D Answers to Selected Odd Numbered Problems

Nerd 2: What are you going to do, Mr. Simpson?
Homer: Actually, I've been working on a plan. During the exam, I'll hide under some coats, and hope that somehow everything will work out.
Nerd 2: Or, with our help, you can cram like you've never crammed before!
Homer: Whatever. Either way is good.

From: *The Simpsons*

Chapter 1

1. $3x^2 - 10x - \frac{3}{x^2} - \frac{1}{x^{3/2}}$

3. $e^x\left(\ln x + \frac{1}{x}\right)$

5. $\frac{2\ln x}{x}$

7. $\frac{e^x(1+x^2-2x)}{(1+x^2)^2}$

9. $15(1+3x-x^3)^4(1-x^2)$

11. $-\frac{6}{(2x+1)^4}$

13. $\frac{1}{2x\sqrt{\ln x}}$

15. $\frac{2xe^{x^2}}{e^{x^2}+1}$

17. ke^{kx}

19. $\frac{x^3}{3} + \frac{2}{x} + 2e^x + C$

21. Expand integrand first:
Ans. $\frac{x^2}{2} - 2x + \ln x + C$

23. Let $u = kx$: Ans. $\frac{1}{k}e^{kx} + C$

25. Expand integrand first
Ans. $\frac{1}{2}e^{2x} + 2x - \frac{1}{2}e^{-2x} + C$

27. Let $u = 2x+1$: Ans. $-\frac{1}{4}(2x+1)^{-2} + C$

29. integrand = $1 - \frac{1}{x+2}$
Ans. $x - \ln(x+2) + C$

31. Let $u = -x^2$: Ans. $-\frac{1}{2}e^{-x^2} + C$

33. Let $u = \sqrt{x}$: Ans. $2e^{\sqrt{x}} + C$

35. Let $u = 1 + x^3$: Ans. $\frac{2}{3}\sqrt{1+x^3} + C$

37. Let $u = e^x - e^{-x}$:
Ans. $\ln(e^x - e^{-x}) + C$

Chapter 2

1. $\frac{dy}{dx} = xe^x + e^x + Ce^x$, $e^x + y = e^x + (xe^x + Ce^x)$. $\therefore \frac{dy}{dx} = e^x + y$.

3. a) $\frac{dy}{dx} = 1 - Ce^{-x}$; $x - y = x - (x - 1 + Ce^{-x}) = 1 - Ce^{-x}$ $\therefore$ $\frac{dy}{dx} = x - y$

   b) $y = x - 1 + 3e^{-x}$

5. a) $\frac{dy}{dx} = \frac{2x - \sqrt{2x^2 + C}}{\sqrt{2x^2 + C}}$; $\frac{x - y}{x + y} = \frac{2x - \sqrt{2x^2 + C}}{\sqrt{2x^2 + C}}$ $\therefore \frac{dy}{dx} = \frac{x - y}{x + y}$.

   b) $y = -x + \sqrt{2x^2 + 2}$

7. a) , c) and d) are separable

9. $y = -\frac{2}{2x^2 - 1}$

11. $y = \frac{1}{2}\ln(t^2 + e^2)$

13. $y = 2e^{x^2/2 + x}$

15. $y = \sqrt{2t^2 + 4t + 4}$

17.

	5 years	10 years
a)	\$1284.03	\$1648.72
b)	\$1343.54	\$1641.01

19. a) 8.1 years

    b) 4.05 years (in fact exactly half as long as the time needed in a))

21. Derive and use formula $t_3 = \frac{\ln 3}{r}$.

r	.03	.05	.07	.10	.15
# years	36.6	22.0	15.7	11.0	7.34

23. a) In time Δt the change in principal

$$\Delta P = \text{Interest + New Deposits} = .06P \times \Delta t + K \times \Delta t .$$

Divide by Δt and take limit as $\Delta t \to 0$.

b) Write the equation in form $\dfrac{dP}{dt} = .06\left(P + \dfrac{K}{.06}\right)$ and then use separation of variables. Alternatively let $\tilde{P} = P + \dfrac{K}{.06}$ and observe that since $\dfrac{d\tilde{P}}{dt} = \dfrac{dP}{dt}$ we have $\dfrac{d\tilde{P}}{dt} = .06\tilde{P}$ and we can solve for $\tilde{P}$ in the usual way. In either case you find that $P = \dfrac{K}{.06}(e^{.06t} - 1)$.

c) $K = \$861.65$, total deposits = $K \times 25 = \$21,541.25$, interest = $\$28,458.75$.

25. a) In time Δt the change in principal $\Delta P = \text{Interest} - \text{Withdrawals} = .05P \times \Delta t - 10 \times \Delta t$. Divide by Δt and take the limit as $\Delta t \to 0$.

b) Write the equation in the form $\dfrac{dP}{dt} = .05(P - 200)$ and then use separation of variables or a change of variables as described for 23b above. Obtain $P = 200 + (P_0 - 200)e^{.05t}$.

c) $P_0 \approx \$126,400$

Chapter 3

1. Using Theorem 3.1 you obtain $\sqrt{26} \approx 5.1$. Calculator value is $\sqrt{26} \approx 5.099$. Error is approximately .001.

3. a) $y'(0) = .25$

b) $y(0.5) \approx y(0) + y'(0)(0.5) = 1.125$

c) $y(1) \approx y(0) + (1)y'(0.5) = 1.158$ (Estimate $y'(0.5)$ using the ODE and the answer to b)

5.

t	y	y'
0	1.000	1.000
0.25	1.250	1.188
0.5	1.547	1.297
0.75	1.871	1.309
1	2.198	1.198

7. a) Solution of the initial value problem is $\ln x$. Value of solution at $x = 2$ is $\ln 2$.

b)

x	y	y'
1	0.000	1.000
1.25	0.250	0.800
1.5	0.450	0.667
1.75	0.617	0.571
2	0.760	0.500

From a calculator $\ln 2 \approx .693$, so error is approximately $.067$.

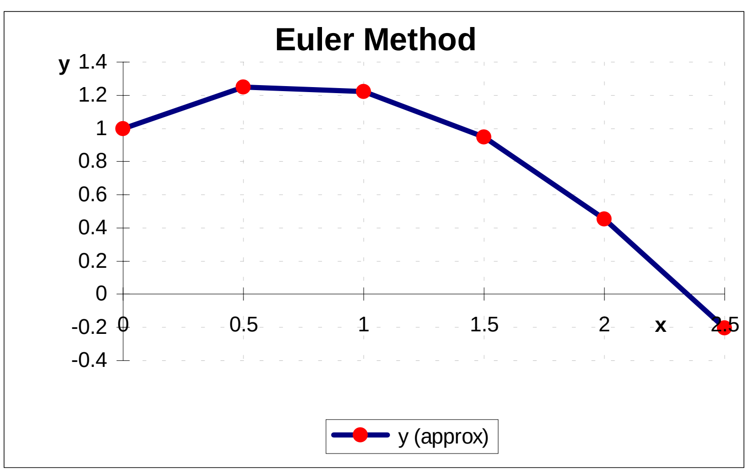
9. a) $\Delta x = 2.5/5 = .5$. The last entry in column 2 is obtained as

$$y_{old} + y'_{old}\Delta x = .454 + (-1.312)(.5).$$

The last entry in column 3 is obtained by substituting $x = 2.5$ and $y = -.202$ into the right side of the differential equation.

x	y	y'
0	1.000	0.500
0.5	1.250	-0.056
1	1.222	-0.550
1.5	0.947	-0.986
2	0.454	-1.312
2.5	-0.202	-1.247

b) Plot the points $(0,1)$, $(.5,1.25)$, $(1,1.222)$, etc. from columns 1 and 2 in the table above. Connect the points with line segments to complete graph, as shown below.



11. First estimate $\sqrt{25.5}$ using Theorem 3.1: Get $\sqrt{25.5} \approx 5.05$. Then Theorem 3.2 gives $\sqrt{26} \approx 5.099009$ compared to calculator value $\sqrt{26} \approx 5.099020$. Error is 1.1×10^{-5}.

13. a)

x	y	y'	$x_{1/2}$	$y_{1/2}$	$y'_{1/2}$
0	1	1	0.125	1.125	1.125
0.25	1.281	1.281	0.375	1.441	1.441
0.5	1.642	1.642	0.625	1.847	1.847
0.75	2.103	2.103	0.875	2.366	2.366
1	2.695	2.695	1.125	3.032	3.032

N.B. The column labeled $x_{1/2}$ is not necessary for this problem since the derivative only depends on the current y value.

The estimate for e is the last value in the second column: 2.695.

b) The value of e from a calculator is ≈ 2.718, so the error in the estimate is about .023.

Chapter 4

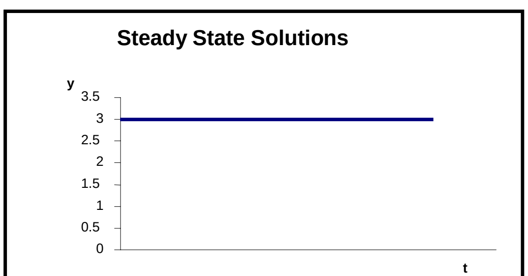
1. $y = 3$ (See graph below)

3. $y = -1$, $y = 3$

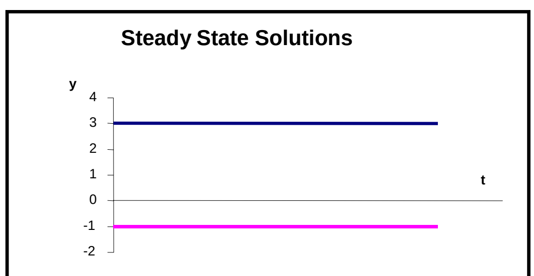
5. $y = 0$, $y = 4$

7. $y(3-y)+10 = -(y-5)(y+2)$, therefore $y = 5$, $y = -2$

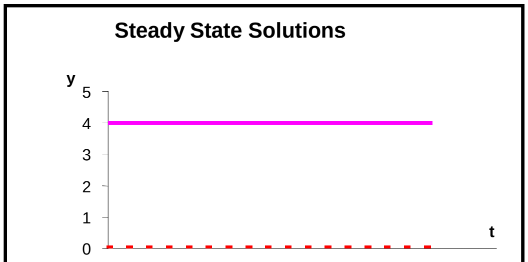
9. No steady state solutions



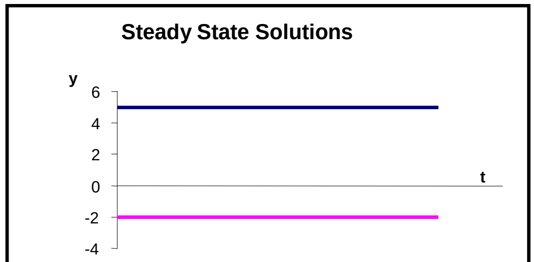
Exercise 1



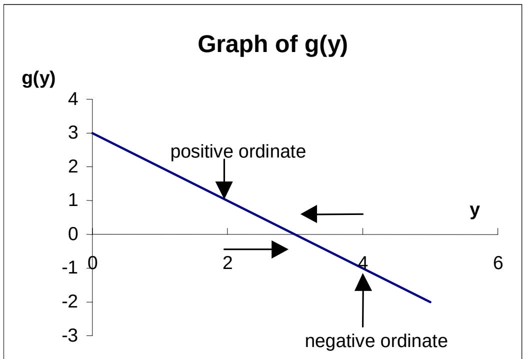
Exercise 3



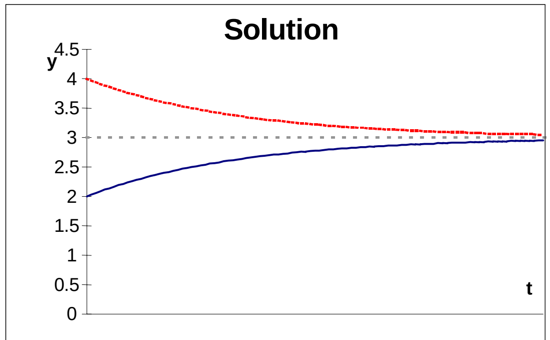
Exercise 5



Exercise 7

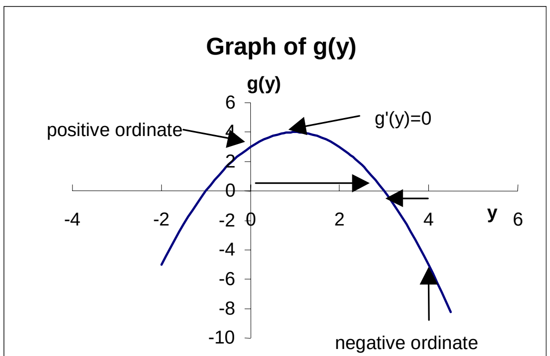


Exercise 11

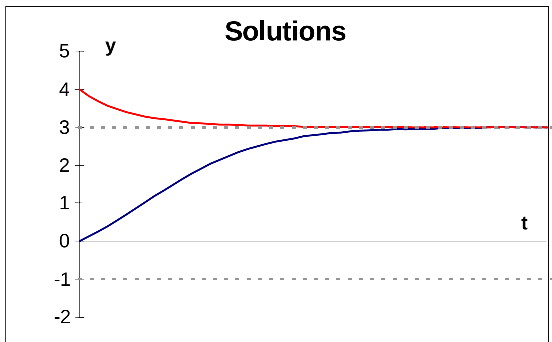


Exercise 11

11. On the graph of $g(y)$ the ordinate at $y=2$ is positive so the solution that begins at two increases approaching the steady state value $y=3$. There are no points of inflection since $g'(y)\neq 0$. The solution curve is concave down since for $y<3$, $g(y)g'(y)<0$. A similar argument applies to the solution with $y(0)=4$.



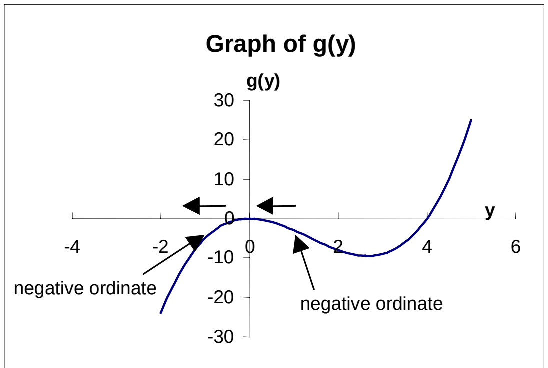
Exercise 13



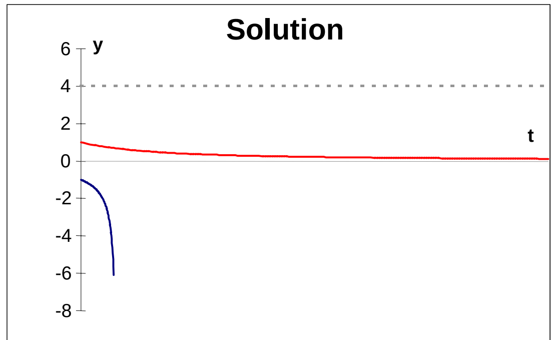
Exercise 13

13. On the graph of $g(y)$ the ordinate at $y=0$ is positive. Therefore the solution with initial value $y(0)=0$ increases towards the steady state solution above it, $y=3$. Since $g'(y)=0$ when $y=1$, the solution curve that begins at $y=0$ will have an inflection point when its ordinate passes through $y=1$. For y between 0 and 1 the solution graph is concave up, since $g(y)g'(y)>0$ for those y values.

