C. I and IV only

D. I, II, and III only

Use the following information to answer the following questions.

Regression Statistics

R squared	0.8537
R sq. adj.	0.8120
Std. error	10.3892
Num obs.	10

ANOVA

	df	SS	MS	F	P-value
Explained	2	4410.4500	2205.2250	20.4309	0.0012
Residual	7	755.5500	107.9357		
Total	9	5166.0000			

	Coefficients	Std. Error	t-Stat	P-value
Intercept	35.5875	6.1737	5.7644	0.0007
X_1	1.8563	1.6681	1.1128	0.3026
X_2	7.4250	1.1615	6.3923	0.0004

Q-58. Based on the results and a 5% level of significance, which of the following hypotheses can be rejected?

I. H_0: $B_0 = 0$

II. H_0: $B_1 = 0$

III. H_0: $B_2 = 0$

IV. H_0: $B_1 = B_2 = 0$

A. I, II, and III

B. I and IV

C. III and IV

D. I, III, and IV

Q-59. Paul Graham, FRM® is analyzing the sales growth of a baby product launched three years ago by a regional company. He assesses that three factors contribute heavily towards the growth and comes up with the following results:

$Y = b + 1.5X_1 + 1.2X_2 + 3X_3$

Sum of Squared Regression [SSR] = 869.76

Sum of Squared Errors [SSE] = 22.12

Determine what proportion of sales growth is explained by the regression results.

A. 0.36

B. 0.98

C. 0.64

D. 0.55

Q-60. Many statistical problems arise when estimating relationships using regression analysis. Some of these problems are due to the assumptions behind the regression model. Which one of the following is NOT one of these problems?

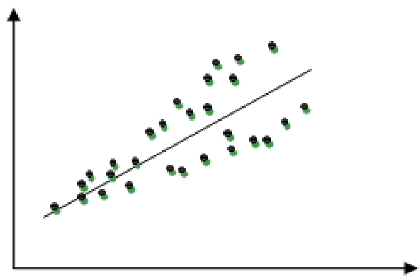
A. Stratification

B. Multicollinearity

C. Heteroscedasticity

D. Autocorrelation

Q-61. An analyst is performing a regression. The dependent variable is portfolio return while the independent variable is the years of experience of the portfolio manager. In his analysis, the resulting scatter plot is as follow:



The analyst can conclude that the portfolio returns exhibit:

A. Heteroskedasticity

B. Homoskedasticity

C. Perfect multicollinearity

D. Non-perfect multicollinearity

Q-62. 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 correct?

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

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

III. If the return on the industry index is 4%, the stock's expected return is 10.3%.

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

A. III only

B. I and II only

C. II and IV only

D. I, II, and IV

Q-63. An analyst is given the data in the following table for a regression of the annual sales for Company XYZ, a maker of paper products, on paper product industry sales.

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 the 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.

Q-64. A risk manager performs an ordinary least squares (OLS) regression to estimate the sensitivity of a stock's return to the return on the S&P 500. This OLS procedure is designed to:

A. Minimize the square of the sum of differences between the actual and estimated S&P 500 returns.

B. Minimize the square of the sum of differences between the actual and estimated stock returns.

C. Minimize the sum of differences between the actual and estimated squared S&P 500 returns.

D. Minimize the sum of squared differences between the actual and estimated stock returns.

Q-65. Using data from a pool of mortgage borrowers, a credit risk analyst performed an ordinary least squares regression of annual savings (in GBP) against annual household income (in GBP) and obtained the following relationship:

$\text{Annual Savings} = 0.24 \times \text{Household Income} - 25.66,\ R^2 = 0.50$

Assuming that all coefficients are statistically significant, which interpretation of this result is correct?

A. For this sample data, the average error term is GBP -25.66.

B. For a household with no income, annual savings is GBP 0.

C. For an increase of GBP 1,000 in income, expected annual savings will increase by GBP 240.

D. For a decrease of GBP 2,000 in income, expected annual savings will increase by GBP 480.

Q-66. A risk manager has estimated a regression of a firm's monthly portfolio returns against the returns of three U.S. domestic equity indexes: the Russell 1000 index, the Russell 2000 index, and the Russell 3000 index. The results are shown below.

Regression Statistics	
Multiple R	0.9
R Square	0.9
Adjusted R Square	0.9
Standard Error	0.0

Observations	192

Regression Output	Coefficients	Standard Error	t-Stat	P-value
Intercept	0.0023	0.0006	3.530	0.0005
Russell 1000	0.1093	1.5895	0.068	0.9452
Russell 2000	0.1055	0.1384	0.762	0.4470
Russell 3000	0.3533	1.7274	0.204	0.8382

Correlation Matrix	Portfolio Returns	Russell 1000	Russell 2000	Russell 3000
Portfolio	1.000			
Russell 1000	0.937	1.000		
Russell 2000	0.856	0.813	1.000	
Russell 3000	0.945	0.998	0.845	1.000

Based on the regression results, which statement is correct?

A. The estimated coefficient of 0.3533 indicates that the returns of the Russell 3000 index are more statistically significant in determining the portfolio returns than the other two indexes.

B. The high adjusted R^2 indicates that the estimated coefficients on the Russell 1000, Russell 2000, and Russell 3000 indexes are statistically significant.

C. The high p-value of 0.9452 indicates that the regression coefficient of the returns of Russell 1000 is more statistically significant than the other two indexes.

D. The high correlations between each pair of index returns indicate that multicollinearity exists between the variables in this regression.

Q-67. Our linear regression produces a high coefficient of determination ($\mathbf{R^2}$) but few significant t ratios. Which assumption is most likely violated?

A. Homoscedasticity

B. Multicollinearity

C. Error term is normal with mean = 0 and constant variance = $sigma^2$

D. No autocorrelation between error terms

Q-68. We observe the variance of the error term is an increasing function of the explanatory variable. Which assumption is violated?

A. Homoscedasticity

B. Multicollinearity

C. Model is linear

D. No autocorrelation between error terms

Q-69. We reject the null hypothesis in a Durbin-Watson D test. Which assumption is violated?

