[a]2. A real number $x$ is represented approximately by 0.6032, and we are told that the relative error is at most 0.1%. What is $x$?
*Note*: There are *two* answers.

[a]3. What is the relative error involved in rounding 4.9997 to 5.000?

[a]4. The value of $\pi$ can be generated by the computer to nearly full machine precision by the assignment statement

$$pi \leftarrow 4.0 \arctan(1.0)$$

Suggest at least four other ways to compute $\pi$ using basic functions on your computer system.

5. A given doubly subscripted array $(a_{ij})_{n\times n}$ can be added in any order. Write the pseudocode segments for each of the following parts. Which is best?

[a]a. $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}$ b. $\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}$

c. $\sum_{i=1}^{n}\left(\sum_{j=1}^{i} a_{ij} + \sum_{j=1}^{i-1} a_{ji}\right)$

[a]d. $\sum_{k=0}^{n-1}\sum_{|i-j|=k} a_{ij}$ e. $\sum_{k=2}^{2n}\sum_{i+j=k}^{n} a_{ij}$

[a]6. Count the number of operations involved in evaluating a polynomial using nested multiplication. Do *not* count subscript calculations.

7. For small $x$, show that $(1+x)^2$ can sometimes be more accurately computed from $(x+2)x+1$. Explain. What other expressions can be used to compute it?

8. Show how these polynomials can be efficiently evaluated:

[a]a. $p(x) = x^{32}$ b. $p(x) = 3(x-1)^5 + 7(x-1)^9$

[a]c. $p(x) = 6(x+2)^3 + 9(x+2)^7 + 3(x+2)^{15} - (x+2)^{31}$

d. $p(x) = x^{127} - 5x^{37} + 10x^{17} - 3x^7$

9. Using the exponential function $\exp(x)$, write an efficient pseudocode segment for the statement $y = 5e^{3x} + 7e^{2x} + 9e^x + 11$.

[a]10. Write a pseudocode segment to evaluate the expression

$$z = \sum_{i=1}^{n} b_i^{-1} \prod_{j=1}^{i} a_j$$

where $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ are linear arrays containing given values.

11. Write segments of pseudocode to evaluate the following expressions efficiently:

a. $p(x) = \sum_{k=0}^{n-1} kx^k$

[a]b. $z = \sum_{i=1}^{n}\prod_{j=1}^{i} x^{n-j+1}$

c. $z = \prod_{i=1}^{n}\sum_{j=1}^{i} x_j$

d. $p(t) = \sum_{i=1}^{n} a_i \prod_{j=1}^{i-1}(t - x_j)$

12. Using summation and product notation, write mathematical expressions for the following pseudocode segments:

a.
```
integer i, n; real v, x
real array (a_i)0:n
v ← a0
for i = 1 to n
    v ← v + x a_i
end for
```

[a]b.
```
integer i, n; real v, x
real array (a_i)0:n
v ← a_n
for i = 1 to n
    v ← v x + a_(n−i)
end for
```

c.
```
integer i, n; real v, x
real array (a_i)0:n
v ← a0
for i = 1 to n
    v ← v x + a_i
end for
```

d.
```
integer i, n; real v, x, z
real array (a_i)0:n
v ← a0
z ← x
for i = 1 to n
    v ← v + z a_i
    z ← x z
end for
```

[a]e.
```
integer i, n; real v
real array (a_i)0:n
v ← a_n
for i = 1 to n
    v ← (v + a_(n−i)) x
end for
```

[a]13. Express in mathematical notation *without* parentheses the final value of $z$ in the following pseudocode segment:

```
integer k, n; real z
real array (b_i)0:n
z ← b_n + 1
for k = 1 to n − 2
    z ← z b_(n−k) + 1
end for
```

[a]**14.** How many multiplications occur in executing the following pseudocode segment?

```
integer i, j, n; real x
real array (a_ij)0:n×0:n, (b_ij)0:n×0:n
x ← 0.0
for j = 1 to n
    for i = 1 to j
        x ← x + a_ij b_ij
    end for
end for
```

**15.** Criticize the following pseudocode segments and write improved versions:

**a.**
```
integer i, n; real x, z; real array (a_i)0:n
for i = 1 to n
    x ← z^2 + 5.7
    a_i ← x/i
end for
```

[a]**b.**
```
integer i, j, n
real array (a_ij)0:n×0:n
for i = 1 to n
    for j = 1 to n
        a_ij ← 1/(i + j − 1)
    end for
end for
```

**c.**
```
integer i, j, n; real array (a_ij)0:n×0:n
for j = 1 to n
    for i = 1 to n
        a_ij ← 1/(i + j − 1)
    end for
end for
```

**16. a.** Solve Example 1.1.2 to full precision.

**b.** Repeat for this augmented matrix

$$\left[\begin{array}{cc|c} 3.5713 & 2.1426 & 7.2158 \\ 10.714 & 6.4280 & 1.3379 \end{array}\right]$$

for a system of two equations and two unknowns $x$ and $y$.

