FRM Part I Exam

By AnalystPrep

Questions with Answers - Quantitative Analysis

Last Updated: Aug 27, 2022

Table of Contents

Reading 12: Fundamentals of Probability

Q.315 The probability of an increase in the annual dividend paid out to shareholders of ABC Limited is 0.4. The probability of an increase in share price given an increase in dividends is 0.7. Determine the joint probability of an increase in dividends and an increase in the share price.

A. 0.28

B. 0.14

C. 0.72

D. 0.3

The correct answer is **A.**

Let:
A be the event that the dividend is increased and,
B be the event that the share price increases
Therefore, $P(A) = 0.4$ and $P(B \mid A) = 0.7$
The joint probability of an increase in dividends and an increase in share price is $P(B \cap A)$
The multiplication rule of probability states that:

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

Hence

$$P(B \cap A) = P(B|A) * P(A) = 0.7 * 0.4 = 0.28 \text{ or } 28\%$$

(Note that $P(A \cap B) = P(B \cap A)$.)

Q.316 Two events are said to be independent if:

A. They cannot occur at the same time.

B. The occurrence of one of the events affects the probability of occurrence of the other event.

C. The occurrence of one of the events does not affect the probability of occurrence of the other event .

D. The occurrence of one of the events means the second event is certain to occur.

The correct answer is **C.**

Independent events are those whose occurrences are uncorrelated. The occurrence of one event does not in any way affect the chances of the other event occurring. In addition, if two events are mutually exclusive, it means they cannot occur at the same time. For example, if a fair coin is tossed, it's impossible to obtain a head and a tail at the same time since the two possible outcomes are mutually exclusive.

Q.317 A financial risk manager exam candidate is asked two questions. The probability that she gets the first question correct is 0.4 and the probability that she gets the second question correct is 0.5. Given that the probability that she gets both questions correct is 0.2, determine the probability that she gets either the first or the second question correct.

A. 0.9

B. 0.7

C. 0.1

D. 0.4

The correct answer is **B.**

Let A = {gets first question right} and B = {gets second question right}

Therefore, $P(A) = 0.4, P(B) = 0.5$, and $P(A \cap B) = 0.2$

We want to determine $P(A \cup B)$;

The addition rule states that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Hence,

$$P(A \cup B) = 0.4 + 0.5 - 0.2 = 0.7$$

Q.318 An empirical study of ABC stock listed on the New York Exchange reveals that the stock has closed higher on one-third of all days in the past few months. Given that up and down days are independent, determine the probability of ABC stock closing higher for six consecutive days.

A. 0.17

B. 0.0137

C. 0.03704

D. 0.00137

The correct answer is **D.**

From the information above, we can establish that the probability of closing higher = $\frac{1}{3}$

Using independence, the probability of 6 consecutive "highs" = $\frac{1}{3} * \frac{1}{3} * \frac{1}{3} * \frac{1}{3} * \frac{1}{3} * \frac{1}{3} = \frac{1}{729}$

(The calculation above follows from the fact that if A and B are independent events, then $P(A \cap B) = P(A) * P(B)$.)

Q.319 A fruit juice shop allows customers to choose apple juice, mango juice or passion juice. The probability of a customer ordering passion juice is 0.45, mango juice and apple juice 0.19, passion juice and mango juice 0.15, passion juice and apple juice 0.25, passion juice or mango juice 0.6, passion juice or apple juice 0.84, and 0.9 for at least one of them.
Find the probability that a customer orders all the three juices.

A. 0.64

B. 0.1

C. 0.3

D. 0.25

The correct answer is **B.**

Let:

A be the event that a customer chooses/orders apple juice

M be the event that a customer chooses mango juice

S be the event that a customer chooses passion fruit

We can easily establish that:

$P(S) = 0.45, P(M \cap A) = 0.19, P(M \cap S) = 0.15,$
$P(A \cap S) = 0.25, P(M \cup S) = 0.6, P(A \cup S) = 0.84, P(A \cup M \cup S) = 0.9$

We need to determine $P(A \cap M \cap S)$:

Borrowing from the addition rule with three sets,

$$P(A \cup M \cup S) = P(A) + P(M) + P(S) - P(M \cap A) - P(M \cap S) - P(A \cap S) + P(A \cap M \cap S) \ldots \ldots \text{equation (}$$

$$P(M \cup S) = P(M) + P(S) – P(M \cap S),$$
$$P(M) = 0.6 + 0.15 – 0.45 = 0.3$$

Similarly,

$$P(A \cup S) = P(A) + P(S) - P(A \cap S)$$
$$P(A) = 0.84 – 0.45 + 0.25 = 0.64$$

Therefore applying equation (I),

$$0.9 = 0.64 + 0.3 + 0.45 – 0.19 – 0.15 – 0.25 + P(A \cap M \cap S)$$

Which gives us $P(A \cap M \cap S) = 0.1$

Q.321 During a lottery, 400 names are fed into a computer program. Five of the names are identical. If a name is drawn from the program at random, what is the probability that one of these 5 names will be drawn?

A. 0.0125

B. 0.25

C. 0.0025

D. 0.0625

The correct answer is **A.**

$$P(\text{name } 1 \cap \text{name } 2 \cap \text{name } 3 \cap \text{name } 4 \cap \text{name } 5) = \frac{1}{400} + \frac{1}{400} + \frac{1}{400} + \frac{1}{400} + \frac{1}{400} = 0.0125$$

Q.322 If two events are not independent, the joint probability of A and B, $P(A \cap B)$ is equal to:

A. $\frac{P(A|B)}{P(B)}$

B. P(A | B) * P(B)

C. P(A) * P(B)

D. None of the above

The correct answer is **B.**

If two events are not independent, then the occurrence of one of the events affects the chances of occurrence of the other event.
As such, $P(A|B) = \frac{P(A \cap B)}{P(B)}$
Which can be rewritten to make $P(A \cap B)$ the subject of the formula.

If A and B were independent, then $P(A \cap B) = P(A) * P(B)$

Q.323 Use the following incomplete probability matrix to compute the joint probability of a poor economy and an increase in interest rate:

Economy	Interest Rate-Increase	Interest Rate-No Increase
Good	20%	10%
Normal	30%	20%
Poor	X	10%

A. 0.2

B. 0.3

C. 0.1

D. 0.05

The correct answer is **C.**

All the joint probabilities must add up to 1. Therefore,

$$X = 100\% - 20\% - 30\% - 10\% - 20\% - 10\% = 10\%$$

Q.324 Calculate the probability that a stock return is either below -5% or above 5%, given:
$P(R < -5\%) = 16\%$
$P(R > +5\%) = 18\%$

A. 0.02

B. 0.32

C. 0.34

D. 0.16

The correct answer is **C.**

Since the two events are mutually exclusive, the return cannot be below -5% or above 5% at the same time. Therefore, the answer is simply 16% + 18% = 34%

Q.356 A financial risk manager has three routes to get to the office. The probability that she gets to the office on time using routes X, Y, and Z are 60%, 65%, and 70%. She does not have a preferred route and is therefore equally likely to choose any of the three routes. Calculate the probability that she chose route Z given that she arrives to work on time.

A. 0.359

B. 0.233

C. 0.216

D. 0.2

The correct answer is **A.**

Define X to be the event "chooses route X." Let Y and Z have similar definitions.
Define O to be the event that she arrives on time

We wish to determine P(Z|O). Then:

$$
\begin{aligned}
P(Z|O) &= \frac{P(Z) * P(O|Z)}{[P(Z) * P(O|Z) + P(Y) * P(O|Y) + P(X) * P(O|X)]} \\
&= \frac{(\frac{1}{3} * 0.7)}{[(\frac{1}{3} * 0.7) + (\frac{1}{3} * 0.65) + (\frac{1}{3} * 0.6)]} \\
&= \frac{0.2333}{(0.2333 + 0.2167 + 0.2)} \\
&= 0.3589
\end{aligned}
$$

Q.357
A life assurance company insures individuals of all ages. A manager compiled the following statistics of the company's insured persons:

Age of insured	Mortality (Probability of death) [arbitrary]	Portion of company's insured persons
16-20	0.04	0.10
21-30	0.05	0.29
31-65	0.10	0.49
66-99	0.14	0.12

If a randomly selected individual insured by the company dies, calculate the probability that the dead client was age 16-20.

A. 0.9525

B. 0.0593

C. 0.0630

D. 0.0475

The correct answer is **D.**

Define the following events:

B = Event of death

B_1= Event the insured's age is in the range 16-20

B_2= Event the insured's age is in the range 21-30

B_3= Event the insured's age is in the range 31-65

B_4= Event the insured's age is in the range 66-99

We wish to determine $P(B_1|B)$

$$
\begin{aligned}
P(B_1|B) &= \frac{(P(B_1) * P(B|B_1))}{[P(B_1) * P(B|B_1) + (P(B_2) * P(B|B_2) + (P(B_3) * P(B|B_3) + (P(B_4) * P(B|B_4)]} \\
&= \frac{(0.1 * 0.04)}{[(0.1 * 0.04) + (0.29 * 0.05) + (0.49 * 0.1) + (0.12 * 0.14)} \\
&= \frac{0.004}{(0.004 + 0.0145 + 0.049 + 0.0168)]} \\
&= 0.04745 \text{ or } 4.7\%
\end{aligned}
$$

Q.358
A life assurance company insures individuals of all ages. A manager compiled the following statistics of the company's insured persons:

Age of insured	Mortality (Probability of death) [arbitrary]	Portion of company's insured persons
16 − 20	0.04	0.10
21 − 30	0.05	0.29
31 − 65	0.10	0.49
66 − 99	0.14	0.12

If a randomly selected individual insured by the company dies, calculate the probability that the dead client was in age range 21-30.

A. 0.172

B. 0.04

C. 0.168

D. 0.145

The correct answer is **A.**

We wish to determine $P(B_2|B)$

$$\begin{aligned} P(B_2|B) &= \frac{(P(B_2) * P(B|B_2))}{[P(B_2) * P(B|B_2) + (P(B_1) * P(B|B_1) + (P(B_3) * P(B|B_3) + (P(B_4) * P(B|B_4)]} \\ &= \frac{(0.29 * 0.05)}{[(0.29 * 0.05) + (0.1 * 0.04) + (0.49 * 0.1) + (0.12 * 0.14)} \\ &= \frac{0.0145}{(0.0145 + 0.004 + 0.049 + 0.0168)} \\ &= 17.2\% \end{aligned}$$

Q.359

A life assurance company insures individuals of all ages. A manager compiled the following statistics of the company's insured persons:

Age of insured	Mortality (Probability of death) [arbitrary]	Portion of company's insured persons
16-20	0.04	0.10
21-30	0.05	0.29
31-65	0.10	0.49
66-99	0.14	0.12

Compute the probability that the dead client was in the age range 31-65.

A. 0.58

B. 0.172

C. 0.168

D. 0.047

The correct answer is **A.**

We wish to determine $P(B_3|B)$

$$P(B_3|B) = \frac{(P(B_3) * P(B|B_3))}{[P(B_3) * P(B|B_3) + (P(B_1) * P(B|B_1) + (P(B_2) * P(B|B_2) + (P(B_4) * P(B|B_4))]}$$

$$= \frac{(0.49 * 0.10)}{[0.49 * 0.10 + 0.1 * 0.04 + 0.29 * 0.05 + 0.12 * 0.14]}$$

$$= \frac{0.049}{(0.049 + 0.004 + 0.0145 + 0.0168)}$$

$$= 58\%$$

Q.360 A life assurance company insures individuals of all ages. A manager compiled the following statistics of the company's insured persons:

Age of insured	Mortality (Probability of death) [arbitrary]	Portion of company's insured persons
16-20	0.04	0.10
21-30	0.05	0.29
31-65	0.10	0.49
66-99	0.14	0.12

Calculate the probability that the dead client was between 66 and 99 years.

A. 0.047

B. 0.172

C. 0.12

D. 0.201

The correct answer is **D.**

Define the following events:

B = Event of death

B_1= Event the insured's age is in the range 16-20

B_2= Event the insured's age is in the range 21-30

B_3= Event the insured's age is in the range 31-65

B_4= Event the insured's age is in the range 66-99

$$
\begin{aligned}
P(B_1|B) &= \frac{(P(B_1) * P(B|B_1))}{[P(B_1) * P(B|B_1) + (P(B_2) * P(B|B_2) + (P(B_3) * P(B|B_3) + (P(B_4) * P(B|B_4)]} \\
&= \frac{(0.1 * 0.04)}{[(0.1 * 0.04) + (0.29 * 0.05) + (0.49 * 0.1) + (0.12 * 0.14)} \\
&= \frac{0.004}{(0.004 + 0.0145 + 0.049 + 0.0168)} \\
&= 0.04745 \text{ or } 4.7\% \\
P(B_2|B) &= \frac{(P(B_2) * P(B|B_2))}{[P(B_2) * P(B|B_2) + (P(B_1) * P(B|B_1) + (P(B_3) * P(B|B_3) + (P(B_4) * P(B|B_4)]} \\
&= \frac{(0.29 * 0.05)}{[(0.29 * 0.05) + (0.1 * 0.04) + (0.49 * 0.1) + (0.12 * 0.14)} \\
&= \frac{0.0145}{(0.0145 + 0.004 + 0.049 + 0.0168)} \\
&= 17.2\% \\
P(B_3|B) &= \frac{(P(B_3) * P(B|B_3))}{[P(B_3) * P(B|B_3) + (P(B_1) * P(B|B_1) + (P(B_2) * P(B|B_2) + (P(B_4) * P(B|B_4)]} \\
&= \frac{(0.49 * 0.10)}{[0.49 * 0.10 + 0.1 * 0.04 + 0.29 * 0.05 + 0.12 * 0.14]} \\
&= \frac{0.049}{(0.049 + 0.004 + h0.0145 + 0.0168)} \\
&= 58\%
\end{aligned}
$$

Thus,

$P(B_4|D) = 100 – 58 – 17.2 – 4.7 = 20.1\%$ (Sum of all the probabilities must add up to 1).

Q.361 An investment firm classifies capital projects into three different categories, depending on risk level: Standard, Preferred, and Ultra-preferred. Of the firm's projects, 60% are standard, 30% are preferred, and 10% are ultra-preferred. The probabilities of a project making a loss are 0.01, 0.005, and 0.001 for categories standard, preferred, and ultra-preferred respectively.
If a capital project makes a loss in the next year, then what is the probability that the project was standard (correct to 2 decimal places)?

A. 0.79

B. 0.73

C. 0.22

D. 0.15

The correct answer is **A.**

Let:

L = Event a project makes a loss

S = Event of a standard project

P_1 = Event of a preferred project

U = Event of a ultra-preferred project

We wish to determine $P(S|L)$

$$\begin{aligned} P(S|L) &= \frac{(P(S) * P(L|S))}{[P(P_1) * P(L|P_1) + P(U) * P(L|U)]} \\ &= \frac{(0.6 * 0.01)}{[(0.6 * 0.01) + (0.3 * 0.005) + (0.1 * 0.001)]} \\ &= \frac{0.006}{[0.006 + 0.0015 + 0.0001]} \\ &= 0.7895 \text{ or } 79\% \end{aligned}$$

Q.363 Upon arrival at a cancer treatment center, patients are categorized into one of four stages, namely: stage 1, stage 2, stage 3, and stage 4. In the past year,
i. 10% of patients arriving were in stage 1
ii. 40% of patients arriving were in stage 2
iii. 30% of patients arriving were in stage 3
iv. The rest of the patients were in stage 4
v. 10% of stage 1 patients died
vi. 20% of stage 2 patients died
vii. 30% of stage 3 patients died
viii. 50% of stage 4 patient died

Given that the patient died, what is the probability that the patient was in stage 4 cancer?

A. 0.86

B. 0.1

C. 0.36

D. 0.5

The correct answer is **C.**

We wish to determine $P(C_4|D)$

$$\begin{aligned}
P(C_4|D) &= \frac{(P(C_4) * P(D|C_4))}{[(P(C_4) * P(D|C_4) + ((P(C_1) * P(D|C_1) + (P(C_2) * P(D|C_2) + (P(C_3) * P(D|C_3)]} \\
&= \frac{(0.2 * 0.5)}{[(0.2 * 0.5) + (0.1 * 0.1) + (0.4 * 0.2) + (0.3 * 0.3)]} \\
&= \frac{0.1}{(0.1 + 0.01 + 0.08 + 0.09)} \\
&= 36\%
\end{aligned}$$

Q.364 You are an analyst at a large mutual fund. After examining historical data, you establish that all fund managers fall into two categories: superstars (S) and ordinaries (O).
Superstars are by far the best managers. The probability that a superstar will beat the market in any given year stands at 70%. Ordinaries, on the other hand, are just as likely to beat the market as they are to underperform it. Regardless of the category in which a manager falls, the probability of beating the market is independent from year to year. Superstars are rare diamonds because only a meager 16% of all recruits turn out to be superstars.

During the analysis, you stumble upon the profile of a manager recruited three years ago, who has since gone on to beat the market every year.

Determine the probability that the manager was a superstar when he was recruited into the fund.

A. 0.5

B. 0.86

C. 0.7

D. 0.16

The correct answer is **D.**

Let:

B = Event that a manager beats the market

S = Event that a superstar is recruited

Therefore,

P(B | S) = 70% = 7/10

P(B | O) = 50% = 1/2

At the time of recruitment, the probability of the manager being a superstar was just the unconditional probability of a manager being a superstar, i.e. P(S) = 16%

Q.365 You are an analyst at a large mutual fund. After examining historical data, you establish that all fund managers fall into 2 categories: superstars (S) and ordinaries (O).
Superstars are by far the best managers. The probability that a superstar will beat the market in any given year stands at 70%. Ordinaries, on the other hand, are just as likely to beat the market as they are to underperform it. Regardless of the category in which a manager falls, the probability of beating the market is independent of year to year. Superstars are rare diamonds because only a meager 16% of all recruits turn out to be superstars.

During the analysis, you stumble upon the profile of a manager recruited 3 years ago, who has since gone on to beat the market every year.

What is the probability that the manager is a superstar as at present?

A. 0.46

B. 0.34

C. 0.84

D. 0.16

The correct answer is **B.**

We need to determine P(S | 3B): The probability that the manager is a superstar given that they have managed to beat the market in three consecutive years. As such, we need to apply Bayes' theorem.

$$P(S|3B) = P(S) * \frac{P(3B|S)}{P(3B)}$$

Now, we already have $P(S) = 16\% = \frac{4}{25}$

$$\begin{aligned} P(3B|S) &= \left(\frac{7}{10}\right)^3 \text{ since performance is independent from one year to the next} \\ &= \frac{343}{1000} \end{aligned}$$

$$\begin{aligned} P(3B) &= \text{unconditional probability of beating the market in 3 consecutive years} \\ &= \text{weighted average probability of 3 marketing-beating years over both superstars and ordinarie} \\ &= P(3B|S) * P(S) + P(3B|O) * P(O) \\ &= \left[\left(\frac{7}{10}\right)^3 * \frac{4}{25}\right] + \left[\left(\frac{1}{2}\right)^3 * \frac{21}{25}\right] \\ &= \left(\frac{343}{1000} * \frac{4}{25}\right) + \left(\frac{1}{8} * \frac{21}{25}\right) \\ &= \frac{1372}{25000} + \frac{21}{200} \\ &= 16\% \end{aligned}$$

Therefore,

$$16\% * \frac{34.3\%}{16\%} = 34.3\% \text{ or } 0.343$$

Q.366 You are an analyst at a large mutual fund. After examining historical data, you establish that all fund managers fall into two categories: superstars (S) and ordinaries (O).
Superstars are by far the best managers. The probability that a superstar will beat the market in any given year stands at 70%. Ordinaries, on the other hand, are just as likely to beat the market as they are to underperform it. Regardless of the category in which a manager falls, the probability of beating the market is independent of year to year. Superstars are rare diamonds because only a meager 16% of all recruits turn out to be superstars.

During the analysis, you stumble upon the profile of a manager recruited three years ago, who has since gone on to beat the market every year.

What is the conditional probability for a non-superstar manager given that he/she has beaten the market for 3 years?

A. 0.66

B. 0.7

C. 0.45

D. 0.64

The correct answer is **A.**

First, we need to determine P(S | 3B): The probability that the manager is a superstar given that they have managed to beat the market in three consecutive years. As such, we need to apply Bayes' theorem.

$$P(S|3B) = P(S) * \frac{P(3B|S)}{P(3B)}$$

Now, we already have $P(S) = 16\% = \frac{4}{25}$

$$\begin{aligned} P(3B|S) &= \left(\frac{7}{10}\right)^3 \text{ since performance is independent from one year to the next} \\ &= \frac{343}{1000} \end{aligned}$$

$$
\begin{aligned}
P(3B) &= \text{unconditional probability of beating the market in 3 consecutive years} \\
&= \text{weighted average probability of 3 marketing-beating years over both superstars and ordinarie} \\
&= P(3B|S) * P(S) + P(3B|O) * P(O) \\
&= [(\frac{7}{10})^3 * \frac{4}{25}] + [(\frac{1}{2})^3 * \frac{21}{25}] \\
&= (\frac{343}{1000} * \frac{4}{25}) + (\frac{1}{8} * \frac{21}{25}) \\
&= \frac{1372}{25000} + \frac{21}{200} \\
&= 16\%
\end{aligned}
$$

Therefore,

$$16\% * \frac{34.3\%}{16\%} = 0.343$$

Finally, $P(O|3B) = 1 – P(S|3B) = 1 – 0.343 = 0.66$

Q.367 A human health organization tracked a group of individuals for five years. At the commencement of the study, 25% were categorized as heavy smokers, 40% as light smokers, and the remaining as nonsmokers. Results revealed that light smokers were twice as likely as nonsmokers to die during the half-decade study, but only half as likely as heavy smokers. During the period, a randomly selected group member passed on.
Compute the probability that the individual who died was a heavy smoker.

A. 19%

B. 53%

C. 47%

D. 18%

The correct answer is **C.**

Let:
D = Event of death
L = Event of light smoker
H = Event of a heavy smoker
N = Event of a nonsmoker
We need to calculate P(H | D)
We also know that:

P(H)=0.25
P(L) = 0.40
P(N) = 0.35

Now, from the information provided, we know that:
P(D | L) = 2P(D | N) and P(D | L) = 1/2P(D | H)

Applying Bayes' theorem,

$$\begin{aligned} P(H|D) &= \frac{P(H)\bullet P(D|H)}{P(H)\bullet P(D|H)+P(L)\bullet P(D|L)+P(N)\bullet P(D|N)} \\ &= \frac{0.25\times 2P(D|L)}{0.25\times 2P(D|L)+0.40\times P(D|L)+0.35\times \frac{1}{2}P(D|L)} \\ &= \frac{P(D|L)[0.25\times 2]}{P(D|L)[0.25\times 2+0.4+0.5\times 0.35]} \\ &= \frac{0.5}{0.5+0.4+0.175} \\ &= 0.4651\approx 47\% \end{aligned}$$

Q.369 Peter selects a coin from a pair of coins and tosses it. While coin 1 is double-headed, coin 2 is a normal unbiased coin. After the toss, the result is a head. Calculate the probability that it was coin 1 which was tossed.

A. $\frac{1}{3}$

B. $\frac{2}{3}$

C. 0.5

D. 0.75

The correct answer is **B.**

We need to determine $P(\text{coin 1}|\text{head})$. By applying Bayes' theorem,

$$P(\text{coin 1}|\text{head}) = \frac{P(\text{coin 1}) \times P(\text{head}|\text{coin 1})}{(P(\text{coin 1}) \times P(\text{head}|\text{coin 1}) + P(\text{coin 2}) \times P(\text{head}|\text{coin 2}))} = \frac{(1 \times \frac{1}{2})}{1 \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2}} = \frac{2}{3}$$

Note:

- $P(\text{coin 1}) = P(\text{coin 2}) = \frac{1}{2}$ since we have two coins (pair) of coins that are equally likely to occur
- $P(\text{head}|\text{coin 1}) = 1$ since coin 1 is double headed
- $P(\text{head}|\text{coin 2}) = \frac{1}{2}$ since coin 2 is unbiased since both the head and the tail are equally likely to occur

Q.3251 The probability that the Eurozone economy will grow this year is 18%, and the probability that the European Central Bank (ECB) will loosen its monetary policy is 26%. Assuming that the joint probability that the Eurozone economy will grow and the ECB will loosen its monetary policy is 12%, then the probability that either the Eurozone economy will grow or the ECB will loosen its the monetary policy is *closest to*:

A. 14%

B. 32%

C. 44%

D. 6%

The correct answer is **B.**

The addition rule of probability is used to solve this question:

P(E) = 0.18 (the probability that the Eurozone economy will grow is 18%)

p(M) = 0.26 (the probability that the ECB will loosen the monetary policy is 26%)

p(EM) = 0.12 (the joint probability that Eurozone economy will grow and the ECB will loosen its monetary policy is 12%)

The probability that either the Eurozone economy will grow or the central bank will loosen its the monetary policy:

$$p(E \text{ or } M) = p(E) + p(M) - p(EM) = 0.18 + 0.26 - 0.12 = 0.32$$

Q.3252 A mathematician has given you the following conditional probabilities:

$p(O\|T) = 0.62$	Conditional probability of reaching the office if the train arrives on time
$p(O\|T^c) = 0.47$	Conditional probability of reaching the office if the train does not arrive on time
$p(T) = 0.65$	Unconditional probability of the train arriving on time
$p(O) = ?$	Unconditional probability of reaching the office

What is the unconditional probability of reaching the office, p(O)?

A. 0.5675

B. 0.4325

C. 0.3265

D. 0.3333

The correct answer is **A.**

This question can be solved using the total probability rule.

If $p(T) = 0.65$ (Unconditional probability of train arriving at time is 0.65), then the unconditional probability of the train not arriving on time $p(T^c) = 1 - p(T) = 1 - 0.65 = 0.35$.

Now, we can solve for $p(O) = p(O|T) * p(T) + p(O|T^c) * p(T^c) = 0.62 * 0.65 + 0.47 * 0.35 = 0.5675$

Q.3254 An investor owns shares of both Apple and Microsoft. The two companies operate independently. The investor assumes that the probability of Apple's share price declining by more than 5% this year is 0.4 while the probability of Microsoft's share price declining by more than 5% is 0.3. What is the probability that either Apple or Microsoft share prices will decline in price by more than 5% this year?

A. 0.58

B. 0.12

C. 0.7

D. 0.67

The correct answer is **A.**

We know that,

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A) * P(B) \\ &= 0.4 + 0.3 - 0.4 * 0.3 = 0.58 \end{aligned}$$

Q.3255 The probabilities that Bond A and Bond X will default in the next two years are 10% and 8%, respectively. The probability that both bonds will default simultaneously in the next two years is 5%. The probability that Bond A will default given that Bond X has already defaulted is *closest to*:

A. 62.50%.

B. 50%.

C. 80%.

D. 37.50%

The correct answer is **A.**

$P(X) = 8\%$

$P(A) = 10\%$

$P(X \cap A) = 5\%$

As per the conditional probability:

$$P(A|X) = \frac{P(X \cap A)}{P(X)} = \frac{5\%}{8\%} = 62.5\%$$

Q.3256 There is a 40% chance that ABX will announce negative quarterly results tomorrow. On any given day, there is a 55% chance that the company's stock price will decrease. If negative quarterly results are announced, the probability that the stock price will decline is 85%. Tomorrow, the probability that ABX will announce negative quarterly results or that the stock will decrease in price is *closest to*:

A. 0.72

B. 0.95

C. 0.85

D. 0.61

The correct answer is **D.**

Let:

$P(N)$ = Event that negative results are announced $= 0.4$

$P(N)$ = Event that stock price declines = 0.55

Therefore $P(D|N)$ = P(stock price decline|negative results) = 0.85

We need:

$$P(D \cup N) = P(D) + P(N) - P(D \cap N)$$

We do not know $P(D \cap N)$. However, based on the information given, we an calculate $P(D \cap N)$ using the conditional probability:

$$\begin{aligned} P(D|N) &= \frac{P(D \cap N)}{P(N)} \\ \Rightarrow P(D \cap N) &= P(D|N) \cdot P(N) \\ &= 0.85 \times 0.4 = 0.34 \end{aligned}$$

Thus,

$$\begin{aligned} P(D \cup N) &= P(D) + P(N) - P(D \cap N) \\ &= 0.4 + 0.55 - 0.34 \\ &= 0.61 \end{aligned}$$

Q.3257 The probability that a portfolio manager reads Business News weekly is 0.50, while the probability that a portfolio manager reads BloomField News is 0.40. If the probability that a portfolio manager reads both Business News and BloomField News is 0.30, then the probability that a portfolio manager does not read any of the two newspapers is *closest* to:

A. 0.30.

B. 0.40.

C. 0.50.

D. 0.6

The correct answer is **B.**

For simplicity, let

A = Business News; and

B = BloomField News.

p(A) = 0.50

p(B) = 0.40

$P(A \cap B) = 0.30$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P(A \cup B) = 0.50 + 0.40 - 0.30 = 0.60$$

The probability that someone reads A or B is 0.60. Therefore, the probability that a person does not read any of the two:

$$p(\text{person does not read any of the two}) = 1 - 0.60 = 0.40$$

Q.3590 An athlete takes part in two different events. The probability that she wins the first event is 0.3 and the probability that she wins the second event is 0.4. Given that the probability that she wins the first and the second event is 0.1, calculate the probability that she wins either the first or the second event.

A. 0.2

B. 0.5

C. 0.6

D. 0.1

The correct answer is **C.**

If A = P(wins first event) and B = P(wins second event), then we are told:

$P(A) = 0.3$; $P(B) = 0.4$; and $P(A \cap B) = 0.1$

We want,

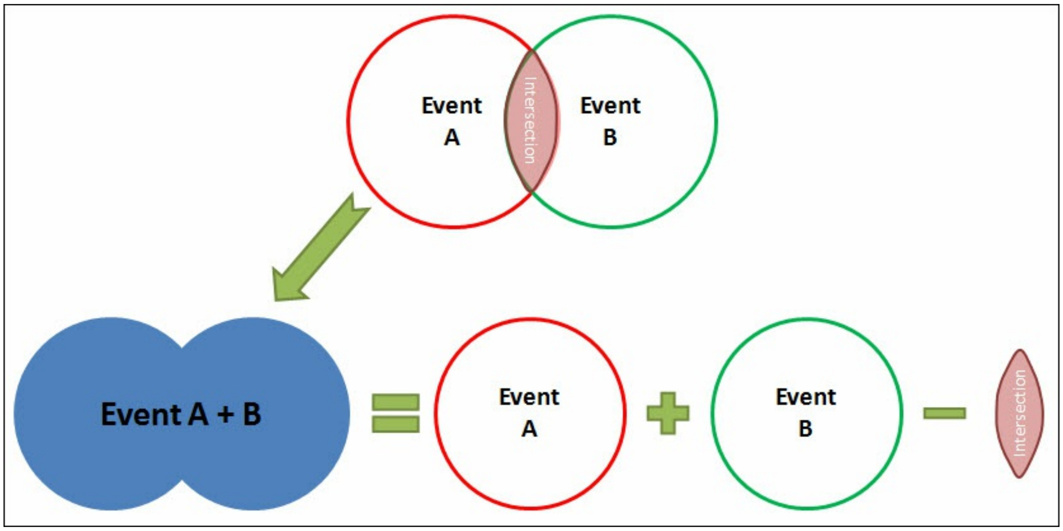
$$P(A \cup B)$$

Recall that:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Thus,

$$P(A \cup B) = 0.3 + 0.4 - 0.1 = 0.6$$



Q.3591 A homeowners insurer offers a discount for homeowners that either have a sprinkler system or live within 5 miles of a fire station. 60% of homeowners qualify for the discount. Only 15% of homeowners have a sprinkler system, and none of those homeowners live within 5 miles of a fire station. If the events are mutually exclusive, what is the probability a randomly selected homeowner lives within 5 miles of a fire station?

A. 15%

B. 25%

C. 30%

D. 45%

The correct answer is **D.**

We use the formula:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

However, since the events are mutually exclusive, then $P(A \cap B) = 0$,

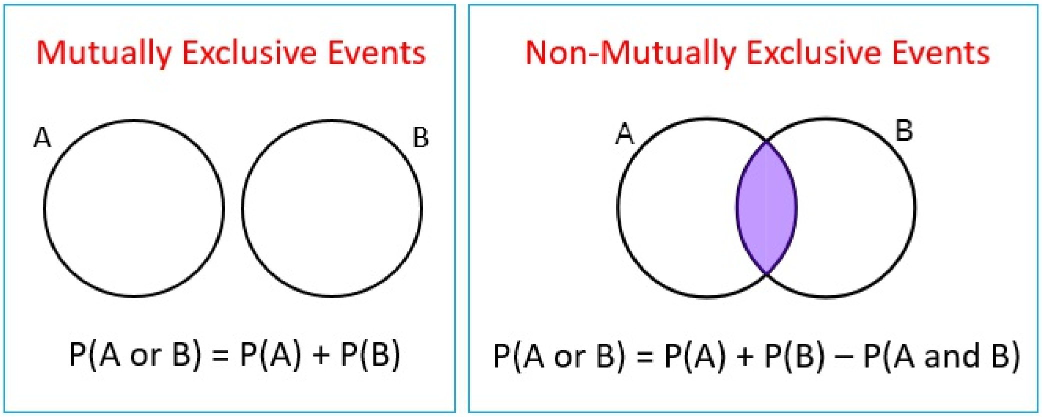
Thus,

$$.60 = .15 + x - 0$$

$$X = .45 = 45\%$$

Things to Remember

- Two events, A and B, are said to be mutually exclusive if the occurrence of A rules out the occurrence of B, and vice versa. For example, a car cannot turn left and turn right at the same time.
- For two mutually exclusive events, A and B, there's **no intersection**, and therefore P(A or B) = P(A) + P(B)
- For non-mutually exclusive events, however, the intersection **contains some outcomes** and has to be subtracted to get P(A or B)



Q.3592 A patient is considered high risk for a heart attack if they either have high cholesterol or high blood pressure. What percentage of patients are NOT considered high risk for a heart attack if 25% have high cholesterol, 30% have high blood pressure and 10% have both high cholesterol and high blood pressure?

A. 45%

B. 50%

C. 55%

D. 60%

The correct answer is **C.**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$.25 + .30 - .10 = .45$$

Therefore, 45% of patients are considered high risk. We now need to find the percentage of patients NOT considered high risk:

$$P(A \cup B)' = 1 - P(A \cup B) = 1 - .45 = .55 \quad \text{or} \quad 55\%$$

Q.3596 55% of an insurer's policyholders are male and 45% are female. The chances of a male having a claim stand at 10% while the chances of a female having a claim stand at 7%. What is the probability that NO ONE will have a claim?

A. 83%

B. 90%

C. 91%

D. 93%

The correct answer is **C.**

We know that:
P(Male) = 55%
P(Female) = 45%
P(Claim|Male)=P(Claim given it's a Male)=10%
P(Claim|Female)=P(Claim given it's a Female)=7%

This question makes use of the conditional probability rule:

$$P(\text{Claim}|\text{Male}) = \frac{P(\text{Claim} \cap \text{Male})}{P(\text{Male})}$$

$$\Rightarrow P(\text{Claim} \cap \text{Male}) = P(\text{Claim}|\text{Male}) \times P(\text{Male}) = .55 \times .10$$

We now have to also add together male and female drivers:

$$.55 \times .10 + .45 \times .07 = .0865$$

And since we're interested in no one having a claim:

$$1 - .0865 = .9135 \quad \text{or} \quad 91\%$$

Q.3599 Which two events are NOT considered independent?

A. Rolling a die; rolling another die

B. Flipping a coin; flipping a coin

C. Rolling a die; flipping a coin

D. Drawing a card; drawing another card from the same deck

The correct answer is **D.**

Drawing two cards from the same deck of cards are not independent events. Note that, two events are said to be independent of each other if the occurrence of one event does not affect the occurrence of the other event.

Option A: Rolling a die is not affected by the event of rolling another die.

Option B: Flipping a coin is also not affected by the event of flipping another coin.

Option C: The event of rolling a die does not dictate the event of flipping a coin.

Option D: The event of drawing a card affects the probability of drawing another card from the same deck, hence, the two events are not independent.

Q.3600 Events A and B are mutually exclusive events and P(A)= .3 while P(B) = .5. Calculate P(A u B).

A. 0.2

B. 0.35

C. 0.5

D. 0.80

The correct answer is **D.**

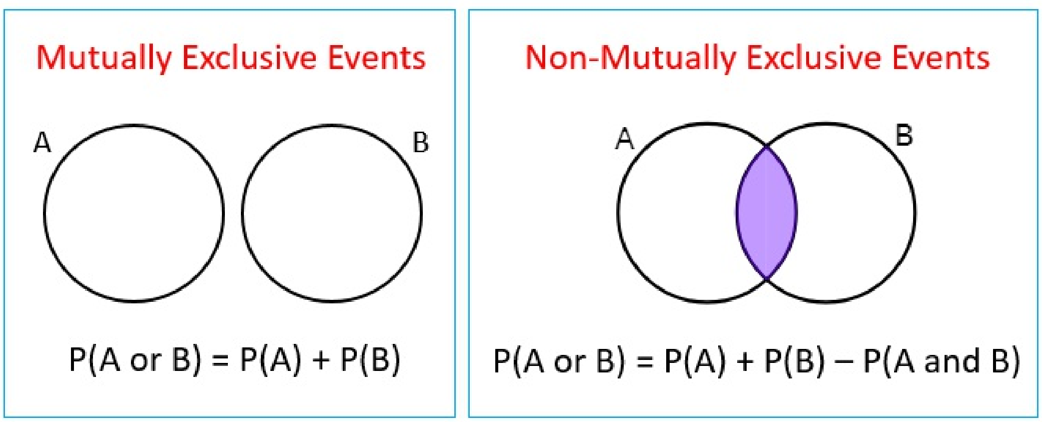
Since the events are mutually exclusive:

$$P\ (A \cup B) = P\ (A) + P\ (B)$$

$$= 0.3 + 0.5 = 0.80$$

Note: If the two events were not mutually exclusive, we would use the formula:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Q.3601 A company insures both male and female drivers. At the moment, the company has insured an equal number of male and female drivers. Males have a 0.15 chance of having a claim during a policy period while females have a 0.10 chance of having a claim. If a driver is randomly selected from the population, what is the probability that the driver has no claim during the policy period?

A. 73.5%

B. 76.5%

C. 77.5%

D. 87.5%

The correct answer is **D.**

We know that:
P(Male) = 50%
P(Female) =50%
P(Claim|Male)=P(Claim given it's a Male)=0.15
P(Claim|Female)=P(Claim given it's a Female)=0.10

This question makes use of the conditional probability rule:

$$P(\text{Claim}|\text{Male}) = \frac{P(\text{Claim} \cap \text{Male})}{P(\text{Male})}$$

$$\Rightarrow P(\text{Claim} \cap \text{Male}) = P(\text{Claim}|\text{Male}) \times P(\text{Male}) = 0.15 \times 0.50$$

$$\Rightarrow P(\text{Claim} \cap \text{Female}) = P(\text{Claim}|\text{Female}) \times P(\text{Female}) = 0.10 \times 0.50$$

We now have to also add together male and female drivers:

$$0.15 \times 0.50 + 0.10 \times 0.50 = 0.125$$

And since we're interested in the probability that the selected driver has no claim:

$$1 - 0.125 = 0.875 \text{ or } 87.5\%$$

Q.3602 For a certain insured, the probability of making no claim during a policy period is .60. The probability of making 1 claim is .25. What is the probability that this insured makes no more than 1 claim during the policy period?

A. 0.15

B. 0.25

C. 0.35

D. 0.85

The correct answer is **D.**

"no more than 1" means either 0 or 1 claim and because these events are mutually exclusive, we can calculate $P(A) + P(B) = .60 + .25 = .85$

Q.3603 For a certain insured, the probability of making no claim during a policy period is .60. The probability of making 1 claim is .25. What is the probability that this insured makes more than 1 claim during the policy period?

A. 0.15

B. 0.25

C. 0.35

D. 0.6

The correct answer is **A.**

"More than 1" is the complement of "no more than 1" which is 0 or 1

$$P(A \cup B) = P(A) + P(B) = .60 + .25 = .85$$

$$P(A \cup B)' = 1 - P(A \cup B) = 1 - .85 = .15$$

Q.3606 60% of an insurer's policyholders are male and 40% are female. The chance of a female having a claim is twice the chance of a male having a claim. Given a randomly selected policyholder has a claim, what's the probability that the policyholder is a male?

A. 35%

B. 65%

C. 57%

D. 43%

The correct answer is **D.**

Let the probability of a male having a claim be P(C|M) and that of a female having a claim be P(C|F).

The chance of a female having a claim is twice the chance of a male having a claim, that is, 2P(C|M) = P(C|F)

Thus,

$$P(M|C) = \frac{[P\ (C|M)^*\ P(M)]}{[P(M)^*P(C|M)+P(F)^*P(C|F)}$$
$$= \frac{[P\ (C|M) * 0.60)]}{[0.60 * P(C|M) + 0.40 * P(C|F)}$$

Replacing 2P(C|M) = P (C|F)

$$P(M|C) = \frac{[P\ (C|M) * 0.60)]}{[0.60 * P(C|M) + 0.40 * 2 * P(C|M)}$$
$$= \frac{0.60}{(0.60 + 0.80)}$$
$$= 0.4286 = 42.86\%$$

Q.3607 55% of an insurer's policyholders are male and 45% are female. The chance of a male having a claim is 10% and the chance of a female having a claim is 7%. Given a randomly selected policyholder has a claim, what's the probability she is a female?

A. 25%

B. 33%

C. 36%

D. 38%

The correct answer is **C.**

This question makes use of Bayes' theorem. We know that: $P(M) = 0.55$ and $P(F) = 0.45$
Also, $P(C|M) = 0.10$ and
$P(C|F) = 0.07$
Now, we need $P(F|C)$. Using Bayes' Theorem,

$$\begin{aligned} P(F|C) &= \frac{P(C|F).P(F)}{P(C|F).P(F) + P(C|M).P(M)} \\ &= \frac{0.07 \times 0.45}{0.07 \times 0.45 + 0.10 \times 0.55} \\ &= \frac{0.0315}{0.0865} \\ &= 0.3642 \approx 36\% \end{aligned}$$

Q.3609 A company insures red and black cars, male and female drivers and writes policies in 2 territories (A and B). There are 300 male drivers and 200 female drivers in total. There are 150 males who drive red cars and 100 females who drive red cars. 100 male and 100 female drivers live in territory A and 50 of each, males and females, drive red cars in territory A. Given that a randomly selected policyholder drives a black car, what is the probability that they are female and live in territory B?

A. 20%

B. 25%

C. 33%

D. 40%

The correct answer is **A.**

		Territory A	Territory B
Male	Red Car	50	100
Male	Black Car	50	100
Female	Red Car	50	50
Female	Black Car	50	50

$$P\ (\text{Female and Territory B} \mid \text{Black Car}) = \frac{P\ (\text{Female, Territory B and Black Car})}{P\ (\text{Black Car})}$$

$$= \frac{50}{250} = .20 \text{ or } 20\%$$

Q.3610 A patient is considered high risk for a heart attack if they either have high cholesterol or high blood pressure and the two events are independent. In a given population, 25% have high cholesterol and 30% have high blood pressure. If a randomly selected person has high blood pressure, what is the probability they also have high cholesterol?

A. 15%

B. 20%

C. 25%

D. 33%

The correct answer is **C.**

$$\begin{aligned} P\,(\text{High Cholesterol}|\text{High Blood Pressure}) &= \frac{P\,(\text{Both})}{P\,(\text{ High BP})} \\ &= \frac{0.25(0.3)}{0.30} = 25\% \end{aligned}$$

Q.3611 A patient is considered high risk for a heart attack if they either have high cholesterol or high blood pressure. In a given population, 45% of people are considered high risk for a heart attack, (25% have high cholesterol, 30% have high blood pressure). If a randomly selected person has high blood pressure, what is the probability they also have high cholesterol?

A. 15%

B. 20%

C. 25%

D. 33%

The correct answer is **D.**

$P\,(\text{High Cholesterol}|\text{High Blood Pressure}) = P\,(\text{Both})/P\,(\text{High BP})$
$P\,(A \cup B) = P\,(A) + P\,(B) - P\,(A \cap B)$

$.45 = .25 + .30 - x$

$P\,(A \cap B) = .10$

$P\,(\text{Both})/P\,(\text{High BP}) = .10/.30 = 1/3 \quad \text{or} \quad 33\%$

Q.3613 Given the following chart describing the claims of an auto insurer during a policy period, calculate P (C|M). (Assume that the number of male and female claims are independent of each other.)

	Male (M)	Female (F)	Total
Claim (C)	100	200	300
No Claim (X)	400	600	1000
Total	500	800	1300

A. 8%

B. 10%

C. 20%

D. 23%

The correct answer is **C.**

$$P(C|M) = \frac{P(C \cap M)}{P(M)}$$

$$= \frac{\frac{100}{1300}}{\frac{500}{1300}}$$

$$= 20\%$$

Q.3615 An insurance company classifies its policyholders into three tiers – standard, preferred, and ultra preferred. 40% of standard tier policyholders are male, 50% of preferred tier policyholders are male and 75% of ultra preferred tier policyholders are male. There is an equal number of policyholders in each tier. If a policyholder is selected at random, what is the chance she is female?

A. 25%

B. 30%

C. 33%

D. 45%

The correct answer is **D.**

Let event F be "female", event S be "standard", event P be "preferred" and event U be "ultrapreferred"

$$P(F) = P(F/S) * P(S) + P(F/P) * P(P) + P(F/U) * P(U)$$

$$P(F) = .60 * (1/3) + .50 * (1/3) + .25 * (1/3)$$

$$P(F) = .45 \quad \text{or} \quad 45\%$$

Q.3616 An insurance company classifies its policyholders into three tiers – standard, preferred, and ultra preferred. 40% of standard tier policyholders are male, 50% of preferred tier policyholders are male and 75% of ultra preferred tier policyholders are male. There is an equal number of policyholders in each tier. If a male policyholder is selected at random, what is the chance he is classified as a standard tier?

A. 15%

B. 24%

C. 30%

D. 33%

The correct answer is **B.**

Using Bayes Theorem,

$$P(S/M) = \frac{P(M/S) \times P(S)}{P(M)}$$

Where

$$\begin{aligned} P(M) &= P(M/S) \times P(S) + P(M/P) \times P(P) + P(M/U) \times P(U) \\ &= 0.40 \times \left(\frac{1}{3}\right) + 0.50 \times \left(\frac{1}{3}\right) + 0.75 \times \left(\frac{1}{3}\right) \\ &= 0.55 \end{aligned}$$

Thus,

$$P(S/M) = \frac{(0.40 \times (1/3))}{(0.55)} = 0.24 \quad \text{or} \quad 24\%$$

Q.3617 An insurance company classifies its policyholders into three tiers – standard, preferred and ultra preferred with a 25%/50%/25% distribution. The chance of a policyholder in the standard tier having a claim is 10%, in the preferred tier it is 5% and in the ultra preferred tier it is 2%. Given a policyholder has a claim, what is the probability they came from the ultra preferred tier?

A. 5%

B. 7%

C. 9%

D. 11%

The correct answer is **C.**

$$P(C) = P(C/S) * P(S) + P(C/P) * P(P) + P(C/U) * P(U)$$
$$P(C) = .10 * .25 + .05 * .50 + .02 * .25 = .055$$

$$P(U/C) = P(C/U) * P(U)/P(C) = .02 * .25/.055 = .09 \quad \text{or} \quad 9\%$$

Q.3618 There are three different bags. The first bag contains 3 square blocks and 2 round blocks. The second bag contains 2 square blocks and 3 round blocks. The third bag contains 5 round blocks. In an experiment, a bag is randomly chosen, and then a block is chosen from the bag. What is the probability that a round block is chosen?

