CFA Institute®
CFA Program

QUANTITATIVE METHODS

CFA® Program Curriculum
2023 • LEVEL 1 • VOLUME 1

SOLUTIONS

1.

A. Investment 2 is identical to Investment 1 except that Investment 2 has low liquidity. The difference between the interest rate on Investment 2 and Investment 1 is 0.5 percentage point. This amount represents the liquidity premium, which represents compensation for the risk of loss relative to an investment's fair value if the investment needs to be converted to cash quickly.

B. To estimate the default risk premium, find the two investments that have the same maturity but different levels of default risk. Both Investments 4 and 5 have a maturity of eight years. Investment 5, however, has low liquidity and thus bears a liquidity premium. The difference between the interest rates of Investments 5 and 4 is 2.5 percentage points. The liquidity premium is 0.5 percentage point (from Part A). This leaves 2.5 − 0.5 = 2.0 percentage points that must represent a default risk premium reflecting Investment 5's high default risk.

C. Investment 3 has liquidity risk and default risk comparable to Investment 2, but with its longer time to maturity, Investment 3 should have a higher maturity premium. The interest rate on Investment 3, r_3, should thus be above 2.5 percent (the interest rate on Investment 2). If the liquidity of Investment 3 were high, Investment 3 would match Investment 4 except for Investment 3's shorter maturity. We would then conclude that Investment 3's interest rate should be less than the interest rate on Investment 4, which is 4 percent. In contrast to Investment 4, however, Investment 3 has low liquidity. It is possible that the interest rate on Investment 3 exceeds that of Investment 4 despite 3's shorter maturity, depending on the relative size of the liquidity and maturity premiums. However, we expect r_3 to be less than 4.5 percent, the expected interest rate on Investment 4 if it had low liquidity. Thus 2.5 percent < r_3 < 4.5 percent.

2. C is correct. The sum of the real risk-free interest rate and the inflation premium is the nominal risk-free rate.

3. C is correct. US Treasury bonds are highly liquid, whereas the bonds of small issuers trade infrequently and the interest rate includes a liquidity premium. This liquidity premium reflects the relatively high costs (including the impact on price) of selling a position.

4. C is correct, as shown in the following (where FV is future value and PV is present value):

$$\mathrm{FV} = \mathrm{PV}\left(1 + \frac{r_s}{m}\right)^{mN}$$

$$\mathrm{FV}_6 = \$75{,}000\left(1 + \frac{0.07}{4}\right)^{(4\times 6)}$$

$$\mathrm{FV}_6 = \$113{,}733.21.$$

5. A is correct. The effective annual rate (EAR) when compounded daily is 4.08%.

$$\mathrm{EAR} = (1 + \text{Periodic interest rate})^{m} - 1$$

$$\text{EAR} = (1 + 0.04/365)^{365} - 1$$

$$\text{EAR} = (1.0408) - 1 = 0.04081 \approx 4.08\%.$$

6. B is correct. The difference between continuous compounding and daily compounding is

   €127,496.85 – €127,491.29 = €5.56, or ≈ €6, as shown in the following calculations.

   With continuous compounding, the investment earns (where PV is present value)

   $$\text{PV}\,e^{r_s N} - \text{PV} = €1{,}000{,}000e^{0.03(4)} - €1{,}000{,}000$$

   $$= €1{,}127{,}496.85 - €1{,}000{,}000$$

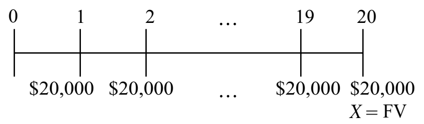
   $$= €127{,}496.85$$

   With daily compounding, the investment earns:

   $€1{,}000{,}000(1 + 0.03/365)^{365(4)} - €1{,}000{,}000 = €1{,}127{,}491.29 - €1{,}000{,}000 = €127{,}491.29$.

7.

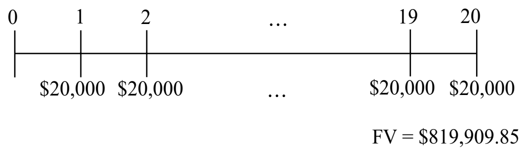
   i. Draw a time line.



   ii. Identify the problem as the future value of an annuity.

   iii. Use the formula for the future value of an annuity.

$$\text{FV}_N = A\left[\frac{(1+r)^N - 1}{r}\right]$$
$$= \$20{,}000\left[\frac{(1+0.07)^{20} - 1}{0.07}\right]$$
$$= \$819{,}909.85$$



   iv. Alternatively, use a financial calculator.

Notation Used on Most Calculators	**Numerical Value for This Problem**
N	20
%*i*	7
PV	n/a (= 0)
FV **compute**	*X*

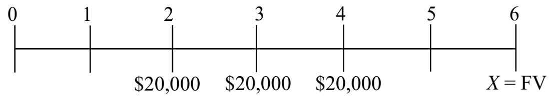
Notation Used on Most Calculators	Numerical Value for This Problem
PMT	\$20,000

Enter 20 for N, the number of periods. Enter 7 for the interest rate and 20,000 for the payment size. The present value is not needed, so enter 0. Calculate the future value. Verify that you get \$819,909.85 to make sure you have mastered your calculator's keystrokes.

In summary, if the couple sets aside \$20,000 each year (starting next year), they will have \$819,909.85 in 20 years if they earn 7 percent annually.

8.

i. Draw a time line.



ii. Recognize the problem as the future value of a delayed annuity. Delaying the payments requires two calculations.

iii. Use the formula for the future value of an annuity (Equation 7).

$$FV_N = A\left[\frac{(1+r)^N - 1}{r}\right]$$

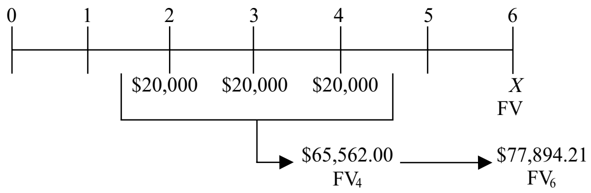
to bring the three \$20,000 payments to an equivalent lump sum of \$65,562.00 four years from today.

Notation Used on Most Calculators	Numerical Value for This Problem
N	3
$\%i$	9
PV	n/a (= 0)
FV **compute**	X
PMT	\$20,000

iv. Use the formula for the future value of a lump sum (Equation 2), $FV_N = PV(1 + r)^N$, to bring the single lump sum of \$65,562.00 to an equivalent lump sum of \$77,894.21 six years from today.

Notation Used on Most Calculators	Numerical Value for This Problem
N	2
$\%i$	9
PV	\$65,562.00
FV **compute**	X

Notation Used on Most Calculators	Numerical Value for This Problem
PMT	n/a (= 0)



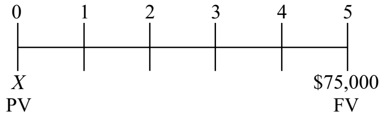
In summary, your client will have $77,894.21 in six years if she receives three yearly payments of $20,000 starting in Year 2 and can earn 9 percent annually on her investments.

9. B is correct. To solve for the future value of unequal cash flows, compute the future value of each payment as of Year 4 at the semiannual rate of 2%, and then sum the individual future values, as follows:

Year	End of Year Deposits ($)	Factor	Future Value ($)
1	4,000	$(1.02)^6$	4,504.65
2	8,000	$(1.02)^4$	8,659.46
3	7,000	$(1.02)^2$	7,282.80
4	10,000	$(1.02)^0$	10,000.00
		Sum =	30,446.91

10.

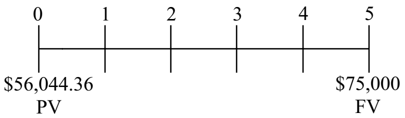
i. Draw a time line.



ii. Identify the problem as the present value of a lump sum.

iii. Use the formula for the present value of a lump sum.

$$\begin{aligned} PV &= FV_N(1+r)^{-N} \\ &= \$75,000(1+0.06)^{-5} \\ &= \$56,044.36 \end{aligned}$$



In summary, the father will need to invest $56,044.36 today in order to have $75,000 in five years if his investments earn 6 percent annually.

11. B is correct. The PV in Year 5 of a $50,000 lump sum paid in Year 20 is $27,763.23 (where FV is future value):

$$PV = FV_N(1+r)^{-N}$$

$PV = \$50{,}000(1 + 0.04)^{-15}$

$PV = \$27{,}763.23$

12. B is correct because £97,531 represents the present value (PV) of £100,000 received one year from today when today's deposit earns a stated annual rate of 2.50% and interest compounds weekly, as shown in the following equation (where FV is future value):

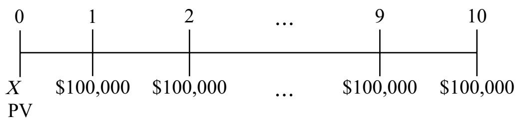
$$PV = FV_N\left(1 + \frac{r_s}{m}\right)^{-mN}$$

$$PV = £100{,}000\left(1 + \frac{0.025}{52}\right)^{-52}$$

$PV = £97{,}531.58.$

13.

i. Draw a time line for the 10 annual payments.



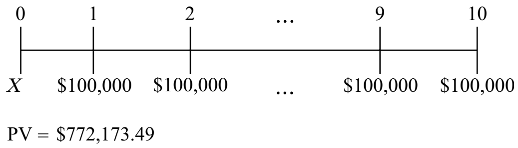
ii. Identify the problem as the present value of an annuity.

iii. Use the formula for the present value of an annuity.

$$PV = A\left[\frac{1 - \frac{1}{(1+r)^N}}{r}\right]$$

$$= \$100{,}000\left[\frac{1 - \frac{1}{(1+0.05)^{10}}}{0.05}\right]$$

$$= \$772{,}173.49$$



iv. Alternatively, use a financial calculator.

Notation Used on Most Calculators	**Numerical Value for This Problem**
N	10
$\%i$	5
PV **compute**	X
FV	n/a (= 0)

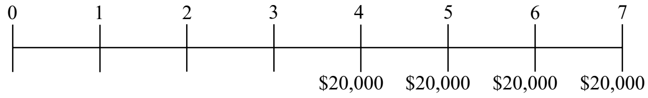
Notation Used on Most Calculators	Numerical Value for This Problem
PMT	$100,000

In summary, the present value of 10 payments of $100,000 is $772,173.49 if the first payment is received in one year and the rate is 5 percent compounded annually. Your client should accept no less than this amount for his lump sum payment.

14.

A. To evaluate the first instrument, take the following steps:

i. Draw a time line.



$$PV_3 = A\left[\frac{1-\frac{1}{(1+r)^N}}{r}\right]$$

$$= \$20,000\left[\frac{1-\frac{1}{(1+0.08)^4}}{0.08}\right]$$

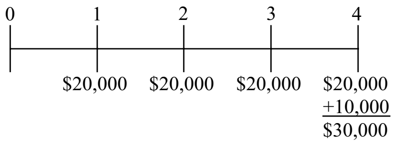
$$= \$66,242.54$$

$$PV_0 = \frac{PV_3}{(1+r)^N} = \frac{\$66,242.54}{1.08^3} = \$52,585.46$$

ii. You should be willing to pay $52,585.46 for this instrument.

B. To evaluate the second instrument, take the following steps:

i. Draw a time line.



The time line shows that this instrument can be analyzed as an ordinary annuity of $20,000 with four payments (valued in Step ii below) and a $10,000 payment to be received at $t = 4$ (valued in Step iii below).

$$PV = A\left[\frac{1-\frac{1}{(1+r)^N}}{r}\right]$$

$$= \$20,000\left[\frac{1-\frac{1}{(1+0.08)^4}}{0.08}\right]$$

$$= \$66,242.54$$

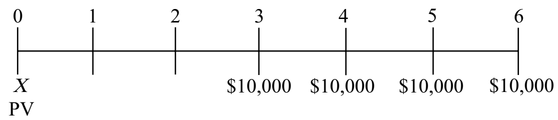
$$PV = \frac{FV_4}{(1+r)^N} = \frac{\$10,000}{(1+0.08)^4} = \$7,350.30$$

ii. Total = $66,242.54 + $7,350.30 = $73,592.84

You should be willing to pay \$73,592.84 for this instrument.

15.

i. Draw a time line.



ii. Recognize the problem as a delayed annuity. Delaying the payments requires two calculations.

iii. Use the formula for the present value of an annuity (Equation 11).

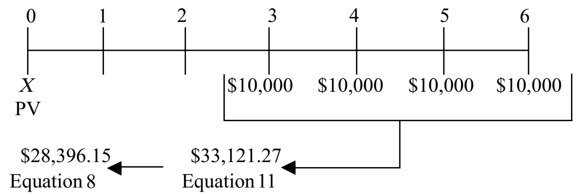
$$\text{PV} = A\left[\frac{1-\frac{1}{(1+r)^N}}{r}\right]$$

to bring the four payments of \$10,000 back to a single equivalent lump sum of \$33,121.27 at t = 2. Note that we use t = 2 because the first annuity payment is then one period away, giving an ordinary annuity.

Notation Used on Most Calculators	**Numerical Value for This Problem**
N	4
%i	8
PV **compute**	X
PMT	\$10,000

iv. Then use the formula for the present value of a lump sum (Equation 8), PV = $\text{FV}_N(1 + r)^{-N}$, to bring back the single payment of \$33,121.27 (at t = 2) to an equivalent single payment of \$28,396.15 (at t = 0).

Notation Used on Most Calculators	**Numerical Value for This Problem**
N	2
%i	8
PV **compute**	X
FV	\$33,121.27
PMT	n/a (= 0)



In summary, you should set aside \$28,396.15 today to cover four payments of \$10,000 starting in three years if your investments earn a rate of 8 percent annually.

16. B is correct, as shown in the following calculation for an annuity (A) due:

$$PV = A\left[\frac{1 - \frac{1}{(1+r)^N}}{r}\right](1+r),$$

where A = €300, r = 0.04, and N = 5.

$$PV = €300\left[\frac{1 - \frac{1}{(1+.04)^5}}{.04}\right](1.04)$$

PV = €1,388.97, or ≈ €1,389.

17. B is correct.

    The present value of a 10-year annuity (A) due with payments of $2,000 at a 5% discount rate is calculated as follows:

$$PV = A\left[\frac{1 - \frac{1}{(1+r)^N}}{r}\right] + \$2,000$$

$$PV = \$2,000\left[\frac{1 - \frac{1}{(1+0.05)^9}}{0.05}\right] + \$2,000$$

PV = $16,215.64.

Alternatively, the PV of a 10-year annuity due is simply the PV of the ordinary annuity multiplied by 1.05:

PV = $15,443.47 × 1.05

PV = $16,215.64.

18. B is correct. First, find the present value (PV) of an ordinary annuity in Year 17 that represents the tuition costs:

$$\$50,000\left[\frac{1 - \frac{1}{(1+0.06)^4}}{0.06}\right]$$

= $50,000 × 3.4651

= $173,255.28.