**c.** Can small changes in the data lead to massive change in the each of these solutions?

**17.** A base 60 approximation (circa 1750 BCE) is

$$\sqrt{2} \approx 1 + \frac{24}{60} + \frac{51}{60^2} + \frac{10}{60^3}$$

Determine how accurate it is. See Sauer [2006] for additional details.

**18.** Use Horner's algorithm to evaluate each of these polynomials at the point indicated.

**a.** $2x^4 + 9x^2 - 16x + 12$ at $-6$

**b.** $2x^4 - 3x^3 - 5x^2 + 3x + 8$ at 2

**c.** $3x^5 - 38x^3 + 5x^2 - 1$ at 4

## Computer Exercises 1.1

**1.** Write and run a computer program that corresponds to the pseudocode program *First* described on pp. 12–13 and interpret the results.

**2.** (Continuation) Select a function $f$ and a point $x$ and carry out a computer experiment like the one given in the text. Interpret the results. Do not select too simple a function. For example, you might consider $1/x$, $\log x$, $e^x$, $\tan x$, $\cosh x$, or $x^3 - 23x$.

**3.** (Continuation) As we saw in the computer experiment *First*, the accuracy of a formula for numerical differentiation may deteriorate as the step-size $h$ decreases. Study the following central difference formula:

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}$$

as $h \to 0$. We learn in Section 4.3 that the **truncation error** for this formula is $-\frac{1}{6}h^2 f'''(\xi)$ for some $\xi$ in the interval $(x - h, x + h)$.

Modify and run the code for the experiment *First* so that approximate values for the rounding error and truncation error are computed. On the same graph, plot the rounding error, the truncation error, and the total error (sum of these two errors) using a log-scale; that is, the axes in the plot should be $-\log_{10}|\text{error}|$ versus $\log_{10} h$. Analyze these results.

[a]**4.** The limit

$$e = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n$$

defines the number $e$ in calculus. Estimate $e$ by taking the value of this expression for $n = 8, 8^2, 8^3, \dots, 8^{10}$.

Compare with $e$ obtained from $e \leftarrow \exp(1.0)$. Interpret the results.

**5.** It is not difficult to see that the numbers

$$p_n = \int_0^1 x^n e^x \, dx$$

satisfy the inequalities $p_1 > p_2 > p_3 > \cdots > 0$. Establish this fact. Next, use integration by parts to show that

$$p_{n+1} = e - (n+1) p_n$$

and that $p_1 = 1$. In the computer, use the recurrence relation to generate the first 20 values of $p_n$ and explain why the inequalities shown are violated. Do not use subscripted variables. (See Dorn and McCracken [1972], pp. 120–129.)

**6.** (Continuation) Let $p_{20} = \frac{1}{8}$ and use the formula in the preceding computer problem to compute $p_{19}, p_{18}, \ldots, p_2$, and $p_1$. Do the numbers generated obey the inequalities $1 = p_1 > p_2 > p_3 > \cdots > 0$? Explain the difference in the two procedures. Repeat with $p_{20} = 20$ or $p_{20} = 100$. Explain what happens.

**7.** Write an efficient routine that accepts as input a list of real numbers $a_1, a_2, \ldots, a_n$ and then computes the following:

**Arithmetic mean** $\quad m = \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} a_k$

**Variance** $\quad v = \dfrac{1}{n-1} \displaystyle\sum_{k=1}^{n} (a_k - m)^2$

**Standard deviation** $\quad \sigma = \sqrt{v}$

Test the routine on a set of data of your choice.

**8.** (Continuation) Show that another formula is

**Variance** $\quad v = \dfrac{1}{n-1}\left[\displaystyle\sum_{k=1}^{n} a_k^2 - nm^2\right]$

Of the two given formulas for $v$, which is more accurate in the computer? Verify on the computer with a data set. *Hint*: Use a large set of real numbers that vary in magnitude from very small to very large.

**[a]9.** Let $a_1$ be given. Write a program to compute for $1 \leqq n \leqq 1000$ the numbers

$$b_n = n a_{n-1}, \qquad a_n = b_n / n$$

Print the numbers $a_{100}, a_{200}, \ldots, a_{1000}$. *Do not* use subscripted variables. What should $a_n$ be? Account for the deviation of fact from theory. Determine four values for $a_1$ so that the computation does deviate from theory on your computer.

*Hint*: Consider extremely small and large numbers and print to full machine precision.