A. 1/5

B. 1/3

C. 2/5

D. 2/3

The correct answer is **D.**

Let:

P (R)=Probability that the block is round

P (R/1)= Probability that the block is round given that it is from the first bag = $\frac{2}{5}$

P (R/2)= Probability that the block is round given that it is from the second bag = $\frac{3}{5}$

P (R/3)= Probability that the block is round given that it is from the third bag = $\frac{5}{5} = 1$

$P\ (R) = P\ (R/1) * P\ (1) + P\ (R/2) * P\ (2) + P\ (R/3) * P\ (3)$

$P\ (R) = (2/5) * (1/3) + (3/5) * (1/3) + 1 * (1/3) = 2/3$

Q.3619 There are three different bags. The first bag contains 3 square blocks and 2 round blocks. The second bag contains 2 square blocks and 3 round blocks. The third bag contains 5 round blocks. In an experiment, a bag is randomly chosen and then a block picked. Given a round block was selected, what is the probability it came from the second bag?

A. 1/5

B. 3/10

C. 1/3

D. 2/5

The correct answer is **B.**

Let "1" be the event that the first bag is picked,
"2" be the event that the second bag is picked,
"3" be the event that the third bag is picked,
and "R" be the event that a round block is drawn
$P\ (2/R) = P\ (R/2) * P\ (2)/P\ (R)$

$P\ (R) = P\ (R/1) * P\ (1) + P\ (R/2) * P\ (2) + P\ (R/3) * P\ (3)$

$P\ (R) = (2/5) * (1/3) + (3/5) * (1/3) + 1 * (1/3) = 2/3$

$P\ (2/R) = (3/5) * (1/3)/(2/3) = 3/10$

Q.3620 There are two bags with red and white balls. The first bag has 5 red and 5 white balls. The second bag has 3 red and 2 white balls. A bag is randomly selected and a ball is drawn. If the ball is red, another ball is selected from the same bag. If the ball is white, another ball is selected from the other bag. What is the probability that the second ball drawn is red?

A. 40%

B. 50%

C. 51%

D. 53%

The correct answer is **C.**

$P\,(2^{nd}\ \ \text{red}) = P\ (2^{nd}\ \ \text{red}/1^{st}\ \ \text{red}) * P\ (1^{st}\ \ \text{red}) + P\ (2^{nd}\ \ \text{red}/1^{st}\ \ \text{white}) * P\ (1^{st}\ \ \text{white})$
$P\ (1^{st}\ \ \text{red}) = \frac{1}{2} * 1/2 + 1/2 * 3/5 = .55$

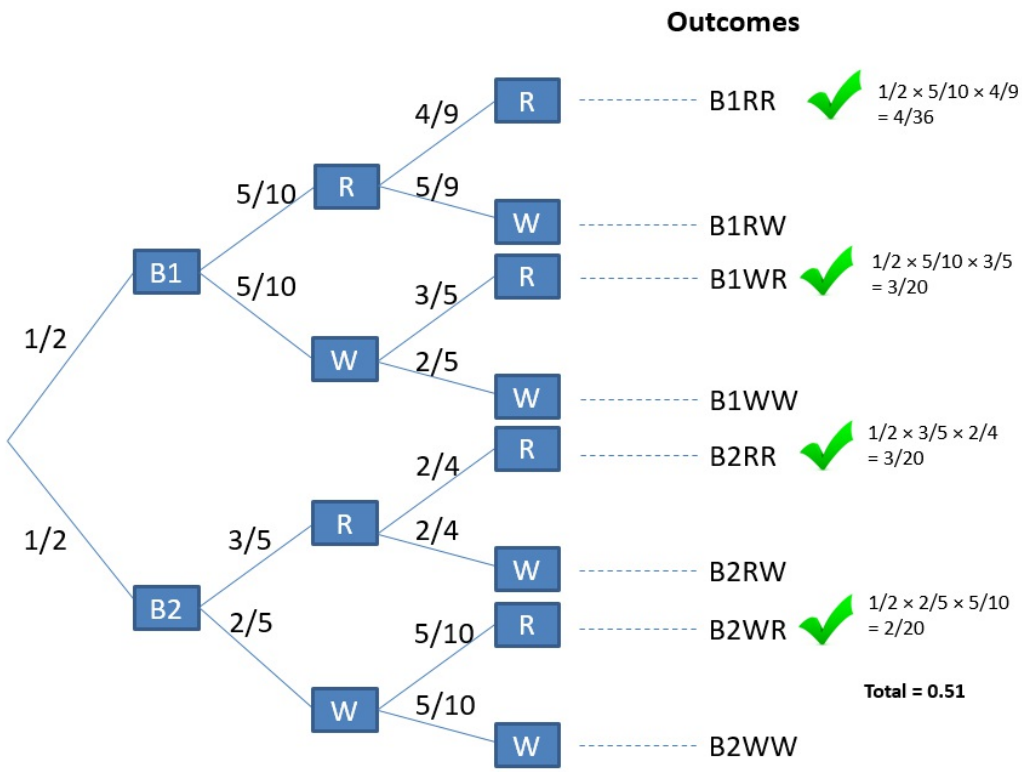
$P\ (2^{nd}\ \ red/1^{st}\ \ red) = (1/2) * (4/9) + (3/5) * (2/4) = .52$

$P\ (2^{nd}\ \ red/1^{st}\ \ white) = (1/2) * (3/5) + (2/5) * (1/2) = .5$

$P\ (1^{st}\ \ white) = \frac{1}{2} * 5/10 + 1/2 * 2/5 = .45$

$P\,(2^{nd}\ \ red) = .55 * .52 + .45 * .50 = .51\ \ \text{or}\ \ 51\%$

Alternative approach using a probability tree



Q.3621 In a game, a coin is flipped. If the coin is heads, the player rolls one die. If the coin turns up tails, the player rolls two dice and the player moves their playing piece the number of spots shown on the die or dice. Given that on a player's turn, he moves 5 spaces, what is the probability he flipped tails on the coin?

A. 1/10

B. 1/5

C. 2/5

D. 1/3

The correct answer is **C.**

$P(5) = P(5/T) * P(T) + P(5/H) * P(H)$
$P(5) = (1/9) * (1/2) + (1/6) * (1/2) = 5/36$

$P(T/5) = P(5/T) * P(T)/P(5)$

$P(T/5) = (1/9) * (1/2)/(5/36) = 2/5$

Further explanation:
P(5/H) is a conditional probability. This simply means that the player has already tossed the coin and it turned up a head (H). From the question, the player can only roll the dice once since his coin is showing a head (H). Remember that rolling dice has 6 outcomes, (1,2,3,4,5,6) and each outcome has a 1/6 probability of occurring. Therefore, the probability that the player rolls the dice and gets a 5 is 1/6. The question states that "If the coin turns up tails, the player rolls two dice ...". Now, if the coin turns out to be a tail, a player rolls two dice. As such, the player will have a sample size of 36. Now, the question conditions us that the player moves 5 times. Now, if the coin turns to be a tail, then we will have (1,4), (2,3), (3,2), and (4,1) out of a sample space of 36. And thus, $P(5|T)=4/36=1/9$

Q.3624 An insurance company writes business in three territories: A, B and C. They have 150 policyholders in territory A, 250 in territory B and 300 in territory C. A person is twice as likely to have a claim in territory B than territory A and 3 times as likely to have a claim in territory C than territory A. On average, 50 people have a claim every policy period. Given a claim occurs, what is the probability it was a policyholder in territory C?

A. 21%

B. 33%

C. 43%

D. 58%

The correct answer is **D.**

$$P(C|\text{Claim}) = \frac{P(\text{Claim}|C) * P(C)}{P(\text{Claim})}$$

$$\begin{aligned} P(\text{Claim}) &= P(\text{Claim}|A) * P(A) + P(\text{Claim}|B) * P(B) + P(\text{Claim}|C) * P(C) \\ &= x * (\frac{150}{700}) + 2x * (\frac{250}{700}) + 3x * (\frac{300}{700}) \\ &= x * (\frac{1550}{700}) \end{aligned}$$

But we also know that the probability of a claim from all territories combined is 50/700. Thus, $\frac{1550x}{700} = \frac{50}{700} \Rightarrow x = 0.032$

But, $P(C) = \frac{300}{700}$

Therefore,

$$\begin{aligned} P(C|\text{Claim}) &= \frac{3 * 0.32 * \frac{300}{700}}{\frac{50}{700}} \\ &= 0.576 \approx 58\% \end{aligned}$$

Q.3625 A test for heart disease results in a correct positive diagnosis 95% of the time and a correct negative diagnosis 99% of the time. 25% of the population has heart disease. What is the probability of a positive test?

A. 0.275

B. 0.150

C. 0.245

D. 0.950

The correct answer is **C.**

Let's define the following events:
+ = positive diagnosis
- = negative diagnosis
H = heart disease
H' = no heart disease
$P\ (+|H) = .95$
$P\ (-|H) = .99$
$P\ (+) = P\ (+|H) * P\ (H) + P\ (-|H) * P\ (H^{'})$
$= .95 * .25 + .01 * .75 = .245 \quad \text{or} \quad 2.45\%$

Q.3626 A test for heart disease results in a false positive 5% of the time. 25% of the population has heart disease and 20% test positive. What is the probability of a negative test given that the patient has no heart disease. ?

A. 0.67

B. 0.07

C. 0.0533

D. 0.9467

The correct answer is **C.**

Let event H be heart disease and event + be a positive test.

We know that, $P(+) = 0.20; \Rightarrow P(-) = 0.80$ and $P(+|H^{'}) = 0.05; \Rightarrow P(-|H^{'}) = 0.95.$

Now, using the conditional probability formula,

$$\begin{aligned} P\left(-|H^{'}\right) &= \frac{P\left(H^{'}|-\right) * P\left(-\right)}{P\left(H^{'}\right)} \\ &= \frac{0.05 \times 0.80}{0.75} = 0.0533 \end{aligned}$$

Q.3643 A company insures red and black cars, male and female drivers and writes policies in 2 territories A and B. There are 300 male drivers and 200 female drivers in total. There are 150 males who drive red cars and 100 females who drive red cars. 100 male and 100 female drivers live in territory A and 50 of each, males and females, drive red cars in territory A. What is the probability that a randomly selected driver is either female or lives in territory B?

A. $\frac{2}{5}$

B. $\frac{7}{10}$

C. $\frac{4}{5}$

D. $\frac{9}{10}$

The correct answer is **C.**

Using:

$$P\ (A \cup B) = P\ (A) + P\ (B) - P\ (A \cap B)$$

We need,

$$P\ (\text{female or living in territory B}) = P\ (\text{Female}) + P\ (\text{Living in territory B}) - P\ (\text{female and living in t}$$

$$= \frac{200}{500} + \frac{300}{500} - \frac{100}{500} = \frac{400}{500} = \frac{4}{5}$$

Reading 13: Random Variables

Q.311 Which of the following can be categorized as continuous random variables?
I. Stock returns on a given day
II. The weight of 20 FRM candidates (in pounds)
III. The total amount of Biannual share dividends received over a 10-year period
IV. The number of holidays in a given year
V. The annual number of FRM exam candidates in the last 10 years

A. I, III, and V

B. I, II, and III

C. I, II, III, and V

D. All the above

The correct answer is **B.**

A continuous random variable is a variable that has infinite possible outcomes, even though lower and upper bounds exist. It differs from a discrete random variable that takes on only a countable number of values.

Stock returns can take on any number and thus stock returns are continuous random variables.

Similarly, weight is a continuous random variable since it can take any value within a given interval, e.g., 180.2 pounds.

The total amount of Biannual share dividends received over a 10-year period can take on any value and therefore it is a continuous random variable

On the other hand, the number of FRM candidates in any given year can only take on an integer/whole number value.

Similarly, there can only be a discrete number of holidays in a year or a discrete number of candidates sitting an exam.

Note that, while stock returns can take on any number, stock prices are discrete as they take on certain values e.g $12.01, $12.02, and not $12.005.

Q.312 Consider the following probability function for a discrete random variable X:

$$P(x) = \frac{x}{100}, X = 10, 20, 30, Y,; \text{ otherwise } P(x) = 0$$

Find the value of Y.

A. 4

B. 40

C. 50

D. 5

The correct answer is **B.**

There are two key properties that any probability function must meet. These are:

I. $0 \leq P(x) \leq 1$

II. $\sum P(x) = 1$

Therefore,

$$\begin{aligned} \frac{10}{100} + \frac{20}{100} + \frac{30}{100} + \frac{Y}{100} &= 1 \\ \Rightarrow \frac{60}{100} + \frac{Y}{100} &= 1 \\ \Rightarrow 60 + Y &= 100 \\ \Rightarrow Y &= 40 \end{aligned}$$

Q.314 The following is the probability mass function for a discrete random variable X,

$$P(x) = \frac{x}{100}; X = 10, 20, 30, 40$$

Determine the CDF at X =30, i.e., F(30)

A. 0.2

B. 0.3

C. 0.7

D. 0.6

The correct answer is **D.**

The cumulative distribution function (cdf) F(x) defines the probability of a random variable X, and assumes a value equal to or less than a specified value, x. As such:

$$\begin{aligned} F(30) &= P(X \leq 30) \\ &= P(X = 10) + P(X = 20) + P(X = 30) \\ &= \frac{10}{100} + \frac{20}{100} + \frac{30}{100} \\ &= 0.6 \text{ or } 60\% \end{aligned}$$

Q.327 The following table presents the probability distribution of the earnings per share (EPS) for a certain company:

Probability	EPS	Interest Rates	Beta
20%	\$1.5	21%	1.15
10%	\$2.0	20%	10%
30%	\$1.3	20%	1.25
40%	\$1.2	10%	

Compute the expected earnings per share.

A. 2

B. 1.5

C. 1.37

D. 1.2

The correct answer is **C.**

$$\begin{aligned} EPS = E(X) &= \sum P_{(xi)} x_i \\ &= 0.2 \times 1.5 + 0.1 \times 2.0 + 0.3 \times 1.3 + 0.4 \times 1.2 \\ &= 1.37 \end{aligned}$$

Q.330 A discrete random variable Y has probability function given by:

Y	0	1	2
P(Y = y)	0.3	0.6	0.1

Calculate Var(Y).

A. 0.2

B. 0.36

C. 0.8

D. 1

The correct answer is **B.**

The variance of any given random variable is given by:

$$Var(Y) = E(Y^2) - E^2(Y)$$

$$E(Y) = \sum Y P(Y = y) = 0 * 0.3 + 1 * 0.6 + 2 * 0.1 = 0.8$$

$$E(Y^2) = \sum Y^2 P(Y = y) = 0^2 * 0.3 + 1^2 * 0.6 + 2^2 * 0.1 = 1$$

Therefore,

$$Var(Y) = 1 - 0.8^2 = 0.36$$

Q.334 Which of the following best describes the concept of skewness in statistics?

A. The degree to which a distribution is symmetric about its mean.

B. The degree to which a distribution is nonsymmetric about its median.

C. The degree to which a distribution is nonsymmetric about its mean.

D. The degree to which a random variable spreads around its mean.

The correct answer is **C.**

Skewness in statistics describes the asymmetry from the normal distribution in a set of data. Such a dataset differs from a normal curve which is bell-shaped and perfectly symmetrical. In layman's language, a symmetrical curve can be divided into two equal halves with the mean in the middle. When this is not possible, the curve (and the underlying data) is said to be skewed. A distribution can either be positively or negatively skewed, depending on where there's a higher concentration of data points.

Q.335 Which of the following is *incorrect* about kurtosis?

A. Excess kurtosis is a measure relative to the uniform distribution, which has a kurtosis of 3.

B. Excess kurtosis that's negative indicates a platykurtic distribution.

C. Excess kurtosis that's positive indicates a leptokurtic distribution.

D. The normal distribution has a kurtosis equal to 3.

The correct answer is **A.**

Kurtosis basically measures the peakedness of a distribution. Data sets with high kurtosis tend to have many data points at the tails (outliers). Kurtosis is measured relative to the **normal** distribution, which has a kurtosis of exactly 3. Therefore, option A is an incorrect statement.

Q.336 Mary Noel, FRM, is tasked with analyzing the returns of two different assets – A and B. She finds that the two assets have the same mean, variance, and skewness, but A has a higher kurtosis than B. Which of the following statements is most likely true?

A. Asset A is riskier than asset B.

B. Asset B is riskier than asset A.

C. Both assets are highly profitable.

D. Assets A and B will earn negative returns in the long term.

The correct answer is **A.**

In finance, Kurtosis affects the riskiness of an asset. The asset with a higher kurtosis is considered riskier than another one with a lower kurtosis. The underlying logic is that a high kurtosis indicates a high number of outliers, meaning that the return for such an asset is highly variable, and therefore highly risky.

Q.337 The following are measures of variability, EXCEPT:

A. Variance

B. Standard deviation

C. Range

D. Median

The correct answer is **D.**

The mean, median, and mode are all measures of central tendency – all of them attempt to describe data by identifying a central position. Measures of variation, such as standard deviation and variance, describe the spread of the data around the mean.

Q.3260 Which of the following statements is *most* accurate?
Skewness refers to the extent a distribution is:

A. Symmetrical. In negatively-skewed distributions, the mean is to the left of the peak.

B. Asymmetrical. In negatively-skewed distributions, the mean is to the right of the peak.

C. Asymmetrical. In positively-skewed distributions, the mean is to the right of the peak.

D. Asymmetrical. In left-skewed distribution, the mean coincides with the peak.

The correct answer is **C.**

Skewness refers to asymmetry in a statistical distribution. It can be quantified to define the extent to which a distribution differs from a normal distribution.
A left-skewed distribution has a long left tail. Left-skewed distributions are also called negatively-skewed distributions. The reason is that there is a long tail in the negative direction on the number line. The mean is to the left of the peak.

A right-skewed distribution has a long right tail. Right-skewed distributions are also called positive-skew distributions. That's because there is a long tail in the positive direction on the number line. The mean is to the right of the peak.

Q.3262 Assume you're a financial risk manager at an investment management firm where you're given the task to estimate the dispersion of a specific equity price around its forecasted value. As a financial risk manager, calculate the variance of equity value using the data provided in the following table.

Probability	Equity Value
0.33	$62.15
0.39	$60.75
0.28	$63

A. 0.87

B. 0.93

C. 0.75

D. 0.78

The correct answer is **A.**

$$
\begin{aligned}
Var(X) &= E(X^2) - [E(x)]^2 \\
E(x) &= 0.33 \times 62.15 + 0.39 \times 60.75 + 0.28 \times 63.0 = 61.842 \\
[E(x)]^2 &= 0.33 \times 62.15^2 + 0.39 \times 60.75^2 + 0.28 \times 63.0^2 = 3,825.3048 \\
Var(X) &= 3,825.3048 - 61.842^2 = 0.8718
\end{aligned}
$$

Q.3268 An equity research analyst forecasts the share price of Equidor Inc.'s stock and the probability of achieving the price target. The forecast made by the analyst is given in the following exhibit.

Exhibit 1: Share Price Forecast

Probability	Share Price
20%	\$32.00
25%	\$28.00
40%	\$34.00
15%	\$38.00

What is the variance of Equidor's stock price?

A. 12.450

B. 10.51

C. 16.324

D. 12.213

The correct answer is **B.**

$$
\begin{aligned}
Var(X) &= E(X^2) - [E(x)]^2 \\
E(X) &= 0.20(32) + 0.25(28) + 0.40(34) + 0.15(38) = 32.7 \\
[E(x)]^2 &= 0.20 \times 32^2 + 0.25 \times 28^2 + 0.40 \times 34^2 + 0.15 \times 38^2 = 1,079.8 \\
Var(X) &= 1,079.8 - 32.7^2 = 10.51
\end{aligned}
$$

Q.3269 Which of the following is an example of a continuous random variable?

A. The number of defective TV sets in a container.

B. The number of visits recorded at a risk management consultancy office on a given day.

C. The amount of time required to run a mile.

D. None of the above.

The correct answer is **C.**

The amount of time required to run a mile can take an infinite number of possible outcomes.
Note: Options A) and B) are examples of discrete random variables as the number of defective TV sets and the number of visits recorded at a risk management consultancy office on a given day have a finite number of possible outcomes.

Q.3739 Let X have the following probability density function:

$$f_X(x) = \begin{cases} 0.15 & x = 1 \\ 0.25 & x = 2 \\ 0.35 & x = 3 \\ C & x = 4 \end{cases}$$

Calculate the mode of the distribution.

A. 1.5

B. 2.0

C. 2.5

D. 3.0

The correct answer is **D.**

First, find C, to see the weight attached to this variable:

$$C = 1 - 0.15 - 0.25 - 0.35 = .25$$

The mode is the most likely value of the distribution (with the highest probability), so in this case it is x=3.

Q.3836 The average salary for an employee at Capital Asset Managers is $50,000 per year. This year, the management has decided to award bonuses to every employee:

- A Christmas bonus of $1,000
- An incentive bonus equal to 15% of the employee's salary

Determine the mean bonus received by employees.

A. $8,500

B. $4,250

C. $500

D. $10,500

The correct answer is **A.**

To compute the bonus, we apply the following linear transformation to each employee's salary:

$$Y = \alpha + \beta x$$
$$Y = 1,000 + 0.15x$$

where Y is the transformed variable (bonus), X is the original variable (salary), β is the scale constant, and α is the shift constant.

The mean bonus, i.e., mean of Y, is given by:

$$E(Y) = E(\alpha + \beta x) = \alpha + \beta E(X)$$

We know that the mean salary is $50,000. Thus,

$$E(Y) = \$1,000 + 0.15(\$50,000) = \$8,500$$

Q.3837 At Capital Bank, the average salary among sales employees is $30,000 per year, and they are also entitled to a bonus of $0.05 for every dollar of sales brought in. Average sales amount to $300,000 per year. Determine the mean compensation received by employees.

A. $165,000

B. $45,000

C. $22,500

D. $330,000

The correct answer is **B.**

To compute the mean compensation, we apply the following linear transformation to sales:

$$Y = \alpha + \beta x$$
$$Y = 30,000 + 0.05x$$

where Y is the transformed variable (mean compensation), X is the sales variable, β is the scale constant ($0.05/$1)s, and α is the shift constant.

The mean compensation, i.e., mean of Y, is given by:

$$E(Y) = E(\alpha + \beta x) = \alpha + \beta E(X)$$

We know that average sales, E(X) = $300,000. Thus,

$$E(Y) = \$30,000 + 0.05(\$300,000) = \$45,000$$

Q.3838 The average salary for an employee at Capital Asset Managers is $50,000 per year, with a variance of 6,000,000. This year, the management has decided to award bonuses to every employee:

- A Christmas bonus of $1,000
- An incentive bonus equal to 15% of the employee’s salary

Calculate the standard deviation of employee bonuses.

A. $8,500

B. $250

C. $367

D. $10,500

The correct answer is **C.**

To compute the bonus, we apply the following linear transformation to each employee's salary:

$$Y = \alpha + \beta x$$
$$Y = 1,000 + 0.15x$$

where Y is the transformed variable (bonus), X is the original variable (salary), β is the scale constant, and α is the shift constant.

The variance of Y, i.e., variance of bonuses, is given by:

$$\text{Var}(Y) = \text{Var}(\alpha + \beta x) = \beta^2 \text{Var}(X) = \beta^2 \sigma^2$$

The variance of salary [Var(X) = σ], is 6,000,000. Thus,

$$\text{Var}(Y) = 0.15^2 \times 6,000,000 = 135,000$$

Standard deviation is equal to the square root of the variance, and therefore the standard deviation of employee bonuses is equal to the square root of 135,000 or $367.42

Q.3839 At Capital Bank, the compensation framework is made up of a basic salary plus bonuses. The average salary among sales employees is $30,000 per year, and they are also entitled to a bonus of $0.05 for every dollar of sales brought in. Average sales amount to $300,000 per year with a variance of 5,000,000. Determine the standard deviation of compensation received by employees.

A. $165

B. $450

C. $222

D. $112

The correct answer is **D.**

To compute the mean compensation, we apply the following linear transformation to sales:

$$Y = \alpha + \beta x$$
$$Y = 30,000 + 0.05x$$

where Y is the transformed variable (mean compensation), X is the sales variable, β is the scale constant ($0.05/$1), and α is the shift constant

The variance of Y, i.e., variance of compensation, is given by:

$$\text{Var}(Y) = \text{Var}(\alpha + \beta x) = \beta^2 \text{Var}(X) = \beta^2 \sigma^2$$

The variance of sales [Var(X) = σ], is 5,000,000. Thus,

$$\text{Var}(Y) = 0.05^2 \times 5,000,000 = 12,500$$

Standard deviation is equal to the square root of variance, and therefore the standard deviation of employee bonuses is equal to the square root of 12,500 or $111.8

Q.340 During a disease outbreak, the probability of surviving after an infection is 60%. Determine the probability that at least 8 out of a group of 9 infected persons will survive.

A. 0.7

B. 0.07

C. 0.007

D. 0.06

The correct answer is **B.**

We note the following:

I. That the infection of any single individual is independent of all other infections for other individuals.
II. The trials are identical (Probability of survival is 60% every time)

Therefore, the number of survivors takes on a binomial distribution with n = 9 and p = 0.6

If X stands for the number of survivors, then:

$$\begin{aligned} P(X \geq 8) &= P(X = 8 \text{ or } 9) = \frac{9!}{8!1!}(0.6)^8 0.4^1 + \frac{9!}{9!0!}(0.6)^9 0.4^0 \\ &= 0.071 \end{aligned}$$

Note: Under the binomial distribution,

$$P(X = x) = {}^nC_x p^x q^{n-x}$$

Q.341 June Barrow, FRM, runs a consultancy firm that offers investment advice to clients in Canada. The number of clients the firm receives in a month follows a Poisson distribution with a mean of 4. What is the probability that the firm receives exactly 44 new clients in a year, assuming every client is independent?

A. 0.025

B. 0.0506

C. 0.24

D. 0.00363

The correct answer is **B.**

We note the following:

I. New clients are received randomly at a rate of 4 per unit time (month)
II. Each event is independent

These observations confirm the Poisson distribution.

Since the question asks us to find a yearly probability, we must convert the monthly Poisson rate to an equivalent yearly rate. Poi(4) monthly distribution is equivalent to Poi(48).

Now, if X represents the number of clients received,

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

$$P\ (X = 44) = \frac{[e^{-48}48^{44}]}{44!} = 0.0506$$

Q.342 Which of the following is NOT true regarding the normal distribution?

A. It's completely described by its mean, μ, and variance, σ^2.

B. Its skewness = 3 and kurtosis = 0.

C. A linear combination of two normally distributed variables also has a normal distribution.

D. The probabilities of extreme events (those further above and below the mean) continually get smaller but extend infinitely without going to zero.

The correct answer is **B.**

Statement B is false but its converse is true: The normal distribution has skewness = 0 and kurtosis = 3. In fact, the kurtosis of other distributions is measured relative to 3, which is the kurtosis of the normal distribution.

Q.343 Consider the following events:
I. Throwing a fair, six-sided die
II. The rate at which customers walk into a banking hall per day
III. The score of 50 FRM exam candidates in a mock test
IV. Tossing a coin
V. Picking an orange from a basket containing 10 equally sized oranges

Which of the events above exhibit uniform distributions?

A. I, IV and V

B. I and II only

C. II and III only

D. None of the above

The correct answer is **A.**

Under the uniform distribution, ALL outcomes are equally likely i.e., they have equal probabilities of occurrence. For example, if we were to throw a fair die, each of the numbers (1 to 6) would have a probability of 1/6. Similarly, a head (or a tail) occurs with probabilities of 0.5 when a coin is tossed.

Q.344 The probability that a patient suffering from typhoid will be treated successfully is 0.8. Forty patients are subjected to treatment. Determine the expected value of the number of patients who are treated successfully.

A. 7

B. 28

C. 8

D. 32

The correct answer is **D.**

This question tests the knowledge of the mean of the binomial distribution (n, θ)
The expected number of cured patients $= E(X) = n\theta = 40 * 0.8 = 32$

Note that $V(X) = n\theta(1-\theta)$

Q.345 The rate of registration for the FRM exam by candidates takes on a Poisson distribution with mean λ. Which of the following statements is correct?

A. Mean equals the standard deviation.

B. Mean equals the variance.

C. Median equals the variance.

D. Median, mean and variance are all equal.

The correct answer is **B.**

An interesting fact about the Poisson distribution is that the mean equals the variance.

Q.347 In the standard normal distribution, what do z-scores represent?

A. Scores below the mean in units of the standard deviation of the distribution.

B. Scores below and above the variance in units of the standard deviation of the distribution.

C. Scores above the mean in units of the standard deviation of the distribution from the mean.

D. Scores below and above the mean in units of the standard deviation of the distribution.

The correct answer is **D.**

The z-score is calculated as $\frac{(X-\mu)}{\sigma}$

It, therefore, gives the location of raw scores above and below the mean in units of the standard deviation of the distribution

In other words, the z-score gives a measure of how far from the mean, a data point is.

Q.351 The normal distribution and the lognormal distribution are related in such a way that:

A. If a random variable X follows a lognormal distribution, ln X is normally distributed.

B. If a random variable X follows a normal distribution, ln X is said to have a lognormal distribution.

C. The mean and variance of a lognormal distribution are twice that of the normal distribution, provided the value of n is the same.

D. The mean and variance of the normal distribution are twice that of the lognormal distribution, provided the value of n is the same.

The correct answer is **A.**

A random variable X follows a lognormal distribution if its natural logarithm, ln X, is normally distributed. In layman's language (for easy understanding), you can view the term "lognormal" as "the log is normal".

Q.352 A motor vehicle production company based in California is assembling its first batch of fully electric cars. After inspecting about 100 newly assembled units, engineers establish that there are a total of 40 defects. While some units have no defects, others have one, two, or more defects.

Assume that the distribution of mechanical defects follows a Poisson distribution. Drawing on the first 100 units produced, how many cars, out of every 10,000 units assembled, would we expect to have at least one defect?

A. About 330

B. About 0.330

C. About 3,300

D. About 1250

The correct answer is **C.**

Let's use X to denote the number of defects in a car.

$X \sim Poi(40/100)$ i.e. $\lambda = 0.4$

$$P(X \text{ is at least } 1) = 1 - P(X = 0) = 1 - \exp(-0.4) = 0.330$$

This is the probability of at least 1 defect in a car. Therefore, for every 10,000 cars, we would expect 0.330 * 10,000 = 3,300 units to have one or more defects.

Further explanation

From the question, the most important information is that we have 40 defect vehicles in every 100 vehicles manufactured by the company. This implies that, probability of defect is equal to 40/100.

Now, to solve the foe the probability that there is at least one defect vehicle, putting in mind that this follows a Poisson distribution, with $\lambda = \frac{40}{100} = 0.4$,

By having at least one defect, means that we expect either 1 defect or more. So that, we don't expect that there will be zero defects, therefore,

$$P(\text{at least one defect}) = 1 - P(\text{No defect}) = 1 - P(x = 0)$$

Now, recall that the pmf of a Poisson distribution is,

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

So that,

$$P(\text{at least one defect}) = 1 - P(\text{No defect}) = 1 - P(x = 0) = 1 - \frac{(e^{-0.4})(0.4^0)}{0!} = 1 - \frac{(e^{-0.4})(1)}{1} = 0.330$$

Now, to get the number of defects out of 10,000 manufactured vehicles, we will multiply that number by the defect probability, i.e., $0.330 * 10,000 = 3,330$

Q.353 The F-distribution and the Chi-square distribution have glaring similarities. Which of the following is not accurate?

A. Both are asymmetrical.

B. Both have a bound equal to zero on the left.

C. Their means are always less than their standard deviations.

D. They are defined by the number of degrees of freedom.

The correct answer is **C.**

There exists no consistent relationship between mean and standard deviation in either the F- or the chi-square distribution.

Q.354 Insurance claims in a certain class of business are modeled using a normal distribution with mean $3,000 and a standard deviation of $400. Calculate the probability that the next claim received will exceed $3,500.

A. 0.8944

B. 0.25

C. 0.75

D. 0.1056

The correct answer is **D.**

$$
\begin{aligned}
X &\sim N(3000, 400^2) \\
P(X > 3500) &= P\left[Z > \frac{(3500 - 3000)}{400}\right] \\
&= P(Z > 1.25) \\
&= 1 - P(Z < 1.25) \\
&= 1 - 0.8944 \\
&= 0.1056
\end{aligned}
$$

Note that the probability, $P(Z < 1.25)$, can be read directly from the standard normal table.

Q.3264 As a portfolio analyst, you're directed to label a fund consisting of 9 stocks out of which 4 stocks should be small-cap stocks, 3 stocks should be blue-chips and 2 stocks should be from emerging markets. Determine how many ways these 9 stocks can be labeled.

A. 1260

B. 362880

C. 60480

D. 112840

The correct answer is **A.**

There is a total of $9! = 362,880$ ways in which these 9 stocks can be sequenced.
However, the number of ways these 9 stocks can be labeled according to the required three categories = $\frac{9!}{(4! * 3! * 2!)} = 1,260$.

Q.3265 A teacher wants to select groups of 3 students out of 15 for group work. How many different groups of 3 are possible?

A. 25

B. 45

C. 225

D. 455

The correct answer is **D.**

Counting problems involve determining the exact number of ways two or more operations or events can be performed together. This particular counting problem can be solved using a combination, which is basically a selection of some given items where the order does not matter. The formula is:

$$_nC_r = \frac{n!}{((n-r)!r!}$$

Where

n=15; and

r=3

$$_nC_r = \frac{15!}{((15-3)! * 3!} = 455$$

Q.3270 A trader purchases one single stock every day during five working days. His risk manager believes that the probability of selecting an underpriced stock at any given time is 52%. Assuming a binomial distribution, what is the probability of selecting exactly two underpriced stocks during the week out of the universe of underpriced and overpriced stocks?

A. 0.395

B. 0.208

C. 0.327

D. 0.299

The correct answer is **D.**

Since it's a binomial distribution, we will solve the question with the help of the Bernoulli trial method.

The probability of having exactly 2 underpriced stocks in 5 trials (5 days), given that the probability of selecting an underpriced stock at any time is 52%, can be expressed as:

$$
\begin{aligned}
&\left(\frac{n!}{x!(n-x)!}\right) * p^x * (1-p)^{n-x} \\
&= \frac{5!}{(2! * 3!)} * 0.52^2 * (1-0.52)^3 \\
&= \left(\frac{120}{12}\right) * 0.2704 * (0.110592) = 0.2990
\end{aligned}
$$

Q.3271 As an investment analyst, your job is to determine how many companies will announce IPOs out of 50 virtual reality startup companies operating in Palo Alto. The annual IPO rate in high-tech industries in all other states of the U.S. is 7.85%. Using a binomial model, what is the standard deviation of the number of virtual reality company IPOs in Palo Alto?

A. 3.616

B. 1.902

C. 1.38

D. 2.125

The correct answer is **B.**

Variance according to the Binomial model $= n * p * (1 - p) = 50 * 0.0785 * (1 - 0.0785) = 3.616$

$$\text{Standard deviation} = \sqrt{3.616} = 1.9018$$

Q.3272 As a research analyst, you're analyzing the probability that the prices of copper will be set below \$44/kg after the upcoming government elections. Suppose that the prices of copper are uniformly distributed with a floor at \$38/kg and a ceiling at \$54/kg imposed by the government, then what is the probability that the prices of copper will be set below \$44/kg?

A. 0.815

B. 0.625

C. 0.375

D. 0.429

The correct answer is **C.**

Since the government has set a floor of \$38/kg (*a* or the lower boundary) and the ceiling of \$54/kg (*b* or the upper boundary.

Thus, n = 54 − 38 = 16

The possible outcomes (prices) of copper that fall below \$44 is \$44 − \$38 = \$6.

Therefore, the probability that the prices of copper will be set under \$44 is:

$$\frac{(X - a)}{(b - a)} = \frac{(44 - 38)}{(54 - 38)} = 0.375$$

Q.3273 Which of the following are the most appropriate properties of normal distribution?

I. The mean, mode, and median are equal in a normal distribution.
II. The linear combination of two or more normally distributed random variables is not necessarily normally distributed.
III. The normal distribution has a skewness of 0 and excess kurtosis of 3.

A. II and III only.

B. I only.

C. II only.

D. All the above.

The correct answer is **B.**

The appropriate properties of a normal distribution are:
The mean, mode and median are equal in a normal distribution
The linear combination of two or more normally distributed random variables is also normally distributed
The normal distribution has a skewness of 0 and a kurtosis of 3. (Note that the normal distribution does not have an **excess** kurtosis of 3, it has an excess kurtosis of 0.)

Q.3274 In Toronto, Canada, there is a 90% chance of having a sunny day. What is the probability that there will be exactly 3 sunny days in the next 7 days?

A. 0.9

B. 0.00255

C. 0.0625

D. 0.00125

The correct answer is **B.**

$$P(3) = \frac{7!}{((7-3)! * 3!)} * 0.9^3 * (1-0.9)^{7-3} = 0.00255$$

Q.3275 A portfolio has an expected return of 9% with a standard deviation of 7%. If the returns are normally distributed, then what is the probability that the return will be greater than 16%?
Click here to view the normal distribution table.

A. 0.1052

B. 0.2241

C. 0.1228

D. 0.1587

The correct answer is **D.**

A 16% return is 1 standard deviation above the mean of 9%, since the standard deviation is 7% (9% + 7% = 16%). The probability of getting a result more than 1 standard deviation above the mean is 1 - Prob(Z≤1) = 1 - 0.8413 = 0.1587 or 15.87%.
Note: 0.8413 is obtained from the Z-table.

Q.3276 The population living in Calgary, Canada has a mean income of CAD 55,000 with a standard deviation of CAD 10,000. If the distribution is assumed to be normal, what is the percentage of the population that makes between CAD 45,000 and CAD 50,000?
Click here to view the normal distribution table.

A. 0.1498

B. 0.1511

C. 0.1624

D. 0.2014

The correct answer is **A.**

$\frac{(45,000-55,000)}{10,000} = -1$, which corresponds to a Z-score of 0.1587.

$\frac{(50,000-55,000)}{10,000} = -0.5$, which corresponds to a Z-score of 0.3085.

The difference is $0.3085 - 0.1587 = 0.1498$ or 14.98%.

Q.3277 A portfolio's expected return is 17% and its standard deviation is 4%. If the returns are normally distributed, then what is the probability that the returns will be greater than 29%? Use the following standard normal table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999							

A. 0.0013

B. 0.01

C. 0.13

D. 0.0525

The correct answer is **A.**

A 29% return is 3 standard deviations above the mean of 17% (17% + 3*4% = 29%). The probability of getting 3 standard deviations above the mean is 1 - Prob($Z \leq 3$) = 1 - 0.9987 = 0.0013 or 0.13%. Note: 0.9987 is obtained from the Z-table.

Q.3278 A portfolio manager's bonus depends on the return generated by the fund. The different bonus bands are listed below:

Band	Bonus %
Return > 5%	2%
Return > 8%	4%
Return > 12%	10%
Return > 20%	14%
Return > 25%	20%

The mean return and the standard deviation of the fund managed by the portfolio manager stood at 8% and 2%, respectively. Assuming that mutual fund returns are normally distributed, what is the probability that the portfolio manager earns a bonus of 4% this year? Click here to view the normal distribution table.

A. 0.6737

B. 0.5

C. 0.4226

D. 0.4772

The correct answer is **D.**

Calculate the z values at two points:

$$8\% - \text{Z value} = \frac{(8\% - 8\%)}{2\%} = 0$$
$$12\% - \text{Z value} - \frac{(12\% - 8\%)}{2\%} = 2$$
$$P(\text{return} < 8\%) = 50\%$$
$$P(\text{return} < 12\%) = 97.72\%$$
$$P(8\% < \text{ return } < 12\%) = 97.72\% - 50\% = 47.72\%$$

Q.3286 Which statistic should you use to compare 2 population's variances, with a sample size smaller than 30?

A. z-test

B. Chi-square test

C. t-test

D. F-test

The correct answer is **D.**

We always use the F-test to compare two population variances even when the sample size is less than 30. The two important conditions for the test are as follows:

I. The populations from which the two samples are drawn are normally distributed.
II. The two populations are independent of each other.

Q.3294 Which of the following is the most common hypothesis test concerning the difference between the variance of a normally distributed population?

A. Chi-squared test

B. F-test

C. Z-test

D. t-test

The correct answer is **B.**

The F-test is used to test the difference between the variance of two populations based on the standard deviations of the samples drawn from the two populations.

Q.3628 If the rate at which accidents occur in a U.S. city is, on average, three every day, calculate the probability that more than 5 accidents occur on a single day.

A. 0.15

B. 0.06

C. 0.05

D. 0.084

The correct answer is **D.**

The number of accidents in a city can be modeled as a Poisson distribution with mean λ
The probability of n accidents can be given by:

$$P(N = n) = \frac{e^{-\lambda}\lambda^{n}}{n!}$$

We can use the recurrence relationship given:

$$\begin{aligned}
P(N > 5) &= 1 - P(N \leq 5) \\
&= 1 - [P(N = 0) + P(N = 1) + P(N = 2) + P(N = 3) + P(N = 4) + P(N = 5)] \\
&= 1 - \frac{e^{-3}3^{0}}{0!} + \frac{e^{-3}3^{1}}{1!} + \frac{e^{-3}3^{2}}{2!} + \frac{e^{-3}3^{3}}{3!} + \frac{e^{-3}3^{4}}{4!} + \frac{e^{-3}3^{5}}{5!} \\
&= 1 - (0.0498 + 0.1494 + 0.2240 + 0.2240 + 0.1680 + 0.1008) \\
&= 0.084
\end{aligned}$$

Q.3629 Given a binomial random variable with P(X=0) = .20 and P(X=1) = .35 and E(X) = 1.5 calculate Var(X).

A. 0.4

B. 0.6

C. 1

D. 1.3

The correct answer is **D.**

$$p(x) = \binom{n}{x} p^{X}(1-p)^{n-x}$$

$$p(0) = \binom{n}{0} p^{0}(1-p)^{n-0} = 1 * 1 * (1-p)^{n} = (1-p)^{n} = .20$$

$$p(1) = \binom{n}{1} p^{1}(1-p)^{n-1} = np(1-p)^{n-1} = .35$$

$$E(X) = np = 1.5$$

$$(1.5)(1-p)^{n-1} = .35$$

$$(1-p)^{n-1} = .2333$$

$$\frac{(1-p)^{n}}{(1-p)} = .233 = .20/(1-p)$$

$$1-p = .86$$

$$Var(X) = np(1-p) = 1.5(.86) = 1.3$$

Q.3630 Given a binomial random variable with E(X) = 2 and Var(X) = 1.5, calculate P(X=5).

A. 0.015

B. 0.023

C. 0.039

D. 0.047

The correct answer is **B.**

$$E(X) = np = 2$$
$$Var(X) = np(1-p) = 1.5$$
$$Var(X) = 2 * (1-p) = 1.5$$
$$p = .25$$
$$n = 8$$
$$p(X = 5) = \binom{8}{5}(.25^5)(.75^3) = .023$$

Q.3631 Given a Poisson random variable with E(X) = .25, calculate P(X=1)

A. 0.1

B. 0.15

C. 0.19

D. 0.25

The correct answer is **C.**

$$E(X) = \lambda = .25$$
$$p(1) = \frac{e^{-\lambda}\lambda^1}{1!} = e^{-.25} * (.25) = .19$$

Q.3632 Given a Poisson random variable with $P(X=1) = .09$ and $P(X=2) = .0045$, calculate $Var(X)$.

A. 0.009

B. 0.01

C. 0.05

D. 0.1

The correct answer is **D.**

$$
\begin{aligned}
&p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \\
&p(1) = \frac{e^{-\lambda}\lambda^{1}}{1!} = .09 \\
&p(2) = \frac{e^{-\lambda}\lambda^{2}}{2!} = .0045 \\
&p(2) = \frac{.09 * \lambda}{2} = .0045 \\
&\lambda = .10 \\
&Var(X) = \lambda = .10
\end{aligned}
$$

Q.3737 The supermarket sells discounted items for an excellent price; this price is modeled uniformly on the interval [0,500].
Calculate the difference between the median and the 10th percentile.

A. 250

B. 200

C. 240

D. 50

The correct answer is **B.**

The probability distribution function of the price is given by

$$f(x) = \frac{1}{500 - 0} = \frac{1}{500}$$

We can find the median by solving for k in:

$$\begin{aligned} P\,(X \text{ less than } k) &= (k - 0) * f(x) = 0.5 \\ \Rightarrow \frac{1}{500}k &= 0.5 \\ \Rightarrow k &= 0.5 \times 500 = 250 \end{aligned}$$

Similarly, we can find the 10th percentile by solving for k in:

$$\begin{aligned} P\,(X \text{ less than } k) &= (k - 0) * f(x) = 0.10 \\ \Rightarrow \frac{1}{500}k &= 0.10 \\ \Rightarrow k &= 0.10 \times 500 = 50 \end{aligned}$$

The difference between these amounts is:

$$250 - 50 = 200$$

Reading 15: Multivariate Random Variables

Q.329 Which of the following statements is NOT true regarding the correlation coefficient?

> A. The correlation coefficient measures the strength of the linear relationship between two random variables.
>
> B. The correlation coefficient has no units.
>
> C. The correlation coefficient ranges from 0 to +1.
>
> D. Random variables with a correlation of +1 are said to be perfectly correlated.

The correct answer is **C.**

In finance, the correlation coefficient attempts to measure the degree to which two random variables, say, returns for different stocks move in relation to each other. The correlation coefficient always lies between -1 and +1. A positive value indicates that the random variables move in the same direction i.e. if an increase (decrease) is recorded in one variable, we expect an increase (decrease) in the other variable, which can either be proportionate or disproportionate depending on the value of the correlation.

Q.332 Two random variables X and Y are such that $V[X] = 4V[Y]$ and $Cov[X,Y] = V[Y]$
Let $E = X + Y$ and $F = X - Y$

Find Cov[E, F].

A. $V[Y] - V[X]$

B. $Cov[X,Y]$

C. $V[Y]$

D. $3V[Y]$

The correct answer is **D.**

$$
\begin{aligned}
Cov[E,F] &= Cov[X+Y, X-Y] \\
&= Cov[X,X] - Cov[X,Y] + Cov[Y,X] - Cov[Y,Y] \\
&= V[X] - V[Y] \\
&= 4V[Y] - V[Y] \\
&= 3V[Y]
\end{aligned}
$$

Logic applied:

I. Given a random variable X, the covariance between X and itself is simply its variance

II. $Cov[X,Y] = Cov[Y,X]$

Q.338 Two stocks, X and Y, have a correlation of 0.50. Stock Y's return has a standard deviation of 0.26. Given that the covariance between X and Y is 0.005, determine the variance of returns for stock X.

A. 0.13

B. 0.00148

C. 0.0385

D. 0.0148

The correct answer is **B.**

Correlation between X and Y,

$$\begin{aligned}
\text{Corr}(X, Y) &= \frac{\text{Cov}(X, Y)}{(\sigma_X * \sigma_Y)} \\
0.50 &= \frac{0.005}{(\sigma_X * 0.26)} \\
0.13\sigma_X &= 0.005 \\
\sigma_X &= 0.0385 \\
\text{Variance}(X) &= \sigma^2 = 0.0385^2 = 0.00148
\end{aligned}$$

Q.349 Which of the following *best describes* the central limit theorem?

A. When the sample size is large, the sum of independent and identically distributed (i.i.d.) random variables are normally distributed.

B. The sum of n independent and identically distributed random variables approaches the normal distribution as n becomes large.

C. For simple random samples of size n from a population with mean μ and finite variance σ^2, the sampling distribution approaches the normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, as the sample size becomes large.

D. For simple random samples of size n from a population with mean μ and finite variance σ^2, the sampling distribution of the sample mean approaches the normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, as the sample size becomes large.

The correct answer is **D.**

The Central Limit Theorem states that the sampling distribution of the sample means approaches a normal distribution as the sample size gets larger — no matter what the shape of the population distribution. This fact holds especially true for sample sizes over 30. All this is saying is that as you take more samples, especially large ones of size 30 or more, your graph of the sample means will look more like a normal distribution.