A. Homoscedasticity

B. Multicollinearity

C. Model is linear

D. No autocorrelation between error terms

2.12. Non-stationary series

2.12.1. 重要知识点

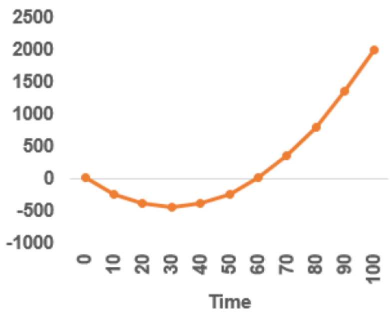
2.12.1.1. Non-stationary series

- **Sources of non-Stationarity**
  - Time trends
    - Linear time trend: $Y_t = \delta_0 + \delta_1 t + \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2)$
    - Nonlinear Trends: $Y_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \cdots + \delta_m t^m + \varepsilon_t$
    - Log-linear trend: $lnY_t = \delta_0 + \delta_1 t + \varepsilon_t$
  - Seasonalities
    - Seasonal Dummy Variables:

      $Y_t = \delta + \gamma_1 l_{1t} + \gamma_2 l_{2t} + \cdots + \gamma_{s-1} l_{s-1t} + \varepsilon_t \quad (l_{it} = 1 \text{ or } 0)$
  - Unit roots (random walks)
    - Testing for Unit Roots: Augmented Dickey-Fuller (ADF) Test

2.12.2. 基础题

Q-70. Consider the following quadratic trend model:



Which of the following functions correctly characterizes this trend?

A. $\text{Tr} = 10 + 0.3 \times \text{TIME} + 0.3 \times \text{TIME}^2$

B. $\text{Tr} = 10 + 30 \times \text{TIME} - 0.3 \times \text{TIME}^2$

C. $\text{Tr} = 10 - 0.4 \times \text{TIME} - 0.4 \times \text{TIME}^2$

D. $\text{Tr} = 10 - 30 \times \text{TIME} + 0.5 \times \text{TIME}^2$

Q-71. Winnie is an analyst in the retail industry. She is modeling a company's sales over time and has noticed a quarterly seasonal pattern. If Winnie includes an intercept term in her model, how many dummy variables should she use to model the seasonality component?

A. 2.

B. 3.

C. 4.

D. 5.

Q-72. Consider the following regression equation utilizing dummy variables for explaining quarterly SALES in terms of the quarter of their occurrence:

$SALES_t = \beta_0+\beta_2 D_{2,t} +\beta_3 D_{3,t}+ \beta_4 D_{4,t} + e_t$

where:

SALES=a quarterly observation of EPS

$D_{2,t}$= 1if period t is the second quarter, $D_{2,t}$= 0 otherwise

$D_{3,t}$=1 if period t is the third quarter, $D_{3,t}$= 0 otherwise

$D_{4,t}$ = 1 if period t is the fourth quarter, $D_{4,t}$ = 0 otherwise

The intercept term β_0 represents the average value of sales for the:

A. First quarter

B. Second quarter

C. Third quarter.

D. Fourth quarter.

Q-73. Which of the following scenarios would produce a forecasting model that exhibits perfect multicollinearity? A model that includes:

I Only one seasonal dummy that is equal to 1.

II A holiday variation variable that accounts for an "Easter dummy variable."

III A trading-day variation variable for modeling trading volume throughout the year.

IV A dummy variable for each season, plus an intercept.

A. II only.

B. I and III.

C. IV only.

D. All not.

2.13. Modeling Cycles: MA, AR, and ARMA Models

2.13.1. 重要知识点

2.13.1.1. Stationary series

- **Covariance Stationarity**
  - $E[Y_t] = \mu$ for all t

- $Cov[Y_t, Y_{t-h}] = \gamma_h$ for all t
- $V[Y_t] = \gamma_0 < \infty$ for all t

➢ **White noise**

- Mean zero
- Constant and finite variance, σ^2
- No autocorrelation or autocovariance, $\gamma(h) = \rho(h) = 0$ for $h \neq 0$

2.13.1.2. AR(1) model

$$\mathbf{y_t = \varphi y_{t-1} + \varepsilon_t \ ; \ \varepsilon_t \sim WN(0, \sigma^2)}$$

$$\mathbf{(1 - \varphi L) y_t = \varepsilon_t}$$

➢ The AR(1) model is capable of capturing much more persistent dynamics than is the MA(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.

2.13.1.3. AR(p) model

$$\mathbf{y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t \ ; \ \varepsilon_t \sim WN(0, \sigma^2)}$$

$$\mathbf{\Phi(L) y_t = \left(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p\right) y_t = \varepsilon_t}$$

2.13.1.4. MA(1) model

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

➢ The current value of the observed series is expressed as a function of current and lagged unobservable shocks.

➢ MA(1) process with parameter θ = 0.95 varies a bit more than the process with a parameter of θ = 0.4.

2.13.1.5. MA(q) model

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

$$\mathbf{\Theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q}$$

2.13.1.6. ARMA(p,q) model

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

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

2.13.2. 基础题

Q-74. In regard to white noise, each of the following statements is true except?

A. If a process is zero-mean white noise, then is must be Gaussian white noise.

B. If a process is Gaussian (aka, normal) white noise, then it must be (zero-mean) white noise.

C. If a process is Gaussian (aka, normal) white noise, then it must be independent white noise.

D. If a process is stationary, has zero mean, has constant variance and it serially uncorrelated, then the process is white noise.

Q-75. A risk manager at a major global bank is conducting a time series analysis of equity returns. The manager wants to know whether the time series is covariance stationary. Which of the following statements describes one of the requirements for a time series to be covariance stationary?

A. The distribution of a time series should have a kurtosis value near 3.0, ensuring no fat tails will distort stationarity.

B. The distribution of a time series should have a skewness value near 0, so that its mean will fall in the center of the distribution.

C. The autocovariance of a covariance stationary time series depends only on displacement, τ, not on time.

D. When the autocovariance function is asymmetric with respect to displacement, τ, forward looking stationarity can be achieved.

Q-76. Which of the following statements is correct regarding the usefulness of an autoregressive (AR) process and an autoregressive moving average (ARMA) process when modeling seasonal data?