The solution of the initial value problem $y(0)=4$ decreases monotonically towards the same steady state $y=3$. The solution graph is concave up since $g(y)g'(y)>0$.

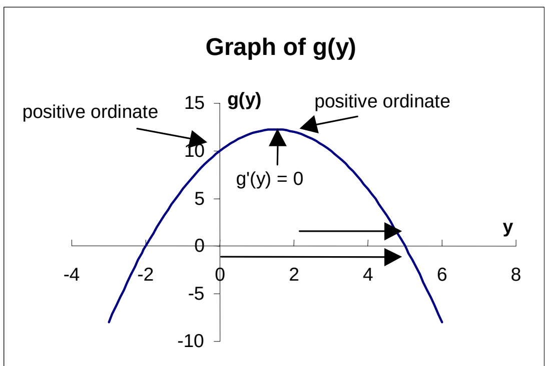


Exercise 15

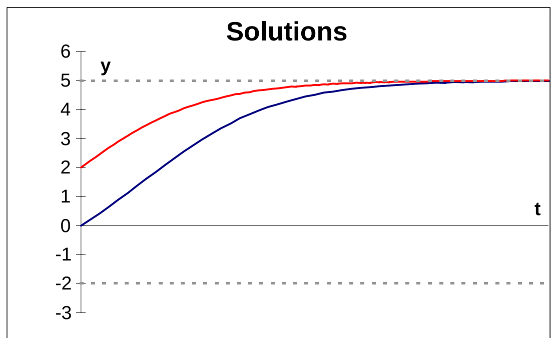


Exercise 15

15. The solution satisfying $y(0)=0$ is steadily decreasing towards $-\infty$. The graph is concave down since for $y<0$ we have $g(y)g'(y)<0$. The solution satisfying $y(0)=1$ also decreases, approaching the steady state solution $y=0$. The graph of the solution is concave up since $g(y)<0$ and $g'(y)<0$ implies that $g(y)g'(y)>0$.

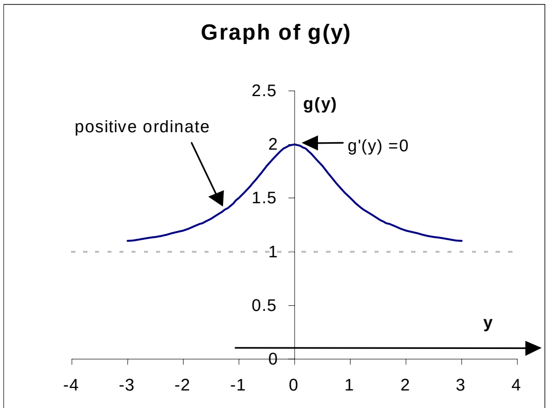


Exercise 17

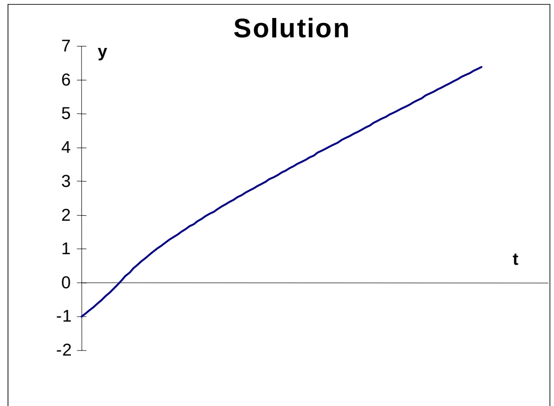


Exercise 17

17. Both solutions increase monotonically towards the steady state solution $y=5$. However, the solution with initial value $y(0)=0$ passes through $y=1.5$ where $g'(y)=0$. Therefore, this solution has an inflection point, where the concavity changes from concave up to concave down. The solution with initial value $y(0)=2$ is concave down since for y between 2 and 5 $g(y)g'(y)<0$.



Exercise 19

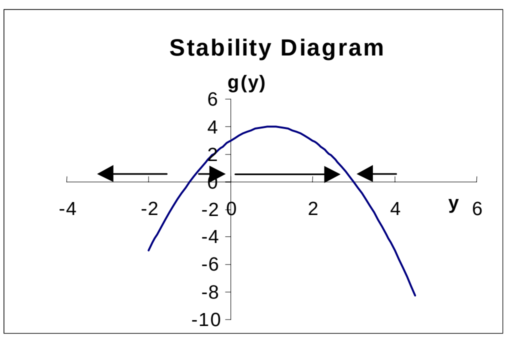


Exercise 19

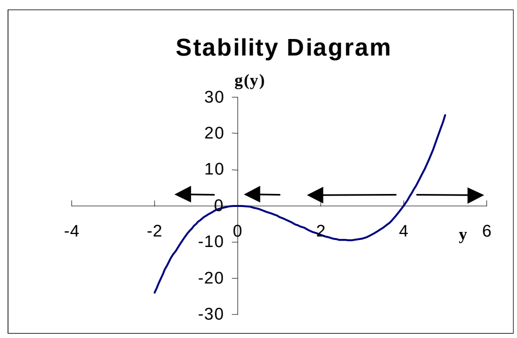
19. The ordinate on the graph of $g(y)$ is positive for all y and there are no steady state solutions. The solution of the initial value problem increases indefinitely. Between $y=-1$ and $y=0$ the solution is concave up. For $y>0$ the graph is concave down. The solution graph appears to look like a straight line as t increases. This is because the derivative (slope) $y'(t)=g(y)$ and $g(y)$ approaches one as y increases.

21. $y=3$ is stable. See graph of $g(y)$ given in exercise 11.

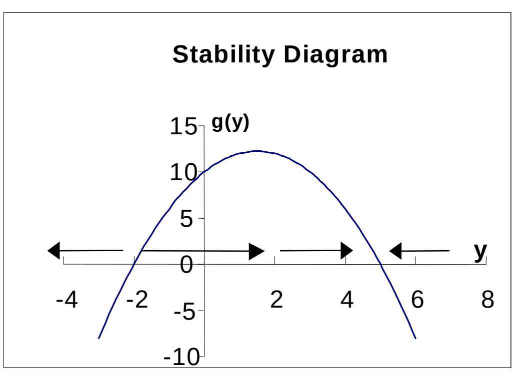
23. $y=-1$ is unstable, $y=3$ is stable



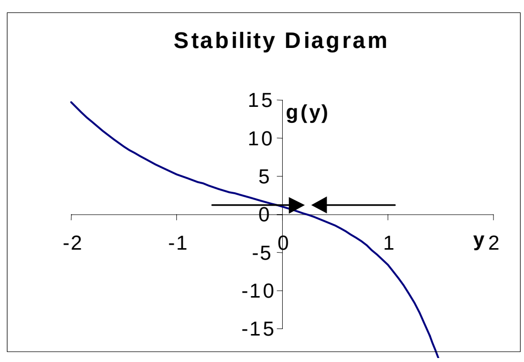
25. $y=0$ is semi-stable. $y=4$ is unstable



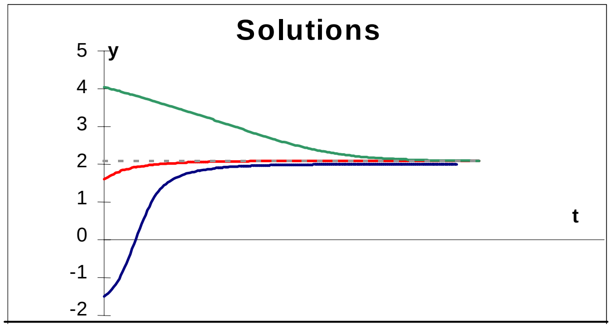
27. $y=-2$ is unstable. $y=5$ is stable



29. The steady state $y=\frac{\ln 2}{3}$ is stable



31.



33. b) $y = \frac{1}{1-t}$. Solution "blows up" at $t = 1$.

Chapter 5

1. a) $N = 57e^{-.00105t}$, units of population N in millions

   b) 53.9 million

   c) annual % decrease is approximately .1 %

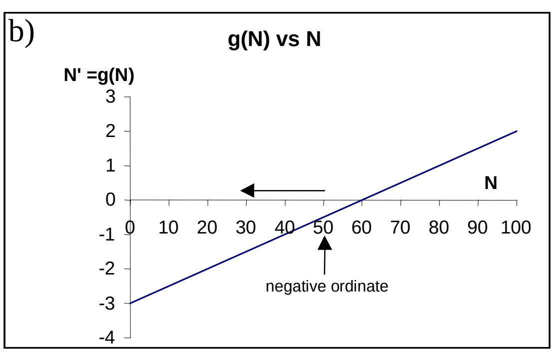
3. Equality of populations in 67.5 years (from 1997). Populations of each country will be 391 million.

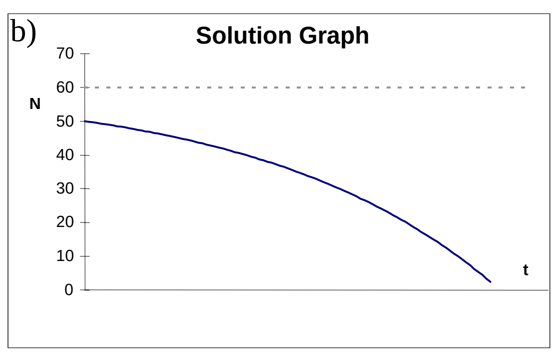
5. a) $r = .0182$ or 18.2 per 1000

   b) $r = .0162$ or 16.2 per 1000. Some support for a decrease in the world's population growth rate.

7. a) $\frac{dN}{dt} = .05N - 3$, where the units of N are in thousands. Assumes that reproduction and harvesting occur uniformly throughout the year. Also assumes that the harvesting does not affect the reproductive rate.

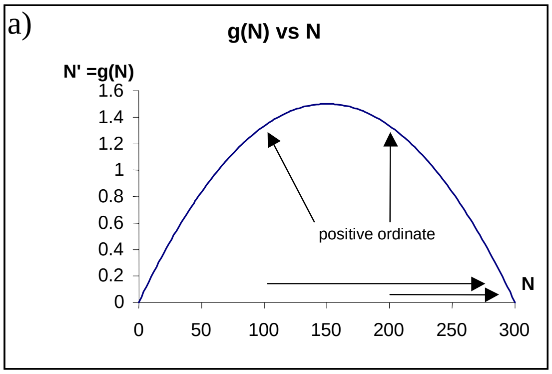
   b) As shown by the graph below of $g(N)$ vs. N, with an initial value of 50 (000) the population will decrease and continue to do so.

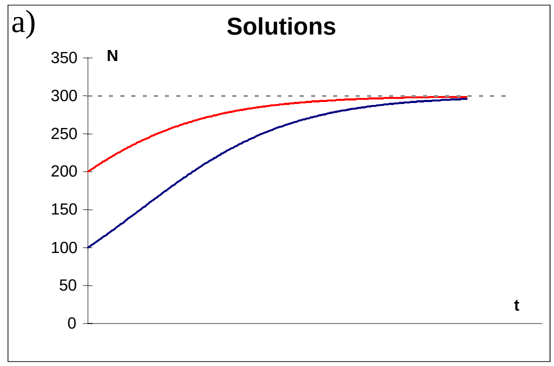




c) The population will eventually be wiped out. The exact solution of the differential equation is $N(t) = 60 - 10e^{.05t}$. Using this formula, the time to extinction is about 35.8 years.

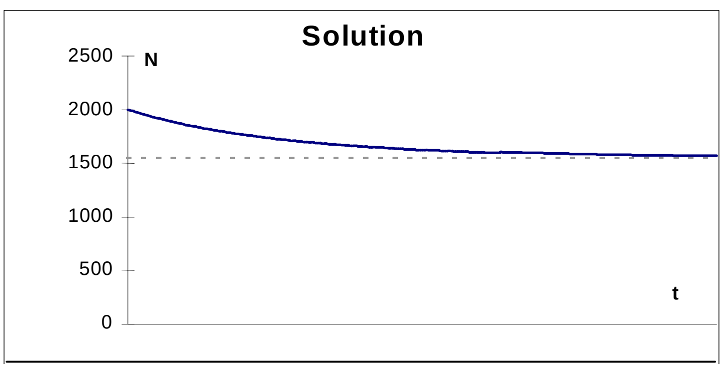
9 .





b) For the solution of the initial value problem $N(0) = 100$, the maximum growth rate occurs when $N = 150$ (see graph of $N' = g(N)$). For the initial value problem with $N(0) = 200$ the maximum growth rate occurs initially when $N = 200$.

13 .New $K = 1540$. Revised logistic is $\frac{dN}{dt} = .03N\left(1 - \frac{N}{1540}\right)$. The population will decrease asymptotically to the new carrying capacity, as shown below.

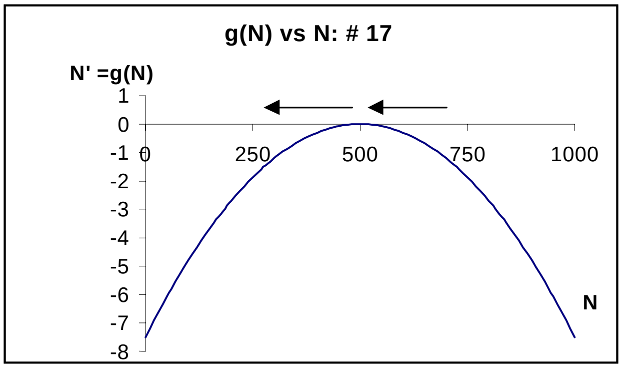


15. a) As written the equation is not autonomous. Using separation of variables the general solution is $V(t) = Ce^{-2e^{-t}}$. With the initial value $V(0) = 0.25$, we obtain the specific solution $V(t) = 1.85e^{-2e^{-t}}$. The tumor volume approaches 1.85 as t increases.

17. a) If we harvest an amount N_1 then the differential equation for the growth rate becomes $\frac{dN}{dt} = g(N) - N_1$. In particular, if $N_1 = g(N_0)$ then the equation is $\frac{dN}{dt} = g(N) - g(N_0)$,

which has as a steady state solution $N \equiv N_0$. Thus, the population size N_0 will persist indefinitely if the harvest size is $g(N_0)$.

b) Since $g(N_0)$ is the sustainable yield for a population of size N_0, to maximize the sustainable yield we need to find N_0 so that $g(N_0)$ is as large as possible. This occurs (for most functions of interest) when $g'(N_0) = 0$. For the model in exercise, 16 the maximum sustainable yield occurs for a population size of 500. The maximum yield is 7.5. After accounting for the harvesting as described in 17a), the graph of $g(N)$ vs. N (below) shows that the steady state solution $N = 500$ is not stable. In particular, if the population randomly drifts below 500 it will continue to decrease.