**[a]10.** In a computer, it can happen that $a + x = a$ when $x \neq 0$. Explain why. Describe the set of $n$ for which $1 + 2^{-n} = 1$ in your computer. Write and run appropriate programs to illustrate the phenomenon.

**11.** Write a program to test the programming suggestion concerning the roundoff error in the computation of $t \leftarrow t + h$ versus $t \leftarrow t_0 + ih$. For example, use $h = \frac{1}{10}$ and compute $t \leftarrow t + h$ in double precision for the correct single-precision value of $t$; print the absolute values of the differences between this calculation and the values of the two procedures. What is the result of the test when $h$ is a machine number, such as $h = \frac{1}{128}$, on a binary computer (with more than seven bits per word)?

**[a]12.** The Russian mathematician P. L. **Chebyshev** (1821–1894) spelled his name Qebywev. Many transliterations from the Cyrillic to the Latin alphabet are possible. *Cheb* can alternatively be rendered as Ceb, Tscheb, or Tcheb. The $y$ can be rendered as $i$. *Shev* can also be rendered as schef, cev, cheff, or scheff. Taking all combinations of these variants, program a computer to print all possible spellings.

**13.** Compute $n!$ using logarithms, integer arithmetic, and double-precision floating-point arithmetic. For each part, print a table of values for $0 \leqq n \leqq 30$, and determine the largest correct value.

**14.** Given two arrays, a real array $v = (v_1, v_2, \ldots, v_n)$ and an integer permutation array $p = (p_1, p_2, \ldots, p_n)$ of integers $1, 2, \ldots, n$, can we form a new permuted array $v = (v_{p_1}, v_{p_2}, \ldots, v_{p_n})$ by overwriting $v$ and *not* involving another array in memory? If so, write and test the code for doing it. If *not*, use an additional array and test. Consider these cases:

**Case 1.** $v = (6.3, 4.2, 9.3, 6.7, 7.8, 2.4, 3.8, 9.7)$,
$p = (2, 3, 8, 7, 1, 4, 6, 5)$

**Case 2.** $v = (0.7, 0.6, 0.1, 0.3, 0.2, 0.5, 0.4)$,
$p = (3, 5, 4, 7, 6, 2, 1)$

**15.** Using a computer algebra system (e.g., Maple, Mathematica, MATLAB, etc.), print 200 decimal digits of $\sqrt{10}$.

**16.** **a.** Repeat the Example 1.1.1 on loss of significant digits of accuracy, but perform the calculations with twice the precision before rounding them. Does this help?

**b.** Use Maple or some other mathematical software system in which you can set the number of digits of precision.
*Hint*: In Maple, use `Digits`.

**17.** In 1706, Machin used the formula

$$\pi = 16 \arctan\left(\frac{1}{5}\right) - 4 \arctan\left(\frac{1}{239}\right)$$

to compute 100 digits of $\pi$. Derive this formula. Reproduce Machin's calculations by using suitable software. *Hint*: Let $\tan\theta = \frac{1}{5}$, and use standard trigonometric identities.

**18.** Using a symbol-manipulating program such as Maple, Mathematica, MATLAB, or Macsyma, carry out the following tasks. Record your work in some manner, for example, by using a `diary` or `script` command.

**a.** Find the Taylor series, up to and including the term $x^{10}$, for the function $(\tan x)^2$, using 0 as the point $x_0$.

**b.** Find the indefinite integral of $(\cos x)^{-4}$.

**c.** Find the definite integral $\int_0^1 \log|\log x|\,dx$.

**d.** Find the first prime number greater than 27448.

**e.** Obtain the numerical value of $\int_0^1 \sqrt{1+\sin^3 x}\,dx$.

**f.** Find the solution of the differential equation $y' + y = (1+e^x)^{-1}$.

**g.** Define the function $f(x, y) = 9x^4 - y^4 + 2y^2 - 1$. You want to know the value of $f(40545, 70226)$. Compute this in the straightforward way by direct substitution of $x = 40545$ and $y = 70226$ in the definition of $f(x, y)$, using first 6-decimal accuracy, then 7-, 8-, and so on up to 24-decimal digits of accuracy. Next, prove by means of elementary algebra that

$$f(x, y) = (3x^2 - y^2 + 1)(3x^2 + y^2 - 1)$$

Use this formula to compute the same value of $f(x, y)$, again using different precisions, from 6-decimal to 24-decimal. Describe what you have learned. To force the program to do floating-point operations instead of integer arithmetic, write your numbers in the form 9.0, 40545.0, and so forth.