I. They both include lagged terms and, therefore, can better capture a relationship in motion.

II. They both specialize in capturing only the random movements in time series data.

A. I only.

B. II only.

C. Both I and II.

D. Neither I nor II.

2.14. Box-Pierce Q-statistic & Ljung-Box Q-statistic

2.14.1. 重要知识点

2.14.1.1. Box-Pierce Q-statistic & Ljung-Box Q-statistic

- H_0: All the correlation observed in the series are independent of each other and hence Autocorrelation of the series is zero.
- For Box-Pierce Q-statistic, the formula used is:

$$Q_{BP} = T\sum_{\tau=1}^{m}\hat{\rho}^2(\tau)$$

- Whereas in case of Ljung-Box Q-statistic, the test statistic is derived as:

$$Q_{LB} = T(T+2)\sum_{\tau=1}^{m}\hat{\rho}^2(\tau)\left(\frac{1}{T-\tau}\right)$$

- Where:
- T = Sample size
- $\hat{\rho}(\tau) = \text{Sample autocorrelation function for } \tau \text{ lags}$
- m = number of lags under observation

2.14.2. 基础题

Q-77. For a certain time series, you have produced a correlogram with an autocorrelation function that includes twenty four monthly observations; m = degrees of freedom = 24. Your calculated Box-Pierce Q-statistic is 19.50 and your calculated Ljung-Box Q-statistic is 27.90. You want to determine if the series is white noise. Which is your best conclusion (given CHISQ.INV(0.95, 24) = 36.41)?

A. With 95% confidence, you accept the series as white noise (more accurately, you fail to reject the null).

B. With 95% confidence, you accept the series as partial white noise (due to Box-Pierce) but reject the null (due to Ljung-Box).

C. With 95% confidence, you reject both null hypotheses and conclude the series is not white noise.

D. With 95% confidence, you reject both null hypotheses but conclude the series is white noise because the sum of the statistics is greater than the critical value.

2.15. Measuring Returns, Volatility and Correlation

2.15.1. 重要知识点

2.15.1.1. Simple Returns

- $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$
- $1 + R_T = \prod_{t=1}^{T}(1 + R_t)$

2.15.1.2. Continuously Compounded Returns

- $r_t = \ln P_t - \ln P_{t-1}$
- $1 + R_t = e^{r_t}$
- $r_T = \sum_{t=1}^{T} r_t$

2.15.1.3. Correlation

- Pearson’s correlation:

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

- Spearman's correlation:

$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

- Kendal’s τ
  - Concordant pair: sgn $(X_2 - X_1)$ = sgn $(Y_2 - Y_1)$
  - Discordant pair: sgn $(X_2 - X_1)$ = - sgn $(Y_2 - Y_1)$

2.15.2. 基础题

Q-78. A risk manager gathers five years of historical returns to calculate the Kendall τ correlation coefficient for stocks X and Y. The stock returns for X and Y from 2010 to 2014 are as follows:

Year	X	Y
2010	5.0%	-10.0%
2011	50.0%	-5.0%
2012	-10.0%	20.0%
2013	-20.0%	40.0%
2014	30.0%	15.0%

What is the Kendall τ correlation coefficient for the stock returns of X and Y?

A. -0.3

B. -0.2

C. -0.6

D. 0.4

Q-79. A risk manager gathers five years of historical returns to calculate the Spearman rank correlation coefficient for stocks X and Y. The stock returns for X and Y from 2010 to 2014 are as follows:

Year	X	Y
2010	5.0%	-10.0%
2011	50.0%	-5.0%
2012	-10.0%	20.0%
2013	-20.0%	40.0%
2014	30.0%	15.0%

What is the Spearman rank correlation coefficient for the stock returns of X and Y?

A. -0.7

B. -0.3

C. 0.3

D. 0.7

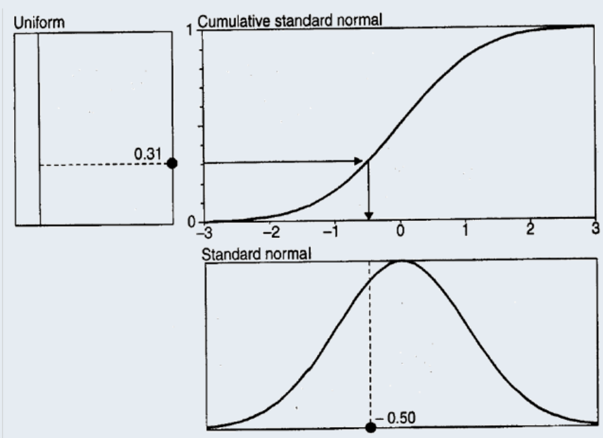
2.16. Monte Carlo Simulation

2.16.1. 重要知识点

Monte Carlo Simulation	
Basic Steps	◻ Specify the data generating process. ◻ Estimate an unknown variable. ◻ Save the estimate from step 2. ◻ Go back to step 1 and repeat this process N times.
Reducing Standard Error $\frac{s}{\sqrt{N}}$	◻ The standard error estimate of a Monte Carlo simulation can be reduced by a factor of 10 by increasing N by a factor of 100. ◻ Variance reduction technique • Antithetic Variates • Control Variates • Random Number Re-Usage across Experiments

Variance Reduction Technique	
Antithetic Variates	◻ Reduces sampling error by rerunning the simulation using a complement set of the original set of random variables.
Control Variates	◻ Replaces a variable x that has unknown properties in a Monte Carlo simulation with a similar variable y that has known properties. The new x* variable estimate will have a smaller sampling error than the original x variable if the control statistic and statistic of interest are highly correlated.

2.16.1.1. Inverse Transform Method:



2.16.2. 基础题

Q-80. A portfolio manager has asked each of four analysts to use Monte Carlo simulation to price a path-dependent derivative contract on a stock. The derivative expires in nine months and the risk-free rate is 4% per year compounded continuously. The analysts generate a total of 20,000 paths using a geometric Brownian motion model, record the payoff for each path, and present the results in the table shown below.

Analyst	Number of Paths	Average Derivative Payoff per Path (USD)
1	2,000	43
2	4,000	44
3	10,000	46
4	4,000	45

What is the estimated price of the derivative?

A. USD 43.33

B. USD 43.77

C. USD 44.21

D. USD 45.10

Q-81. Which of the following statements about Monte Carlo simulation is incorrect?