Chapter 6

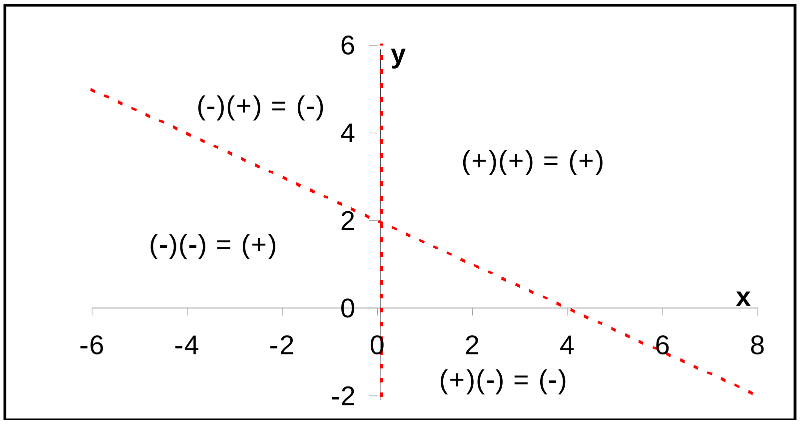
1. a) $x = 3/2,\ y = 1/2$

   b) $x = 25,\ y = 75$

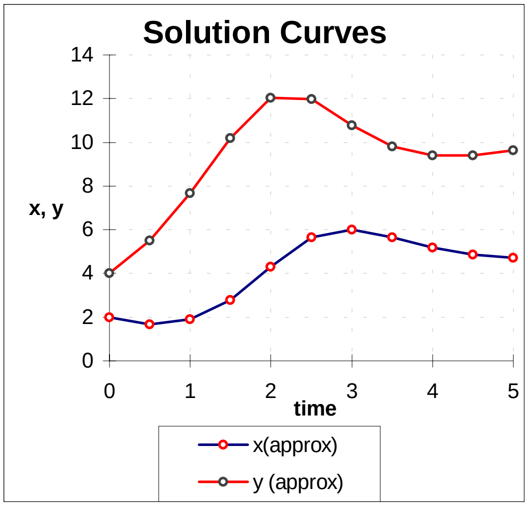
   c) $\{x = 0,\ y = 0\},\ \{x = 5,\ y = 10\}$

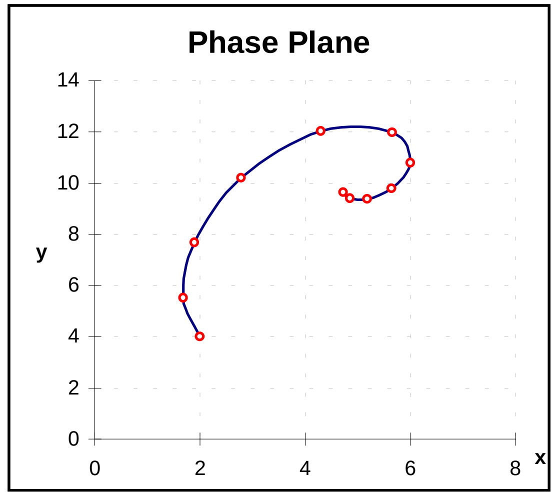
   d) $\{x = 0,\ y = 0\},\ \{x = 0,\ y = 100\},\ \{x = 200,\ y = 0\},\ \{x = 200/3,\ y = 100/3\}$

3.



5 a)





b) Steady states are $x=0$, $y=0$ and $x=5$, $y=\pm 10$. The solution in a) appears to be approaching the steady state $x=5$, $y=10$.

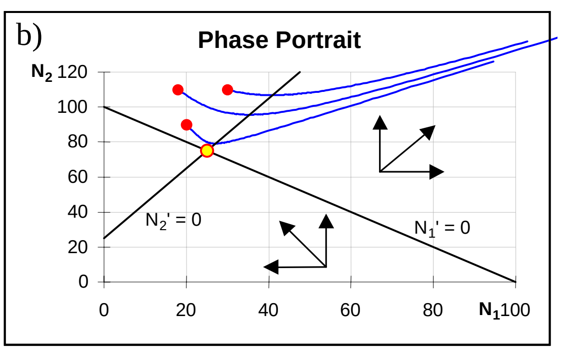
7. A corresponds to III, since the solutions in A appear to be periodic. The curve with the larger amplitude is y.

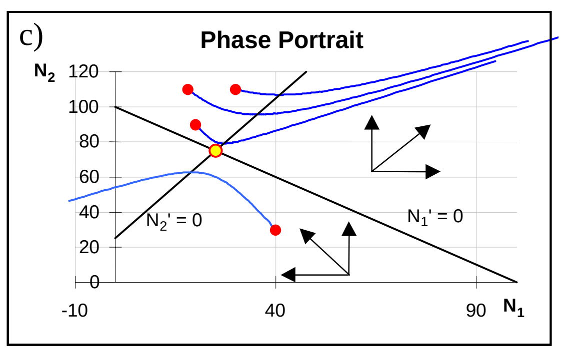
B corresponds to I, since the solutions oscillate but appear to be approaching a limiting value. The curve with the larger amplitude is y.

C corresponds to IV, since the x value is initially decreasing and then both increase. The lower curve in IV corresponds to x.

D corresponds to II, since both solutions increase initially. The lower curve in II corresponds to y, since the y values in the phase plot are much smaller than the x values.

9. a) See the solution to 1b.



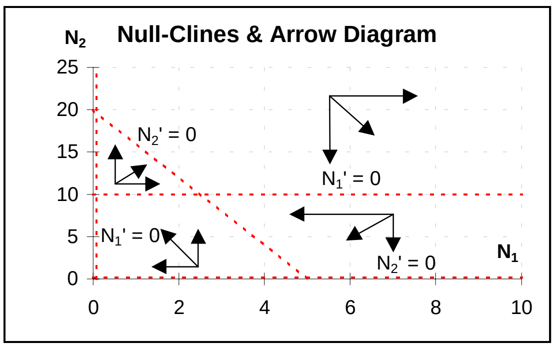


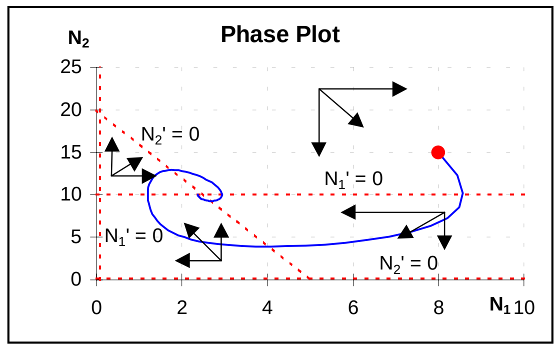
c) The actual solution to the initial value problem is shown in the figure above, labeled c). It is also conceivable that the solution moves up and to the left as shown, but then passes vertically through the N_1 null-cline and heads off to the upper right. Deciding between these two possibilities on the basis of theory requires more advanced techniques than we have studied.

11. From the differential equations in exercise 5 we have $x'(0) = -2 + .05(4)^2 = -1.2$ and $y'(0) = 4 - .2(2)(4) = 2.4$. Using a $\Delta t = .5$ the Euler update equation applied to each function gives $x(.5) \approx x(0) + .5x'(0) = 1.4$ and $y(.5) \approx y(0) + .5y'(0) = 5.2$.

13. a) 20

b) Steady states are $\{N_1 = 0, N_2 = 0\}$, $\{N_1 = 0, N_2 = 20\}$, $\{N_1 = 2.5, N_2 = 10\}$. The N_1 null-clines are $N_1 = 0$ (vertical axis), $N_2 = 10$, and the N_2 null-clines are $N_2 = 0$ (horizontal axis) and $\frac{N_1}{5} + \frac{N_2}{20} = 1$. The arrow diagram is shown in the left panel below.

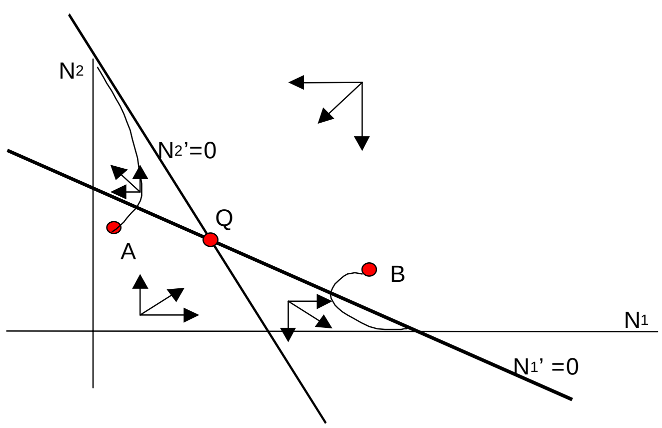




The actual solution is shown on the right panel above. Other possibilities are also consistent with the arrow diagram.

15. See the solution to 1d.

17. a)

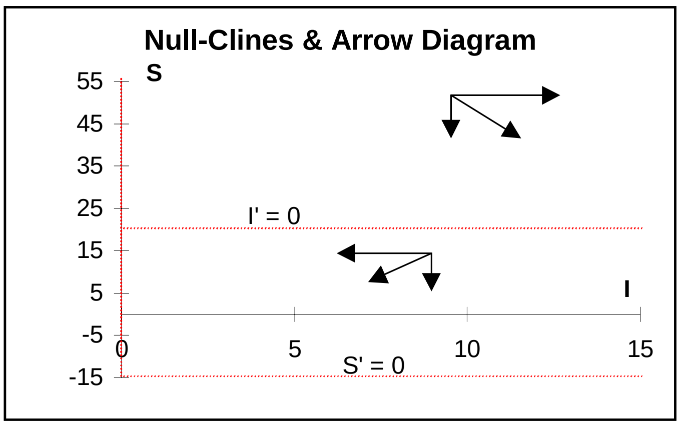


In each case the phase curves approach a limit. One species is pushed to extinction and the other approaches its environmental carrying capacity.

b) The intercepts of the two slanted null-clines on the N_1 axis are at $N_1 = K_1$, the intercept of the line $N_1' = 0$, and $N_1 = \frac{K_2}{a_{12}}$, the intercept of $N_2' = 0$. From the picture we must have $\frac{K_2}{a_{12}} < K_1$ or $\frac{a_{12}}{K_2} > K_1$. This means that species 1 affects species 2 more than it affects itself. Examining the intercepts on the N_2 axis gives a similar result with species 1 and 2 interchanged. Thus the species are competitive but the outcome of the competition will depend on the initial sizes of the populations. The two cases drawn illustrate that different final scenarios are possible depending on the initial values of N_1 and N_2.

19. a) All points on the vertical (S axis) are steady states.

b) The null-clines are the lines (I null-clines): $S = 20,\ I = 0$ (S null-clines): $S = -.15,\ I = 0$

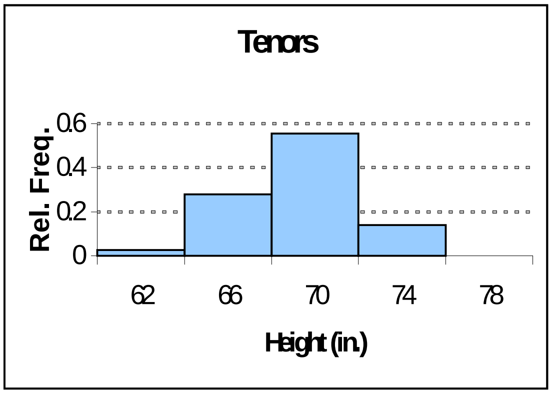


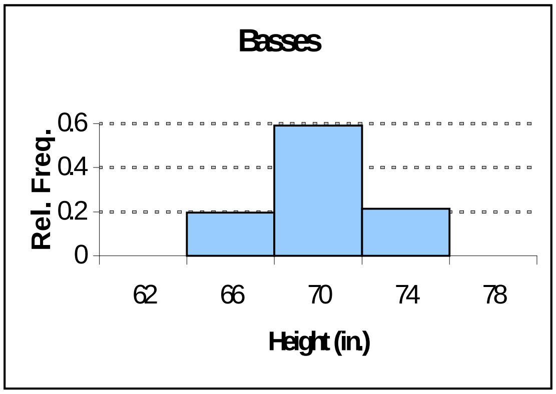
If the initial value of S exceeds 20 the disease will spread. This is the same epidemic threshold as before the immunization program. The immunization program does not affect this level because no action is taken until the disease actually appears.

21. Using the arrow diagram (Figure 6.8), we see that solutions that begin near a value $I = 0$, $S > 20$ will wind up approaching a steady state with $S < 20$. Thus the steady states with $S > 20$ are unstable. For initial values for which I is close to zero and S is near a value that is below 20, the solution will move to a nearby steady state with $I = 0$ and S near its original value. Thus the steady states with $I = 0$ and $S \leq 20$ are stable.

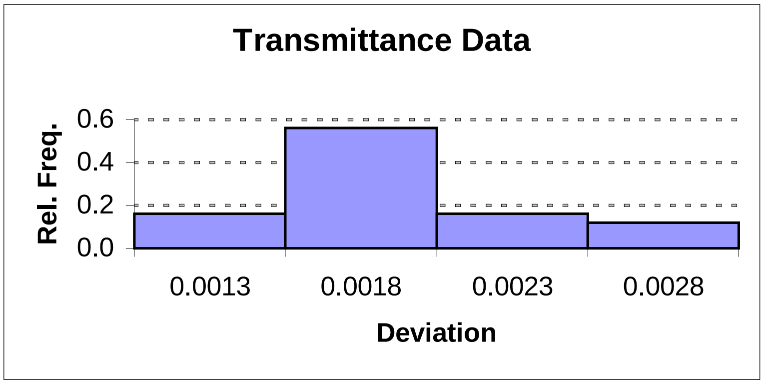
Chapter 7

1. The relative frequency histograms for the two groups are shown below. In order to make a comparison we use the same bin intervals for each group. The histograms show some evidence that basses are taller than tenors, since the latter histogram is slightly skewed to the left.





3.