Please note that it's the sample mean that's normally distributed, not the sample itself. This is why option C is incorrect. This handy result considerably simplifies the computation of probabilities and the construction of a statistical hypothesis. So long as we have sample statistics, we can draw relevant conclusions about the actual population regardless of the population's distribution, provided n is sufficiently large (n is usually taken to be ≥ 30).

Q.3263 Assuming that the covariance of returns of Stock X and Stock Y is $Cov(R_X, R_Y) = 0.093$, the variance of $R_X = 0.69$, and the variance of $R_Y = 0.36$, the correlation of returns of Stock X and Stock Y is *closest to*:

A. 0.155.

B. 0.1865.

C. 0.1119.

D. 0.2133

The correct answer is **B.**

$$Corr(R_X, R_Y) = \frac{Cov(R_X, R_Y)}{\sigma(R_X)\sigma(R_Y)}$$

Since

$$\begin{aligned}
\text{Variance} &= \sigma^2 \\
\sigma(R_X) &= \sqrt{0.69} = 0.8306, \text{ and} \\
\sigma(R_Y) &= \sqrt{0.36} = 0.6 \\
Corr R_X, R_Y &= \frac{Cov R_X, R_Y}{\sigma(R_X)\sigma(R_Y)} = \frac{0.093}{0.8306 * 0.6} = 0.1865
\end{aligned}$$

Q.3266 The covariance matrix of two stocks is given in the following exhibit.

Exhibit: Covariance Matrix

Stock	X	Y
X	650	120
Y	120	450

What is the correlation of returns for stocks X and Y?

A. 0.45

B. 0.22.

C. 0.37.

D. 0.33

The correct answer is **B.**

We can start by calculating the standard deviation from the given variance.

$$\sigma(X) = (650)^{0.5} = 25.50$$
$$\sigma(Y) = (450)^{0.5} = 21.21$$
$$\text{Covariance}(X,Y) = 120$$
$$\text{Correlation}(X, Y) = \frac{\text{Cov}(R_X, R_Y)}{\sigma(R_X)\sigma(R_Y)} = \frac{(120)}{(25.50 * 21.21)} = 0.22$$

Q.3267 A portfolio consists of two funds A and B. The weights of the two funds in the portfolio and the covariance matrix of the two funds are given in the following two exhibits.

Exhibit 1: Weight of the Funds in the Portfolio

Fund	A	B
Weight	60%	40%

Exhibit 2: Covariance Matrix

Fund	A	B
A	700	200
B	200	500

What is the portfolio variance?

A. 428

B. 500

C. 324

D. 328

The correct answer is **A.**

Based on the covariance matrix:

$$
\begin{aligned}
\text{Variance}_{\text{Portfolio}} &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{1,2}\sigma_1\sigma_2 \\
&= 0.60^2 \times 700 + 0.40^2 \times 500 + 2 \times 0.60 \times 0.40 \times 200 \\
&= 428
\end{aligned}
$$

Fund	A	B
A	700	200
B	200	500

Q.3627 Living in a certain city is really expensive. The mean spent by a citizen monthly is $3,000 with a variance of $500. Nonetheless, the city mayor decided to put a tax in order to make people regulate their monthly expenses adding 20% to all daily living articles. Find the new variance of this city.

A. 520

B. 580

C. 600

D. 720

The correct answer is **D.**

For this exercise, we need to find the variance of new living costs. Let X be the old living costs and Y the new living costs, then we have:

$$\begin{aligned} Y &= 1.2X \\ Var(Y) &= Var(1.2X) \\ &= 1.44Var(X) \\ &= 1.44(500) = 720 \end{aligned}$$

Note: $Var(aX) = a^2 Var(X)$

Q.3742 The two random variables X_1 and X_2 have well-defined variances, and they are independent. What is the correlation of these random variables?

A. 0

B. 1

C. -1

D. 0.5

The correct answer is **A.**

The correlation between independent random variables is always 0. This is true because the variances of the random variables are well defined.

Q.3743 The yearly profits of the two firms A and B can be summarized in the following probability matrix.

		Company A (X_1) Profits			
		-1 Million	0 Million	2 Million	4 Million
Company B (X_2) Profits	-50 Million	0.0197	0.0395	0.010	0.002
	0 Million	0.0390	0.230	0.124	0.0298
	10 Million	0.011	0.127	0.144	0.0662
	100 Million	0	0.0309	0.0656	0.0618

What is the marginal distribution of company A?

Table A: Marginal Distribution of Company A

Company A(X_1) Profits	-1 Million	0 Million	2 Million	4 Million
$P(X_1 = x_1)$	0.0697	0.4274	0.3436	0.1598

Table B: Marginal Distribution of Company A

Company A(X_1) Profits	-1 Million	0 Million	2 Million	4 Million
$P(X_1 = x_1)$	0.0697	0.5274	0.3436	0.0593

Table C: Marginal Distribution of Company A

Company A(X_1) Profits	-1 Million	0 Million	2 Million	4 Million
$P(X_1 = x_1)$	0.0593	0.5274	0.3436	0.0697

Table D: Marginal Distribution of Company A

Company A(X_1) Profits	-1 Million	0 Million	2 Million	4 Million
$P(X_1 = x_1)$	0.0697	0.5274	0.1235	0.2794

A. Table A

B. Table B

C. Table C

D. Table D

The correct answer is **A.**

Recall that the marginal distribution gives the distribution of a single component in a joint distribution. In the case of bivariate distribution, the marginal PMF of X_1 is computed by summing up the probabilities for X_1 across all the values at which X_2 is realized.

For example, the first probability $P(X_1 = -1M)$ is calculated as:

$$0.0197 + 0.0390 + 0.011 + 0 = 0.0697$$

So, there is 6.97% that company A will go at a loss of 1 million.

Note that we are adding along the columns. For the second probability $P(X_1 = 0M)$ is calculated as:

$$= 0.0395 + 0.230 + 0.127 + 0.0309 = 0.4274$$

The third and the fourth probabilities are calculated in a similar manner.

Q.3744 The yearly profits of the two firms A and B can be summarized in the following probability matrix.

		Company A (X_1) Profits			
		-1 Million	0 Million	2 Million	4 Million
Company B (X_2) Profits	-50 Million	0.0197	0.0395	0.010	0.002
	0 Million	0.0390	0.230	0.124	0.0298
	10 Million	0.011	0.127	0.144	0.0662
	100 Million	0	0.0309	0.0656	0.0618

What is the marginal distribution of company B?

A.

Company B(X_2) Profits	-50 Million	0 Million	10 Million	100 Million
$P(X_2 = x_2)$	0.1325	0.4244	0.3599	0.0832

B.

Company B(X_2) Profits	-50 Million	0 Million	10 Million	100 Million
$P(X_2 = x_2)$	0.0235	0.4856	0.3254	0.1655

C.

Company B(X_2) Profits	-50 Million	0 Million	10 Million	100 Million
$P(X_2 = x_2)$	0.0712	0.4228	0.3482	0.1583

D.

Company B(X_2) Profits	-50 Million	0 Million	10 Million	100 Million
$P(X_2 = x_2)$	0.0633	0.4423	0.3658	0.1286

A. Table A

B. Table B

C. Table C

D. Table D

The correct answer is **C.**

To compute the marginal distribution for company B, we need to sum across the rows. For instance, $P(X_2 = -50)$ is given by:

$$0.0197 + 0.0395 + 0.010 + 0.002 = 0.0712$$

So, there is a 7.12% chance that company B will go at a loss of 50 million.

Similarly, $P(X_2 = 0)$:

$$0.0390 + 0.230 + 0.124 + 0.0298 = 0.4228$$

That is, there is 42.28% chance that company B will not make any profit.

Q.3745 The yearly profits of the two firms A and B can be summarized in the following probability matrix.

		Company A (X_1) Profits			
		-1 Million	0 Million	2 Million	4 Million
Company B (X_2) Profits	-50 Million	0.0197	0.0395	0.010	0.002
	0 Million	0.0390	0.230	0.124	0.0298
	10 Million	0.011	0.127	0.144	0.0662
	100 Million	0	0.0309	0.0656	0.0618

What is the conditional distribution of company A given that company B made a profit of 100 Million?

A.

Company A(X_1) Profits	-1 Million	0 Million	2 Million	4 Million
$P(X_1 \mid X_2 = 100)$	0.0697	0.4274	0.3436	0.1598

B.

Company A(X_1) Profits	-1 Million	0 Million	2 Million	4 Million
$P(X_1 \mid X_2 = 100)$	0.0697	0.5274	0.6436	0.1598

C.

Company A(X_1) Profits	-1 Million	0 Million	2 Million	4 Million
$P(X_1 \mid X_2 = 100)$	0.0697	0.5274	0.3436	0.2598

D.

Company A(X_1) Profits	-1 Million	0 Million	2 Million	4 Million
$P(X_1 \mid X_2 = 100)$	0	0.1952	0.4144	0.3904

A. Table A

B. Table B

C. Table C

D. Table D

The correct answer is **D.**

The conditional distributions describe the probability of an outcome of a random variable conditioned on the other random variable taking a particular value.

Recall that given any two events A and B, then:

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

This result can be applied in bivariate distributions. That is, the conditional distribution of X_1 given X_2 is defined as:

$$f_{(X_1|X_2)}(x_1|X_2 = x_2) = \frac{f_{(X_1,X_2)}(x_1, x_2)}{f_{X_2}(x_2)}$$

So, in this case, we need:

$$f_{(X_1|X_2)}(x_1|X_2 = 100) = \frac{f_{(X_1,X_2)}(x_1, x_2)}{f_{X_2}(X_2 = 100)}$$

The marginal distribution of company B($f_{X_2}(x_2)$) given by:

Company B(X_2) Profits	-50 Million	0 Million	10 Million	100 Million
$P(X_2 = x_2)$	0.0712	0.4228	0.3482	0.1583

We are simply summing up the probabilities given in the table from left to right:

Thus:

P(Profit = -50) = 0.0197 + 0.0395 + 0.010 + 0.002 = 0.0712

P(Profit = 0) = 0.0390 + 0.230 + 0.124 + 0.0298 = 0.4228

P(Profit = 10) = 0.011 + 0.127 + 0.144 + 0.0662 = 0.3482

P(Profit = 100) = 0 + 0.0309 + 0.0656 + 0.0618 = 0.1583

The joint distribution is the following table (given at the introduction of this question)

		Company A (X_1) Profits			
		-1 Million	0 Million	2 Million	4 Million
Company B (X_2) Profits	-50 Million	0.0197	0.0395	0.010	0.002
	0 Million	0.0390	0.230	0.124	0.0298
	10 Million	0.011	0.127	0.144	0.0662
	100 Million	0	0.0309	0.0656	0.0618

To calculate the conditional distribution, we divide the individual components of the last column by the the marginal distribution(0.1583). So, the conditional distribution is:

Company A(X_1) Profits	-1 Million	0 million	2 Million	4 Million
$P(X_1 \mid X_2 = 100)$	$\frac{0}{0.1583} = 0$	$\frac{0.0309}{0.1583} = 0.1952$	$\frac{0.0656}{0.1583} = 0.4144$	$\frac{0.0618}{0.1583} = 0.3904$

Q.3746 The yearly profits of the two firms A and B can be summarized in the following probability matrix.

		Company A (X_1) Profits			
		-1 Million	0 Million	2 Million	4 Million
Company B (X_2) Profits	-50 Million	0.0197	0.0395	0.010	0.002
	0 Million	0.0390	0.230	0.124	0.0298
	10 Million	0.011	0.127	0.144	0.0662
	100 Million	0	0.0309	0.0656	0.0618

Looking at the marginal distribution of companies A and B, which of the following of the statements is true?

A. The marginal distributions are independent.

B. The marginal distributions are not independent.

C. The distributions are not ideal probability distributions.

D. None of the above.

The correct answer is **B.**

Recall that if the distributions of the components of the bivariate distributions are independent, then:

$$f_{(X_1,X_2)}(x_1, x_2) = f_{(X_1)}(x_1)f_{(X_2)}(x_2)$$

If we calculate the first value (upper-left cell) using the corresponding marginal distributions, we have:

$$f_{(X_1)}(x_1)f_{(X_2)}(x_2) = 0.0712 \times 0.0697 = 0.00496264$$

The corresponding joint distribution at the upper left cell is:

$$f_{(X_1,X_2)}(x_1, x_2) = 0.0197$$

Therefore,

$$f_{(X_1,X_2)}(x_1, x_2) \neq f_{(X_1)}(x_1)f_{(X_2)}(x_2)$$

And thus, the marginal distributions are not independent. Option C is incorrect because, for an ideal discrete distribution function, the following properties must be met:

$$\sum_x f_X(x) = 0$$

And

$$f_X(x) \geq 0$$

Both the marginal distributions meet these properties, and a slight deviation from 1 may be due to rounding off.

Q.3747 The yearly profits of the two firms A and B can be summarized in the following probability matrix.

		Company A (X_1) Profits			
		-1 Million	0 Million	2 Million	4 Million
Company B	-50 Million	0.0197	0.0395	0.010	0.002
(X_2)	0 Million	0.0390	0.230	0.124	0.0298
Profits	10 Million	0.011	0.127	0.144	0.0662
	100 Million	0	0.0309	0.0656	0.0618

What is the covariance of company A and B given that $E(X_1X_2) = 43.23$?

A. 24.56

B. 23.43

C. 21.45

D. 22.45

The correct answer is **B.**

We know the covariance is given by:

$$\text{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$$

Now,

Using the marginal distribution:

Company A(X_1) Profits	-1 Million	0 Million	2 Million	4 Million
$P(X_1 = x_1)$	0.0697	0.4274	0.3436	0.1598

$$\begin{aligned} E(X_1) &= \sum_{\forall x1} x_1 P(X_1 = x_1) \\ &= -1 \times 0.0697 + 0 \times 0.4274 + 2 \times 0.3436 + 4 \times 0.1598 \\ &= 1.2567 \end{aligned}$$

Similarly, using the marginal distribution:

Company B(X_2) Profits	-50 Million	0 Million	10 Million	100 Million
$P(X_2 = x_2)$	0.0712	0.4228	0.3482	0.1583

$$\begin{aligned} E(X_2)E(X_1) &= \sum_{\forall x_2} x_2 P(X_2 = x_2) \\ &= -50 \times 0.0712 + 0 \times 0.4228 + 10 \times 0.3482 + 100 \times 0.1583 \\ &= 15.752 \end{aligned}$$

So that the covariance of companies A and B is given by:

$$\text{Cov}(A, B) = E(X_1X_2) - E(X_1)E(X_2) = 43.23 - 1.2567 \times 15.752 = 23.43$$

Q.3748 The yearly profits of the two firms A and B can be summarized in the following probability matrix.

		Company A Profits	(X_1)		
		-1 Million	0 Million	2 Million	4 Million
Company B	-50 Million	0.0197	0.0395	0.010	0.002
(X_2)	0 Million	0.0390	0.230	0.124	0.0298
Profits	10 Million	0.011	0.127	0.144	0.0662
	100 Million	0	0.0309	0.0656	0.0618

What is the correlation coefficient between the two companies A and B if Cov(A, B) = 23.43 ?

A. 0.4553

B. 0.3827

C. 0.4562

D. 0.5651

The correct answer is **B.**

Recall that the correlation coefficient is given by:

$$\text{Corr}(X_1, X_1) = \rho_{X_1 X_2} = \frac{\text{Cov}(X_1, X_2)}{\sqrt{s_1^2}\sqrt{s_2^2}} = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

We have $\text{Cov}(A, B) = \text{Cov}(X_1, X_2) = 23.43$. We need to calculate the variance of each component, and we are good to go.

Recall that:

$$Var(X) = \sigma_X^2 = E(X^2) - [E(X)]^2$$

So that,

$$Var(X_1) = \sigma_{X_1}^2 = E(X_1^2) - [E(X_1)]^2$$

Using the marginal distribution of company, A:

$$E(X_1^2) = (-1)^2 \times 0.0697 + 0^2 \times 0.4274 + 2^2 \times 0.3436 + 4^2 \times 0.1598 = 4.0009$$

We had calculated $E(X_1) = 1.2567$ then:

$$Var(X_1) = 4.0009 - [1.2567]^2 = 2.4216$$

Similarly, using the marginal distribution for company B

$$Var(X_2) = \sigma_{X_2}^2 = E(X_2^2) - [E(X_2)]^2$$

$$\begin{aligned} E(X_2^2) &= (-50)^2 \times 0.0712 + 0 \times 0.4228 + 10^2 \times 0.3482 + 100^2 \times 0.1583 \\ &= 1795.82 \end{aligned}$$

We had calculated $E(X_2) = 15.752$

So, that:

$$Var(X_2) = 1795.82 - [15.752]^2 = 1547.6945$$

So, if we substitute these in our correlation coefficient, we get:

$$Corr(X_1, X_1) = \rho_{X_1X_2} = \frac{23.43}{\sqrt{2.4216}\sqrt{1547.6945}} = 0.3827$$

Q.3749 An investor invests 30% of his assets in security A and 70% in security B. The variance of returns for security in security A is 1234.56, and that of B is 243.56. The covariance between securities A and B is 25.56. What is the standard deviation of the combined returns from these securities?

A. 18.89

B. 14.78

C. 15.53

D. 13.45

The correct answer is **C.**

The standard deviation is just the square root of the variance. The variance is given by:

$$\begin{aligned}\sigma_{A+B}^2 &= w_A^2\sigma_A^2 + w_B^2\sigma_B^2 + 2w_Aw_B\text{Cov}(A, B)\\ &= 0.3^2 \times 1234.56 + 0.7^2 \times 243.56 + 2 \times 0.3 \times 0.7 \times 25.56\\ &= 241.19\end{aligned}$$

So the standard deviation of the portfolio (combined returns from these securities) is given by:

$$\sigma_{A+B} = \sqrt{241.19} = 15.53$$

Q.3750 The resulting probability matrix displays the amount of returns of two income-generating sections of bank: Loans and Stock Market

Loans	Returns(X_1)	−20%	0%	20%
Return	Probability	30%	55%	15%
Stock Market	Returns(X_2)	−5%	0%	9%
Returns	Probability	40%	31%	29%

Assuming that the two income-generating avenues are independent of each other, what is the joint probability distribution (matrix)?

A.

	Loan Return (X_1)			
		−20%	0%	20%
Stock	−5%	12%	22%	6%
Market	0%	9.3%	17.05%	4.65%
Returns(X_2)	9%	8.7%	15.95%	4.35%

B.

	Loan Return (X_1)			
		−20%	0%	20%
Stock	−5%	12%	12%	7%
Market	0%	10.3%	17.05%	4.65%
Returns(X_2)	9%	8.7%	15.95%	4.35%

C.

	Loan Return (X_1)			
		−20%	0%	20%
Stock	−5%	12%	12%	6%
Market	0%	9.3%	27.05%	5.65%
Returns(X_2)	9%	8.7%	25.95%	4.35%

D.

	Loan Return (X_1)			
		−20%	0%	20%
Stock	−5%	12%	22%	6%
Market	0%	9.3%	14.05%	4.65%
Returns(X_2)	9%	7.7%	55.95%	4.35%

A. Table A

B. Table B

C. Table C

D. Table D

The correct answer is **A.**

Recall that if two marginal distributions are independent then:

$$f_{(X_1,X_2)}(x_1,x_2) = f_{(X_1)}(x_1)f_{(X_2)}(x_2)$$

We are given the marginal distributions so that the joint distributions are given by multiplying their corresponding PMFs. For example, the joint probability that loan return is -20%, and the stock return is -5% is 30%×40%=12.

The other joint distributions are given in the table below:

Loans Return	Returns(X_1)	−20%	0%	20%
	Probability	30%	55%	15%
Stock Market Returns	Returns(X_2)	−5%	0%	9%
	Probability	40%	31%	29%

		Loan Return(X_1)		
		−20%	0%	20%
Stock	5%	$40\% \times 30\% = 12\%$	$40\% \times 55\% = 22\%$	$40\% \times 15\% = 6\%$
Market	0%	$31\% \times 30\% = 9.3\%$	$31\% \times 55\% = 17.05\%$	$31\% \times 15\% = 4.65\%$
Returns(X_2)	9%	$29\% \times 30\% = 8.7\%$	$29\% \times 55\% = 15.95\%$	$29\% \times 15\% = 4.35\%$

The resulting probability matrix displays the amount of returns of two income-generating sections of bank: Loans and Stock Market

		Loan Return (X_1)		
		−20%	0%	20%
Stock	−5%	12%	22%	6%
Market	0%	9.3%	17.05%	4.65%
Returns(X_2)	9%	8.7%	15.95%	4.35%

Q.3751 The resulting probability matrix displays the amount of returns of two independent income-generating sections of bank: Loans and Stock Market

Loans Return	Returns(X_1)	20%	0%	20%
	Probability	30%	55%	15%
Stock Market Returns	Returns(X_2)	−5%	0%	9%
	Probability	40%	31%	29%

What is the conditional distribution of loan returns given that the return from the stock market is 9%?

A.

Loans Return(X_1)	−20%	0%	20%
$P(X_1 \mid X_2 = 9\%)$	40%	31%	29%

B.

Loans Return(X_1)	−20%	0%	20%
$P(X_1 \vert X_2 = 9\%)$	9.3%	17.05%	4.65%

C.

Loans Return(X_1)	−20%	0%	20%
$P(X_1 \vert X_2 = 9\%)$	30%	55%	15%

D.

Loans Return(X_1)	−20%	0%	20%
$P(X_1 \vert X_2 = 9\%)$	8.7%	15.95%	4.35%

A. Table A

B. Table B

C. Table C

D. Table D

The correct answer is **C.**

The conditional distribution of X_1 given X_2 is defined as:

$$f_{(X_1 \mid X_2)}(x_1 \mid X_2 = x_2) = \frac{f_{(X_1, X_2)}(x_1, x_2)}{f_{(X_2)}(x_2)}$$

So, in this case, we need:

$$f_{(X_1 \mid X_2)}(x_1 \mid X_2 = 10\%) = \frac{f_{(X_1, X_2)}(x_1, x_2)}{f_{(X_2)}(X_2 = 9\%)}$$

Note that we are given the marginal distribution of stock market return $f_{(X_2)}(x_2)$ given by:

Stock Market	Returns(X_2)	−5%	0%	9%
Returns	Probability	40%	31%	29%

We calculated the joint distribution as:

		Loan Return (X_1)		
		−20%	0%	20%
Stock	−5%	12%	22%	6%
Market	0%	9.3%	17.05%	4.65%
Returns(X_2)	9%	8.7%	15.95%	4.35%

To calculate the conditional distribution, we divide the last column by the corresponding marginal distribution(29%). So, the conditional distribution is:

Loans Return(X_1)	-20%	0%	20%
$P(X_1 \mid X_2 = 9\%)$	$\frac{8.7\%}{29\%} = 30\%$	$\frac{15.95\%}{29\%} = 55\%$	$\frac{4.35\%}{29\%} = 15\%$

Q.3752 Three random variables X, Y, and Z have the equal variance of $\sigma^2 = 2$. X is independent of both Y and Z, and that Y and Z are correlated with a correlation coefficient of 0.8. What is the covariance between X and K given that K=Y+Z?

A. 0

B. 2

C. 3

D. 4

The correct answer is **A.**

Recall from the properties of the covariance that:

$$\text{Cov}(A, B + C) = \text{Cov}(A, B) + \text{Cov}(A, C)$$

So, we need:

$$\text{Cov}(X, K) = \text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z) = 0 + 0 = 0$$

This is true because the X in uncorrelated with Y and Z!

Q.3753 Three random variables X, Y, and Z have equal variance of $\sigma^2 = 2$. X is independent of both Y and Z, and that Y and Z are correlated with a correlation coefficient of 0.8. What is the covariance between Z and V given that $V = 3X - 2Y$.

A. 4.1

B. -4.3

C. 3.2

D. -3.2

The correct answer is **D.**

Using the same property that:

$$\text{Cov}(A, B + C) = \text{Cov}(A, B) + \text{Cov}(B, C)$$

Then,

$$\text{Cov}(Z, L) = \text{Cov}(Z, 3X - 2Y) = 3\text{Cov}(Z, X) - 2\text{Cov}(Z, Y)$$

But $\text{Cov}(Z, X) = 0$. So,

$$\text{Cov}(Z, 3X - 2Y) = 0 - 2\text{Cov}(Z, Y) = -2\text{Cov}(Z, Y)$$

Also, recall that the correlation coefficient between two random variables A and B

$$\rho_{AB} = \frac{\text{Cov}(A, B)}{\sigma_A \sigma_B} \Rightarrow \text{Cov}(A, B) = \rho_{AB} \times \sigma_A \sigma_B$$

We are given $\rho_{YZ} = 0.8$ with equal variances. So, using the above reasoning, we have:

$$\text{Cov}(Z, Y) = \rho_{ZY} \times \sigma_Z \sigma_Y = \rho_{ZY} \times \sigma^2 = 0.8 \times 2 = 1.6$$

Therefore,

$$\text{Cov}(Z, 3X - 2Y) = -2\text{Cov}(Z, Y) = -2 \times 1.6 = -3.2$$

Note that if X and Y are independent variables, then their covariance is 0.

Q.3754 The amount of profit (X) for a sales company is continuously distributed uniformly with the parameters 0 and 1,500. However, a financial analyst believes that the actual profit (Y) is a minimum of X. What is the conditional distribution of X given X<1,300?

A. Continuous uniform with parameters 0 and 1,300.

B. Continuous uniform with parameters 0 and 1,500.

C. Continuous uniform with parameters 0 and 1,000.

D. Continuous uniform with parameters 0 and 2,800.

The correct answer is **A.**

We are given that: $X \sim U(0, 1500)$ and that $Y = \min(X, 1300)$

We need:

$$P(X < x | X < 1300)$$

Recall the conditional distribution is given by:

$$f_{(X_1 \mid X_2)}(x_1 \mid X_2 = x_2) = \frac{f_{(X_1, X_2)}(x_1, x_2)}{f_{(X_2)}(x_2)}$$

So that,

$$\begin{aligned} P(X < x \mid X < 1300) &= \frac{P(X < x, X < 1300)}{P(X < 1300)} \\ &= \frac{P(X < x)}{\frac{1300}{1500}} = \frac{\left(\frac{x}{1500}\right)}{\left(\frac{1300}{1500}\right)} = \frac{x}{1300} \end{aligned}$$

Therefore, the marginal distribution is a continuous uniform with the parameters 0 and 1300. This should be intuitive since we are computing conditional distribution, which is the scaled version of the original distribution.

This can be seen where $P(X < x, X < 1300) = P(X < x)$ since 1300 is less than 1500.

Q.3755 Which one of these correlation coefficients shows the weakest linear relationship?

A. -0.8

B. 0.65

C. 0.30

D. 0.56

The correct answer is **C.**

The closer to -1 or +1, the stronger the relationship. The closer to 0, the weaker. Out of the three options, 0.3 is the weakest.

Q.3756 The random variables X and Y have a discrete joint distribution with joint probability function:

$$P(X = x, Y = y) = c(x + 2y); x = 0, 1, 2; y = 0, 1, 2$$

Determine the value of c.

A. $\frac{1}{10}$

B. $\frac{1}{27}$

C. 1

D. $\frac{1}{8}$

The correct answer is **B.**

An ideal joint distribution is:

$$\sum_{x}\sum_{y} f_X f_Y(x, y) = 1$$

Using this concept, we get:

$$\begin{aligned}\sum_{x}\sum_{y} f_X f_Y(x,y) &= \sum_{(x=0)}^{2}\sum_{(y=0)}^{2}(c(x + 2y)\\ &= \sum_{(x=0)}^{2}(cx + cx + 2c + cx + 4c) = 3cx + 6c = 1\\ &= \sum_{(x=0)}^{2}(3cx + 6c = 6c + 3c + 6c + 6c + 6c) = 27c = 1 \Rightarrow c = \frac{1}{27}\end{aligned}$$

Q.3758 Let X represent the age of an insured automobile involved in an accident. Let Y denote the length of time the insurance contract has been in place at the time of the accident. X and Y have joint probability density function

$$f(x, y) = \begin{cases} \frac{1}{64}(10 - xy^2), & 2 < x < 10, 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

What is the expected length of time the contract has been in place for an insured automobile involved in an accident?

A. 0.4563

B. 0.5500

C. 0.4375

D. 0.2010

The correct answer is **C.**

We need to calculate the marginal distribution of Y.

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(xy)\, dx = \int_2^{10} \frac{1}{64}(10 - xy^2)dx = \frac{1}{64}\left[10x - \frac{x^2y^2}{2}\right]_2^{10} = \frac{1}{64}[80 - 48y^2]$$

So,

$$f(y) = \begin{cases} \frac{1}{64}[80 - 48y^2], & 0 < y > 1 \\ 0, & \text{elsewhere} \end{cases}$$

We need to calculate the expectation. Recall that for a continuous random variable, the expectation is given by:

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy$$

So, in this case,

$$E(Y) = \int_0^1 y.\left(\frac{1}{64}[80 - 48y^2]\right) = \frac{1}{64}[40y^2 - 12y^4]_0^1 = \frac{1}{64}[40 - 12] = \frac{28}{64} = \frac{7}{16} = 0.4375$$

Q.3759 What is the correlation of returns between these two portfolios?

- Portfolio A's variance of returns: 52.5%
- Portfolio B's variance of returns: 63%
- The covariance of return between the two portfolios: 0.315

A. 0.8257

B. 0.0011

C. 0.5477

D. 0.9524

The correct answer is **C.**

$$\begin{aligned}\text{Corr}(R_A, R_B) &= \frac{\text{Cov}(R_A, R_B)}{\sigma_A \sigma_B} \\ &= \frac{0.315}{\sqrt{0.525} \times \sqrt{0.63}} = 0.5477\end{aligned}$$

Note: We had to transform the variance into standard deviation by using $\sqrt{0.525}$ and $\sqrt{0.63}$

Q.3760 Given the following joint probability density function of two random variables:

$$f(x_1, x_2) = \begin{cases} 8x_1x_2, & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Find $E(X_1|X_2 = 0.5)$

A. 0.01

B. 0.54

C. 0.33

D. 0.67

The correct answer is **D.**

We need to calculate the conditional probability $f(x_1|X_2 = 0.5)$. Recall that:

The conditional distribution is given by:

$$f_{(X_1 \mid X_2)}(X_1 \mid X_2 = x_2) = \frac{f_{(X_1,X_2)}(x_1, x_2)}{f_{(X_2)}(x_2)}$$

So,

$$f(x_1 \mid X_2 = 0.5) = \frac{f_{(X_1,X_2)}(x_1, x_2 = 0.5)}{f_{(X_2)}(x_2 = 0.5)}$$

The denominator $f_{(X_2)}(x_2 = 0.5)$ requires that we calculate the marginal distribution of X_2 which is given by:

$$\begin{aligned} f_{(X_2)}(x_2) &= \int_{-\infty}^{\infty} f_{(X_1,X_2)}(x_1, x_2)dx_1 \\ &= \int_0^1 8x_1x_2dx_1 = [4x_1^2x_2]_0^1 = 4x_2 - 0 = 4x_2 \end{aligned}$$

So, $f_{(X_2)}(x_2 = 0.5) = 4 \times 0.5 = 2$

Thus,

$$f(x_1 \mid X_2 = 0.5) = \frac{f_{(X_1,X_2)}(x_1, x_2 = 0.5)}{f_{(X_2)}(x_2 = 0.5)} = \frac{(8x_1 \times 0.5)}{2} = 2x_1$$

Given that

$$f(x_1 \mid X_2 = 0.5) = 2x_1, 0 < x_1 < 1$$

We know that $E(X) = \int_{-\infty}^{\infty} xf_X(x)dx$. Then:

$$E(X_1 \mid X_2 = 0.5) = \int_{-\infty}^{\infty} x_1 f_{X_1 \mid X_2 = 0.5}(X_1 \mid X_2 = 0.5)dx_1 = \int_0^1 x_1 \times 2x_1dx_1 = [\frac{2}{3}x_1^3]_0^1 = \frac{2}{3} - 0 = \frac{2}{3} \approx 0.67$$

Q.3761 A financial risk manager believes that the prevailing interests rate (X_1) and the return in a stock market (X_2) can be modeled using the following joint probability function:

$$f(x_1, x_2) = \begin{cases} \frac{1}{8} x_1 x_2, & 0 \le x_1 \le 1, 0 \le x_2 \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

What is the covariance between the interest rate and the return in the stock market?

A. 0.0972

B. 0.0444

C. 0.2222

D. 0.0555

The correct answer is **A.**

We need $\text{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$

So, from the joint distribution, we need to $E(X_1X_2)$, $E(X_1)$ and $E(X_2)$ (of course, after evaluating the marginal distributions).

Let us start with the interest rate (X_1):

The marginal distribution of (X_1) is given by:

$$\begin{aligned} f_{(X_1)}(x_1) &= \int_{-\infty}^{\infty} f_{(X_1,X_2)}(x_1, x_2) dx_2 \\ &= \int_0^2 \frac{1}{8} x_1 x_2 dx_2 = \left[\frac{1}{16} x_1 x_2^2\right]_0^2 = \frac{1}{4} x_1 - 0 = \frac{1}{4} x_1 \\ \Rightarrow f_{(X_1)}(x_1) &= \frac{1}{4} x_1, \quad 0 \le x_1 \le 1 \end{aligned}$$

Now,

$$E(X_1) = \int_{-\infty}^{\infty} x_1 . f_{(X_1)}(x_1) = \int_0^1 x_1 . \frac{1}{4} x_1 = \left[\frac{1}{12} x_1^3\right]_0^1 = \frac{1}{12}$$

For the stock return X_2,

$$
\begin{aligned}
f_{(X_2)}(x_2) &= \int_{-\infty}^{\infty} f_{(X_1,X_2)}(x_1, x_2)dx_1 \\
&= \int_0^1 \frac{1}{8}x_1x_2dx_2 = [\frac{1}{16}x_1^2x_2]_0^1 = \frac{1}{16}x_2 - 0 = \frac{1}{16}x_2 \\
\Rightarrow f_{(X_1)}(x_1) &= \frac{1}{16}x_2, \quad 0 \leq x_1 \leq 2
\end{aligned}
$$

Now,

$$
E(X_2) = \int_{-\infty}^{\infty} x_2.f_{(X_2)}(x_2) = \int_0^2 x_2.\frac{1}{16}x_2 = [\frac{1}{48}x_2^3]_0^2 = \frac{1}{6}
$$

We also need to calculate $E(X_1X_2)$:

$$
E(X_1X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2.f_{(X_1X_2)}(x_1,x_2)dx_1dx_2
$$

For this case,

$$
\begin{aligned}
E(X_1X_2) &= \int_0^1\int_0^2 x_1x_2.\frac{1}{8}x_1x_2dx_1dx_2 \\
\int_0^1 \frac{1}{8}[\frac{1}{3}x_1^2x_2^3]_0^2 dx_1 &= \int_0^1 \frac{1}{3}x_1^2dx_1 \\
&= \frac{1}{3}[\frac{1}{3}x_1^3]_0^1 = \frac{1}{9} \\
\Rightarrow E(X_1X_2) &= \frac{1}{9}
\end{aligned}
$$

So, the covariance between the X_1 and X_2 is:

$$
\begin{aligned}
Cov(X_1,X_2) &= E(X_1X_2) - E(X_1)E(X_2) \\
&= \frac{1}{9} - \frac{1}{12} \times \frac{1}{6} = \frac{7}{72} = 0.0972
\end{aligned}
$$

Reading 16: Sample Moments

Q.325 Compute the sample standard deviation given the following sample data: $\sum x = 31,353$ $n = 100$ $\sum x^2 = 10,687,041$

A. 86

B. 93

C. 71

D. 75

The correct answer is **B.**

The formula for calculation of sample variance is:

$$s^2 = \frac{1}{(n-1)}\left[\sum_{i=1}^{n} x^2 - n\bar{x}^2\right]$$

where:

$\sum x$ = sum of x

n = sample size

$\sum x^2$ =sum of x squared.

From the data,sample mean is given by:

$$\bar{x} = \frac{31353}{100} = 313.53$$
$$\Rightarrow s^2 = \frac{1}{(99)}[10,687,041 - 100 \times 313.53^2] = 8655.9082$$

The sample standard deviation is given by:

$$= \sqrt{s^2} = \sqrt{8655.9082} = 93.0371 \approx 93$$

Q.326 On Tuesday, an insurance company receives a total of 10 claims for automobile policies. After the first-round assessment, it's found that the mean claim amount of the 10 claims is $426 while the standard deviation is 112. On Tuesday, the chief claims analyst authorizes the removal of one of the claims for $545 from the list on grounds that it's fraught with fraud. Compute the standard deviation for the remaining set of 9 claims.

A. 110.2

B. 12145.2

C. 421.8

D. 420

The correct answer is **A.**

$\sum x$ = the total of the original set of 10 claims $= 426 \times 10 = 4260$

After removing the claim worth $545, the new value of $\sum x = 4260 – 545 = 3,715$

Thus, the new mean $= \frac{3715}{9} = \$412.8$

Since $s^2 = \frac{1}{(n-1)}[\sum x^2 - n\bar{x}^2]$, using data for all the ten claims,

$$112^2 = \frac{1}{9}[\sum x^2 - 10 \times 426^2]$$

Making $\sum x^2$ the subject of the formula,

$$\sum x^2 = 112^2 \times 9 + 10 \times 426^2 = 1,927,656$$

Removing the fraudulent claim gives,

$$\sum x^2 = 1,927,656 - 545^2 = 1,630,631$$

Now, using the data for the remaining 8 claims,

$$S^2 = \frac{1}{(9-1)}[1,630,631 - 9 \times 412.8^2] = 12,145.2$$

Therefore,

$$s = \sqrt{12,145.2} = 110.2$$

Q.328 A renowned economist has calculated that the Canadian economy will be in one of 3 possible states in the coming year: Boom, Normal, or Slow. The following table gives the returns of stocks A and B under each economic state.

State	Probability State	Return for Stock A	Return for Stock B
Boom	40%	12%	18%
Normal	35%	10%	15%
Slow	25%	8%	12%

Which of the following is closest to the covariance of the returns for stocks A and B?

A. 0.103

B. 0.0001734

C. 0.1545

D. 0.0003765

The correct answer is **D.**

$$\text{Cov}(A,B) = \sum P(s) * [R_A - E(R_A)] * [R_B - E(R_B)]$$

First, you have to determine the expected return for every stock:

$$E(R_A) = 0.4 * 0.12 + 0.35 * 0.1 + 0.25 * 0.08 = 0.103$$
$$E(R_B) = 0.4 * 0.18 + 0.35 * 0.15 + 0.25 * 0.12 = 0.1545$$

State	PS	RA	RB	P(S) * [RA − E(RA)] * [RB − E(RB)]
Boom	0.40	0.12	0.18	0.4 * [0.12–0.103] * [0.18–0.1545] = 0.0001734
Normal	0.35	0.1	0.15	0.35 * [0.1–0.103] * [0.15–0.1545] = 0.000004725
Slow	0.25	0.08	0.12	0.25 * [0.08–0.103] * [0.12–0.1545] = 0.0001984
				Cov(A, B) = 0.0001734 + 0.000004725 + 0.0001984 = 0.0003765

Q.331 The mean height of female FRM exam candidates over the years is 1.671m while that of male candidates is 1.757. Given that the mean height of ALL of the exam candidates is 1.712m, calculate the percentage of the candidates who are female:

A. 0.523

B. 0.46

C. 0.087

D. 0.477

The correct answer is **A.**

Let F be the proportion of females.

Applying the idea of a weighted mean,

$$
\begin{aligned}
1.671F + 1.757(1-F) &= 1.712 \\
\Rightarrow 1.671F + 1.757 - 1.757F &= 1.712 \\
\Rightarrow 1.757 - 1.712 &= 1.757F - 1.671F \\
\Rightarrow 0.045 &= 0.086F \\
\Rightarrow F &= 0.523 \text{ or } 52.3\%
\end{aligned}
$$

Q.339 Which of the following best describes the concept of an unbiased estimator?

A. One for which the accuracy of the parameter estimate increases as the sample size increases.

B. One that has the least variance compared to all other estimators.

C. One for which the accuracy of the parameter estimate increases as the sample size decreases.

D. One for which the expected value of the estimator is equal to the value of the parameter being estimated.

The correct answer is **D.**

If $\bar{x}$ is an unbiased estimator of μ then the expected value of $\bar{x}$ is equal to μ.

Q.370 At a certain investment firm, each of the firm's 5 managers is tasked with overseeing a project. During a given one-year period, the managers reported the following individual returns from their projects:
[24%, 26%, 30%, 18%, 20%]

Calculate the population variance of these returns.

A. 0.1824%

B. 18.24%

C. 22.8%

D. 0.228%

The correct answer is **A.**

Note that the data given is comprised of the entire population and NOT a sample. As such, we should use the formula for calculating the population variance.

We know that $\sigma^2 = \frac{[\sum(X_i - \mu)^2]}{N}$ where N is the size of the population

And $\mu = \frac{(\sum(X_i))}{N} = \frac{(0.24+0.26+0.30+0.18+0.20)}{5} = 0.236$

$$\sigma^2 = \frac{[(0.24 - 0.236)^2 + (0.26 - 0.236)^2 + (0.30 - 0.236)^2 + (0.18 - 0.236)^2 + (0.20 - 0.236)^2]}{5}$$
$$= \frac{(0.000016 + 0.000576 + 0.004096 + 0.003136 + 0.001296)}{5}$$
$$= 0.001824 \text{ or } 0.1824\%$$

Note: Had we been given sample data, the formula for the mean would remain unchanged but when calculating the variance, we would divide the sum of squared deviations by (n – 1) to remove bias.

Q.3258 The returns generated by a sample of five stocks from the Karachi Stock Exchange are given in the exhibit below.

Stock	Return
A	12%
B	13%
C	5%
D	4%
E	20%

What is the standard deviation of this sample?

A. 6.53%.

B. 5.84%.

C. 7.53%.

D. 8.34%

The correct answer is **A.**

$$\text{Mean} = \frac{(0.12 + 0.13 + 0.05 + 0.04 + 0.2)}{5} = 0.108$$

Stock	Return	$X - \text{Mean}$	$(X - \text{Mean})^2$
A	12%	1.2%	0.000144
B	13%	2.2%	0.000484
C	5%	−5.8%	0.003364
D	4%	−6.9%	0.004624
E	20%	9.2%	0.008464
Total			0.017080

$$\text{Sample deviation} = \left(\frac{1}{(n-1)}(X - \text{Mean})^2\right)^{\frac{1}{2}}$$

$$= \left(\frac{0.017080}{4}\right)^{\frac{1}{2}} = 6.53\%$$

Note: The standard deviation calculated with a divisor of $n - 1$ is a standard deviation calculated from the sample as an estimate of the standard deviation of the population from which the sample was drawn.

Q.3261 Which of the following statements is the *most* accurate? Lognormal Distributions are:

A. Skewed to the left and rarely used to model asset prices.

B. Skewed to the left and often used to model asset prices.

C. Skewed to the right and often used to model asset prices.

D. Skewed to the right and rarely used to model asset prices.

The correct answer is **C.**

Lognormal distributions are skewed to the right because they are bound by zero to the left. For this reason, they are often used to model stock prices since the prices cannot take negative values.

Q.3279 What is the correlation of returns between these two portfolios?

- Portfolio A's variance of returns: 52.5%
- Portfolio B's variance of returns: 63%
- Covariance of return between the two portfolios: 0.315

A. 0.8257

B. 0.0011

C. 0.5477

D. 0.9524

The correct answer is **C.**

$$\begin{aligned}\text{Corr}(R_A, R_B) &= \frac{\text{Cov}(R_A, R_B)}{\sigma_A \sigma_B} \\ &= \frac{0.315}{(\sqrt{0.525} * \sqrt{0.63})} \\ &= 0.5477\end{aligned}$$

Note: We had to transform the variance into standard deviation by using $\sqrt{0.525}$ and $\sqrt{0.63}$.

Q.3762 The following are the data on the financial analysis of a sales company's income over the last 200 months:

$$n = 200, \sum_{i=1}^{n} (x_i - \hat{\mu})^2 = 774,759.90 \text{ and } \sum_{i=1}^{n} (x_i - \hat{\mu})^3 = -13,476.784$$

What is the value of skewness?

A. --0.0002795

B. -0.00051738

C. -0.00031736

D. -0.00021733

The correct answer is **A.**

The skewness is given by:

$$\frac{\hat{\mu}^3}{\hat{\sigma}^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^3}{[\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2]^{\frac{3}{2}}} = \frac{\frac{1}{200}(-13,476.784)}{[\frac{1}{200} \times 774,759.90]^{\frac{3}{2}}} = -0.0002795$$

The value of the skewness is slightly negative and not far from 0, implying the data is symmetrical.

Q.3763 A sample of 100 monthly profits gave out the following data:

$$\sum_{i=1}^{100} x_i = 3,453 \text{ and } \sum_{i=1}^{100} x_i^2 = 800,536$$

What is the sample mean and standard deviation of the monthly profits?

A. Sample Mean=33.53, Standard deviation=85.99

B. Sample Mean=34.53, Standard deviation=82.96

C. Sample Mean=43.53, Standard deviation=89.99

D. Sample Mean=33.63, Standard deviation=65.99

The correct answer is **B.**

Recall that the sample mean is given by:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$\Rightarrow \bar{X} = \frac{1}{100} \times 3,453 = 34.53$$

The variance is given by:

$$s^2 = \frac{1}{n-1}\{\sum_{i=1}^{n} X_i^2 - n\hat{\mu}^2\} = \frac{1}{99}(800,536 - 100 \times 34.53^2) = 6881.857677$$

So that the standard deviation is given to be:

$$s = \sqrt{6881.857677} = 82.96$$

Q.3764 The following contains the amounts of monthly profits for a certain company for the year 2019:

$$1576, 1595, 1754, 1464, 1850, 1698, 1614, 1524, 4320, 1650, 1440, 1602$$

What is the mean?

A. 1840.58

B. 2007.91

C. 1564.33

D. 1785.44

The correct answer is **A.**

We know the sample mean is given by:

$$\begin{aligned} \hat{\mu} = \bar{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ \Rightarrow \bar{X} &= \frac{1}{12}(1576 + 1595 + 1754 + 1464 + 1850 + 1698 + 1614 + 1524 + 4320 \\ &\quad + 1650 + 1440 + 1602) \\ &= 1840.5833 \end{aligned}$$

Q.3766 The following data represents a sample of daily profit of a sales company for six weeks in a particular year.

Week	Amount of the Profit($)
1	3,800
2	2,800
3	2,700
4	9,900
5	2,600
6	4,300

What is the median of the profit?

A. 3,000

B. 5,000

C. 3,300

D. 3,700

The correct answer is **C.**

The first step would be to sort the data as follows:

Week	Amount of the Profit($)
5	2,600
3	2,700
2	2,800
1	3,800
6	4,300
4	9,900

Since the number of observations is even, the median is estimated by averaging the two consecutive central points. That is:

$$\text{Med}(x) = \frac{1}{2}[x_{\frac{6}{2}} + x_{\frac{6}{2}+1}] = \frac{1}{2}[x_3 + x_4] = \frac{1}{2}[2800 + 3800] = 3,300$$

Q.3767 The following data represents a sample of daily profit of a sales company for six weeks in a particular year.

Week	Amount of the Profit($)
1	3,800
2	2,800
3	2,700
4	9,900
5	2,600
6	4,300

What is the 25% quantile of the profits?

A. 2750

B. 2,700

C. 2,650

D. 2,725

The correct answer is **C.**

The first step would be to sort the data as follows:

Week	Amount of the Profit($)
5	2,600
3	2,700
2	2,800
1	3,800
6	4,300
4	9,900

We need to estimate the α-quantile using the data point in location $\alpha \times n$, i.e., $0.25 \times 6 = 1.5$

Since the number of observations is even, the lower quartile is the average of ordered data points 1 and 2.

$$q_{25} = 0.5(2,600 + 2,700) = 2,650$$

Q.3768 The following data represents a sample of daily profit of a sales company for six weeks in a particular year.