A. Correlations among variables can be incorporated into a Monte Carlo simulation.

B. Monte Carlo simulations can handle time-varying volatility.

C. Monte Carlo methods can be used to estimate value-at-risk (VaR) but cannot be used to price options.

D. For estimating VaR, Monte Carlo methods generally require more computing power than historical simulations.

Q-82. Consider a stock that pays no dividends, has a vol. of 30% per annum, and provide an expected return of 15% per annum with continuous compounding. The stock price movements follow GBM. Consider a time interval of 1 week and the initial stock price is 100, then the stock price increase has a normal distribution with:

A. Mean = 0.268%, standard deviation = 4.03%

B. Mean = 0.278%, standard deviation = 4.13%

C. Mean = 0.288%, standard deviation = 4.16%

D. Mean = 0.288%, standard deviation = 4.27%

Q-83. A risk manager has been requested to provide some indication of accuracy of a Monte Carlo simulation. Using 1,000 replications of a normally distributed variable S, the relative error in the one-day 99% VaR is 5%. Under these conditions:

A. Using 1,000 replications of a long option position on S should create a larger relative error.

B. Using 10,000 replications should create a larger relative error.

C. Using another set of 1,000 replications will create an exact measure of 5.0% for relative error.

D. Using 1,000 replications of a short option position on S should create a larger relative error.

2.17. The Bootstrap

2.17.1. 重要知识点

2.17.1.1. An alternative to generating random numbers from a hypothetical distribution is to Sample from historical data. $\mathbf{S_{t+1} = S_t\left[1 + R_{m(1)}\right]}$

- **Advantage of bootstrap:**
  - Can include fat tails, jumps, or any departure from the normal distribution.
  - Account for correlations across series because one draw consists of the simultaneous returns for N series, such as stock, bonds, and currency prices.
- **Limitation of bootstrap:**
  - For small sample sizes, it may be a poor approximation of the actual one.
  - Relies heavily on the assumption that returns are independent.

2.17.2. 基础题

Q-84. Which of the following statements about simulation is invalid?

A. The historical simulation approach is a nonparametric method that makes no specific assumption about the distribution of asset returns.

B. When simulating asset returns using Monte Carlo simulation, a sufficient number of trials must be used to ensure simulated returns are risk neutral.

C. Bootstrapping is an effective simulation approach that naturally incorporates correlations between asset returns and non-normality of asset returns, but does not generally capture autocorrelation of asset returns.

D. Monte Carlo simulation can be a valuable method for pricing derivatives and examining asset return scenarios.

SOLUTIONS

2. Quantitative Analysis

1. Solution: C

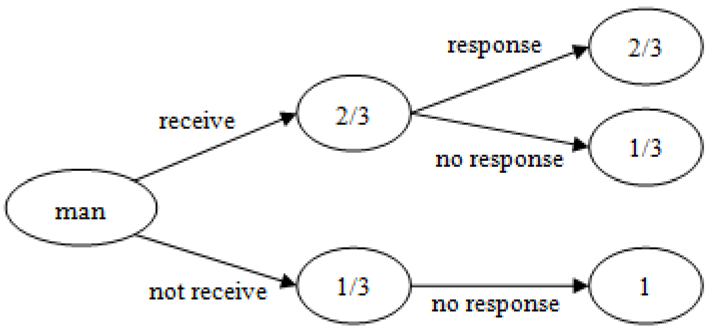
$\mathrm{P(AB) = P(A)P(B|A) = 4\% = P(A) \times 80\% \Rightarrow P(A) = 5\% = P(B)}$

$\mathrm{1 - [P(A) + P(B) - P(AB)] = 94\%}$

2. Solution: C

Given that the economy is good, the probability of a poor economy and a bull market is zero. The other statements are true. The P(normal market) = (0.60 × 0.30) + (0.40 × 0.30) = 0.30. P(good economy and bear market) = 0.60 × 0.20 = 0.12. Given that the economy is poor, the probability of a normal or bull market = 0.30 + 0.20 = 0.50.

3. Solution: D



(Notices: if the wife did not receive the letter, she would not send response to the man.)

Note:

A: the wife received the letter;

$\bar{A}$: the wife did not received the letter;

B: the man received a response letter from his wife;

$\bar{B}$:: the man did not receive a response letter from his wife

The question "If the man does not receive a response letter from his wife, what is the probability that his wife received his letter?" is equal to figure out P(A|$\bar{\mathrm{B}}$).

$$\mathrm{P(A|\bar{B}) = \frac{P(A\bar{B})}{P(\bar{B})} = \frac{P(A) \times P(\bar{B}|A)}{P(A) \times P(\bar{B}|A) + P(\bar{A}) \times P(\bar{B}|\bar{A})}} = \frac{\frac{2}{3} \times \frac{1}{3}}{\frac{2}{3} \times \frac{1}{3} + \frac{1}{3} \times 1} = \frac{2}{5}$$

4. Solution: D

Base on the information given:

P (both auto policy and homeowner policy) = 0.15

P (only home policy) = 0.65 – 0.15 = 0.5

P (only auto policy) = 0.5 – 0.15 = 0.35

Therefore, the proportion of policyholders that will renew at least one policy is shown below:

= 0.4×0.5 + 0.6×0.35 + 0.8×0.15 = 0.53

5. Solution: D

In order to solve this conditional probability question, first calculate the probability that any one mortgage in the portfolio is late.

This is: P(Mortgage is late) = (200+48)/(1000+600) = 15.5%.

Next, use the conditional probability relationship as follows:

P(Subprime mortgage | Mortgage is late) = P(Subprime mortgage and late)/P(Mortgage is late).

Since P(Subprime mortgage and late) = 200/1600 = 12.5%;

P(Mortgage subprime | Mortgage is late) = 12.5% / 15.5% = 0.81 = 81%.

Hence the probability that a random late mortgage selected from this portfolio turns out to be subprime is 81%.

6. Solution: D

Bayes applies to both, although practicing applications are almost always using simple discrete random variables.

7. Solution: C

The sum of the pmf probabilities must be 100.0% such that $30b = 1.0$ or $b = 1/30$.

Therefore the $P[X < 2] = P[X \leq 1] = P[X = 0] + P[X = 1] = 12/30 + 7/30 = 19/30 =$ 63.33%.

8. Solution: C

The anti-derivative is $F(X) = x^2/16 - 0.75x + c$.