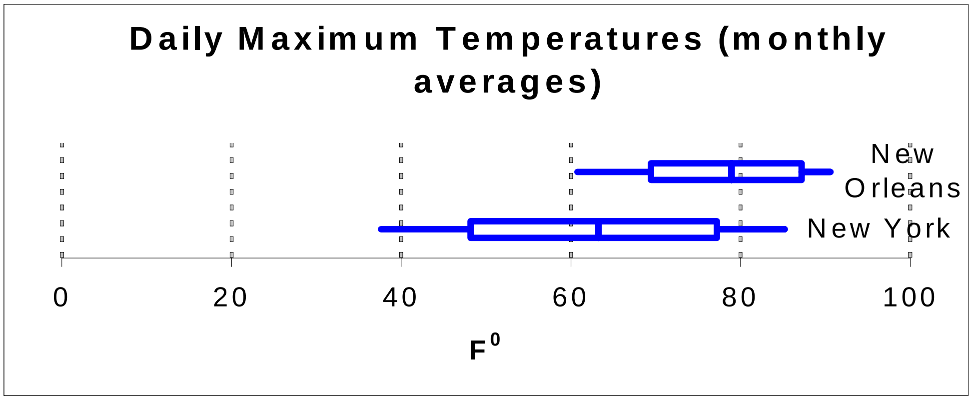


5. a) median = 4, mean = 4.33, standard deviation = 4.08, variance = 16.65
   b) median = 19, mean = 19.33, standard deviation = 4.08, variance = 16.65
   c) median =-20, mean = -21.67, standard deviation = 20.4, variance = 416.2
   d) Adding a constant k to all data values adds the same constant to the mean, but leaves the variance and standard deviation unchanged. Multiplying all data values by a constant k, multiplies the mean by the same constant, multiplies the variance by k^2, and the standard deviation by $|k|$.

7. The relevant intervals are: one standard deviation around the mean, [.773, .931]. Contains 67 out of 100 data values or 67%. Two standard deviations around the mean, [.694, 1.01]. Contains 97 data values or 97%. Three standard deviations around mean, [.615, 1.09]. Contains all data values.

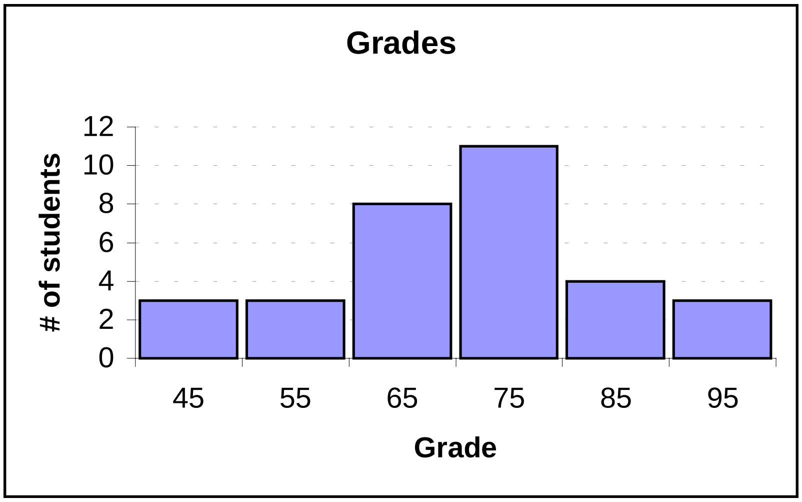
9. In this case the average is not a very good estimate for a “typical” value. Most values tend to be either 15 to 20 minutes more or less than the average. In either situation, your advice would probably generate complaints. We might try to predict the time to next eruption from the duration of the previous eruption. See exercise 16 in chapter 9.

11.



New York daily maximum shows a much larger variation than the corresponding statistic for New Orleans.

13. a)



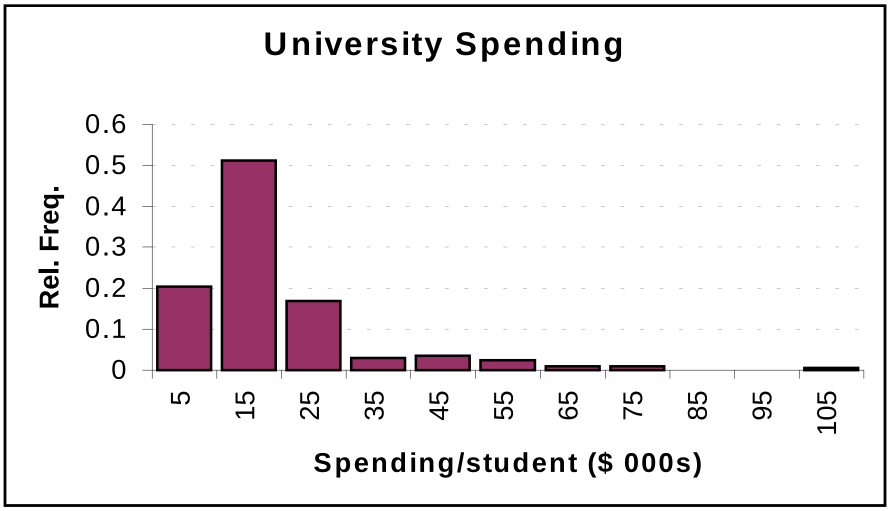
Note that the horizontal scale does not begin at zero.

b) There are 32 grades so the median is the average of the 16th and 17th when sorted. This would probably be approximately 71 or 72. The value of q_1 is approximately the 8th sorted grade from the bottom, placing it around 63. By similar reasoning $q_3 \approx 79$.

15. a) $m \approx \$16\text{ K}$. Reasoning: the median (50th percentile) is 30 percentage points above the 20th percentile (located at $10 K) and is therefore about 3/5 of the way to the 70th percentile located at $ 20K. Three-fifths of the way from $10 K to $20 K is $ 16 K. (Actual $ 15.2 K)

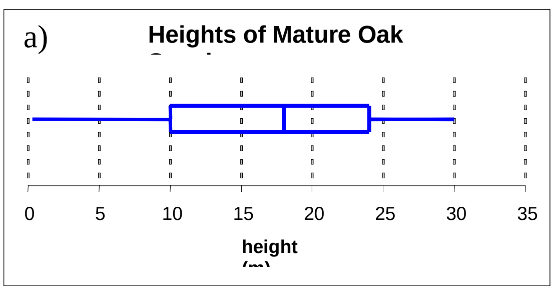
Using similar reasoning $q_1 \approx 11$ (actual 11), $q_3 \approx 21.5$ (actual 21.6).

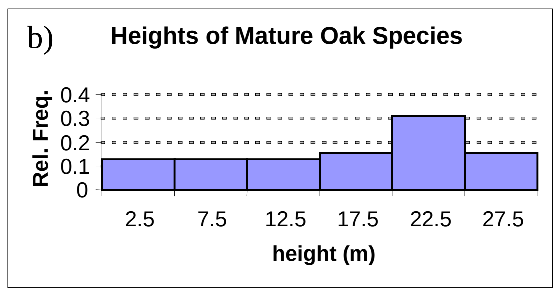
The estimate for $\bar{x}$ is $3785/201 \approx 18.8$ (actual 19.0) and the estimate for $s \approx 14.4$ (actual 13.9) b)



The overall distribution is skewed to the right, as is the distribution of expenses for the top 50 schools. However, the distribution of the lower 90 % (everything in the box labeled 35 or below) is somewhat bell-shaped.

17. a) $m = 18$ m., $q_1 = 9$ m., $q_3 = 24$ m.





c) Both plots show a left skewed distribution. The mean will be smaller than the median (actual $\bar{x} = 17.3$ m).

19. 84.2

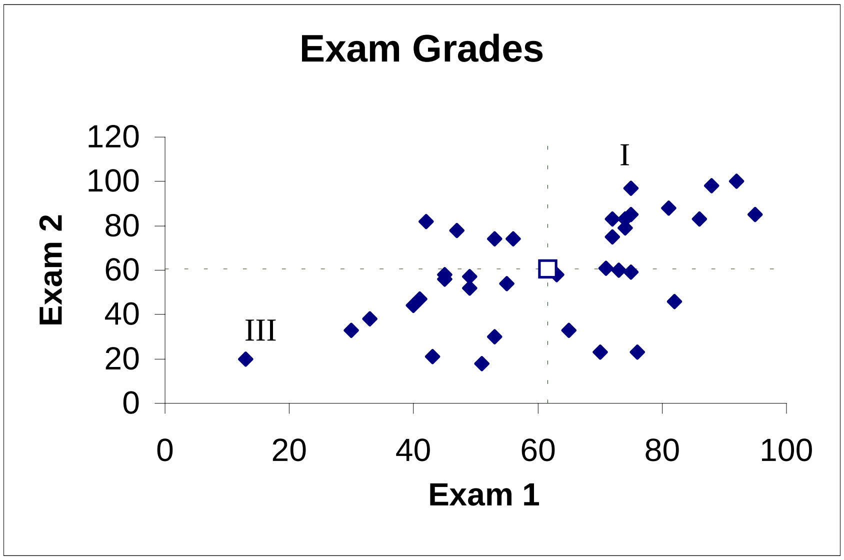
Chapter 8

1. a) $s_x = 1.98$, $s_y = 2.04$, $c_{xy} = 1.21$ each rounded to two decimal places. The value of $c_{xy}/(s_x s_y) = .30$, to the second decimal place, thus verifying the answer provided by *Excel*.

   b) The new x values equal the old values plus one. The new y values equal the old values plus one. Therefore, all four quantities computed in a) are the same for the new x and y.

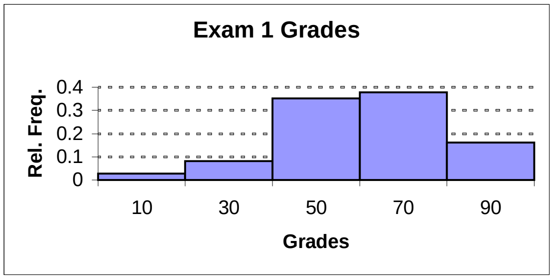
c) The x values are twice the values in a) and the y values are three times the y values in a). Therefore $s_x = 2\times1.98 = 3.96$, $s_y = 3\times2.04 = 6.12$, $c_{xy} = 6\times1.21 = 7.26$. The value of r will remain unchanged.

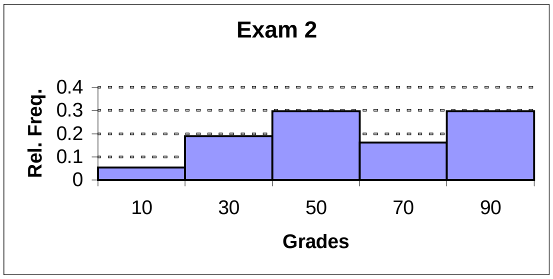
3. a) The graph below shows the point $(\bar{x}, \bar{y})$ marked with a □. The number of points in the regions labeled I and III is approximately 25 out of 37. These points exhibit positive correlation. The remaining 12 points show negative correlation. We would thus expect only a moderate positive value of r. In fact $r = .585$.



b) range x ≈95 - 15 =80, range y ≈100 - 15 = 85.

c) To apply the range rule, the x and y grades need to have an approximately bell-shaped distribution. This is somewhat difficult to deduce from the scatter plot. Histograms for the grade distributions on each exam are given below.





The grade distribution for exam 1 is approximately bell-shaped, but the exam 2 grades appear bimodal. We would expect the range rule to work well for exam 1, but not so well for exam 2. Indeed we find that $\text{range}_x / 4 \approx 80/4 = 20$ compared to the actual value of 19.1, while $\text{range}_y / 4 \approx 85/4 = 21.25$, compared to the actual value of $s_y = 24.5$.

5. Moderately positive. As the car gets older it will require more maintenance. However, the expenditures may rise more rapidly with the car's age so a non-linear pattern would probably better describe the observed relationship.

7. a) The correlation coefficient for each group should be close to zero.

   b) Pooling the data there are 23 data points. The average of the x values will equal

$$\frac{\text{sum of } x \text{ values}}{23} = \frac{\text{sum of } x \text{ for squares} + \text{sum of } x \text{ for circles}}{23}$$
$$= \frac{11 \times \bar{x}_{sq} + 12 \times \bar{x}_{circ}}{23} = \frac{11(0) + 12(3.4)}{23} \approx 1.8$$

   Similarly the average of the pooled y values is $\frac{11(-.5) + 12(4.1)}{23} = 1.9$. Almost all points show strong positive correlation with respect to the point (1.8, 1.9). The value of r for the combined data should therefore be very close to one.

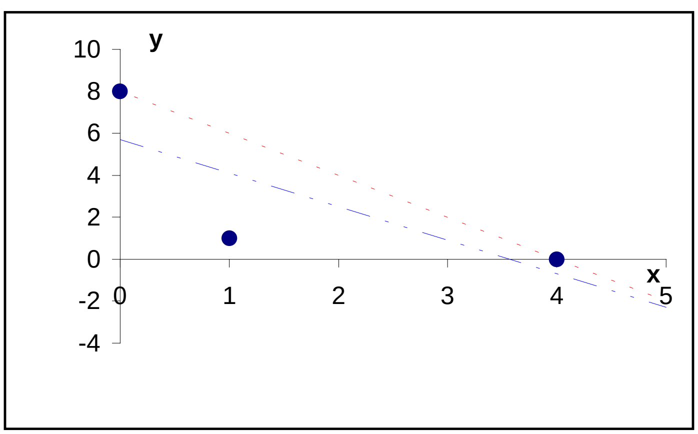
   c) In this case the value of r obtained by pooling the data masks the behavior in the two subgroups. Using the value of r derived from the entire set of data would not be appropriate.

Chapter 9

1. a) $fit = (a + b - 1)^2 + (a + 4b)^2 + (a - 8)^2$

   b) i) Equation of line is $y = 8 - 2x$ so we use $a = 8$ and $b = -2$ in the fitness formula in a), giving a fitness of 25.

   ii) Put $a = 5.7$ and $b = -1.6$ into the fitness formula, giving 15.39. This line has a better fit.



3. From the point-slope formula the equation of the line is $y = 2x + 1$. The equation of the regression line is determined from (rounding to the third decimal place) $\bar{x} = 3.5$, $\bar{y} = 8$, $s_y = 6.245$, $s_x = 3.122$, $c_{xy} = 19.5$. From these we obtain $r = 1.000$, $s_y / s_x = 2.000$ and the regression line is $y - 8 = 2.000(x - 3.5)$, which works out to the same equation we found above. For collinear points the common line containing the points minimizes the fit and therefore must coincide with the regression line.

5. Only b) shows a negative trend (y decreasing with increasing x) and therefore corresponds to iii). The points in c) appear more tightly clustered around the regression line and would therefore yield a higher value of r. Thus c) corresponds to i. and a) corresponds to ii.

7. a) Low positive correlation, perhaps $r \approx .2$

   b) $\hat{y} = 18.1\,m.$. This represents the average height for species whose acorn volume is 5 cm^3. This is barely different than the average of an arbitrary oak species (17.3 m.), reflecting the small size of r.

   c) Using the slope of the regression line we obtain $.42 = r(8.5/3.45)$, giving $r \approx .17$ and thus $r^2 \approx .03$. Since r^2 is close to zero we have $s_{y|x} \approx s_y$. Thus knowing the acorn volume x scarcely reduces the variation in height of species, so the statement appears to be supported by the data.