Week	Amount of the Profit($)
1	3,800
2	2,800
3	2,700
4	9,900
5	2,600
6	4,300

What is the 75% quantile profit?

A. 4,175

B. 4,234

C. 4,050

D. 4,654

The correct answer is **C.**

The first step would be to sort the data as follows:

Week	Amount of the Profit($)
5	2,600
3	2,700
2	2,800
1	3,800
6	4,300
4	9,900

We need to estimate the α-quantile using the data point in location $\alpha \times n$ i.e., $0.75 \times 6 = 4.5$

When $\alpha \times n$ is not an integer value, then the usual practice is to take the average of the points immediately above and below $\alpha \times n$. In this case, we ought to find the average of the the 4th and 5th ordered observations.

$$q_{25} = 0.5(3,800 + 4,300) = 4,050$$

Q.3769 The following data represents a sample of daily profit of a sales company for six weeks in a particular year.

Week	Amount of the Profit($)
1	3,800
2	2,800
3	2,700
4	9,900
5	2,600
6	4,300

What is the interquartile range?

A. 1,600

B. 1,542

C. 1,475

D. 1,400

The correct answer is **D.**

Recall that:

$$IQR = q_3 - q_1$$

The first step would be to sort the data as follows:

Week	Amount of the Profit($)
5	2,600
3	2,700
2	2,800
1	3,800
6	4,300
4	9,900

We estimate the α-quantile using the data point in location alpha $\times$ n.

To find q_1 we have to take, $0.25 * 6 = 1.5$, since this is not an integer, then we will find the average of the 1st and 2nd values in the ordered list. This gives,

$$q_1 = \frac{2600 + 2700}{2} = 2650$$

Similarly, to find q_3, we take $0.75 * 6 = 4.5$, which mean, we have to find the average of 4th and 5th values in the ordered list, that is

$$q_3 = \frac{3800 + 4300}{2} = 4050$$

Therefore, the interquartile range is given by:

$$IQR = 4050 - 2650 = 1400$$

Q.3770 What are the conventional values of skewness and kurtosis of a normal random variable?

A. Skewness=0, kurtosis=3

B. Skewness=1, kurtosis=3

C. Skewness=0, kurtosis=2

D. Skewness=0, kurtosis=4

The correct answer is **A.**

The skewness of a standard normal random variable is symmetrical, and thus it has a skewness of 0. The value kurtosis is 3 which is a reference point for other random variables. A kurtosis of 3 implies that the distributions lack heavy tails i.e., the random variable are distributed equally around the mean.

Q.3771 A sample amount of profit for a certain company for the first 15 weeks of the year is given below:

Weeks	Amount of the Profit($)
1	0
2	7,000
3	13,000
4	13,000
5	20,000
6	23,000
7	25,000
8	27,000
9	34,000
10	41,000
11	60,000
12	66,000
13	76,000
14	77,000
15	96,000

What is the difference between the biased and an unbiased estimator of the sample mean?

A. 0

B. 1.0

C. 3.2

D. 4.6

The correct answer is **A.**

This question does not require any calculation whatsoever since the sample always mean unbiased. So, the difference is just by itself, which is zero.

Q.3772 A sample amount of profit for a certain company for the first 15 weeks of the year is given below:

Weeks	Amount of the Profit($)
1	0
2	70
3	130
4	130
5	200
6	230
7	250
8	270
9	340
10	410
11	600
12	660
13	760
14	770
15	960

What is the difference between the biased and unbiased estimators of the variance?

A. 5,867.50

B. 5,503.90

C. 5,767.49

D. 5,867.87

The correct answer is **C.**

Study the prefilled table below

Weeks	Amount of the Profit($) = X_i	$(X_i - \hat{\mu})^2$
1	0	148481.778
2	70	99435.111
3	130	65195.111
4	130	65195.111
5	200	34348.444
6	230	24128.444
7	250	18315.111
8	270	13301.778
9	340	2055.111
10	410	608.444
11	600	46081.778
12	660	75441.778
13	760	140375.111
14	770	147968.444
15	960	330241.778
	385.3333	1211173.333

We know that biased estimator of the sample variance is given by:

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

Therefore, we need to calculate the sample mean of the data:

$$\bar{X} = \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \$385.33$$

So that,

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 = \frac{1}{15} \times 1211173.333 = 80744.89$$

Recall that the unbiased estimator of the sample variance is given by:

$$s^2 = \frac{n}{n-1}\hat{\sigma}^2 = \frac{15}{14} \times 80744.89 = 86512.38$$

So, the difference is given by:

$$86512.38 - 80744.89 = 5,767.49$$

Q.3773 The following are the two series of data X and Y.

X	Y
700	318
130	304
140	317
200	305
230	309
250	307
270	316
340	309
400	315
620	327
620	450
760	324
650	500
750	699

Assuming the Central limit theorem, what is the correct representation of the sample means distributions of the random variables above?

A.

$$\begin{bmatrix} 432.86 \\ 264.29 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} 4058.71 & 1106.37 \\ 3106.37 & 914.81 \end{bmatrix}\right)$$

B.

$$\begin{bmatrix} 432.86 \\ 464.29 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} 4058.71 & 1106.37 \\ 2106.37 & 914.81 \end{bmatrix}\right)$$

C.

$$\begin{bmatrix} 432.86 \\ 364.29 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} 4058.71 & 1106.37 \\ 1106.37 & 914.81 \end{bmatrix}\right)$$

D.

$$\begin{bmatrix} 432.86 \\ 324.29 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} 4458.71 & 1146.37 \\ 1406.37 & 914.81 \end{bmatrix}\right)$$

The correct answer is **C.**

Recall that according to CLT for bivariate data, the vector of means is usually distributed. That is:

$$\begin{bmatrix} \hat{\mu}_x \\ \hat{\mu}_y \end{bmatrix} \rightarrow N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \frac{\sigma_X^2}{n} & \frac{\sigma_{XY}}{n} \\ \frac{\sigma_{XY}}{n} & \frac{\sigma_Y^2}{n} \end{bmatrix}\right)$$

This implies that we need to calculate the mean and variance of each series and their covariance. Study the following table:

X	Y	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
700	318	71365.3061	2142.36735	−12364.9
130	304	91722.449	3634.36735	18257.96
140	317	85765.3061	2235.93878	13847.96
200	305	54222.449	3514.79592	13805.1
230	309	41151.0204	3056.5102	11215.1
250	307	33436.7347	3281.65306	10475.1
270	316	26522.449	2331.5102	7863.673
340	309	8622.44898	3056.5102	5133.673
400	315	1079.59184	2429.08163	1619.388
620	327	35022.449	1390.22449	−6977.76
620	450	35022.449	7346.93878	16040.82
760	324	107022.449	1622.93878	−13179.2
650	500	47151.0204	18418.3673	29469.39
750	699	100579.592	112033.653	106152.2

Starting with the mean, we have:

$$\hat{\mu}_X = \bar{X}_X = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{6060}{14} = 432.8571$$

$$\hat{\mu}_Y = \bar{X}_Y = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{5100}{14} = 364.2857$$

We go to the variances (we shall use the unbiased estimator)

$$\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{738685.714}{13} = 56821.978$$

$$\hat{\sigma}_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \frac{166494.857}{13} = 12807.2967$$

The last part involves computing the sample covariance given by:

$$\text{Cov}(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \frac{201358.60}{13} = 15489.12$$

We can now apply the CLT:

$$\begin{bmatrix} 432.86 \\ 364.29 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \frac{56821.978}{14} & \frac{15489.12}{14} \\ \frac{15489.12}{14} & \frac{12807.2967}{14} \end{bmatrix}\right)$$

$$\begin{bmatrix} 432.86 \\ 364.29 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} 4058.71 & 1106.37 \\ 1106.37 & 914.81 \end{bmatrix}\right)$$

Q.3774 What is the difference between the skewness and the kurtosis?

A. Skewness measures the tendency of a larger amount of values moving in a given direction or with respect to the mean while the kurtosis is the peakedness of a distribution with respect to the normal distribution.

B. Skewness ranges from 0 and 1, while kurtosis ranges from 0 and 3.

C. Skewness is always positive relative to the kurtosis.

D. None of the above

The correct answer is **A.**

Skewness measures the tendency of a larger amount of values moving in a given direction or with respect to the mean, while kurtosis measures the peakedness of a distribution with respect to the normal distribution.

Option B is incorrect since its skewness can take any value; Same as Option C.

Q.3775 Which of the following statements describes the Central Limit Theorem (CLT)?

A. If $X_1, X_2, \ldots, X_n$ is a sequence of iid random variables with a finite mean μ and finite non-zero variance σ^2 then the distribution of $\frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}}$ tends to a standard normal distribution as $n \to \infty$.

B. If $X_1, X_2, \ldots, X_n$ is a sequence of iid random variables with a finite mean μ and finite non-zero variance σ^2 then the distribution of $\frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}}$ tends to a standard normal distribution as $n \to 0$.

C. that If $X_1, X_2, \ldots, X_n$ is a sequence of iid random variables with a finite mean μ and finite non-zero variance σ^2 then the distribution of $\frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}}$ tends to a standard normal distribution as $n \to 1$.

D. None of the above

The correct answer is **A.**

The Central Limit Theorem (CLT) states that if $X_1, X_2, \ldots, X_n$ is a sequence of iid random variables with a finite mean μ and finite non-zero variance σ^2 then the distribution of $\frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}}$ tends to a standard normal distribution as $n \to \infty$.

Put simply,

$$\frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1)$$

Note that $\hat{\mu} = \bar{X}$ = Sample Mean

Q.3776 All of the following statements are true, except:

A. If the return distributions of two investments have the same mean and standard deviation, the one which has a more negatively skewed distribution will be considered to be riskier.

B. Distributions with positive excess kurtosis are known as leptokurtic.

C. If two investments have the same mean, standard deviation and skewness, then the one with the lower kurtosis will be considered riskier.

D. The normal distribution has a kurtosis of 3.

The correct answer is **C.**

If two investments have the same return distributions in terms of mean and standard deviation and skewness, then the one with a higher kurtosis will be considered riskier as excess kurtosis implies more extreme points, which results in more risk. (The possibility of experiencing extreme and financially devastating losses is higher)

Option A is a true statement. The positive skewness of a distribution indicates that an investor may expect frequent small losses and a few large gains . This tilts the investment toward upside risk. On the other hand, the negative skewness of a distribution indicates that an investor may expect frequent small gains and a few large losses which tilt the investment more toward downside risk. As a result, the probability of loss increases with negative skewness, and that of gain increases with positive skewness.

Option B is a true statement. The presence of positive excess kurtosis indicates a leptokurtic distribution. Such a distribution has fatter tails than a normal distribution on either side, increasing the probability of larger extreme outcomes (i.e., more risk). When excess kurtosis is negative, the variable has a platykurtic distribution.

Option D is a true statement. A standard normal distribution has kurtosis of 3 and is recognized as mesokurtic.

Q.3777 An analyst gathers monthly data about the returns of a stock for the past five years. If the mean monthly return is 6% and the standard deviation of the series of returns is 1.8%, then what is the standard deviation of the mean over the period?

A. 6.24%

B. 0.23%

C. 13.94%

D. 4.02%

The correct answer is **B.**

The standard deviation of the sample mean is given by:

$$\frac{\sigma}{\sqrt{n}} = \frac{1.8\%}{\sqrt{5 \times 12}} = 0.2324\%$$

Note that we multiplied n by 12 since we are dealing with monthly data.

Q.3778 Assume we have equally invested in two different companies; ABC and XYZ. We anticipate that there is a 15% chance that next year's stock returns for ABC Corp will be 6%, a 60% probability that they will be 8% and a 25% probability that they will be 10%. In addition, we already know the expected value of returns is 8.2%, and the standard deviation is 1.249%. We also anticipate that the same probabilities and states are associated with a 4% return for XYZ Corp, a 5% return, and a 5.5% return. The expected value of returns is then 4.975, and the standard deviation is 0.46%. Calculate the portfolio standard deviation:

A. 0.0000561

B. 0

C. 0.00849

D. 0.00897

The correct answer is **C.**

The portfolio variance is given by:

$$\text{Portfolio variance} = W_A^2 \times \sigma^2(R_A) + W_B^2 \times \sigma^2(R_B) + 2 \times (W_A) \times (W_B) \times \text{Cov}(R_A, R_B)$$

First, we must calculate the covariance between the two stocks. Now using

$$\begin{aligned}
\sigma_{R_iR_j} &= \sum_{i=1}^{n} P(R_i)[R_i - E(R_i)][R_j - E(R_j)] \\
\text{cov}(ABC, XYZ) &= 0.15(0.06\text{–}0.082)(0.04\text{–}0.04975) \\
&+ 0.6(0.08\text{–}0.082)(0.05\text{–}0.04975) \\
&+ 0.25(0.10\text{–}0.082)(0.055\text{–}0.04975) \\
&= 0.0000555
\end{aligned}$$

Since we already have the weight and the standard deviation of each asset, we can proceed and calculate the portfolio variance:

$$\begin{aligned}
&= 0.5^2 \times 0.01249^2 + 0.5^2 \times 0.0046^2 + 2 \times 0.5 \times 0.5 \times 0.0000555 \\
&= 0.000072040
\end{aligned}$$

Therefore, the standard deviation is $(0.000072040)^{\frac{1}{2}} = 0.00849$

Q.3779 A sample of 36 working days was analyzed; for the amount of income of a company. If the income has a standard deviation of 7, what is the approximate probability that the mean of this sample is greater than 44.50 and that the mean of the yearly income (population) is $\mu=42$?

A. 0.045

B. 0.016

C. 0.065

D. 0.042

The correct answer is **B.**

According to the Central Limit Theorem,

$$\frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0,1)$$

Note that $\hat{\mu} = \bar{X}$ = Sample Mean

We need:

$$P(\bar{X} > 44.5) = P\left[\frac{\hat{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{44.5-42}{\frac{7}{\sqrt{36}}}\right]$$
$$\Rightarrow P(\bar{X} > 44.5) \approx P(Z > 2.143) = 1 - \varphi(2.143) = 1 - 0.9840 = 0.016$$

Q.3797 A portfolio is composed of 60% equities and 40% bonds. The variance of equities is 320, the variance of bonds is 110, and the covariance is 90. What is the portfolio's variance?

A. 154.4

B. 176

C. 192

D. 279.2

The correct answer is **B.**

$$\begin{aligned}\text{Portfolio variance} &= w_A^2 * s^2(R_A) + w_B^2 * s^2(R_B) + 2 * (w_A) * (w_B) * \mathrm{Cov}(R_A, R_B)\\ &= (0.6)^2 * (320) + (0.4)^2 * (110) + 2 * (0.6) * (0.4) * (90) = 115.2 + 17.6 + 43.2\\ &= 176\end{aligned}$$

Q.3798 Alan West, a portfolio manager, created the following portfolio:

Security	Security Weight (%)	Expected Standard deviation(%)
A	20	4
B	80	10

If the correlation of returns between the two securities is 0.60, then what is the expected standard deviation of the portfolio?

A. 9.50%

B. 8.10%

C. 9.15%.

D. 8.50%.

The correct answer is **D.**

$$\begin{aligned}\text{Portfolio standard deviation} &= [(0.2)^2(4\%)^2 + (0.8)^2(10\%)^2 + 2(0.2)(0.8)(0.6)(4\%)(10\%)]^{0.5}\\ &= 8.50\%\end{aligned}$$

Here, we're simply using the portfolio variance formula and raising it to the 0.5 power to get the portfolio standard deviation.

Q.3799 Raul Perez, a portfolio manager, created the following portfolio:

Security	Security Weight (%)	Expected Standard deviation(%)
A	40	7
B	60	12

If the covariance of returns between the two securities is -0.004, then what is the expected standard deviation of the portfolio?

A. 6.36%

B. 6.56%

C. 8.14%

D. 6.10%

The correct answer is **A.**

$$\begin{aligned} \sigma^2{}_P &= w^2{}_1\sigma^2{}_1 + w^2{}_2\sigma^2{}_2 + 2w_1w_2\text{Cov}(R_1, R_2) \\ &= [(0.4)^2(7\%)^2 + (0.6)^2(12\%)^2 + 2(0.4)(0.6)(-0.004)] \\ &= 0.004048 \\ \sigma_P &= 0.004048^{0.5} = 0.063624 \end{aligned}$$

Q.3800 Tina Fer, a portfolio manager, created the following portfolio:

Security	Security Weight (%)	Expected Standard deviation(%)
A	10	6
B	90	15

If the standard deviation of the portfolio is 14.1%, then what is the covariance between the two securities?

A. 0.009

B. 0.090

C. 0.008

D. -0.011

The correct answer is **A.**

We know that:

$$Var(A,B) = w_A^2\sigma_A^2 + w_B^2\sigma_B^2 + 2W_AW_BCov(A,B)$$

Where
w_A^2 = weight of security A.
w_B^2 = weight of security B.
σ_A^2 = variance of security A
σ_B^2 = variance of security B
Cov(A, B) = covariance between securities A and B.
Plugging in the values given in the table, we have:

$$(0.141)^2 = 0.10^2 \times 0.06^2 + 0.90^2 \times 0.15^2 + 2 \times 0.10 \times 0.90 \times Cov(A,B)$$
$$\Rightarrow Cov(A,B) = \frac{(0.141)^2 - (0.10^2 \times 0.06^2 + 0.90^2 \times 0.15^2)}{2 \times 0.10 \times 0.90} = 0.009$$

Q.3801 Carla Mayes, a portfolio manager created the following portfolio:

Security	Expected Return (%)	Expected Standard Deviation(%)
A	5	8
B	10	14

If the correlation of returns between the two securities is -0.20, then what is the standard deviation of a portfolio invested 75% in Security A and 25% in Security B?

A. 0.51%

B. 0.81%

C. 5.12%

D. 6.31%

The correct answer is **D.**

We know that:

$$\begin{aligned}\sigma_P^2 &= w_A^2\sigma_A^2 + w_B^2\sigma_B^2 + 2w_Aw_B\rho_{AB}\sigma_A\sigma_B \\ &= (0.75)^2(0.08)^2 + (0.25)^2(0.14)^2 + 2(0.75)(0.25)(-0.20)(0.08)(0.14) \\ &= 0.003985 = 0.3985\%\end{aligned}$$

The standard deviation of the portfolio = 6.3120%

Q.3802 Which of the following statements is *most* accurate regarding a negative covariance between two assets?

A. The returns of two assets move together in the negative direction.

B. The returns of two assets move in opposite directions.

C. The returns of two assets move in the positive direction.

D. None of the above.

The correct answer is **B.**

Positive covariance indicates that two variables tend to move in the same direction. Negative covariance indicates that two variables tend to move in inverse (opposite) directions

---

Q.3803 Which of the following properties of covariance is INCORRECT?

A. Covariance measures how one random variable moves with another random variable.

B. Covariance of (R,R) = Variance of R

C. Covariance ranges from -1 to +1.

D. None of the above.

The correct answer is **C.**

Covariance may range from negative infinity to positive infinity, whereas CORRELATION ranges from -1 to +1. Options A) and B) are appropriate properties of covariance.

Q.3804 Which of the following is the most appropriate explanation of a -1 correlation between two random variables?

A. There is no correlation between two random variables.

B. Both variables move together in a negative direction.

C. The movement in one variable will result in the exact opposite proportional movement in the other variable.

D. None of the above.

The correct answer is **C.**

A correlation of -1 means that the movement in one variable will result in the exact opposite proportional movement in the other variable.
Note: A correlation of +1 means both variables move in the same direction in the same proportion.
A correlation of 0 means there is no linear relationship between the two random variables.

Q.3805 Assuming that the covariance of returns of Stock X and Stock Y is $Cov(R_X, R_Y) = 0.093$, the variance of $R_X = 0.69$, and the variance of $R_Y = 0.36$, what is the correlation of returns of Stock X and Stock Y?

A. 0.155

B. 0.1865

C. 0.1713

D. 0.1119

The correct answer is **B.**

$Corr (R_X, R_Y) = Cov\ R_X,R_Y / \sigma(R_X)\ \sigma(R_Y)$

Since Variance = σ^2,
$\sigma(R_X) = \sqrt{0.69} = 0.8306$, and
$\sigma(R_Y) = \sqrt{0.36} = 0.6$

$Corr\ R_X,R_Y = Cov\ R_X,R_Y / \sigma(R_X)\ \sigma(R_Y) = 0.093/0.8306 * 0.6 = 0.1865$

Q.3806 Which of the following is an INCORRECT interpretation of the correlation coefficient?

A. A correlation coefficient of +1 means that the mean returns of two assets move proportionately in the same direction.

B. A correlation coefficient of -1 means that the mean returns of two assets move proportionately in a negative direction.

C. A correlation coefficient of 0 means that the mean returns of two assets move proportionately in a negative direction.

D. None of the above.

The correct answer is **C.**

A positive correlation of +1 means that the returns move in the same direction in exactly the same proportions. In contrast, a perfect negative correlation of -1 suggests that the deviation in the returns of two assets is inversely proportionate.

Q.3807 A junior fund manager at Dapper Assets Management is constructing a portfolio consisting of two large-cap stocks that trade on the London stock exchange. In a meeting with the investment committee, the manager was asked to present the covariance of both stocks. Using the data given in the following table, calculate the covariance if the population mean is unknown.

Year	Stock A Return	Stock B Return
1	17%	45%
2	21%	20%
3	-8%	-2%
4	-1%	2%
5	4%	-19%
6	19%	2%
7	-7%	13%

A. 0.0113

B. 0.1010

C. 0.1156

D. 0.00907

The correct answer is **A.**

Below are the calculations:

$$\bar{X}_A = \frac{0.17 + 0.21 - 0.08 - 0.01 - 0.04 + 0.19 - 0.07}{7} = 0.0642$$

$$\bar{X}_B = \frac{0.45 + 0.20 - 0.02 + 0.02 - 0.19 + 0.02 - 0.13}{7} = 0.0871$$

Stock A Return Mean Return of A	Stock B Return minus Mean Return of B	Covariance = (Stock A Return minus Mean Return of A) × (Stock B Return minus Mean Return of B)
0.1057	0.3629	0.0384
0.1457	0.1129	0.0164
-0.1443	-0.1071	0.0155
-0.0743	-0.0671	0.0050
-0.0243	-0.2771	0.0067
0.1257	-0.0671	-0.0084
-0.1343	0.0429	-0.0058

$$\begin{aligned} \text{Cov} &= \frac{0.0384 + 0.0164 + 0.0155 + 0.0050 + 0.0067 - 0.0084 - 0.0058}{(n-1)} \\ &= \frac{0.0384 + 0.0164 + 0.0155 + 0.0050 + 0.0067 - 0.0084 - 0.0058}{6} \\ &= 0.0113 \end{aligned}$$

The reason we use "n-1" and not "n" is essentially that the population mean E(X) is not known and is replaced by the sample mean x-bar.

Note: You can also do the problem with the help of the financial calculator by using the STAT function and using the sample standard deviation.

Q.3808 A fund manager is constructing a portfolio consisting of two stocks. Which of the following equations can the manager use to calculate the correlation coefficient if the covariance is 0.0168, the standards deviation of stock A is 0.125 and the standard deviation of stock B is 0.2?

A. $0.0168/(0.125*0.2)^2$

B. 0.0168/(0.125*0.2)

C. $0.0168/(0.125*0.2)^{1/2}$

D. $0.0168/(0.125^2*0.2^2)$

The correct answer is **B.**

Correlation = Covariance/(Standard deviation A * Standard deviation B) = 0.0168/(0.125 * 0.2)

Q.3809 Hakim Ahmed has recently joined Lampard Investment Inc. He was given the data related to the assets of a portfolio provided in the following table. If the weight of Asset X is 35% and the weight of Asset Z is 65%, then what is the variance of the portfolio?

Variance Asset X	0.1225
Variance Asset Z	0.4225
Covariance	0.19

A. 0.2800

B. 0.1156

C. 0.2245

D. 0.2587

The correct answer is **A.**

$$\begin{aligned}\text{Portfolio Variance} &= w_X^2\sigma_X^2 + w_Z^2\sigma_Z^2 + 2 * w_X w_Z * \text{Cov}(XZ)\\ &= 0.35^2 * 0.1225 + 0.65^2 * 0.4225 + 2 * 0.35 * 0.65 * 0.19\\ &= 0.2800\end{aligned}$$

Q.3810 Hakim Ahmed has recently joined Lampard Investment Inc. He has been given data related to the assets of a client's portfolio provided in the following table:

Variance Asset X	0.1225
Variance Asset Z	0.3721
Covariance	0.19

If the weight of Asset X is 35% and the weight of Asset Z is 65%, then what is the correlation coefficient between Assets X and Z?

A. 0.8899

B. 0.0469

C. 0.4412

D. 0.4168

The correct answer is **A.**

tandard deviation of X = $0.1225^{1/2}$ = 0.35
Standard deviation of Z = $0.3721^{1/2}$ = 0.61
Correlation coefficient = Covariance(X,Z)/(Standard deviation of X * Standard deviation of Z) = 0.19/(0.35 * 0.61) = 0.8899

Note: The correlation coefficient does not take into account the weights.

Q.3811 The covariance matrix of two stocks is given in the following exhibit.
Exhibit: Covariance Matrix

Stock	X	Y
X	650	120
Y	120	450

What is the correlation of returns for stocks X and Y?

A. 0.45

B. 0.22

C. 0.28

D. 0.37

The correct answer is **B.**

$\sigma(X) = (650)^{0.5} = 25.50$
$\sigma(Y) = (450)^{0.5} = 21.21$

Covariance(X,Y) = 120

Correlation(X, Y) = $Cov(R_X,R_Y) / \sigma(R_X)\ \sigma(R_Y)$ = (120) / (25.50* 21.21) = 0.22

Q.3812 A portfolio consists of two funds A and B. The weights of the two funds in the portfolio and the covariance matrix of the two funds are given in the following two exhibits.
Exhibit 1: Weight of the Funds in the Portfolio

Fund	A	B
Weight	60%	40%

Exhibit 2: Covariance Matrix

Fund	A	B
A	700	200
B	200	500

What is the portfolio variance?

A. 428.04

B. 500.00

C. 512

D. 324.80

The correct answer is **A.**

Based on the covariance matrix:
$\sigma(A) = (700)^{0.5} = 26.46$

$\sigma(B) = (500)^{0.5} = 22.36$

Covariance (A,B) = 200

$\text{Variance}_{\text{Portfolio}} = \sigma_p^2 = w_1^2\ \sigma_1^2 + w_2^2\ \sigma_2^2 + 2w_1\ w_2\ \rho_{1,2}\ \sigma_1\ \sigma_2$
$= (0.60)^2 * (26.46)^2 + (0.40)^2 * (22.36)^2 + 2 * 0.60 * 0.40 * 200$
= 428.04

Q.3813 Which of the following statements is INCORRECT regarding the correlation coefficient?

A. The correlation coefficient measures the strength of the linear relationship between two random variables

B. The correlation coefficient has no units

C. The correlation coefficient ranges from -1 to +1

D. None of the above.

The correct answer is **D.**

In finance, the correlation coefficient attempts to measure the degree to which two random variables, say, returns for different stocks move in relation to each other. The correlation coefficient always lies between -1 and +1. A positive value indicates that the random variables move in the same direction, i.e., if an increase (decrease) is recorded in one variable, we expect an increase (decrease) in the other variable, which can either be proportionate or disproportionate depending on the value of the correlation.

Reading 17: Hypothesis Testing

Q.371 The mean hourly wage for coal workers in the U.S. is $15.5 with a population standard deviation of $3.2. Calculate the standard error of the sample mean if the sample size is 30.

A. 3.2

B. 0.3413

C. 0.1067

D. 0.5842

The correct answer is **D.**

Since the standard deviation for the population is known,
Standard error of the mean = $\frac{\sigma}{\sqrt{n}} = \frac{3.2}{\sqrt{30}} = 0.5842$
Interpretation: If we were to take a number of samples of size 30 from U.S. coal workers population and proceed to prepare a sampling distribution of the sample means, the distribution would have a mean of $15.5 and a standard error of $0.58.

Note: In most cases, σ is unknown in which case we work with the sample standard deviation, s.

Q.373 An auto insurance company intends to establish the mean claim amount demanded by policyholders who own SUVs. After extensive analysis of its records, the company believes the standard deviation of such claims is about $200.
The company wishes to construct a 95% confidence interval for the mean claim amount such that the interval is of width "±$50". Determine the value of *n*, the sample size that would be required to achieve this.

A. 62

B. 100

C. 30

D. 124

The correct answer is **A.**

A 95% CI for the mean μ, assuming normal distribution = $\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$

Since σ = 200, we need to find the value of *n* such that:

$$\begin{aligned} 1.96 \times \frac{\sigma}{\sqrt{n}} &= 50 \\ 1.96 \times 200 &= 50\sqrt{n} \\ 392 &= 50\sqrt{n} \\ \sqrt{n} &= 7.84 \\ n &= 62 \end{aligned}$$

Q.374 After 72 FRM Part 1 students took a mock exam, the mean score was 75. Assuming that the population standard deviation is 10, construct a 99% confidence interval for the mean score on the mock exam.

A. (75, 85)

B. (65, 75)

C. (71.96, 78.04)

D. (75, 78.04)

The correct answer is **C.**

A 99% CI for the mean, $\mu = \bar{x} \pm Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$ where $\alpha = 0.01$

From the normal dist. table, $Z_{0.005} = 2.58$

$$\mu = 75 \pm 2.58 \times \frac{10}{\sqrt{72}} = 75 \pm 3.04$$

Giving us a CI $= 71.96 \leq \mu \leq 78.04$

Q.375 Which of the following best defines the term “hypothesis” as used in statistics?

A. An assumption about a problem.

B. The current state of knowledge or belief about the value of a population parameter.

C. A test that divides the sample space into a region of acceptance and the critical region.

D. A statement about the value of a population parameter developed to test a theory or belief.

The correct answer is **D.**

A hypothesis describes an assumption about the true value of a population parameter, such as the mean. It is developed to test the validity and accuracy of a belief about the parameter. For example, suppose a risk manager intends to establish the mean monthly return on a stock option. They may formulate a hypothesis such as
“The mean monthly return on stock options is positive” or;
“The mean monthly return on stock options is greater than 10%”

Q.376 An investment firm intends to carry out a test to determine whether bonuses have any significant effect on job performance. The head of the human resource department develops the following sets of possible hypotheses.

I. H_0: Bonuses do not have any effect on job performance H_1: Bonuses improve job performance
II. H_0: Bonuses do not have any effect on job performance H_1: Bonuses reduce job performance
III. H_0: Bonuses do not have any effect on job performance H_1: Bonuses affect job performance
IV. H_0: Bonuses have no effect on job performance H_1: Bonuses improve job performance

Which of the above pairs of hypotheses implies a two-sided test?

A. I

B. II

C. III

D. IV

The correct answer is **C.**

The difference between a one-sided test and a two-sided test is that while the alternative hypothesis in the former explores the possibility of a change in only one direction (increase or decrease), the latter explores the possibility of a change in either direction. While the alternative hypothesis in each of sets I, II, and IV explore increases or decreases, the word "affect" in the H_1 of set III leaves open the possibility of either an increase or a decrease in job performance.

Q.377 Which of the following choices correctly defines type I error, type II error, and p-value?

	Type I error	Type II error	P-value
I.	The probability of rejecting H_0 when it is in fact true	The probability of accepting H_0 when it is in fact false	The highest level at which H_0 can be rejected
II.	The probability of accepting H_0 when it is in fact false	The probability of rejecting H_0 when it is in fact true	The highest level at which H_0 can be rejected
III.	The probability of rejecting H_0 when it is in fact true	The probability of accepting H_0 when it is in fact false	The lowest level at which H_0 can be rejected
IV.	The probability of rejecting H_0 when it is in fact true	The lowest level at which H_0 can be rejected	The probability of accepting H_0 when it is in fact false

A. I

B. II

C. III

D. IV

The correct answer is **C.**

Let's borrow a little from the justice system: Suppose a judge sends a murder suspect to jail when the suspect is, in fact, innocent; that would be a type I error. On the other hand, suppose the same judge frees the suspect when they are, in fact, guilty; that's a type II error. In both cases the presumption (H_0) would be "innocent".

Similarly, let's say we've got $H_0 : \mu = 0$ and $H_1 : \mu \neq 0$. Rejecting H_0 when μ is in fact equal to zero constitutes a type I error. Failing to reject H_0 when μ is in fact not equal to zero (either greater than or less than zero) constitutes a type II error.

Lastly, the probability value (p-value) is the probability, assuming H_0 is true, of observing a test statistic at least as extreme as the value observed.

Q.378 A random sample of 50 FRM exam candidates was found to have an average IQ of 125. The standard deviation among candidates is known (approximately 20). Assuming that IQs follow a normal distribution, carry out a statistical test (5% significance level) to determine whether the average IQ of FRM candidates is greater than 120. Compute the test statistic and give a conclusion.

A. Test statistic: 1.768; Reject H_0

B. Test statistic: 2.828; Reject H_0

C. Test statistic: 1.768; Fail to reject H_0

D. Test statistic: 1.0606; Fail to reject H_0

The correct answer is **A.**

The first step: Formulate H_0 and H_1

$H_0 : \mu = 120$

$H_1 : \mu > 120$

Note that this is a one-sided test because H_1 explores a change in one direction only

Under H_0, $\frac{(\bar{x}-120)}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N(0,1)$

Next, compute the test statistic:

$$= \frac{(125-120)}{\left(\frac{20}{\sqrt{50}}\right)} = 1.768$$

Next, we can confirm that $P(Z > 1.6449) = 0.05$, which means our critical value is the upper 5% point of the normal distribution i.e. 1.6449. Since 1.768 is greater than 1.6449, it lies in the rejection region. As such, we have sufficient evidence to reject H_0 and conclude that the average IQ of FRM candidates is indeed greater than 120.

Alternatively, we could go the "p-value way"

$$P(Z > 1.768) = 1 - P(Z < 1.768) = 1 - 0.96147 = 0.03853 \text{ or } 3.853\%$$

This probability is less than 5% meaning that we have sufficient evidence against H_0. This approach leads to a similar conclusion.

Q.379 A stock has an initial market price of $80. Exactly one year from now, its price will be given by:
P = 80 * exp(i) where *i* is the rate of return.

i is normally distributed with mean 0.2 and standard deviation 0.3. Construct a 95% confidence interval for the price of the stock after one year.

A. ($54.27, $80)

B. ($80, $175.91)

C. ($54.27, $175.91)

D. ($54.27, $140)

The correct answer is **C.**

A 95% CI for the price $= \mu \pm Z_{\frac{\alpha}{2}} \times \sigma$ where $\alpha = 0.05$

$$\begin{aligned} &= \mu \pm 1.96 \times 0.3 \\ &= \mu \pm 0.588 \end{aligned}$$

Therefore,

$$\begin{aligned} P &= 80 * \exp(0.2 \pm 0.588) \\ &= (\$54.27, \$175.91) \end{aligned}$$

Note: “exp” stands for “exponential”

Note:

When considering the confidence interval for the price of an asset, we only consider the assets expected price, its standard deviation, and critical values from the normal distribution, i.e.

CI = Expected price +/- critical value * standard deviation.

Q.380 A manager conducts a hypothesis test at the 1% significance level. What does this mean?

A. P(reject H_0 | H_0 is false) = 0.01

B. P(reject H_0 | H_0 is true) = 0.01

C. P(not reject H_0 | H_0 is false) = 0.01

D. P(not reject H_0 | H_0 is true) = 0.01

The correct answer is **B.**

The significance level represents the probability of committing a type I error or rejecting a correct model. This is equivalent to P(reject H_0 | H_0 is true). Choice C is actually the probability of committing a type II error.

Q.381 Suppose you conducted a hypothesis test. What would happen if you decrease the level of significance of the test?

A. The likelihood of committing a type II error decreases

B. The likelihood of a type I error increases

C. The likelihood of rejecting the null hypothesis when it's in fact true decreases

D. The likelihood of frequently committing a type I error increases, even when it's in fact true

The correct answer is **C.**

Having seen that the significance level gives the probability of rejecting a true null hypothesis, it follows that a decrease in α (the level of significance) effectively decreases this probability. That means a decrease of α from, say, 5% to 1%, would mean less frequent rejection of a true null hypothesis.

Q.383 At a 95% confidence interval, the value at risk (VaR) of a portfolio is approximately $10 million. During 100 days, the VaR was exceeded on 9 different occasions. Based on this information:

A. This model is overestimating risk.

B. This model is underestimating risk.

C. This model is appropriate for estimating the risk.

D. The model is accurate.

The correct answer is **B.**

The 95% CI means that we expect to have losses exceeding $10 million just 5% of the time. Thus, during a 100-day period, we would expect a maximum of 5 exceedances. That 9 of these were recorded means the model is underestimating risk and is therefore NOT appropriate.

Q.384 A population has a known mean of 100. Suppose 36 samples are randomly drawn from this population with replacement. The observed mean is 97.8 and the standard deviation is 10. Calculate the standard error of the sample mean.

A. 0.8165

B. 1.667

C. 1.8165

D. 16.67

The correct answer is **B.**

The standard error of the sample mean = $\frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{36}} = 1.6667$

Q.3280 Which of the following statement(s) are accurate?
I. The t-distribution is similar but not identical to the normal distribution in shape. It has fatter tails compared to the normal distribution.
II. Degrees of freedom for the t-distribution are equal to n. Students' t-distribution is closer to a normal distribution when the degrees of freedom are lower.

A. Statement I is correct and Statement II is incorrect.

B. Statement I and Statement II are both correct.

C. Statement I and Statement II are both are incorrect.

D. Statement I is incorrect and Statement II is correct.

The correct answer is **A.**

Statement I is correct. The t-distribution is similar, but not identical to the normal distribution in shape. It has fatter tails compared to the normal distribution.
Statement 2 is incorrect. Degrees of freedom for the t-distribution are equal to n - 1. A student's t-distribution tends to the normal distribution as the degrees of freedom increase.

Q.3281 A sample of 121 applicants received the Canadian travel visa in 45 days on average. Suppose the population is normally distributed, and the standard deviation of the sample is 19, then what is the 95% confidence interval for the population mean?

A. [44.7 days; 45.3 days]

B. [41.6 days; 48.4 days]

C. [42.2 days; 47.8 days]

D. [40.1 days; 49.8 days]

The correct answer is **B.**

Before calculating the confidence interval, we will calculate the Standard error of the sample mean:
Standard error of the sample mean = Standard deviation of the sample mean/Sample size = $19/\sqrt{121}$ = 1.73
The z-static at the 95% confidence interval is 1.96.

The confidence interval is calculated as Mean +/- Reliability factor * Standard error
Lower limit of the confidence interval = 45 - (1.96 * 1.73) = 41.60
Upper limit of the confidence interval = 45 + (1.96 * 1.73) = 48.40

Q.3282 A sample of 100 students is currently renting rooms in the mean distance of 18 miles from a small U.S. College. Assuming that the population is normally distributed and the standard deviation of the sample is 14 miles, what is the 99% confidence interval for the population mean?

A. [15.26 miles; 20.74 miles]

B. [16.6 miles; 19.4 miles]

C. [14.4 miles; 21.6 miles]

D. [12.8 miles; 23.6 miles]

The correct answer is **C.**

$$\text{Standard error of the sample} = \frac{\text{Standard deviation of sample mean}}{\sqrt{\text{Sample size}}} = \frac{14}{\sqrt{100}} = 1.4$$

Z-static (Reliability factor) at the 99% confidence interval = 2.58

Lower limit of the confidence interval = $18 - (2.58 * 1.4) = 14.39$ miles

Upper limit of the confidence interval = $18 + (2.58 * 1.4) = 21.61$ miles

---

Q.3283 The mean return of a sample of 28 BB+ corporate bonds is 7.5%, and the sample's standard deviation is 14%. Assuming that the population is normally distributed and the population variance is unknown, what is the 95% confidence interval for the population mean?

A. [2.77%; 12.23%]

B. [2.07%; 12.93%]

C. [2.93%; 11.43%]

D. [4.12%; 13.3%]

The correct answer is **B.**

Since the population variance is unknown and the population is normally distributed, and the sample size is less than 30, we will use a t-statistic. The t-statistic for a 95% confidence interval and 27 degrees of freedom (df=n-1) is 2.052.

$$\text{The standard error of the sample} = \frac{\text{Standard Deviation of sample mean}}{\sqrt{\text{Sample size}}}$$
$$= \frac{14}{\sqrt{28}} = 2.646$$

The confidence interval is $7.5 - (2.052 * 2.646) = 2.07$ and $7.5 + (2.052 * 2.646) = 12.93$

Using a reliability factor based on the t-distribution is essential for a small sample size. Using a t reliability factor is appropriate when the population variance is unknown, and when the sample size is less or equal to 30.

Below is a t distribution table:

t distribution table

Q.3284 Which of the following is INCORRECT regarding the t-statistic?

A. As the degree of freedom increases, the t-statistic decreases.

B. A 90% confidence interval with n-1 degrees of freedom will be calculated at $\frac{\alpha}{2}$ *or* $t_{0.05}$.

C. The t-statistic is used for sample sizes smaller than 30 observations.

D. The t-statistic has thinner tails than the normal distribution.

The correct answer is ***D.***

The t-statistic has fatter tails than the normal distribution as it has more probability in its tails. Options A, B and C are accurate. As the degree of freedom increases, the t-statistic decreases. The t-statistic is used for sample sizes smaller than 30 observations, as compared to the z-statistic which is used for sample sizes of more than 30 observations. A 90% confidence interval with n-1 degrees of freedom will be calculated at $\frac{\alpha}{2}$ *or* $t_{0.05}$.

Q.3285 From a population of 10,000 observations, a researcher chooses a sample of 1,000. If the population's standard deviation is 100, then what is the standard error of the mean?

A. 0.1

B. 3.16

C. 27.6

D. 0.01

The correct answer is ***B.***

$$\begin{aligned}\text{Standard error of the sample mean} &= \frac{\text{Population standard deviation}}{\sqrt{\text{Sample size}}}\\ &= \frac{100}{\sqrt{1,000}}\\ &= \frac{100}{31.6228}\\ &= 3.16228\end{aligned}$$

Q.3287 Which of the following best represents a 99 percent confidence interval if the mean score from 40 students in an exam is 85 and the population's standard deviation is 18?

A. [77.658; 92.342]

B. [73.658; 92.342]

C. [77.658; 90.342]

D. [80.212; 93.526]

The correct answer is ***A.***

The mean is 85, the standard deviation is 18, and the sample size is 40.

The 99% z-value is 2.58.

$$\text{Confidence interval} = \bar{x} \pm z * \left(\frac{\sigma}{\sqrt{n}}\right)$$

$$= 85 \pm 2.58 \left(\frac{18}{\sqrt{40}}\right) = [77.658; 92.342]$$

Q.3288 While performing a hypothesis test, Albert Khan is told that his analysis suffers from a Type I error. We can therefore conclude that:

A. Khan rejected the null hypothesis when it was actually false.

B. Khan failed to reject the null hypothesis when it was actually false.

C. Khan failed to reject the null hypothesis when it was actually true.

D. Khan rejected the null hypothesis when it was actually true.

The correct answer is ***D.***

A Type I error means that Khan rejected the null hypothesis when it was actually true.

Q.3289 A two-tailed hypothesis test at the 95% confidence level has a p-value of 2.14%. This means that:

A. At a 5% significance level, we cannot reject the null hypothesis.

B. At a 2% significance level, we can reject the null hypothesis.

C. At a 5% significance level, we can reject the null hypothesis.

D. We have insufficient information to provide any kind of conclusion.

*The correct answer is **C.***

The p-value is the probability of obtaining results at least as extreme as the observed results of a hypothesis test, assuming that the null hypothesis is correct. In other words, the p-value is the lowest level at which we can reject the null hypothesis.

At 95% significance level, any p-value less than 5%(1-95%) will lead to rejection of the null hypothesis. Hence, in this case, a p-value of 2.14% will lead to rejection of the null hypothesis

---

Q.3290 Which of the following statement(s) is/are correct?

I. *Decreasing the significance level will decrease the probability of rejecting a true null*
II. *Increasing the significance level will increase the power of the test.*

A. Both statements are correct.

B. Only Statement I is correct.

C. Only Statement II is correct.

D. Both statements are incorrect.

The correct answer is ***A.***

Statement I is correct*: Increasing the significance level increase the probability of rejecting a true null hypothesis i.e committing a Type I error. Decreasing the significance level will therefore decrease the probability of rejecting a true null hypothesis. The probability of making Type I error and Type II error is therefore affected by the level of significance. Lower significance levels reduce the frequency of Type I errors, while large significance levels reduce the frequency of Type II errors.*

Statement II is correct*: Large significance levels use small critical values that are close to zero and which have larger rejection regions than lower significance levels. Smaller critical values increase the probability that the sample mean falls within the rejection region., thus increasing the power of the test.*

Q.3291 A survey is conducted to determine if the average starting salary of investment bankers is equal to or greater than $57,000 per year. Given a sample of 115 newly employed investment bankers with a mean starting salary of $65,000 and a standard deviation of $4,500, and assuming a normal distribution, what is the test statistic?

A. 204.44

B. 19.06

C. 1.78.

D. 746

The correct answer is ***B.***

$$\begin{aligned}\text{Standard error of the sample mean} &= \frac{\text{Standard deviation}}{\sqrt{\text{Sample size}}}\\ &= \frac{\$4,500}{\sqrt{115}}\\ &= 419.6272\\ \text{Test statistic} &= \frac{(\text{Sample mean} - \text{Hypothesized value})}{\text{Standard error of the sample mean}}\\ &= \frac{(65,000 - 57,000)}{(419.6272)} = 19.06\end{aligned}$$

Q.3292 Hilda believes that the average return on equity in the consumer durables industry is greater than 80%. What are the null (H_0) *and the alternative* (H_1) *hypotheses for this study?*

A. $H_0: M = 0.8$ versus $H_1: M \neq 0.8$

B. $H_0: M \geq 0.8$ versus $H_1: M < 0.8$

C. $H_0: M \leq 0.8$ versus $H_1: M > 0.8$

D. $H_0: M < 0.8$ versus $H_1: M \geq 0.8$

The correct answer is ***C.***

This is a one-sided alternative (so we cannot use =) because of the "greater than" belief. As always, we expect to reject the null hypothesis.

Q.3293 The average return on the Dow Jones Industrial Average for 121 quarterly observations is 1.5%. If the standard deviation of the returns can be assumed to be 8%, what is the 99% confidence interval for the quarterly returns of the Dow Jones?

A. [-0.4%; 3.4%]

B. [0.1%; 2.9%]

C. [-6.5%; 9.5%]

D. [-0.1%; 2.9%]

The correct answer is ***A.***

$$\begin{aligned}\text{Standard error of the sample mean} &= \frac{\text{Standard deviation}}{\sqrt{\text{Sample size}}} \\ &= \frac{8\%}{\sqrt{121}} \\ &= 0.00727\end{aligned}$$

The critical value for the z-statistic for a 99% confidence interval is 2.575.

$$\text{Confidence Interval} = 1.5\% \pm 2.575[0.00727] = [-0.37\%; 3.37\%]$$

Q.3295 If a researcher wants to test that the mean return of 50 small-cap stocks from the Singapore Exchange is greater than 14%, which of the following would be the alternative hypothesis for the test?

A. $H_1 : \mu \neq 14\%$

B. $H_1 : \mu > 14\%$

C. $H_1 : \mu < 14\%$

D. $H_0 : \mu < 14\%$

The correct answer is ***B.***

Since the researcher wants to test that if the mean of 50 small-cap stocks is greater than 14%, the null hypothesis is $H_0 : \mu \leq 14\%$ *and the alternative hypothesis is* $H_1 : \mu > 14\%$.

We always want to reject the null hypothesis and accept the alternative. Since the researcher wants to prove that the mean returns are greater than 14%, $H_1 : \mu > 14\%$.

Q.3298 A portfolio manager observes that the weekly return generated by a portfolio of high-beta stocks stood at 5%. The standard deviation of the portfolio return stood at 1.50%. However, the manager observes that the standard deviation of the portfolio return for the past 15 weeks stood at 2.00%. The portfolio manager wants to determine whether the standard deviation of the portfolio return has increased from 1.50% to 2.00%.
What is the test statistic to test the above hypothesis?

A. 34.89

B. 25.89

C. 31.55

D. 24.89

*The correct answer is **D.***

The chi-square test is used for hypothesis tests regarding population variance.

$$\text{Test statistic} = \frac{[(n-1)*s^2]}{\sigma^2}$$
$$= \frac{[(15-1)*0.02^2]}{1.5\%^2} = 24.89$$

Q.3299 A portfolio manager believes that returns on pharmaceutical stocks are more volatile than the returns generated on e-commerce stocks. To check this hypothesis, the portfolio manager collects the data summarized in exhibit 1.

Exhibit 1: Volatility in Pharmaceutical vs. e-Commerce Stocks

	Pharma Stock	e-Commerce Stocks
Standard Deviation	1.50%	2.10%
Sample Size	20	25

What is the value of the test statistic?

A. 1.51

B. 1.96

C. 1.70

D. 2.14

The correct answer is ***B.***

As the test requires testing the equality of variances of two populations, the appropriate test is the F-test.

$$\text{Test statistic} = \frac{(\text{std. dev. ecommerce})^2}{(\text{std. dev. pharma})^2}$$
$$= \frac{(2.10\%)^2}{(1.50\%)^2}$$
$$= 1.96$$

Note: A convention, or usual practice, is to use the larger of the two standard deviations on top (in the numerator). When we follow this convention, the value of the test statistic is always greater than or equal to 1; tables of critical values of F then need to include only values greater than or equal to 1. Under this convention, the rejection point for any formulation of hypotheses is a single value on the right-hand side of the relevant F-distribution. However, even without following this convention, we would still arrive at the same conclusion (on whether or not to reject the null).

Q.3300 An investor is planning to invest in mutual funds. He intends to maximize his chances of earning a return in excess of 20%. The list of mutual funds available to the investor is listed in *exhibit 1.*

Exhibit 1: Potential List of Mutual Funds

Fund	Mean Return	Std. Dev. of Return
X	15%	2.%
Y	15.20%	3%
Z	14%	4%

Assuming that mutual fund returns are normally distributed and using a z-table, what is the correct probability of earning a return in excess of 20%?

A. *1.60% for Fund Y.*

B. *5.78% for Fund Z.*

C. *0.62% for Fund X.*

D. *11.6% for Fund Y.*

The correct answer is ***C.***

The first step is calculating the z-values:

Fund	Mean Return	Std. Dev. of Return	Z-value
X	15%	2.%	$= (20\% - 15\%)/2\% = 2.5$
Y	15.20%	3%	$= (20\% - 15.20\%)/3\% = 1.60$
Z	14%	4%	$= (20\% - 14\%)/4\% = 1.50$

The z-table provides the cumulative probability for each of the fund:

Fund	Z-value	Cumulative Probability of Return > 20%
X	2.50	$= 100\% - 99.38\% = 0.62\%$
Y	1.60	$= 100\% - 94.52\% = 5.48\%$
Z	1.50	$= 100\% - 93.32\% = 6.68\%$

Standard Normal Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999							

Q.3330 For a sample of the past 28 monthly stock returns for Bidco Inc., the mean return is 5% and the sample standard deviation is 15%. Assume that the population variance is unknown.
The related t*-table values are given below, where* (t_{ij}) *denotes the* $(100-j)^{th}$ *percentile of* t*-distribution value with* i *degrees of freedom):*

$t_{27,0.025}$	2.05
$t_{27,0.05}$	1.70
$t_{26,0.025}$	2.06
$t_{26,0.05}$	1.71

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

A. [0.00181, 0.0989]

B. [-0.0084, 0.1084]

C. [-0.00811, 0.10811]

D. [0.02135, 0.07835]

*The correct answer is **C.***

$$\text{CI} = \text{Mean return} \pm (\text{Reliability factor} \times \text{Standard error})$$

Here, we use the t*-reliability factor since the population variance is unknown. Since there are 28 observations, the degrees of freedom are* $28-1=27$*. The* t*-test is a two-tailed test, and confidence intervals are two-sided. So the correct critical* t*-value is* $t_{27,0.025}=2.05$

$$\text{CI} = 0.05 \pm 2.05 \times \frac{0.15}{\sqrt{28}} = 0.05 \pm 0.05811 = [-0.00811, 0.10811]$$

Q.3331 Using returns observed over the past 18 monthly, an analyst has estimated the mean monthly return of stock A *to be 2.85% with a standard deviation of 1.6%.*

One-tailed t-distribution table

Degrees of Freedom		a	
	0.1	0.05	0.025
14	1.35	1.78	2.15
15	1.34	1.75	2.13
16	1.34	1.75	2.12
17	1.33	1.74	2.11
18	1.33	1.73	2.10

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

A. *[0.02031, 0.03688]*

B. *[0.02051, 0.03650]*

C. *[0.02194, 0.03506]*

D. *[0.02054, 0.03646]*

The correct answer is ***D.***

$$\text{CI} = \text{Mean return} \pm (\text{Reliability factor} \times \text{Standard error})$$

Here, we use the t*-reliability factor since the population variance is unknown. Since there are 18 observations, the degrees of freedom are* $18 - 1 = 17$.
This t*-test is two-tailed and confidence intervals are two-sided, implying that the 0.025 column must be used. So the correct critical* t*-value is* $t_{17,0.025} = 2.11$

$$\text{CI} = 0.0285 \pm 2.11 \times \frac{0.016}{\sqrt{18}} = 0.0285 \pm 0.00796 = [0.02054, 0.03646]$$

Q.3332 A risk analyst wishes to establish the VaR of a hedge fund. She has gathered return data spanning 24 weeks. From her analysis, the mean and standard deviation of weekly returns are 10% and 12%, respectively. Assuming that weekly returns are independent and identically distributed, what is the standard error of the mean of the weekly returns?

A. 10%

B. 2%

C. 2.45%

D. 12%

The correct answer is ***C.***

The standard error of the ***mean*** *measures how far the sample mean of the data is likely to be from the true* ***population mean****. It should not be confused with the standard deviation which measures the variability of a set of data around the mean.*

$$\text{Standard error of the mean} = \frac{\text{Standard deviation of weekly returns}}{\text{Square root of sample size}}$$

$$\frac{s}{\sqrt{n}} = \frac{0.12}{\sqrt{24}}$$

$$= \frac{0.12}{4.89898} = 0.02449$$

Reading 18: Linear Regression

Q.386 The relationship between two variables can be explained by the following regression function:
$Y_i = B_0 + B_1 \times X_i + \varepsilon_i$

What does ε_i represent?

A. The difference between total variation and the unexplained variation.

B. The effects of independent variables not included in the model.

C. The slope coefficient.

D. The intercept coefficient.

The correct answer is ***B.***

ε_i denotes the effects of independent variables other than the variable of interest to the researcher, which nonetheless impact the dependent variable. For example, assume you want to assess the effect of class size (independent variable) on student performance (dependent variable): the error term might comprise factors such as teacher quality, student economic background, or even luck.

Q.387 What is the difference between regression analysis and correlation analysis?

A. Regression enables us to measure the association between two or more variables in specified units of the dependent variable.

B. Regression enables us to establish a line of best fit through the data.

C. With regression, we can predict the values of the dependent variable.

D. All of the above.

The correct answer is ***D.***

While correlation merely informs us of the existence of a linear relationship, regression analysis goes further to enable us to predict distinctive values of the dependent variable given distinctive values of the independent variable. The regression line is very useful in estimation and prediction.

Q.388 A hospital uses ultrasound technology to measure the weight of unborn babies as follows:

Gestation period in weeks	30	32	34	36	38	40
Estimated weight of foetus	1.6	1.7	2.5	2.8	3.2	3.5

Further information: $S_{XX} = 70, S_{YY} = 3.015, S_{XY} = 14.3$ *Calculate the least square estimator of the slope and the Y-intercept (in that order).*

A. Least square estimator: 0.2043; Y-intercept: -4.6

B. Least square estimator: 0.20; Y-intercept: -4

C. Least square estimator: 2.55; Y-intercept: 35

D. Least square estimator: 0.2043; Y-intercept: 35

The correct answer is ***A.***

Under OLS estimation,
The regression equation is of the form: $y_i = \alpha + \beta x_i$
Where Y *is the dependent variable (foetal weight),* X *is the independent variable (gestation period),* α *= the y-intercept, and* beta *= the slope.*

$$\beta = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{(x_i - \bar{x})^2} = \frac{S_{XY}}{S_{XX}} = \frac{14.3}{70} = 0.2043$$

$$\alpha = \bar{y} - \beta\bar{x}$$

$$\bar{y} = \frac{(1.6 + 1.7 + 2.5 + 2.8 + 3.2 + 3.5)}{6} = 2.55$$

$$\bar{x} = \frac{(30 + 32 + 34 + 36 + 38 + 40)}{6} = 35$$

Therefore,

$$\alpha = 2.55 - 0.2043 \times 35 = -4.60$$

Q.389 Given a regression equation is of the form: $y_i = \alpha + \beta x_i$ *where* Y *is the dependent variable (foetal weight),* X *is the independent variable (gestation period, in weeks),* α *= the y-intercept, and* β *= the slope.*
If $\alpha = -4.60$ *and* $\beta = 0.2043$*, then estimate the weight of the foetus at exactly 31 weeks.*

A. 1.533kg

B. 1.733kg

C. 1.722kg

D. 1.8kg

The correct answer is ***B.***

The regression equation is of the form: $y_i = \alpha + \beta x_i$

*Therefore, weight = -4.6 + 0.2043*31 = 1.733kgs*

Q.391 A limited liability company uses the ordinary least squares method to estimate a linear relationship between total monthly revenue and the total promotional expenditure. The linear function is found to have a positive slope that's significantly different from zero. Assuming that other variables, like product price and supply region, remain constant during the period covered by the data set, this implies that:

A. The company should significantly increase promotional expenditure.

B. The company should significantly reduce promotional expenditure.

C. Promotional expenditures have no effect on demand.

D. Promotional expenditures have a significant influence on demand.

The correct answer is ***D.***

The slope indicates the change noted in the dependent variable (sales) per unit change in the independent variable (promotional expenditure). Since the slope is positive, it confirms what economic theory might suggest: promotional events have an influence on demand for a product. The decision to increase promotional expenditure should be reviewed taking into consideration all the other business aspects.

Q.392 Ordinary least squares is used to estimate the relationship between foetal weight and the number of weeks of gestation in a group of women. The exercise gives the following results:

- *Total sum of squares, $ST_{TT} = 3.125$*
- *Regression sum of squares, $SS_{REG} = 2.925$*
- *Residual sum of squares, $SS_{RES} = 0.2$*

This implies that:

A. The variability explained by the model is 0.2

B. The variability unexplained by the model is 3.125

C. The variability explained by the model is 2.925

D. The variability explained by the model is 3.125

*The correct answer is **C.***

The total variation (total sum of squares) in responses is given by:

$$\text{Syy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

While some of this variation could be attributed to the relationship with the independent variable (x), some of it is beyond that and cannot be modeled. The "attributable" variation is said to be explained by the model and is the regression sum of squares. The unexplained variation is the residual sum of squares.

Q.394 The annual returns of two stocks, X and Y , are jointly normally distributed. Each stock has a marginal distribution with mean = 2%, standard deviation = 10%. The correlation coefficient between X and Y is 0.9. If the annual return on stock X is 4%, what is the expected annual return on stock Y ?

A. 0.038

B. 0.029

C. 0.0038

D. 0.4

*The correct answer is **A.***

We can use the information given above to construct a regression model of Y *on* X.

$$\mathrm{R_Y} = 2\% + 0.9\left(\frac{10}{10}\right) * (\mathrm{R_X} - 2\%) + \epsilon$$

Replacing $\mathrm{R_X}$ *with 4% gives:*

$$\mathrm{R_Y} = 2\% + 0.9\left(\frac{10}{10}\right) * (4\% - 2\%) = 3.8\%$$

Note: The intercept is simply the marginal distribution of each stock.

Remember that by definition, the marginal distribution gives the probabilities of various values of the variables in the subset without reference to the values of the other variables.

Alternative Approach

This is a classic example of exam-style questions that test your understanding of concepts not the mere application of formulas.

We have two stocks whose performance appears to be linearly related. With a positive correlation of 0.9, the returns of the two stocks move in the same direction; when the return of stock 1, the return of stock 2 also increases, and the opposite is true. This begs the question, can we try to model this as a simple linear regression? Yes

The general format of simple regression is:

$$\mathrm{Y} = \mathrm{a} + \mathrm{bX}$$

Where Y and X are our variables, a is the intercept, and b is the slope

What we are trying to model here is that the return on Y has two components:

(I) a = intercept, a component that independent of X's performance, i.,e, the return earned by Y even if X earns a zero return. That's effectively the marginal distribution

(II) bX = a component that depends on the performance of X multiplied by a factor b

If we know the values of a and b, we can easily get the expected annual return on stock Y.

Recall that in linear regression,

$$\begin{aligned}
a &= Y_{mean} - b(X_{mean}) \\
Y_{mean} &= X_{mean} = 2\% \\
b &= \text{slope} = \text{beta (Y with respect to X)} = \frac{\text{cov(X,Y)}}{\text{var(X)}} \\
&= \text{correlation(X,Y)} * \text{standard dev(X)} * \frac{\text{standard dev(Y)}}{\text{variance(X)}} \\
&= \text{correlation(X,Y)} * \frac{\text{standard dev(Y)}}{\text{standard dev(X)}} \\
&= 0.9 * \frac{10\%}{10\%} = 0.9
\end{aligned}$$

Thus, $a = 2\% - 0.9(2\%) = 0.2\%$

With these, we can comfortably predict the expected return on Y

$$Y = a + bX = 0.2\% + 0.9(4\%) = 3.8\%$$

Q.395 If the value of the independent variable = 0, what would be the expected value of the dependent variable?

A. The slope coefficient

B. The residual value

C. The intercept coefficient

D. 0

The correct answer is ***C.***

$$E(Y|X) = b_0 + b_1 * X$$

Therefore, $E(Y|0) = b_0$ *which is the y-intercept/intercept coefficient*

Q.396 A graduate school constructs a linear regression model to estimate the effect of increased study hours on the average performance of students in a test. The slope coefficient is found to be equal to 2. What does this mean?

A. The average score when the number of study hours is zero is 2.

B. The predicted score when the number of study hours is zero is 2%.

C. For every one unit change in the number of study hours, the model predicts that the average score will change by 2 units.

D. For every one unit change in the number of study hours, the model predicts that the average score will change by 2%.

*The correct answer is **C.***

The slope coefficient represents the expected change in the dependent variable for a 1-unit change in the independent variable. If the coefficient is 2 and the independent variable changes by 1 unit, the dependent variable will change by 2 units. Candidates usually have difficulties differentiating between the slope coefficient and the intercept coefficient. The latter is simply the value of the dependent variable (Y) when the independent variable (X) is equal to zero.

Q.397 The estimated slope coefficient (β_1) *for a certain stock is 0.8823 with a standard error equal to 0.0931. Assuming that the sample had 10 observations, carry out a statistical test to determine if the slope coefficient is statistically different than zero. Quote the test statistic and the decision rule using a 5% level of significance.*

A. 9.477, reject H_0

B. 9.477, do not reject H_0

C. 2.307, reject H_0

D. 2.307, do not reject H_0

The correct answer is ***A.***

The first step entails formulating the hypothesis:

$$H_0 : \beta_1 = 0 \text{ vs } H_a : \beta_1 \neq 0$$

The test statistic for the slope coefficient takes the form: $t_{\alpha, n-2} = \frac{(\beta_1 - \beta_0)}{se(\beta_1)}$

In this case, $t_{0.025,8} = \frac{(0.8823 - 0)}{0.0931} = 9.477(\alpha = \frac{5\%}{2}$ *since this is a 2-tailed test)*

The critical 2-tailed values $(t_{0.025,8})$ *are* ± 2.306

Our test statistic lies outside the non-rejection region (-2.306, 2.306). As such, we have sufficient evidence to reject the null hypothesis and conclude that the slope coefficient is statistically different than zero.

Q.398 An analyst obtained the following linear regression relationship between 2 variables, X *and* Y *:*

$$Y = \alpha + \beta_1 X$$

where $\alpha = 0.45$ *and* $\beta = 0.8823$ *He proceeded to construct a 2-sided 95% confidence interval for the slope coefficient* (β_1) *and obtained the following interval:*

$$\beta = 0.8823 \pm 0.2147$$

Suppose the analyst decided to test the hypothesis $H_0 : \beta_1 = 1$ vs $H_a : \beta_1 \neq 1$ *at 5% significance, what would be the inference?*

- *A. Reject* H_0
- *B. Do not reject* H_0
- *C. The slope coefficient is statistically different than "1"*
- *D. Cannot tell from the information provided*

The correct answer is ***B.***

The 95% 2-sided confidence interval contains value "1". Therefore, if the analyst were to conduct a 2-sided test at the 5% level, he would end up not rejecting the null hypothesis.
Further explanation
In general, if the Y% confidence interval contains the hypothesized parameter, then a hypothesis test at the 1-Y% level of significance will always fail to reject the null hypothesis, affirming the assumption that the parameter can take on that particular value. If the interval does not contain the hypothesized parameter, then the hypothesis test at the 1-Y% level of significance will always reject the null hypothesis, discrediting the assumption that the parameter can take on that particular value.

Q.399 Two variables have a linear relationship of the form:

$$Y = -4.6 + 0.2043X$$

The slope coefficient has a standard error equal to 0.01828. Carry out a test of $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 > 0$ *at 0.5% significance, quoting the test statistic and the conclusion if the number of observations,* n*, is 6.*

A. $11.2, \beta > 0$

B. $11.2, \beta = 0$

C. $0.01828, \beta \neq 0$

D. $11.2, \beta \neq 0$

*The correct answer is **A.***

The test statistic for the slope coefficient takes the form: $t_{\alpha,n-2} = \frac{(\beta_1 - \beta_0)}{se(\beta_1)}$

In this case, $t_{0.005,4} = \frac{(0.2043-0)}{0.01828} = 11.1761$

Since the alternative hypothesis has a ">" sign, this is a one-tailed test. We should, therefore, compare our test statistic to the upper 0.5% point of the t-distribution with 4 degrees of freedom.

$T_{0.005,4} = 4.604$

*Since 11.1761 is greater than 4.604, we have **sufficient evidence** to reject* H_0 *and conclude that* $\beta_1 > 0$

Q.400 During a statistical test to determine if the mean return on an asset is different from zero, an FRM Part 1 candidate obtains a p-value of 1.4%. With a significance level of 1%, she would:

A. Reject the null hypothesis.

B. Fail to reject the null hypothesis.

C. Conclude that the mean return is different from zero.

D. Conclude that the mean return is negative (loss).

*The correct answer is **B.***

In statistical work, the p-value is the smallest level of significance at which the null hypothesis can be rejected. If the p-value = 1.4%, she would reject the null hypothesis only if the significance level were to be 1.4% or more, say, 5%. A significance level of less than 1.4% would result in the failure to reject H_0*.*

Q.402 An organization estimates that the effect of increasing the number of qualified Financial Risk Managers hired by 1 will improve the stock's annual return by 2.8% with a standard error of 0.52%. Construct a 90% 2-sided confidence interval for the size of the slope coefficient, assuming the stock's returns are normally distributed.

A. (1.9%, 2.8%)

B. (1.4%, 3.1%)

C. (1.9%, 3.5%)

D. (1.9%, 3.7%)

*The correct answer is **D.***

The CI will take the form:

$$\beta_1 \pm Z_{\frac{\alpha}{2}} \times \text{se}(\beta_1)$$
$$= \beta_1 \pm Z_{0.05} = 2.8\% \pm 1.645 \times 0.52\% = 2.8\% \pm 0.8554\% = (1.9446\%, 3.6554\%)$$

Q.404 A linear regression model gave the following results:

$$S_{yy} = 10.6; S_{xx} = 12.0; S_{xy} = 8.0; n = 18$$

Test (at 1% significance) whether β *is significantly different from zero, given that its standard error = 0.16 and give the value of the test statistic and the conclusion.*

A. 0.667, β is not significantly different from zero

B. 0.4169, β is not significantly different from zero

C. 0.667, β is significantly different from zero

D. 4.169, β is significantly different from zero

*The correct answer is **D.***

We wish to test:

$$H_0 : \beta = 0 \text{ vs } H_1 \neq 0$$

Under $H_0, \frac{(\beta-\beta_0)}{\text{se}(\beta)}$ *has a* t_{n-2} *distribution.*

$$\beta = \frac{S_{xy}}{S_{xx}} = \frac{8}{12} = 0.667$$

The observed test statistic is:

$$\frac{(0.667-0)}{0.16} = 4.169$$

This is a two-tailed test and therefore we should compare the test statistic to the upper 0.5% point of the t-distribution.

$$T_{0.005,16} = 2.921$$

$4.169 > 2.921$ *so we have sufficient evidence to reject* H_0 *at the 1% level of significance. What's the conclusion?* β *is significantly different from zero.*

Note that, we can tell whether to use one-tail or two-tail tests by considering what the question wants you to get.

A two-tailed test is used if you want to determine if there is any difference between the groups. For example, let's say you want to determine if Group X performed higher or lower than Group Y, then you should use a two-tailed test. In our case, we have been asked to determine whether β is significantly different from zero. In this case, we will use a two-tailed test.

One-tailed test, on the other hand, is used when you want to determine if there is any difference between two groups, but now in a given particular direction. Let's say you are only interested in determining if Group X performed lower than group Y, in such a case, you will need to use a one-tailed test.

Q.406 A 95% confidence interval for β_1 is determined to be (20, 25). This means that:

A. We can be 95% confident that the mean value of Y lies between 20 and 25 units.

B. We can be 95% confident that the value of X will increase by between 20 and 25 units for every one unit increase in Y.

C. We can be 95% confident that the value of Y will increase by between 20 and 25 units for every one unit increase in X.

D. At the 5% level of significance, we would not find evidence of a linear relationship between X and Y.

*The correct answer is **C.***

First, it's important to remember that β_1 represents the change in Y for every unit change in X. As such, if we are 95% confident that beta_1 *lies somewhere between 20 and 25, what we actually mean is that we are 95% confident that every unit change in X is accompanied by a change of between 20 and 25 units in Y.*

Q.408 You have been given the following regression equation:

$$\overline{WPO} = -3.2\% + 0.49(\overline{S\&P\ 500})$$

Calculate the predicted value of WPO excess returns if forecasted S&P 500 excess returns are 10%.

A. 0.017

B. 0.12

C. 0.17

D. 0.0017

*The correct answer is **A.***

You should determine the predicted value of WPO excess returns by:

$$\overline{WPO} = -3.2\% + 0.49 * 10\% = 1.7\%$$

Q.409 Use the regression equation "$\overline{WPO}$ = -3.2% + 0.49($\overline{S\&P\ 500}$) " to calculate a 95% confidence interval on the predicted value of WPO. You have been given that n = 30, the standard error of the forecast is 3.76%, and the forecasted value of S&P 500 excess return is 10%.

A. (1.7%, 9.37%)

B. (-5.97%, 1.7%)

C. (4.9%, 9.37%)

D. (-5.97%, 9.37%)

*The correct answer is **D.***

The predicted value for WPO is:

$$\overline{WPO} = -3.2\% + 0.49 * 10\% = 1.7\%$$

$$T_{0.025,30} = 2.04 (\text{Degrees of freedom} = n - 2)$$

Therefore,

$$\begin{aligned} CI_{95\%} &= \overline{WPO} \pm T_{0.025,30} * \text{Standard error of forecast} = 1.7\% \pm 2.04 * 3.76\% \\ &= 1.7\% \pm 7.67 \text{ or } (-5.97\%, 9.37\%) \end{aligned}$$

Q.410 Sometimes the explanatory power of regression analysis can be overstated. Under which of the following scenarios would that most likely happen?

A. If the residual term is normally distributed.

B. If the explanatory variables are not correlated with one another.

C. The omission of a crucial explanatory variable, which has significant influence on the explanatory variables included as well as the dependent variable..

D. If there are only two explanatory variables.

The correct answer is ***C.***

If the true regression includes an additional variable z *that has an effect on both* y *and* x*, that will be a violation of one of the assumptions of the OLS model, since the error term will not be conditionally independent of* x*. Thus, this will incorrectly increase the explanatory power of the regression.*

Additional Explanation

In the context of omitted variable bias, we have left out (excluded) some explanatory (independent) variables that are correlated with the included explanatory variables. The coefficients of the included variables that are correlated with the omitted variable will partly (depending on the correlation between them) pick up the impact of the omitted variable. As a result, these coefficients will have an inflated (higher than actual) effect on the dependent variable. The uncorrelated portion of the omitted variable's influence on the dependent variable will be captured by the error, which, again, will be higher than its true value.

Q.3333 An analyst is trying to establish the relationship between the return on stock $X(R_X)$ *and the return on stock* $S(R_s)$*. Stock* X *is listed on the Bombay Stock Exchange (BSE). The analyst has assumed a linear relationship as follows.*

$$R_X = a + b \times R_S + \epsilon_t$$

Furthermore, the analyst has gathered the following historical data.

Expected return on stock X	15%
Expected return on S	10%
Standard deviation of return on stock X	20%
Standard deviation of return on stock S	15%
Correlation between returns on stock X and S	0.3

Which of the following is the correct model that can be deduced using the ordinary least squares technique?

A. $E(R_X) = 0.40 + 0.40 \times E(R_S)$

B. $E(R_X) = 0.11 + 0.40 \times E(R_S)$

C. $E(R_X) = 0.40 + 0.11 \times E(R_S)$

D. None of the above

The correct answer is ***B.***

As we know that the expected value of the error is zero, the equation from the question is reduced to:

$$E(R_X) = a + b \times E(R_S)$$

$$b = \frac{Cov(S, X)}{Var(S)} = \frac{[Corr(S, X) * SD(S) * SD(X)]}{Var(S)} = Corr(S, X) * \frac{SD(X)}{SD(S)}$$

$$b = 0.3 * \frac{0.20}{0.15} = 0.40$$

$$a = \bar{x} - b\bar{s} = 0.15 - 0.4 * 0.10 = 0.11$$

Note: You may be used to the use of variables coded X *and* Y*, where we regress* Y *on* X *(use the values of variable* X *to predict those of* Y*, with* X *as the independent variable). In such a case, we usually formulate the regression as,*

$$Y = a + bX$$

The same logic applies here only that X *has been used as the dependent variable* $(X = a + bS)$. *The question presents a good way to test your understanding not just the application of formulas.*

Q.3335 Which of the following is/are correct regarding the assumption(s) required in OLS to draw a valid conclusion?

A. The expected value of the error term, conditional on the independent variable, $E(\epsilon_i|X_i)$*, is zero.*

B. The error term, ϵ*, is uncorrelated across observations.*

C. The error term, ϵ*, is normally distributed.*

D. All of the above

*The correct answer is **D.***

OLS regression requires three key assumptions:

I. *The expected value of the error term, conditional on the independent variable, is zero* ($E(\epsilon_i|X_i) = 0$).
II. *All* (X, Y) *observations are independent and identically distributed (i.i.d.).*
III. *There are no large outliers in the data, which would have the potential to create misleading regression results.*

Other assumptions include:

- *The error term is normally distributed;*
- *A linear relationship exists between the independent and dependent variables;*
- *The independent variable is uncorrelated with the error terms; and*
- *There are no omitted variables.*

Q.3336 The return on a stock (R) *exhibits the following relationship with the market return (*MR*).*

$$\mathrm{R} = -1.15 \times \mathrm{MR} + 2\%; \mathrm{R}^2 = 81\%$$

Assuming all coefficients are significant, which of the following interpretations is correct?

A. A 1% increase in MR results into a 1.15% increase in R.

B. A 1% increase in MR results into 2% increase in R.

C. The correlation between the return on the stock and the return on the market is 0.81.

D. The correlation between the return on the stock and the return on the market is -0.90.

The correct answer is ***D.***

When MR increases by 1%, R *changes by* $-1.15 * 1\% = -1.15\%$ *(or* R *decreases by 1.15%).*
The coefficient of determination (R^2) *measures the fraction of the total variation in the dependent variable that is explained by the independent variable. About 81% of the variation in the return on stock is explained by the independent variable (return on the market).*

The correlation coefficient is given by:

$$\mathrm{r} = [\text{Sign of estimated slope coefficient}]\sqrt{\mathrm{R}^2} = -\sqrt{0.81} = -0.9$$

Note: Given a regression $\mathrm{Y} = \mathrm{a} + \mathrm{bX}$*, the sign of* r *depends on the sign of the estimated slope coefficient* b*:*

- *If* b *is negative, then* r *takes a negative sign.*
- *If* b *is positive, then* r *takes a positive sign.*

Q.3337 The return on a stock (R) *exhibits the following relationship with the market return (MR).*

$$\mathrm{R} = -1.15 \times \mathrm{MR} + 2\%; \mathrm{R}^2 = 81\%$$

Compute the ratio of the standard deviation of stock return to standard deviation of the market return.

A. *1.15*

B. *1.28*

C. *0.81*

D. *0.90*

The correct answer is ***B.***

The coefficient of determination (R^2) *measures the fraction of the total variation in the dependent variable that is explained by the independent variable. The return on* R *explains approximately 81% of the variation from the return on MR.*

The correlation coefficient is given by:

$$\mathrm{r} = [\text{Sign of estimated slope coefficient}]\sqrt{\mathrm{R}^2} = -\sqrt{0.81} = -0.9$$

From OLS, the slope coefficient (b) *is given by:*

$$\mathrm{b} = \frac{\mathrm{Cov}(\mathrm{MR}, \mathrm{R})}{\mathrm{Var}(\mathrm{MR})} = \frac{[\mathrm{Corr}(\mathrm{MR}, \mathrm{R}) * \mathrm{SD}(\mathrm{MR}) * \mathrm{SD}(\mathrm{R})]}{\mathrm{Var}(\mathrm{MR})} = \mathrm{Corr}(\mathrm{MR}, \mathrm{R}) * \frac{\mathrm{SD}(\mathrm{R})}{\mathrm{SD}(\mathrm{MR})}$$

$$-1.15 = -0.9 * \frac{\mathrm{SD}(\mathrm{R})}{\mathrm{SD}(\mathrm{MR})}$$

$$\frac{\mathrm{SD}(\mathrm{R})}{\mathrm{SD}(\mathrm{RM})} = \frac{1.15}{0.9} = 1.28$$

Note: Given a regression $\mathrm{Y} = \mathrm{a} + \mathrm{bX}$*, the sign of* r *depends on the sign of the estimated slope coefficient* b*:*

- *If* b *is negative, then* r *takes a negative sign.*
- *If* b *is positive, then* r *takes a positive sign.*

Q.3338 An analyst has attempted to get some insight into the relationship between the return on stock A($R_{A,t}$) *and the return on the Nasdaq Composite index* ($R_{NC,t}$)*. The analyst gathers historical data and comes up with the following estimates:*

Expected mean return for A	10%
Annual mean return for Nasdaq Composite	6%
Annual volatility for Nasdaq Composite	15%
Covariance between the returns of A and Nasdaq Composite	5%

The analyst goes ahead and formulates the following regression model using the data:

$$R_{A,t} = \alpha + \beta R_{NC,t} + \epsilon_t$$

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

A. $R_{A,t} = -0.03333 + 2.2222 R_{NC,t} + \epsilon_t$

B. $R_{A,t} = -0.05 + 2.2222 R_{NC,t} + \epsilon_t$

C. $R_{A,t} = 2.2222 + 0.06 R_{NC,t} + \epsilon_t$

D. $R_{A,t} = 1.80 - 0.05 R_{NC,t} + \epsilon_t$

The correct answer is ***A.***

The regression coefficients for a model specified by $Y = \hat{a} + \hat{b}X + \varepsilon$ *, where* ε *represents the error term, are obtained using the formula:*

$$\hat{b} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{0.05}{0.15^2} = 2.2222$$
$$\hat{\alpha} = E(Y) - \hat{b}E(X) = 0.1 - 2.2222(0.06) = -0.03333$$

Q.3339 The return on a stock (R) *exhibits the following relationship with the market return (MR).*

$$R = \hat{a} + \hat{b} \times MR$$

Where $\hat{b}$ *is the slope coefficient and* $\hat{a}$ *is the intercept. After gathering 36 observations, an analyst computed the estimated slope coefficient as 0.6 with a standard error of 0.2. Compute a 95% confidence interval for the slope coefficient,* b.

A. $0.208 < b < 0.992$

B. $0.194 < b < 1.006$

C. $0.6 < b < 0.95$

D. $0.2 < b < 0.95$

The correct answer is ***A.***

$$\begin{aligned} CI &= \text{Point estimate} \pm (\text{Reliability Factor} \times \text{Standard Error}) \\ CI &= \hat{b} \pm Z_{\frac{\alpha}{2}, n-2} s_{\hat{b}} \\ CI &= 0.6 \pm 1.96 \times 0.2 = 0.6 \pm 0.392 = [0.208, 0.992] \end{aligned}$$

Note: confidence intervals are two-sided. We use the z-test because the sample size is greater than 30.

Q.3340 The return on a stock R exhibits the following relationship with the market return (MR).

$$R = \hat{a} + \hat{b} \times MR$$

Where $\hat{b}$ *is the slope coefficient and* $\hat{a}$ *is the intercept. After gathering 36 observations, an analyst computed the estimated slope coefficient as 0.6 with a standard error of 0.2. Determine whether the estimated slope coefficient is different from 0 at a 95% confidence level with reference to the critical t-value.*

A. The slope coefficient is not significant.

B. The slope coefficient is statistically significant with a t-statistic of 2.03.

C. The slope coefficient is statistically significant with a t-statistic of 3.

D. The slope coefficient is statistically significant with a t-statistic of 1.015.

The correct answer is ***C.***

Null Hypothesis: b = 0

Alternate Hypothesis: b ≠ 0

This is a two-tailed test.

$$\text{Test Statistic} = \frac{\text{Sample statistic} - \text{Hypothesized value}}{\text{Standard error of the sample statistic}}$$

$$t = \frac{(0.6 - 0)}{0.2 = 3}$$

Critical test-statistic for two-tailed test is ±2.03

We need to consider the t-statistic in a 2-tailed test.

$\alpha = 5$

$\frac{\alpha}{2} = 0.025;$

Degrees of freedom = n − 2 = 36 − 2 = 34;

$t_{0.025,34} = \pm 2.03$ from statistics table.

As the test statistic is greater than its critical value, we can reject the null hypothesis and conclude that the slope coefficient is statistically different from zero. In other words, we can say that slope coefficient is statistically significant.

Q.3341 An analyst has regressed the annual return on a stock (R_{stock}) *against the annual return on the NIFTY 50*(R_{index}) *for 30 years. The NIFTY is the National Stock Exchange (NSE) index in India.The results are as shown below. Regression equation:*

$$R_{index,\,t} = \hat{a} + \hat{b} \times R_{stock,\,t} + \varepsilon_t$$

Coefficient	Coefficient Estimate	Standard Error
a	0.002	0.001
b	1.223	0.063

Interpret whether the regression coefficients are statistically different from zero at a 95% confidence level?

A. *Intercept term (a): Yes; Slope coefficient (b): Yes*

B. *Intercept term (a): No; Slope coefficient (b): No*

C. *Intercept term (a): No; Slope coefficient (b): Yes*

D. *Intercept term (a): Yes; Slope coefficient (b): No*

The correct answer is ***C.***

The relevant hypotheses are:

$H_0 : a = 0$ vs. $H_1 : a \neq 0$, and

$H_0 : b = 0$ vs. $H_1 : b \neq 0$

Degrees of freedom $= n - 2 = 30 - 2 = 28$

Critical value for t-test = 2.048 under 2-tailed 5% significance level.

$$\text{Test Statistic} = \frac{\text{Sample Statistic} - \text{Hypothesized Value}}{\text{Standard Error of Sample Statistic}}$$

The decision rule is to reject H_0 *if test statistic > upper critical value or if test statistic < lower critical value*

$$\text{Estimated t-statistic for the intercept} = \frac{0.002 - 0}{0.001} = 2(< 2.048)$$
$$\text{Estimated t-statistic for the slope} = \frac{1.223 - 0}{0.063} = 19.413(> 2.048)$$

The intercept is not statistically significant while the slope is statistically significant at the 5% level.

Q.3342 An analyst has regressed the annual return on a stock (R_{stock}) *against the annual return on the NIFTY 50* (R_{index}) *for 36 years. The NIFTY is the index of the National Stock Exchange (NSE), India. Results are shown below. Regression equation:*

$$R_{index,\,t} = \hat{a} + \hat{b} \times R_{stock,\,t} + \varepsilon_t$$

Coefficient	Coefficient Estimate	Standard Error
a	0.002	0.001
b	1.223	0.063

What is the 90% confidence interval for the slope coefficient?

A. [1.1165; 1.3295]

B. [1.223; 1.3295]

C. [0.002; 1.223]

D. [0.063; 1.223]

The correct answer is ***A.***

$$\text{CI} = \text{point estimate} \pm (\text{reliability factor x standard error})$$

When constructing confidence intervals for the slope coefficient, we make use of the t-distribution to come up with the reliability factor.

$$\text{CI} = \hat{b} \pm t_{\frac{\alpha}{2},n-2} s_{\hat{b}}$$

alpha = 0.1, *which implies that,* $\frac{\text{alpha}}{2} = 0.05$;

Degrees of freedom $= n - 2 = 36 - 2 = 34$;

$t_{0.05,34} = \pm 1.69$ from statistical tables.

The interval of the slope coefficient is $1.223 \pm 1.69 * 0.063 = 1.1165$ *to* 1.3295.

Q.3343 An analyst has regressed the annual return on a stock (R_{stock}) *against the annual return on the NIFTY 50*(R_{index}) *for 36 years. The NIFTY is the National Stock Exchange (NSE) index in India. Results are shown below. Regression equation:*

$$R_{index,\,t} = \hat{a} + \hat{b} \times R_{stock,\,t} + \varepsilon_t$$

Coefficient	Coefficient Estimate	Standard Error
a	0.002	0.001
b	1.223	0.063

An analyst wants to test the hypothesis given below at 5% significance level: $H_0 : b \leq 1$ $H_a : b > 1$ *Which of the following statement is correct about slope coefficient?*

A. *Estimated t-statistic: 1.223; Hypothesis: Fail to reject* H_0

B. *Estimated t-statistic: 3.54; Hypothesis: Reject* H_0

C. *Estimated t-statistic: 3.54; Hypothesis: Fail to reject* H_0

D. *Estimated t-statistic: 1.223; Hypothesis: Reject* H_0

*The correct answer is **B.***

Degrees of freedom $= n - 2 = 36 - 2 = 34$
Estimated t-statistic for the slope coefficient $= \frac{1.223-1}{0.063} = 3.54$
As an upper one-tailed test, the decision rule is to reject H_0 *if test statistic > upper critical value from the t-distribution. Critical value for a one-tailed t-test at a 5% significance level = 1.69*

The estimated t-statistic (3.54) is greater than the upper 95% point of the t-distribution (1.69), so the analyst would reject the null hypothesis.

Q.3344 An analyst wishes to establish the relationship between corporate revenue (Y_t) *and the average years of experience per employee* (X_t) *and comes up with the following model.*

$$Y_t = 0.45 + 0.78X_t$$

The analyst also observes that the standard error of the coefficient of the average years of experience per employee is 0.65. In order to test the null hypothesis that the average years of experience per employee have no effect on corporate revenue, what is the correct statistic to calculate?

A. F-test

B. t-test

C. Chi-square test

D. Durbin Watson test

The correct answer is ***B.***

In both single and multiple regression, we use the t-test to determine whether a specific independent variable has a significant effect on the dependent variable at a given level of confidence. If we take the coefficient of the average years of experience per employee to be β_1*, then the test would be formulated as follows:*

$H_0 : \beta_1 = 0$
$H_0 : \beta_1 \neq 0$

In this case, the test statistic would be computed as follows:

$$t = \frac{\text{estimated value of } \beta_1 - \text{value of } \beta_1 \text{ under } H_0}{\text{standard error of estimated value of } \beta_1}$$
$$= \frac{0.78 - 0}{0.65} = 1.2$$

Q.3966 A financial analyst develops a Capital Pricing Model that regresses the expected monthly return of a company on the prevailing interest rates. The coefficients are $\beta_0 = 0.064$ *and* $\beta = 0.65$ *where* β_0 *is the intercept. What is the value of the monthly expected return for the company if the interest rate at a particular month is 5%?*

A. 0.0965

B. 0.0856

C. 0.0778

D. 0.0567

The correct answer is ***A.***

Using the stated regression parameters, the regression equation is:

$$\text{Expected Monthly Return} = \beta_0 + \beta(\text{Interest rates})$$

So, when the interest rate is 5%, then the corresponding expected return is:

$$\text{Expected Monthly Return} = 0.064 + 0.05 \times 0.65 = 0.0965 = 9.65\%$$

Q.3968 A regression analysis of monthly returns of a sales company on the market return over ten years gives an intercept of $\hat{\beta}_0 = 0.65$*, the slope* $\hat{\beta} = 1.65$*. Other quantities include:* $s^2 = 20.45$*,* $\hat{\sigma}_X^2 = 18.65$ *and* $\hat{\mu}_X = 0.61$*. What is the standard error estimate of* $\hat{\beta}_0$*?*

A. 0.5463

B. 0.56435

C. 0.4552

D. 0.4169

The correct answer is ***D.***

We know that:

$$\text{SEE}_{\beta_0} = \sqrt{\frac{s^2(\hat{\mu}_X^2 + \hat{\sigma}_X^2)}{n\hat{\sigma}_X^2}}$$

Therefore,

$$\text{SEE}_{\beta_0} = \sqrt{\frac{20.45(0.61^2 + 18.65)}{120 \times 18.65}} = 0.4169$$

Q.3969 A regression analysis of monthly returns of a sales company on the market return over ten years gives an intercept of $\hat{\beta}_0 = 0.65$*, the slope* $\hat{\beta} = 1.65$*. Other quantities include:* $s^2 = 20.45$*,* $\hat{\sigma}_X^2 = 18.65$ *and* $\hat{\mu}_X = 0.61$*. The analyst wishes to test whether the slope coefficient is different from 0. What is the test statistic of* $\hat{\beta}$ *?*

- *A. 17.2594*
- *B. 10.1891*
- *C. 24.3234*
- *D. 20.3232*

The correct answer is ***A.***

We start by starting the hypothesis:

$$H_0 : \beta = 0 \text{ vs } H_0 \neq 0$$

We know that:

$$T = \frac{\hat{\beta} - \beta_{H_0}}{SEE_{\hat{\beta}}}$$

Where:

$$SEE_{\beta} = \sqrt{\frac{s^2}{n\hat{\sigma}_X^2}} = \sqrt{\frac{20.45}{120 \times 18.65}} = 0.0956$$

So,

$$T = \frac{\hat{\beta} - \beta_{H_0}}{SEE_{\hat{\beta}}} = \frac{1.65 - 0}{0.0956} = 17.2594$$

Q.3970 A regression analysis of monthly returns of a sales company on the market return over ten years gives an intercept of $\hat{\beta}_0 = 0.65$*, the slope* $\hat{\beta} = 1.65$*. Other quantities include:* $s^2 = 20.45$*,* $\hat{\sigma}_X^2 = 18.65$ *and* $\hat{\mu}_X = 0.61$*. The analyst wishes to test whether the slope coefficient is different from 0. What is 99% confidence interval for* $\hat{\beta}$ *?*

A. *[1.6034,1.8906]*

B. *[1.3034,1.8966]*

C. *[1.3997,1.9002]*

D. *[1.5034,1.6976]*

*The correct answer is **C.***

The confidence interval is given by:

$$[\hat{\beta} - C_t \times SEE_{\hat{\beta}}, \hat{\beta} + C_t \times SEE_{\hat{\beta}}]$$

The critical value (C_t) *at 1% (0.5% on each tail) level is 2.618. (From the t-distribution table)*

Now, $SEE_{\hat{\beta}}$ *is given by,*

$$SEE_{\beta} = \sqrt{\frac{s^2}{n\hat{\sigma}_X^2}} = \sqrt{\frac{20.45}{120 \times 18.65}} = 0.0956$$

So,

$$\begin{aligned}[\hat{\beta} - C_t \times SEE_{\hat{\beta}}, \hat{\beta} + C_t \times SEE_{\hat{\beta}}] &= [1.65 - 2.618 \times 0.0956, 1.65 + 2.618 \times 0.0956] \\ &= [1.3997, 1.9002]\end{aligned}$$

Q.3971 You want to develop a model that regresses the number of options on the number of underlying stocks. You estimate the following regression equation.

$$\text{Number of Options} = \hat{\alpha} + \hat{\beta}\,(\text{amount of underlying stock})$$

The statistical software you used returns a 90% confidence interval of [0.30,1.60] for the slope coefficient. If the 10% critical value for the t-test is 1.70, what is the likely value of the p-value corresponding to your slope coefficient if you wanted to test whether the slope is different from 0?

A. 0.0132

B. 0.0164

C. 0.0192

D. 0.0186

*The correct answer is **A.***

Clearly, this a two-tailed test due to confidence interval. So we need:

$$p - \text{value} = 2[1 - \Phi(|T|)]$$

Recall that the confidence interval is given by:

$$[\hat{\beta} - C_t \times SEE_{\hat{\beta}}, \hat{\beta} + C_t \times SEE_{\hat{\beta}}]$$

$\hat{\beta}$ *is equivalent to the midpoint of the confidence interval. That is:*

$$\frac{1}{2}(\hat{\beta} - C_t \times SEE_{\hat{\beta}}, \hat{\beta} + C_t \times SEE_{\hat{\beta}}) = \hat{\beta}$$

So,

$$\hat{\beta} = \frac{1}{2}(0.30 + 1.60) = 0.95$$

The critical value C_t *is for two-tailed 10% is 1.70. So, using the lower bound of the confidence interval then:*

$$0.95 - 1.70 \times SEE_{\hat{\beta}} = 0.30 \Rightarrow SEE_{\hat{\beta}} = \frac{0.95 - 0.30}{1.70} = 0.3824$$

We know that the test statistic is given by:

$$T = \frac{\hat{\beta} - \beta_{H_0}}{SEE_{\hat{\beta}}} = \frac{0.95 - 0}{0.3824} = 2.4843$$

We need:

$$2[1 - \Phi(|T|)] = 2[1 - \Phi(|2.4843|)] = 2[1 - 0.9934] = 0.0132$$

Q.3972 Which of the following is the correct sequence of steps in hypothesis testing?

A. State the hypothesis, select the level of significance, compute the test statistic, formulate the decision rule, and make a decision.

B. State the hypothesis, select the level of significance, formulate the decision rule, compute the test statistic, and make a decision.

C. State the hypothesis, formulate the decision rule, select the level of significance, compute the test statistic, and make a decision.

D. Formulate the decision rule, state the hypothesis, select the level of significance, compute the test statistic, and make a decision.

The correct answer is ***B.***

The correct sequence is:

1. *State the hypothesis*
2. *Select the level of significance*
3. *Formulate the decision rule*
4. *Compute the test statistic*
5. *Make a decision*

Q.3973 The covariance between the 10-year money supply growth rates and the inflation rate is 0.007668, and the variance of the money supply growth rates is 0.02320. An investment analyst wants to explain the inflation rates using the money supply growth rates and predict the inflation rate when the money supply rate is 25%. The 10-year means for the money supply growth rate and inflation rate are 9% and 3%, respectively. The predicted inflation rate is closest to:

A. 7.234%

B. 8.289%

C. 6.345%

D. 8.756%

*The correct answer is **B.***

Intuitively, the dependent variable (Y) is the inflation rate, and the independent variable (X) is the money supply growth rate.