Then, find the PV of the annuity in today's dollars (where FV is future value):

$$PV_0 = \frac{FV}{(1+0.06)^{17}}$$

$$PV_0 = \frac{\$173,255.28}{(1+0.06)^{17}}$$

PV_0 = $64,340.85 ≈ $64,341.

19. B is correct, as shown in the following table.

Year	Cash Flow (€)	Formula CF × $(1 + r)^t$	PV at Year 0
1	100,000	$100,000(1.12)^{-1}$ =	89,285.71
2	150,000	$150,000(1.12)^{-2}$ =	119,579.08
5	–10,000	$–10,000(1.12)^{-5}$ =	–5,674.27

Year	Cash Flow (€)	Formula CF × (1 + *r*)^t	PV at Year 0
			203,190.52

20. B is correct. The value of the perpetuity one year from now is calculated as:

    PV = A/*r*, where PV is present value, A is annuity, and *r* is expressed as a quarterly required rate of return because the payments are quarterly.

    $PV = \$2.00/(0.06/4)$

    $PV = \$133.33.$

    The value today is (where FV is future value)

    $PV = FV_N(1 + r)^{-N}$

    $PV = \$133.33(1 + 0.015)^{-4}$

    $PV = \$125.62 \approx \$126.$

21. C is correct. As shown below, the present value (PV) of a £2,000 per month perpetuity is worth approximately £400,000 at a 6% annual rate compounded monthly. Thus, the present value of the annuity (A) is worth more than the lump sum offers.

    $A = £2{,}000$

    $r = (6\%/12) = 0.005$

    $PV = (A/r)$

    $PV = (£2{,}000/0.005)$

    $PV = £400{,}000$

22. A is correct. The effective annual rate (EAR) is calculated as follows:

    $EAR = (1 + \text{Periodic interest rate})^m - 1$

    $EAR = (1 + 0.03/365)^{365} - 1$

    $EAR= (1.03045) - 1 = 0.030453 \approx 3.0453\%.$

    Solving for *N* on a financial calculator results in (where FV is future value and PV is present value):

    $(1 + 0.030453)^N = FV_N/PV = ¥1{,}000{,}000/¥250{,}000$

    = 46.21 years, which multiplied by 12 to convert to months results in 554.5, or ≈ 555 months.

23. C is correct, as shown in the following (where FV is future value and PV is present value):

    If:

    $$FV_N = PV\left(1 + \frac{r_s}{m}\right)^{mN},$$

    Then:

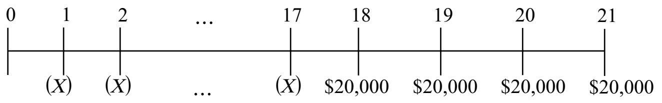
$$\left(\frac{\mathrm{FV}_N}{\mathrm{PV}}\right)^{\frac{1}{mN}} - 1 = \frac{r_s}{m}$$

$$\left(\frac{800,000}{500,000}\right)^{\frac{1}{2\times 6}} - 1 = \frac{r_s}{2}$$

r_s = 0.07988 (rounded to 8.0%).

24.

i. Draw a time line.



ii. Recognize that you need to equate the values of two annuities.

iii. Equate the value of the four $20,000 payments to a single payment in Period 17 using the formula for the present value of an annuity (Equation 11), with r = 0.05. The present value of the college costs as of t = 17 is $70,919.

$$\mathrm{PV} = \$20,000\left[\frac{1-\frac{1}{(1.05)^4}}{0.05}\right] = \$70,919$$

Notation Used on Most Calculators	Numerical Value for This Problem
N	4
%i	5
PV **compute**	X
FV	n/a (= 0)
PMT	$20,000

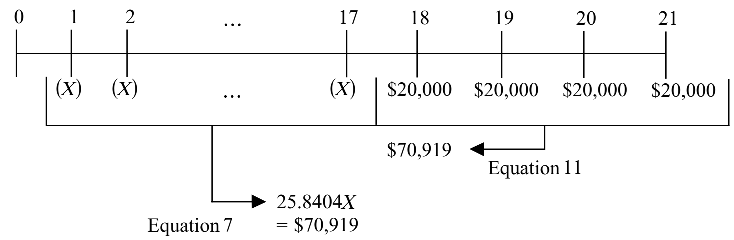
iv. Equate the value of the 17 investments of X to the amount calculated in Step iii, college costs as of t = 17, using the formula for the future value of an annuity (Equation 7). Then solve for X.

$$\$70,919 = \left[\frac{(1.05)^{17}-1}{0.05}\right] = 25.840366X$$

$$X = \$2,744.50$$

Notation Used on Most Calculators	Numerical Value for This Problem
N	17
%i	5
PV	n/a (= 0)
FV	$70,919

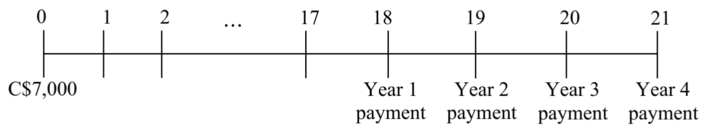
Notation Used on Most Calculators	Numerical Value for This Problem
PMT **compute**	X



In summary, your client will have to save $2,744.50 each year if she starts next year and makes 17 payments into a savings account paying 5 percent annually.

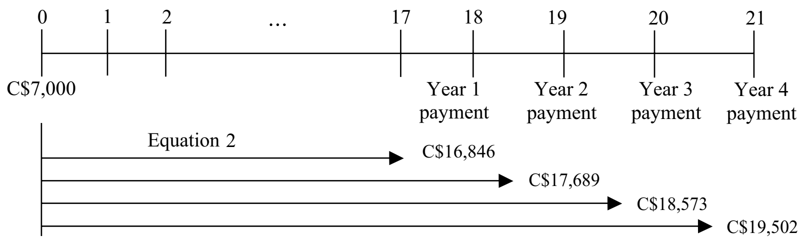
25.

i. Draw a time line.



ii. Recognize that the payments in Years 18, 19, 20, and 21 are the future values of a lump sum of C$7,000 in Year 0.

iii. With r = 5%, use the formula for the future value of a lump sum (Equation 2), $FV_N = PV(1 + r)^N$, four times to find the payments. These future values are shown on the time line below.



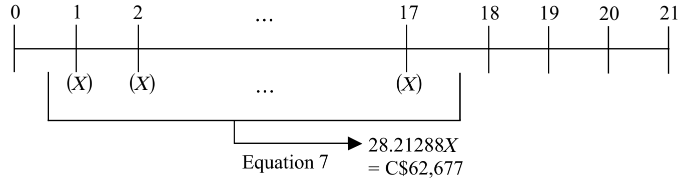
iv. Using the formula for the present value of a lump sum (r = 6%), equate the four college payments to single payments as of t = 17 and add them together. C$16,846$(1.06)^{-1}$ + C$17,689$(1.06)^{-2}$ + C$18,573$(1.06)^{-3}$ + C$19,502$(1.06)^{-4}$ = C$62,677

v. Equate the sum of C$62,677 at t = 17 to the 17 payments of X, using the formula for the future value of an annuity (Equation 7). Then solve for X.

$$\text{C\$62,677} = X\left[\frac{(1.06)^{17} - 1}{0.06}\right] = 28.21288X$$
$$X = \text{C\$2,221.58}$$

Notation Used on Most Calculators	Numerical Value for This Problem
N	17
$\%i$	6

Notation Used on Most Calculators	Numerical Value for This Problem
PV	n/a (= 0)
FV	C$62,677
PMT **compute**	X



In summary, the couple will need to put aside C$2,221.58 each year if they start next year and make 17 equal payments.

26. B is correct, calculated as follows (where A is annuity and PV is present value):

$$\text{A} = (\text{PV of annuity}) \Big/ \left[\frac{1 - \dfrac{1}{(1 + r_s/m)^{mN}}}{r_s/m}\right]$$

$$= (£200,000) \Big/ \left[\frac{1 - \dfrac{1}{(1 + r_s/m)^{mN}}}{r_s/m}\right]$$

$$(£200,000) \Big/ \left[\frac{1 - \dfrac{1}{(1 + 0.06/12)^{12(5)}}}{0.06/12}\right]$$

$$= (£200,000)/51.72556$$
$$= £3,866.56$$

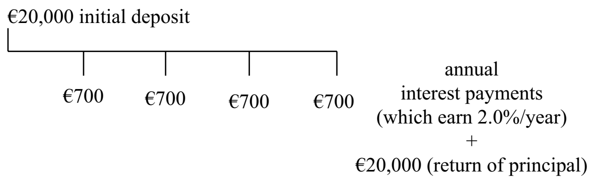
27. A is correct. To solve for an annuity (A) payment, when the future value (FV), interest rate, and number of periods is known, use the following equation:

$$\text{FV} = \text{A}\left[\frac{\left(1 + \frac{r_s}{m}\right)^{mN} - 1}{\frac{r}{m}}\right]$$

$$£25,000 = \text{A}\left[\frac{\left(1 + \frac{0.06}{4}\right)^{4\times10} - 1}{\frac{0.06}{4}}\right]$$

A = £460.68

28. B is correct, as the following cash flows show:



The four annual interest payments are based on the CD's 3.5% annual rate.

The first payment grows at 2.0% compounded monthly for three years (where FV is future value):

$$FV_N = €700\left(1 + \frac{0.02}{12}\right)^{3\times12}$$

$FV_N = 743.25$

The second payment grows at 2.0% compounded monthly for two years:

$$FV_N = €700\left(1 + \frac{0.02}{12}\right)^{2\times12}$$

$FV_N = 728.54$

The third payment grows at 2.0% compounded monthly for one year:

$$FV_N = €700\left(1 + \frac{0.02}{12}\right)^{1\times12}$$

$FV_N = 714.13$

The fourth payment is paid at the end of Year 4. Its future value is €700.

The sum of all future value payments is as follows:

€20,000.00	CD
€743.25	First payment's *FV*
€728.54	Second payment's *FV*
€714.13	Third payment's *FV*
€700.00	Fourth payment's *FV*
€22,885.92	Total *FV*

SOLUTIONS

1. A is correct. Ordinal scales sort data into categories that are ordered with respect to some characteristic and may involve numbers to identify categories but do not assure that the differences between scale values are equal. The buy rating scale indicates that a stock ranked 5 is expected to perform better than a stock ranked 4, but it tells us nothing about the performance difference between stocks ranked 4 and 5 compared with the performance difference between stocks ranked 1 and 2, and so on.

2. C is correct. Nominal data are categorical values that are not amenable to being organized in a logical order. A is incorrect because ordinal data are categorical data that can be logically ordered or ranked. B is incorrect because discrete data are numerical values that result from a counting process; thus, they can be ordered in various ways, such as from highest to lowest value.

3. B is correct. Categorical data (or qualitative data) are values that describe a quality or characteristic of a group of observations and therefore can be used as labels to divide a dataset into groups to summarize and visualize. The two types of categorical data are nominal data and ordinal data. Nominal data are categorical values that are not amenable to being organized in a logical order, while ordinal data are categorical values that can be logically ordered or ranked. A is incorrect because discrete data would be classified as numerical data (not categorical data). C is incorrect because continuous data would be classified as numerical data (not categorical data).

4. C is correct. Continuous data are data that can be measured and can take on any numerical value in a specified range of values. In this case, the analyst is estimating bankruptcy probabilities, which can take on any value between 0 and 1. Therefore, the set of bankruptcy probabilities estimated by the analyst would likely be characterized as continuous data. A is incorrect because ordinal data are categorical values that can be logically ordered or ranked. Therefore, the set of bankruptcy probabilities would not be characterized as ordinal data. B is incorrect because discrete data are numerical values that result from a counting process, and therefore the data are limited to a finite number of values. The proprietary model used can generate probabilities that can take any value between 0 and 1; therefore, the set of bankruptcy probabilities would not be characterized as discrete data.

5. A is correct. Ordinal data are categorical values that can be logically ordered or ranked. In this case, the classification of sentences in the earnings call transcript into three categories (negative, neutral, or positive) describes ordinal data, as the data can be logically ordered from positive to negative. B is incorrect because discrete data are numerical values that result from a counting process. In this case, the analyst is categorizing sentences (i.e., unstructured data) from the earnings call transcript as having negative, neutral, or positive sentiment. Thus, these categorical data do not represent discrete data. C is incorrect because nominal data are categorical values that are not amenable to being organized in a logical order. In this case, the classification of unstructured data (i.e., sentences from the earnings call transcript) into three categories (negative, neutral, or positive) describes ordinal (not nominal) data, as the data can be logically ordered from positive to negative.

6. B is correct. Time-series data are a sequence of observations of a specific variable collected over time and at discrete and typically equally spaced intervals of time,

such as daily, weekly, monthly, annually, and quarterly. In this case, each column is a time series of data that represents annual total return (the specific variable) for a given country index, and it is measured annually (the discrete interval of time). A is incorrect because panel data consist of observations through time on one or more variables for multiple observational units. The entire table of data is an example of panel data showing annual total returns (the variable) for three country indexes (the observational units) by year. C is incorrect because cross-sectional data are a list of the observations of a specific variable from multiple observational units at a given point in time. Each row (not column) of data in the table represents cross-sectional data.

7. C is correct. Cross-sectional data are observations of a specific variable from multiple observational units at a given point in time. Each row of data in the table represents cross-sectional data. The specific variable is annual total return, the multiple observational units are the three countries' indexes, and the given point in time is the time period indicated by the particular row. A is incorrect because panel data consist of observations through time on one or more variables for multiple observational units. The entire table of data is an example of panel data showing annual total returns (the variable) for three country indexes (the observational units) by year. B is incorrect because time-series data are a sequence of observations of a specific variable collected over time and at discrete and typically equally spaced intervals of time, such as daily, weekly, monthly, annually, and quarterly. In this case, each column (not row) is a time series of data that represents annual total return (the specific variable) for a given country index, and it is measured annually (the discrete interval of time).

8. A is correct. Panel data consist of observations through time on one or more variables for multiple observational units. A two-dimensional rectangular array, or data table, would be suitable here as it is comprised of columns to hold the variable(s) for the observational units and rows to hold the observations through time. B is incorrect because a one-dimensional (not a two-dimensional rectangular) array would be most suitable for organizing a collection of data of the same data type, such as the time-series data from a single variable. C is incorrect because a one-dimensional (not a two-dimensional rectangular) array would be most suitable for organizing a collection of data of the same data type, such as the same variable for multiple observational units at a given point in time (cross-sectional data).

9. B is correct. In a frequency distribution, the absolute frequency, or simply the raw frequency, is the actual number of observations counted for each unique value of the variable. A is incorrect because the relative frequency, which is calculated as the absolute frequency of each unique value of the variable divided by the total number of observations, presents the absolute frequencies in terms of percentages. C is incorrect because the relative (not absolute) frequency provides a normalized measure of the distribution of the data, allowing comparisons between datasets with different numbers of total observations.