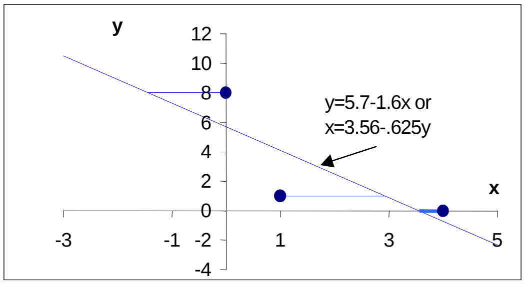
9. a) $s_{\text{Tom}} = 10 s_{\text{Jerry}}$. Since Tom is recording the data in cm, his numbers are ten times larger than those recorded by Jerry.

   b) No. The correlation coefficient is not affected by scaling.

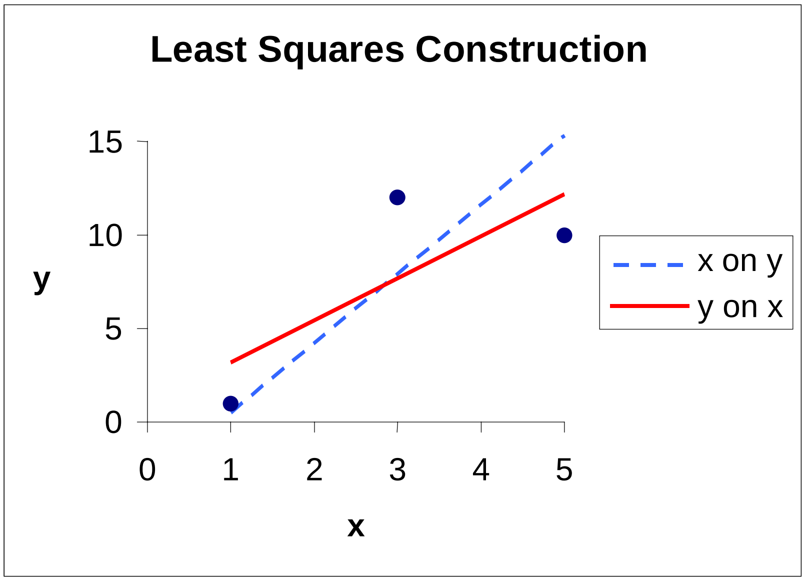
   c) The slope of the regression line is $r s_y / s_x$. By a) and b) this quantity is ten times as large for Tom than for Jerry.

11. a) $fit = (a + b - 1)^2 + (a - 4)^2 + (a + 8b)^2$

    b) i.) The equation of the proposed line is $x = 4 - 0.5y$, so that $a = 4$ and $b = -.5$. The value of the *fit* is 6.25. For ii.) the line has equation $x = 3.56 - .625y$. The value of the *fit* is 6.03, so this line provides the better fit. The sketch is the same as in exercise 1, except that we use the square of the horizontal distances from the given points to the proposed line. The construction is illustrated below.



13. From Example 9.2 we have that $\bar{x}=3$, $s_x=2$, $\bar{y}=7.67$, $s_y=5.86$ and $r=.768$. Thus the equations of the two regression lines are $y=.92+2.25x$ (regression of y on x) and $y=3.82x-3.77$ (regression of x on y).



15. a) The average weight for a man of height 64 inches is obtained by substituting $x=64$ into the regression equation. This yields $\hat{y}=148.8$ lbs. A man of this height who weighed 165 pounds would be above average for this height.

b) The value of $r=slope\times s_x/s_y=4.2\times 2.3/15=.644$. Thus

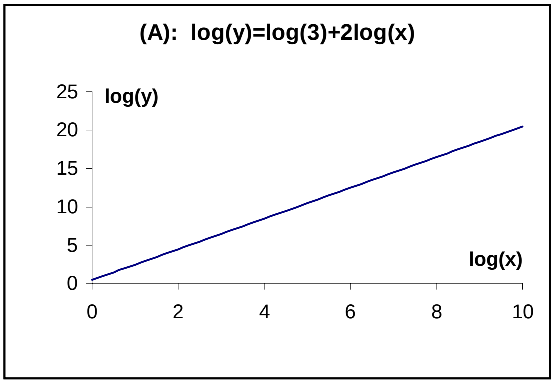
$$s_{y|x}\approx\sqrt{1-r^2}s_y=(.765)15=11.47\ .$$

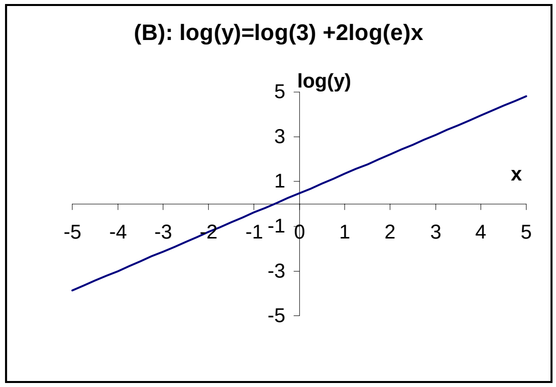
A weight of 165 lbs. would be 17 pounds above the average weight for males of that height. This is smaller than 2 standard deviations above the mean, so the man would not be considered obese.

c) You cannot find the constant term because this requires knowing the mean values of x and y. The slope is determined from the formula $slope=s_y/(rs_x)$. Using the slope of the given regression line the value of $r=.644$. The regression line of height on weight will have a slope of 10.1.

d) The x values for children would probably be out of the range of the data on which the regression was based.

21.





c) Since $10 = e^{\ln(10)}$ we have $y = ce^{\ln(10)bx} = ce^{kx}$, where $k = b\ln(10)$.

Chapter 10

1. (number of flavors)×(number of toppings) = $3 \times 4 = 12$

3. a) Since a codon consists of three bases, selecting the particular bases amounts to a three stage experiment in which there are four choices available at each stage. Thus the number of amino acids is $4 \times 4 \times 4 = 64$.

   b) With two bases, we can code for $4 \times 4 = 16$ amino acids. Since we need 20 amino acids to synthesize any protein, two base codons would not be adequate.

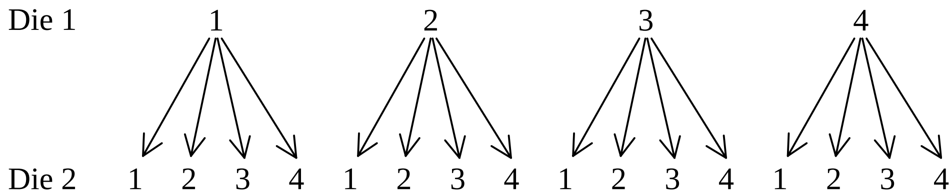
5. The theory part of Table 10.2 is obtained as follows: for each possible sum s we calculate all possible ways to obtain s. This number is then divided by the total number of outcomes (36) for tossing two dice. For example, a sum of 4 can be obtained from the outcomes: (1,3), (2,2) and (3,1). Therefore the probability of a sum of 4 is 3/36 or 1/12.

7. a) Let P(5+) be the probability of patients surviving for more than 5 years. From the Relative Frequency Method for assigning probabilities, P(5+) = frequency with which a patient survives more than 5 years. This is given as 62% = 0.62. Similarly, P(10+) = .51

   b) The people who survive between 5 and 10 years consist of those who survive more than 5 years, but not more than 10. Thus, from the 62% who survive past 5 years, we must remove the 51% who live more than 10 years. The resulting 11% are precisely the class we are interested in. Thus the probability is 0.11

9. Let E be the event that a person is killed in auto accident in U.S. during a given year. We are given that $P(E)$ = 0.00016. Since the U.S. population is 270 million, the Frequency Rule implies that the number of fatalities will be approximately (.00016)(270,000,000) = 43200.

11. a)



Die #2 → Die #1 ↓	1	2	3	4
1	(1,1)	(1,2)	(1,3)	(1,4)
2	(2,1)	(2,2)	(2,3)	(2,4)
3	(3,1)	(3,2)	(3,3)	(3,4)
4	(4,1)	(4,2)	(4,3)	(4,4)

b) Let E be the sum of two dice above. $P(E) = \dfrac{\#\text{ of outcomes with sum } E}{\text{total }\#\text{ of outcomes}}$. For example, $P(3) = 2/16 = 0.125$. The complete list of probabilities is given by

c)

Sums	Theory
2	.0625
3	.125
4	.1875
5	.25
6	.1875
7	.125
8	.0625

13. a) These two outcomes usually do not have the same chance of occurring. (Usually more students pass an exam than fail it.) The probabilities of passing or failing could be assessed by tracking the pass rate for a large number of students.

b) These two outcomes are unlikely to have equal probability. The frequency with which a "positive" occurs depends on the frequency of the disease, as well as the reliability (sensitivity) of the test.

c) The various blood groups do not occur with equal frequency in the population. If 40% of the population has type A blood, then the probability of that type would be .40.

d) From the areas of the figures, one would expect that the probability of the spin landing in the sections marked 1 or 2 would be .25, while the probability of landing in section 3 would be .5.

e) Unless the two colors occur with equal frequency, the probabilities of selecting each color would not be the same.

15. a) Find the probability of the complementary event E^c: The other 9 birthdays occur on some day other than yours. We have $P(E^c)=\frac{\overbrace{364\times364\times\cdots\times364}^{\text{9 times}}}{365^9}=\left(\frac{364}{365}\right)^9=.976$. Therefore, $P(E)=.024$ or 2.4%.

b) The probability of the complementary event is $P(E^c)=\frac{365\times364\times363\times\cdots\times341}{365^{25}}\approx.431$. Therefore the probability for the stated event is 1-.431 = .569, or more than 50%.

c) As in a) the complementary event has probability $P(E^c)=\frac{\overbrace{364\times364\times\cdots\times364}^{\text{24 times}}}{365^{24}}=.936$. The probability of E is .064 or 6.4%. Compare with the answer to b).

17. There are 32 possible outcomes for the sequence of heads and tails when tossing a coin 5 times in succession.

a) Consider the complementary event: all five tosses are heads, HHHHH, or all are tails, TTTTT. Thus the probability of the complement is 2/32 = 1/16. The probability of the given event is therefore $1-{}^1\!/_{16}={}^{15}\!/_{16}$.

b) We need to enumerate how many of the sequences of five tosses contain a streak of three or more heads. This can be done by going through a tree diagram, but the enumeration can be carried out more directly by classifying each streak according to the toss on which it begins.

Streak begins on first toss: **HHH**HT, **HHH**HH, **HHH**TH, and **HHH**TT.

Streak begins on second toss: T**HHH**H, T**HHH**T.

Streak begins on third toss: HT**HHH**, TT**HHH.**

Thus there are 8 outcomes favorable to the event, so the probability is $\frac{8}{32}=\frac{1}{4}$**.**

19. a) Any sequence of three tosses is as likely to occur as any other. By the Multiplication Rule the number of possible outcomes is $6\times6\times6=216$.

b) The sum of three dice can assume any value between 3 and 18. To illustrate the probability computation, consider obtaining a sum of 6 with three dice. One die, say die A, must land 1, 2, 3, or 4 (why not 5 or 6?). Correspondingly, the remaining two dice must add up to 5, 4, 3, or 2. Table 10.1 gives the number of ways this can happen. There are 4 ways to obtain a sum of 5, 3 ways to obtain a sum of 4, two ways to get a sum of three, and one way to obtain a sum of two. Altogether, there are 10 ways that a sum of 6 can be obtained, so the probability is $\frac{10}{216}\approx.0463$. The complete list of probabilities is given in the table below.

Sum	Prob.	Sum	Prob.	Sum	Prob.	Sum	Prob.
3	.0046	7	.0694	11	.125	15	.0463
4	.0139	8	.0972	12	.1157	16	.0278
5	.0278	9	.1157	13	.0972	17	.0139
6	.0463	10	.125	14	.0694	18	.0046

Chapter 11

1. a) $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$. We have $P(B) = 1 - P(B^c) = 1 - 0.75 = .25$. Since A and B are independent, $P(A \text{ and } B) = P(A)P(B) = 0.3 \times 0.25 = 0.075$. Thus $P(A \text{ or } B) = .475$.

   b) Whatever the relationship between the events A and B, we can certainly say that $P(A \text{ and } B) \geq 0$. Therefore

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \leq P(A) + P(B) = 0.55 .$$

3. a) Using independence gives $P(A \text{ and } Rh^+) = 0.40 \times 0.85 = 0.34$

   b) $P(O \text{ and } Rh^-) = 0.45 \times 0.15 = 0.0675$

   c) Let E designate the event $((O \text{ and } Rh^+) \text{ or } (A \text{ and } Rh^+))$. If a donor satisfies event E then he or she is suitable for the recipient. Since A, B, O blood type and Rh factor are independent we find

$$P(E) = P(O)P(Rh^+) + P(A)P(Rh^-) \approx .72$$

   We want the probability that at least one of five donors satisfies event E. Let's call this event S. The probability of the complementary event S^c, that none of the donors is satisfactory, can be computed using the product rule, assuming the donors' types are independent. We have

$$P(S^c) = \overbrace{P(E^c) \times P(E^c) \times \cdots \times P(E^c)}^{\text{5 times}} = (1 - .72)^5 \approx 0.0017$$

   Therefore $P(S) \approx .998$, so it is almost certain that one of the five prospective donors will be satisfactory.

5. a) $P(M \text{ or } W \text{ or } F) = \frac{1}{7} + \frac{1}{7} + \frac{1}{7} = \frac{3}{7}$

   b) Letting M_1 and M_2 denote that the first person's (respectively 2nd person's) birthday is on a Monday and assuming these events are independent, we have

$$P(M_1 \text{ and } M_2) = P(M_1)P(M_2) = \frac{1}{49} \approx .02$$

c) $P(M_1 \text{ or } M_2) = P(M_1) + P(M_2) - P(M_1 \text{ and } M_2) = \frac{13}{49} \approx 0.27$

d) The event in question is $E = (M_1 \text{ and } M_2) \text{ or } (Tu_1 \text{ and } Tu_2) \text{ or } \ldots \text{ or } (Sun_1 \text{ and } Sun_2)$. The probability of each joint occurrence is the same as in b) and the outcomes are mutually exclusive. Therefore, the probability of E is $\frac{7}{49} = \frac{1}{7}$.

7. a) $P(C \text{ or } D) = .65$

b) We want $P(A_1 \text{ and } A_2)$, A_1 designating less than 20 inches of rain in year one and similarly for A_2. Assuming the rainfall amounts in successive years are independent (perhaps not correct) we have $P(A_1 \text{ and } A_2) = P(A_1)P(A_2) = 0.0225$.

c) The complementary event is that the rainfall is in excess of 20 inches in each of the three years. The probability of this event in any one year is $1 - .15 = .85$. For this to happen in three consecutive years we can multiply the probabilities obtaining $.85^3 \approx .61$. So the event we are interested in has probability $1 - .61 = .39$.