We can confirm it is a probability by evaluating it on the domain

$$[x = 6, x = 10 = \frac{10^2}{16} - 0.75 \times 10 - \frac{6^2}{16} + 0.75 \times 6 = -1.25 - (-2.25) = 1.0$$

$$P[8 \leq x \leq 9] = [\frac{9^2}{16} - 0.75 \times 9] - [\frac{8^2}{16} - 0.75 \times 8] = -1.68750 - (-2.000) = 31.250\%$$

9. Solution: A

$$E(Y|X = 7) = 10 \times \frac{0.05}{0.32} + 20 \times \frac{0.03}{0.32} + 30 \times \frac{0.13}{0.32} + 40 \times \frac{0.11}{0.32} = 29.375.$$

$$E(Y^2|X = 7) = 10^2 \times \frac{0.05}{0.32} + 20^2 \times \frac{0.03}{0.32} + 30^2 \times \frac{0.13}{0.32} + 40^2 \times \frac{0.11}{0.32} = 968.75$$

$\text{Variance}(Y|X = 7) = 968.75 - 29.375^2 = 105.8594.$

$\text{Std. Dev}(Y|X = 7) = \sqrt{105.8594} = 10.289.$

10. Solution: B

Although we can sum the product of $f(x) \times X \times Y$ for all nine cells, it is much easier to seize on the property of independence: if X and Y are independent, then by definition $E(XY) = E(X) \times E(Y) = 3.0 \times 7.0 = 21.0$.

11. Solution: D

Variance is a measure of the mean deviation. In the above four graphs, it can be seen that D has the highest proportion of the distribution that deviates from the mean, and it also has a relatively higher density in both tails. Hence, D has the highest variance.

12. Solution: A

Covariance is a measure of how the variables move together.

$\mathrm{Cov}(A,B) = \beta_{A,1}\beta_{B,1}\sigma_{F1}^2 + \beta_{A,2}\beta_{B,2}\sigma_{F2}^2 + (\beta_{A,1}\beta_{B,2} + \beta_{A,2}\beta_{B,1})\mathrm{Cov}(F_1, F_2)$

$= 0.7 \times 0.85 \times 0.3424 + 0.3 \times 0.55 \times 0.0079 + (0.7 \times 0.55 + 0.3 \times 0.85) \times 0.0122$

$= 0.213$

13. Solution: A

$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 11 - 3 \times 4 = -1$

14. Solution: C

Expected return of Stock B =0.4×0.5 + 0.3×0.2 + 0.3×(-0.3) = 0.17

Standard deviation (R_B) =$\sqrt{0.4 \times (0.5-0.17)^2 + 0.3 \times (0.2-0.17)^2 + 0.3 \times (-0.3-0.17)^2} = 0.3318$

15. Solution: B

$\mathrm{Cov}(R_A,R_B)$ = 0.4×(-0.1 – 0.08) ×(0.5 – 0.17) + 0.3×(0.1 – 0.08) ×(0.2 – 0.17) + 0.3×(0.3 – 0.08) ×(-0.3 – 0.17) = -0.0546

16. Solution: D

The analyst's statement is incorrect in reference to either portfolio. Portfolio A has a kurtosis of less than 3, indicating that it is less peaked than a normal distribution (platykurtic). Portfolio B is positively skewed (long tail on the right side of the distribution).

17. Solution: C

A lognormal distribution is positively skewed because it cannot contain negative values. The returns on a long call position cannot be more negative than the premium paid for the option but has unlimited potential positive value, so it will also be positively skewed.

18. Solution: D

A is incorrect. A leptokurtic distribution has fatter tails than the normal distribution. The kurtosis indicates the level of fatness in the tails, the higher the kurtosis, the fatter the tails. Therefore, the probability of exceeding a specified extreme value will be higher.

B is incorrect. Since answer A. has a lower kurtosis, a distribution with a kurtosis of 8 will necessarily produce a larger probability in the tails.

C is incorrect. By definition, a normal distribution has thinner tails than a leptokurtic distribution and larger tails than a platykurtic distribution.

D is correct. By definition, a platykurtic distribution has thinner tails than both the normal distribution and any leptokurtic distribution. Therefore, for an extreme value X, the lowest probability of exceeding it will be found in the distribution with the thinner tails.

19. Solution: C

20. Solution: C

The question considers a skewed leptokurtic distribution. To measure the magnitude of these skewed tails, the analyst needs to consider both the skewness and kurtosis.

21. Solution： C

Since the bond defaults are independent and identically distributed Bernoulli random variables, the Binomial distribution can be used to calculate the probability of exactly two bonds defaulting. The correct formula to use is:

$$P(K = k) = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

where n is the number of bonds in the portfolio, p is the probability of default of each individual bond, and K is the number of bond defaults over the next year. Thus, this question requires： P (K=2) with n = 5 and p = 0.17. Entering the variables into the equation, this simplifies to 10×0.172×0.833 = 0.1652

22. Solution: B

$P(X = 0 \text{ successes}) = 21.46\%$

$P(X = 1 \text{ success}) = 33.89\%$

such that $P(X \leq 1) = 21.46\% + 33.89\% = 55.35\%$.

23. Solution: B

Letting n equal the number of bonds in the portfolio and p equal the individual default probability, the formulas to use are as follows:
Mean = E(K) = n x p = 5 x 0.17 = 0.85.
Variance = Variance(K) = n x p x (1-p) = 5 x 0.17 x (0.83) = 0.7055
Standard deviation =$\sqrt{0.7055} = 0.8399$

24. Solution: C

The result would follow a Binomial distribution as there is a fixed number of random variables, each with the same annualized probability of default. It is not a Bernoulli distribution, as a Bernoulli distribution would describe the likelihood of default of one of the individual bonds rather than of the entire portfolio (i.e. a Binomial distribution essentially describes a group of Bernoulli distributed variables). A normal distribution is used to model continuous variables, while in this case the number of defaults within the portfolio is discrete.

25. Solution: D

The probability of an event is between 0 and 1. If these are mutually exclusive events, the probability of individual occurrences are summed. This probability follows a binomial distribution with a p-parameter of 0.2. The probability of getting at least three questions correct is 1 – [p(0) + p (1) + p (2)] = 32.2%.

26. Solution: A

To solve this question, we first need to realize that the expected number of phone calls in an 8-hour day is λ = 2 × 8 = 16. Using the Poisson distribution, we solve for the probability that X will be 20.