10. A is correct. The relative frequency is the absolute frequency of each bin divided by the total number of observations. Here, the relative frequency is calculated as: $(12/60) \times 100 = 20\%$. B is incorrect because the relative frequency of this bin is $(23/60) \times 100 = 38.33\%$. C is incorrect because the cumulative relative frequency of the last bin must equal 100%.

11. C is correct. The cumulative relative frequency of a bin identifies the fraction of observations that are less than the upper limit of the given bin. It is determined by summing the relative frequencies from the lowest bin up to and including the given bin. The following exhibit shows the relative frequencies for all the bins of

the data from the previous exhibit:

Lower Limit (%)	Upper Limit (%)	Absolute Frequency	Relative Frequency	Cumulative Relative Frequency
−9.19 ≤	< −5.45	1	0.083	0.083
−5.45 ≤	< −1.71	2	0.167	0.250
−1.71 ≤	< 2.03	4	0.333	0.583
2.03 ≤	< 5.77	3	0.250	0.833
5.77 ≤	≤ 9.47	2	0.167	1.000

The bin −1.71% ≤ x < 2.03% has a cumulative relative frequency of 0.583.

12. C is correct. The marginal frequency of energy sector bonds in the portfolio is the sum of the joint frequencies across all three levels of bond rating, so 100 + 85 + 30 = 215. A is incorrect because 27 is the relative frequency for energy sector bonds based on the total count of 806 bonds, so 215/806 = 26.7%, not the marginal frequency. B is incorrect because 85 is the joint frequency for AA rated energy sector bonds, not the marginal frequency.

13. A is correct. The relative frequency for any value in the table based on the total count is calculated by dividing that value by the total count. Therefore, the relative frequency for AA rated energy bonds is calculated as 85/806 = 10.5%.

    B is incorrect because 31.5% is the relative frequency for AA rated energy bonds, calculated based on the marginal frequency for all AA rated bonds, so 85/(32 + 25 + 85 + 100 + 28), not based on total bond counts. C is incorrect because 39.5% is the relative frequency for AA rated energy bonds, calculated based on the marginal frequency for all energy bonds, so 85/(100 + 85 + 30), not based on total bond counts.

14. A is correct. Twenty observations lie in the interval "0.0 to 2.0," and six observations lie in the "2.0 to 4.0" interval. Together, they represent 26/48, or 54.17%, of all observations, which is more than 50%.

15. A is correct. A bar chart that orders categories by frequency in descending order and includes a line displaying cumulative relative frequency is called a Pareto Chart. A Pareto Chart is used to highlight dominant categories or the most important groups. B is incorrect because a grouped bar chart or clustered bar chart is used to present the frequency distribution of two categorical variables. C is incorrect because a frequency polygon is used to display frequency distributions.

16. C is correct. A word cloud, or tag cloud, is a visual device for representing unstructured, textual data. It consists of words extracted from text with the size of each word being proportional to the frequency with which it appears in the given text. A is incorrect because a tree-map is a graphical tool for displaying and comparing categorical data, not for visualizing unstructured, textual data. B is incorrect because a scatter plot is used to visualize the joint variation in two numerical variables, not for visualizing unstructured, textual data.

17. C is correct. A tree-map is a graphical tool used to display and compare categorical data. It consists of a set of colored rectangles to represent distinct groups, and the area of each rectangle is proportional to the value of the corresponding group. A is incorrect because a line chart, not a tree-map, is used to display the change in a data series over time. B is incorrect because a scatter plot, not a tree-map, is used to visualize the joint variation in two numerical variables.

18. B is correct. An important benefit of a line chart is that it facilitates showing

changes in the data and underlying trends in a clear and concise way. Often a line chart is used to display the changes in data series over time. A is incorrect because a scatter plot, not a line chart, is used to visualize the joint variation in two numerical variables. C is incorrect because a heat map, not a line chart, is used to visualize the values of joint frequencies among categorical variables.

19. B is correct. A heat map is commonly used for visualizing the degree of correlation between different variables. A is incorrect because a word cloud, or tag cloud, not a heat map, is a visual device for representing textual data with the size of each distinct word being proportional to the frequency with which it appears in the given text. C is incorrect because a histogram, not a heat map, depicts the shape, center, and spread of the distribution of numerical data.

20. B is correct. A bubble line chart is a version of a line chart where data points are replaced with varying-sized bubbles to represent a third dimension of the data. A line chart is very effective at visualizing trends in three or more variables over time. A is incorrect because a heat map differentiates high values from low values and reflects the correlation between variables but does not help in making comparisons of variables over time. C is incorrect because a scatterplot matrix is a useful tool for organizing scatterplots between pairs of variables, making it easy to inspect all pairwise relationships in one combined visual. However, it does not help in making comparisons of these variables over time.

21. C is correct. Because 50 data points are in the histogram, the median return would be the mean of the 50/2 = 25th and (50 + 2)/2 = 26th positions. The sum of the return bin frequencies to the left of the 13% to 18% interval is 24. As a result, the 25th and 26th returns will fall in the 13% to 18% interval.

22. C is correct. The mode of a distribution with data grouped in intervals is the interval with the highest frequency. The three intervals of 3% to 8%, 18% to 23%, and 28% to 33% all have a high frequency of 7.

23. C is correct. The median of Portfolio R is 0.8% higher than the mean for Portfolio R.

24. C is correct. The portfolio return must be calculated as the weighted mean return, where the weights are the allocations in each asset class:

    $(0.20 \times 8\%) + (0.40 \times 12\%) + (0.25 \times -3\%) + (0.15 \times 4\%) = 6.25\%$, or $\approx 6.3\%$.

25. A is correct. The geometric mean return for Fund Y is found as follows:

    Fund Y = $[(1 + 0.195) \times (1 - 0.019) \times (1 + 0.197) \times (1 + 0.350) \times (1 + 0.057)]^{(1/5)} - 1$

    $= 14.9\%$.

26. A is correct. The harmonic mean is appropriate for determining the average price per unit. It is calculated by summing the reciprocals of the prices, then averaging that sum by dividing by the number of prices, then taking the reciprocal of the average:

    $4/[(1/62.00) + (1/76.00) + (1/84.00) + (1/90.00)] = €76.48$.

27. B is correct. The geometric mean compounds the periodic returns of every period, giving the investor a more accurate measure of the terminal value of an investment.

28. B is correct. The sum of the returns is 30.0%, so the arithmetic mean is 30.0%/10 = 3.0%.

29. B is correct.

Year	Return	1+ Return
1	4.5%	1.045
2	6.0%	1.060
3	1.5%	1.015
4	−2.0%	0.980
5	0.0%	1.000
6	4.5%	1.045
7	3.5%	1.035
8	2.5%	1.025
9	5.5%	1.055
10	4.0%	1.040

The product of the 1 + Return is 1.3402338.
Therefore, $\overline{X}_G = \sqrt[10]{1.3402338} - 1 = 2.9717\%$.

30. A is correct.

Year	Return	1+ Return	1/(1+Return)
1	4.5%	1.045	0.957
2	6.0%	1.060	0.943
3	1.5%	1.015	0.985
4	−2.0%	0.980	1.020
5	0.0%	1.000	1.000
6	4.5%	1.045	0.957
7	3.5%	1.035	0.966
8	2.5%	1.025	0.976
9	5.5%	1.055	0.948
10	4.0%	1.040	0.962
Sum			**9.714**

The harmonic mean return = (n/Sum of reciprocals) − 1 = (10 / 9.714) − 1.

The harmonic mean return = 2.9442%.

31. B is correct.

Year	Return	Deviation	Deviation Squared
1	4.5%	0.0150	0.000225
2	6.0%	0.0300	0.000900
3	1.5%	−0.0150	0.000225
4	−2.0%	−0.0500	0.002500
5	0.0%	−0.0300	0.000900
6	4.5%	0.0150	0.000225

Year	Return	Deviation	Deviation Squared
7	3.5%	0.0050	0.000025
8	2.5%	−0.0050	0.000025
9	5.5%	0.0250	0.000625
10	4.0%	0.0100	0.000100
Sum		0.0000	**0.005750**

The standard deviation is the square root of the sum of the squared deviations divided by $n - 1$:

$s = \sqrt{\frac{0.005750}{9}} = 2.5276\%.$

32. B is correct.

Year	Return	Deviation Squared below Target of 2%
1	4.5%	
2	6.0%	
3	1.5%	0.000025
4	−2.0%	0.001600
5	0.0%	0.000400
6	4.5%	
7	3.5%	
8	2.5%	
9	5.5%	
10	4.0%	
Sum		0.002025

The target semi-deviation is the square root of the sum of the squared deviations from the target, divided by $n - 1$:

$s_{\text{Target}} = \sqrt{\frac{0.002025}{9}} = 1.5\%.$

33. B is correct. The median is indicated within the box, which is the 100.49 in this diagram.

34. C is correct. The interquartile range is the difference between 114.25 and 79.74, which is 34.51.

35. B is correct. Quintiles divide a distribution into fifths, with the fourth quintile occurring at the point at which 80% of the observations lie below it. The fourth quintile is equivalent to the 80th percentile. To find the yth percentile (P_y), we first must determine its location. The formula for the location (L_y) of a yth percentile in an array with n entries sorted in ascending order is $L_y = (n + 1) \times (y/100)$. In this case, $n = 10$ and $y = 80\%$, so

$L_{80} = (10 + 1) \times (80/100) = 11 \times 0.8 = 8.8.$

With the data arranged in ascending order (−40.33%, −5.02%, 9.57%, 10.02%, 12.34%, 15.25%, 16.54%, 20.65%, 27.37%, and 30.79%), the 8.8th position would be between the 8th and 9th entries, 20.65% and 27.37%, respectively. Using linear interpolation, $P_{80} = X_8 + (L_y - 8) \times (X_9 - X_8)$,

$P_{80} = 20.65 + (8.8 - 8) \times (27.37 - 20.65)$

$= 20.65 + (0.8 \times 6.72) = 20.65 + 5.38$

$= 26.03\%$.

36. A is correct. The formula for mean absolute deviation (MAD) is

$$\text{MAD} = \frac{\sum_{i=1}^{n}\left|X_i - \overline{X}\right|}{n}.$$

Column 1: Sum annual returns and divide by n to find the arithmetic mean $(\overline{X})$ of 16.40%.

Column 2: Calculate the absolute value of the difference between each year's return and the mean from Column 1. Sum the results and divide by n to find the MAD.

These calculations are shown in the following exhibit:

	Column 1		Column 2
Year	**Return**		$\left\|X_i - \overline{X}\right\|$
Year 6	30.79%		14.39%
Year 7	12.34%		4.06%
Year 8	−5.02%		21.42%
Year 9	16.54%		0.14%
Year 10	27.37%		10.97%
Sum:	82.02%	Sum:	50.98%
n:	5	n:	5
$\overline{X}$:	16.40%	MAD:	10.20%

37. C is correct. The mean absolute deviation (MAD) of Fund ABC's returns is greater than the MAD of both of the other funds.

$$\text{MAD} = \frac{\sum_{i=1}^{n}\left|X_i - \overline{X}\right|}{n},$$ where $\overline{X}$ is the arithmetic mean of the series.

MAD for Fund ABC =

$$\frac{|-20-(-4)| + |23-(-4)| + |-14-(-4)| + |5-(-4)| + |-14-(-4)|}{5} = 14.4\%.$$

MAD for Fund XYZ =

$$\frac{|-33-(-10.8)| + |-12-(-10.8)| + |-12-(-10.8)| + |-8-(-10.8)| + |11-(-10.8)|}{5}$$
$= 9.8\%$.

MAD for Fund PQR =

$$\frac{|-14-(-5)| + |-18-(-5)| + |6-(-5)| + |-2-(-5)| + |3-(-5)|}{5} = 8.8\%.$$

A and B are incorrect because the range and variance of the three funds are as follows:

	Fund ABC	Fund XYZ	Fund PQR
Range	43%	44%	24%
Variance	317	243	110

The numbers shown for variance are understood to be in "percent squared" terms so that when taking the square root, the result is standard deviation in percentage terms. Alternatively, by expressing standard deviation and variance in decimal form, one can avoid the issue of units. In decimal form, the variances for Fund ABC, Fund XYZ, and Fund PQR are 0.0317, 0.0243, and 0.0110, respectively.

38. A is correct. The more disperse a distribution, the greater the difference between the arithmetic mean and the geometric mean.

39. B is correct. The coefficient of variation (CV) is the ratio of the standard deviation to the mean, where a higher CV implies greater risk per unit of return.

$$CV_{UTIL} = \frac{s}{\bar{X}} = \frac{1.23\%}{2.10\%} = 0.59.$$

$$CV_{MATR} = \frac{s}{\bar{X}} = \frac{1.35\%}{1.25\%} = 1.08.$$

$$CV_{INDU} = \frac{s}{\bar{X}} = \frac{1.52\%}{3.01\%} = 0.51.$$

40. B is correct. The coefficient of variation is the ratio of the standard deviation to the arithmetic average, or $\sqrt{0.001723}/0.09986 = 0.416$.

41. C is correct. The skewness is positive, so it is right-skewed (positively skewed).

42. C is correct. The excess kurtosis is positive, indicating that the distribution is "fat-tailed"; therefore, there is more probability in the tails of the distribution relative to the normal distribution.

43. B is correct. The distribution is thin-tailed relative to the normal distribution because the excess kurtosis is less than zero.

44. B is correct. The correlation coefficient is positive, indicating that the two series move together.

45. C is correct. Both outliers and spurious correlation are potential problems with interpreting correlation coefficients.

46. C is correct. The correlation coefficient is positive because the covariation is positive.

47. A is correct. The correlation coefficient is negative because the covariation is negative.

48. C is correct. The correlation coefficient is positive because the covariance is positive. The fact that one or both variables have a negative mean does not affect the sign of the correlation coefficient.

SOLUTIONS

1. C is correct. The term "exhaustive" means that the events cover all possible outcomes.

2. C is correct. A subjective probability draws on personal or subjective judgment that may be without reference to any particular data.

3. A is correct. Given odds for E of a to b, the implied probability of $E = a/(a + b)$. Stated in terms of odds a to b with $a = 1$, $b = 5$, the probability of $E = 1/(1 + 5) = 1/6 = 0.167$. This result confirms that a probability of 0.167 for beating sales is odds of 1 to 5.

4. C is correct. The odds for beating the benchmark = P(*beating benchmark*) / [1 – P(*beating benchmark*)]. Let $P(A)$ = P(*beating benchmark*). Odds for beating the benchmark = $P(A)$ / [1 – $P(A)$].

   $3 = P(A) / [1 - P(A)]$

   Solving for $P(A)$, the probability of beating the benchmark is 0.75.

5. Use this equation to find this conditional probability: P(*stock is dividend paying* | *telecom stock that meets criteria*) = P(*stock is dividend paying and telecom stock that meets criteria*)/P(*telecom stock that meets criteria*) = 0.01/0.05 = 0.20.