9. a) We imagine that the balls are distinguished from each other with numbers, so we have balls R_1, R_2, R_3 and B_1, B_2, $\ldots$ B_5. We are performing a three-stage experiment. The total number of outcomes is $8 \times 7 \times 6 = 336$. The number of outcomes in which a black ball is selected each time is $5 \times 4 \times 3 = 60$. Therefore the probability of this event is $\frac{60}{336} \approx .179$.

The problem may also be solved using conditional probabilities. We want

$$P(B_1 \text{ and } B_2 \text{ and } B_3) = P(B_3 \mid B_1 \text{ and } B_2)P(B_1 \text{ and } B_2).$$

The term $P(B_3 \mid B_1 \text{ and } B_2)$ asks for the probability that B_3 occurs, given that B_1 and B_2 have occurred. If the latter event has occurred then there are only 6 balls remaining, of which 3 are black. Therefore, $P(B_3 \mid B_1 \text{ and } B_2) = \frac{3}{6} = \frac{1}{2}$. Similarly we can reason that

$$P(B_1 \text{ and } B_2) = P(B_2 \mid B_1)P(B_1) = \tfrac{4}{7} \times \tfrac{5}{8}$$

and therefore

$$P(B_1 \text{ and } B_2 \text{ and } B_3) = P(B_3 \mid B_1 \text{ and } B_2)P(B_1 \text{ and } B_2) = \tfrac{3}{6} \times \tfrac{4}{7} \times \tfrac{5}{8} = \tfrac{60}{336} \approx .179.$$

Note that the probability $P(B_2 \mid B_1)$ is not the same as $P(B_2)$. The latter measures the chance that a black ball will be picked on the second pick, without regard to the outcome of the first selection. The reader might wish to show that this value is $\frac{5}{8}$, compared to the value of $\frac{4}{7}$ for $P(B_2 \mid B_1)$.

b) Let us denote the event in question by A. The complementary event A^c is that all the balls are black or all the balls are red. In a) we found the probability that all balls selected are black. Using similar reasoning, the probability that all are red is $\frac{6}{336}$. Therefore, the probability of A^c, that all balls are red or all are black, is $\frac{6}{336}+\frac{60}{336}=\frac{66}{336}\approx 0.196$. Finally, $P(A)=1-P(A^c)=\frac{270}{336}\approx .803$

11\. a) Assuming the drugs work independently, the probability that all will fail (the complement of the desired event) is $.6\times.6\times.7\approx.25$. Therefore, the probability that at least one will succeed is .75.

b) If the drugs did not work independently, the answer might be larger or smaller than the answer in part a). For example, if the drugs affected the body through the same pathway, then if one failed it would be highly likely that the others would fail as well. In this case the probability that at least one worked would be considerably smaller than .75. However, one drug might enhance the action of another, and so the combination might have a greater chance of succeeding than the independence hypothesis predicts. In effect, if events A and B are not independent then $P(A\,|\,B)$ may be larger or smaller than $P(A)$.

13\. a) Let A_X and A_Y denote that events of receiving an "A" grade in the respective courses. We want to find $P(A_X \text{ and } A_Y)$. Since we do not know whether the events are independent, we cannot use the product rule for independent events. Instead we use that

$$P(A \text{ or } B)=P(A)+P(B)-P(A \text{ and } B).$$

We have $P(A_X)=.15$, $P(A_Y)=.10$ and $P(A_X \text{ or } A_Y)=.235$. Using the equation above we obtain that $P(A_X \text{ and } A_Y)=.015$.

b) $P(A_Y\,|\,A_X)=\dfrac{P(A_Y \text{ and } A_X)}{P(A_X)}=\dfrac{.015}{.15}=.1$

c) The events are independent since $P(A_Y\,|\,A_X)=P(A_Y)$ or equivalently $P(A_Y \text{ and } A_X)=P(A_Y)P(A_X)$.

15\. a) $1-.003=.997$

b) $(.997)^{40}\approx .887$

c) The event is the complement of the event whose probability is given in b). Therefore, the probability is $1-.887=.113$ or 11.3%.

17\. a) The death rate per household is obtained by dividing the number of reported deaths in each category by the number of households. For example, for Southwark & Vauxhall (S&V) we find

$$\text{S\&V:}\quad \textit{death rate}=\frac{1263}{40{,}046}\approx 3.2\times10^{-2}.$$

Similarly:

$$\text{L:}\quad \textit{death rate}=\frac{98}{26{,}107}\approx 3.8\times10^{-3}$$

$$\text{Rest of London:} \quad \textit{death rate} = \tfrac{1422}{256,423} \approx 5.5\times10^{-3}$$

One might initially assert that these numbers represent the probability of a cholera death for a single household. However, since multiple deaths might occur in each household (due to high occupancy and spread from one infected to another) these numbers do not represent a proper frequency count to be interpreted as probabilities. (For instance, if the disease killed virtually everyone and there were multiple occupants in each dwelling the ratios above would have a value in excess of one. This is allowed for a death rate, but not allowed as a probability.)

b) With the given data we cannot compute the risk of death for an individual. For a precise estimate of this probability we would need to know the number of people who resided in the households under study. This information was not known to Snow, who relied on coroner's death reports (showing a cause of death and the deceased's address) and water company records showing the households connected to their systems.

c) We can still define a relative risk ratio as the ratio of the household death rates for each water company, compared with the household death rate for the remainder of London. The values are

$$\text{Rel. Risk S\&V:} = \frac{3.2\times10^{-2}}{5.5\times10^{-3}} \approx 5.8$$
$$\text{Rel. Risk L:} = \frac{3.8\times10^{-3}}{5.5\times10^{-3}} \approx .69$$

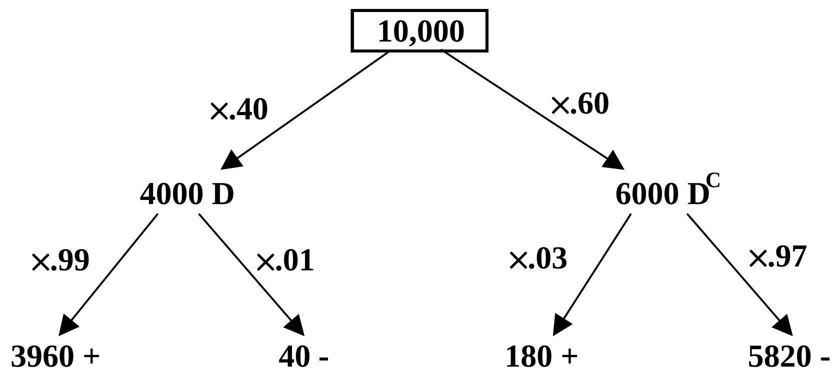
The water supply of Southwark & Vauxhall would appear to pose a serious risk of death. The Lambeth company's supply appears as safe as the remainder of London's.

d) The water supply might be suspected, particularly as the two companies drew water from different parts of the Thames. However, the high relative risk for S&V computed in c) might be due to other causes: high occupancy rate, poor sanitation (other than water), age differences in the populations served by the two companies, different occupational exposures to some other source of infection. Fortunately for Snow, the two water systems served customers in the same neighborhoods, often on the same street, so that these confounding factors could be ignored, i.e. they were likely to be the same for each group.

19. a) Using a tree diagram as in Example 11.14 of chapter 11, but with a 10% incidence of disease, we find that starting with 10,000 people we will have $990+270=1260$ individuals who will give a positive reaction to the test, i.e. $P(+)=.126$. Out of the 1260 persons giving a positive reaction, 990 are individuals who also have the disease. Thus $P(D\,|\,+) = \dfrac{P(D\text{ and }+)}{P(+)} = \tfrac{990}{1260} = .786$.

b) We must find the incidence of disease among people who give a positive reaction. This is $P(D\,|\,+) = \dfrac{P(D\text{ and }+)}{P(+)}$. From Figure 11.1 in chapter 11 we find that the number of positive reactions to the test is 198 + 294 =492, out of 10,000. Thus $P(+) = .0492$. Of the 492 who gave a positive reaction, 198 also had the disease so that $P(D \text{ and } +) = .0198$. Therefore, $P(D \mid +) = \frac{.0198}{.0492} \approx .40$.

c) Suppose we start with 10,000 individuals who have shown a positive reaction to the test. We can use a tree diagram as in Figure 11.1, chapter 11, to determine how many positive and negative reactions we should expect when we give a second test to these individuals. Note that now, according to b) the incidence of the disease in our population is .40.



From this we obtain $P(D \mid +) = \frac{3960}{4140} \approx .956$ and $P(D^c \mid -) = \frac{5820}{5860} \approx .993$. Thus, if the first test is positive the outcome of the second test has a very accurate predictive value.

Chapter 12

1. The following table lists the six genotypes and the corresponding blood groups.

genotype	phenotype	genotype	phenotype
$I_A I_A$	A	$I_B i$	B
$I_A I_B$	AB	$I_B I_B$	B
$I_A i$	A	ii	O

3. a) There are 4 homozygous genotypes. There are $4 \times 3 = 12$ pairs of the form $A_i A_j$, where $i \neq j$. By the Principle of Equivalence these give rise to six more genotypes, since a pair such as $A_1 A_2$ defines the same genotype as $A_2 A_1$. Thus, we have in total 10 genotypes.

   b) With n alleles there will be n homozygous genotypes. There are $n(n-1)$ pairs of the form $A_i A_j$, with $i \neq j$. Half of these need to be counted towards the totality of genotypes. This gives the stated result.

5. Denote by A, respectively B, the phenotype associated with the dominant gene and by a and b the recessive trait. Four phenotypes are possible: AB, Ab, aB, and ab. Note that the same answer holds if the genes were on different chromosomes.

7. The probability that a single offspring is heterozygous (Aa) is ½. If each of 5 consecutive offspring fails to be heterozygous then all are homozygous. Assuming the gene selection for each child is independent of the selection of any other, the chance of 5 consecutive homozygotes is $\left(\frac{1}{2}\right)^5 = \frac{1}{32}$. Thus the probability of at least one heterozygous offspring is $\frac{31}{32}$.

9. Four genotypes are possible for the offspring: $Aa\,//\,BB$, $Aa\,//\,Bb$, $aa\,//\,BB$, $aa\,//\,Bb$. Each occurs with probability ¼.

11. a) If the parental genes were heterozygous at either locus, then under self-fertilization some progeny would be formed that were homozygous of type AA and aa. Hence both phenotypes should appear, which contradicts the assumptions stated in the problem.

    b) The F_1 generation all have genotype $Aa\,//\,Bb$ if the loci are on different chromosomes or AB/ab if they are on the same chromosome pair. In the first case the F_1 parent will contribute gametes $AB, Ab, aB,$ and ab. The pure-breeding recessive parent contributes only ab gametes. Hence the zygotes can have only four possible genotypes: $Aa\,//\,Bb, Aa\,//\,bb, aa\,//\,Bb,$ and $aa\,//\,bb$. If the loci are on the same chromosome pair, then because of crossover the F_1 parent can still produce the same types of gametes AB, ab (parental types) and Ab, aB (recombinant types). Again, we get four genotypes when such a parent is crossed with a pure-breeding recessive. These would be denoted $AB/ab, Ab/ab, aB/ab$ and ab/ab. In each case, exactly one genotype corresponds to each phenotype given in the problem data.

    c) If the gene loci were on different chromosome pairs then the principle of independent assortment would imply that each of the four gametes of the F_1 parent would occur with equal frequency (.25). From this fact all four genotypes ($Aa\,//\,Bb, Aa\,//\,bb, aa\,//\,Bb,$ and $aa\,//\,bb$) would occur with equal frequency. Since each such genotype corresponds to a specific phenotype, the latter should also occur with equal frequency. However, this is greatly contradicted by the data.

    d) Since c) implies that the loci are on the same homologous pair, some crossover must have occurred. The recombinant types come from the recombinant gametes Ab and aB. These give the combinations Ab/ab and aB/ab with phenotypes Ab and aB respectively. Thus the recombination frequency is $\dfrac{65+75}{400} = .35$

13. a) The six possible genotypes are $A_1A_1, A_2A_2, A_3A_3, A_1A_2, A_1A_3,$ and A_2A_3. These will have genotypes $p^2, q^2, r^2, 2pq, 2pr,$ and $2qr$ respectively.

b) An A_1 gamete in the next generation arises by selecting a parental genotype that includes the A_1 gene and then selecting the A_1 gamete. We can thus compute the probability from

$$\begin{aligned} P(A_1) &= P(A_1A_1 \text{ and } A_1) + P(A_1A_2 \text{ and } A_1) + P(A_1A_3 \text{ and } A_3) \\ &= P(A_1 \mid A_1A_1)P(A_1A_1) + P(A_1 \mid A_1A_2)P(A_1A_2) + P(A_1 \mid A_1A_3)P(A_1A_3) \\ &= p^2 + pq + pr \\ &= p(p+q+r) = p \end{aligned}$$

A similar computation holds for the other alleles. Thus the gene frequencies are the same as the original p, q and r.

c) The frequencies given in a) are stable.

15. First add the columns:

column 1 = $p^2 + 2pt + t^2 = (p+t)^2$

column 2 = $2ps + 2(pq+st) + 2qt = 2(ps + pq + st + qt)$
$= 2(p(s+q) + t(s+q)) = 2(s+q)(p+t)$

column 3 = $s^2 + 2qs + q^2 = (s+q)^2$

Now observe that

$$\begin{aligned} \text{column 1} + \text{column 2} + \text{column 3} &= (p+t)^2 + 2(s+q)(p+t) + (q+s)^2 \\ &= ((p+t)+(q+s))^2 = (p+q+s+t)^2 = 1, \end{aligned}$$

since $p+q+s+t=1$.