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

P(X = 20) = 0.0559 = 5.59%

27. Solution: C

$P(X > 1) = 1 - P(X = 0) - P(X = 1)$ where $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$ such that

$$P(X = 0) = \frac{(8/5)^0 e^{-\frac{8}{5}}}{0!} = 20.19\%, \quad P(X = 1) = \frac{(8/5)^1 e^{-\frac{8}{5}}}{1!} = 32.3\%$$

and $P(X > 1) = 1 - 20.19\% - 32.30\% = 47.51\%$. Notice we need to translate the mean frequency from eight per week to 8/5 per day because we are seeking the probability of the number of occurrences in a single day.

28. Solution: C

Bernoulli (lowest) and Poisson (highest)

In order:

$\text{Bernoulli has variance} = p(1-p) = 4\% \times 96 = 0.0384$

$\text{Binomial has variance} = p(1-p)n = 2\% \times 98\% \times 50 = 0.980$

$\text{Standard normal has, by definition, mean} = 0 \text{ and variance} = 1.0$

$\text{Student's t has variance} = df/(df-2) = 10/8 = 1.25$

$\text{Poisson has lambda} = \text{variance} = \text{mean} = 5$

29. Solution: D

Normal variables are stable under addition, so that (I) is true. For lognormal variables X_1 and X_2, we know that their logs, $\ln(X_1)$ and $\ln(X_2)$ are normally distributed. Hence, the sum of their logs, or $\ln(X_1) + \ln(X_2) = \ln(X_1 \times X_2)$must also be normally distributed. The product is itself lognormal, so that (IV) is true.

30. Solution: D

A binomial distribution is a probability distribution, and it refers to the various probabilities associated with the number of correct answers out of the total sample.

The correct approach is to find the cumulative probability for 8 in a binomial distribution with N = 20 and p = 0.5. The cumulative probability is to be calculated on the basis of a binomial distribution with number of questions (n) equaling 20 and probability of a single event occurring being 50% (p = 0.5).

31. Solution: B

Letting n equal the number of bonds in the portfolio and p equal the individual default probability, the formulas to use are as follows:

Mean = n ×p = 3 ×15% = 0.45. Variance = np (1-p) = 3 ×0.15 ×0.85 = 0.3825

32. Solution: D

$$\text{Pro}(32 \leq X \leq 116) = \text{Pro}\left(\frac{32-80}{24} \leq \frac{X-80}{24} \leq \frac{116-80}{24}\right)$$

$$\text{Pro}(-2 \leq z \leq 1.5) = 0.9104$$

33. Solution: C

Since these are independent normally distributed random variables, the combined expected mean return is: $\mu = 0.2 \times 3\% + 0.8 \times 7\% = 6.2\%$

Combined volatility is: $\sigma = \sqrt{0.2^2 0.07^2 + 0.8^2 0.15^2} = 0.121 = 12.1\%$

The appropriate Z-statistic is $Z = \frac{26\% - 6.2\%}{12.1\%} = 1.64$

And therefore P (Z>1.64) = 1-0.95 = 0.05 = 5%

34. Solution: B

Both approach a normal distribution; all of the so-called sampling distributions (student t, chi-square, F) approach normal as d.f. increase.

35. Solution: C

The standard error of the sample mean is the standard deviation of the distribution of the sample means. And it is calculated as:

$\sigma_{\bar{X}} = \sigma/\sqrt{n}$, where σ, the population standard deviation is known.

$S_{\bar{X}} = S/\sqrt{n}$, where S, is the sample standard deviation.

So, standard error of the mean $= S/\sqrt{n} = 0.64$

36. Solution: C

In order to calculate the standard deviation of the mean weekly returns, we must divide the standard deviation of the return series by the square root of the sample size. Therefore, the correct answer is $\frac{15\%}{\sqrt{25}} = 3\%$

37. Solution: D

Here the t-reliability factor is used since the population variance is unknown. Since there are 30 observations, the degrees of freedom are 30 – 1 = 29. The t-test is a two-tailed test. So the correct critical t-value is $t_{29.25} = 2.045$, thus the 95% confidence interval for the mean return is:

$$\left[4\% - 2.045\left(\frac{20\%}{\sqrt{30}}\right), 4\% + 2.045\left(\frac{20\%}{\sqrt{30}}\right)\right] = [-3.464\%, 11.464\%]$$

38. Solution: A

The confidence interval is equal to the mean monthly return plus or minus the t-statistic times the standard error. To get the proper t-statistic, the 0.025 column must be used since this is a two-tailed interval. Since the mean return is being estimated using the sample observations, the appropriate degrees of freedom to use is equal to the number of sample observations minus 1. Therefore we must use 11 degrees of freedom and therefore the proper statistic to use from the t-distribution is 2.201.

The proper confidence interval is: -0.75% + /- (2.201 × 2.70%) or -6.69% to + 5.19%.

39. Solution: B

When the population mean and population variance are not known, the t-statistic can be used to analyze the distribution of the sample mean.

$\text{Sample mean} = (10 + 80 + 90 - 60 + 30)/5 = 30$

$$\text{Unbiased sample variance} = \frac{(10-30)^2 + (80-30)^2 + (90-30)^2 + (-60-30)^2 + (30-30)^2}{4}$$

$$= 3650$$

$\text{Unbiased sample standard deviation} = 60.4152$

$\text{Sample standard error} = (\text{sample standard deviation})/\sqrt{5} = 27.0185$

$\text{Population mean of return distribution} = 4.5\ (\text{million HKD})$

$\text{Therefore the t} - \text{statistic} = (\text{Sample mean} - \text{population mean})/\text{Sample standard error}$

$$= (30 - 4.5)/27.02 = 0.9438.$$

Because we are using the sample mean in the analysis, we must remove 1 degree of freedom before consulting the t-table; therefore 4 degrees of freedom are used. According to the table, the closest possibility is 0.2 = 20%.

40. Solution: D

The correct test is:

Null Hypothesis	Alternative Hypothesis	Critical Region, reject the null if:
$\sigma^2 = 4\%^2 = 0.0016$	$\sigma^2 \neq 0.0016$	$\frac{(24-1)(0.05)^2}{(0.04)^2} > \chi^2_{2.5,24-1} = 35.9375 < 38.0757$

Therefore, you would not reject the null hypothesis. A chi-square test is a statistical hypothesis test whereby the sampling distribution of the test statistic is a chi-squared distribution when the null hypothesis is true.