We know that:

$$\hat{\beta} = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X^2} = \frac{0.007668}{0.02320} = 0.33051$$

And

$$\hat{\beta}_0 = \bar{Y} - \beta\bar{X} = 0.03 - 0.33051 \times 0.09 = 0.0002541$$

So, the estimated regression equation is:

$$\hat{Y} = 0.0002541 + 0.33051X$$

Now, the analyst wants to predict the inflation rate ($\hat{Y}$) when the money supply growth rate (X) is 0.25 (25%). That is:

$$\hat{Y} = 0.0002541 + 0.33051 \times 0.25 = 0.0828816 \approx 8.289\%$$

Reading 19: Regression with Multiple Explanatory Variables

Q.385 Which of the following is NOT true regarding a scatter plot?

> *A. It's a visual representation of the relationship between the explained variable and the independent/explanatory variable.*
>
> *B. It's a standard two-dimensional graph where the values of the explained variable are plotted on the Y-axis while those of the explanatory variable are plotted on the X-axis.*
>
> *C. It reveals the kind of relationship between the explained variable and the explanatory variable – linear or non-linear which may be positive or negative. .*
>
> *D. None of the above.*

*The correct answer is **D.***

It's imperative to note that the independent/explanatory variable takes on the X-axis while the explained/dependent variable takes on the Y-axis. For example, when assessing the relationship between productivity and the age of employees, age would be our independent variable (X-axis) while productivity would be the dependent variable (Y-axis).

Q.393 Which of the following is true regarding the coefficient of determination?

A. Can take values between 0% and 100% inclusive

B. Will generally increase when additional independent variables are added to the regression model

C. It is maximized by ordinary least squares.

D. All of the above are correct

The correct answer is ***D.***

The coefficient of determination refers to the proportion of variability of the responses explained by a model.

$$R^2 = \frac{SS_{REG}}{SS_{TOT}} = \frac{S(xy)^2}{SS_{xx}SS_{yy}}$$

Besides taking values between 0 and 1 inclusive, it increases as the number of "independents" in a model is increased.

Q.407 The following table represents the return of a portfolio over the return of its benchmark.

Portfolio Parameter	Value
Alpha	0.25
Coefficient of Determination	0.77
Standard Deviation of Error	2.40
Beta	1.20

Which of the following statements are correct?
I. The correlation is 0.69
II. The dependent variable is the portfolio
III. About 23% of the variation noted in the portfolio return is explained by variation in benchmark return
IV. For an estimated portfolio return of 10%, the 95% confidence interval is (5.296%, 14.704%)

A. I and II

B. II and III

C. II only

D. II and IV

*The correct answer is **D.***

The correlation is $\sqrt{0.77} = 0.88$ *so statement (I) is incorrect. The dependent variable is indeed the portfolio while the independent variable is the benchmark, so statement (II) is correct. The benchmark explains about 77% of the variation, so (III) is incorrect.*

Lastly, if we estimate the portfolio return to be 10%, the 95% CI for the return would be:

$$10\% \pm Z_{\frac{a}{2}} * 2.4 = 10\% \pm 1.96 * 2.4 = (5.296\%, 14.704\%)$$

So statement (IV) is correct.

Q.436 Which of the following best explains why multiple linear regression analysis may be preferrec to single linear regression?

A. It is simpler to model using modern software and computer programming.

B. It reduces the omitted variable bias.

C. It's easier to model and establishes the relationship between the dependent variable anc important independent variables.

D. None of the above.

The correct answer is ***B.***

Omitted variable bias occurs when a regression model leaves out independent variables that have a major bearing on the dependent variable. The omitted variable can either be correlated with the movement of the independent variable already included in the model, or it could be a determinant ot the dependent variable. The omission of certain variables may overstate the explanatory power ot the regression analysis.

Q.439 Under multiple linear regression, a residual is defined as:

A. A type 1 error

B. The error sum of squares

C. The regression sum of squared deviations from the mean value of the dependent variable

D. $Y - \hat{Y}$

*The correct answer is **D.***

The residual, e_i*, is the difference between the observed value,* Y_i*, and the predicted value from the regression,* $\hat{Y}_i$.

$$E_i = Y_i - \hat{Y}_i = Y_i - (b_0 + b_1X_{1i} + b_2X_{2i} + \ldots + b_kX_{ki})$$

A is incorrect. Type I error is the rejection of a true null hypothesis.

B is incorrect. The error (residual) sum of squares (RSS), also known as the sum of squared residuals (SSR) or the sum of the squared estimate of errors (SSE), is the sum of the squares of residuals. In this case, the examiner asks for just the residual without the square.

$$\text{Residual} = \text{observed Y} - \text{predicted Y}$$

$$\text{RSS} = (\text{observed Y} - \text{predicted Y})^2$$

Q.440 Assume we have the following multiple regression model:

$$Y = b_0 + 0.25X_1 + 0.14X_2 + \epsilon$$

It would be correct to say that:

- *A. If the independent variable* X_1 *increases by 1 unit, we would expect* Y *to increase by 0.25 units.*

- *B. If the independent variable* X_1 *increases by 1 unit, we would expect* Y *to increase by 0.25/0.14 units.*

- *C. If independent variable* X_1 *increases by 1 unit, we would expect* Y *to increase by 0.25 units, holding* X_2 *constant.*

- *D. If independent variable* X_2 *increases by 1 unit, we would expect* Y *to increase by 0.25 + 0.14 units.*

*The correct answer is **C.***

The interpretation of the estimated regression coefficients in both single and multiple regression models is the same but only with respect to the intercept term. Under multiple regression models, each slope coefficient represents the estimated change in the dependent variable, for a one-unit change in that independent variable, holding the other independent variables constant. In fact, the slope coefficients in multiple regression models are also known as partial slope coefficients.

Q.441 Which of the following best describes how OLS estimators are derived in multiple regression models?

- *A. By minimizing the absolute difference of the residuals.*

- *B. By minimizing the sum of squared prediction mistakes.*

- *C. Minimizing the distance between the actual and fitted values.*

- *D. By equating the sum of squared errors to zero.*

*The correct answer is **B.***

The multiple regression model estimates the slope coefficients and the intercept (OLS estimators) by minimizing the sum of the squared error terms, $\sum \epsilon_i^2$*.*

Q.443 In a problem involving three independent variables and one dependent variable, assume that the computed coefficient of determination is 0.59. This result means that:

A. 59% of the total variation is explained by the dependent variable.

B. The correlation coefficient is 0.59 as well.

C. At least two of the three independent variables are highly correlated.

D. 59% of the total variation in the dependent variable is explained by the independent variables.

The correct answer is ***D.***

In multiple regression, the coefficient of determination refers to the total variation in the dependent variable that's collectively explained by all of the independent variables. This interpretation is actually similar to that for linear regression with one regressor.

Q.444 Under multiple linear regression models, there's always the risk of overestimating the impact of additional variables on the explanatory power of the resulting model, which is why most researchers recommend using the adjusted R^2, $\overline{R^2}$, *instead of* R^2 *itself. This adjusted* R^2*:*

A. Is always positive.

B. Will never be greater than the regression R-Squared.

C. Cannot increase when an additional independent variable is incorporated into the model.

D. Is always negative.

The correct answer is ***B.***

$\overline{R^2}$ *is always less than or equal to* R^2*. While adding a new independent variable increases* R^2*, it may either increase or decrease* $\overline{R^2}$*. If* R^2 *is very small, then an additional explanatory variable may actually result in a negative value for* R^2*.*

C is incorrect. The adjusted R-squared can increase, but that happens only if the new variable improves the model more than would be expected by chance. If the added variable improves the model by less than expected by chance, then the adjusted R-squared decreases.

Q.445 When an important variable is omitted from a regression model, the assumption that $E(\epsilon_i|X_i = 0)$ *is violated. This implies that:*

A. The OLS estimator is biased.

B. The product of the residuals and any of the independent variables is no longer zero.

C. The sum of the residuals is no longer equal to zero.

D. The coefficient of determination is zero.

The correct answer is ***A.***

One of the key assumptions of the multiple linear regression model stipulates that the conditional distribution of ϵ_i *given* $X_{1i}, \ldots, X_{ki}$ *has a mean of zero. The interpretation is that sometimes* Y_i *is above the population regression line, other times it's below the line but on average* Y_i *falls on the population regression line. In case an important independent variable is left out, this is no longer true and this implies that your estimators are biased.*

Q.446 Study the following table:

Source	Sum of squares
Explained	825
Residual	625

The total sum of squares is closest to:

A. 1.32

B. 200

C. 1450

D. 0.7576

The correct answer is ***C.***

$$TSS = 825 + 625 = 1450$$

Q.447 An analyst uses the following regression model to explain stock returns:
Dependent variable:
ASR = Annual stock returns (%)

Independent variables:
MCP = Market capitalization (divided by $1million to simplify modeling)
SEF = Stock exchange firm, where SEF = 1 if the stock is that of a firm listed on the New York Stock Exchange and SEF = 0 if not listed
FMR = Forbes magazine ranking (FMR = 4 is the highest ranking)

The following table presents the regression results:

	Coefficient	Standard Error
Intercept	0.6330	1.11
MCP	0.0840	0.0130
SEF	0.5101	0.1235
FMR	0.7000	0.3241

Based on the results in the table above, which of the following is the correct regression equation?

A. 0.0840(MCP) + 0.5101(SEF) + 0.7(FMR)

B. 0.6330 + 0.0840(MCP) + 0.5101(SEF) + 0.7(FMR)

C. 1.11 + 0.0840(MCP) + 0.5101(SEF) + 0.7(FMR)

D. 1.11 + 0.0130(MCP) + 0.1235(SEF) + 0.3241(FMR)

The correct answer is ***B.***

All the regression parameters are contained in the coefficients column. Standard errors would only be used to conduct hypothesis tests or construct confidence intervals about the independent variables.

Q.448 An analyst uses the following regression model to explain stock returns:
Dependent variable:
ASR = Annual stock returns (%)

Independent variables:
MCP = Market capitalization (divided by $1million to simplify modeling)
SEF = Stock exchange firm, where SEF = 1 if the stock is that of a firm listed on the New York Stock Exchange and SEF = 0 if not listed
FMR = Forbes magazine ranking (FMR = 4 is the highest ranking)

If the regression equation is 0.6330 + 0.0840(MCP) + 0.5101(SEF) + 0.7(FMR), then what is the expected amount of stock return that would be attributed to it being a listed stock?

A. 1.11 + 0.5101

B. 0.1235

C. 0.5101

D. 1.11 + 0.1235

*The correct answer is **C.***

The regression equation is 0.6330 + 0.0840(MCP) + 0.5101(SEF) + 0.7(FMR)
The coefficient on SEF is the amount of stock return that would be attributed to it being a listed stock.

Q.449 The following are assumptions of the multiple linear regression model EXCEPT:

A. The independent variables are random, and there is an exact linear relationship between any two or more independent variables.

B. The error term has an expected value equal to zero and is normally distributed.

C. A linear relationship exists between the dependent and independent variables.

D. The error for one observation is uncorrelated with that of a different observation.

*The correct answer is **A.***

*One of the most important assumptions of multiple regression states that independent variables **must not** be random, and there should be **no exact linear relationship** between any two or more independent variables.*

Q.450 To construct a confidence interval for a regression coefficient, we need the estimated regression coefficient, the appropriate test statistic, and:

A. The F-statistic

B. The standard error of the regression coefficient

C. The coefficient of determination

D. The adjusted R-squared

The correct answer is ***B.***

A confidence interval for a regression coefficient under multiple linear regression modeling is given by:

$$CI = \beta_j \pm (t_{a,n-k-1} * Se(\beta_j))$$

Where B_j *is the estimated coefficient of the regression parameter, and* $Se(\beta_j)$ *is the standard error of the coefficient.*

Q.451 An analyst believes that future 15-year real earnings of the S&P 500 are a function of the trailing dividend payout ratio of the stocks in the index (DB) *and the yield curve slope* (YC). *She collects data and obtains the following multiple regression results:*

	Coefficient	Standard Error
Intercept	−10.8%	1.567%
DB	0.27	0.029
YC	0.12	0.210

Test the statistical significance of the independent variable DB *at the 5% level of significance, quoting the value of the test statistic and the conclusion. (Number of observations = 43)*

A. Test statistic = 2.021; DB regression coefficient is statistically different from zero

B. Test statistic = 9.310; DB regression coefficient is statistically different from zero

C. Test statistic = 0.018; DB regression coefficient is not statistically different from zero

D. Test statistic = 9.310; DB regression coefficient has little effect on the returns of S&P 500

The correct answer is ***B.***

We are testing the following hypothesis:

$$H_0 : \beta_{DB} = 0 \text{ vs } H_1 : \beta_{DB} \neq 0$$

The test statistic is $\frac{0.27}{0.029} = 9.310$

$T_{\frac{a}{2}, n-k-1} = t_{0.025,40} = 2.021$

The test statistic (9.310) is greater than the upper critical value (2.021) of the t-distribution with 40 degrees of freedom. Therefore, we reject the null hypothesis and conclude that the DB regression coefficient is statistically different from zero at the 5% level of significance.

Q.453 The following table presents regression results for a linear regression model with 3 independent random variables:

Variable	Coefficient	Standard Error	t-statistic	p-value
Intercept	2.0	2.0	1.0	0.3215
X	−1.8	0.56	−3.2	0.0022
Y	16.4	4.10	4.0	0.0002
Z	0.12	0.54	0.22	0.0319

Determine the regression parameters that are significantly different from zero at the 1% level of significance, assuming $n = 60$.

A. X *and* Y

B. X *only*

C. Y *and* Z

D. All of X, Y *, and* Z

The correct answer is ***A.***

The p-value is the smallest level of significance for which the null hypothesis can be rejected. In other words, if the $p-\text{value}$ *is less than the significant level, we can reject the null hypothesis. Considering the p-values for each of the variables,* X, Y*, and* Z*, only those of* X *and* Y *are less than 1% (0.01). This means that* X *and* Y *are statistically significantly different from zero.*

Q.454 A multiple regression model has 4 independent variables such that:

$$Y_i = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

An analyst carries out a joint hypothesis test to determine the statistical significance of the independent variable coefficients, incorporating all the 3 variables. The null hypothesis is such that each variable coefficient is equated to zero. The results reveal that the F*-statistic is greater than the one-tailed critical* F*-value. This implies that:*

> *A. At least one of the coefficients is statistically significantly different from zero.*
>
> *B. Each of the independent variable coefficients is statistically significantly different from zero.*
>
> *C. None of the coefficients is statistically different from zero.*
>
> *D. Only one of the independent variable coefficients is statistically different from zero.*

The correct answer is ***A.***

A joint hypothesis test is used to assess whether ***at least one*** *of the independent variables explains a significant portion of the total variation exhibited by the dependent variable. To determine the statistical significance, the* F*-statistic calculated is compared with the always one-tailed critical* F*-value. The null hypothesis is rejected when the* F*-statistic > one-tailed critical value, indicating that at least one of the coefficients is statistically significantly different from zero.*

Q.455 Peter Bridge, FRM, runs a regression of monthly stock returns on four independent variables over 65 months. The total sum of squares is 540, and the sum of squared residuals is 250. Carry out a statistical test at the 5% significance level with the null hypothesis that all four of the independent variables are equal to zero. Quote the F*-statistic and the conclusion.*

> *A.* F*-statistic = 17.40; At least one of the 4 independent variables is significantly different from zero.*
>
> *B.* F*-statistic = 2.525; At least one of the 4 independent variables is significantly different from zero.*
>
> *C.* F*-statistic = 72.5; All the 4 independent variables are significantly different from zero.*
>
> *D.* F*-statistic = 17.40; None of the independent variables is significantly different from zero.*

The correct answer is ***A.***

The set of hypotheses is such that:

$H_0 : B_1 = B_2 = B_3 = B_4 = 0$ vs H_1 *At least one* $B_j \neq 0$

The F-statistic is given by:

$$F = \frac{\left[\frac{ESS}{k}\right]}{\left[\frac{SSR}{(n-k-1)}\right]}$$

Where:

TSS =total sum of squares;

ESS = explained sum of squares;

RSS = residual sum of squares;

k = number of independent variabls; and

n=number of observations.

$$ESS = TSS - SSR = 540 - 250 = 290$$

Therefore,

$$F = \frac{\left[\frac{290}{4}\right]}{\left[\frac{250}{(60)}\right]} = 17.4$$

The one-tailed critical F*-value at 5% level with 4 and 60 degrees of freedom in the numerator and denominator respectively, is approximately 2.525. Since 17.40 > 2.525, we can reject the null hypothesis and conclude that at least one of the four independent variables is significantly different from zero.*

Q.456 Which of the following statements is INCORRECT regarding the use if R^2 *and* $\overline{\mathrm{R}^2}$ *in multiple regression analysis?*

A. An increase in the R^2 *or* $\overline{\mathrm{R}^2}$ *always means that an added variable is statistically significant*

B. A high R^2 *or* $\overline{\mathrm{R}^2}$ *does not mean that the regressors are the true cause of the dependent variable*

C. A high R^2 *or* $\overline{\mathrm{R}^2}$ *does not necessarily indicate that you have the most relevant set of regressors, nor does a low* R^2 *or* $\overline{\mathrm{R}^2}$ *necessarily indicate the presence of inappropriate regressors*

D. A high R^2 *or* $\overline{\mathrm{R}^2}$ *does not mean that we do not have omitted variable bias*

The correct answer is ***A.***

Regardless of the significance of the added variable, the value of R^2 *will always increase when a regressor is added. The value of* $\overline{\mathrm{R}^2}$ *does not always increase, but when it does, this does not always indicate that the added regressor is statistically significant. To be able to determine the significance of an added regressor confidently, a t-test must be performed.*

Q.459 Elizabeth Graham, FRM, runs a regression of monthly stock returns on five independent variables over 66 months. The explained sum of squares is 270, and the sum of squared residuals is 250. Graham then performs a statistical test at the 10% significance level with the null hypothesis that all five of the independent variables are equal to zero. Quote the F *-statistic and the conclusion.*

> *A.* F *-statistic = 12.96; At least one of the 5 independent variables is significantly different from zero.*
>
> *B.* F *-statistic = 1.946; At least one of the 5 independent variables is significantly different from zero.*
>
> *C.* F *-statistic = 72.5; All the 5 independent variables are significantly different from zero.*
>
> *D.* F *-statistic = 17.40; None of the independent variables is significantly different from zero.*

*The correct answer is **A.***

The set of hypotheses is such that:

$H_0 : B_1 = B_2 = B_3 = B_4 = 0$ vs H_1 *At least one* $B_j \neq 0$

$$\text{The F-statistic} = \frac{\left[\frac{ESS}{k}\right]}{\left[\frac{SSR}{(n-k-1)}\right]}$$

$$F = \frac{\left[\frac{270}{5}\right]}{\left[\frac{250}{60}\right]}$$

$$= 12.96$$

The one-tailed critical F-value at 10% level with 5 and 60 degrees of freedom in the numerator and denominator respectively, is approximately 1.946. Since 12.96 > 1.946, we can reject the null hypothesis and conclude that at least one of the five independent variables is significantly different from zero.

Q.460 An analyst runs a regression of monthly value-stock returns on 8 independent variables . Given the following information: Explained Sum of Squares=1435 Residual sum of Squares=1335 Number of obsrvations=28 R^2 *and the* F*-statistic, respectively, are closest to:*

A. 53%, 4

B. 51%, 3.8

C. 52%, 2.6

D. 50%, 4

The correct answer is ***C.***

$$R^2 = \frac{ESS}{TSS} = \frac{1435}{2770} = 52\%$$

$$F = \frac{\left[\frac{ESS}{df}\right]}{\left[\frac{SSR}{n-(df+1)}\right]} = \frac{\left[\frac{1435}{8}\right]}{\left[\frac{1335}{28-8-1}\right]} = 2.5529$$

Q.461 A Financial Risk Manager exam candidate conducts a hypothesis test using a 10% significance level. Which of the following statements are correct?

A. The significance level is equal to the size of the test.

B. 5% of the total distribution will be in each tail rejection region if the test is 2-sided.

C. 10% of the whole distribution will be in the rejection region if the test is 2-sided.

D. All the above.

*The correct answer is **D.***

Option A is true since the significance level and the size of a test mean the same thing: they measure the proportion of the total distribution which is placed in the rejection region.

Option B is true since in case the test is 2-sided, this implies the significance level must be split into two equal halves, with 5% of the distribution located in each of the rejection regions.

Option C is true since If a 10% significance level is used, this implies that in total, 10% of the distribution must be in the rejection region.

Thus we conclude that D is the correct answer.

Q.462 Under multiple linear regression, if an estimator is said to be consistent, what does this imply?

A. On average, the estimated values of the coefficients will be equal to the true values.

B. The coefficient estimates will be as close to their true values as possible, regardless of the sample size.

C. The estimates will converge upon the true values as the sample size, n, increases.

D. The OLS estimator will also be unbiased and efficient.

*The correct answer is **C.***

By definition, a consistent estimator is one where the sample estimates converge on their true population values as the sample size increases. Choice A is actually the definition of an unbiased estimator, while choice B is also a differently worded description of unbiasedness.

Q.463 Assume that after constructing a multiple linear regression model, you find that $R^2 = 0.97$*. How confident would you be in using the line of best fit for the purposes of prediction?*

A. Not confident.

B. Very confident.

C. The relationship would be too weak at predicting using a linear function.

D. The relationship would be random and hence impossible to predict.

The correct answer is ***B.***

If $R^2 = 0.97$*, this implies that 97% of the total variation is 'explained' or 'accounted for' by the regression line. In other words, the regression line passes through almost all the data points on the scatter*

Q.480 A multiple regression model with three variables has the following formula:

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i}$$

A FRM exam candidate performs a joint hypothesis test where:

$$H_0 : b_1 = b_2 = b_3 = 0$$

If the computed F-statistic is less than the tabulated one-tailed F *-value, this implies that:*

A. All the three coefficients of the independent variables are statistically different from zero.

B. Only b_1 *is statistically different from zero.*

C. The effects of the independent variables on Y *are statistically significant.*

D. None of the coefficients are statistically significantly different from zero.

The correct answer is ***D.***

In joint hypothesis tests, the aim is always to establish the statistical significance of ***at least one*** *of the regression coefficients. If the computed* F *-statistic is greater than the 1-tailed F-value at a given level of significance, at least one of the coefficients is statistically significantly different from zero; otherwise, none of the coefficients are statistically significant.*

Q.481 Suppose you performed the F*-test on a multiple regression model and established that a significant amount of variation in the* Y *variable is explained by the set of* X *variables; then:*

A. You should transform the Y *variable.*

B. You could perform another test with an indicator variable so as to establish the significance level of the test.

C. You should discard the initial model in favor of a different one with more X *variables.*

D. You should perform t*-tests on each* X *variable to establish whether there's any of them whose effect on* Y *is not statistically significant.*

The correct answer is ***D.***

The F*-test establishes whether at least one of the independent variables is statistically significant, but stops short of specifying which one. To establish the significance of each variable, we perform a* t*-test on each and discard the ones that appear insignificant.*

Q.482 A market analyst has established that future 10-year growth of earnings in the S&P 500 can be explained by a combination of two factors: the slope of the yield curve (YCS) and the preceding dividend payout ratio (PR) of stocks that have been featured in the index. The analyst carries out a regression and obtains the following results:

	Coefficient	Standard error
Intercept	−10.6%	1.525%
YCS	0.20	0.024
PR	0.12	0.230

Test the statistical significance of YCS at the 10% level of significance, quoting the t-statistic and the conclusion if n *= 46.*

A. 16.60; The YCS coefficient is statistically significantly different from zero.

B. 16.60; The PR coefficient is statistically significantly different from zero.

C. 8.333; The YCS coefficient is statistically significantly different from zero.

D. 1.68; The YCS coefficient is not statistically significantly different from zero.

The correct answer is ***C.***

You should test the hypothesis:

$$H_0 : PR = 0 \text{ vs } H_1 : PR \neq 0$$

The t-statistic = $\frac{0.20}{0.024} = 8.333$ *$*

From statistical tables, we find that the 10% 2-tailed critical t*-value with a total of (46 – 2 – 1) degrees of freedom = 1.68 (approximately). We should reject* H_0 *only if the t-statistic is greater than the tabulated* t*-value.*

$$8.333 > 1.68$$

Therefore, we can reject H_0 *and conclude that the YCS regression coefficient is statistically significantly different from zero.*

Q.483 The following statements regarding R^2 and the adjusted R^2 are correct EXCEPT:

A. If R^2 improves after the addition of an independent variable, that does not necessarily mean that the variable is statistically significant

B. A high R^2 means that the independent variables are the definite causes of the movement seen in the dependent variable

C. Even with a high value of R^2, it's incorrect to assume that all the relevant independent variables have been found

D. The R^2 cannot and does not give evidence that the most or least statistically significant variables have been selected

*The correct answer is **B.***

A high R^2 does not mean that the included independent variables are the exact/definite cause of the movement seen in the dependent variable. In some cases, the R^2 could just be spurious, which rules out exactness.

Q.484 Which of the following best defines the omitted variable bias under multiple regression?

A. The bias that emerges whenever an omitted determinant of the dependent variable is correlated with at least one of the included regressors.

B. The bias that emerges whenever two or more included regressors are correlated with an omitted variable.

C. The bias that emerges whenever one or more included regressors are uncorrelated with an omitted variable.

D. The bias that emerges whenever one or more included regressors are positively correlated with an omitted variable.

*The correct answer is **A.***

The omitted variable bias exists whenever an omitted determinant of Y (the dependent variable) is correlated with at least one of the included regressor variables.

Q.3334 Tom Well, FRM, works for a trading company. Using historical data, he has computed the following variables considering one independent and one dependent variable.

- *Explained Sum of Squares (ESS) = 60*
- *Sum of Squared Residuals (SSR) = 15*

If we are dealing with a sample size of 62 observations, what is the coefficient of determination and the standard error of the estimate, respectively.

A. Coefficient of Determination = 0.50, Standard Error of the Estimate = 0.50

B. Coefficient of Determination = 0.80, Standard Error of the Estimate = 0.80

C. Coefficient of Determination = 0.80, Standard Error of the Estimate = 0.50

D. Coefficient of Determination = 0.25, Standard Error of the Estimate = 0.25

*The correct answer is **C.***

$$\text{Total variation (TSS)} = \text{Explained variation (ESS)} + \text{Unexplained variation (SSR)}$$
$$TSS = 60 + 15 = 75$$

$$\text{Coefficient of determination: R-square} = \frac{ESS}{TSS} = \frac{60}{75} = 0.80$$

Standard Error of the Estimate (SEE) is given by:

$$SEE = \sqrt{\frac{\text{Unexplained variation}}{n-2}} = \sqrt{\frac{15}{62-2}} = 0.5$$

Additional tips:

$$\text{Total variation} = \sum (Y - \bar{Y})^2$$
$$\text{Unexplained variation} = \sum (Y - \hat{Y})^2 = \sum (\hat{\epsilon})^2$$
$$SEE = \sqrt{\frac{\sum (\text{Actual Y} - \text{Predicted Y})^2}{n-2}} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{\sum (\hat{\epsilon})^2}{n-2}}$$

Q.3345 During the course of building a model using multiple linear regression, an analyst tried to judge the model based on its coefficient of determination (R^2) *and adjusted* R^2*. Which of the following interpretation is correct?*

A. The adjusted R^2 *is always greater than the* R^2

B. Both the adjusted R^2 *and the* R^2 *always have positive values*

C. The adjusted R^2 *is always less than the* R^2

D. The adjusted R^2 *always increases with an increase in the number of independent*

The correct answer is ***C.***

R^2 *is not a reliable indicator of the explanatory power of a multiple regression model. Why?* R^2 *almost always increases as new independent variables are added to the model, even if the marginal contribution of the new variable is not statistically significant.*
Thus, a high R^2 *may reflect the impact of a large set of independents rather than how well the set explains the dependent.*

This problem is solved by the use of the adjusted R^2*. The adjusted R-squared increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance. The adjusted R-squared can be negative, but it's usually not. Negative values can occur when the model contains terms that do not help to predict the dependent. It is always lower than the R-squared.*

It can be calculated as shown below

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1 - R^2)$$

$$1 - \bar{R}^2 = \left(\frac{n-1}{n-k-1}\right)(1 - R^2)$$

Where n *is the number of observations and* k *is the number of independent variables*

Q.3346 An analyst performed a regression of monthly returns on a stock with 4 independent variables over a 50 month period. The analyst calculated the total sum of squares (TSS) and the sum of square residuals or error (SSR) as 500 and 100, respectively.
What is the adjusted R^2*?*

- *A. 0.80*
- *B. 0.78*
- *C. 0.20*
- *D. 0.75*

The correct answer is ***B.***

$$R^2 = \frac{\text{Explained variation}}{\text{Total variation}} = 1 - \frac{\text{Unexplained variation}}{\text{Total variation}} = 1 - \frac{100}{500} = 0.8$$

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1 - R^2) = 1 - \left(\frac{50-1}{50-4-1}\right)(1-0.8) = 1 - \frac{49}{45} \times 0.2 = 0.7822 \approx 0.78$$

Q.3350 In a hypothetical world, GDP is regressed against interest rate and inflation, and regression results are shown below.

$$\text{GDP} = \text{a} + \text{b (Interest rate)} + \text{c (Inflation)} + \text{Error term}$$

	Coefficient	p-Value
a	9	0.042
b	2	0.035
c	1.5	0.012

ANOVA	df	SS
Regression	2	240
Residual	37	1070
Total	39	1300
Total	0.428	
R2	0.183	
Observation	40	

Which of the test is relevant to determine whether the regression model as a whole is significant?

A. $\text{F}-\text{test}$; H_0*: All slope coefficients = 0;* H_a*: At least one slope coefficient* $\neq$ *0*

B. $\text{F}-\text{test}$; H_0*: All slope coefficients* ≥ 0*;* H_a*: At least one slope coefficient < 0*

C. $\text{t}-\text{test}$; H_0*: All slope coefficients = 0;* H_a*: At least one slope coefficient* $\neq$ *0*

D. $\text{t}-\text{test}$; H_0*: All slope coefficients* ≥ 0*;* H_a*: At least one slope coefficient < 0*

The correct answer is ***A.***

An F-test is used to test whether any of the independent variables explain the variation in the dependent variable (test of overall model significance). It is used to determine whether the regression model as a whole is significant.

H_0*: All slope coefficients = 0*

H_a*: At least one slope coefficient* $\neq 0$

Q.3352 In a hypothetical world, GDP is regressed against interest rate, and inflation and regression results are shown below.

$$GDP = a + b\,(\text{Interest Rate}) + c\,(\text{Inflation}) + \text{Error Term}$$

	Coefficient	p-Value
a	9	0.042
b	2	0.035
c	1.5	0.012

ANOVA	df	SS
Regression	2	240
Residual	37	1070
Total	39	1310
Total	0.428	
R2	0.183	
Observation	40	

Assume that on a certain significance level, the critical value of the F-statistic is 4. Which of the following is correct?

A. *The value of F-statistic is less than its critical value.*

B. *We fail to reject the null hypothesis:* H_0*: All slope coefficients = 0.*

C. *The model as a whole is statistically significant.*

D. *The model as whole is not statistically significant.*

The correct answer is ***C.***

The F-statistic is given by:

$$F = \frac{\left[\frac{ESS}{k}\right]}{\left[\frac{SSR}{(n-k-1)}\right]}$$

Where:

TSS *=total sum of squares;*

ESS *= explained sum of squares;*

RSS *= residual sum of squares;*

k *= number of independent variables; and*

n=*number of observations.*

Therefore,

$$F = \frac{\frac{240}{2}}{\frac{1070}{(40-2-1)}} = \frac{120}{\frac{1070}{37}} = 4.1495$$

The F-statistic is 4.1495 and its critical value is 4.

If F-statistic > F critical, we reject the null hypothesis.

H_0*: All slope coefficients = 0*

H_a*: At least one slope coefficient* $\neq 0$

Model as a whole is significant.

Q.3353 Suppose that the following regression is estimated using 25 quarterly observations:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

What is the appropriate critical value for a 2-sided 5% size of test of $H_0 : \beta_3 = 1$*?*

A. 1.72

B. 2.06

C. 2.07

D. 1.64

The correct answer is ***C.***

When we have a total of k *"regressors" (including a constant) and* n *observations, the t-test statistic will follow a t-distribution with* $n - k$ *degrees of freedom.*
In this case, $n = 25, k = 3$

Also, this is a two-tailed test.

Thus, we would be interested in $t_{0.025,22}$*. We would be looking in the t-tables in the degrees of freedom = 22 rows and the 2.5% column (so that 2.5% is in each tail for a 5% 2-tailed test). The critical value would be 2.07.*

Option A is incorrect. 1.72 would be the critical value for a 5% one-tailed test or a 10% two-tailed test.

Option B is incorrect. This is the critical value obtained if you forget to subtract the number of parameters estimated (3) from the number of observations to get the degrees of freedom, i.e., $t_{0.025,25} = 2.06$*.*

Option D is incorrect. 1.64 is the 5% one-sided critical value from the normal distribution.

Note: the fact that observations are quarterly is irrelevant in our calculations.

Also, note that when testing a hypothesis about a single coefficient in a model with multiple regressors is just the same as testing in a model with a single explanatory variable. Test of the null hypothesis, $H_0 : \beta_i = \beta_\delta$ *can be achieved using a* t-test*. However, the t-test is* ***NOT*** *suitable to test hypotheses that involve more than one parameter. In such scenarios, then we have to use the* F-test*.*

Q.3354 Consider the following 2 regression models built to estimate a common phenomenon:

Model	R^2	Adjusted R^2
Model 1	0.75	0.73
Model 2	0.81	0.72

If a variable x_4 is introduced in model 2 and that the coefficient estimate β_4 was zero, which one of the following is most likely correct?

A. The researcher must have made a mistake because the adjusted R^2 for model 2 must be greater than the adjusted R^2 for model 1.

B. Variable x_4 is statistically significant.

C. The coefficient estimate β_4 is non-zero but not significant.

D. None of the above.

*The correct answer is **C.***

The fact that R-squared is higher for model 2 but the adjusted R-squared is lower for model 2 seems to suggest that the extra variable x_4 is not statistically significant (i.e. it doesn't have a big impact on the dependent variable y).<|p> ***Option A is incorrect<|b>: The adjusted -R^2 need not increase when a new independent variable is added. In fact, adding new independent variables always increases R^2 except where $\beta = 0$. On the other hand, adding independent variables may either increase or decrease the adjusted-R^2, depending on whether or not the additional variable is significant.***

Q.3355 Consider the following 2 regression models:

$$\text{Model 1}: y_t = \beta_1 + \beta_2 x_{2t} + u_t$$
$$\text{Model 2}: y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

A researcher determines that the two models have identical R-squared values. This most likely implies that:

A. Model 2 must have a lower value of adjusted R-squared.

B. Model 2 must have a higher value of adjusted R-squared.

C. Model 1 and 2 will also have identical values of adjusted R-squared.

D. Variable x_3 ***is statistically significant.***

The correct answer is A.

If the two models have identical values of R-squared, the implication is that the additional variable x_3 ***has zero explanatory power for*** y ***and that its coefficient estimate,*** β_3 ***must be exactly zero. It is also not statistically significant.***

$$\overline{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1 - R^2)$$

In the formula above, k ***represents the number of regressors, which have increased from 2 to 3 in this case. As such, model 2 will definitely have a lower adjusted R-squared.***

Q.3356 For a sample of 40 years, the relationship between GDP growth** (Y_i)**, inflation** (X_1) **and interest rates** (X_2) **is modeled as follows:

$$y_t = \beta_1 + \beta_2 x_{1t} + \beta_3 x_{2t} + u_t$$

An economist wishes to test the joint hypothesis that** $\beta_1 = 0, \beta_2 = 0,$ **and** $\beta_3 = 0$ **at the 90% confidence level.
The p-value for the t-statistic for** β_1 **is 0.11, and the p-value for the t-statistic for** β_2 **is 0.12. The p-value for the F-statistic for the regression is 0.09. Which of the following statements is correct?

> ***A. We cannot reject the null hypothesis because none of the** β_s **is different from zero at 90% confidence.***
>
> ***B. We can reject the null hypothesis because each of the** β_s **is different from zero at 90% confidence.***
>
> ***C. We cannot reject the null hypothesis because the F-statistic is not significant at 90% confidence.***
>
> ***D. We can reject the null hypothesis because the F-statistic is significant at 90% confidence.***

The correct answer is D.

To test, the joint hypothesis, we need the F-test. The t-tests are not sufficient - individual tests do not account for the effects of interactions among the independent variables Thus, the information on the p-value of each** β **is not relevant in this case.
P-value is the smallest level of significance at which we can reject the null hypothesis. In other words, we always reject the null hypothesis whenever the p-value is less than the level of significance of the test. In this case, the P-value of the F-test is 0.09 which is less than 10% (level of significance). As such, we can reject the null hypothesis and conclude that at least one** β **is different from zero.

Reading 20: Regression Diagnostics

Q.401 What is the implication of having heteroskedastic regression?

> ***A. The variance of the error terms is constant.***
>
> ***B. The variance of the error terms is not constant.***
>
> ***C. Subsamples are equally spread out.***
>
> ***D. The independent variable has no linear relationship with the dependent variable.***

The correct answer is B.

Heteroskedasticity occurs when the variance of the residuals, commonly known as error terms, is not the same across all observations in the sample. It is the opposite of homoskedasticity, which occurs when the variance of the residuals is constant across all observations.

Q.403 Which of the following statements regarding linear regression is incorrect?

> ***A. Homoskedasticity occurs when the variance of the residuals is constant across all observations.***
>
> ***B. Heteroskedasticity occurs when the variance of the residuals, commonly known as error terms, is not the same across all observations in the sample.***
>
> ***C. If residual terms are correlated with each other, this can lead to serial correlation.***
>
> ***D. Heteroskedasticity does not lead to problems with inference and estimation.***

The correct answer is D.

Heteroskedastic data still provides an unbiased estimate, but standard errors – and potentially, statistical inferences – are suspect.

Q.437 One of the following assumptions is applied in the multiple least squares regression model. Which one?

A. The independent variables included in the model are homoskedastic.

B. The residual terms are heteroskedastic.

C. The dependent variable is unique and stationary.

D. There is no perfect multi-collinearity..

The correct answer is D.

Multiple linear regression assumes that there's no perfect multicollinearity, i.e., the explanatory variables are not perfectly linearly dependent (i.e., each explanatory variable must have some variation that cannot be perfectly explained by the other variables in the model). Even if there's multicollinearity, it doesn't pose problems for hypothesis testing or parameter estimation as long as it's not perfect.
Choices A, B, and C are not relevant assumptions of the multiple regression model.

Q.442 Jessica Pearson, FRM, builds a model to study the annual salaries of individuals in a certain developed country. The model incorporates just 2 independent variables – age and experience. She is surprised for ending up with a negative value for the coefficient of experience, which seems to be somewhat counterintuitive. Furthermore, the coefficients have low t-statistics but otherwise the model has a high R^2***. Which of the following is the most likely cause of such results?***

A. Heteroskedasticity

B. Multicollinearity

C. Homoskedasticity

D. Serial correlation

The correct answer is B.

Multicollinearity exists when two or more of the independent variables, (or their linear combinations) in a multiple regression model are highly correlated with each other. Such a scenario distorts the standard error of the regression as well as the standard errors for the coefficients which might cause problems when carrying out tests for significance.

In this particular case, age and experience are most likely highly correlated. Even if R^2 ***is high, multicollinearity will still be present as long as the standard errors for the coefficients are high.***

Q.457 Which of the following conditions must be met for omitted variable bias to occur under multiple linear regression?
I. The value of $\overline{R^2}$ ***must be less than that of*** R^2
II. At least one of the included regressors must be correlated with the omitted variable
III. The omitted variable must be a determinant of the dependent variable
IV. The residuals must be homoskedastic
V. The number of included regressors must be less than or equal to 5

A. I and II

B. II and III only

C. I, III, and V

D. All the above

The correct answer is B.

Omitted variable bias is the bias that emerges whenever one or more included regressors are correlated with an omitted variable. For omitted variable bias to arise, conditions II and III above must be met.

Q.3347 What will be the properties of the OLS estimator in the presence of near multicollinearity?

A. It will be consistent, unbiased, and efficient.

B. It will not be consistent.

C. It will be consistent, unbiased, but not efficient.

D. It will be consistent but not unbiased.

The correct answer is A.

In the presence of near multicollinearity, the OLS estimator will still be consistent, unbiased, and efficient. This happens because none of the four (Gauss-Markov) assumptions of the CLRM have been violated.
(These assumptions are: Linearity, non-collinearity, exogeneity, and homoscedasticity)

Note: Near multicollinearity is defined as the situation where there is a high, but not perfect, the correlation between two or more of the explanatory variables. However, there is no definitive answer as to how big the correlation has to be before it is defined as "high"

Q.3349 Which of the following statements is/are correct?

I. ***Homoskedasticity means that the variance of the error terms is constant across all observations in the sample***
II. ***Heteroskedasticity means that the variance of error terms varies over the sample***
III. ***The presence of conditional heteroskedasticity leads to biased standard error estimates***

A. Only I is correct.

B. Only II and III are correct.

C. All statements are correct.

D. None of the statements is correct.

The correct answer is C.

All statements are correct
If the variance of the residuals is constant across all observations in the sample, the regression is said to be homoskedastic. When the opposite is true, the regression is said to exhibit heteroskedasticity, i.e., the variance of the residuals is not the same across all observations in the sample. The presence of conditional heteroskedasticity poses a major problem: it introduces a bias into the estimators of the standard error of the regression coefficients. As such, it understates the standard error.

Q.3780 When is a set of data termed as homoscedastic? When:

A. the variance of the error terms is the same across all the observations.

B. the observations are i.i.d. random variables.

C. the variance of the errors varies with the independent variables.

D. None of the above.

The correct answer is A.

When developing a regression model, homoscedasticity is a model's property where the model error has a constant variance.

Q.3782 Assume that you want to test for heteroskedasticity in a model with one explanatory variable, using a sample size of 100. What is the value of R^2 ***at which the null hypothesis will be rejected at a 5% level of significance?***

A. 0.0399

B. 0.0112

C. 0.0563

D. 0.0599

The correct answer is D.

We use the White test statistic calculated as nR^2 ***where*** R^2 ***is calculated in the second regression. The test statistic has a*** $\chi^2_{\frac{k(k+3)}{2}}$ ***(chi-distribution), where k is the number of explanatory variables in the first-step model.***

In this case, we have one explanatory variable so that $\chi^2_{\frac{k(k+3)}{2}} = \chi^2_2$ ***whose critical value at 5% size is 5.99.***

Therefore, we would not reject the null hypothesis

$$\Rightarrow nR^2 = 5.99$$

$$R^2 = \frac{5.99}{100} = 0.0599$$

A detailed breakdown of the chi-square test for heteroskedasticity

R^2 ***refers to the coefficient of determination. The question is built upon the chi-square test for heteroskedasticity, which follows the following steps:***

I. ***Regress*** Y ***on*** Xs ***(independents) and generate squared residuals***
II. ***Regress squared residuals on*** Xs ***(or a subset of*** Xs***) (This is what is called the second regression).***
III. ***Calculate the*** R^2 ***for the model in step 2 and use it to calculate the chi-squared test statistic:*** $\chi^2 = nR^2$
IV. ***The chi-squared statistic is compared to its critical value with*** $[k \times \frac{(k+3)}{2}]$

degrees of freedom, where k ***= number of independent variables.***

V. ***If the calculated*** χ^2 ***> critical*** χ^2***, we reject the null hypothesis of no conditional heteroskedasticity***

Q.3783 A regression model is estimated as:

$$Y_i = 3 + 1.5X_{1i} - 2X_{2i} + \epsilon_i$$

What is the value of $\hat{\beta}_1$ ***if the model is reduced to*** $Y_i = \alpha + \hat{\beta}_1 X_1 + \epsilon_i$ ***given that*** $\rho_{X_1X_2} = 0.7$ ***,*** $\sigma^2_{X_1} = 25$ ***and*** $\sigma^2_{X_2} = 36$***?***

A. -0.45

B. 0.67

C. -0.18

D. 0.23

The correct answer is C.

Recall the formula of the omitted variable.Given large samples sizes, the OLS estimator $\hat{\beta}_1$ ***converges to:***

$$\beta_1 + \beta_2\delta$$

Where:

$$\delta = \frac{Cov(X_1, X_2)}{Var(X_1)}$$

So, we need to calculate the value of δ***. Recall also that:***

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

Where

ρ_{XY} ***= the correlation between X and Y***

σ_X ***=standard deviation of X***

σ_Y ***=standard deviation of Y***

$Cov(X,Y)$ ***=covariance between X and Y***

Using this analogy,

$$\Rightarrow \delta = \frac{Cov(X_1, X_2)}{Var(X_1)} = \frac{\rho_{X_1X_2}\sigma_{X_1}\sigma_{X_2}}{Var(X_1)} = \frac{0.7 \times \sqrt{25} \times \sqrt{36}}{25} = 0.84$$

So,

$$\hat{\beta}_1 = \beta_1 + \beta_2\delta = 1.5 + (-2 \times 0.84) = -0.18$$

Note that β_1 ***and*** β_2 ***are obtained from the original model***

Q.3785 Given a model with two independent variables: $Y_i = Y_i = \alpha + \beta_1 X_1 i + \beta_2 X_2 i + \epsilon_i$, ***what is the value of the correlation coefficient between*** X_1 ***and*** X_2 ***so that the value of the variance inflation factor is 15?***

- ***A. 0.9595***
- ***B. 0.9654***
- ***C. 0.9661***
- ***D. 0.4563***

The correct answer is C.

Recall that the VIF is given by:

$$\frac{1}{1 - R_j^2}$$

Where R_j^2 ***measures how other variables explain a variable j. But in this case, we have two explanatory variable such that:***

$$R_j^2 = \rho_{X_1 X_2}^2$$

Thus:

$$\frac{1}{1 - R_j^2} = \frac{1}{1 - \rho_{X_1 X_2}^2} = 15$$
$$\therefore \rho_{X_1 X_2} = \sqrt{1 - \frac{1}{15}} = 0.9661$$

Q.3786 Consider the following data sets (We are using a small sample size for illustration purposes. In an exam situation, it might involve large sample sizes)

Y	X_1	X_2
−2	−0.41	−0.01
−0.11	0.40	−1.2
−1.68	−0.86	−0.91
−0.36	1.69	0.37
−0.08	0.46	−0.64
−0.74	1.40	−1.09

What are the estimated values of the parameters ($\hat{\alpha}$ and $\hat{\beta}_1$) ***in the model:***

$$Y = \alpha + \beta_1 X_1$$

A. $\hat{\alpha}$=-1.080 ,$\hat{\beta}_1$=0.5633

B. $\hat{\alpha}$=-1.280 ,$\hat{\beta}_1$=0.3433

C. $\hat{\alpha}$=-1.5797 ,$\hat{\beta}_1$=0.6633

D. $\hat{\alpha}$=-1.780 ,$\hat{\beta}_1$=0.9933

The correct answer is A.

We need to use the standard formula:

$$\hat{\beta}_1 = \frac{Cov(Y, X_1)}{Var(X_1)}$$

We need to calculate:

$$Cov(Y, X_1) = \frac{1}{(n-1)} \sum_{i=1}^{n} (Y - \bar{Y})(X_1 - \bar{X}_1)$$

Consider the following table:

Y	X_1	X_2	$Y-\bar{Y}$	$X_1-\bar{X}_1$	$(Y-\bar{Y})(X_1-\bar{X}_1)$
−2	−0.41	−0.01	−1.17167	−0.85667	1.003728
−0.11	0.4	−1.2	0.718333	−0.04667	−0.03352
−1.68	−0.86	−0.91	−0.85167	−1.30667	1.112844
−0.36	1.69	0.37	0.468333	1.243333	0.582294
−0.08	0.46	−0.64	0.748333	0.013333	0.009978
−0.74	1.4	−1.09	0.088333	0.953333	0.084211
Mean= −0.82833	Mean= 0.446667	Mean= −0.58		Total	2.759533

$$Cov(Y, X_1) = \frac{1}{6-1}(2.7595) = 0.5519$$

And if you calculate correctly,

$$Var(X_1) = \frac{1}{n-1} \sum_{i=1}^{n} (X_1 - \bar{X}_1)^2 = 0.9797$$

So that:

$$\hat{\beta}_1 = \frac{0.5519}{0.9797} = 0.5633$$

And from the regression equation:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}_1\bar{X}_1 = -0.82833 - 0.5633 \times 0.446667 = -1.079937 \approx -1.080$$

Q.3787 Consider the following data sets (We are using a small sample size for illustration purposes. In an exam situation, it might involve large sample sizes)

Y	X_1	X_2
−2	−0.41	−0.01
−0.11	0.40	−1.2
−1.68	−0.86	−0.91
−0.36	1.69	0.37
−0.08	0.46	−0.64
−0.74	1.40	−1.09

What is the estimated regression equation

$$\hat{Y} = \hat{\alpha} + \hat{\beta}_1 X_1$$

A. $\hat{Y}$ ***=0.8967+0.9633***X_1

B. $\hat{Y}$ ***=-1.0799 +0.5633***X_1

C. $\hat{Y}$ ***=-1.8967+0.7633***X_1

D. $\hat{Y}$ ***=-1.5745+0.6633***X_1

The correct answer is B.