6. According to the multiplication rule for independent events, the probability of a company meeting all three criteria is the product of the three probabilities. Labeling the event that a company passes the first, second, and third criteria, A, B, and C, respectively, $P(ABC) = P(A)P(B)P(C) = (0.20)(0.45)(0.78) = 0.0702$. As a consequence, (0.0702)(500) = 35.10, so 35 companies pass the screen.

7. C is correct. Let event A be a stock passing the first screen (Criterion 1) and event B be a stock passing the second screen (Criterion 2). The probability of passing each screen is $P(A) = 0.50$ and $P(B) = 0.50$. If the two criteria are independent, the joint probability of passing both screens is $P(AB) = P(A)P(B) = 0.50 \times 0.50 = 0.25$, so 25 out of 100 stocks would pass both screens. However, the two criteria are positively related, and $P(AB) \neq 0.25$. Using the multiplication rule for probabilities, the joint probability of A and B is $P(AB) = P(A \mid B)P(B)$. If the two criteria are not independent, and if $P(B) = 0.50$, then the contingent probability of $P(A \mid B)$ is greater than 0.50. So the joint probability of $P(AB) = P(A \mid B)P(B)$ is greater than 0.25. More than 25 stocks should pass the two screens.

8. Use the equation for the multiplication rule for probabilities $P(AB) = P(A \mid B) P(B)$, defining A as the event that *a stock meets the financial strength criteria* and defining B as the event that *a stock meets the valuation criteria*. Then $P(AB) = P(A \mid B)P(B) = 0.40 \times 0.25 = 0.10$. The probability that a stock meets both the financial and valuation criteria is 0.10.

9. C is correct. A conditional probability is the probability of an event given that another event has occurred.

10. B is correct. Because the events are independent, the multiplication rule is most appropriate for forecasting their joint probability. The multiplication rule for independent events states that the joint probability of both A and B occurring is $P(AB) = P(A)P(B)$.

11. B is correct. The probability of the occurrence of one is related to the occurrence

of the other. If we are trying to forecast one event, information about a dependent event may be useful.

12. C is correct. The total probability rule for expected value is used to estimate an expected value based on mutually exclusive and exhaustive scenarios.

13.

A. *Outcomes associated with Scenario 1:* With a 0.45 probability of a \$0.90 recovery per \$1 principal value, given Scenario 1, and with the probability of Scenario 1 equal to 0.75, the probability of recovering \$0.90 is 0.45 (0.75) = 0.3375. By a similar calculation, the probability of recovering \$0.80 is 0.55(0.75) = 0.4125.

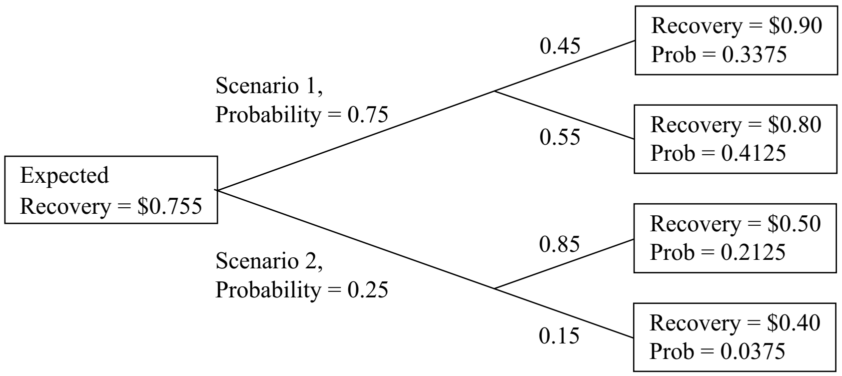
Outcomes associated with Scenario 2: With a 0.85 probability of a \$0.50 recovery per \$1 principal value, given Scenario 2, and with the probability of Scenario 2 equal to 0.25, the probability of recovering \$0.50 is 0.85(0.25) = 0.2125. By a similar calculation, the probability of recovering \$0.40 is 0.15(0.25) = 0.0375.

B. $E(recovery \mid Scenario\ 1) = 0.45(\$0.90) + 0.55(\$0.80) = \$0.845$

C. $E(recovery \mid Scenario\ 2) = 0.85(\$0.50) + 0.15(\$0.40) = \$0.485$

D. $E(recovery) = 0.75(\$0.845) + 0.25(\$0.485) = \$0.755$

E.



14. C is correct. If Scenario 1 occurs, the expected recovery is 60% (\$50,000) + 40% (\$30,000) = \$42,000, and if Scenario 2 occurs, the expected recovery is 90% (\$80,000) + 10%(\$60,000) = \$78,000. Weighting by the probability of each scenario, the expected recovery is 40%(\$42,000) + 60%(\$78,000) = \$63,600. Alternatively, first calculating the probability of each amount occurring, the expected recovery is (40%)(60%)(\$50,000) + (40%)(40%)(\$30,000) + (60%)(90%)(\$80,000) + (60%)(10%)(\$60,000) = \$63,600.

15. A is correct. The analyst must first calculate expected sales as $0.05 \times \$70 + 0.70 \times \$40 + 0.25 \times \$25 = \$3.50$ million + \$28.00 million + \$6.25 million = \$37.75 million.

After calculating expected sales, we can calculate the variance of sales:

$$\sigma^2(\text{Sales}) = P(\$70)[\$70 - E(\text{Sales})]^2 + P(\$40)[\$40 - E(\text{Sales})]^2 + P(\$25)[\$25 - E(\text{Sales})]^2$$

$$= 0.05(\$70 - 37.75)^2 + 0.70(\$40 - 37.75)^2 + 0.25(\$25 - 37.75)^2$$

= \$52.00 million + \$3.54 million + \$40.64 million = \$96.18 million.

The standard deviation of sales is thus σ = (\$96.18)$^{1/2}$ = \$9.81 million.

16. A is correct. The covariance is the product of the standard deviations and correlation using the formula Cov(US bond returns, Spanish bond returns) = σ(US bonds) × σ (Spanish bonds) × ρ(US bond returns, Spanish bond returns) = 0.64 × 0.56 × 0.24 = 0.086.

17. C is correct. The covariance of returns is positive when the returns on both assets tend to be on the same side (above or below) their expected values at the same time, indicating an average positive relationship between returns.

18. B is correct. Correlations near +1 exhibit strong positive linearity, whereas correlations near –1 exhibit strong negative linearity. A correlation of 0 indicates an absence of any linear relationship between the variables. The closer the correlation is to 0, the weaker the linear relationship.

19. C is correct. The correlation between two random variables R_i and R_j is defined as $\rho(R_i,R_j) = \text{Cov}(R_i,R_j)/[\sigma(R_i)\sigma(R_j)]$. Using the subscript i to represent hedge funds and the subscript j to represent the market index, the standard deviations are $\sigma(R_i) = 256^{1/2} = 16$ and $\sigma(R_j) = 81^{1/2} = 9$. Thus, $\rho(R_i,R_j) = \text{Cov}(R_i,R_j)/[\sigma(R_i)\ \sigma(R_j)] = 110/(16 \times 9) = 0.764$.

20. A is correct. As the correlation between two assets approaches +1, diversification benefits decrease. In other words, an increasingly positive correlation indicates an increasingly strong positive linear relationship and fewer diversification benefits.

21. A is correct. A covariance matrix for five stocks has 5 × 5 = 25 entries. Subtracting the 5 diagonal variance terms results in 20 off-diagonal entries. Because a covariance matrix is symmetrical, only 10 entries are unique (20/2 = 10).

22. C is correct. The covariance of returns is positive when the returns on both assets tend to be on the same side (above or below) their expected values at the same time.

23. B is correct. The covariance between the returns for the two stocks is $\text{Cov}(R_1,R_2) = \rho(R_1,R_2)\ \sigma(R_1)\ \sigma(R_2) = 0.20\ (12)\ (25) = 60$. The portfolio variance is

$$\sigma^2(R_p) = w_1^2\sigma^2(R_1) + w_2^2\sigma^2(R_2) + 2w_1w_2\text{Cov}(R_1,R_2)$$
$$= (0.30)^2(12)^2 + (0.70)^2(25)^2 + 2(0.30)(0.70)(60)$$
$$= 12.96 + 306.25 + 25.2 = 344.41$$

The portfolio standard deviation is

$$\sigma^2(R_p) = 344.41^{1/2} = 18.56\%$$

24. C is correct. For a three-asset portfolio, the portfolio variance is

$$\sigma^2(R_p) = w_1^2\sigma^2(R_1) + w_2^2\sigma^2(R_2) + w_3^2\sigma^2(R_3) + 2w_1w_2\text{Cov}(R_1,R_2)$$
$$+2w_1w_3\text{Cov}(R_1,R_3) + 2w_2w_3\text{Cov}(R_2,R_3)$$
$$= (0.20)^2(196) + (0.30)^2(225) + (0.50)^2(400) + 2(0.20)(0.30)(105)$$
$$+(2)(0.20)(0.50)(140) + (2)(0.30)(0.50)(150)$$
$$= 7.84 + 20.25 + 100 + 12.6 + 28 + 45 = 213.69$$

The portfolio standard deviation is

$\sigma^2(R_p) = 213.69^{1/2} = 14.62\%$

25. B is correct. The covariance is 26.56, calculated as follows. First, expected returns are

$E(R_{FI}) = (0.25 \times 25) + (0.50 \times 15) + (0.25 \times 10)$

$= 6.25 + 7.50 + 2.50 = 16.25$ and

$E(R_{DI}) = (0.25 \times 30) + (0.50 \times 25) + (0.25 \times 15)$

$= 7.50 + 12.50 + 3.75 = 23.75.$

Covariance is

$$\text{Cov}(R_{FI},R_{DI}) = \sum_i \sum_j P\left(R_{FI,i}, R_{DI,j}\right)\left(R_{FI,i} - ER_{FI}\right)\left(R_{DI,j} - ER_{DI}\right)$$

$= 0.25[(25 - 16.25)(30 - 23.75)] + 0.50[(15 - 16.25)(25 - 23.75)] + 0.25[(10 - 16.25)(15 - 23.75)]$

$= 13.67 + (-0.78) + 13.67 = 26.56.$

26.

A. We can set up the equation using the total probability rule:

$$P(\textit{pass test}) = P(\textit{pass test}|\textit{survivor})P(\textit{survivor}) + P(\textit{pass test}|\textit{non-survivor})P(\textit{non-survivor})$$

We know that $P(\textit{survivor}) = 1 - P(\textit{non-survivor}) = 1 - 0.40 = 0.60$. Therefore, $P(\textit{pass test}) = 0.55 = 0.85(0.60) + P(\textit{pass test} \mid \textit{non-survivor})(0.40)$.

Thus, $P(\textit{pass test} \mid \textit{non-survivor}) = [0.55 - 0.85(0.60)]/0.40 = 0.10$.

$$P(\textit{survivor}|\textit{pass test}) = [P(\textit{pass test}|\textit{survivor})/P(\textit{pass test})]P(\textit{survivor}) = (0.85/0.55)0.60 = 0.927273$$

B. The information that a company passes the test causes you to update your probability that it is a survivor from 0.60 to approximately 0.927.

C. According to Bayes' formula, $P(\textit{non-survivor} \mid \textit{fail test}) = [P(\textit{fail test} \mid \textit{non-survivor})/P(\textit{fail test})]P(\textit{non-survivor}) = [P(\textit{fail test} \mid \textit{non-survivor})/0.45]0.40$.

We can set up the following equation to obtain $P(\textit{fail test} \mid \textit{non-survivor})$:

$$P(\textit{fail test}) = P(\textit{fail test}|\textit{non-survivor})P(\textit{non-survivor}) + P(\textit{fail test}|\textit{survivor})P(\textit{survivor})$$

$$0.45 = P(\textit{fail test}|\textit{non-survivor})0.40 + 0.15(0.60)$$

where $P(\textit{fail test} \mid \textit{survivor}) = 1 - P(\textit{pass test} \mid \textit{survivor}) = 1 - 0.85 = 0.15$. So $P(\textit{fail test} \mid \textit{non-survivor}) = [0.45 - 0.15(0.60)]/0.40 = 0.90$.

Using this result with the formula above, we find $P(\textit{non-survivor} \mid \textit{fail test}) = [0.90/0.45]0.40 = 0.80$. Seeing that a company fails the test causes us to update the probability that it is a non-survivor from 0.40 to 0.80.

D. A company passing the test greatly increases our confidence that it is a survivor. A company failing the test doubles the probability that it is a non-survivor. Therefore, the test appears to be useful.

27. B is correct. With Bayes' formula, the probability of failure given a "good" rating is

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

where

$P(A)$ = 0.20 = probability of failure

$P(B)$ = 0.70 = probability of a "good" rating

$P(B \mid A)$ = 0.50 = probability of a "good" rating given failure

With these estimates, the probability of failure given a "good" rating is

$$P(A|B) = \frac{P(B|A)}{P(B)}P(A) = \frac{0.50}{0.70} \times 0.20 = 0.143$$

If the analyst uses the bankruptcy prediction model as a guide, the probability of failure declines from 20% to 14.3%.

28. C is correct. With Bayes' formula, the probability of the CEO being fired given a "good" rating is

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

where

$P(A)$ = 0.05 = probability of the CEO being fired

$P(B)$ = 0.50 = probability of a "good" rating

$P(B \mid A)$ = 0.30 = probability of a "good" rating given that the CEO is fired

With these estimates, the probability of the CEO being fired given a "good" rating is

$$P(A|B) = \frac{P(B|A)}{P(B)}P(A) = \frac{0.30}{0.50} \times 0.05 = 0.03$$

Although 5% of all CEOs are fired, the probability of being fired given a "good" performance rating is 3%.

29. C is correct. The combination formula provides the number of ways that *r* objects can be chosen from a total of *n* objects, when the order in which the *r* objects are listed does not matter. The order of the bonds within the portfolio does not matter.

30. A is correct. The answer is found using the combination formula

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

Here, n = 4 and r = 2, so the answer is 4!/[(4 – 2)!2!] = 24/[(2) × (2)] = 6. This result can be verified by assuming there are four vice presidents, VP1–VP4. The six possible additions to the investment committee are VP1 and VP2, VP1 and VP3, VP1 and VP4, VP2 and VP3, VP2 and VP4, and VP3 and VP4.

31. A is correct. The permutation formula is used to choose *r* objects from a total of *n* objects when order matters. Because the portfolio manager is trying to rank the four funds from most recommended to least recommended, the order of the funds matters; therefore, the permutation formula is most appropriate.