17. a)

Zygote	Freq.	Zygote	Freq.
AB/AB	p^2	AB/aB	$2ps$
ab/ab	q^2	AB/Ab	$2pt$
aB/aB	s^2	ab/aB	$2qs$
Ab/Ab	t^2	ab/Ab	$2qt$
AB/ab	$2pq$	aB/Ab	$2st$

b) Let the crossover frequency be denoted by r and, to avoid confusion with the frequency p of genotype AB, let $1-r$, rather than p, denote the frequency of parental types. The gamete genotype AB arises from the following parental genotypes: AB/AB, AB/ab, AB/aB, AB/Ab, and through recombination from aB/Ab. The frequency of the AB gamete from each of these parental types is 1, $(1-r)/2$, $1/2$, $1/2$ and $r/2$, taking into account the two situations where crossover introduces recombinant types. Combining this with the zygote frequencies given in a) we find that

$$\begin{aligned} p' = P(AB) &= p^2 + (1-r)pq + ps + pt + rst \\ &= p^2 + pq + ps + pt + r(st - pq) \\ &= p(p+q+s+t) + r(st - pq) \\ &= p + r(st - pq), \end{aligned}$$

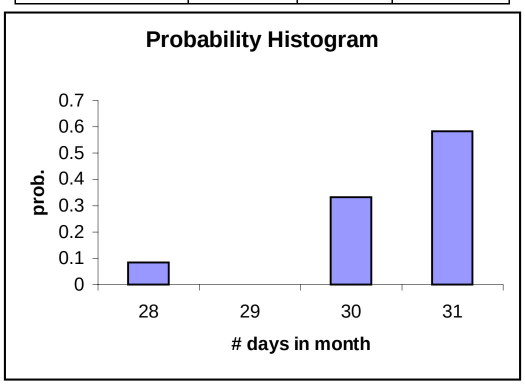
which are similar to the equations given in Table 12.2. A similar analysis holds for the other genotypes

Chapter 13

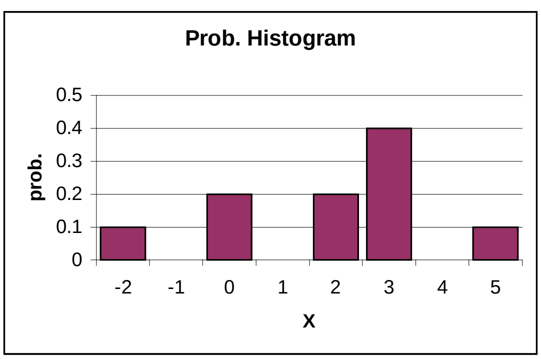
1. a) Not a random variable.
   b) A discrete random variable taking on the values 28, 30 and 31.
   c) A continuous random variable.
   d) Not a random variable.
   e) Before the final exam has been given, a discrete random variable. Sometimes treated as a continuous random variable.

3. a)

X	28	30	31
$P(X = x)$	$1/12$	$1/3$	$7/12$



5. a)



b) $P(X=0)+P(X=2)+P(X=3)=0.8$

c) $\mu=1.9$, var $=3.69$, $\sigma=1.92$

7. a) The expected value is

$$\mu=2(\frac{1}{36})+3(\frac{2}{36})+\cdots+7(\frac{6}{36})+8(\frac{5}{36})+\cdots+12(\frac{1}{36})=7.$$

The computation of the standard deviation is best organized using a table, as we did when computing the related quantity for data. In the table below x_i denotes a possible value for the sum of two dice and $\mu=7$.

x_i	p_i	$p_i(x_i-\mu)^2$	x_i	p_i	$p_i(x_i-\mu)^2$
2	1/36	$\frac{25}{36}$	8	5/36	$\frac{5}{36}$
3	2/36	$\frac{32}{36}$	9	4/36	$\frac{16}{36}$
4	3/36	$\frac{27}{36}$	10	3/36	$\frac{27}{36}$
5	4/36	$\frac{16}{36}$	11	2/36	$\frac{32}{36}$
6	5/36	$\frac{5}{36}$	12	1/36	$\frac{25}{36}$
7	6/36	0			

Adding the numbers in the two columns headed $p_i(x_i-\mu)^2$ yields the square of the standard deviation. Note that the values in these two columns are identical. Thus

$$\sigma^2=2\left(\frac{25}{35}+\frac{32}{36}+\frac{27}{36}+\frac{16}{36}+\frac{5}{36}\right)=\frac{210}{36}$$

and $\sigma=\sqrt{\frac{210}{36}}\approx 2.415$.

9. a) $P(2\le X\le 6)=P(X=2)+P(X=3)+\cdots+P(X=6)=.876$

b) $P(6\le X\le 9)=P(X\le 9)-P(X\le 5)=.811-.193=.618$

11. a) On each pick the probability of selecting a red ball is $p=7/10$. With 12 independent trials we have $P(X=4)={}_{12}C_4(.7)^4(.3)^8=495(.7)^4(.3)^8\approx .0078$

b) Using the table for $n=12$ trials we have

$$P(3\le X\le 6)=P(X=3)+P(X=4)+P(X=5)+P(X=6)=.118$$

13. a) i) $P(X=2)=\dfrac{4.5^2e^{-4.5}}{2!}\approx .112$

ii) $P(X>2)=1-P(X\le 2)$. Since $P(X\le 2)=P(X=0)+P(X=1)+P(X=2)\approx .173$, we have $P(X>2)=1-.173=.827$

iii) $P(2 \le X \le 5) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$
$= .112 + .169 + .190 + .171 = .642$

15. a) $p = .4$. The means and standard deviation are given in the next table.

n	10	20	40	80	160
μ_X	4	8	16	32	64
σ_X	1.55	2.19	3.10	4.38	6.20

b) For example, when $n = 20$ the relevant probability in row two is

$$P(\mu_X - \sigma_X \le X \le \mu_X + \sigma_X) = P(8 - 2.19 \le X \le 8 + 2.19)$$
$$= P(5.81 \le X \le 10.19)$$
$$= P(6 \le X \le 10) \approx .747$$

17. a) For example, a value $Y = 8/25$ means that 8 successes were obtained in 25 trials. If X denotes the number of successes, then the probability that $Y = 8/25 = .32$ is the same as $P(X = 8)$. The latter probability is found using the Excel function =*binomdist*(8, 25, .6, false).

19. a) In the worst case scenario the probability that a randomly selected computer in the shipment is defective is $p = 0.01$. The number of defectives X in a shipment of 20 has a binomial distribution with $p = 0.01$, and $n = 20$. The complement of the event that there is at least one defective is the event that all the computers are OK, i.e. that $X = 0$. We find that $P(X = 0) = .99^{20} = .818$ and therefore the probability of at least one defective in a shipment of 20 is $1 - .818 = .182$.

b) Use the random number generator to obtain random outcomes from a suitable binomial distribution. Sort the results and count how often a value greater than or equal to one was obtained.

Chapter 14

1. a) (A) Using the formula for the area of a rectangle, the area under the graph between $x = 0$ and $x = 2$ is one. (B) Using the formula for the area of a triangle, the area under the graph between $y = 0$ and $y = 2$ is one.

b) (A) $P(1 \le X \le 1.5) = .5 \times .5 = .25$ (B) The equations of the two lines making up the sloping boundary of the density curve are $z = y$ and $z = 2 - y$. Using these formulas we find that the region between $y = 1$ and $y = 1.5$ is a trapezoid of height 0.5 and with bases of length one and length $z = 2 - 1.5 = 0.5$. The formula for the area of a trapezoid gives $P(1 \le Y \le 1.5) = .5(\frac{1+.5}{2}) = .375$.

c) We need a value x_0 such that $.5x_0 = .9$. This gives $x_0 = 1.8$.

d) The area from $y = 0$ to $y = 1$ is 0.5. Therefore the value y_0 must be chosen so that the trapezoid lying over the interval $[1, y_0]$ has an area of 0.4. Dropping the subscript on the y_0 we need to solve the equation $.4 = \text{area} = \text{ht}\left(\frac{\text{base 1+base 2}}{2}\right) = (y-1)\left(\frac{1+(2-y)}{2}\right)$. We obtain a quadratic equation for y, namely $y^2 - 4y + 3.8 = 0$. Using the quadratic formula we find $y \approx 1.56$ as the only solution in the interval [1,2].

3. a) The average occupancy number is obtained by adding the occupancy numbers of each cell and dividing by the total number of cells. This can be readily achieved using the grouping principle discussed in section 7.6. We obtain

$$\bar{x} = \frac{0(40)+1(32)+2(19)+3(7)+4(1)+5(1)}{100} = 1,$$

which agrees with our intuition, since we are scattering 100 points among 100 cells.

We also need the square of the standard deviation or variance. This can also be computed using the grouping technique.

$$s^2 = \frac{40(0-1)^2 + 32(1-1)^2 + 19(2-1)^2 + 7(3-1)^2 + 1(4-1)^2 + 1(5-1)^2}{99} = 1.13$$

The index of dispersion is then $\frac{1.13}{1} = 1.13$

b) $\bar{x} = \frac{200}{60} = 3.33$. The variance $s^2 = 3.14$. The index of dispersion is $\frac{3.14}{3.33} = .94$.

c) The indices in both cases are fairly close to one, indicative that the occupancy patterns follow a Poisson distribution.

5. a) $P(Z < 1.6) = .9452$

b) $P(Z > -.52) = 1 - P(Z < -.52) = .6985$

c) $P(1.2 \le Z \le 2.24) = P(Z \le 2.24) - P(Z \le 1.2) = .9875 - .8849 = .1026$

d) $P(|Z| < 2.5) = P(-2.5 < Z < 2.5) = P(Z < 2.5) - P(Z < -2.5) = .9876$

7. a) The mean of the binomial distribution is $\mu_X = np = 50(.3) = 15$ and the standard deviation is $\sigma_X = \sqrt{npq} = \sqrt{50(.3)(.7)} = 3.24$. We approximate probabilities for X using a random variable $Y = N(15, 3.24)$. With the continuity correction we have

$$P(X \le 12) \approx P(Y \le 12.5) = P(Z \le \frac{12.5-15}{3.24}) = P(Z \le -.77) = .2266$$

9. Let X equal the number of defectives in a shipment of 1000. Based on the manufacturer's claim X has a binomial distribution with $p = .03$ and $n = 1000$. A shipment of 1000 that contains more than 45 defectives will be returned. We therefore want $P(X > 45)$. We can use the normal approximation to estimate this. The mean of X is 30 and the standard deviation is 5.39. Therefore using $Y = N(30, 5.39)$ we obtain

$$P(X > 45) \approx P(Y > 44.5) = P(Z > \frac{44.5-30}{5.39}) = P(Z > 2.69) = 1 - P(Z < 2.69) = .0036$$

11. The probability that a random person appears for the flight is $p = .90$. If the airline sells 235 tickets then the number of people who actually show up has a binomial distribution with $p = .9$ and $n = 235$. The expected value of X is 211.5 with a standard deviation $\sigma = \sqrt{235(.9)(.1)} \approx 4.6$. Using a normal approximation $Y = N(211.5, 4.6)$ we have

$$P(X > 220) \approx P(Y > 219.5) = P\left(Z > \frac{219.5-211.5}{4.60}\right) = P(Z > 1.74) = .0409$$

Thus approximately 5% of flights of the type described will be overbooked.

13. a) Let X denote the income of a random worker. Expressing all units in thousands of dollars we have $X = N(30, 4.25)$. Thus $P(25 < X < 40) = P(-1.13 < Z < 2.35) = .8614$ (Note: No "continuity correction" is used since the random variable X is assumed to have a normal distribution.)

b) $P(X > 25) = P(Z > -1.18) = .881$.

c) We must find a value of I so that $P(X > I) = .01$. Standardizing the probabilities yields the equivalent statement that $P(Z > \frac{I-30}{4.25}) = .01$. From the table of the standard normal distribution we find that for $z_0 = 2.33$ we have $P(Z < z_0) \approx .99$ and therefore $P(Z > z_0) \approx .01$. Therefore we must have

$$\frac{I-30}{4.25} = 2.33,$$

yielding $I = \$39.9$ thousand.

15. a) Let X denote the weight of a randomly selected candy bar. We have $X = N(6.25, 0.1)$. The probability that a random bar will have a weight below 6 oz. is given by $P(X < 6) = P\left(Z < \frac{6 - 6.25}{0.1}\right) = P(Z < -2.5) = .0062$. From this calculation, approximately 6 out of 1000 bars will have weight shortfalls.

    b) From the answer to a) it is clear that the manufacturer must increase the average weight of the candy bars. We need to find a value for μ so that $P(X < 6) = P\left(Z < \frac{6 - \mu}{0.1}\right) \approx .001$. The value of z_0 such that $P(Z < z_0) = .001$ is $z_0 = -3.1$. Thus for μ we find the equation

$$\frac{6 - \mu}{0.1} = -3.1,$$

which gives $\mu = 6.31$ oz.

17. The pressure reading is normally distributed with mean 105 and standard deviation 7, i.e. $X = N(105, 7)$. Therefore $P(X < 95) = P\left(Z < \frac{95 - 105}{7}\right) = P(Z < -1.43) = .0764$. There is approximately an 8% chance that the pressure reading will be significantly below its mean value.

Chapter 15

1. a) The bowl contains 500 balls labeled 0 for a red flowering plant and 1 for a white flowering. The investigator needs to compute the fraction of the balls that have each number.

   b) A bowl with 3000 balls each of which has a number representing the beer consumption of a particular student. The average beer consumption would be estimated using a random sample of these balls.

   c) A bowl containing a ball for each student. A “1” on a ball indicates a right-handed student, a “0” a left-handed student. Need to estimate the number of balls with a “0”.

   d) The bowl contains balls, one for each person. The differences between the right and left pressure readings are written on each ball. Need to estimate the average value of this difference.

   e) Here we need two bowls, one with balls representing the treated patients and the other the controls. Each ball would contain the number of polyps found on that person during the study period. Must estimate the average values in the two groups and compare.

3. a) In both box plots the series showing the smallest spread came from the sample of size 100. this is Series 1 in box plot A and Series 2 in box plot B

   b) The random variable Y has a skewed distribution (to the right). The sample average involving the smaller number of samples will reflect this skewness; averages based on larger sample sizes become increasingly symmetric (Central Limit Theorem). Therefore, boxplot (A) gives the result of sampling Y.

5. a) The fraction of heads obtained is $\hat{p}$. In both cases the values of $\hat{p}$ will be centered around the expected value of 0.5. When making 100 tosses the values of $\hat{p}$ will more likely be close to 0.5 than will happen when we make 25 tosses. Therefore, a value of $\hat{p}=0.6$ will be more likely to occur when 25 tosses are made.

   b) The hospital that makes 25 deliveries, for the reasons given in a).

7. a) The standard deviation measures the variation among individual values of a random variable. The standard error measures the variation among values of a sample mean.

   b) Sampling error refers to any variation from an average value arising from random variation. In expressing a confidence interval for a mean, the width of the interval represents the likely sampling error of the estimate.

   c) A confidence interval for a certain parameter is an interval constructed from a sample that, in repeated application, has a specified probability of containing the true parameter value.

9. Assume the current class represents a random sample representative of prior students. The class average $\overline{X}$ should be 75 and the standard deviation $\sigma_{\overline{X}} = 10/\sqrt{40} \approx 1.58$. Since $\overline{X} \approx N(75, 1.58)$ the probability $P(\overline{X} < 72) \approx P(Z < \frac{72-75}{1.58}) = P(Z < -1.90) \approx .029$ or about 3%.

11. For example the confidence coefficient for 99% confidence level is determined as the value of z_0 such that $P(Z < z_0) = .995$. Using Table section B.3 we find that $z_0 = 2.57$ satisfies this condition.