We need to calculate the estimated parameters first

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(Y, X_1)}{\mathrm{Var}(X_1)}$$

We need to calculate:

$$Cov(Y, X_1) = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_1 - \bar{X}_1)(Y - \bar{Y})$$

Consider the following table:

Y	X_1	X_2	$Y - \bar{Y}$	$X_1 - \bar{X}_1$	$(Y - \bar{Y})(X_1 - \bar{X}_1)$
−2	−0.41	−0.01	−1.17167	−0.85667	1.003728
−0.11	0.4	−1.2	0.718333	−0.04667	−0.03352
−1.68	−0.86	−0.91	−0.85167	−1.30667	1.112844
−0.36	1.69	0.37	0.468333	1.243333	0.582294
−0.08	0.46	−0.64	0.748333	0.013333	0.009978
−0.74	1.4	−1.09	0.088333	0.953333	0.084211
Mean= −0.82833	Mean= 0.446667	Mean= 0.58		Total	2.759533

$$Cov(Y, X_1) = \frac{1}{6-1}(2.7595) = 0.5519$$

And if you calculate correctly,

$$Var(X_1) = \frac{1}{n-1} \sum_{i=1}^{n} (X_1 - \bar{X}_1)^2 = 0.9797$$

So that:

$$\hat{\beta}_1 = \frac{0.5519}{0.9797} = 0.5633$$

And from the regression equation:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}_1 \bar{X}_1 = 0.82833 - 0.5633 \times 0.446667 = -1.0799$$

So that the estimated regression equation is stated as:

$$\hat{Y} = -1.0799 + 0.5633X_1$$

Q.3788 Assume that you have estimated two regression equations: $\hat{Y} = 0.5767 + 0.5633X_1$

and $\hat{Y} = 0.6767 - 0.7633X_2$ ***and that covariance between explanatory variables*** X_1 ***and*** X_2 ***is 0.603*** $(\mathrm{Cov}(X_1, X_2) = 0.603)$ ***and*** $\mathrm{Var}(X_1) = 0.874$ ***and*** $\mathrm{Var}(X_2) = 0.75$ ***What is the estimated expression the intercept*** $(\hat{\alpha})$ ***for*** $Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon_i$ ***?***

- ***A.*** $\hat{\alpha} = \hat{Y}_i - 4.875X_1i + 1.627X_2i$
- ***B.*** $\hat{\alpha} = \hat{Y}_i - 3.585X_1i + 1.907X_2i$
- ***C.*** $\hat{\alpha} = \hat{Y}_i - 2.585X_1i + 3.827X_2i$
- ***D.*** $\hat{\alpha} = \hat{Y}_i - 2.4476X_1i + 2.7312X_2i$

The correct answer is D.

This a tricky question that needs application of the omitted variables formula. Recall that if the regression model is stated as:

$$Y_i = \alpha + \beta_i X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

If we omit X_2 ***from the estimated model, then the model is given by:***

$$Y_i = \alpha + \beta_1 X_{1i} + \epsilon_i$$

Now, in large samples sizes, the OLS estimator $\hat{\beta}_1$ ***converges to:***

$$\beta_1 + \beta_2 \delta_1$$

Where:

$$\delta_1 = \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)}$$

Maintaining this line of thought, the first regression equation suggests that:

$$\beta_1 + \beta_2 \delta_1 = 0.5633$$

And the second equation suggests that:

$$\beta_2 + \beta_1 \delta_2 = -0.7633$$

Moreover,

$$\delta_1 = \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)} = \frac{0.603}{0.874} = 0.6899$$

And

$$\delta_2 = \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_2)} = \frac{0.603}{0.750} = 0.804$$

So that:

$$\beta_1 + 0.6899\beta_2 = 0.5633 \ldots. (1)$$
$$\beta_2 + 0.804\beta_1 = -0.7633 \ldots.. (2)$$

If we solve the two equations simultaneously, we get:

$$\beta_1 = 2.4476 \text{ and } \beta_2 = -2.7312$$

So that the combined regression model is:

$$Y_i = \alpha + 2.4476X_{1i} - 2.7312X_{2i}$$

So that the expression of the intercept is:

$$\hat{\alpha} = \hat{Y}_i - 2.4476X_{1i} + 2.7312X_{2i}$$

Q.3789 A financial analyst wishes to come up with Ordinary Least Squares Estimation (OLS) regression model to analyze the financial performance of a company. However, the analyst is aware that some of the explanatory variables might be excluded (and hence omitted variable bias). What is the cause of omitted variables bias?

> ***A. Omitted variable bias happens when the omitted variable is independent of the included independent variables but is not a determinant of the dependent variable.***
>
> ***B. Omitted variable bias occurs when the omitted variable is correlated with all of the included independent variables and is a determinant of the dependent variable.***
>
> ***C. Omitted variable bias occurs when the omitted variable is independent of the included independent variables and is a determinant of the dependent variable.***
>
> ***D. Omitted variable bias occurs when the omitted variable is correlated with at least one of the included independent variables and is a determinant of the dependent variable.***

The correct answer is D.

Omitted variable bias occurs when a model unnecessarily excludes one or more variables that are significant determinants of the dependent variable and are correlated with one or more of the other included independent variables. Omitted variable bias leads to an over- or under-estimation of the regression parameters (intercept and the coefficients).

Q.3790 Consider the following data sets:

Observation	Y	X
1	3.67	1.85
2	1.88	0.65
3	1.35	0.63
4	0.34	1.24
5	0.89	2.45

The regression analysis was done on the entire data set, and the regression equation was estimated as:

$$\hat{Y} = 1.4110 + 0.1512X_1$$

Additionally, first four observations were used, leading to the following estimated regression equation:

$$\hat{Y} = 0.3169 + 1.3667X_1$$

What is Cook's distance for the 5th observation?

- ***A. 3.3923***
- ***B. 1.6268***
- ***C. 0.6458***
- ***D. 1.3667***

The correct answer is B.

Cook's distance is given by:

$$D_j = \frac{\sum_{i=1}^{n} \left(\hat{Y}_i^{(-j)} - \hat{Y}_i\right)^2}{ks^2}$$

Where:

$\hat{Y}_i^{(-j)}$ ***=fitted value of*** $\hat{Y}_i$ ***when the observed value*** j ***is excluded, and the model is approximated using*** $n-1$ ***observations.***

k ***=number of coefficients in the regression model***

s^2 ***=estimated error variance from the model using all observations***

So, in our case we have:

$$\hat{Y}_i = 1.4110 + 0.1512X_1$$

And

$$\hat{Y}_i^{(-j)} = 0.3169 + 1.3667X_1$$

Now study the following table:

Oservation	Y	X	$\hat{Y}_i$	$\hat{Y}_i^{(-j)}$	$\left(\hat{Y}_i^{(-j)} - \hat{Y}_i\right)^2$	$(\hat{Y}_i - y)^2$
1	3.67	1.85	1.69072	2.845295	1.333043	3.917549
2	1.88	0.65	1.50928	1.205255	0.092431	0.137433
3	1.35	0.63	1.506256	1.177921	0.107804	0.024416
4	0.34	1.24	1.598488	2.011608	0.170668	1.583792
5	0.89	2.45	1.78144	3.665315	3.548985	0.794665
				Sum	5.252932	6.457855

Now,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{Y}_i - y)^2 = \frac{1}{4} \times 6.457856 = 1.614464$$

And:

$$\sum_{i=1}^{5}\left(\hat{Y}_i^{(-j)} - \hat{Y}_i\right)^2 = 5.252932$$

And thus:

$$D_j = \frac{5.252932}{2 \times 1.614464} = 1.626834$$

The value of the Cook's distance is greater than 1 implying that the 5th observation is an outlier.

Q.3791 What is the extraneous variable in regression diagnostics?

A. It is a variable which is eliminated to increase the effectiveness of a model.

B. It is one that is unnecessarily included in the model, whose actual coefficient and consistently approximated value is 0 in large sample sizes. If we add these variables is costly.

C. It is a variable that is included in a model in case the sample size is not appropriate.

D. None of the above.

The correct answer is B.

An extraneous variable is one that is unnecessarily included in the model, whose actual coefficient and consistently approximated value is 0 in large sample sizes. If we add these variables is costly.

Q.3792 Which of the following is least likely to be the method of handling data with heteroscedastic data?

A. Use of weighted least squares (WLS).

B. Ignoring the heteroskedasticity when approximating the parameters and then utilize the White covariance estimator in hypothesis tests.

C. Transforming the data in an attempt to remove heteroskedasticity.

D. None of the above.

The correct answer is D.

The three common methods of handling data with heteroskedastic shocks include:

I. *Ignoring the heteroskedasticity when approximating the parameters and then utilize the White covariance estimator in hypothesis tests.*

   However simple, this method leads to less accurate model parameter estimates compared to other methods that address the heteroskedasticity.

II. *Transformation of data.*

   For instance, positive data can be log-transformed to try and remove heteroskedasticity and give a better view of data. Another transformation can be in the form of dividing the dependent variable by another positive variable.

III. *Use of weighted least squares (WLS).*

   This is a complicated method that applies weights to the data before approximating the parameters. That is if we know that $\text{Var}(\epsilon_i) = w_i^2\sigma^2$ *where* w_i *is known then we can transform the data by dividing by* w_i *to remove heteroskedasticity from the errors. In other words, the WLS regresses* $\frac{Y_i}{w_i}$ *on* $\frac{X_i}{w_i}$ *such as:*

$$\frac{Y_i}{w_i} = \alpha\frac{1}{w_i} + \beta\frac{X_i}{w_i} + \frac{\epsilon_i}{w_i}$$
$$\bar{Y}_i = \alpha\bar{C}_i + \beta\bar{X}_i + \bar{\epsilon}_i$$

Q.3793 Multicollinearity occurs when several variables can significantly explain one or more independent variables. Which one of the following is most likely to be true about multicollinearity?

> ***A. Multicollinearity does pose technical problems in parameter estimation, and data modeling.***
>
> ***B. When there is multicollinearity in a model, the coefficients tend to be jointly statistically significant.***
>
> ***C. Multicollinearity can be detected using the Variance Inflation Factor, where a variable with high VIF is considered for inclusion in the model.***
>
> ***D. All of the above.***

The correct answer is B.

When there is multicollinearity in a model; the coefficients tend to be jointly statistically significant, which is evidenced by the F-statistic of the regression.

Option A is incorrect because multicollinearity does not pose any technical problems in parameter approximation but rather in data modeling.

Option C is incorrect because a variable with high VIF indicates that it is highly correlated with other variables and thus should be excluded from the model.

Q.3794 There is a bias-variance tradeoff that amounts to choosing between including irrelevant variables and excluding relevant variables. There are many methods of choosing a final model from a set of explanatory variables. Among them is the general-to-specific model. Which of the following is right about the general-to-specific (GtS) model selection?

> ***A. The appropriate variables are chosen from a pool of random variables.***
>
> ***B. The variables with the smallest t-statistics are excluded in the model.***
>
> ***C. The model is re-estimated with the random variables with the smallest t-statistic.***
>
> ***D. None of the above.***

The correct answer is B.

General- to- specific method starts by stating the large model that incorporates all appropriate variables. If there is a statistically insignificant coefficient in the estimated model, the variables with the coefficient with the smallest t-statistic are removed. The model is then readjusted to include the remaining independent variables.

Q.3795 What is the main reason the m-fold cross-validation method of model selection is mostly used in modern data science?

> ***A. It is relatively easy to execute.***
>
> ***B. The method is appropriate in modeling observations that can be used for out-of-sample predictions.***
>
> ***C. This method is suitable in large sample sizes.***
>
> ***D. All of the above.***

The correct answer is B.

The m-fold validation method chooses the model that performs the out-of-sample prediction, that is, a model that fits the observations not included in the estimation of the parameters. It picks the variables that consistently predict the dependent variable and excludes variables with small coefficients that fail to perform out-of-sample predictions.

Q.469 Study the following statements regarding information criteria:
I. Akaike's information criteria are consistent
II. Akaike's information criterion always gives model orders that are at least as large as those obtained under the Schwarz's information criterion
III. If the residual sum of squares falls after the addition of an extra term, the value of the information criterion must fall
IV. The adjusted R-squared is an example of an information criterion

Which of the above statements is/are true?

A. I, II, and III

B. II and III

C. II and IV

D. All the above

The correct answer is C.

An information criterion can be defined as a measure of model fit that balances a closer fit to a set of data and an increasing number of parameters. Clearly, the adjusted R-squared falls under such a definition, although it's rarely used as a consequence of having a very weak penalty term for adding extra parameters. It can also be proved that Akaike's information criterion always gives model orders that are at least as large as those obtained under the Schwarz's information criterion. It's important to note that if the RSS falls only by a small amount, the information criteria will rise. Akaike's information criterion is also NOT consistent.

Q.505 A time series is said to be stationary if:

> ***A. Its statistical properties including the mean and variance do not change over time.***
>
> ***B. Its mean, variance, and covariances with lagged and leading values change over time.***
>
> ***C. Its mean remains constant but variance and covariances with lagged and leading values change over time.***
>
> ***D. Its mean and variance are variables but covariances with lagged and leading values do not change over time.***

The correct answer is A.

For a time series to be stationary, the expected value of the series must be not only finite but also constant across time. This also applies to the variance, which must be finite and constant across time. Time series that are not covariance stationary have linear regression estimates that have no economic meaning.

Q.506 Financial asset return time series have one common characteristic:

> ***A. They are not weakly stationary.***
>
> ***B. They are highly correlated.***
>
> ***C. They do not exhibit any trend.***
>
> ***D. Their distributions have very thin tails.***

The correct answer is C.

A majority of asset return distributions are leptokurtic – more data points are in the tails. Their returns have no trend, either stochastic or deterministic. They are also lowly correlated.

Q.507 Which of the following characteristics apply to a white noise process?
I. Zero mean
II. Autocovariances that are constant
III. Autocovariances that are zero except at lag zero
IV. Constant variance

- ***A. I and III***
- ***B. II and III***
- ***C. I and IV***
- ***D. I, III and IV***

The correct answer is D.

A white noise must have zero mean and constant variance and no autocovariance structure except at lag zero, which is actually the variance.

Q.508 Distinguish between independent white noise and normal (Gaussian white noise).

- ***A. An independent white noise is a time series that exhibits both serial independence and a lack of serial correlation while a normal white process is a time series that's serially independent, serially uncorrelated, and is normally distributed.***
- ***B. A normal white noise is a time series that exhibits both derail independence and a lack of serial correlation while an independent white noise is a time series that's serially independent, serially uncorrelated, and is normally distributed.***
- ***C. An independent white noise is a time series with equal mean and variance while a normal white noise is a time series where the mean is not equal to the variance.***
- ***D. An independent white noise is discrete while a normal white noise is continuous.***

The correct answer is A.

In addition to being serially independent and uncorrelated, a normal white noise is normally distributed.

Q.509 Which of the following is not a characteristic describing the dynamic nature of a white noise process?

> ***A. The unconditional mean and variance must be constant for any covariance stationary process.***
>
> ***B. The absence of any correlation means that all autocovariance and autocorrelations are not zero beyond displacement zero.***
>
> ***C. Events in a white noise process do not exhibit any correlation between the past and the present.***
>
> ***D. Both conditional and unconditional means and variances are the same for an independent white noise process.***

The correct answer is B.

The lack of any correlation means that all autocovariance and autocorrelations are zero beyond displacement zero.

Displacement is the distance covered by a moving body from a central point. Zero displacements implies the time lag is zero and therefore there is zero (no) movement.

Displacement beyond zero implies the time lag is non zero (time lag = 1, 2,3..etc). At zero time lag (displacement) i.e. $\tau = 0$***, the autocorrelation will be equivalent to 1 because the series will be perfectly correlated with itself:***

$$\rho(0) = \frac{\gamma(0)}{\gamma(0)}$$

At non zero time lag, however, the autocorrelation is zero because white noise is uncorrelated.

Q.510 Which of the following statements is most likely correct regarding lag operators? Lag operators:

A. only use lagged future values.

B. are of limited use in modeling a time series.

C. consider only infinite-order polynomials

D. quantify how a time series evolves by lagging a data series.

The correct answer is D.

A lag operator represents the standard medium for representing the results of forecasting models. They quantify how a time series evolves by typically lagging present values upon past values. The lag operator (also called backshift operator) operates on an element of a time series to produce the previous element:

$$Ly_t = y_{t-1}$$

Lag operators may use finite-order polynomials and are an essential tool to model a time series

---

Q.511 Unlike structural models, pure time series models do not incorporate any explanatory variable. Which of the following is a disadvantage of pure time series models when compared to the structural models?

> ***A. They are not theoretically motivated.***
>
> ***B. They cannot produce forecasts easily.***
>
> ***C. They cannot be used when the data has a very high frequency.***
>
> ***D. It's difficult to select the most appropriate explanatory variables to include in a pure time-series model.***

The correct answer is A.

Pure time series models have very weak theoretical backing. For instance, it might be difficult to see why the current value of a stock return should be related to its past values and to the values of a random error process. It would be more theoretical to explain return fluctuations using some macroeconomic variables that influence profitability, such as the state of the economy as a whole.

Note that:

A pure time series model (univariate model) is one that strictly considers the variable in question that is, it does not include external predictors. It usually tries to estimate some kind of auto-regression, trend, and seasonality in the time series.While structural time series models are models in which the explanatory variables are functions of time with coefficients which change over time.

Q.512 The following sample autocorrelation estimates are obtained using 200 data points:

Lag	1	2	3
Coefficient	0.3	−0.15	−0.10

Compute the value of the Box-Pierce Q-statistic.

A. 12.6

B. 28.0

C. 18.2

D. 24.5

The correct answer is D.

The Box-Pierce Q-statistic is given by:

$$Q_{BP} = n\sum \rho_k^2$$

Where n ***is the sample size and*** ρ ***represents the autocorrelations.***

Therefore, $Q_{BP} = 200(0.3^2 + (-0.15)^2 + (-0.10)^2) = 24.5$

Q.513 The following sample autocorrelation estimates are obtained using 250 data points:

Lag	1	2	3
Coefficient	0.3	−0.15	−0.10

Compute the value of the Ljung Box Q statistic.

- ***A. 18***
- ***B. 31.04***
- ***C. 30***
- ***D. 35***

The correct answer is B.

The Ljung Box statistic, also known as the modified Box Pierce statistic, is a function of the accumulated autocorrelations, ρ_i***, up to time lag*** m***. It's calculated as:***

$$Q(m) = n(n+2)\sum_{i=1}^{\tau}\left(\frac{\rho_i^{2}}{n-i}\right)$$

In this case, time lag = 3

Thus,

$$Q(3) = 250 \times 252\left(\frac{0.3^2}{249}\right) + 250 \times 252\left(\frac{-0.15^2}{248}\right) + 250 \times 252\left(\frac{-0.1^2}{247}\right) = 31.04$$

Q.514 The Box Pierce and the Ljung Box Q-statistics are used to measure the degree to which autocorrelations vary from zero and to establish whether white noise might be present in a set of data. Which of the following statements is incorrect regarding these two test statistics?

> ***A. The Box Pierce test has better small sample properties compared to the Ljung Box test.***
>
> ***B. Asymptotically (as the sample size increases), the values of the two test statistics will be equal.***
>
> ***C. The Box Pierce test is sometimes oversized for small samples.***
>
> ***D. Both tests show a tendency to reject the null hypothesis of zero autocorrelations as n tends towards infinity.***

The correct answer is A.

The Ljung Box test was developed in part because the Ljung Box test has better small sample properties compared to the Box Pierce test

Option A contradicts this fact.

Q.515 All the following characteristics indicate a time series that's covariance stationary EXCEPT:

> ***A. Stability of the autocorrelation.***
>
> ***B. Stability of the mean.***
>
> ***C. Stability of the covariance structure.***
>
> ***D. A nonconstant variance.***

The correct answer is D.

The distribution of the individual observations around the mean (volatility) remains unchanged over time.

Q.516 Which of the following statements is most likely correct regarding lag operators?

A. They only use lagged future values.

B. They are of limited use in modeling a time series.

C. They consider only infinite-order polynomials.

D. They quantify how a time series evolves by lagging a data series.

The correct answer is D.

Lag operators show how a time series evolves by lagging present values upon past values. They may also use finite-order polynomials

---

Q.517 Which of the following statements are true with regard to sample partial correlations?

A. They are identical to sample autocorrelations.

B. They utilize non-linear regressions.

C. They typically fall within a one-standard-error band.

D. They differ from sample autocorrelation in the size of the dataset to which they apply.

The correct answer is D.

The sample partial correlation differs from sample autocorrelation in that it performs linear regression on a finite data series.

Q.518 The following sample autocorrelation estimates are obtained using 300 data points:

Lag	1	2	3
Coefficient	0.25	−0.1	−0.05

Compute the value of the Ljung Box Q statistic.

- ***A. 20***
- ***B. 22.74***
- ***C. 24***
- ***D. 18.45***

The correct answer is B.

$$Q(m) = n\sum_{j=1}^{h}\left(\frac{n+2}{n-j}\right)\rho_j^2 = n(n+2)\sum_{j=1}^{h}\left(\frac{\rho_j^2}{n-j}\right)$$

In this case, time lag = 3
Thus,

$$Q(3) = 300(302)\left[\frac{0.25^2}{299} + \frac{(-0.1)^2}{298} + \frac{(-0.05)^2}{297}\right] = 22.74$$

Q.519 The following sample autocorrelation estimates are obtained using 300 data points:

Lag	1	2	3
Coefficient	0.25	−0.1	−0.05

Calculate the value of the Box Pierce Q-statistic.

- ***A. 22.5***
- ***B. 22.74***
- ***C. 21.5***
- ***D. 18***

The correct answer is A.

$$Q_{BP} = n \sum \rho_k^2$$

Where n ***is the sample size and*** ρ ***represents the autocorrelations.***

Therefore,

$$Q_{BP} = 300(0.25^2 + (-0.1)^2 + (-0.05)^2) = 22.5$$

Note: Provided the sample size is large, the Box Pierce and the Ljung Box tests typically arrive at the same result.

Q.521 Assume you have an MA(1) with zero mean and 0.5 as the moving average coefficient. Determine the value of the autocovariance at lag 1.

> ***A. 0.5***
>
> ***B. 0.25***
>
> ***C. 1***
>
> ***D. It's impossible to determine the values of autocovariances without knowing disturbing variances.***

The correct answer is D.

The autocovariance at lag 1 is impacted by the variance of the disturbances. Hence, the former cannot be calculated in this case.

Q.522 Which of the following is an autoregressive model with order 2?

> ***A.*** $x_t = b_0 + b_1 x_{t-1} + \epsilon_t$
>
> ***B.*** $x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + \epsilon_t$
>
> ***C.*** $x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + b_3 x_{t-3} + \epsilon_t$
>
> ***D.*** $x_t = b_0 + b_1 + \epsilon_t$

The correct answer is B.

Given the ρ***th autoregressive process,*** $\text{AR}(\rho)$ ***has*** ρ ***lagged values of the dependent variable. In other words, the order of an AR model refers to the number of prior values used in the model. In model choice B above,*** ρ ***= 2.***

Q.524 The partial autocorrelation function is necessary for distinguishing between:

- ***A. Models linked to the ARMA family.***
- ***B. An MA and an ARMA model.***
- ***C. An AR and an ARMA model.***
- ***D. None of the above.***

The correct answer is C.

The most important use of the partial autocorrelation function is in differentiating between AR and ARMA processes. For AR processes, the pacf would be zero after p lags. For the ARMA processes, the decline in the pacf would assume a geometric form.

Distinguishing between AR and an MA; ARMA(p,q) and an MA(q) process can be achieved by ACF and therefore PACF is not required to distinguish between these processes.

Any values of $p \geq 1$ ***and*** $q \geq 1$ ***would lead to decreasing value of ACF and PACF hence it is difficult to use the ACF or the PACF to distinguish between models related to the ARMA(p,q) family.***

Q.526 Consider the time series $Y_t = 1 + 0.5t + Z_t$ ***where*** $Z_t \sim WN(0, 1)$***Determine the mean function of*** Y_t***.***

A. 0.5t

B. 1

C. 1 + 0.5t

D. 1.5

The correct answer is C.

$$\mu_Y(t) = E(Y_t) = 1 + 0.5t + EZ_t$$

But $E(Z_t) = 0$

Hence $\mu_Y(t) = 1 + 0.5t$

Q.527 If a time series is reasonably approximated as white noise, then each of the following is true EXCEPT:

A. Observations in the time series are normally distributed.

B. Serial correlations (autocorrelations) are zero.

C. In a large sample, the distribution of the sample autocorrelations is approximately normal with a variance of s^2***.***

D. In a large sample, the distribution of the sample autocorrelations is approximately normal with mean of zero.

The correct answer is A.

Weak white noise processes do NOT require Gaussian observations. They only require uncorrelated, stationary observations with zero mean. However, the variables may be i.i.d.,i.e. independent and identically distributed, in which case the process is called strict white noise but the distribution need not be normal.

Q.528 An autoregressive process of order q is considered stationary if:

- ***A. The roots of the characteristic equation lie on the unit circle.***
- ***B. The roots of the characteristic equation lie outside the unit circle.***
- ***C. The roots of the characteristic equation lie inside the unit circle.***
- ***D. The characteristic equation is of order 1.***

The correct answer is B.

An AR(q) model is stationary if all the roots of its characteristic equation lie outside the unit circle, that is, all the roots are greater than 1 in absolute value.

Suppose we have a simple AR(1) model,

$$Y_t = \alpha + \rho_1 Y_{t-1} + \epsilon_t$$

Where:

α =***the intercept***

ρ =***AR parameter***

ϵ_t =***white-noise shock***

The associated 1st order polynomial is given by: $1 - \rho_1 L$

and the associated characteristics equation is: $1 - \rho_1 z = 0$

The solution here is: $z = \frac{1}{\rho_1}$

Now if $\rho_1 > 1$***, then the AR(1) is said to be stationary.***

Q.529 Which of the following statements best explains the main setback of the moving average representation of a first-order moving average process, MA(1)?

> ***A. It does not incorporate observable shocks, so the solution is to use an autoregressive representation.***
>
> ***B. It does not show evidence of autocorrelation cutoff.***
>
> ***C. It only incorporates observable shocks, so the solution is to use an autoregressive representation.***
>
> ***D. The process is highly complicated and can only be carried out using a computer program.***

The correct answer is A.

The main challenge of a moving average representation of an MA(1) process is the fact that it attempts to estimate a variable in terms of random, unobservable white shocks. The only way to make it more useful for estimation is to invert it into an autoregressive representation because an observable item can now be used.

Q.530 The key difference between a moving average representation and an autoregressive process is that:

> ***A. An autoregressive process is never covariance stationary.***
>
> ***B. An autoregressive process shows evidence of autocorrelation cutoff.***
>
> ***C. Unlike the autoregressive process, a moving average representation shows evidence of gradual decay.***
>
> ***D. A moving average representation shows evidence of autocorrelation cutoff.***

The correct answer is D.

The main differentiator between a moving average (MA) representation and an autoregressive (AR) process is that while the AR process shows a gradual decay in autocorrelations, the MA process shows abrupt autocorrelation cutoff

---

Q.531 Consider the following statements regarding the modeling of seasonal data:
I. Both the autoregressive process and the autoregressive moving average process include lagged terms and are therefore appropriate for a relationship in motion
II. Both the autoregressive process and the autoregressive moving average process specialize in capturing only the random movements in data

- ***A. I and II are both correct.***
- ***B. Only II is correct.***
- ***C. Only I is correct.***
- ***D. I and II are both incorrect.***

The correct answer is C.

Both the autoregressive process and the autoregressive moving average process include lagged terms and are therefore appropriate for a relationship in motion. The process best suited for capturing only the random movements in data is the moving average representation.

Q.532 Consider the following statements regarding an autoregressive moving average (ARMA) process:
I. The process combines the lagged unobservable random shock characteristic of the MA process with the observed lagged time series characteristic of the AR process
II. The process involves gradually-decaying autocorrelations
Which of the above statement(s) is /are correct?

- ***A. Only I is correct.***
- ***B. Only II is correct.***
- ***C. Both I and II are correct.***
- ***D. Both I and II are incorrect.***

The correct answer is C.

The ARMA process forms an important part of time series analysis since it captures a very robust picture of the variable being estimated thanks to the inclusion of lagged random shocks and lagged observations. It also combines the lagged unobservable random shock characteristic of the MA process with the observed lagged time series characteristic of the AR process.

Q.533 What is the purpose of a q^{th}-order moving average process?

> ***A. To add a fifth error term to an MA(1) process.***
>
> ***B. To add a third error term to an MA(1) process.***
>
> ***C. To add as many additional lagged variables as needed so as to produce a robust set of estimates for the time series.***
>
> ***D. To invert the moving average representation and make it more useful.***

The correct answer is C.

More lagged operators usually provide a set of estimates that's more robust.

Q.534 An analyst is charged with developing a model to analyze employment data from a developing nation. Which of the following models would be the most appropriate?

> ***A. A moving average process.***
>
> ***B. An autoregressive process.***
>
> ***C. Both AR and ARMA models.***
>
> ***D. None of the above.***

The correct answer is C.

Employment data follows some pattern of seasonality and autocorrelations would decay gradually, making both the AR and ARMA models appropriate for the analysis. The moving average model would be most appropriate if the autocorrelation were cut off abruptly.

Q.3367 The following sample autocorrelation estimates have been obtained using 200 data points:

Lag	1	2	3	4
Coefficient	0.15	−0.14	−0.1	−0.08

What is the value of the Box-Pierce Q-statistic?

> ***A. 1.05***
>
> ***B. 17.1***

C. 11.7

D. 12.5

The correct answer is C.

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

- ***T = Sample size***
- $\hat{\rho}^2(\tau)$ ***= Sample auto correlation function for*** τ ***lags***
- ***m = number of lags under observation***

$$Q_{BP} = 200(0.15^2 + -0.14^2 + -0.1^2 + -0.08^2) = 11.7$$

Q.3368 A covariance stationary time series must satisfy which of the following requirements?

*I. **The expected value of the time series must be constant and finite in all periods***
*II. **The variance of the time series must be constant and finite in all periods***
*III. **The covariance of the time series with itself for a fixed number of periods in the past or future must be constant and finite in all periods***

Which of them is correct?

A. Only III

B. I, II and III

C. I and III

D. II and III

The correct answer is B.

To be covariance stationary, a time series has to satisfy the following three conditions:

1. ***Constant and finite expected value. The expected value of the time series should be constant over time.***
2. ***Constant and finite variance. The time series volatility around its mean (i.e., the distribution of the individual observations around the mean) should not change over time.***
3. ***Constant and finite covariance between values at any given lag. The covariance of the time series with leading or lagged values of itself is constant.***

Q.3370 Each of the following is a requirement for a series to be covariance stationary (aka, weak stationarity), EXCEPT:

> ***A. The mean of the series is stable over time, i.e.,*** $Ey_t = \mu \{vs. \mu_t\}$
>
> ***B. The autocovariance at displacement (0) is finite***
>
> ***C. The autocovariance depends on time*** (t)***, but does not depend on the displacement*** (π)
>
> ***D. The covariance structure of the series is stable over time***

The correct answer is C.

A time series, y_t***, is said to be covariance stationary if it meets the following conditions:***

1. ***Constant mean***

$$Ey_t = \mu \{vs. \mu_t\}$$

   [for all t***. That is, for each*** t***,*** y_t ***is drawn from a population with the same mean]***

2. ***Stable Covariance***

   The covariance structure should be stable during the observation period and the autocovariance can be written as: $\gamma(t, \tau) = \gamma(\tau)$

   The autocovariance must depend only on the displacement factor, τ ***, and not on*** t***.***

   when $\tau = 0$***, the autocovariance is equivalent to the variance of the series itself.***

3. ***Stable Autocovariance:***

   The third condition for covariance stationarity is that the autocovariance function is stable at all displacements. The resulting autocovariance at zero displacement, i.e., when $\tau = 0$***, should be finite.***

Q.3371 Which of the following statements is (are) correct?

I. ***The Q-statistic measures the degree to which autocorrelations vary from zero and whether white noise is present in a dataset***
II. ***The Box-Pierce Q-statistic represents the weighted sum of squared autocorrelations***
III. ***The Ljung-Box Q-statistic represents the sum of squared autocorrelations***

A. Only I

B. I, II and III

C. Only II

D. II and III

The correct answer is A.

Statement I is correct. The Q-statistic measures the degree to which autocorrelations vary from zero and whether white noise is present in a dataset
Statement II is incorrect. The Box-Pierce Q-statistic represents the sum of squared autocorrelations.

Statement III is incorrect. The Ljung-Box Q-statistic represents the weighted sum of squared autocorrelations

---

Q.3372 The first-order moving average MA(1) process has zero mean and constant variance defined as:

$$y_t = \varepsilon_t + 0.2\varepsilon_{t-1}$$

Based on the above assumption, the autocorrelation can be deduced as:

- ***A. 0.1923***
- ***B. 0.2000***
- ***C. 0.0400***
- ***D. None***

The correct answer is A.

$$\rho = \frac{\theta}{1+\theta^2} = \frac{0.2}{1+0.2^2} = 0.1923$$

Q.3373 The first-order autoregressive AR(1) is defined as:

$$y_t = \varepsilon_t + 0.25y_{t-1}$$

Using the Yule-Walker equation, compute the autocorrelation of the AR(1).

- ***A. 0.50***
- ***B. 0.25***
- ***C. 0.0625***
- ***D. None***

The correct answer is B.

From the Yule-Walker equation,

$$\rho_t = \varphi^t = 0.25^1$$

The first period autocorrelation = 0.25

Q.3374 Assume the shock in a time series is approximated by Gaussian white noise. Yesterday's realization, $y_{(t)}$ ***was 0.015 and the lagged shock was -0.160. Today's shock is 0.170.***
If the weight parameter theta, θ***, is equal to 0.70, determine today's realization under a first-order moving average, MA(1), process.***

- ***A. 0.254***
- ***B. 0.075***
- ***C. 0.062***
- ***D. 0.058***

The correct answer is D.

Today's shock = ε_t;
Yesterday's shock = ε_{t-1};
Today's realization = y_t;
Yesterday's realization = y_{t-1}
The MA(1) is given by:

$$\begin{aligned} y_t &= \varepsilon_t + \theta\varepsilon_{t-1} \\ &= 0.170 + 0.7(-0.160) = 0.058 \end{aligned}$$

Q.3375 Consider the following AR(1) model with the disturbances having zero mean and unit variance

$$y_t = 0.2 + 0.3y_{t-1} + u_t$$

The (unconditional) variance of y ***will be given by:***

- ***A. 1.1500***
- ***B. 1.0989***
- ***C. 0.2198***
- ***D. 0.2145***

The correct answer is B.

The (unconditional) variance of an AR(1) process is given by the variance of the disturbances divided by (1 minus the square of the autoregressive coefficient), which in this case is $1/(1-0.3^2) = 1.0989$

Q.3376 An analyst is analyzing the sales and fitting time series model and found that sales are periodically spiked in autocorrelations as they gradually decay. Such behavior is most likely indicative of:

- ***A. Seasonality in sales data***
- ***B. A regime change structural sales data series***
- ***C. A structural shift in sales data series***
- ***D. A differencing lag***

The correct answer is A.

In the presence of significant seasonality, the autocorrelation plot should show spikes at lags equal to the period. For monthly data, for instance, a seasonality effect would yield significant peaks at lag 12, 24, 36, 48, and so forth.

Q.3960 The AR(2) model is defined as: $Y_t = 0.4 + 1.5Y_{t-1} - 0.7Y_{t-2} + \epsilon_t$ ***where*** ϵ_t ***is a white noise. What is the long-term mean of the time series?***

- ***A. 3.0***
- ***B. 1.0***
- ***C. 2.1***
- ***D. 2.0***

The correct answer is D.

We know that for an AR(p) model $Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-1} + \ldots + \beta_p Y_{t-p} + \epsilon_t$***, the mean is given by***

$$E(Y_t) = \frac{\alpha}{1 - \beta_1 - \beta_2 - \ldots - \beta_p}$$

So, in this case:

$$E(Y_t) = \frac{\alpha}{1 - (\beta_1 + \beta_2)} = \frac{0.4}{1 - (1.5 - 0.7)} = 2$$

Q.3962 The lag operator is applied to the AR times series as follows:

$$(1-0.2L)(1-0.6L^4)Y_t = \epsilon_t$$

What is the resulting time series?

- ***A.*** $0.2Y_{t-1} + 0.6Y_{t-4} - 0.12Y_{t-5} + \epsilon_t$
- ***B.*** $Y_{t-2} + 0.6Y_{t-4} - 0.12Y_{t-5} + \epsilon_t$
- ***C.*** $Y_{t-2} + 0.6Y_{t-4} - 0.12Y_{t-4} + \epsilon_t$
- ***D.*** $0.1Y_{t-1} + 0.6Y_{t-4} - 0.2Y_{t-5} + \epsilon_t$

The correct answer is A.

This question requires the application of the properties of the Lag operator. Recall that:

$$LY_t = Y_{t-1}$$

And that the lag operator is multiplicative property. So,

$$\begin{aligned}(1-0.2L)(1-0.6L^4)Y_t &= (1-0.6L^4-0.2L+0.12L^5)Y_t \\ &= Y_t - 0.6Y_tL^4 - 0.2Y_tL + 0.12Y_tL^5 = \epsilon_t \\ &= Y_t - 0.6Y_{t-4} - 0.2Y_{t-1} + 0.12Y_{t-5} = \epsilon_t\end{aligned}$$

Rearranging we get:

$$Y_t = 0.2Y_{t-1} + 0.6Y_{t-4} - 0.12Y_{t-5} + \epsilon_t$$

Q.3963 The MA(2) model is defined as $Y_t = 0.1 + 0.8\epsilon_{t-1} + 0.16\epsilon_{t-2} + \epsilon_t$. ***What is the corresponding lag polynomial?***

> ***A.*** $\epsilon_t(0.4L^2 + 0.16L^2 + 3)$
>
> ***B.*** $\epsilon_t(0.5L + 0.15L^2 + 3)$
>
> ***C.*** $0.1 + \epsilon_t(0.8L^2 + 0.16L^3 + L)$
>
> ***D.*** $\epsilon_t(0.2L + 0.14L^2 + 2)$

The correct answer is C.

Recall that

$$LY_t = Y_{t-1}$$

So that

$$\begin{aligned} LY_t &= L(0.1) + 0.8L(\epsilon_{t-1}) + 0.16L(\epsilon_{t-2}) + L(\epsilon_t) \\ &= L(0.1) + 0.8L * (L\epsilon_t) + 0.16L * L * L(\epsilon_t) + L(\epsilon_t) \\ &= 0.1 + \epsilon_t(0.8L^2 + 0.16L^3 + L) \end{aligned}$$

Note: Note that 0.1 is a constant and hence is not affected by the lag operator.

Q.3964 The sample autocorrelations for a time series are estimated to be $\hat{\rho}_1 = 0.20, \hat{\rho}_2 = -0.03$ and $\hat{\rho}_3 = 0.05$ ***from a sample size of 90. What is the value Ljung-Box Q statistic?***

- ***A. 3.45***
- ***B. 3.87***
- ***C. 20.54***
- ***D. 4.04***

The correct answer is D.

The Ljung Box statistic, also known as the modified Box Pierce statistic, is a function of the accumulated autocorrelations, ρ_i, up to time lag m. It's calculated as:

$$Q(m) = n(n+2)\sum_{i=1}^{\tau}\left(\frac{\rho_i{}^2}{n-i}\right)$$

In this case, time lag = 3
Thus,

$$Q(3) = 90 \times 92\left(\frac{0.2^2}{89}\right) + 90 \times 92\left(\frac{-0.03^2}{88}\right) + 90 \times 92\left(\frac{0.05^2}{87}\right) = 4.04$$

Q.3965 An investment analyst wishes to forecast the future returns based on the prevailing interest rate then. The analyst chooses AR times series to model the monthly interest rates movement over 20 years. The equivalent AR(1) model has an intercept of 0.24 and an AR parameter of 0.65. What is the mean-reverting value of the times series used by the analyst?

- ***A. 0.69***
- ***B. 0.56***
- ***C. 0.65***
- ***D. 0.54***

The correct answer is A.

Recall that a stationary time series tend to revert to its historical mean. So the mean-reverting level of the AR series is its mean:

$$\lim_{\tau\to\infty} ?E_T(Y_{T+\tau}) = E(Y_t)$$

$$\Rightarrow \text{Mean-reverting level} = \frac{0.24}{1-0.65} = 0.6857$$

Reading 22: Nonstationary Time Series

Q.470 Forecasting involves using sample data to predict future movements. Which of the following is correct regarding forecasting?

A. Forecasts are only possible in the presence of time-series data.

B. Forecasts will always improve whenever the number of parameters is increased.

C. As the number of variables incorporated in a regression equation increases, the risk of over-fitting the in-sample data reduces.

D. In-sample forecasting ability is a very poor test of model appropriateness and adequacy.

The correct answer is D.

In essence, using in-sample data to estimate a model and evaluate the forecasts actually amounts to cheating. Fitting a larger model would be an easy way to improve the accuracy of the in-sample forecasts. It would be akin to guessing which playing card your poker opponent has picked when you have already seen it.

Q.472 A Financial Risk Manager exam candidate suggests that a model based on financial theory is likely to lead to a high degree of out-of-sample forecast accuracy. Which of the following best explains why the candidate is correct?

A. A solid financial background significantly increases the chances of the model working in the out-of-sample period as well as for the sample data used to estimate the model's parameters.

B. A financial background increases the chances of use of authentic input data.

C. Financial theory incorporates industry-wide variables.

D. Financial theory would be easy to understand and research on.

The correct answer is A.

A model based on a solid financial background is likely to bring about good, realistic forecasts since there are high chances that the model will work in the out-of-sample period as well as for the sample data used to estimate the model's parameters.

Q.477 Joel Matip, FRM, is running a regression model to forecast in-sample data. He's worried about data mining and over-fitting the data. The criterion that provides the highest penalty factor based on degrees of freedom is the:

A. Schwarz information criterion.

B. Akaike information criterion.

C. Unbiased mean squared error.

D. Mean squared error.

The correct answer is A.

The highest penalty factor is found under the Schwarz information criterion. MSE does not penalize the regression model based on the increased number of parameters, k. The penalty factors are:

$T * (T - k)$, $e^{\frac{2k}{T}}$***, and*** $T^{\frac{k}{T}}$ ***for*** s^2, AIC***, and*** SIC ***respectively.***

Q.479 Which of the following criteria is most consistent?

A. MSE criterion

B. Akaike's information criterion

C. Schwarz information criterion

D. None of the above

The correct answer is C.

The SIC is considered the most consistent information criterion. It has the greatest penalty factor for degrees of freedom.

Q.485 Define trend as used in business and economics.

> ***A. A curve that represents the change in a series of data points collected over a given period of time.***
>
> ***B. A straight line extrapolated from past and present values to forecast future values of a given variable.***
>
> ***C. A pattern of gradual change in output as a result of data points moving in a given direction over time, and which can be represented by a line, curve, or graph.***
>
> ***D. A pattern of gradual change in output as a result of data points moving in a positive direction over time, and which can be represented by a line, curve, or graph.***

The correct answer is C.

"Trend" refers to a gradual change in the output of a random variable as a result of a tendency of data points of the variable to move in a given direction (positive or negative) over time. The trend can take the form of a line or even a nonlinear form such as an exponential form.

Q.486 An FRM exam candidate studies labor participation rate for males 18 years and over, and obtains the following linear model:

$$P_t = \beta_0 + \beta_1 t$$

Where p_t ***stands for participation rate and*** t ***= time in years. The fact that*** β_1 ***(the regression slope) is positive means:***

> ***A. Participation rate decreases with increase in time.***
>
> ***B. Participation rate increases with time.***
>
> ***C. Time is the only explanatory variable.***
>
> ***D. As males grow older, less and less of them engage in labor.***

The correct answer is B.

β_1 ***is the regression slope and represents the change in*** p_t ***for every unit change in*** t***. The regression slope is always positive if the trend is increasing and negative if the trend is decreasing.***

Q.487 In which of the following scenarios would we expect to have a non-linear trend?

> ***A. When a variable increases at a constant rate only.***
>
> ***B. When variable increases or decreases at either a constant or non-constant rate.***
>
> ***C. When the explanatory variable increases only at discrete intervals***
>
> ***D. When a linear trend has a negative regression slope***

The correct answer is B.

A non-linear trend occurs if the rate of increase or decrease is not constant. Such a trend may appear curved. For example, a model may show an increase of 10 units at time t = 1, 7 units at t = 2, 4 units at t = 3, and so on.

A non-linear trend may also occur if the rate of increase or decrease is constant. For example, Suppose that a series increases at a constant rate of 10% and the initial value is 10. So the series will be - 10,11, 12.1, 13.31...

Q.488 Suppose we have the following linear trend model which holds for any time t***:***

$$Y_t = \beta_0 + \beta_1 TIME_t + \epsilon_t$$

Assuming that ϵ ***is independent zero-mean random noise, which of the following accurately represents the model at time*** $T+h$***, if the forecast is made at time*** T ***?***

> ***A.*** $Y_t = \beta_0 + \beta_1 TIME_{(T+h)} + \epsilon_t$
>
> ***B.*** $Y_{(T+h,T)} = \beta_0 + \beta_1 TIME_T$
>
> ***C.*** $Y_{(T+h,T)} = \beta_0 + \beta_1 TIME_{(T+h)}$
>
> ***D.*** $Y_{(T+h,T)} = \beta_0 + \beta_1 TIME_{(T+h)} + \epsilon_{(T+h)}$

The correct answer is C.

Since the forecast is for time $T+h$***, the model must have the subscript*** $T+h$***. If we assume that*** ϵ ***is independent zero-mean random noise, then the optimal forecast of*** ϵ_{T+h} ***is zero for any time period. The subscript "***T***" on y, next to "***$T+h$***" reminds us that the forecast was made at time*** T ***.***

Q.489 A linear trend is calculated as $T = 17.5 + 0.65t$***.Determine the trend projection for period 10.***

A. 28.15

B. 181.5

C. 82.5

D. 24

The correct answer is D.

$$T = 17.5 + 0.65 * 10 = 24$$

Q.490 A seasonal time series component:

A. Reflects variability due to natural disasters.

B. Reflects variability during a single year.

C. Reflects a regular, multi-year pattern of being above and below the trend line.

D. Reflects gradual variability over a long time period.

The correct answer is B.

A seasonal time series is a time series component that reflects variability during a single year.

Option A refers to an irregular time series.

Option C refers to a cyclical time series.

Option D: refers to a trend time series.

Q.491 Which one of the following can bring about seasonality?

> ***A. Technologies linked to the calendar.***
>
> ***B. Preferences with links to the calendar.***
>
> ***C. Social events with links to the calendar.***
>
> ***D. All of the above.***

The correct answer is D.