32. A is correct. The number of combinations is the number of ways to pick four mutual funds out of 10 without regard to order, which is

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

$$_{10}C_4 = \frac{10!}{(10-4)!4!} = \frac{10 \times 9 \times 8 \times 7}{4 \times 3 \times 2 \times 1} = 210$$

33. B is correct. The number of permutations is the number of ways to pick five stocks out of 30 in the correct order.

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

$$_{30}P_5 = \frac{30!}{(30-5)!} = \frac{30!}{25!} = 30 \times 29 \times 28 \times 27 \times 26 = 17,100,720$$

The contestant's chance of winning is one out of 17,100,720.

34. A is correct. The number of combinations is the number of ways to pick five stocks out of 30 without regard to order.

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

$$_{30}C_5 = \frac{30!}{(30-5)!5!} = \frac{30 \times 29 \times 28 \times 27 \times 26}{5 \times 4 \times 3 \times 2 \times 1} = 142,506$$

The contestant's chance of winning is one out of 142,506.

35. This contest does not resemble a usual lottery. Each of the 30 stocks does not have an equal chance of having the highest returns. Furthermore, contestants may have some favored investments, and the 30 stocks will not be chosen with the same frequencies. To guard against more than one person selecting the winners correctly, Sturm may wish to stipulate that if there is more than one winner, the winners will share the $1 million prize.

SOLUTIONS

1.

A. The put's minimum value is \$0. The put's value is \$0 when the stock price is at or above \$100 at the maturity date of the option. The put's maximum value is \$100 = \$100 (the exercise price) − \$0 (the lowest possible stock price). The put's value is \$100 when the stock is worthless at the option's maturity date. The put's minimum price increments are \$0.01. The possible outcomes of terminal put value are thus \$0.00, \$0.01, \$0.02, . . . , \$100.

B. The price of the underlying has minimum price fluctuations of \$0.01: These are the minimum price fluctuations for terminal put value. For example, if the stock finishes at \$98.20, the payoff on the put is \$100 − \$98.20 = \$1.80. We can specify that the nearest values to \$1.80 are \$1.79 and \$1.81. With a continuous random variable, we cannot specify the nearest values. So, we must characterize terminal put value as a discrete random variable.

C. The probability that terminal put value is less than or equal to \$24 is $P(Y \leq 24)$, or $F(24)$ in standard notation, where F is the cumulative distribution function for terminal put value.

2. C is correct. The rate of return is a random variable because the future outcomes are uncertain, and it is continuous because it can take on an unlimited number of outcomes.

3. B is correct. The function $g(x)$ satisfies the conditions of a probability function. All of the values of $g(x)$ are between 0 and 1, and the values of $g(x)$ all sum to 1.

4. B is correct. The value of the cumulative distribution function lies between 0 and 1 for any x: $0 \leq F(x) \leq 1$.

5. B is correct. The probability of any outcome is 0.05, $P(1) = 1/20 = 0.05$. The probability that X is greater than or equal to 3 but less than 6 is expressed as $P(3 \leq X < 6) = P(3) + P(4) + P(5) = 0.05 + 0.05 + 0.05 = 0.15$.

6.

A. The expected value of fourth-quarter sales is €14,500,000, calculated as (€14,000,000 + €15,000,000)/2. With a continuous uniform random variable, the mean or expected value is the midpoint between the smallest and largest values.

B. The probability that fourth-quarter sales will be less than or equal to €14,125,000 is 0.125, or 12.5%, calculated as (€14,125,000 − €14,000,000)/(€15,000,000 − €14,000,000).

7. A is correct. The probability that X will take on a value of 4 or less is $F(4) = P(X \leq 4) = p(1) + p(2) + p(3) + p(4) = 0.60$. The probability that X will take on a value of 3 or less is $F(3) = P(X \leq 3) = p(1) + p(2) + p(3) = 0.50$. So, the probability that X will take on a value of 4 is $F(4) - F(3) = p(4) = 0.10$. The probability of $X = 2$ can be found using the same logic: $F(2) - F(1) = p(2) = 0.25 - 0.15 = 0.10$. The probability of X taking on a value of 2 or 4 is $p(2) + p(4) = 0.10 + 0.10 = 0.20$.

8. A is correct. The probability of generating a random number equal to any fixed point under a continuous uniform distribution is zero.

9. A binomial random variable is defined as the number of successes in n Bernoulli trials (a trial that produces one of two outcomes). The binomial distribution is used to make probability statements about a record of successes and failures or about anything with binary (twofold) outcomes.

10. C is correct. The binomial distribution is symmetric when the probability of success on a trial is 0.50, but it is asymmetric or skewed otherwise. Here, it is given that p = 0.50.

11.

A. The probability of an earnings increase (success) in a year is estimated as 7/10 = 0.70, or 70%, based on the record of the past 10 years.

B. The probability that earnings will increase in 5 of the next 10 years is about 10.3%. Define a binomial random variable X, counting the number of earnings increases over the next 10 years. From Part A, the probability of an earnings increase in a given year is p = 0.70 and the number of trials (years) is n = 10. Equation 2 gives the probability that a binomial random variable has x successes in n trials, with the probability of success on a trial equal to p:

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

For this example,

$$\binom{10}{5} 0.7^5 0.3^{10-5} = \frac{10!}{(10-5)!5!} 0.7^5 0.3^{10-5}$$
$$= 252 \times 0.16807 \times 0.00243 = 0.102919.$$

We conclude that the probability that earnings will increase in exactly 5 of the next 10 years is 0.1029, or approximately 10.3%.

C. The expected number of yearly increases is E(X) = np = 10 × 0.70 = 7.

D. The variance of the number of yearly increases over the next 10 years is σ^2 = $np(1 - p)$ = 10 × 0.70 × 0.30 = 2.1. The standard deviation is 1.449 (the positive square root of 2.1).

E. You must assume that (1) the probability of an earnings increase (success) is constant from year to year and (2) earnings increases are independent trials. If current and past earnings help forecast next year's earnings, Assumption 2 is violated. If the company's business is subject to economic or industry cycles, neither assumption is likely to hold.

12. B is correct. To calculate the probability of four years of outperformance, use the formula

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

Using this formula to calculate the probability in four of five years, n = 5, x = 4, and p = 0.60.

Therefore,

$$p(4) = \frac{5!}{(5-4)!4!} 0.6^4 (1-0.6)^{5-4} = [120/24](0.1296)(0.40) = 0.2592.$$

$$p(5) = \frac{5!}{(5-5)!5!} 0.6^5 (1-0.6)^{5-5} = [120/120](0.0778)(1) = 0.0778.$$

The probability of outperforming four or more times is $p(4) + p(5)$ = 0.2592 + 0.0778 = 0.3370.

13. The observed success rate is 4/7 = 0.571, or 57.1%. The probability of four or fewer successes is $F(4) = p(4) + p(3) + p(2) + p(1) + p(0)$, where $p(4)$, $p(3)$, $p(2)$, $p(1)$, and $p(0)$ are, respectively, the probabilities of 4, 3, 2, 1, and 0 successes, according to the binomial distribution with $n = 7$ and $p = 0.70$. We have the following probabilities:

    $p(4) = (7!/4!3!)(0.70^4)(0.30^3) = 35(0.006483) = 0.226895.$

    $p(3) = (7!/3!4!)(0.70^3)(0.30^4) = 35(0.002778) = 0.097241.$

    $p(2) = (7!/2!5!)(0.70^2)(0.30^5) = 21(0.001191) = 0.025005.$

    $p(1) = (7!/1!6!)(0.70^1)(0.30^6) = 7(0.000510) = 0.003572.$

    $p(0) = (7!/0!7!)(0.70^0)(0.30^7) = 1(0.000219) = 0.000219.$

    Summing all these probabilities, you conclude that $F(4)$ = 0.226895 + 0.097241 + 0.025005 + 0.003572 + 0.000219 = 0.352931, or 35.3%.

14. C is correct. The probability that the performance is at or below the expectation is calculated by finding $F(3) = p(3) + p(2) + p(1) + p(0)$ using the formula:

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

    Using this formula,

    $$p(3) = \frac{4!}{(4-3)!3!} 0.75^3 (1-0.75)^{4-3} = [24/6](0.42)(0.25) = 0.42.$$

    $$p(2) = \frac{4!}{(4-2)!2!} 0.75^2 (1-0.75)^{4-2} = [24/4](0.56)(0.06) = 0.20.$$

    $$p(1) = \frac{4!}{(4-1)!1!} 0.75^1 (1-0.75)^{4-1} = [24/6](0.75)(0.02) = 0.06.$$

    $$p(0) = \frac{4!}{(4-0)!0!} 0.75^0 (1-0.75)^{4-0} = [24/24](1)(0.004) = 0.004.$$

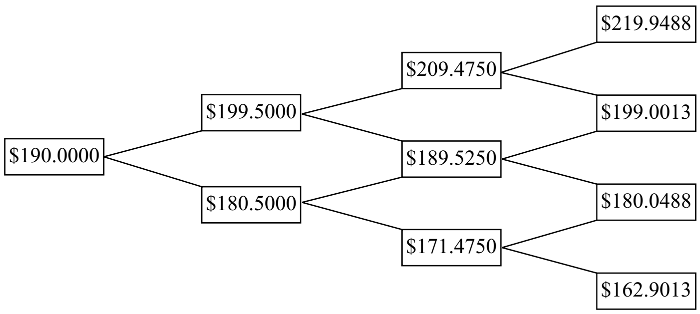
    Therefore,

    $F(3) = p(3) + p(2) + p(1) + p(0) = 0.42 + 0.20 + 0.06 + 0.004 = 0.684$, or approximately 68%.

15. A is correct. A trial, such as a coin flip, will produce one of two outcomes. Such a trial is a Bernoulli trial.

16. C is correct. The probability of an up move (p) can be found by solving the equation $(p)uS + (1 - p)dS = (p)105 + (1 - p)97 = 102$. Solving for p gives $8p = 5$, so $p = 0.625$.

17. A is correct. Only the top node value of \$219.9488 exceeds \$200.



18.

- **A.** Approximately 68% of all outcomes of a normal random variable fall within plus or minus one standard deviation of the mean.
- **B.** Approximately 95% of all outcomes of a normal random variable fall within plus or minus two standard deviations of the mean.
- **C.** Approximately 99% of all outcomes of a normal random variable fall within plus or minus three standard deviations of the mean.

19.

- **A.** The probability of exhausting the liquidity pool is 4.7%. First, calculate $x = \lambda / \left(\sigma\sqrt{T}\right) = \$2,000 / \left(\$450\sqrt{5}\right) = 1.987616$. By using Excel's NORM.S.DIST() function, we get NORM.S.DIST(1.987616) = 0.9766. Thus, the probability of exhausting the liquidity pool is 2[1 – N(1.99)] = 2(1 – 0.9766) = 0.0469, or about 4.7%.
- **B.** The probability of exhausting the liquidity pool is now 32.2%. The calculation follows the same steps as those in Part A. We calculate $x = \lambda / \left(\sigma\sqrt{T}\right) = \$2,000 / \left(\$450\sqrt{20}\right) = 0.993808$. By using Excel's NORM.S.DIST() function, we get NORM.S.DIST(0.993808)= 0.8398. Thus, the probability of exhausting the liquidity pool is 2[1 – N(0.99)] = 2(1 – 0.8398) = 0.3203, or about 32.0%. This is a substantial probability that you will run out of funds to meet marking to market.

  In their paper, Kolb et al. called the probability of exhausting the liquidity pool the probability of ruin, a traditional name for this type of calculation.

20. B is correct. The normal distribution has a skewness of 0, a kurtosis of 3, and a mean, median, and mode that are all equal.

21. B is correct. Multivariate distributions specify the probabilities for a group of related random variables. A portfolio of technology stocks represents a group of related assets. Accordingly, statistical interrelationships must be considered, resulting in the need to use a multivariate normal distribution.

22. C is correct. A bivariate normal distribution (two stocks) will have two means, two variances, and one correlation. A multivariate normal distribution for the returns on n stocks will have n means, n variances, and $n(n – 1)/2$ distinct correlations.

23. A is correct. P(8% ≤ Portfolio return ≤ 11%) = N(Z corresponding to 11%) – N(Z corresponding to 8%). For the first term, NORM.S.DIST((11% – 8%)/14%) = 58.48%. To get the second term immediately, note that 8% is the mean, and for the normal distribution, 50% of the probability lies on either side of the mean. Therefore, N(Z corresponding to 8%) must equal 50%. So, P(8% ≤ Portfolio return ≤ 11%) = 0.5848 – 0.50 = 0.0848, or approximately 8.5%.

24. B is correct. By using Excel's NORM.S.DIST() function, we get NORM.S.DIST((4% – 7%)/13%) = 40.87%. The probability that the portfolio will underperform the target is about 41%.

25. B is correct. A normal distribution has a skewness of zero (it is symmetrical around the mean). A non-zero skewness implies asymmetry in a distribution.

26. A is correct. The chance of a negative return falls in the area to the left of 0% under a standard normal curve. By standardizing the returns and standard deviations of the two assets, the likelihood of either asset experiencing a negative return may be determined: Z-score (standardized value) = $(X - \mu)/\sigma$.

    Z-score for a bond return of 0% = (0 – 2)/5 = –0.40.

    Z-score for a stock return of 0% = (0 – 10)/15 = –0.67.

    For bonds, a 0% return falls 0.40 standard deviations below the mean return of 2%. In contrast, for stocks, a 0% return falls 0.67 standard deviations below the mean return of 10%. A standard deviation of 0.40 is less than a standard deviation of 0.67. Negative returns thus occupy more of the left tail of the bond distribution than the stock distribution. Thus, bonds are more likely than stocks to experience a negative return.

27.

A. Because £50,000/£1,350,000 is 3.7%, for any return less than 3.7% the client will need to invade principal if she takes out £50,000. So R_L = 3.7%.

B. To decide which of the allocations is safety-first optimal, select the alternative with the highest ratio $[E(R_P) - R_L]/\sigma_P$:

Allocation A: 0.5125 = (16 – 3.7)/24.

Allocation B: 0.488235 = (12 – 3.7)/17.

Allocation C: 0.525 = (10 – 3.7)/12.

Allocation D: 0.481818 = (9 – 3.7)/11.

Allocation C, with the largest ratio (0.525), is the best alternative according to the safety-first criterion.

C. To answer this question, note that $P(R_C < 3.7) = N(0.037 - 0.10)/0.12$ $= N(-0.525)$. By using Excel's NORM.S.DIST() function, we get NORM.S.DIST((0.037 – 0.10)/0.12) = 29.98%, or about 30%. The safety-first optimal portfolio has a roughly 30% chance of not meeting a 3.7% return threshold.

28. B is correct. Allocation B has the highest safety-first ratio. The threshold return level, R_L, for the portfolio is £90,000/£2,000,000 = 4.5%;, thus, any return less than R_L = 4.5% will invade the portfolio principal. To compute the allocation that is safety-first optimal, select the alternative with the highest ratio:

$$\frac{[E(R_P - R_L)]}{\sigma_P}.$$

$$\text{Allocation A} = \frac{6.5 - 4.5}{8.35} = 0.240.$$

$$\text{Allocation B} = \frac{7.5 - 4.5}{10.21} = 0.294.$$

Allocation C $= \frac{8.5 - 4.5}{14.34} = 0.279.$

29. A is correct. The continuously compounded return of an asset over a period is equal to the natural log of the asset's price change during the period. In this case,

    ln(120/112) = 6.90%.