41. Solution: B

The formula for the Chi-squared test statistic is:

(n - 1) ×(sample variance / hypothesis variance)

Since we are given a daily standard deviation, we must first annualize it by multiplying it by the square root of the number of trading days. Therefore:

Sample volatility = sqrt (260) 2% = 32.25%

And the Chi-squared test statistic = (21 - 1) $\times 0.3225^2/0.25^2$ = 33.28

42. Solution: A

The required test for testing the variance is the chi-squared test.

$$\text{test statistic} = (n-1)\frac{\text{sample variance}}{\text{hypothesized variance}} = 60 \times \frac{21\%^2}{14\%^2} = 135$$

To test whether the standard deviation is higher (H0: standard deviation is lower than or equal to 14%), the critical value of chi-squared will be 79.08 (using df = 60 and p = 0.05). Since the test statistic is higher than the critical value, the analyst can reject the null hypothesis and concludes that the standard deviation of returns is higher than 14%.

43. Solution: B

A is incorrect. The denominator of the z-test statistic is standard error instead of standard deviation. If the denominator takes the value of standard deviation 4.5, instead of standard error 4.5/sqrt(81), the z-test statistic computed will be z = 0.29, which is incorrect.

B is correct. The population variance is unknown but the sample size is large (>30). The test statistics is: z = (46.3 - 45) / (4.5 / (sqrt (81)) = 2.60. Decision rule: reject H0 if z(computed)>z(critical). Therefore, reject the null hypothesis because the computed test statistics of 2.60 exceeds the critical z-value of 2.33.

C is incorrect because z-test (instead of t-test) should be used for sample size (81) ≥30.

D is incorrect because z-test (instead of t-test) should be used for sample size (81)≤30.

44. Solution: D

Tests of the variance (or standard deviation) of a population requires the chi-squared test.

45. Solution: B

If the probability distribution of an estimator has an expected value equal to the parameter it is supposed to be estimating, it is said to be unbiased.

Between two candidate estimators, the one with a smaller variance is said to use the information in the data more efficiently.

When the probability that an estimator is within a small interval of the true value approaches 1, it is said to be a consistent estimator.

46. Solution: D

47. Solution: C

48. Solution: A

Decreasing the level of significance of the test decreases the probability of making a Type I error and hence makes it more difficult to reject the null when it is true. However, 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, because the null is rejected less frequently, even when it is actually false.

49. Solution: B

A Type I error is the error of rejecting a hypothesis when it is true. A Type II error is the error of accepting a false hypothesis.

A decrease in the level of significance decreases P(Type I error) but increases P(Type II error). Reducing the sample size increases P(Type II error)..

50. Solution: D

The power of the tat refers to the probability of rejecting an incorrect model, which is one minus the probability of not rejecting an incorrect model. Given that the power of the test is 87 percent, the probability of a type II error, the probability of not rejecting the incorrect model is 1.0 - 0.87 = 13%.

51. Solution: B

The regression coefficients for a model specified by $\mathrm{Y} = \mathrm{a} + \mathrm{bX} + \varepsilon_i$ are obtained using the formula: $\mathrm{b} = \mathrm{S_{XY}}/\mathrm{S_X^2}$, In this example: $\mathrm{S_{XY}} = 0.06$, $\mathrm{S_X} = 0.18$, $\mathrm{E(Y)} = 0.11$, Then: $\mathrm{b} = 0.06/(0.18)^2 = 1.85$, $\mathrm{a} = \mathrm{E(Y)} - \mathrm{b} \times \mathrm{E(X)} = 0.11 - 1.85 \times 0.07 = -0.02$Where $\varepsilon_i$represents the error term.

52. Solution: C

The T-test would not be sufficient to test the joint hypothesis. In order to test the joint null hypothesis, examine the F-statistic, which in this case is statistically significant at the 95% confidence level. Thus the null can be rejected.

53. Solution: A

The correct test is the t test. The t statistic is defined by:

$$\mathrm{t} = \frac{\beta_{\mathrm{estimated}} - \beta}{\mathrm{SE(estimated\ \beta)}} = \frac{0.86 - 1}{0.8}$$

In this case t = -0.175. Since |t| < 1.96 we cannot reject the null hypothesis.

54. Solution: A

Omitted variable bias occurs when a model improperly omits one or more variables that are critical determinants of the dependent variable and are correlated with one or more of the other included independent variables. Omitted variable bias results in an over-or under-estimation of the regression parameters.

55. Solution: B

Added a regressor has an unclear impact on the adjusted R^2 (however, the adjusted R^2 is always LESS THAN the R^2).

56. Solution: B

One of the assumptions of the multiple regression model of least squares is that no perfect multicollinearity is present. Perfect multicollinearity would exist if one of the regressors is a perfect linear function of the other regressors.

None of the other choices are assumptions of the multiple least squares regression model.

57. Solution: C

In the OLS model, the variance of the independent variable is assumed to be uncorrelated with the variance of the error term.

In the OLS model, it is assumed that the correlation between the dependent variable and the random error term is constant.

58. Solution: D

The t-statistics for the intercept and coefficient on X_{2i} are significant as indicated by the associated p-values being less than 0.05: 0.0007 and 0.0004 respectively. Therefore, H0: B_0 = 0 and H_0: B2 = 0 can be rejected. The F-statistic on the ANOVA table has a p-value equal to 0.0012; therefore, H0: B_1 = B_2 = 0 can be rejected. The p-value for the coefficient on X_{1i} is greater than five percent; therefore, H_0: B_1 = 0 cannot be rejected.

59. Solution: B

60. Solution: A

Stratification is not related to regression analysis. Choices B, C, and D describe situations that can produce inaccurate descriptions of the relationship between the independent and dependent variables. Multicollinearity occurs when the independent variables are themselves correlated, Heteroscedasticity occurs when the variances are different across observations, and autocorrelation occurs when successive observations are influenced by the proceeding observations.

61. Solution: A

62. Solution: B

The R^2 of the regression is calculated as ESS/TSS = (92.648/117.160) = 0.79, which means that the variation in industry returns explains 79% of the variation in the stock return. By taking the square root of R^2, we can calculate that the correlation coefficient (r) = 0.889. The t-statistic for the industry return coefficient is 1.91/0.31 = 6.13, which is sufficiently large enough for the coefficient to be significant at the 99% confidence interval. Since we have the regression coefficient and intercept, we know that the regression equation is R_{stock} = 1.9X + 2.1. Plugging in a value of 4% for the industry return, we get a stock return of 1.9 (4) + 2.1 = 9.7%.

63. Solution: C

64. Solution: C

An estimated coefficient of 0.24 from a linear regression indicates a positive relationship between income and savings, and more specifically means that a one unit increase in the independent variable (household income) implies a 0.24 unit increase in the dependent variable (annual savings). Given the equation provided, a household with no income would be expected to have negative annual savings of GBP 25.66. The error term mean is assumed to be equal to 0.

65. Solution: D

The OLS procedure is a method for estimating the unknown parameters in a linear regression model.

The method minimizes the sum of squared differences between the actual, observed, returns and the returns estimated by the linear approximation. The smaller the sum of the squared differences between observed and estimated values, the better the estimated regression line fits the observed data points.

66. Solution: D

This is an example of multicollinearity, which arises when one of the regressors is very highly correlated with the other regressors. In this case, all three regressors are highly correlated with each other, so multicollinearity exists between all three. Since the variables are not perfectly correlated with each other this is a case of imperfect, rather than perfect, multicollinearity.

67. Solution: B

High R^2 (implies F test will probably reject null) but low significance of partial slope coefficients is the classic symptom of multicollinearity

68. Solution: A

The error terms have a non-constant variance so they are heteroscedastic.

69. Solution: D

Durbin-Watson is classic test for autocorrelation. Null hypothesis is No positive/negative autocorrelation.

70. Solution: D

71. Solution: B

Whenever we want to distinguish between s seasons in a model that incorporates an intercept, we must use s- 1 dummy variables. For example, if we have quarterly data, s = 4, and thus we would include s- 1 = 3 seasonal dummy variables.

72. Solution: A

The intercept term represents the average value of EPS for the first quarter. The slope coefficient on each dummy variable estimates the difference in EPS (on average) between the respective quarter (i.e., quarter 2, 3or 4) and the omitted quarter (the first quarter, in this ease).

73. Solution: C

Including the full set of dummy variables and an intercept term would produce a forecasting model that exhibits perfect multicollinearity. For Option I/II/III, there is only one input in the model, so there will not exist multicollinearity effect.

74. Solution: A

First, zero-mean white noise may be uncorrelated but not necessarily serially independent (the difference between correlation and independence). Second, white noise (aka, zero-mean white noise) is not necessarily normally distributed.

75. Solution: C

One requirement for a series to be covariance stationary is that its covariance structure be stable over time. If the covariance structure is stable, then the autocovariances depend only on displacement, τ, not on time, t. Also, covariance stationarity does not place restrictions on other aspects of the distributions or the series, such as kurtosis and skewness.

76. Solution: A

Both autoregressive (AR) models and autoregressive moving average (ARMA) models are good at forecasting with seasonal patterns because they both involve lagged observable variables, which are best for capturing a relationship in motion. It is the moving average representation that is best at capturing only random movements.

77. Solution: A

CHISQ.INV(0.95, 24) = 36.41 such that both statistics are less than the critical values; i.e., fall into the acceptance region of the chi-squared distribution.

78. Solution: C

The following table provides the ranking of pairs with respect to X.

Year	X	Y	X Rank	Y Rank
2013	-20.0%	40.0%	1	5
2012	-10.0%	20.0%	2	4
2010	5.0%	-10.0%	3	1
2014	30.0%	15.0%	4	3
2011	50.0%	-5.0%	5	2

There are two concordant pairs and eight discordant pairs shown as follows:

Concordant Pairs:

{(3,1), (4,3)}; {(3,1), (5,2)}

Discordant Pairs:

{(1,5), (2,4)};{(1,5), (3,1)};{(1,5), (4,3)};{(1,5), (5,2)};{(2,4), (3,1)};{(2,4), (4,3)};{(2,4), (5,2)};{(4,3), (5,2)}

Thus, the Kendall τ correlation coefficient is -0.2

$$\tau = \frac{n_c - n_d}{n\,(n-1)\,\div 2}$$

79. Solution: A

The following table illustrates the calculations used to determine the sum of squared ranking deviations:

Year	X	Y	X Rank	Y Rank	d_i	d_i^2

2013	-20.0%	40.0%	1	5	-4	16
2012	-10.0%	20.0%	2	4	-2	4
2010	5.0%	-10.0%	3	1	2	4
2014	30.0%	15.0%	4	3	1	1
2011	50.0%	-5.0%	5	2	3	2
					Sum	34

Thus, the Spearman rank correlation coefficient is -0.7:

$$\rho_S = 1 - \frac{6 \times \sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6 \times 34}{5 \times (25-1)} = -0.7$$

80. Solution: B

Following the risk neutral valuation methodology, the price of the derivative is obtained by calculating the weighted average nine month payoff and then discounting this figure by the risk free rate.

Average payoff calculation: (2000×43 + 4000×44 + 10000×46 + 4000×45)/20000 = 45.10

Discounted payoff calculation: $45.10 \times \mathrm{e}^{-0.04\times(9/12)} = 43.77$

81. Solution: C

Monte Carlo simulations cannot price options with early exercise accurately. All of the other statements are correct. Correlation can be incorporated using the method of Cholesky decomposition, Monte Carlo simulations can be designed to handle time varying volatility, and Monte Carlo simulations are computationally more intensive than historic simulations.

82. Solution: C

$$\text{Mean} = \frac{\mu}{\text{n}} = \frac{15\%}{52} = 0.288\%$$

$$\text{SD} = \frac{\sigma}{\sqrt{\text{n}}} = \frac{30\%}{\sqrt{52}} = 4.16\%$$

83. Solution: D

Short option positions have long left tails, which makes it more difficult to estimate a left-tailed quantile precisely. Accuracy with independent draws increases with the square root of K. Thus increasing the number of replications should shrink the standard error, so answer B is incorrect.

84. Solution: B

Risk neutrality has nothing to do with sample size.