Seasonality arises from technologies, preferences, and social events that have links to the normal calendar year. For example, any agricultural technology that relies on weather patterns is likely to be seasonal. In addition, most people tend to prefer summer vacations to winter ones, and gasoline prices show seasonality as a result, with prices increasing significantly throughout the summer in places like the U.S. and U.K. Social events like Christmas and Eid Mubarak normally lead to spikes in retail goods prices.

Q.492 An FRM exam candidate wishes to determine the type of variation in a time series and singles out non-seasonal variation while at the same time discarding seasonal variation. The candidate's approach:

> ***A. Is appropriate since seasonality has minimal impact on most phenomena, including economic ones.***
>
> ***B. Is appropriate because seasonal variation is very easy to establish without the need for intensive statistical analysis.***
>
> ***C. Is inappropriate because seasonality may account for a large part of the variation.***
>
> ***D. Is inappropriate because non-seasonal variation is irrelevant while studying economic relationships.***

The correct answer is C.

It’s common to find economic researchers centering on non-seasonal variation at the expense of seasonality. However, disregarding the possibility of seasonal variation can be quite irresponsible because a large part of the variation in a time series could actually be seasonal.

Q.493 An analyst wishes to use seasonal dummy variables (0s and 1s) to model seasonality. Suppose there are four seasons in a year. Which of the arrangements below would represent the third season (third quarter)?

> ***A. 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, ...;***
>
> ***B. 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, ...;***
>
> ***C. 0, 0, 3, 0, 0, 3, 0, 0, 3, ...;***
>
> ***D. 0, 1, 0, 0, 1, 0, 0, 1, 0 ...;***

The correct answer is B.

Dummy variables work in a way such that it's 1 in the quarter/season of interest and 0 otherwise. At any given time, we can be in only one season.

Q.494 Which one of the following statements best explains the idea of holiday variation?

> ***A. The dates of certain holidays may change from one country to another.***
>
> ***B. The number of holidays in a year differs from one country to another.***
>
> ***C. The number of holidays varies from year to year.***
>
> ***D. The dates of certain holidays change from year to year.***

The correct answer is D.

Holiday variation has much to do with the fact that some holidays' dates change from year to year. Although they arrive at approximately the same period of the year, the exact dates differ. A good example might be the Easter holiday observed by Christians.

Q.495 Which of the following is not an example of seasonal variation?

- ***A. Flower sales***
- ***B. Sales of Suntan oil***
- ***C. Use of electricity***
- ***D. Annual earnings of a limited liability company***

The correct answer is D.

Annual earnings may show a given trend over time, but they cannot show seasonality. However, quarterly earnings may exhibit seasonality due to certain in-year events or conditions.

Q.497 Quarter 1 sales for an automobile manufacturer were $180 million. If the quarter 1 seasonal index was 1.8, what is the estimate of annual sales for this firm?

- ***A. $400 million***
- ***B. $180 million***
- ***C. $720 million***
- ***D. $100 million***

The correct answer is A.

An index of 1.8 means that quarter 1 sales are 80% above the norm (average).

Thus, each quarter has average sales worth $\frac{\$180,000,000}{1.8} = \$100,000,000.$

Therefore, the annual estimate = 4 * Quarter average = 4 * $100,000,000 = $400,000,000.

Q.499 Which one of the following is correct? A time series data set with quarterly seasonality can be handled by:

> ***A. Using three dummy variables.***
>
> ***B. Using two dummy variables – one for each season.***
>
> ***C. Deseasonalizing the data and then applying non-seasonal methods.***
>
> ***D. All the three choices above are incorrect.***

The correct answer is A.

When the number of seasons is* s*, the number of dummy variables that should be used for analysis should be (s – 1) i.e. one fewer than the number of seasons. Thus, with four seasons, we need 3 dummy variables.

Q.500 A season in the seasonality index can only occur:

> ***A. Daily***
>
> ***B. Monthly***
>
> ***C. Weekly***
>
> ***D. All of the above***

The correct answer is D.

A season does not necessarily have to be quarterly/monthly. Sometimes it can also be day-long, week-long, or even extend for a larger period. For example, in poor countries where citizens heavily rely on salaries for survival, there's a notable surge in demand for retail goods during the fourth week of every month when most people receive their salaries.

Q.502 An analyst applies the following time series model to daily data:** R_t **= returns** D_1 **– dummy variable for Monday and zero otherwise** D_2 **– dummy variable for Tuesday and zero otherwise** D_3 **– dummy variable for Wednesday and zero otherwise** D_4 **– dummy variable for Thursday and zero otherwiseWhat is the interpretation of the parameter estimate for the intercept?

A. It's the average return for the 5 days from Monday to Friday.

B. It's the Friday deviation from the mean return for the week.

C. It's the average return on Monday.

D. It's the average return on Friday.

The correct answer is D.

For any Friday observation, please note that all of the included dummy variables will be zero. It means that the estimated value of** β_0 **will give an estimate of the average return on Friday.** $(\beta_0 + \beta_1)$ **gives the average Monday return,** $(\beta_0 + \beta_2)$ **gives the average Tuesday return and so on.

Q.523 An analyst intends to use linear regression to model the relationship between two-time series. After some testing, she finds out that one of the time series has a unit root. She should:

A. Not use linear regression if the two time series are not co-integrated.

B. Not use linear regression.

C. Perform another test on a higher level of significance before proceeding to use linear regression.

D. Only use linear regression if the time series are co-integrated.

The correct answer is B.

Since one of the time series contains a stochastic trend, i.e a unit root, then the series can not be modeled using a linear regression model. To model such a time series, the series must be differenced first in order to remove the trend.

Q.3358 Roderick Jaynes, FRM, analyzed historical sales** (S) **for over 20 years and found that sales are increasing but its growth rate over the period is relatively constant. Which model is most suitable to forecast out-of-sample sales?

A. $S_t = \beta_0 + \beta_1 \times S_{t-1}$

B. $S_t = \beta_0 + \beta_1 \times t$

C. $\ln S_t = \beta_0 + \beta_1 \times t$

D. $S_t = \beta_0 + \beta_1 \times t + \beta_2 \times t^2$

The correct answer is C.

The log-linear model exhibits best fit for data having constant growth rates.

An example of such a model, in this case, would be:

$$\ln S_t = \beta_0 + \beta_1 \times t$$

Q.3362 After finding that sales of a company vary seasonally over each quarter, an analyst is trying to incorporate such seasonality effect and build the regression model using dummy variables. Which of the following statements is (are) correct?

I. There are four dummy variables required

II. There are three dummy variables required

III. A dummy variable, takes on a value of 1 if a particular condition is true and 0 if that condition is false.

A. Only II

B. I and III

C. II and III

D. Only I

The correct answer is C.

Statement II is correct. The required dummy variables are $n-1$ ***to avoid multicollinearity. Hence 3 dummy variables are right.***
Statement III is correct. A dummy variable takes on a value of 1 if a particular condition is true and 0 if that condition is false.

Q.3363 A mortgage analyst produced a model to predict housing starts (given in thousands) within Florida in the US. The time series model contains both a trend and a seasonal component and is given by the following:

$$y_t = 0.2 \times Time_t + 10.5 + 3.0 \times D_{2t} + 5.4 \times D_{3t} + 0.7 \times D_{4t}$$

The trend component is reflected in variable $TIME_t$***, where*** (t) ***= month.***
Seasons are defined as follows :

Season	Months	Dummy
Winter	December,January and February	−
Spring	March, April and May	D_{2t}
Summer	June, July and August	D_{3t}
Fall	September, October and November	D_{4t}

The model starts in May 2019, i.e., $y_{(T+1)}$ ***refers to June 2019. What does the model predict for September 2020?***

- ***A. 23***
- ***B. 14***
- ***C. 13***
- ***D. 16***

The correct answer is B.

The model is given as:

$$y_t = 0.2 \times Time_t + 10.5 + 3.0 \times D_{2t} + 5.4 \times D_{3t} + 0.7 \times D_{4t}$$

Since we have three dummies and an intercept, quarterly seasonality is reflected by the intercept (10.5) plus the three seasonal dummy variables $(D_2, D_3$ ***and*** $D_4)$
If $y_{(T+1)}$ ***= June 2019, then Sept 2020 =*** $y_{(T+16)}$

Finally, note that Sept. falls under D_{4t}

$$y_{(T+16)} = 0.20 \times 16 + 10.5 + 0.7 \times 1 = 14.4$$

Thus, the model predicts 14 housing starts in Sept. 2020.

Q.3364 A ski resort has come up with a model to predict the number of guests (given in hundreds) checking in throughout the year. The time series model contains both a trend and a seasonal component and is given by the following:

$$y_t = 0.2 \times Time_t + 10.5 + 3.0 \times D_{2t} + 2.1 \times D_{3t} + 2.8 \times D_{4t}$$

The trend component is reflected in variable $TIME_{(t)}$***, where*** (t) ***= month.***
Seasons are defined as follows :

Season	Months	Dummy
Winter	December,January and February	−
Spring	March, April and May	D_{2t}
Summer	June, July and August	D_{3t}
Fall	September, October and November	D_{4t}

The model starts in April 2019, i.e., $y_{(T+1)}$ ***refers to May 2019. How many more guests are expected in April 2020 than in July of the same year?***

- ***A. 10***
- ***B. 30***
- ***C. 9***
- ***D. 15***

The correct answer is B.

The model is given as:

$$y_t = 0.2 \times Time_t + 10.5 + 3.0 \times D_{2t} + 2.1 \times D_{3t} + 2.8 \times D_{4t}$$

Since we have three dummies and an intercept, quarterly seasonality is reflected by the intercept (10.5) plus the three seasonal dummy variables $(D_2, D_3$ ***and*** $D_4)$
If $y_{(T+1)}$ ***= May 2019, then April 2020 =*** $y_{(T+12)}$ ***and July 2020 =*** $y_{(T+15)}$

Finally, note that April falls under $D_{(2t)}$***, while July falls under*** $D_{(3t)}$

$$y_{(T+12)} = 0.20 \times 12 + 10.5 + 3.0 \times 1 = 15.9$$
$$y_{(T+15)} = 0.20 \times 15 + 10.5 + 2.1 \times 1 = 15.6$$
$$y_{(T+12)} - y_{(T+15)} = 15.9 – 15.6 = 0.3$$

Thus, the model predicts 1,590 guests in April and 1,560 guests in July, representing a difference of 30.

Q.3366 A pure seasonal dummy model is constructed as below:

$$y_t = \sum_{i=1}^{s} \gamma_i D_{i,t} + \varepsilon_t$$

If all seasonal factors (γ_i) ***in the model are equal, it can be concluded that:***

- ***A. A seasonally adjusted time series is to be constructed***
- ***B. There is a lack of seasonality***
- ***C. There is a need for additional dummy variables***
- ***D. Both A and C***

The correct answer is B.

If all seasonal factors (γ_i) ***in the model are equal, we can conclude that there is no effect of seasonality. For example, sales could be equal in all four quarters.***

Q.3953 An investment analyst wants to fit the weekly sales (in millions) of his company by using the sales data from Jan 2018 to Feb 2019. The regression equation is defined as:

$$\ln Y_t = 5.105 + 0.044t, \quad t = 1, 2, \ldots, 100$$

What is the trend value estimate of the sales in the 90th week?

- ***A. 8,647.28 Million***
- ***B. 7,947.26 Million***
- ***C. 8,537.38 Million***
- ***D. 8,237.48 Million***

The correct answer is A.

From the regression equation, $\hat{\beta}_0 = 5.105$ ***and*** $\hat{\beta} = 0.044$***. We know that, under log-linear trend models, the predicted trend value is given by:***

$$\begin{aligned} \ln Y_t &= \hat{\beta}_0 + \hat{\beta}_1 t \\ \Rightarrow Y_t &= e^{\hat{\beta}_0 + \hat{\beta}_1 t} \\ \therefore Y_{90} &= e^{5.105 + 0.044 \times 90} \\ &= 8,647.28 \text{ Million} \end{aligned}$$

Q.3954 A financial analyst wishes to conduct an ADF test on the log of 20-year real GDP from 1999 to 2019. The result of the tests is shown below:

Deterministic	γ	δ_0	δ_1	Lags	5%CV	1%CV
None	−0.004 (−1.555)			8	−1.940	−2.570
Constant	−0.008 (−1.422)	0.010 (1.025)		4	−2.860	−3.445
Trend	−0.084 (−4.376)	0.188 (−4.110)		3	−3.420	−3.984

The output of the ADF reports the results at the different number deterministic terms (first column), and the last three columns indicate the number of lags according to AIC and the 5% and 1% critical values that are appropriate to the underlying sample size and the deterministic terms. The quantities in the parenthesis (below the parameters) are the test-statistics. What deterministic term is most preferred for this model?

- ***A. Constant***
- ***B. Trend***
- ***C. None (No deterministic components)***
- ***D. Both the constant and the trend***

The correct answer is B.

The recommended method of choosing appropriate deterministic terms is by including the deterministic terms that are significant at the 10% level. The trend deterministic is chosen because the absolute value of its test statistic is greater than 1.645. That is, $|-4.376| > |1.645|$.

Also, since the absolute value of the trend test statistic is greater than both 5% critical value (CV) and 1% CV, then there is no evidence of unit roots and thus, the null hypothesis is rejected.

Q.3955 The seasonal dummy model is generated on the quarterly growth rates of mortgages. The model is given by:

$$Y_t = \beta_0 + \sum_{j=1}^{s-1} \gamma_j D_{jt} + \epsilon_t$$

The estimated parameters are $\hat{\gamma}_1 = 6.25, \hat{\gamma}_2 = 50.52, \hat{\gamma}_3 = 10.25$ ***and*** $\hat{\beta}_0 = -10.42$ ***using the data up to the end of 2019. What is the difference between the forecasted value of the growth rate of the mortgages in the second and third quarters of 2020?***

- ***A. 24.56***
- ***B. 32.45***
- ***C. 40.27***
- ***D. 30.32***

The correct answer is C.

For the second quarter, define the set of dummy variables:

$$D_{jt} = \begin{cases} 1, & \text{for } Q_2 \\ 0, & \text{for } Q_1, Q_3 \text{ and } Q_4 \end{cases}$$

So,

$$E(\hat{Y}_{Q2}) = \beta_0 + \sum_{j=1}^{3} \gamma_j D_{jt} = -10.42 + 0 \times 6.25 + 1 \times 50.52 + 0 \times 10.25 = 40.1$$

Analogously, the dummy variables for the third quarter is defined as:

$$D_{jt} = \begin{cases} 1, & \text{for } Q_3 \\ 0, & \text{for } Q_1, Q_2 \text{ and } Q_4 \end{cases}$$

$$E(\hat{Y}_{Q3}) = \beta_0 + \sum_{j=1}^{3} \gamma_j D_{jt} = -10.42 + 0 \times 6.25 + 0 \times 50.52 + 1 \times 10.25 = -0.17$$

So that:

$$Q_2 - Q_3 = 40.10 - -0.17 = 40.27$$

Q.3956 A log trend model approximated on the interest rate (in %) movement is given as:

$$\ln Y_t = -0.1567 + 0.00134t + \hat{\epsilon}_t$$

for the last 20 years. The standard deviation of the residual is 0.0342. Assuming that the residuals are white noise, what is the point forecast of the interest rate after 3 years from now?

- ***A. 0.8589%***
- ***B. 0.8453%***
- ***C. 0.7890%***
- ***D. 0.7945%***

The correct answer is A.

The forecast under the log model is given by:

$$\begin{aligned} E_T(Y_{T+h}) &= e^{\beta_0+\beta_1(Y_{T+h})+\frac{\sigma^2}{2}} \\ &= e^{\beta_0+\beta_1(3)+\frac{\sigma^2}{2}} \\ &= e^{-0.1567+0.00134(3)+\frac{0.0342^2}{2}} = 0.8589\% \end{aligned}$$

Q.3957 A log trend model is approximated on the interest rate (in %) movement in a certain market based on data from 2000 until 2020. The estimated model is given as:

$$\ln Y_t = -0.1567 + 0.00134t + \hat{\epsilon}_t$$

The standard deviation of the residual is 0.0342. Assuming that the residuals are white noise, what is the 95% confidence interval for interest rate movement 3 years from now.

- ***A. [12.030, 13.754]***
- ***B. [12.018, 13.739]***
- ***C. [-0.0584, 0.0756]***
- ***D. [11.994, 13.713]***

The correct answer is A.

In this case,

$$E_T[Y_T] = \exp(E_T[\ln Y_T] + \frac{\sigma^2}{2})$$

And the error bounds on the In are $\pm 1.96 * 0.0342$***.***

Thus the bounds-multiplier is given by $\exp(\pm 1.96 * 0.0342) = 0.935, 1.069$

$E[\ln Y_T]$ ***can be calculated as follows:***

$$E_T[\ln Y_T] = -0.1567 + 0.00134 * 2023 = 2.554$$

We also need

$$\frac{\sigma^2}{2} = \frac{0.0342^2}{2} = 0.0006$$

Thus,

$$E[Y_{2023}] = \exp(2.554 + 0.0006) = 12.866$$

Thus, the 95% confidence interval for interest rate movement 3 years from now, is given by:

$$95\%CI_{2023} = [0.935 * 12.866, 1.069 * 12.866] = [12.030, 13.754]$$

Q.3959 Which of the following statements best describes the time series with a unit-roots?

> ***A. It is a random walk time series with a drift described using AR(1) model whose lag coefficient is 1***
>
> ***B. It is a random walk time series with a drift described using AR(1) model whose lag coefficient is 0***
>
> ***C. Time series with unit roots are covariance stationary.***
>
> ***D. All of the above***

The correct answer is A.

Option B is incorrect because the lag coefficient is 1.
Option C is incorrect because unit roots are the cause of non-stationarity.

Reading 23: Measuring Return, Volatility, and Correlation

Q.550 Define volatility in the context of risk management.

> ***A. The level of specific risk of an asset.***
>
> ***B. The standard deviation of the return provided by the variable per unit time, when the return is expressed using continuous compounding.***
>
> ***C. The standard deviation of returns on an asset per unit time.***
>
> ***D. The variance of the return provided by the variable per unit time, when the return is expressed using continuous compounding.***

The correct answer is B.

Volatility refers to the standard deviation of returns on an asset per unit time when the return is expressed using continuous compounding. For example, when pricing options, volatility is measured per year.

Q.551 The price of an asset is $80, and its daily volatility is 2%. Determine the one-standard-deviation move in the asset's price over a one-day period.

> ***A. 1***
>
> ***B. 2***
>
> ***C. 1.6***
>
> ***D. 16***

The correct answer is C.

The one-standard-deviation move is simply 80 * 0.02 = $1.6

Q.552 If an asset has volatility equal to s, its variance rate is equal to:

> ***A. S – 1***
>
> ***B.*** S^2
>
> ***C.*** $\sqrt{s}$
>
> ***D. S%***

The correct answer is B.

Variance rate is defined as the square of volatility. The easiest way to differentiate the two might be to recall that volatility represents the standard deviation.

Q.553 Distinguish between historical and implied volatility.

> ***A. Historical volatility measures the standard deviation of past price movements while implied volatility gives an estimate of future volatility in the price of the asset.***
>
> ***B. Historical volatility measures the standard deviation of past price movements while implied volatility is the immeasurable volatility in the future price of an asset.***
>
> ***C. Historical volatility is the volatility of an asset that has been recorded as at present, while implied volatility is the future volatility.***
>
> ***D. Historical volatility measures the total standard deviation of past price movements while implied volatility gives the historical volatility beyond a given benchmark.***

The correct answer is A.

Implied volatility is the estimated volatility of a security's price. It gives clues as to how the markets expect the price to move in the future. It is a probabilistic estimate of future prices, not an accurate prediction. In other words, there are no guarantees price movements will follow the predicted pattern.

Q.554 Suppose that we know from experience that $\alpha = 4$ ***for a certain financial variable and we observe that*** $P(v > 20) = 0.01$***. Apply the Power Law and find the probability that*** $v > 10$***.***

A. 22

B. 16,000

C. 20

D. 0.16

The correct answer is D.

The Power Law states that for many variables, it's approximately true that the value of the variable, v***, is such that:***

$$Prob(v > x) = Kx^{-\alpha}$$

Therefore, $0.01 = K * 20^{-4}$

k=1600

Thus, $P(v > 10) = 1600 * 10^{-4} = 0.16$

Q.555 Which of the following statements is false?

> ***A. Correlation measures the strength of the linear relationship between two variables.***
>
> ***B. Mathematically, we determine correlation by dividing the covariance between two random variables by the product of their standard deviations.***
>
> ***C. Two variables are independent if the knowledge of one variable does not impact the probability distribution of the other variable.***
>
> ***D. If two variables have a correlation of zero, this implies they also have zero dependence between them.***

The correct answer is D.

Just because two variables have a correlation of zero does not mean that there's zero dependence between them. It means that the two have zero linear relationship but there could still be some kind of a non-linear relationship between them.

Q.3381 Assumed that asset prices are normally distributed. The expected value of an asset price is $80 with daily volatility of 2%. Compute the 95% confidence interval of the asset price at the end of 4 days.

> ***A. 80 ± 2.000***
>
> ***B. 80 ± 3.200***
>
> ***C. 80 ± 6.272***
>
> ***D. 80 ± 3.136***

The correct answer is C.

4-day Volatility = $2 * \sqrt{4} = 4\%$
Compute one standard deviation move in dollar = $80 * 0.04 = 3.2$

95% confidence interval of asset price = $80 \pm 1.96 * 3.2 = 80 \pm 6.272$

Q.3382 A zero correlation between two variables indicates that:

- ***A. They are independent of each other***
- ***B. They do not have any linear relationship between them***
- ***C. They are linearly dependent on each other***
- ***D. They have a negative linear relationship***

The correct answer is B.

Correlation measures the linear relationship. A zero correlation indicates a linear relationship does not exist but other relationships could exist.
$\text{Corr}(X, Y)$ ***of a symmetric parabola is zero.***
$\text{Corr}(X, Y)$ ***of a symmetric*** V ***shape graph is zero.***

Therefore, a zero correlation does not indicate independence but indicates the absence of linear dependence.

Q.4036 After arranging the data of a portfolio comprising of two assets X and Y from the period 2012 to 2016, it is found that the number of concordant data pairs is 2 and the number of discordant pairs is 4. On the basis of this information, which of the following is closest to the Kendall τ?

- ***A. -0.13***
- ***B. -0.20***
- ***C. -0.08***
- ***D. 0.08***

The correct answer is B.

Since the data is analyzed from 2012 to 2016, we have five observation pairs and therefore $5 * \frac{(5-1)}{2} = 10$ ***combinations of pairs to evaluate. By using the Kendall equation:***

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

$$\tau = \frac{(2-4)}{10} = -0.2$$

Q.4037 Which of the following statements is (are) correct?
I. The Spearman's approach is also known as the Spearman's correlation coefficient for ranked variables.
II. Both the Spearman's and the Kendall measures are nonparametric. They both use the numerical values of the elements in the formula for calculating the correlation coefficient, not the rating of the elements.
III. For calculating Kendall's τ, finding the number of concordant pairs and number of discordant pairs is necessary.

- ***A. I, II and III***
- ***B. I and III***
- ***C. I and II***
- ***D. II and III***

The correct answer is B.

The Spearman's approach is often referred to as the Spearson correlation coefficient for ranked variables. The Kendall τ formula is:

$$\tau = \frac{(nc - nd)}{\left[\frac{n(n-1)}{2}\right]}$$

Where nc is the number of concordant data pairs and nd ***is the number of discordant pairs. Statement II is not correct because both models consider the ranking of elements instead of numerical values.***

Q.4038 A portfolio manager is trying to determine the correlation between the return of two assets. Given the following data about the yearly returns of the stocks, he decides to calculate the Kendall's τ correlation coefficient for the returns of these assets.

Year	Return of Asset X	Return of asset Y
1	3%	8%
2	1%	5%
3	−4%	6%
4	5%	2%
5	2%	9%

What is Kendall's τ correlation coefficient for the returns of the two assets?

A. -0.1

B. -0.2

C. 0

D. 0.1

The correct answer is B.

First: order the return set pairs of X and Y with respect to the set X. This is done in columns 2 and 3 of the table below.

Year	Ranked Return of X_i(Assigned same year)	Ranked Return of Y_i	Rank of X_i	Rank of Y_i
3	−4%	6%	1	3
2	1%	5%	2	2
5	2%	9%	3	5
1	3%	8%	4	4
4	5%	2%	5	1

Then, derive the ranks of X; and Y;. This is done in columns 4 and 5.
We have five observation pairs and therefore 5 x (5 - 1)/ 2 = 10 combinations of pairs to evaluate. These are:
[(1,3),(2,2)], [(1,3),(3,5)], [(1,3),(4,4)], [(1,3),(5,1)]
[(2,2),(3,5)], [(2,2),(4,4)], [(2,2),(5,1)]
[(3,5),(4,4)], [(3,5),(5,1)]
and [(4,4),(5,1)]

Concordant pairs
[(1,3),(3,5)], [(1,3),(4,4)], [(2,2),(3,5)], and [(2,2),(4,4)]
Discordant pairs
[(1,3),(2,2)], [(1,3),(5,1)], [(2,2),(5,1)], [(3,5),(4,4)], [(3,5),(5,1)], and [(4,4),(5,1)]
None of the pairs is neither concordant nor discordant

$$\text{The Kendall} \quad \tau = \frac{n_c - n_d}{\frac{n(n-1)}{2}}$$

Where:
n_c = number of concordant pairs
n_d = number of discordant pairs
n = number of time periods

$$\tau = \frac{4-6}{\frac{5(5-1)}{2}} = \frac{-2}{10} = -0.2$$

Additional Information

These are the steps to follow in order to get it right:

The first step is to arrange each set of returns in order - from the lowest to the highest.

Second, make a list of all pairs possible without being repetitive. For instance, [(1,3),(2,2)] and [(2,2),(1,3)] is just one pair, not two. The total number of non-repetitive pairs should be equal to [n *(n-1)]/2 where n is the number of observations. In this case, n = 5, so the # of pairs is [5 *(4)]/2 = 10

Concordant pairs

A pair is said to be concordant if it agrees about the relative position of X and Y. This means that if Xi > Xj, then Yi > Yj or if Xi < Xj then Yi < Yj.

Examples: [(1,3),(2,5)], [(2,2),(4,4)]

Discordant pairs

A pair is said to be discordant if it does not agree about the relative position of X and Y. This means that if Xi > Xj, then Yi < Yj or if Xi < Xj then Yi > Yj.

Examples: [(1,3),(2,2)], [(3,5),(5,1)]

Neither Concordant nor Discordant

A pair is neither concordant nor discordant if Xi = Xj or Yi = Yj

Examples: [(1,5), (2,5)], [(3,5), (3,7)]

Q.4039 A survey is conducted to find investors' perceptions of the financial market. The survey data takes the form (AAA, AA, A, ..., D). Which of the following statistical models would you apply to analyze such data?
I. The Pearson model
II. The Kendall model
III. The Spearman model

A. I and II

B. I and III

C. II and III

D. I, II and III

The correct answer is C.

The data is given in ranking order, which is actually ordinal. Therefore, both the Spearman and the Kendall models can be used to analyze the data because both of them measure ordinal relationships.

Q.4040 The application of statistical correlation models to access financial correlation is limited because of the following reasons:
I. The Spearman and the Kendall models work best with cardinal observations and consider the extreme value of outliers.
II. Both the Spearman and the Kendall approaches take the order of the elements into consideration while ignoring numerical values.
III. The Kendall t works best with only a few concordant and discordant pairs.
IV. Among all models, the Pearson approach is the best statistical model and is widely used because it measures nonlinear relationships and financial variables are mostly nonlinear.

A. I and IV

B. III and IV

C. II and III

D. II, III and I

The correct answer is C.

Statements II and III are correct because both the Spearman and the Kendall approaches measure ordinal rank correlations. The problem with applying ordinal rank correlations to cardinal observations is that ordinal correlations are less sensitive to outliers making the application of the model limited.
A problem with the Kendall t is that when many non-concordant and many non-discordant pairs occur, some have to be omitted from the calculation. Working with only a few concordant and discordant pairs can distort the Kendall t coefficient. Statement IV is incorrect because the Pearson model is used to measure the linear relationship between the variables.

Q.4041 An investor is analyzing the data of two assets X and Y for a period of 7 years. He applied all three statistical models to measure the correlation coefficient. The results were as follows:
Pearson correlation coefficient = -0.8501
Spearman correlation coefficient = -0.9
Kendall's t = -0.4

He again analyzed the same data but changed two values of asset X without affecting its rating. What would be the impact of this change on the results?

> ***A. The Spearman results would change but the results of Pearson and Kendall approaches would remain unchanged.***
>
> ***B. The Pearson and the Kendall results would change but the Spearman results would remain unchanged.***
>
> ***C. The Kendall results would change but the results of the Spearman and Pearson approaches would remain unchanged.***
>
> ***D. The Pearson results would change but the results of the Spearman and the Kendall approaches would remain unchanged.***

The correct answer is D.

Only the Pearson result would change because this particular approach considers numeric values while the Spearman and the Kendall approach only consider the rating or order of the elements.

Q.4042 Consider the following data.

Time	Price
0	100
1	98.65
2	98.50
3	97.50
4	95.67
5	96.54

What is the value of the simple return at time 4?

- ***A. 1.88%***
- ***B. -1.97%***
- ***C. -1.88%***
- ***D. 1.97%***

The correct answer is C.

The simple return is given by:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

So that:

$$R_4 = \frac{P_4 - P_3}{P_3} = \frac{95.67 - 97.50}{97.50} = -0.0188 = -1.88\%$$

Q.4043 Consider the following data.

Time	Price
0	100
1	98
2	98
3	97
4	99
5	P

If the simple return for the 5th period is 1.75%, what is the value of p?

- ***A. 99.56***
- ***B. 100.50***
- ***C. 100.73***
- ***D. 99.98***

The correct answer is C.

Using the analogy:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Then

$$0.0175 = \frac{p - 99}{99} \Rightarrow p = (0.0175 \times 99) + 99 = 100.7325$$

Q.4044 The prices of a portfolio at different times are as shown in the table below.

Time	Price
0	105.62
1	104.10
2	105.54
3	103.68
4	103.56

What is the value of continuously compounded return at time 4?

> ***A. 0.81%***
>
> ***B. -0.0081***
>
> ***C. 0.0091***
>
> ***D. -0.0012***

The correct answer is D.

The continuously compounded returns is given by:

$$r_t = \ln P_t - \ln P_{t-1}$$

So that:

$$r_4 = \ln P_4 - \ln P_3 = \ln 103.56 - \ln 103.68 = -0.0012\%$$

Q.4045 A simple return for an investment in a particular holding period is 15%. What is the equivalent continuously compounded return over the same holding period?

A. 0.1256

B. 0.1578

C. 0.1398

D. 0.1278

The correct answer is C.

Using the relationship:

$$1 + R_t = e^{r_t}$$
$$\Rightarrow r_t = \ln(1 + R_t) = \ln 1.15 = 0.1398 = 13.98\%$$

Q.4046 If the daily volatility of the price of gold is 0.3% in a given year. What is the annualized volatility of the gold price?

A. 4.67%

B. 3.56%

C. 2.56%

D. 4.76%

The correct answer is D.

Using the scaling analogy, the corresponding annualized volatility is given by:

$$\sigma_{\text{annual}} = \sqrt{252 \times \sigma^2_{\text{daily}}} = \sqrt{252 \times 0.003^2} = 0.047623 = 4.7623\%.$$

Q.4047 A financial analyst wishes to model the returns from investment using the normal distribution. The analyst approximates the skewness of the data to 0.35 and kurtosis of 3.04. The analyst performs the JB test at a 95% confidence level. What is the value of the test statistic as per the analyst's results if the sample size is 100? You might need to use the following chi-square table:

df	0.1	0.05	0.025	0.01	0.005
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838
4	7.779	9.488	11.143	13.277	14.860
5	9.236	11.070	12.833	15.086	16.750
6	10.645	12.592	14.449	16.812	18.548
7	12.017	14.067	16.013	18.475	20.278
8	13.362	15.507	17.535	20.090	21.955
9	14.684	16.919	19.023	21.666	23.589
10	15.987	18.307	20.483	23.209	25.188
11	17.275	19.675	21.920	24.725	26.757
12	18.549	21.026	23.337	26.217	28.300

A. 2.028

B. 5.991

C. 1.0256

D. 6.484

The correct answer is A.

According to the Jarque-Bera (JB) test, the test statistic is given by:

$$JB = (T-1)\left(\frac{\hat{S}^2}{6} + \frac{(\hat{k}-3)^2}{24}\right)$$

So,

$$JB = (100-1)\left(\frac{0.35^2}{6} + \frac{(3.04-3)^2}{24}\right) = 2.028$$

Recall that $JB \sim \chi_2^2$ ***so that critical value at 5% with 2 degrees of freedom (df) is 5.991 (which can be seen from a chi-square table).***

Since the test statistic is less than the critical value, we fail to reject the null hypothesis that the returns are not normal, and hence the analyst can use the normal distributions in analyzing the returns.

Q.4049 Which of the following statements is true about the simple and continuously compounded returns?

> ***A. The continuously compounded return is always less than the simple return.***
>
> ***B. The simple return is always less than the continuously compounded return.***
>
> ***C. The log return (continuously compounded return) approximates the simple return with the approximation error decreasing with an increase in simple return.***
>
> ***D. The simple return can go below -100% while the continuously compounded returns do not.***

The correct answer is A.

The log return(The continuously compounded return) is always less than the simple return, hence this contradicts option B.

Option C is incorrect because the approximation error becomes more significant as the simple return increases. This is true because the continuously compounded returns are always less than the simple returns

Option D incorrect. Simple returns can never go below the total loss! However, log returns can go below -100%.

Reading 24: Simulation and Bootstrapping

Q.570 Which one of the following is NOT a method of choosing a probability distribution for a simulation model?

> ***A. Parameter estimate technique.***
>
> ***B. Sampling technique.***
>
> ***C. Best fit technique.***
>
> ***D. Bootstrapping technique.***

The correct answer is B.

The four ways of choosing a probability distribution for a simulation model are: Bootstrapping technique, parameter estimate technique, best fit technique, and the subjective guess technique.

Q.571 Consider the following basic steps involved when conducting a Monte Carlo simulation:
I. Carry out regression and calculate the test statistic
II. Generate the data following the desired data generation process, drawing the errors from some given distribution
III. Save the test statistic or the parameter of interest
IV. Go back to stage 1 and repeat N times

Which of the following is the correct order of the steps above?

> ***A. I, II, III, IV***
>
> ***B. I, III, II, IV***
>
> ***C. II, I, III, IV***
>
> ***D. II, I, IV, III***

The correct answer is C.

When conducting a Monte Carlo simulation, the first stage entails specifying the model to use to generate the data, such as a pure time series model or even a structural model. Second, the parameter of interest is estimated through regression. For instance, it could be a coefficient of the independent variable. Third, the parameter of interest or the test statistic is saved. Finally, these steps are repeated N times to eliminate the possibility of obtaining "odd" combinations of random variables that could be unfit for the purpose at hand.

Q.572 In which of the following situations would bootstrapping be ineffective?

- ***A. Use of non-independent data.***
- ***B. Sampling with replacement.***
- ***C. If there are no outliers in the data.***
- ***D. Re-sampling from regression residuals.***

The correct answer is A.

Bootstrapping implicitly assumes that the data are independent of one another. Therefore, if there are, say, correlations in the data, bootstrapping would be ineffective.

Q.573 Consider the following statements regarding Monte Carlo (MC) simulation:
I. One of the downsides of MC simulation is that it can be computationally intensive and sometimes necessitate the use of expensive software
II. MC simulation can be applied in the presence of data that takes on the lognormal distribution
III. Mc allows for a wider variety of scenarios than those that can be deduced from historical data by itself
IV. It can only be applied if portfolios contain linear positions

Which of these statements is (are) incorrect?

- ***A. I and III***
- ***B. II only***
- ***C. IV only***
- ***D. I, II, and IV***

The correct answer is C.

MC simulations can also generate distributions for portfolios that also contain nonlinear positions. They do account for options, where the first step entails simulation of the risk factor and the second step prices the option, thereby accounting for nonlinearity.

Q.574 Construct a 95% confidence interval for the ending mutual fund capital amount where the number of simulations is 100, the mean ending capital is $200,000, and the standard deviation is $34,456.

> ***A. [$193,247, $206,753]***
>
> ***B. [$193,177.7, $200,000]***
>
> ***C. [$193, $206]***
>
> ***D. [$180,000, $220,000]***

The correct answer is A.

We need to find the 2.5th percentile and the 97.5th percentile for the t-distribution with 100 observations. The formula to apply is:

$$\bar{X} - 1.96 * \left(\frac{s}{\sqrt{N}}\right), \bar{X} + 1.96 * \left(\frac{s}{\sqrt{N}}\right)$$

Where N = 100

$$= \$200,000 - 1.96\left(\frac{\$34,456}{\sqrt{100}}\right), \$200,000 + 1.96\left(\frac{\$34,456}{\sqrt{100}}\right)$$
$$= [\$193,246.624, \$206,753.376]$$

Q.575 A researcher happens to use a very small number of replications during a Monte Carlo study. Which of the following statements will be true in this scenario?
I. Standard errors of the estimated quantities may be too large and quite unacceptable
II. The process may yield a statistic that's imprecise
III. The results of the process may be affected by unrepresentative combinations ot random draws

A. All the above

B. I and II only

C. II only

D. II and III only

The correct answer is A.

If a very small sample is used, estimation of the relationships between the variables will be inaccurate and the sampling error will also be large. "Odd" combinations ot the random draws could produce results that do not depict the system's average behavior.

Q.576 Which of the following statements is correct regarding the use of antithetic variates during a Monte Carlo simulation exercise?

A. The antithetic variates reduce the sampling error through the correlation coefficient.

B. The antithetic variates method involves taking one over each random draw and then repeating the experiment using those values as the draws.

C. The antithetic variates method reduces the variance of the simulation results.

D. All of the above.

The correct answer is D.

Antithetic variates are quite helpful during a Monte Carlo experiment since they work by reducing the number of replications required to cover the whole probability space. Covering the entire probability space may be difficult to achieve naturally. The successive replications generate outcomes that encompass the entire space ot possibilities.

Q.577 Under which of the following situations would you prefer bootstrapping to pure simulation?

> ***A. If you have a very small sample of actual data.***
>
> ***B. If the distribution of the actual data is unknown.***
>
> ***C. If the distribution of the data is known exactly.***
>
> ***D. All the above.***

The correct answer is B.

When the properties of the actual data are unknown, bootstrapping may be preferred to pure simulation since it would be very difficult to find an appropriate distribution from which to make the random draws needed under pure simulation.

Q.578 The following are advantages of simulation EXCEPT:

> ***A. It allows for the study of "what if" questions.***
>
> ***B. It saves a considerable amount of time for analysts.***
>
> ***C. It incorporates multiple relationships and serial correlations between random variables.***
>
> ***D. It's a computationally expensive method of analyzing future scenarios.***

The correct answer is D.

Sometimes the number of replications required to generate precise solutions may be very large and unachievable unless the analyst invests in the latest and most developed computer systems which obviously come with extra costs.

Q.579 Which of the following statements best explains why Monte Carlo simulations are considered exceptionally effective compared to probability trees when appraising a capital project?

> ***A. Monte Carlo simulations incorporate scenarios that span the entire probability space.***
>
> ***B. Monte Carlo simulations consume less time.***
>
> ***C. Monte Carlo simulations are embedded within most modern computer programs.***
>
> ***D. Monte Carlo simulations allow for the comparison of only the net present values that are positive.***

The correct answer is A.

The main reason why Monte Carlo simulations outshine other historical simulation methods, including the use of probability trees, is because the process encompasses all the interrelationships between variables as well as serial correlations that would otherwise not be identifiable.

Q.580 John Neur, FRM, runs a Monte Carlo simulation to estimate the ending amount of capital in 25 years using monthly returns for three investments as the basis. Investments A and B are highly correlated while C has zero correlation with both A and B. In order to compute the output of the Monte Carlo simulation, John:

> ***A. Cannot measure the correlations between the three investments.***
>
> ***B. Must accurately determine the probability distribution of the output.***
>
> ***C. Can easily examine effects on output variables when changing scenarios.***
>
> ***D. Must assume that the output is normally distributed.***

The correct answer is C.

The effects on output variables can easily be established when changing strategies or scenarios during simulations. It simplifies complex functions and allows for correlation between inputs since the probability distribution of the output variable needs not be identified.

Q.582 An analyst runs a simulation to estimate the future value of an investment of $10,000 today over a 40-year period. He uses random monthly returns that are normally distributed. How does the analyst's situation create a discretization error bias?

> ***A. By using normally distributed returns.***
>
> ***B. By using a simulation period that's too long (40 years).***
>
> ***C. By assuming that returns are random.***
>
> ***D. By assuming that returns are generated on a monthly basis instead of a continuous basis.***

The correct answer is D.

Discretization bias crops up when a variable is incorrectly assumed to take on only discrete values.

Q.584 Construct a 95% confidence interval for the future value of a pension fund where the number of simulations is 100, the mean ending value is $400,000, and the standard deviation is $23,300.

> ***A. ($395,433.2, $404,566.8)***
>
> ***B. ($400,000, $404,613)***
>
> ***C. ($395,456, $404,456)***
>
> ***D. ($395, $404)***

The correct answer is A.

The confidence interval is constructed using the normal distribution, not the student's t-distribution because n is large (in line with the central limit theorem)

The interval is given by:

$$\bar{X} - 1.96 * \left(\frac{s}{\sqrt{N}}\right), \bar{X} + 1.96 * \left(\frac{s}{\sqrt{N}}\right)$$

Thus,

$$\begin{aligned} CI &= \$400,000 - 1.96\left(\frac{\$23,300}{\sqrt{100}}\right), \$400,000 + 1.96\left(\frac{\$23,300}{\sqrt{100}}\right) \\ &= (\$395,433.2, \$404,566.8) \end{aligned}$$

Q.3387 Tom Breitling, FRM, is working on building a model using a Monte Carlo Simulation. However, he is concerned about the accuracy of the simulation. Which of the following is not a way of increasing the accuracy of the simulation?

A. Increasing the number of generated scenarios

B. Variance reduction techniques

C. Decreasing the standard error

D. Controlling the standard deviation

The correct answer is D.

Standard deviation cannot be controlled. For accuracy, we either reduce the standard error estimate by increasing the number of replication or we use variance reduction techniques.

Q.3388 Which of the following variance reduction techniques is (are) required to reduce the standard error estimate which ultimately is important for the correctness of the simulation?

A. Antithetic variate technique

B. Control variate technique

C. Both options A and B

D. Neither options A nor B

The correct answer is C.

The standard deviation cannot be controlled and increasing the number of generated scenarios is too costly. The alternative is to use variance reduction techniques for reducing the standard error estimate. The two most commonly used techniques are antithetic variates and control variates.

Q.3389 Which of the following is NOT a disadvantage of Monte Carlo simulations in solving financial problems?

> ***A. It might be computationally expensive***
>
> ***B. The results might not be precise***
>
> ***C. The results are often easy to replicate***
>
> ***D. Simulation results are experiment-specific***

The correct answer is C.

The following are the disadvantages of Monte Carlo simulations:

i. ***It might be computationally expensive***
ii. ***The results might not be precise***
iii. ***The results are often hard to replicate***
iv. ***Simulation results are experiment-specific***

Q.3390 Tim Yang, FRM, is working on building a model using a Monte Carlo simulation. However, he is concerned about the accuracy of the simulation which is measured by its standard error. Tim initially runs a model with 81 simulations and the standard deviation was found to be 27%. He then runs the model with 144 simulations and the standard deviation is still 27%.
What are the standard errors for the simulations?

A. Standard error for the first simulation: 0.33%; Standard error for the second simulation: 0.19%

B. Standard error for the first simulation: 27%; Standard error for the second simulation: 27.00%

C. Standard error for the first simulation: 2.25%; Standard error for the second simulation: 3%

D. Standard error for the first simulation: 3%; Standard error for the second simulation: 2.25%

The correct answer is D.

$\text{Standard error estimate} = \frac{\text{Standard deviation}}{\sqrt{n}}$

$\text{Standard error estimate} = \frac{27}{\sqrt{81}} = 3\%$ ***for the first simulation;***

And $\frac{27}{\sqrt{144}} = 2.25\%$ ***for the second simulation***

Q.4196 Which of the following statements correctly describes the difference between Monte Carlo Simulation and bootstrapping?

A. Monte Carlo Simulation generates variables and shocks from a particular distribution while bootstrapping generates the variables from observed data through random sampling.

B. Monte Carlo Simulation generates variables and shocks from observed data while bootstrapping generates the variables from a particular distribution..=

C. In Monte Carlo Simulation, random samples are used to draw indices when selecting data to be included in the simulation sample.

D. None of the above.

The correct answer is A.

The simulation uses the draws from a particular distribution to simulate the model quantities while a bootstrap uses random sampling from the observed data to draw a new dataset that has similar features.

Option B is incorrect because it contradicts Option A.

Option C is incorrect because it is only bootstrapping that draws the indices when selecting the data to be included in the bootstrap sample.

Q.4197 Which of the following is the most significant limitation of bootstrapping?

A. The "Black Swan" problem.

B. Bootstrapping can potentially construct samples that are significantly larger than historically observed datasets if the assumed distribution possesses the same feature.

C. Bootstrapping is suitable where the data being bootstrapped has a time dependence feature.

D. Bootstraps requires the use of more antithetic variables.

The correct answer is A.

The Black Swan problem is where the bootstrap cannot generate the data that has not occurred in the sample.

Option B is incorrect because it describes the limitation of a simulation.

Option C is an advantage of the circular block bootstrap (CBB)bootstrap.

Option D is incorrect because antithetic variables apply to Monte Carlo Simulation.

Q.4198 The estimated standard error for a Monte Carlo simulation without antithetic variables is 4.21. The antithetic variables are now included so that the correlation between the pairs is 0.22 and the simulation is repeated 144 times. What is the percentage change in standard error?

- ***A. 10.65%***
- ***B. 46.65%***
- ***C. 40.84%***
- ***D. 10.45%***

The correct answer is D.

Recall that the standard error of the simulation when antithetic variables are included, the expectation is given by:

$$\frac{\sigma_g\sqrt{1+\rho}}{\sqrt{b}}$$

We are told that,

$$\frac{\sigma_g}{\sqrt{b}} = 4.21 \Rightarrow \sigma_g = 4.21 \times \sqrt{144} = 50.52$$

So the new standard error is given by:

$$\frac{\sigma_g\sqrt{1+\rho}}{\sqrt{b}} = \frac{50.52 \times \sqrt{1+0.22}}{\sqrt{144}} = 4.65$$

So the percentage change is

$$\frac{4.65}{4.21} - 1 = 10.45\%$$

Q.4199 Assume that you want to generate random variables from U(-1,5) using random variables from U(0,1). What is the corresponding random variable of $0.10 \sim U(0,1)$***?***

A. -0.80

B. 0.50

C. -0.40

D. 0.70

The correct answer is C.

The distribution function for $X \sim U(-1,5)$ ***is given by:***

$$F(x) = \frac{x+1}{6}$$

this is due to the fact that the CDF of a uniform distribution is given by:

$$\frac{x-a}{b-a}, \quad \text{for} \quad a \leq x \leq b$$

Now, define the random variable U, so that:

$$U = \frac{x+1}{6} \Rightarrow X = 6U - 1$$

So for, U=0.10, the corresponding random variable is

$$X = 6 \times 0.10 - 1 = -0.40$$

Q.4200 Which of the following is the first step in a typical iid bootstrapping?

A. Generating data according to an assumed distribution.

B. Constructing a bootstrap sample.

C. Selecting an appropriate block size.

D. Generating a set of a random collection of m integers (1,2,3...n) with replacement.

The correct answer is D.

The first step in iid bootstrapping is generating a set of m integers $(i_1, i_2, i_4, \ldots, i_m)$ ***from {1,2,3...,n} with replacement and then followed by the construction of bootstrap sample.***

Option A is incorrect because it is the first step in the Monte Carlo simulation.

Option C is incorrect because it is the first step in CBB bootstrapping.