13. We must check section B.4 and find the confidence intervals that contain the true p, which is 0.4 in this case. When $n = 20$ and confidence level is 80% we find that $k = 5$ to $k = 10$ successes produce confidence intervals that contain the true value of p. The probability this happens can be found from the binomial distribution for X = # of successes in 20 trials, with probability of success equal to 0.4. Using section B.2 we find

$$P(5 \le X \le 10) = P(X \le 10) - P(X \le 4) = .872 - .051 = .821$$

As expected the result is close to 0.8

15. b) To find the sample size solve the equation $2.33\sqrt{\frac{(.4)(.6)}{n}} = (.03)^2$ for n. The solution is approximately $n = 1450$

17. a) False. By construction, both intervals have a 90% chance of containing the true value of p.

c) True. The width of the confidence interval is determined by the expression $z^*\sqrt{pq/n}$, which decreases as n increases.

d) False. The coefficient z^* is given in Table 15.5. The higher the confidence level the larger this number and therefore the wider the confidence interval.

19. a) The description of the poll indicates ("… 19 out of 20 cases …") that the investigators are constructing a 95% confidence interval with a spread of $\pm.03$. Using the hint (which can be verified using calculus or by graphing the expression $x(1-x)$ on the interval [0,1]), the sample size n should satisfy the equation $\frac{(1.96)(0.5)}{\sqrt{n}} = 0.03$. The solution for n is about 1067. Typically, the polls will include a larger number so that some additional information regarding subgroups can be extracted as well.

b) Using a sample size of $n = 1100$ we obtain for the 98% confidence interval the bounds

LCL = $0.4 - 2.33\sqrt{\frac{(.4)(.6)}{1100}} = .365$ to UCL = $.4 + 2.33\sqrt{\frac{(.4)(.6)}{1100}} = .434$.

21. The confidence interval would have boundaries $70 - 2.33\frac{\sigma}{\sqrt{n}} = 70 - 2.33\frac{2}{\sqrt{125}} = 69.6$ to $70 + 2.33\frac{\sigma}{\sqrt{n}} = 70.4$. Note that this is an estimate for the <u>mean</u> height. If the heights were normally distributed it would be very likely that an <u>individual</u> height would fall in the interval 66 to 74 inches, i.e. 2σ around the approximate mean.

23. a) (i): 1.81 (ii): 2.95 (iii): 2.02

b) (i) Enter =tinv(0.2,5). Answer: 1.48 (ii) Enter =tinv(.25,10). Answer: 1.22

(iii) Enter =tinv(0.2, 25). Answer 1.32

25. a) The average of the four measurements is 85 mm Hg with a standard deviation of approximately 7. The number of degrees of freedom is $n-1=3$. The 95% confidence interval based on Student's distribution (see section B.5) is $85 \pm 3.18\frac{7}{\sqrt{4}} = 85 \pm 11$, i.e. the interval [74, 96]. Since this interval includes pressure readings that are classified as mildly hypertensive, we should classify the patient in the latter category.

We are assuming that the readings are normally distributed.

Chapter 16

1. a) The null hypothesis H_0 is that $p = 0.8$ versus the alternative $H_a \neq 0.8$. If H_0 were true we would expect 24 = 30(0.8) trains to be on time. If the number of on time trains was ≤ 20 or ≥ 28 this might be considered adequate evidence against H_0. Using this criterion and Table B.2 we find that $\alpha = 0.061 + 0.044 = .105$.

   b) Using Table B.2 again we find that for the alternative $p = 0.7$ the power γ is 0.413.

3. a) $H_0 : \mu = 1.25$ hrs. versus $H_a : \mu \neq 1.25$. We use the sample average $\bar{x}_{50}$ or equivalently its *z* score.

   b) With a two-sided significance test and $\alpha = 0.05$ the rejection criterion is $z < -1.96$ or $z > 1.96$. We find here that $z = 2.47$ so we would reject the null hypothesis and conclude that the medication prolongs sleep even more than was claimed. Since the sample is large we do not have to make any assumptions on the distribution of the # of hours of sleep gained.

5. a) If *X* is the mileage for a tire then $H_0 : \mu_X = 40{,}000$ (or 40 if we use units of 1000), versus the alternative $\mu \neq 40$. Assuming *X* has a normal distribution we use the *t* test with $f = 9$. From Table B.5 the *t* rejection criterion is $t < -2.26$ or $t > 2.26$. We find that $t = -2.08$. The data does not refute the null hypothesis at the 0.05 significance level.

   b) $P = .067$. For any significance level above 0.067 the data will be accepted as evidence refuting the null hypothesis. At lower levels the data is not strong enough.

7. a) $H_0 : p_A = p_B$, where these are the probabilities of death for the Zovant patients (A) and the placebo group (B). We use the *z* statistic to evaluate the claim. The *z* score is -2.8 which is significant at the 0.05 significance level.

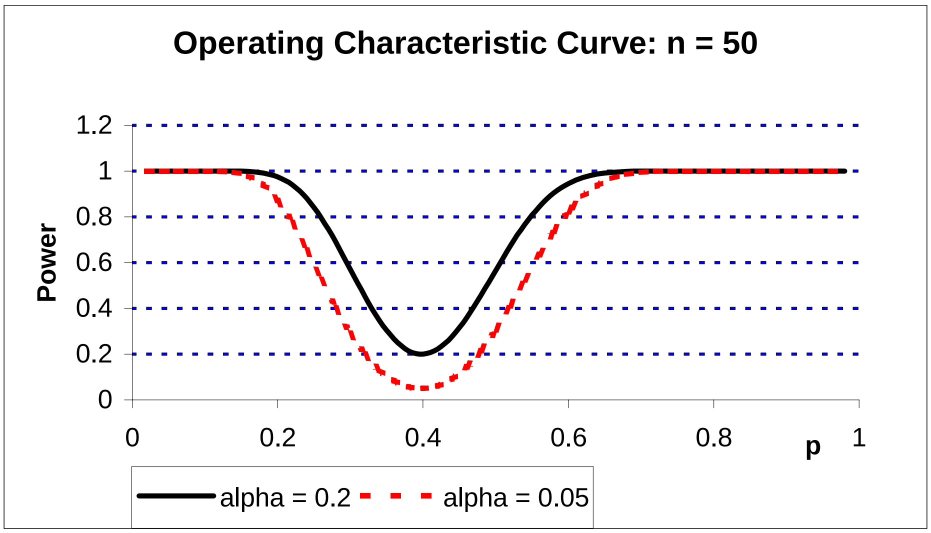
   b) $P = .0052$, so the data is highly significant and very unlikely to have occurred by chance if the null hypothesis were true.

   c) Both statements are correct, but they refer to different fractions. If we let $D = \frac{a}{b}$ denote the death rate, where *a* is the number of deaths and *b* is the number of cases, then the absolute decrease of 6% in *D* refers (essentially) to $\frac{\Delta a}{b}$, while the statement regarding the % reduction in *D* refers to $\frac{\Delta a}{a}$. There is merit to both numbers, particularly when *a* is large. Unfortunately, people who wish to confuse will opt to report the larger or smaller number, as fits the case they wish to present.

9. We test $H_0 : p_A = p_B$, where p_A and p_B refer to the frequency of vertebral fractures in the two groups. Since the samples are large we can use the *z* score to test the hypothesis. We find $z = -5.09$. The *P* value of this score is certainly smaller than .001 ($z = \pm 3.3$). The exact *P*

value can be found using software: the result is approximately 5×10^{-7}, so the result is quite strong evidence that the medication reduces these fractures.

11. a) If $\alpha > \alpha'$ then the rejection region for the α' test is farther into the tails of the distribution under H_0 than the rejection region for the α test. The same will true for the distribution under any alternative hypothesis so the powers must have the same relationship, $\gamma > \gamma'$. The relationship is illustrated in the graph below which shows the operating characteristic curves for two different values of α



b) As we increase the sample size we reduce variability. The distribution for the null hypothesis becomes more tightly centered around its central zone. Therefore, the rejection criterion for a given significance level will become closer and closer to the hypothesized value. If some alternative is true, no matter how close to the hypothesized value, for a large enough *n* we will obtain an outcome that is far enough from the hypothesized value to trigger rejection. As a result, even when the difference between the true and hypothesized value is quite close, and where the difference is not likely to matter much in practice, the null hypothesis may be rejected. Reporting a confidence interval will at least make clear that the observed result is quite close to the hypothesized value.

13. a) If the true *p* is p_0 then we know that for 95% of experiments the 95% confidence interval will contain the true *p*. Therefore, there is only a 5% chance that our rejection criterion (p_0 outside the confidence interval) will be met, when the null hypothesis is true. However, this is exactly the definition of significance level.

b) Recall that the 95% confidence interval is constructed using the endpoints $\hat{p} \pm 1.96\sigma_{\hat{p}} = \hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$. If the hypothesized value p_0 is outside the interval then we have either $p_0 > \hat{p} + 1.96\sigma_{\hat{p}}$ or $p_0 < \hat{p} - 1.96\sigma_{\hat{p}}$. These are equivalent to

$z = \dfrac{\hat{p} - p_0}{\sigma_{\hat{p}}} < -1.96$ or $z = \dfrac{\hat{p} - p_0}{\sigma_{\hat{p}}} > 1.96$, which are precisely the requirements for rejection using the conventional test.

Chapter 17

1. $N_1 = 3$, $N_2 = 7$, $N_3 = 15$, $N_4 = 31$, $N_5 = 63$

3. $N_1 = -0.790$, $N_2 = -1.376$, $N_3 = -0.107$, $N_4 = -1.989$, $N_5 = 1.954$

5. $N_1 = 0.6$, $N_2 = 0.794$, $N_3 = 1.160$, $N_4 = 1.721$, $N_5 = 2.243$

7. (I) corresponds to (c); (II) corresponds to (a); (III) corresponds to (b)

9. a) $M_{t+1} = N_{t+1} - \frac{h}{r} = (1+r)N_t - h - \frac{h}{r} = (1+r)N_t - \dfrac{h}{r}(1+r) = (1+r)M_t$. Notice that the quantity M_t gives the "displacement" of the quantity N_t from its steady state value.

   b) Theorem 17.4 implies that $M_t = (1+r)^t M_0 = (1+r)^t (N_0 - \frac{h}{r})$. Using that $M_t = N_t - \frac{h}{r}$ we obtain that $N_t = (1+r)^t (N_0 - \frac{h}{r}) + \dfrac{h}{r}$.

   c) Introduce the new function $y(t) = N(t) - \frac{h}{r}$. Observe that $\dfrac{dy}{dt} = \dfrac{dN}{dt} = rN - h = r(y + \frac{h}{r}) - h = ry$. See also exercise 23b, Chapter 2.

11. Let $f(N) = 2N + 1$: Steady state: $N^* = -1$. Stability: $f'(N^*) = 2$, therefore unstable.

13. Let $f(N) = N^2 - 2$: Steady state: $N^* = 2, -1$.: Stability $f'(N) = 2N$, therefore both steady states are unstable.

15. Let $f(N) = \dfrac{3N^2}{1+N^2}$: Steady states: $N_1^* = 0$, $N_2^* = \dfrac{3+\sqrt{5}}{2} \approx 2.618$, $N_3^* = \dfrac{3-\sqrt{5}}{2} \approx 0.382$. Stability: $f'(N) = \dfrac{6N}{(N^2+1)^2}$, therefore $N_1^* = 0$ is stable, $f'(N_2^*) \approx 0.255$, so N_2^* is stable, and $f'(N_3^*) \approx 1.75$ and so N_3^* is unstable.

17. a) Divide both sides of the equation $N_{t+1} = (1+r)N_t e^{(1-\frac{N_t}{K})}$ by the quantity K. Since $M_t = \dfrac{N_t}{K}$ we obtain $M_{t+1} = (1+r)M_t e^{1-M_t} = (1+r)M_t e e^{-M_t} = e(1+r)M_t e^{-M_t}$.

b) Solve the equation $sMe^{-M} = M$. This gives $M = 0$ or $se^{-M} = 1$. The latter equation has solution $M = \ln s$ (take logs of both sides).

c) If $f(M) = sMe^{-M}$ then $f'(M) = se^{-M}(1-M)$. Then $f'(0) = s > e > 1$ so $N^* = 0$ is unstable. For the other steady state we obtain $f'(\ln s) = 1 - \ln s$. The steady state is stable if $|1-\ln s| < 1$ or $-1 < 1 - \ln s < 1$. This yields $0 < \ln s < 2$. Since $s > e$ we have $\ln s > 1$ so stability requires $\ln s < 2$ or $s < e^2$.

21. a) $N_2 = 1.195$, $N_3 = 0.131$, $N_4 = 0.239$, $N_5 = 1.255$.

b) If N^* is a constant solution then $N^* = sN^* e^{-N^*}$, so N^* is a fixed point of the function $f(N) = sNe^{-N}$. See the solution to Exercise 17c.

Chapter 18

1 a) $N_1 = 100, N_2 = 50, N_3 = 100, N_4 = 200$.

b) Using the binomial distribution with $n = 4$ trials and probability of success $p = 0.75$ we obtain the table below:

N_4	3.125	12.5	50	200	800
$\log_{10} N_4$	0.495	1.097	1.699	2.301	2.903
Probability	0.004	0.047	0.211	0.422	0.316

c) Using the probability distribution in b) we have $E(N_4) = 348.6$ and $E(\log_{10} N_4) = 2.30$. (Note it is not true that $E(\log_{10} N_4) = \log_{10}(E(N_4))$!) $\sigma_{N_4} = 313.6$, $\sigma_{\log_{10} N_4} = 0.52$

d) Identical argument as in the text.

e) Use that the binomial distribution X has mean $0.75t$. The mean of $\log_{10} N_t$ increases over time. In other words there is an underlying upward drift in the population. This is a reflection of the more frequent occurrence of good years.

f) Formula follows the method in the text, but uses the result $\sigma_X = \frac{\sqrt{3}}{2}\sqrt{t}$. Since the expected value increases like t and the standard deviation like $\sqrt{t}$, in a short time it becomes unlikely that N_t will drop below one. Of course, if the population is small this may happen by chance for small t.

3 a) $E(X) = 1$ and $\sigma_X = 1$

b) For the distribution defined by Table 18.5 we have $E(X) = 1$ and $\sigma_X = 1/\sqrt{2} \approx .707$.

c) Both distributions have the same mean, so each has a tendency to leave the population invariant. The less variable distribution would produce on average smaller swings in the population, thereby lowering the possibility of extinction for small populations.

Appendix A

1. A workbook is a collection of one or more worksheets.

3. With *Excel's* usual settings, when a cell contains a formula the formula is displayed in the formula bar and the value of the formula is displayed on the worksheet.

5. a) 16 b) 4 c) 6 d) 9 e) EXP(0) (text) f) 0.4 g) 2/5 (text) h) 1

7. Cell C3: =A1 +C1 +D1, value 2 Cell C4: =SUM(C1:D2), value 3

9. D1 has formula =B1+C1, value 3 E1 has formula =C1+D1, value 5,

   F1 has formula =D1+E1, value 8 G1 has formula =E1+F1, value 13