30. B is correct. By definition, lognormal random variables cannot have negative values.

31. C is correct. A lognormal distributed variable has a lower bound of zero. The lognormal distribution is also right skewed, which is a useful property in describing asset prices.

32. A is correct. The continuously compounded return from $t = 0$ to $t = 1$ is $r_{0,1} = \ln(S_1/S_0) = \ln(186.75/208.25) = -0.10897 = -10.90\%$.

33. A is correct, since it is false. Student's t-distribution has longer (fatter) tails than the normal distribution and, therefore, it may provide a more reliable, more conservative downside risk estimate.

34. C is correct, since it is false. Both chi-square and F-distributions are bounded from below by zero, so the domains of their pdfs are restricted to positive numbers.

35.

    A. Monte Carlo simulation involves the use of computer software to represent the operation of a complex financial system. A characteristic feature of Monte Carlo simulation is the generation of a large number of random samples from a specified probability distribution (or distributions) to represent the role of risk in the system. Monte Carlo simulation is widely used to estimate risk and return in investment applications. In this setting, we simulate the portfolio's profit and loss performance for a specified time horizon. Repeated trials within the simulation produce a simulated frequency distribution of portfolio returns from which performance and risk measures are derived. Another important use of Monte Carlo simulation in investments is as a tool for valuing complex securities for which no analytic pricing formula is available. It is also an important modeling resource for securities with complex embedded options.

    B. *Strengths:* Monte Carlo simulation can be used to price complex securities for which no analytic expression is available, particularly European-style options.

    Weaknesses: Monte Carlo simulation provides only statistical estimates, not exact results. Analytic methods, when available, provide more insight into cause-and-effect relationships than does Monte Carlo simulation.

36. C is correct. A characteristic feature of Monte Carlo simulation is the generation of a large number of random samples from a specified probability distribution or distributions to represent the role of risk in the system. Therefore, it is very useful for investigating the sensitivity of a model to changes in assumptions—for example, on distributions of key variables.

37. C is correct. Monte Carlo simulation is a complement to analytical methods. Monte Carlo simulation provides statistical estimates and not exact results. Analytical methods, when available, provide more insight into cause-and-effect relationships.

SOLUTIONS

1. A is correct. With judgmental sampling, Kinzua will use his knowledge and professional judgment as a seasoned auditor to select transactions of interest from the population. This approach will allow Kinzua to create a sample that is representative of the population and that will provide sufficient audit coverage. Judgmental sampling is useful in cases that have a time constraint or in which the specialty of researchers is critical to select a more representative sample than by using other probability or non-probability sampling methods. Judgement sampling, however, entails the risk that Kinzua is biased in his selections, leading to skewed results that are not representative of the whole population.

2. A is correct. Because non-probability sampling is dependent on factors other than probability considerations, such as a sampler's judgment or the convenience to access data, there is a significant risk that non-probability sampling might generate a non-representative sample

3. C is correct. Stratified random sampling involves dividing a population into subpopulations based on one or more classification criteria. Then, simple random samples are drawn from each subpopulation in sizes proportional to the relative size of each subpopulation. These samples are then pooled to form a stratified random sample.

4. No. First the conclusion on the limit of zero is wrong; second, the support cited for drawing the conclusion (i.e., the central limit theorem) is not relevant in this context.

5. In many instances, the distribution that describes the underlying population is not normal or the distribution is not known. The central limit theorem states that if the sample size is large, regardless of the shape of the underlying population, the distribution of the sample mean is approximately normal. Therefore, even in these instances, we can still construct confidence intervals (and conduct tests of inference) as long as the sample size is large (generally $n \geq 30$).

6. The statement makes the following mistakes:

   - Given the conditions in the statement, the distribution of $\overline{X}$ will be approximately normal only for large sample sizes.
   - The statement omits the important element of the central limit theorem that the distribution of $\overline{X}$ will have mean μ.

7.

   A. The standard deviation or standard error of the sample mean is $\sigma_{\overline{X}} = \sigma/\sqrt{n}$. Substituting in the values for $\sigma_{\overline{X}}$ and σ, we have 1% = 6%/$\sqrt{n}$, or $\sqrt{n}$ = 6. Squaring this value, we get a random sample of n = 36.

   B. As in Part A, the standard deviation of sample mean is $\sigma_{\overline{X}} = \sigma/\sqrt{n}$. Substituting in the values for $\sigma_{\overline{X}}$ and σ, we have 0.25% = 6%/$\sqrt{n}$, or $\sqrt{n}$ = 24. Squaring this value, we get a random sample of n = 576, which is substantially larger than for Part A of this question.

8. B is correct. Given a population described by any probability distribution (normal or non-normal) with finite variance, the central limit theorem states that the sampling distribution of the sample mean will be approximately normal, with the

mean approximately equal to the population mean, when the sample size is large.

9. B is correct. Taking the square root of the known population variance to determine the population standard deviation (σ) results in

$$\sigma = \sqrt{2.45} = 1.565$$

The formula for the standard error of the sample mean (σ_X), based on a known sample size (n), is

$$\sigma_X = \frac{\sigma}{\sqrt{n}}$$

Therefore,

$$\sigma_X = \frac{1.565}{\sqrt{40}} = 0.247$$

10. B is correct. An unbiased estimator is one for which the expected value equals the parameter it is intended to estimate.

11. A is correct. A consistent estimator is one for which the probability of estimates close to the value of the population parameter increases as sample size increases. More specifically, a consistent estimator's sampling distribution becomes concentrated on the value of the parameter it is intended to estimate as the sample size approaches infinity.

12.

A. Assume the sample size will be large and thus the 95% confidence interval for the population mean of manager returns is $\overline{X} \pm 1.96 s_{\overline{X}}$, where $s_{\overline{X}} = s/\sqrt{n}$. Munzi wants the distance between the upper limit and lower limit in the confidence interval to be 1%, which is

$$\left(\overline{X} + 1.96 s_{\overline{X}}\right) - \left(\overline{X} - 1.96 s_{\overline{X}}\right) = 1\%$$

Simplifying this equation, we get $2\left(1.96 s_{\overline{X}}\right) = 1\%$. Finally, we have $3.92 s_{\overline{X}} =$ 1%, which gives us the standard deviation of the sample mean, $s_{\overline{X}} = 0.255\%$. The distribution of sample means is $s_{\overline{X}} = s/\sqrt{n}$. Substituting in the values for $s_{\overline{X}}$ and s, we have $0.255\% = 4\%/\sqrt{n}$, or $\sqrt{n} = 15.69$. Squaring this value, we get a random sample of $n = 246$.

B. With her budget, Munzi can pay for a sample of up to 100 observations, which is far short of the 246 observations needed. Munzi can either proceed with her current budget and settle for a wider confidence interval or she can raise her budget (to around $2,460) to get the sample size for a 1% width in her confidence interval.

13.

A. For a 99% confidence interval, the reliability factor we use is $t_{0.005}$; for df = 20, this factor is 2.845.

B. For a 90% confidence interval, the reliability factor we use is $t_{0.05}$; for df = 20, this factor is 1.725.

C. Degrees of freedom equals $n - 1$, or in this case 25 − 1 = 24. For a 95% confidence interval, the reliability factor we use is $t_{0.025}$; for df = 24, this factor is 2.064.

D. Degrees of freedom equals 16 − 1 = 15. For a 95% confidence interval, the reliability factor we use is $t_{0.025}$; for df = 15, this factor is 2.131.

14. Because this is a small sample from a normal population and we have only the sample standard deviation, we use the following model to solve for the confidence interval of the population mean:

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

    where we find $t_{0.025}$ (for a 95% confidence interval) for df = n − 1 = 24 − 1 = 23; this value is 2.069. Our solution is 1% ± 2.069(4%)/$\sqrt{24}$ = 1% ± 2.069(0.8165) = 1% ± 1.69. The 95% confidence interval spans the range from −0.69% to +2.69%.

15. If the population variance is known, the confidence interval is

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

    The confidence interval for the population mean is centered at the sample mean, $\overline{X}$. The population standard deviation is σ, and the sample size is n. The population standard deviation divided by the square root of n is the standard error of the estimate of the mean. The value of z depends on the desired degree of confidence. For a 95% confidence interval, $z_{0.025}$ = 1.96 and the confidence interval estimate is

    $\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$

    If the population variance is not known, we make two changes to the technique used when the population variance is known. First, we must use the sample standard deviation instead of the population standard deviation. Second, we use the t-distribution instead of the normal distribution. The critical t-value will depend on degrees of freedom n − 1. If the sample size is large, we have the alternative of using the z-distribution with the sample standard deviation.

16. A is correct. As the degree of confidence increases (e.g., from 95% to 99%), a given confidence interval will become wider. A confidence interval is a range for which one can assert with a given probability 1 − α, called the degree of confidence, that it will contain the parameter it is intended to estimate.

17. B is correct. The confidence interval is calculated using the following equation:

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

    Sample standard deviation (s) = $\sqrt{245.55}$ = 15.670.
    For a sample size of 17, degrees of freedom equal 16, so $t_{0.05}$ = 1.746.
    The confidence interval is calculated as

    $116.23 \pm 1.746\frac{15.67}{\sqrt{17}} = 116.23 \pm 6.6357$

    Therefore, the interval spans 109.5943 to 122.8656, meaning its width is equal to approximately 13.271. (This interval can be alternatively calculated as 6.6357 × 2.)

18. A is correct. To solve, use the structure of Confidence interval = Point estimate ± Reliability factor × Standard error, which, for a normally distributed population with known variance, is represented by the following formula:

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

    For a 99% confidence interval, use $z_{0.005}$ = 2.58.
    Also, σ = $\sqrt{529}$ = 23.
    Therefore, the lower limit = $31 - 2.58\frac{23}{\sqrt{65}} = 23.6398$.

19. B is correct. All else being equal, as the sample size increases, the standard error of the sample mean decreases and the width of the confidence interval also decreases.

20. B is correct.

    The estimate of the standard error of the sample mean with bootstrap resampling is calculated as follows:

$$s_{\overline{X}} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_b - \overline{\theta})^2} = \sqrt{\frac{1}{200-1}\sum_{b=1}^{200}(\hat{\theta}_b - 0.0261)^2} = \sqrt{\frac{1}{199} \times 0.835}$$

$$s_{\overline{X}} = 0.0648$$

21. B is correct. For a sample of size n, jackknife resampling usually requires n repetitions. In contrast, with bootstrap resampling, we are left to determine how many repetitions are appropriate.

22. A is correct. The discrepancy arises from sampling error. Sampling error exists whenever one fails to observe every element of the population, because a sample statistic can vary from sample to sample. As stated in the reading, the sample mean is an unbiased estimator, a consistent estimator, and an efficient estimator of the population mean. Although the sample mean is an unbiased estimator of the population mean—the expected value of the sample mean equals the population mean—because of sampling error, we do not expect the sample mean to exactly equal the population mean in any one sample we may take.

23. No, we cannot say that Alcorn Mutual Funds as a group is superior to competitors. Alcorn Mutual Funds' advertisement may easily mislead readers because the advertisement does not show the performance of all its funds. In particular, Alcorn Mutual Funds is engaging in sample selection bias by presenting the investment results from its best-performing funds only.

24. Spence may be guilty of data mining. He has used so many possible combinations of variables on so many stocks, it is not surprising that he found some instances in which a model worked. In fact, it would have been more surprising if he had not found any. To decide whether to use his model, you should do two things: First, ask that the model be tested on out-of-sample data—that is, data that were not used in building the model. The model may not be successful with out-of-sample data. Second, examine his model to make sure that the relationships in the model make economic sense, have a story, and have a future.

25. B is correct. A report that uses a current list of stocks does not account for firms that failed, merged, or otherwise disappeared from the public equity market in previous years. As a consequence, the report is biased. This type of bias is known as survivorship bias.

26. B is correct. An out-of-sample test is used to investigate the presence of data-mining bias. Such a test uses a sample that does not overlap the time period of the sample on which a variable, strategy, or model was developed.

27. A is correct. A short time series is likely to give period-specific results that may not reflect a longer time period.

SOLUTIONS

1. C is correct. Together, the null and alternative hypotheses account for all possible values of the parameter. Any possible values of the parameter not covered by the null must be covered by the alternative hypothesis (e.g., H_0: μ ≤ 5 versus H_a: μ > 5).

2.

   A. As stated in the text, we often set up the "hoped for" or "suspected" condition as the alternative hypothesis. Here, that condition is that the population value of Willco's mean annual net income exceeds $24 million. Thus, we have H_0: μ ≤ 24 versus H_a: μ > 24.

   B. Given that net income is normally distributed with unknown variance, the appropriate test statistic is $t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} = 1.469694$ with $n - 1 = 6 - 1 = 5$ degrees of freedom.

   C. We reject the null if the calculated t-statistic is greater than 2.015. The calculated t-statistic is $t = \frac{30 - 24}{10/\sqrt{6}} = 1.469694$. Because the calculated test statistic does not exceed 2.015, we fail to reject the null hypothesis. There is not sufficient evidence to indicate that the mean net income is greater than $24 million.

3. A is correct. The null hypothesis and the alternative hypothesis are complements of one another and together are exhaustive; that is, the null and alternative hypotheses combined consider all the possible values of the population parameter.

4. B is correct. One-tailed tests in which the alternative is "greater than" or "less than" represent the beliefs of the researcher more firmly than a "not equal to" alternative hypothesis.

5. C is correct. For a one-tailed hypothesis test, there is a 5% rejection region in one tail of the distribution.

6. A is correct. Calculated using a sample, a test statistic is a quantity whose value is the basis for deciding whether to reject the null hypothesis.

7. A is correct. The definition of a Type I error is when a true null hypothesis is rejected.

8. B is correct. A Type II error occurs when a false null hypothesis is not rejected.

9. B is correct. The level of significance is used to establish the rejection points of the hypothesis test.

10. B is correct. Specifying a smaller significance level decreases the probability of a Type I error (rejecting a true null hypothesis) but increases the probability of a Type II error (not rejecting a false null hypothesis). As the level of significance decreases, the null hypothesis is less frequently rejected.

11. B is correct. The power of a test is the probability of rejecting the null hypothesis when it is false.

12. B is correct. The power of a hypothesis test is the probability of correctly rejecting the null when it is false. Failing to reject the null when it is false is a Type II

error. Thus, the power of a hypothesis test is the probability of not committing a Type II error.

13. We make the decision either by comparing the calculated test statistic with the critical values or by comparing the *p*-value for the calculated test statistic with the level of significance.

    A. Reject the null hypothesis because the calculated test statistic is outside the bounds of the critical values.

    B. The calculated *t*-statistic is in the rejection region that is defined by +1.679, so we reject the null hypothesis.

    C. The *p*-value corresponding to the calculated test statistic is less than the level of significance, so we reject the null hypothesis.

    D. We fail to reject because the *p*-value for the calculated test statistic is greater than what is tolerated with a 2% level of significance.

14. B is correct. The critical value in a decision rule is the rejection point for the test. It is the point with which the test statistic is compared to determine whether to reject the null hypothesis, which is part of the fourth step in hypothesis testing.

15. C is correct. When a statistically significant result is also economically meaningful, one should further explore the logic of why the result might work in the future.

16. A is correct. The hypothesis is a two-tailed formulation. The *t*-statistic of 2.802 falls outside the critical rejection points of less than −2.756 and greater than 2.756. Therefore, the null hypothesis is rejected; the result is statistically significant. However, despite the statistical results, trying to profit on the strategy is not likely to be economically meaningful because the return is near zero after transaction costs.

17. C is correct. When directly comparing the *p*-value with the level of significance, it can be used as an alternative to using rejection points to reach conclusions on hypothesis tests. If the *p*-value is smaller than the specified level of significance, the null hypothesis is rejected. Otherwise, the null hypothesis is not rejected.

18. C is correct. The *p*-value is the smallest level of significance at which the null hypothesis can be rejected for a given value of the test statistic. The null hypothesis is rejected when the *p*-value is less than the specified significance level.

19. C is correct. The *p*-value is the smallest level of significance (α) at which the null hypothesis can be rejected.

20. B is correct. The *p*-value is the smallest level of significance (α) at which the null hypothesis can be rejected. If the *p*-value is less than α, the null is rejected. In Test 1, the *p*-value exceeds the level of significance, whereas in Test 2, the *p*-value is less than the level of significance.

21.

    A. The appropriate test statistic is a *t*-statistic, $t = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}}$, with $n - 1 = 15 - 1 =$ 14 degrees of freedom. A *t*-statistic is correct when the sample comes from an approximately normally distributed population with unknown variance.

- **B.** The appropriate test statistic is a t-statistic, $t = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}}$, with 40 − 1 = 39 degrees of freedom. A t-statistic is theoretically correct when the sample comes from a normally distributed population with unknown variance.
- **C.** The appropriate test statistic is a z-statistic, $z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$, because the sample comes from a normally distributed population with a known variance.
- **D.** The appropriate test statistic is chi-square, $x^2 = \frac{s^2(n-1)}{\sigma_0^2}$, with 50 − 1 = 49 degrees of freedom.
- **E.** The appropriate test statistic is the F-statistic, $F = \sigma 12/\sigma 22$, with 29 and 39 degrees of freedom.
- **F.** The appropriate test statistic is a t-statistic using a pooled estimate of the population variance: $t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$, where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$. The t-statistic has 25 + 30 − 2 = 53 degrees of freedom. This statistic is appropriate because the populations are normally distributed with unknown variances; because the variances are assumed to be equal, the observations can be pooled to estimate the common variance. The requirement of independent samples for using this statistic has been met.

22. We make the decision either by comparing the calculated test statistic with the critical values or by comparing the p-value for the calculated test statistic with the level of significance.

- **A.** The calculated chi-square falls between the two critical values, so we fail to reject the null hypothesis.
- **B.** The p-value for the calculated test statistic is less than the level of significance (the 5%), so we reject the null hypothesis.
- **C.** The calculated F-statistic falls outside the bounds of the critical F-values, so we reject the null hypothesis.
- **D.** The calculated chi-square exceeds the critical value for this right-side test, so we reject the null hypothesis.

23.

- **A.** H_0: μ = 0 versus H_a: μ ≠ 0.
- **B.** The t-test is based on $t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$ with n − 1 = 10 − 1 = 9 degrees of freedom. At the 0.05 significance level, we reject the null if the calculated t-statistic is outside the bounds of ±2.262 (from the table for 9 degrees of freedom and 0.025 in the right side of the distribution). For Analyst A, we have a calculated test statistic of $t = \frac{0.05 - 0}{0.10/\sqrt{10}} = 1.58114$. We, therefore, fail to reject the null hypothesis at the 0.05 level.
- **C.** For Analyst B, the t-test is based on t with 15 − 1 = 14 degrees of freedom. At the 0.05 significance level, we reject the null if the calculated t-statistic is outside the bounds of ±2.145 (from the table for 14 degrees of freedom). The calculated test statistic is $t = \frac{0.02 - 0}{0.09/\sqrt{10}} = 0.86066$. Because 0.86066 is within the range of ±2.145, we fail to reject the null at the 0.05 level.

24.

- **A.** Stating the suspected condition as the alternative hypothesis, we have

H_0: $\mu_A - \mu_B \leq 0$ versus H_a: $\mu_A - \mu_B > 0$,

where

μ_A = the population mean value of Analyst A's forecast errors

μ_B = the population mean value of Analyst B's forecast errors

B. We have two normally distributed populations with unknown variances. Based on the samples, it is reasonable to assume that the population variances are equal. The samples are assumed to be independent; this assumption is reasonable because the analysts cover different industries. The appropriate test statistic is t using a pooled estimate of the common variance: $t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$, where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$. The number of degrees of freedom is $n_A + n_B - 2 = 10 + 15 - 2 = 23$.

C. For df = 23, according to the table, the rejection point for a one-sided (right side) test at the 0.05 significance level is 1.714.

D. We first calculate the pooled estimate of variance:

E. $s_p^2 = \frac{(10 - 1)0.01 + (15 - 1)0.0081}{10 + 15 - 2} = 0.0088435.$

We then calculate the t-distributed test statistic:

$$t = \frac{(0.05 - 0.02) - 0}{\sqrt{\frac{0.0088435}{10} + \frac{0.0088435}{15}}} = \frac{0.03}{0.0383916} = 0.78142.$$

Because 0.78142 < 1.714, we fail to reject the null hypothesis. There is not sufficient evidence to indicate that the mean for Analyst A exceeds that for Analyst B.

25. A is correct. If the population sampled has unknown variance and the sample is large, a z-test may be used. Hypotheses involving "greater than" or "less than" postulations are one sided (one tailed). In this situation, the null and alternative hypotheses are stated as H_0: $\mu \leq 6\%$ and H_a: $\mu > 6\%$, respectively. A one-tailed t-test is also acceptable in this case, and the rejection region is on the right side of the distribution.

26. B is correct. The z-test is theoretically the correct test to use in those limited cases when testing the population mean of a normally distributed population with known variance.

27. B is correct. A t-statistic is the most appropriate for hypothesis tests of the population mean when the variance is unknown and the sample is small but the population is normally distributed.

28. A is correct. The calculated t-statistic value of 0.4893 falls within the bounds of the critical t-values of ±1.984. Thus, H_0 cannot be rejected; the result is not statistically significant at the 0.05 level.

29. B is correct. The assumption that the variances are equal allows for the combining of both samples to obtain a pooled estimate of the common variance.

30.

A. We test H_0: $\mu_d = 0$ versus H_a: $\mu_d \neq 0$, where μ_d is the population mean difference.

B. This is a paired comparisons t-test with $n - 1 = 480 - 1 = 479$ degrees of freedom. At the 0.05 significance level, we reject the null hypothesis if the calculated t is less than −1.96 or greater than 1.96.

$$t = \frac{\bar{d} - \mu_{d0}}{s\bar{d}} = \frac{-0.258 - 0}{3.752/\sqrt{480}} = \frac{-0.258}{0.171255} = -1.506529, \text{ or} -1.51.$$

Because the calculate t-statistic is between ±1.96, we do not reject the null hypothesis that the mean difference between the returns on the S&P 500 and small-cap stocks during the entire sample period was zero.

C. This t-test now has $n - 1 = 240 - 1 = 239$ degrees of freedom. At the 0.05 significance level, we reject the null hypothesis if the calculated t is less than −1.96 or greater than 1.96.

$$t = \frac{\bar{d} - \mu_{d0}}{s\bar{d}} = \frac{-0.640 - 0}{4.096/\sqrt{240}} = \frac{-0.640}{0.264396} = -2.420615, \text{ or} -2.42.$$

Because −2.42 < −1.96, we reject the null hypothesis at the 0.05 significance level. We conclude that during this subperiod, small-cap stocks significantly outperformed the S&P 500.

D. This t-test has $n - 1 = 240 - 1 = 239$ degrees of freedom. At the 0.05 significance level, we reject the null hypothesis if the calculated t-statistic is less than −1.96 or greater than 1.96. The calculated test statistic is

$$t = \frac{\bar{d} - \mu_{d0}}{s\bar{d}} = \frac{0.125 - 0}{3.339/\sqrt{240}} = \frac{0.125}{0.215532} = 0.579962, \text{ or } 0.58.$$

At the 0.05 significance level, because the calculated test statistic of 0.58 is between ±1.96, we fail to reject the null hypothesis for the second subperiod.

31. C is correct. A paired comparisons test is appropriate to test the mean differences of two samples believed to be dependent.

32. A is correct. A chi-square test is used for tests concerning the variance of a single normally distributed population.

33.

A. We have a "less than" alternative hypothesis, where σ^2 is the variance of return on the portfolio. The hypotheses are H_0: $\sigma^2 \geq 400$ versus H_a: $\sigma^2 < 400$, where 400 is the hypothesized value of variance, σ_0^2. This means that the rejection region is on the left side of the distribution.

B. The test statistic is chi-square distributed with 10 − 1 = 9 degrees of freedom: $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$.

C. The test statistic is calculated as

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{9 \times 15^2}{400} = \frac{2{,}025}{400} = 5.0625, \text{ or } 5.06.$$

Because 5.06 is not less than 3.325, we do not reject the null hypothesis; the calculated test statistic falls to the right of the critical value, where the critical value separates the left-side rejection region from the region where we fail to reject.

We can determine the critical value for this test using software:

Excel [CHISQ.INV(0.05,9)]

R [qchisq(.05,9)]

Python [from scipy.stats import chi2 and chi2.ppf(.05,9)]

We can determine the p-value for the calculated test statistic of 17.0953 using software:

Excel [CHISQ.DIST(5.06,9,TRUE)]

R [pchisq(5.06,9,lower.tail=TRUE)]

Python [from scipy.stats import chi2 and chi2.cdf(5.06,9)]

34.

A. We have a "not equal to" alternative hypothesis:

$$H_0: \sigma^2_{Before} = \sigma^2_{After} \text{ versus } H_a: \sigma^2_{Before} \neq \sigma^2_{After}$$

B. To test a null hypothesis of the equality of two variances, we use an F-test:

$$F = \frac{s_1^2}{s_2^2}.$$

F = 22.367/15.795 = 1.416, with 120 – 1 = 119 numerator and 120 – 1 = 119 denominator degrees of freedom. Because this is a two-tailed test, we use critical values for the 0.05/2 = 0.025 level. The calculated test statistic falls within the bounds of the critical values (that is, between 0.6969 and 1.4349), so we fail to reject the null hypothesis; there is not enough evidence to indicate that the variances are different before and after the disruption. Note that we could also have formed the F-statistic as 15.796/22.367 = 0.706 and draw the same conclusion.

We could also use software to calculate the critical values:

Excel [F.INV(0.025,119,119) and F.INV(0.975,119,119)]

R [qf(c(.025,.975),119,119)]

Python from scipy.stats import f and f.ppf
[(.025,119,119) and
f.ppf(.975,119,119)]

Additionally, we could use software to calculate the p-value of the calculated test statistic, which is 5.896% and then compare it with the level of significance:

Excel [(1-F.DIST(22.367/15.796,119,119,TRUE))*2 or
F.DIST(15.796/22.367,119,119,TRUE)*2]

R [(1-pf(22.367/15.796,119,119))*2 or
pf(15.796/22.367,119,119)*2]

Python from scipy.stats import f and f.cdf
[(15.796/22.367,119,119)*2 or
(1-f.cdf(22.367/15.796,119,119))*2]

35. B is correct. An F-test is used to conduct tests concerning the difference between the variances of two normally distributed populations with random independent

samples.

36. A is correct. A nonparametric test is used when the data are given in ranks.

37. B is correct. There are only 12 (monthly) observations over the one year of the sample and thus the samples are small. Additionally, the funds' returns are non-normally distributed. Therefore, the samples do not meet the distributional assumptions for a parametric test. The Mann–Whitney U test (a nonparametric test) could be used to test the differences between population means.

38. The hypotheses are H_0: ρ = 0 and H_a: ρ ≠ 0. The calculated test statistics are based on the formula $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$. For example, the calculated t-statistic for the correlation of Fund 3 and Fund 4 is

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.3102\sqrt{36-2}}{\sqrt{1-0.3102^2}} = 1.903.$$

Repeating this calculation for the entire matrix of correlations gives the following:

Calculated *t*-Statistics for Correlations

	Fund 1	Fund 2	Fund 3	Fund 4	S&P 500
Fund 1					
Fund 2	13.997				
Fund 3	3.165	2.664			
Fund 4	5.897	6.116	1.903		
S&P 500	8.600	8.426	4.142	6.642	

With critical values of ±2.032, with the exception of the correlation between Fund 3 and Fund 4 returns, we reject the null hypothesis for these correlations. In other words, there is sufficient evidence to indicate that the correlations are different from zero, with the exception of the correlation of returns between Fund 3 and Fund 4.

We could use software to determine the critical values:

Excel [T.INV(0.025,34) and T.INV(0.975,34)]

R [qt(c(.025,.975),34)]

Python [from scipy.stats import t and t.ppf(.025,34) and t.ppf(.975,34)]

We could also use software to determine the p-value for the calculated test statistic to enable a comparison with the level of significance. For example, for t = 2.664, the p-value is 0.01172:

Excel [(1-T.DIST(2.664,34,TRUE))*2]

R [(1-pt(2.664,34))*2]

Python [from scipy.stats import t and (1-t.cdf(2.664,34))*2]

39.

A. We have a "not equal to" alternative hypothesis:

H_0: ρ = 0 versus H_a: ρ ≠ 0

B. Mutual fund expense ratios are bounded from above and below; in practice, there is at least a lower bound on alpha (as any return cannot be less than −100%), and expense ratios cannot be negative. These variables may not be normally distributed, and the assumptions of a parametric test are not likely to be fulfilled. Thus, a nonparametric test appears to be appropriate.

We would use the nonparametric Spearman rank correlation coefficient to conduct the test: $r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$ with the t-distributed test statistic of $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$.

C. The calculation of the Spearman rank correlation coefficient is given in the following table.

Mutual Fund	Alpha	Expense Ratio	Rank by Alpha	Rank by Expense Ratio	Difference in Rank	Difference Squared
1	−0.52	1.34	6	6	0	0
2	−0.13	0.40	1	9	−8	64
3	−0.50	1.90	5	1	4	16
4	−1.01	1.50	9	2	7	49
5	−0.26	1.35	3	5	−2	4
6	−0.89	0.50	8	8	0	0
7	−0.42	1.00	4	7	−3	9
8	−0.23	1.50	2	2	0	0
9	−0.60	1.45	7	4	3	9
						151

$$r_s = 1 - \frac{6(151)}{9(80)} = -0.25833.$$

The calculated test statistic, using the t-distributed test statistic $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ is $t = \frac{-0.25833\sqrt{7}}{\sqrt{1-0.066736}} = \frac{-0.683486}{0.9332638} = -0.7075$. On the basis of this value falling within the range of ±2.306, we fail to reject the null hypothesis that the Spearman rank correlation coefficient is zero.

40. A is correct. The calculated test statistic is

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$
$$= \frac{-0.1452\sqrt{248-2}}{\sqrt{1-(-0.1452)^2}} = -2.30177.$$

Because the value of $t = -2.30177$ is outside the bounds of ±1.96, we reject the null hypothesis of no correlation and conclude that there is enough evidence to indicate that the correlation is different from zero.

41.

A. The hypotheses are as follows:

H_0: Dividend and financial leverage ratings are not related, so these groupings are independent.

H_a: Dividend and financial leverage ratings are related, so these groupings are not independent.

B. The appropriate test statistic is $\chi^2 = \sum_{i=1}^{m} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, where O_{ij} represents the observed frequency for the i and j group and E_{ij} represents the expected frequency for the i and j group if the groupings are independent.

C. The expected frequencies based on independence are as follows:

	Dividend Group			
Financial Leverage Group	**1**	**2**	**3**	**Sum**
1	38.4	48	33.6	120
2	19.2	24	16.8	60
3	22.4	28	19.6	70
Sum	80	100	70	250

The scaled squared deviations for each combination of financial leverage and dividend grouping are:

	Dividend Group		
Financial Leverage Group	**1**	**2**	**3**
1	0.06667	1.33333	1.21905
2	6.07500	8.16667	0.60952
3	6.86429	17.28571	4.70204

The sum of these scaled squared deviations is the calculated chi-square statistic of 46.3223. Because this calculated value exceeds the critical value of 9.4877, we reject the null hypothesis that these groupings are independent.

42. A is correct. The test statistic comprises squared differences between the observed and expected values, so the test involves only one side, the right side. B is incorrect because the null hypothesis is that the groups are independent, and C is incorrect because with three levels of groups for the two categorical variables, there are four degrees of freedom.

SOLUTIONS

1. C is correct. Homoskedasticity is the situation in which the variance of the residuals is constant across the observations.

2.

   A. The coefficient of determination is 0.4279:

$$\frac{\text{Explained variation}}{\text{Total variation}} = \frac{60.16}{140.58} = 0.4279.$$

   B. $F = \frac{60.16/1}{(140.58 - 60.16)/(60 - 2)} = \frac{60.16}{1.3866} = 43.3882.$

   C. Begin with the sum of squares error of 140.58 – 60.16 = 80.42. Then calculate the mean square error of 80.42 ÷ (60 – 2) = 1.38655. The standard error of the estimate is the square root of the mean square error: $s_e = \sqrt{1.38655} = 1.1775$.

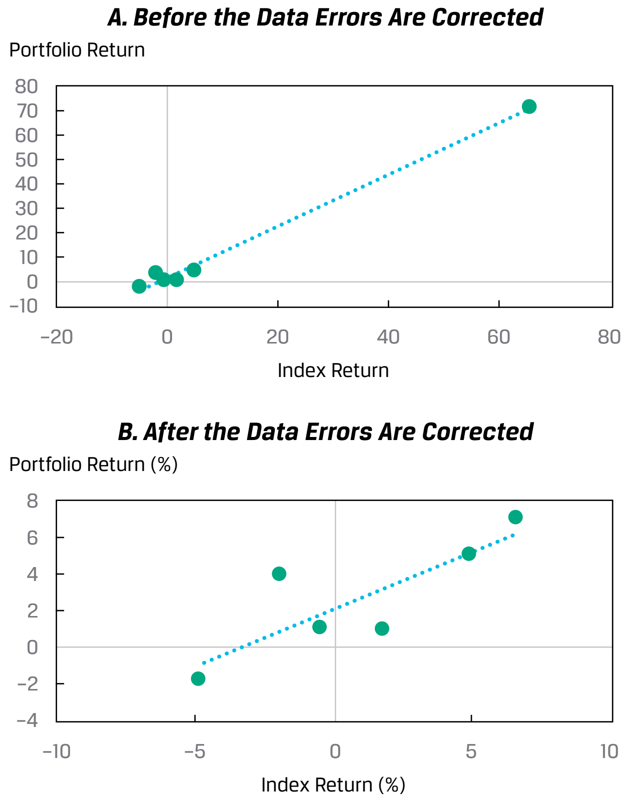
   D. The sample variance of the dependent variable uses the total variation of the dependent variable and divides it by the number of observations less one:

$$\sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{n - 1} = \frac{\text{Total variation}}{n - 1} = \frac{140.58}{60 - 1} = 2.3827.$$

   The sample standard deviation of the dependent variable is the square root of the variance, or $\sqrt{2.3827} = 1.544$.

3. The Month 2 data point is an outlier, lying far away from the other data values. Because this outlier was caused by a data entry error, correcting the outlier improves the validity and reliability of the regression. In this case, revised R^2 is lower (from 0.9921 to 0.6784). The outliers created the illusion of a better fit from the higher R^2; the outliers altered the estimate of the slope. The standard error of the estimate is lower when the data error is corrected (from 2.8619 to 2.0624), as a result of the lower mean square error. However, at a 0.05 level of significance, both models fit well. The difference in the fit is illustrated in Exhibit 1.

Exhibit 1: The Fit of the Model with and without Data Errors



4. A is correct. SHIFT is an indicator or dummy variable because it takes on only the values 0 and 1.

5. C is correct. In a simple regression with a single indicator variable, the intercept is the mean of the dependent variable when the indicator variable takes on a value of zero, which is before the shift in policy in this case.

6. C is correct. Whereas the intercept is the average of the dependent variable when the indicator variable is zero (that is, before the shift in policy), the slope is the difference in the mean of the dependent variable from before to after the change in policy.

7. A is correct. The null hypothesis of no difference in the annual growth rate is rejected at the 0.05 level: The calculated test statistic of −8.16188 is outside the bounds of ±2.048.

8.

   A. The sample variance of the dependent variable is the sum of squares total divided by its degrees of freedom ($n - 1 = 5 - 1 = 4$, as given). Thus, the sample variance of the dependent variable is 95.2 ÷ 4 = 23.8.

   B. The coefficient of determination = 88.0 ÷ 95.2 = 0.92437.

   C. The F-statistic tests whether all the slope coefficients in a linear regression are equal to zero.

D. The calculated value of the F-statistic is 36.667, as shown in the table. The corresponding p-value is less than 0.05, so you reject the null hypothesis of a slope equal to zero.

E. The standard error of the estimate is the square root of the mean square error: $s_e = \sqrt{2.4} = 1.54919$.

9. A is correct. The coefficient of determination is the same as R^2, which is 0.7436 in the table.

10. C is correct. Because the slope is positive, the correlation between X and Y is simply the square root of the coefficient of determination: $\sqrt{0.7436} = 0.8623$.

11. C is correct. To make a prediction using the regression model, multiply the slope coefficient by the forecast of the independent variable and add the result to the intercept. Expected value of CFO to sales = 0.077 + (0.826 × 5) = 4.207.

12. C is correct. The p-value is the smallest level of significance at which the null hypotheses concerning the slope coefficient can be rejected. In this case, the p-value is less than 0.05, and thus the regression of the ratio of cash flow from operations to sales on the ratio of net income to sales is significant at the 5% level.

13. A is correct. The data are observations over time.

14. C is correct. From the regression equation, Expected return = 0.0138 + (−0.6486 × −0.01) = 0.0138 + 0.006486 = 0.0203, or 2.03%.

15. C is correct. R^2 is the coefficient of determination. In this case, it shows that 2.11% of the variability in Stellar's returns is explained by changes in CPIENG.

16. B is correct. The standard error of the estimate is the standard deviation of the regression residuals.

17. C is the correct response because it is a false statement. The slope and intercept are both statistically different from zero at the 0.05 level of significance.

18. C is correct. The slope coefficient (shown in Exhibit 2) is negative. We could also determine this by looking at the cross-product (Exhibit 1), which is negative.

19. B is correct. The sample covariance is calculated as

$$\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{n-1} = -9.2430 \div 49 = -0.1886.$$

20. A is correct. In simple regression, the R^2 is the square of the pairwise correlation. Because the slope coefficient is negative, the correlation is the negative of the square root of 0.0933, or −0.3054.

21. C is correct. Conclusions cannot be drawn regarding causation; they can be drawn only about association; therefore, Interpretations 1 and 2 are incorrect.

22. C is correct. Liu explains the variation of the short interest ratio using the variation of the debt ratio.

23. A is correct. The degrees of freedom are the number of observations minus the number of parameters estimated, which equals 2 in this case (the intercept and the slope coefficient). The number of degrees of freedom is 50 − 2 = 48.

24. B is correct. The t-statistic is −2.2219, which is outside the bounds created by

the critical t-values of ±2.011 for a two-tailed test with a 5% significance level. The value of 2.011 is the critical t-value for the 5% level of significance (2.5% in one tail) for 48 degrees of freedom. A is incorrect because the mean of the short interest ratio is 192.3 ÷ 50 = 3.846. C is incorrect because the debt ratio explains 9.33% of the variation of the short interest ratio.

25. A is correct. The predicted value of the short interest ratio = 5.4975 + (–4.1589 × 0.40) = 5.4975 – 1.6636 = 3.8339.

26. C is correct because $F = \frac{\text{Mean square regression}}{\text{Mean square error}} = \frac{38.4404}{7.7867} = 4.9367.$

27. C is correct. The assumptions of the linear regression model are that (1) the relationship between the dependent variable and the independent variable is linear in the parameters b_0 and b_1, (2) the residuals are independent of one another, (3) the variance of the error term is the same for all observations, and (4) the error term is normally distributed. Assumption 3 is incorrect because the dependent variable need not be normally distributed.

28. B is correct. The standard error of the estimate for a linear regression model with one independent variable is calculated as the square root of the mean square error:

$$s_e = \sqrt{\frac{0.071475}{34}} = 0.04585.$$

29. C is correct. Crude oil returns explain the Amtex share returns if the slope coefficient is statistically different from zero. The slope coefficient is 0.2354, and the calculated t-statistic is

$$t = \frac{0.2354 - 0.0000}{0.0760} = 3.0974,$$

which is outside the bounds of the critical values of ±2.728.

Therefore, Vasileva should reject the null hypothesis that crude oil returns do not explain Amtex share returns, because the slope coefficient is statistically different from zero.

A is incorrect because the calculated t-statistic for testing the slope against 0.15 is $t = \frac{0.2354 - 0.1500}{0.0760} = 1.1237$, which is less than the critical value of +2.441.
B is incorrect because the calculated t-statistic is $t = \frac{0.0095 - 0.0000}{0.0078} = 1.2179$, which is less than the critical value of +2.441.

30. B is correct. The predicted value of the dependent variable, Amtex share return, given the value of the independent variable, crude oil return, –0.01, is calculated as $\hat{Y} = \hat{b}_0 + \hat{b}_1 X_i = 0.0095 + [0.2354 \times (-0.01)] = 0.0071.$

31. C is correct. The predicted share return is 0.0095 + [0.2354 × (–0.01)] = 0.0071. The lower limit for the prediction interval is 0.0071 – (2.728 × 0.0469) = –0.1208, and the upper limit for the prediction interval is 0.0071 + (2.728 × 0.0469) = 0.1350.

A is incorrect because the bounds of the interval should be based on the standard error of the forecast and the critical t-value, not on the mean of the dependent variable.

B is incorrect because bounds of the interval are based on the product of the standard error of the forecast *and* the critical t-value, not simply the standard error of the forecast.

32. A is correct. We fail to reject the null hypothesis of a slope equal to one, and we fail to reject the null hypothesis of an intercept equal to zero. The test of the slope equal to 1.0 is $t = \frac{0.9830 - 1.000}{0.0155} = -1.09677$. The test of the intercept equal to

0.0 is $t = \frac{0.0001 - 0.0000}{.00002} = 0.5000$. Therefore, we conclude that the forecasts are unbiased.

33. A is correct. The forecast interval for inflation is calculated in three steps:

    Step 1. Make the prediction given the US CPI forecast of 2.8:

    $$\begin{aligned}\widehat{Y} &= b_0 + b_1 X\\ &= 0.0001 + (0.9830 \times 2.8)\\ &= 2.7525.\end{aligned}$$

    Step 2. Compute the variance of the prediction error:

    $$s_f^2 = s_e^2 \left\{1 + (1/n) + \left[\left(X_f - \overline{X}\right)^2\right] / \left[(n-1) \times s_x^2\right]\right\}.$$

    $$s_f^2 = 0.0009^2 \left\{1 + (1/60) + \left[(2.8 - 1.3350)^2\right] / \left[(60-1) \times 0.7539^2\right]\right\}.$$

    $$s_f^2 = 0.00000088.$$

    $$s_f = 0.0009.$$

    Step 3. Compute the prediction interval:

    $$\hat{Y} \pm t_c \times s_f$$

    $2.7525 \pm (2.0 \times 0.0009)$

    Lower bound: $2.7525 - (2.0 \times 0.0009) = 2.7506$.

    Upper bound: $2.7525 + (2.0 \times 0.0009) = 2.7544$.

    So, given the US CPI forecast of 2.8, the 95% prediction interval is 2.7506 to 2.7544.

34. B is correct. The confidence level influences the width of the forecast interval through the critical t-value that is used to calculate the distance from the forecasted value: The larger the confidence level, the wider the interval. Therefore, Observation 1 is not correct.

    Observation 2 is correct. The greater the standard error of the estimate, the greater the standard error of the forecast.

35. B is correct. The coefficient of determination is 102.9152 ÷ 105.1303 = 0.9789.

36. A is correct. The standard error is the square root of the mean square error, or $\sqrt{0.0692} = 0.2631$.

37. B is correct. The p-value corresponding to the slope is less than 0.01, so we reject the null hypothesis of a zero slope, concluding that the fixed asset turnover explains the natural log of the net profit margin.

38. C is correct. The predicted natural log of the net profit margin is 0.5987 + (2 × 0.2951) = 1.1889. The predicted net profit margin is $e^{1.1889} = 3.2835\%$.

39. C is correct. Under the weighted average cost method:

October purchases	10,000 units	$100,000
November purchases	5,000 units	$55,000
Total	15,000 units	$155,000

$155,000/15,000 units = $10.3333

$10.3333 × 12,000 units = $124,000

APPENDICES