**19.** Consider the following pseudocode segments:

**a.**
```
integer i; real x, y, z
for i = 1 to 20
    x ← 2 + 1.0/8^i
    y ← arctan(x) − arctan(2)
    z ← 8^i y
    output x, y, z
end for
```

**b.**
```
real epsi ← 1
while 1 < 1 + epsi
    epsi ← epsi/2
    output epsi
end while
```

What is the purpose of each program? Is it achieved? Explain. Code and run each one to verify your conclusions.

**20.** Consider some oversights involving assignment statements.

[a]**a.** What is the difference between the following two assignment statements? Write a code that contains them and illustrate with specific examples to show that sometimes $x = y$ and sometimes $x \neq y$.

```
integer m, n; real x, y
x ← real(m/n)
y ← real(m)/real(n)
output x, y
```

**b.** What value does $n$ receive?

```
integer n;   real x, y
x ← 7.4
y ← 3.8
n ← x + y
output n
```

What happens when the last statement is replaced with the following?

```
n ← integer(x) + integer(y)
```

**21.** Write a computer code that contains the following assignment statements exactly as shown. Analyze the results.

**a.** Print these values first using the default format and then with an extremely large format field:

```
real p, q, u, v, w, x, y, z
x ← 0.1
y ← 0.01
z ← x − y
p ← 1.0/3.0
q ← 3.0p
u ← 7.6
v ← 2.9
w ← u − v
output x, y, z, p, q, u, v, w
```

**b.** What values would be computed for $x$, $y$, and $z$ if this code is used?

```
integer n; real x, y, z
for n = 1 to 10
    x ← (n − 1)/2
    y ← n^2/3.0
    z ← 1.0 + 1/n
    output x, y, z
end for
```

**c.** What values would the following assignment statements produce?

```
integer i, j; real c, f, x, half
x ← 10/3
i ← integer(x + 1/2)
half ← 1/2
j ← integer(half)
c ← (5/9)(f − 32)
f ← 9/5c + 32
output x, i, half, j, c, f
```

**d.** Discuss what is wrong with the following pseudocode segment:

```
real area, circum, radius
radius ← 1
area ← (22/7)(radius)^2
circum ← 2(3.1416)radius
output area, circum
```

**22.** Criticize the following pseudocode for evaluating $\lim_{x\to 0} \arctan(|x|)/x$. Code and run it to see what happens.

```
integer i; real x, y
x ← 1
for i = 1 to 24
    x ← x/2.0
    y ← arctan(|x|)/x
    output x, y
end for
```

**23.** Carry out some computer experiments to illustrate or test the programming suggestions in Appendix A. Specific topics to include are these:

**a.** When to avoid arrays.

**b.** When to limit iterations.

**c.** Checking for floating-point equality.

**d.** Ways for taking equal floating-point steps.

**e.** Various ways to evaluate functions.

*Hint*: Comparing single and double precision results may be helpful.

**24.** **(Easy/Difficult Problem Pairs)** Write a computer program to obtain the power form of a polynomial from its roots. Let the roots be $r_1, r_2, \ldots, r_n$. Then (except for a scalar factor) the polynomial is the product

$$p(x) = (x - r_1)(x - r_2)\cdots(x - r_n).$$

Find the coefficients in the expression

$$p(x) = \sum_{j=0}^{n} a_j x^j$$

Test your code on the **Wilkinson polynomials** in Computer Exercises 3.1.10 and 3.3.9. Explain why this task of getting the power form of the polynomial is *trivial*, whereas the inverse problem of finding the roots from the power form is quite difficult.

**25.** A **prime number** is a positive integer that has no integer factors other than itself and 1. How many prime numbers are there in each of these open intervals: $(1, 40)$, $(1, 80)$, $(1, 160)$, and $(1, 2000)$? Make a guess as to the percentage of prime numbers among all numbers.

**26.** Mathematical software systems such as Maple, Mathematica, and MATLAB are able to do both numerical calculations and symbolic manipulations. Verify symbolically that a nested multiplication is correct for a general polynomial of degree ten.

**27.** In MATLAB, the `rat` function finds a rational fraction approximation (numerator and denominator) within a certain tolerance to a given floating-point number. For example, `[a,b]=rat(pi, 8000e-6)` return `a=22` and `b=7`. However, the relative error between $19/6$ and $\pi$ is 0.007981306248670 in format `long`, which is less than the tolerance 0.008. What's going on here? In terms of absolute and relative errors, is $19/6$ or $22/7$ the better approximation to $\pi$?

**28.** Use mathematical software to reproduce the three solutions to Example 1.1.2.
*Hint*: In MATLAB, use commands `str2nun(num2str(x,4))` for rounding to four significant decimal digits as well as `format long`.

**29.** Explain the results from coding and executing the following pseudocode using mathematical software such as in MATLAB with `format long`: