Chapter 5

Exercises for Section 5.2

1. Probabilities must lie between 0 and 1 and add to 1. The value 1.10 is clearly incorrect for a probability. If it is changed to 0.11 then the set of probabilities add to 1.

2. $\text{pr}(X = 1) = \frac{176}{200} = 0.88$, $\text{pr}(X = 2) = \frac{22}{200} = 0.11$, etc. Placing these into a table, we get the following.

x	1	2	3
$\text{pr}(x)$	0.88	0.11	0.01

3. We have, using independence, $\text{pr}(BBB) = \text{pr}(B) \times \text{pr}(B) \times \text{pr}(B) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$. Every other outcome has the same probability. (This is the same probability distribution as in Example 4.4.6(b).) $X = 0$ corresponds to one outcome (BBB) and thus has probability $\frac{1}{8}$. $X = 1$ corresponds to three outcomes (GBB, BGB, and BBG) and thus has probability $\frac{3}{8}$, and so on. We obtain the following.

x	0	1	2	3
$\text{pr}(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

4. $\text{pr}(4 < X < 8) = \text{pr}(X = 5) + \text{pr}(X = 6) + \text{pr}(X = 7) = 0.844$ or $\text{pr}(4 < X < 8) = \text{pr}(X \leq 7) - \text{pr}(x \leq 4) = 0.882 - 0.038 = 0.844$

5. **(a)** The cumulative probabilities are given below.

x	3	4	5	6	7	8	9	10	11	12	13
$\text{pr}(X = x)$	.07	.01	.09	.01	.16	.25	.20	.03	.02	.11	.05
$\text{pr}(X \leq x)$	.07	.08	.17	.18	.34	.59	.79	.82	.84	.95	1.0

(b) $\text{pr}(X \leq 5) = \text{pr}(3) + \text{pr}(4) + \text{pr}(5) = .07 + .01 + .09 = 0.17$. We also read the value directly off the cumulative probabilities row.

(c) $\text{pr}(X > 9) = \text{pr}(10) + \text{pr}(11) + \text{pr}(12) + \text{pr}(13) = .03 + .02 + .11 + .05 = 0.21$, or, using the cumulative probabilities row, $\text{pr}(X > 9) = 1 - \text{pr}(x \leq 9) = 1 - 0.79 = 0.21$.

(d) $\text{pr}(X \geq 9) = \text{pr}(9) + \ldots + \text{pr}(13) = 0.41$, or, using the cumulative probabilities row, $\text{pr}(X \geq 9) = 1 - \text{pr}(x \leq 8) = 1 - 0.59 = 0.41$.

(e) $\text{pr}(X < 12) = \text{pr}(3) + \ldots + \text{pr}(11) = 0.84$, or using the cumulative probabilities table, $\text{pr}(X < 12) = \text{pr}(X \leq 11) = 0.84$.

(f) $\text{pr}(5 \leq X \leq 9) = \text{pr}(5) + \ldots + \text{pr}(9) = 0.71$, or using the cumulative probabilities table, $\text{pr}(5 \leq X \leq 9) = \text{pr}(X \leq 9) - \text{pr}(X \leq 4) = 0.79 - 0.08 = 0.41$.

(g) $\text{pr}(4 < X < 11) = \text{pr}(5) + \ldots + \text{pr}(10) = 0.74$, or using the cumulative probabilities table, $\text{pr}(4 < X < 11) = \text{pr}(X \leq 10) - \text{pr}(X \leq 4) = 0.82 - 0.08 = 0.74$.

(h) and **(i)** The probabilities for the values of x listed in the table add to 1 so these take up all the possible values. Everything outside this has probability 0. Thus $\text{pr}(X = 14) = 0$ and $\text{pr}(X < 3) = 0$.

6. Arguing exactly as in the Case Study, $0.9^3 = 0.729$. As $\text{pr}(X \leq 6) = 0.469$ and $\text{pr}(X \leq 7) = 0.522$, she should try seven times.

Exercises for Section 5.3

1. **(a)** 0.2668. **(b)** $1 - \text{pr}(X \leq 3) = 0.3504$. **(c)** $1 - \text{pr}(X \leq 6) = 0.01059$.
(d) $\text{pr}(X \leq 6) = 0.9894$. **(e)** 0.1493. **(f)** $\text{pr}(X \leq 7) - \text{pr}(X \leq 3) = 0.3488$.
(g) $\text{pr}(X \leq 7) - \text{pr}(X \leq 3) = 0.3488$. **(h)** 0 as X only takes values $0, 1, \ldots, 10$.
(i) $\text{pr}(X \leq 10) = 1$.

2. **(a)** Here the 20% refers to a conceptual population rather an actual one, so that the coin-tossing model is a candidate. The urn model does not apply, as the 10 chosen cars are not a simple random sample of all the cars parking. You need to have the probability of overstaying remaining constant at 0.2. Here $n = 10$ and $p = 0.2$. The arrival of each car is like a binomial trial. You can use Binomial$(n = 10, p = 0.2)$.

(b) Urn model. The 50 cars watched must be a simple random sample of the cars. Here $N = 10,000$; M, the total number of overstayers, is unknown; and $n = 50$. You can approximate by Binomial$(n = 50, p)$, where $p = \frac{M}{N}$, as $\frac{n}{N} < 0.1$. You may be able to use $p = 0.2$ from part (a).

(c) Urn model. The 50 people dialed must be a simple random sample of subscribers. Here $N = 7400$, $M = 2730$, and $n = 50$. You can approximate by Binomial$(n = 50, p = \frac{M}{N} = 0.3689)$ as $\frac{n}{N} < 0.1$.

(d) Urn model. Here N and M are large unknown numbers, $\frac{M}{N} = 0.45$, and $n = 100$. You can approximate by Binomial$(n = 100, p = 0.45)$ as you can expect $\frac{n}{N} < 0.1$.

(e) Urn model. Here $N = 100$, $M = 12$ and $n = 7$. You can approximate by Binomial$(n = 100, p = \frac{M}{N} = 0.12)$ as $\frac{n}{N} = \frac{7}{100} < 0.1$.

(f) Urn model. Here N and M are large unknown numbers, $\frac{M}{N} = 0.64$, and $n = 50$. You can approximate by Binomial$(n = 50, p = 0.64)$ as you can expect $\frac{n}{N} < 0.1$.

(g) Urn model. Here $N = 188$, $M = 99$, and $n = 30$. You cannot use the Binomial as $\frac{n}{N} > 0.1$.

(h) Coin-tossing model. Here $n = 10$ and $p = \frac{1}{6}$. Can use Binomial$(n = 10, p = \frac{1}{6})$

(i) Urn model. Here $N = 52$, $M = 4$, and $n = 7$. You need random shuffling of the pack before dealing. You cannot use the Binomial as $\frac{n}{N} > 0.1$.

(j) Urn model. Here N and M are large unknown numbers, $\frac{M}{N} = 0.1$ and $n = 30$. You can approximate by Binomial$(n = 30, p = 0.1)$ as you can expect $\frac{n}{N} < 0.1$.
Either model is a candidate. For an urn model, the 30% must refer to the actual population sampled with M and N unknown and $n = 20$. For the coin-tossing model, you must have p constant. Since you can expect $\frac{n}{N} < 0.1$, either model will lead to Binomial$(n = 20, p = 0.3)$.

(k) Either model is a candidate, both leading to Binomial$(n = 20, p = 0.3)$. For example, urn model applies if we think of the 20 bearings being sampled from the existing population of bearings of which 30% will function for a year of continuous use. If, however, we think in terms of the bearings being random items from a continuous manufacturing process producing items with the "30% will function" being stable (or constant) over time, the coin-tossing model applies.

(l) Like (a), either model is a candidate. The urn model, however, will apply only if the 50 patients can be regarded as a simple random sample of patients, which is probably not the case. The coin-tossing model is dubious as p may not be constant.

Exercises for Section 5.4.1

1. $\mathrm{E}(X) = \sum x\,\mathrm{pr}(x) = 2 \times 0.2 + 3 \times 0.1 + 5 \times 0.3 + 7 \times 0.4 = 5.0.$

2. $\mathrm{E}(X) = \sum x\,\mathrm{pr}(x) = 0 \times 0.49 + 1 \times 0.42 + 2 \times 0.09 = 0.6 \quad (= n\,p).$

Exercises for Section 5.4.2

1. $\mathrm{E}(X-\mu)^2\mathrm{pr}(x) = (2-5)^2 \times 0.2 + (3-5)^2 \times 0.1 + (5-5)^2 \times 0.3 + (7-5)^2 \times 0.4 = 3.8.$
$\mathrm{sd}(X) = \sqrt{3.8} = 1.9494.$

2. $\mathrm{E}(X-\mu)^2\mathrm{pr}(x) = (0-1.25)^2 \times \frac{1}{8} + (1-1.25)^2 \times \frac{5}{8} + (2-1.25)^2 \times \frac{1}{8} + (3-1.25)^2 \times \frac{1}{8}$
$= 0.6875, \qquad \mathrm{sd}(X) = \sqrt{0.6875} = 0.8292.$

3. $\mathrm{E}(X-\mu)^2\mathrm{pr}(x) = (0-0.6)^2 \times 0.49 + (1-0.6)^2 \times 0.42 + (2-0.6)^2 \times 0.09 = 0.42.$
$\mathrm{sd}(X) = \sqrt{0.42} = 0.6481.$

Exercises for Section 5.4.3

(a) $\mathrm{E}(2X) = 2\,\mathrm{E}(X) = 6, \quad \mathrm{sd}(2X) = 2\,\mathrm{sd}(X) = 4.$

(b) $\mathrm{E}(4+X) = 4 + \mathrm{E}(X) = 7, \quad \mathrm{sd}(4+X) = \mathrm{sd}(X) = 2.$

(c) As for (b), since $\mathrm{E}(X+4) = \mathrm{E}(4+X)$ and $\mathrm{sd}(X+4) = \mathrm{sd}(4+X)$.

(d) $\mathrm{E}(3X+2) = 3\,\mathrm{E}(X) + 2 = 11, \quad \mathrm{sd}(3X+2) = 3\,\mathrm{sd}(X) = 6.$

(e) $\mathrm{E}(4+5X) = 4 + 5\,\mathrm{E}(X) = 19, \quad \mathrm{sd}(4+5X) = 5\,\mathrm{sd}(X) = 10.$

(f) $\mathrm{E}(-5X) = -5\,\mathrm{E}(X) = -15, \quad \mathrm{sd}(-5X) = 5\,\mathrm{sd}(X) = 10.$

(g) $\mathrm{E}(-5X+4) = -5\,\mathrm{E}(X) + 4 = -11, \quad \mathrm{sd}(-5X+4) = 5\,\mathrm{sd}(X) = 10.$

(h) As for (g).

(i) $\mathrm{E}(-7X-9) = -7\,\mathrm{E}(X) - 9 = -30, \quad \mathrm{sd}(-7X-9) = 7\,\mathrm{sd}(X) = 14.$

Review Exercises 5

1. In the following "N/A" is used when neither model is applicable.

(a) Coin-tossing model. $X_1 \sim \text{Binomial}(n = 20, p = 0.2)$.

(b) Urn model with $N = 1000$, $M = 100$, and $n = 20$. You can use the Binomial approximation, $X_2 \sim \text{Binomial}(n = 20, p = 0.1)$ as $\frac{n}{N} < 0.1$.

(c) N/A.

(d) Coin-tossing model. $X_4 \sim \text{Binomial}(n = 120, p = 0.6)$.

(e) Urn model with $N = 120$, $M = 70$ and $n = 10$. You can use the Binomial approximation, $X_5 \sim \text{Binomial}(n = 10, p = \frac{7}{12})$, as $\frac{n}{N} < 0.1$.

(f) Urn model with $N = 20$, $M = 9$ and $n = 15$. You cannot use a Binomial approximation as $\frac{n}{N} > 0.1$.

(g) Coin-tossing model. $X_7 \sim \text{Binomial}(n = 12, p = \frac{1}{6})$.

(h) N/A. (Not the number of "heads" in a fixed number of "tosses".)

(i) Coin-tossing model. $X_9 \sim \text{Binomial}(n = 12, p = \frac{1}{36})$.

(j) Urn model with $N = 98$, $M = 44$, and $n = 7$. You can use the Binomial approximation, $X_{10} \sim \text{Binomial}(n = 7, p = \frac{44}{98})$, as $\frac{n}{N} < 0.1$.

2. (a) No. He does not guess all the questions so that p is not constant.

(b) Yes. His little brother will guess them all so that $p = 0.2$.

(c) Yes, provided the probability of having an error free page is constant.

(d) Probably no. You can expect the probability of an error-free page to depend on the number of mathematical symbols and numbers on the page.

(e) No, as the number of trials is not fixed in advance.

(f) No, as what happens in successive months (trials) will almost certainly not be independent.

3. (a) $p = \dfrac{\text{sampled area}}{\text{population area}} = \frac{20\times100\times100}{2000\times2000} = \frac{1}{20}$.

(b) The 420 animals may be regarded as 420 independent Binomial trials each with probability of success, where success means "found in the sample area"and failure means "found outside the sample area." The four assumptions are satisfied because the animals (trials) are independent.
$X \sim \text{Binomial}(n = 400, p = \frac{1}{20})$.

(c) For a single plot $p = \frac{100\times100}{2000\times2000} = \frac{1}{400}$ and $W \sim \text{Binomial}(n = 420, p = \frac{1}{400})$.

(d) Any two of the following.

(i) Animals tend to exhibit social tendencies and so are not generally independent.

(ii) Animals do not move randomly but usually have well-defined territories or "home ranges."

(iii) The presence of observers may disturb the animals so that they move out of the area.

(iv) Some animals may be missed. Deer, for example, are very hard to spot.

4. **(a)** The ten fish are a simple random sample of fish. Fish do not lose their tags. All the fish stay alive and no new fish are born.

(b) Binomial$(n = 10, p)$ where $p = \frac{M}{N} = \frac{50}{1000} = 0.05$.

(c) $\text{pr}(X \geq 1) = 1 - \text{pr}(X = 0) = 0.4013$.

5. **(a)** $L \sim \text{Binomial}(n = 160, p = \frac{120}{80{,}000})$.

(b) $\text{E}(L) = np = 0.24$ (about one every four years).

(c) Each ship has the same probability of being lost and ships are lost or not lost independently. This is probably not true, but it may be a reasonable approximation.

6. Let X = number of women from the 30 who become pregnant in the first year. Then $X \sim \text{Binomial}(n = 30, p = 0.11)$.

(a) $\text{pr}(X = 0) = 0.0303$.

(b) $\text{pr}(X \leq 2) = 0.3442$.

(c) pr(*No pregnancies in 2 years*)

$$
\begin{aligned}
&= \text{pr}(\textit{No pregnancies in 1st year} \text{ and } \textit{no pregnancies in 2nd year}) \\
&= \text{pr}(\textit{No pregnancies in 1st year}) \times \text{pr}(\textit{No pregnancies in 2nd year}) \\
&= 0.0303 \times 0.0303 \quad \text{(Assuming independence and constant probs)} \\
&= 0.0009.
\end{aligned}
$$

7. **(a)** Binomial$(n = 12, p = 0.18)$.

(b) pr(*No failures*) $= \text{pr}(X = 0) = 0.09242$.

8. The drugs will be discredited if less than 7 of the 12 patients recover. Let X = the number that recover and assume that $X \sim \text{Binomial}(n = 12, p = 0.5)$. Then $\text{pr}(X \leq 6) = 0.6128$. (What assumptions have been made?)

9. If X = number of rods that perform satisfactorily, you assume that $X \sim \text{Binomial}(n = 10, p = 0.80)$. Then $\text{pr}(X \leq 4) = 0.006369$.

10. **(a)** If X = number of sixes when 6 dice are rolled, then $X \sim \text{Binomial}(n = 6, p = \frac{1}{6})$. $\text{pr}(X \geq 1) = 1 - \text{pr}(X = 0) = 0.6651$.

(b) If Y = number of sixes when 12 dice are rolled, then $Y \sim \text{Binomial}(n = 12, p = \frac{1}{6})$. $\text{pr}(Y \geq 2) = 1 - \text{pr}(Y \leq 1) = 0.6187$.

(c) If Z = number of sixes when 18 dice are rolled, then $Z \sim \text{Binomial}(n = 18, p = \frac{1}{6})$. $\text{pr}(Z \geq 3) = 1 - \text{pr}(Z \leq 2) = 0.5974$.

11. **(a)** The value -0.39 is clearly in error (probabilities cannot be negative). We replace it with $1 - (0.23 + 0.18 + 0.17 + 0.13) = 0.29$.

(b) $\text{pr}(X \geq 1) = 0.18 + 0.17 + 0.13 = 0.48$.

(c) $\text{pr}(X \leq 0) = 0.23 + 0.29 = 0.52$.

(d) $\mathrm{E}(X) = \sum x\,\mathrm{pr}(x) = (-3)\times 0.23 + 0\times 0.29 + 1\times 0.18 + 3\times 0.17 + 8\times 0.13 = 1.04$,
$\mathrm{E}[(X-\mu)^2] = \sum (x-\mu)^2 \mathrm{pr}(x) = (-3-1.04)^2\times 0.23 + (0-1.04)^2\times 0.29.$
$+(1-1.04)^2\times 0.18 + (3-1.04)^2\times 0.17 + (8-1.04)^2\times 0.13 = 11.0184.$
$\mathrm{sd}(X) = \sqrt{11.0184} = 3.32.$

12. **(a)** Let B = "black face shows uppermost" = $\{2,5\}$.
Let E = "even numbered face" = $\{2,4,6\}$. Then $\mathrm{pr}(B \text{ or } E) = \frac{4}{6} = \frac{2}{3}$.

(b) $\mathrm{pr}(X=-10) = \mathrm{pr}(\{5\}) = \frac{1}{6}$; $\mathrm{pr}(X=-4) = \mathrm{pr}(\{2\}) = \frac{1}{6}$;
$\mathrm{pr}(X=0) = \mathrm{pr}(\{1\}) = \frac{1}{6}$; $\mathrm{pr}(X=3) = \mathrm{pr}(\{3\}) = \frac{1}{6}$;
$\mathrm{pr}(X=4) = \mathrm{pr}(\{4\}) = \frac{1}{6}$; $\mathrm{pr}(X=6) = \mathrm{pr}(\{6\}) = \frac{1}{6}$.

(c) Expected amount in dollars won:
$\mathrm{E}(X) = \sum x\,\mathrm{pr}(x) = -10\times\frac{1}{6} - 4\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 6\times\frac{1}{6} = -\frac{1}{6}$. No, since one expects to lose.

13. **(a)** In the table, $a(b) = a\times 10^b$.

x	0	1.61	7.85	25.90	34.29	224.13	558.00
$\mathrm{pr}(x)$	0.9333	6.511(−2)	1.190(−3)	2.381(−4)	9.524(−5)	1.720(−5)	1.190(−5)

(b) $\mathrm{E}(X) = \sum x\mathrm{pr}(x)$ is most quickly calculated by noting that here $\sum x\ \mathrm{pr}(x) = \sum \frac{x\times n(x)}{3{,}780{,}000}$, where $n(x)$ is the number of vouchers with value x. Thus
$\mathrm{E}(X) = \frac{1.61\times 246130 + 7.85\times 4500 + \ldots + 558.00\times 45}{3{,}780{,}000} = 0.1341$, or 13.41 cents.

(c) We calculate $\mathrm{E}[(X-\mu)^2] = \sum (x-\mu)^2\mathrm{pr}(x) = \frac{\sum (x-\mu)^2\times n(x)}{3{,}780{,}000}$
$= \frac{(1.61-0.1341)^2\times 246130 + (7.85-0.1341)^2\times 4500 + \ldots + (558.00-0.1341)^2\times 45}{3{,}780{,}000}$ and find the square root of the answer to obtain $\mathrm{sd}(X) = 2.2509$.

(d) The expected redeemable value in cents is calculated by (since we only incur a postage cost if we get a voucher)
$\mathrm{E}(X) - (\textit{Postage Cost}) \times \mathrm{pr}(\textit{getting a voucher})$
$= 13.41 - 40\times 0.0677 = 10.74 \approx 11$ cents. Thus the expected cost of a box of Almond Delight is \$1.84 – \$0.11 = \$1.73. At \$1.60, the alternative brand is better value.

14. Eventually you must get a prize or nothing. Dividing the probabilities by 0.9 you get

Prize	\$ 0	\$3	\$7	\$11	\$ 21	\$ 2100
Prob.	0.9333	0.0370	0.0111	0.011	7.407×10^{-3}	3.703×10^{-6}

$\mathrm{E}(X) = \$0\times 0.933 + \$3\times 0.037 + \ldots + \$2100\times 3.7\times 10^{-6} = 0.4742.$
$\mathrm{sd}(X) = \sqrt{(0-0.4774)^2\times 0.933 + \ldots + (2100-0.4774)^2\times 3.7\times 10^{-6}}$
$= 4.6473.$

15. Let X = number of defective rivets in the sample.

(a) $X \sim \mathrm{Binomial}(n=8, p=0.01)$. Then $\mathrm{pr}(X\geq 2) = 1 - \mathrm{pr}(X\leq 1) = 0.00269$.

(b) $Y \sim \mathrm{Binomial}(n=8, p=0.02)$. Then $\mathrm{pr}(X\leq 1) = 0.9897$.

***16.** **(a)** $X \sim \mathrm{Binomial}(n, p = \frac{1}{12{,}000{,}000})$. The following four assumptions support the Binomial: (i) the total number of trials or number of couples is fixed; (ii) a given couple either fits or doesn't fit the description; (iii) the chance of any couple matching the description is constant, and (iv) independence is assumed.

(b) $$\text{pr}(X \geq 2 \,|\, X \geq 1) = \frac{\text{pr}(X \geq 2 \text{ and } X \geq 1)}{\text{pr}(X \geq 1)} = \frac{\text{pr}(X \geq 2)}{\text{pr}(X \geq 1)} = \frac{1 - \text{pr}(X \leq 1)}{1 - \text{pr}(X = 0)}.$$

(c) $n = 10^6$, pr $= 0.04109$; $n = 4 \times 10^6$, pr $= 0.1574$; $n = 10^7$, pr $= 0.3595$.

17. (a) Let X = number of attempts made.
From the case study:

x	1	2	3	4
$\text{pr}(X = x)$	0.1	0.09	0.081	0.729

(b) $\text{E}(X) = \sum x \, \text{pr}(x) = 1 \times 0.1 + 2 \times 0.09 + 3 \times 0.081 + 4 \times 0.729 = 3.439.$

$\text{sd}(X) = \sqrt{\sum (x - \mu)^2 \text{pr}(x)} = \sqrt{(1 - 3.439)^2 \times 0.1 + ... + (4 - 3.439)^2 \times 0.729}$
$= \sqrt{1.0263} = 1.0131.$

(c) $\text{E}(Cost) = \$7,000 \times \text{E}(X) = \$7,000 \times 3.439 = \$24,073.$

(d) $\text{pr}(\textit{Still childless}) = (\frac{9}{10})^4 = 0.6561.$ [Alternatively, use $\text{pr}(Y = 0)$ where $Y \sim \text{Binomial}(n = 4, p = 0.1)$.]

(e) and **(f)** The frequencies are computed in the following table.

Group	Starting number	Yes 1st	No	Yes 2nd	No	Yes 3rd	No	Yes 4th
1	30,000	6,000	24,000	4,800	19,200	3,840	15,360	3,072
2	30,000	3,000	27,000	2,700	24,300	2,430	21,870	2,187
3	40,000	400	39,600	396	39,204	392	38,812	388
Total	100,000	9,400	90,600	7,896	82,704	6,662	76,042	5,647

Here "Yes" means the number who were successful, "No" means the number unsuccessful and "1st" means first attempt, and so on.

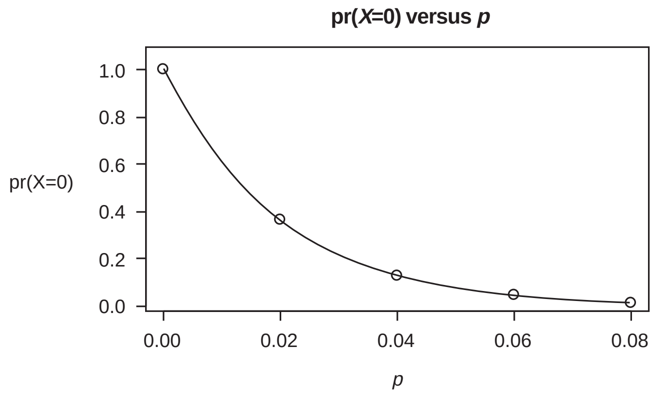
(g) Proportion who succeed on first attempt $= \frac{9{,}400}{100{,}000} = 0.094.$
Proportion who succeed on second attempt $= \frac{7{,}896}{90{,}600} = 0.0872.$
Proportion who succeed on third attempt $= \frac{6{,}662}{82{,}704} = 0.0806.$
Proportion who succeed on fourth attempt $= \frac{5{,}647}{76{,}042} = 0.0743.$

18. (a) You can use either an urn model with $\frac{n}{N} < 0.1$ and $p = \frac{M}{N}$ the proportion of defectives, or a coin tossing model with p the probability of getting a defective item. Either way you have $X \sim \text{Binomial}(n = 50, p)$.

(b) Implicit in our answers to (a) are the assumptions that the manufacturing process is stable or is in control, that no time drifts occur in the process which may otherwise lead to deterioration resulting from the wearing of machine parts or the loss of accuracy in machine settings. This implies that the probability of a defective item produced remains constant. You also assume that the sampling is taken with a sufficient time interval between selections to preserve the independence condition.

(c) Using $X \sim \text{Binomial}(n = 50, p)$, for $p = 0, 0.02, 0.04, 0.06, 0.08$ you obtain:

p	0	0.02	0.04	0.06	0.08
$\text{pr}(X = 0)$	1	0.3642	0.1299	0.0453	0.0155



(d) $\text{pr}(X = 0) = (1 - p)^{50}$. If $\text{pr}(X = 0) = 0.1$, then $1 - p = 0.1^{1/50} = 0.9550$. Hence $p = 0.045$.

19. Let X = number of incorrect identifications out of 91 calves. Then $X \sim \text{Binomial}(n = 91, p)$ and you can compute $\text{pr}(X \leq 1)$ for different p, namely:

p	0	0.02	0.04	0.06	0.08
$\text{pr}(X \leq 1)$	1	0.4545	0.1167	0.0244	0.0045

By calculating a couple more points between 0.04 and 0.06, and plotting a graph, you can estimate that the value of p giving $\text{pr}(X \leq 1) = 0.1$ is 0.042.

20. **(a)** Let X = number delivered in a day. Then $X \sim \text{Binomial}(n = 15, p = 0.95)$, and $\text{pr}(X \leq 11) = 0.00547$.

(b) Independence of the delivery of each of the 15 letters.

(c) If N.Z. Post's claim was true, we would be highly unlikely to get so few letters delivered the next day. 99.5% of the time $(1 - .005)$, we would get more than 11 letters delivered the next day. The data does not support N.Z. Post's claim.

21. If X is the number of mutations, then $X \sim \text{Binomial}(n = 20,000, p = \frac{1}{10,000})$.

(a) $\text{pr}(X = 0) = 0.1353$.

(b) $\text{pr}(X \geq 1) = 1 - \text{pr}(X = 0) = 0.8647$.

(c) $\text{pr}(X \leq 3) = 0.8571$.

22. If X is the number of cases for 150,000 people, then $X \sim \text{Binomial}(n = 150{,}000, p = \frac{3.1}{100{,}000})$. $E(X) = np = 4.65$, i.e., 4 or 5 claims per year. We also have, $\text{pr}(X > 6) = 1 - \text{pr}(X \leq 6) = 0.1886$.

23. **(a)** $X \sim \text{Binomial}(n = 10, p = \frac{1}{50})$. The Binomial distribution is appropriate since the person either wins or does not win a bottle, sampling is with replacement so that the draws are (supposedly) independent, and the probability of winning is the same for each draw, namely, $\frac{1}{50}$.

(b) **(i)** $\text{pr}(X = 0) = 0.81707$, $\text{pr}(X \leq 2) = 0.999136$. **(ii)** $\text{pr}(X \geq 3) = 1 - \text{pr}(X \leq 2) = 0.000864$.

(c) Since the probability in (b)(ii) is so small, the fact that one person has won three bottles would lead us to suspect that the names were not properly stirred.

(d) By b(ii), $\text{pr}(E_i) = \text{pr}(X \geq 3) \approx 0.00086 \neq 0$. If the event E_i has occurred for at least 3 people, then the event E_j $(j \neq i)$ cannot occur for the other people (since there are only 10 bottles to be won), so $\text{pr}(E_j) = 0$. This establishes that the outcome of E_i depends on the occurrence or otherwise of the other outcomes. Hence the E_i 's are not independent.

(e) Let Y = number of people who win 3 or more bottles. Assuming independence of the E_i's and regarding E_i as a binomial trial you have $Y \sim \text{Binomial}(n = 50, p = 0.000864)$. Hence $\text{pr}(Y \geq 1) = 1 - \text{pr}(Y = 0) = 0.0423$.

(f) Although the last probability in (e) is somewhat larger (slightly over 4 chances in 100), our opinion in (c) is not changed.

24. (a) The forty weeks of data is given below. Each row corresponds to a week, 1 corresponds to a day with rain and 0 corresponds to a day without rain.

Week 1 :	1	0	0	1	1	0	0
Week 2 :	1	0	0	0	0	0	0
Week 3 :	0	0	1	0	0	1	0
Week 4 :	0	1	1	0	0	1	1
Week 5 :	0	1	0	0	0	0	0
Week 6 :	0	0	1	0	1	0	0
Week 7 :	0	1	1	0	1	0	0
Week 8 :	1	1	0	0	0	1	0
Week 9 :	1	0	0	0	0	0	0
Week 10:	1	0	0	0	0	0	0
Week 11:	1	0	0	1	1	1	1
Week 12:	0	1	1	1	1	0	1
Week 13:	0	0	0	0	1	0	1
Week 14:	1	1	0	1	0	1	0
Week 15:	1	0	0	1	0	0	0
Week 16:	0	0	0	0	1	1	1
Week 17:	0	0	1	0	0	1	1
Week 18:	1	1	1	0	0	0	1
Week 19:	1	0	0	1	0	0	0
Week 20:	0	0	0	0	0	0	0
Week 21:	1	0	0	0	1	0	0
Week 22:	1	0	0	0	1	1	0
Week 23:	0	1	1	0	0	0	0
Week 24:	1	1	0	1	0	0	1
Week 25:	1	1	0	1	0	0	1
Week 26:	0	1	0	0	0	1	0
Week 27:	0	1	1	1	1	0	0
Week 28:	0	0	1	0	1	0	1
Week 29:	1	0	1	1	0	0	1
Week 30:	1	0	1	0	0	0	0
Week 31:	0	0	1	0	0	0	0
Week 32:	0	1	0	1	0	1	1
Week 33:	1	0	1	0	1	1	1
Week 34:	1	0	0	1	0	1	0
Week 35:	0	1	0	1	1	0	0
Week 36:	1	0	0	0	0	1	1
Week 37:	1	1	0	0	0	1	1
Week 38:	1	0	1	0	0	1	0
Week 39:	1	1	0	0	1	1	1
Week 40:	1	0	1	1	0	1	1

In our simulated weeks given above, we obtain the following (your's will be somewhat different):

- There were **26** weeks with more dry days than wet days.
- There were **14** weeks with more wet days than dry days.
- The longest run of wet days was **4** days (weeks 11, 12 and 27).
- The longest run of dry days was **7** days (week 20).
- The proportion of wet days was $\frac{117}{280} = 0.4178$.

(b) The number of wet days for each of the forty weeks is as follows.

3 1 2 4 1 2 3 3 1 1 5 5 2 4 2 3 3 4 2 0 2 3 2 4 4 2 4 3 4 2 1 4 5 3 3 3 4 3 5 5.

Each simulated day has a probability 0.4 of being wet, days are generated independently and we count the number of wet days ("heads") in a fixed number of "tosses", namely 7, so the distribution is Binomial($n = 7, p = 0.4$).
Forty random observations from a Binomial($n = 7, p = 0.4$) distribution were generated and are given below

0 5 3 3 3 2 1 4 2 4 5 3 2 2 3 4 3 3 2 2 4 3 2 1 4 3 6 3 2 4 5 3 1 3 5 2 1 4 5 2.

(c) Two hundred Binomial($n = 7, p = 0.4$) random numbers were generated; the numbers and the relative-frequency table are given below.

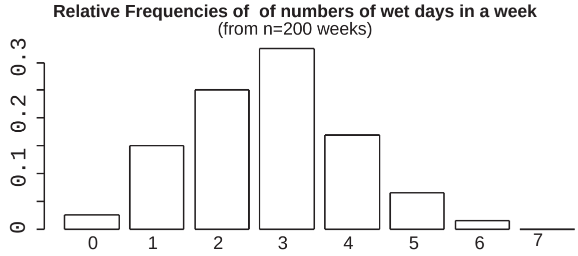
0	4	4	3	2	3	3	4	3	3	4	3	4	3	6	1	4	2	4	3	4	1	3	2	4
2	3	3	2	2	4	2	3	2	4	2	2	2	4	3	3	6	3	4	3	4	4	4	2	1
2	4	4	2	3	4	1	3	1	4	3	3	2	2	4	6	2	4	2	5	2	1	2	4	1
3	3	1	4	2	1	4	3	3	3	3	3	3	2	5	2	1	4	1	3	4	4	3	3	4
4	3	5	3	2	3	3	3	4	1	3	2	6	4	3	4	2	2	4	4	2	3	1	2	2
4	2	2	2	3	5	3	5	2	2	4	2	3	2	2	3	3	5	1	3	2	3	1	2	1
2	4	5	4	2	3	3	2	4	3	2	5	5	5	5	2	2	4	0	4	2	4	3	6	4
1	4	4	3	2	3	2	1	4	5	4	3	1	4	1	2	3	4	1	1	4	3	2	4	4

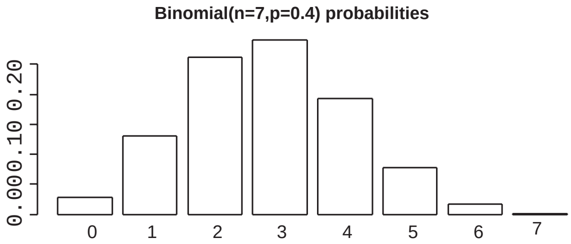
Days wet per week	0	1	2	3	4	5	6
Frequency	5	30	50	65	34	13	3
Rel. frequency	0.025	0.150	0.25	0.325	0.17	0.065	0.015

The probabilities for the Binomial($n = 7, p = 0.4$) distribution are shown in the following table. You will see that they are reasonably similar to the relative frequencies above. [The following probabilities have been rounded to 3 decimal places.]

x	0	1	2	3	4	5	6	7
pr(x)	0.028	0.131	0.261	0.290	0.194	0.077	0.017	0.002

A bar graph of the relative frequency of wet days in each of 200 weeks and a bar graph of the Binomial($n = 7, p = 0.4$) probabilities are given below.





The bar graphs are quite similar in shape. (The most distinct differences come in the bars relating to 2 and 4.). If we had used, say, 10,000 Binomial random numbers instead of only 200, there would be no visible difference.

(d) A relative frequency table of the longest run of dry days in each of 200 weeks is given below.

Longest dry	1	2	3	4	5	6	7
Frequency	5	16	8	6	1	3	1
Rel. Frequency	0.125	0.4	0.2	0.15	0.25	0.075	0.025

(e) Using the table in (d) the expected value of the longest run of dry days can be estimated by

$$1 \times 0.125 + 2 \times 0.4 + 3 \times 0.2 + 4 \times 0.15 + 5 \times 0.025 + 6 \times 0.075 + 7 \times 0.025 = 2.875.$$

Chapter 6

Exercises on Section 6.2.2

Note: Recall that for continuous distributions like the Normal, we do not have to worry whether interval endpoints are included or excluded.

2. We use $X \sim \text{Normal}(\mu = 266, \sigma = 16)$.

(a) $\text{pr}(X < 252) = \text{pr}(X \leq 252) = 0.1908$.
(b) $\text{pr}(260 < X \leq 280) = \text{pr}(X \leq 280) - \text{pr}(X \leq 260) = 0.4554$.
(c) $\text{pr}(X > 280) = 1 - \text{pr}(X \leq 280) = 0.1908$.

3. We use $X \sim \text{Normal}(\mu = 100, \sigma = 15)$.

(a) $\text{pr}(X < 80) = 0.09121$.
(b) $\text{pr}(85 < X \leq 110) = \text{pr}(X \leq 85) - \text{pr}(X \leq 110) = 0.5889$.
(c) $\text{pr}(X > 120) = 1 - \text{pr}(X \leq 120) = 0.09121$.

Exercises on Section 6.2.3

2. **(a)** 249.4. **(b)** The 98th percentile, which is 298.9.
(c) The 10th percentile, which is 245.5. **(d)** The 70th percentile, which is 274.4.

3. IQs are given to the nearest whole number. **(a)** 113.
(b) The 99th percentile, which is 135. **(c)** The 30th percentile, which is 92.

Exercises on Section 6.2.4

1. **(a)** From the 10th percentile to the 90th percentile, or $[154.75, 170.65]$.
(b) From the 5th percentile to the 95th percentile, or $[152.50, 172.90]$.

2. **(a)** From the 20th percentile to the 80th percentile, or $[23.85, 30.75]$.
(b) From the 10th percentile to the 90th percentile, or $[22.05, 32.55]$.

Exercises on Section 6.3.1

1. **(a)** 280 is $\frac{280-266}{16} = 0.875$ sd's above the mean.
(b) 250 is $\frac{250-266}{16} = -1$, i.e., 1 sd below the mean.
(c) 270 is $\frac{270-266}{16} = 0.25$ sd's above the mean.

2. **(a)** 80 is $\frac{80-100}{15} = -1.3333$, i.e., 1.3333 sd's below the mean.
(b) 110 is $\frac{110-100}{15} = 0.6667$ sd's above the mean.
(c) 90 is $\frac{90-100}{15} = -0.6667$, i.e., 0.6667 sd's below the mean.

Exercises on Section 6.3.2

1. **(a)** $z = 1.9600$, $[\mu - z\sigma, \mu + z\sigma] = [234.6, 297.4]$. $z = 1.2816$, $[\mu - z\sigma, \mu + z\sigma] = [245.5, 286.5]$.

(b) $z = 1.6449$, $\mu + z\sigma = 292.3$.
$z = 2.3263$, $\mu + z\sigma = 303.2$.

(c) $z = 1.6449$, $\mu - z\sigma = 239.7$.
$z = 2.3263$, $\mu - z\sigma = 228.8$.

2. **(a)** $z = 1.96$, $[\mu - z\sigma, \mu + z\sigma] = [71, 129]$.
$z = 1.2816$, $[\mu - z\sigma, \mu + z\sigma] = [81, 119]$.

(b) $z = 1.6449$, $\mu + z\sigma = 125$.
$z = 2.3263$, $\mu + z\sigma = 135$.

(c) $z = 0.6745$, $\mu - z\sigma = 90$.
$z = 1.2816$, $\mu - z\sigma = 82$.

Exercises on Section 6.3.3

1. **(a)** **(i)** $\frac{-5-3}{4} = -2$. **(ii)** $\frac{11-3}{4} = 2$. **(iii)** $\frac{5-3}{4} = 0.5$. **(iv)** $\frac{1.4-3}{4} = -0.4$.

(b) **(i)** 2 sd's below. **(ii)** 2 sd's above. **(iii)** 0.5 sd's above. **(iv)** 0.4 sd's below.

2. **(a)** $\text{pr}(Z \geq \frac{5-7}{6} = -0.33) = 1 - \text{pr}(Z \leq -0.33) = 1 - 0.371 = 0.629$.

(b) $\text{pr}(Z \leq \frac{9-7}{6} = 0.33) = 0.629$.

(c) $\text{pr}(-0.33 = \frac{5-7}{6} \leq Z \leq \frac{11-7}{6} = 0.67) = \text{pr}(Z \leq 0.67) - \text{pr}(Z \leq -0.33) = 0.749 - 0.371 = 0.378$.

(d) $\text{pr}(-0.67 = \frac{3-7}{6} \leq Z \leq \frac{6-7}{6} = -0.17) = \text{pr}(Z \leq -0.17) - \text{pr}(Z \leq -0.67) = 0.433 - 0.251 = 0.182$.

(d) $\text{pr}(Z \leq \frac{3-7}{6} = -0.67) = 0.251$.

3. **(a)** $\text{pr}(Z \leq \frac{-4+3}{2} = -0.5) = 0.309$.

(b) $\text{pr}(Z \geq \frac{3}{2} = 1.5) = 1 - \text{pr}(Z \leq 1.5 = 1 - 0.933 = 0.067$.

(b) $\text{pr}(0 = \frac{-3+3}{2} \leq Z \leq \frac{-1+3}{2} = 1) = \text{pr}(Z \leq 1) - \text{pr}(Z \leq 0) = 0.841 - 0.5 = 0.341$.

Exercises on Section 6.4.3

1. **(a)** $\text{E}(Y) = 2.5 + 1.5 = 4.0$, $\text{sd}(X) = \sqrt{3^2 + 2^2} = 3.61$.

(b) $\text{E}(Y) = 2.5 - 1.5 = 1.0$, $\text{sd}(X) = \sqrt{3^2 + 2^2} = 3.61$.

(c) $\text{E}(Y) = 2.5 + 5 - 4 = 3.5$, $\text{sd}(X) = \sqrt{3^2 + 5^2 + 3^2} = 6.56$.

(d) $\text{E}(Y) = 1.5 + 5 = 6.5$, $\text{sd}(X) = \sqrt{2^2 + 5^2} = 5.39$.

(e) $\text{E}(Y) = 1.5 - 5 = -3.5$, $\text{sd}(X) = \sqrt{2^2 + 5^2} = 5.39$.

(f) $\text{E}(Y) = 2.5 + 1.5 - 5 = -1.0$, $\text{sd}(X) = \sqrt{3^2 + 2^2 + 5^2} = 6.16$.

(g) $E(Y) = 5 - 4 = 1.0$, $\text{sd}(X) = \sqrt{5^2 + 3^2} = 5.83$.

(h) $E(Y) = 5 - (-4) = 9.0$, $\text{sd}(X) = \sqrt{5^2 + 3^2} = 5.83$.

(i) $E(Y) = 2.5 + 5 - (-4) = 11.5$, $\text{sd}(X) = \sqrt{3^2 + 5^2 + 3^2} = 6.56$.

(j)

Variable	Mean		Standard deviation	
$W_1 = 2X_1$	2×2.5	$= 5.0$	2×3	$= 6$
$W_2 = 3X_2$	3×1.5	$= 4.5$	3×2	$= 6$
$Y = W_1 + W_2$	$5.0 + 4.5$	$= 9.5$	$\sqrt{6^2 + 6^2}$	$= 8.49$

(k)

Variable	Mean		Standard deviation	
$W_1 = 5X_1$	5×2.5	$= 12.5$	5×3	$= 15$
$W_2 = 4X_2$	4×1.5	$= 6.0$	4×2	$= 8$
$Y = W_1 - W_2$	$12.5 - 6.0$	$= 6.5$	$\sqrt{15^2 + 8^2}$	$= 17$

(l)

Variable	Mean		Standard deviation	
$W_1 = 2X_1$	2×2.5	$= 5.0$	2×3	$= 6$
$W_2 = 3X_2$	3×1.5	$= 4.5$	3×2	$= 6$
$W_3 = 4X_3$	4×5	$= 20.0$	4×5	$= 20$
$Y = W_1 + W_2 + W_3$	$5.0 + 4.5 + 20.0$	$= 29.5$	$\sqrt{6^2 + 6^2 + 20^2}$	$= 21.73.$

2. Let D be the score for a depressed child, and N be the score for a child not depressed. We have $D \sim \text{Normal}(\mu = 11.2, \sigma = 6.8)$ and $N \sim \text{Normal}(\mu = 8.5, \sigma = 7.8)$.

(a) We require $\text{pr}(D < N)$ or $\text{pr}(N - D \geq 0)$. Now $X = N - D \sim \text{Normal}(\mu = 8.5 - 11.2, \sigma = \sqrt{7.8^2 + 6.8^2})$, i.e., $X \sim \text{Normal}(\mu = -2.7, \sigma = 10.3480)$. Thus, $\text{pr}(X > 0) = 1 - \text{pr}(X \leq 0) = 0.3971$.

(b) Some of them may be depressed or have a high CDI score.

Review Exercises 6

Note: All of the probabilities and "inverse" values required in these review exercises were obtained directly using a computer.

1. (a) (i) $\text{pr}(X > 141) = 1 - \text{pr}(X \leq 141) = 0.06681$.

(ii) $\text{pr}(120 < X < 132) = \text{pr}(X < 132) - \text{pr}(X < 120) = 0.4515$.

(iii) $\text{pr}(X < 118.5) = 0.2266$.

(b) We have $z = -0.6$ which tells us that 120 is 0.6 sd's below the mean.

(c) We have $z = 1.4$ which tells us that 140 is 1.4 sd's above the mean.

(d) We need the 85th percentile, i.e., a so that $0.85 = \text{pr}(X \leq a)$. We find $a = 136.4$.

(e) We need the 90th percentile, i.e., b so that $0.9 = \text{pr}(X \leq b)$, We find $b = 138.8$.

(f) We need the 1st percentile, i.e., c so that $0.01 = \text{pr}(X \leq c)$, We find $c = 102.7$.

(g) From the 5th to the 95th percentile, or $[109.6, 142.4]$.

(h) From the 20th to the 80th percentile, or $[117.6, 134.4]$.

(i) $\text{Total} = T \sim \text{Normal}(\mu = 20 \times 126, \sigma = \sqrt{20} \times 10)$, or $\text{Normal}(2520, 44.72)$.
$\overline{X} = \frac{T}{20} \sim \text{Normal}(\mu = 126, \sigma = \frac{10}{\sqrt{20}})$, or $\text{Normal}(126, 2.24)$.

2. **(a)** Let X = survival time, then $X \sim \text{Normal}(\mu = 31.1, \sigma = 16.0)$.

(i) $\text{pr}(X \leq 12) = 0.1163$.

(ii) $\text{pr}(12 \leq X \leq 24) = \text{pr}(X \leq 24) - \text{pr}(X \leq 12) = 0.2123$.

(iii) We need the 20th percentile, i.e., a so that $0.2 = \text{pr}(X \leq a)$. We find that $a = 17.63$.

(iv) From the 10th to the 90th percentile, or $[10.60, 51.60]$.

(b) `Units 1|0=10`

```
0 | 1159
1 | 0347889
2 | 12555679
3 | 6899
4 | 0113445669
5 | 00449
```

(c) Yes. The plot looks bimodal and not bell shaped.

3. Let X be the distance reached. Then $X \sim \text{Normal}(\mu = 125, \sigma = 10)$.

(a) $\text{pr}(X \geq 120) = 1 - \text{pr}(X < 120) = 0.6915$.

(b) We need the 5th percentile, i.e., x so that $\text{pr}(X \leq x) = 0.05$. We find that $x = 108.55$, i.e., about 109 cm.

(c) The pilot has a reach which is 1.5 sd's above the mean reach, namely $125 + 1.5 \times 10 = 140$ cm.

4. Let X be the amount of dye discharged. Then $X \sim \text{Normal}(\mu, \sigma = 0.40)$.
We are told $\text{pr}(X \geq 6) = 0.01$. Equivalently, 6 is the 99th percentile of the distribution. The 99th percentile of the standard Normal is $z = 2.3263$. Thus, 6 must be 2.3263 standard deviations above the mean, i.e., $6 = \mu + 2.3263 \times 0.40$. Solving for μ we obtain $\mu = 6 - z \times 0.40 = 5.069$.

5. Here $X \sim \text{Normal}(\mu_X = 266, \sigma_X = 16)$.

(a) $\text{pr}(X > 349) = 1 - \text{pr}(X \leq 349) = 1.0657 \times 10^{-7}$.

(b) Looks like a wrong decision! Gestation periods are virtually never as long as 349 days.

(c) Let Y be the number of gestation periods lasting 349 or more days.
Then $Y \sim \text{Binomial}(n = 10{,}000{,}000, p = 1.0657 \times 10^{-7})$
and $\text{pr}(Y \geq 1) = 1 - \text{pr}(Y = 0) = 0.6555$. It reasonably likely that this would happen in at least one of 10 million births.
[We would put very little trust in these calculations. The weakest point is (a). It relies critically on a Normal approximation to the unknown true distribution of gestation periods working well in the far tail of the distribution; this approach is highly suspect.]

6. Let X_H and X_D denote the fasting glucose levels for healthy and diabetic people respectively. Then $X_H \sim \text{Normal}(\mu_H = 5.31, \sigma_H = 0.58)$, and $X_D \sim \text{Normal}(\mu_D = 11.74, \sigma_D = 3.50)$.

(a) The probability that a diabetic is correctly diagnosed as having diabetes is given by $\text{pr}(X_D > 6.5) = 1 - \text{pr}(X_D \leq 6.5) = 0.9328$.

(b) The probability that a non-diabetic is correctly diagnosed as not having diabetes is given by $\text{pr}(X_H < 6.5) = 0.9799$.

(c) **(i)** The sensitivity of the test at $C = 5.7$ is given by $\text{pr}(X_D > 5.7) = 1 - \text{pr}(X_D \leq 5.7) = 0.9578$.

(ii) The specificity of the test at $C = 5.7$ is given by $\text{pr}(X_H < 5.7) = 0.7493$.

(d) We need the 2nd percentile, i.e., c so that $\text{pr}(X_D \leq c) = 0.02$. We find that $c = 4.552$. The specificity at $c = 4.552$ is given by $\text{pr}(X_H < 4.55) = 0.0956$.

7. Let X_H be the serum acid of a randomly chosen healthy person.
Let X_G be the serum acid of a randomly chosen gout sufferer.
Then $X_H \sim \text{Normal}(\mu = 5.0, \sigma = 1)$, and $X_G \sim \text{Normal}(\mu = 8.5, \sigma = 1)$.

(a) $\text{pr}(X_G < 6.75) = 0.0401$.

(b) $\text{pr}(X_H > 6.75) = 0.0401$.

(c) $X \sim \text{Binomial}(n = 50, p = 0.0401)$.

(d) Let u be the new cutoff level for serum uric acid. Then $\text{pr}(X_G > u) = 0.90$ and $u = 7.2184$.

(e) $\text{pr}(X_H > 7.2184) = 0.0133$.

8. Let W be the weight of a fish, then $W \sim \text{Normal}(\mu = 1.3, \sigma = 0.4)$.

(a) $\text{pr}(W < 0.5) = 0.02275$.

(b) $\text{pr}(0.5 \leq W \leq 2.06) = \text{pr}(W \leq 2.06) - \text{pr}(W < 0.5) = 0.9485$.

(c) We need the 95th percentile, i.e., w so that $\text{pr}(W \leq w) = 0.95$. We find that $w = 1.958\,\text{kg}$.

(d) $Y \sim \text{Binomial}(n = 900, p = 0.02275)$. [The value of p is our answer from (a).]

(e) $\text{pr}(25 \leq Y \leq 35) = \text{pr}(Y \leq 35) - \text{pr}(Y < 25) = 0.1810$.

9. **(a)** We use the $\text{Normal}(\mu = 106, \sigma = 5)$ distribution. Then, $\text{pr}(X \leq 100) = 0.1151$, or 11.5%.

(b) $Y \sim \text{Binomial}(n = 100, p = 0.1151)$. [The value of p is our answer from (a).]

(c) Here Y is the number sold at discount giving a profit of $(65 - 25)Y$. The remainding $100 - Y$ are sold at full price giving a profit of $(70 - 26.5) \times (100 - Y)$. The total profit is $P = (65 - 25)Y + (70 - 26.5)(100 - Y) = 4350 - 3.5Y$.
$E(P) = 4350 - 3.5E(Y) = 4350 - 3.5 \times 11.51 = 4310$, or \$43.10.

***(d)** $\text{pr}(X < 95 \,|\, X < 100) = \frac{\text{pr}(X<95 \text{ and } X<100)}{\text{pr}(X<100)} = \frac{\text{pr}(X<95)}{\text{pr}(X<100)} = \frac{0.01390}{0.1151} = 0.1208$, or 12.1%.

10. Let X_D be the questionnaire score for a randomly chosen "delinquent" and Let X_N be the questionnaire score for a randomly chosen "nondelinquent".

(a) pr(*A nondelinquent is classified as delinquent*) $= \text{pr}(X_N > 75) = 0.06681$, using $X_N \sim \text{Normal}(\mu_N = 60, \sigma_N = 10)$.

(b) pr(*A delinquent is classified as nondelinquent*) $= \text{pr}(X_D < 75) = 0.1587$, using $X_D \sim \text{Normal}(\mu = 80, \sigma_D = 5)$.

(c) We construct the following two-way table.

	Classified as		
Reality	Delinquent	Nondelinquent	Total
Delinquent		0.1587×0.05	0.05
Nondelinquent	0.0668×0.95		0.95
Total			1.0

We do not need to go any further. Misclassification corresponds to falling into one of the two cells of the table that we have already filled in.

pr(*Misclassified*) $= 0.1587 \times 0.05 + 0.0668 \times 0.95 = 0.0714$.

11. Let F = length of randomly chosen female and M = the length of a randomly chosen male. Assume $F \sim \text{Normal}(\mu_F = 89.2, \sigma_F = 6.6)$ and $M \sim \text{Normal}(\mu_M = 92.0, \sigma_M = 6.7)$. Using these two distributions we find the following.

(a) We need the 90th percentile of the female distribution, i.e., we need a so that $\text{pr}(F \leq a) = 0.9$. We find that $a = 97.66\,\text{cm}$. Moreover, using the male distribution, we find that $\text{pr}(M \geq 97.66) = 1 - \text{pr}(M < 97.66) = 0.1992$.

(b) $\text{pr}(F > M) = \text{pr}(Y > 0)$ where $Y = F - M$. Now, $Y \sim \text{Normal}(\mu_Y = 89.2 - 92.0 = -2.8, \sigma_Y = \sqrt{6.6^2 + 6.7^2} = 9.4048)$ and $\text{pr}(Y \geq 0) = 1 - \text{pr}(Y < 0) = 0.3830$.

(c) $\text{pr}(M - F \geq 10) = \text{pr}(Y \leq -10) = 0.2220$.

(d) Human couples tend to be related to some extent in their physical dimensions rather than independent. For example, very tall women seldom partner with very short men. Perhaps similar dynamics affect coyotes.

12. Let T_M and T_E be the morning and evening travel times respectively. Then, $T_M \sim \text{Normal}(\mu_M = 31, \sigma_M = 3)$ and $T_E \sim \text{Normal}(\mu_E = 35.5, \sigma_E = 3.5)$.

(a) Let $T = T_M + T_E$ be one day's total travelling time. Then, $T \sim \text{Normal}(\mu_T = 31 + 35.5 = 66.5, \sigma_T = \sqrt{3^2 + 3.5^2} = 4.610)$. We need
$\text{pr}(T > 60) = 1 - \text{pr}(T \leq 60) = 0.9207$.

(b) $\text{pr}(T_M > T_E) = \text{pr}(D > 0)$ where $D = T_M - T_E \sim \text{Normal}(\mu_D = 31 - 35.5 = -4.5, \sigma_D = 4.61)$. Now, $\text{pr}(D > 0) = 1 - \text{pr}(D \leq 0) = 0.1645$.

(c) $\text{pr}(T_E \geq T_M + 5) = \text{pr}(T_E - T_M \geq 5) = \text{pr}(D \geq 5) = 1 - \text{pr}(D < 5) = 0.4568$.

(d) Morning total = $M_{Tot} \sim \text{Normal}(5 \times 31, \sqrt{5} \times 3)$, or $\text{Normal}(151, 6.708)$.
Evening total= $E_{Tot} \sim \text{Normal}(5 \times 35.5, \sqrt{5} \times 3.5)$, or $\text{Normal}(177.5, 7.826)$.
$M_{Tot} + E_{Tot} \sim \text{Normal}(151 + 177.5, \sqrt{5 \times 9 + 5 \times 12.25})$, or $\text{Normal}(328.5, 10.31)$.

(e) $T \sim \text{Normal}(\mu, \sigma = 3)$, but we do not know μ. We are told that 30 minutes is the 88th percentile of the distribution. The 88th percentile of the standard Normal distribution is $z = 1.1750$ so 30 must be 1.1750 standard deviations above μ, i.e., $30 = \mu + 1.1750 \times 3$. Solving for μ we obtain $\mu = 30 - 3 \times 1.175 = 26.475$. The mean time for morning trip, using the new route, is 26.5 minutes

13. (a) Let X be the weight of an adult. Then $X \sim \text{Normal}(\mu = 73, \sigma = 13)$. Then the distribution of the sum of the weights of 11 randomly sampled people is Normally distributed with mean $n\,\mu = 11 \times 73$ and standard deviation $\sqrt{N}\,\sigma = \sqrt{11} \times$

13), i.e., $Sum \sim \text{Normal}(803, 43.1161)$. Then $\text{pr}(Sum > 800) = 1 - \text{pr}(Sum \leq 800) = 0.5277$. We have assumed that the weights of adults satisfy a single Normal distribution, and that the weights are independent. Both assumptions are doubtful because of sex differences in weight, and friends or acquaintances often take a lift together.

Let F the weight of a random female, and M the weight of a random male. Then, $F \sim \text{Normal}(\mu_F = 68, \sigma_F = 12)$ and $M \sim \text{Normal}(\mu_M = 78, \sigma_M = 12)$.

(b) The distribution of the sum of the weights of 11 randomly sampled men, B, is Normal with $\mu_B = 11 \times 78 = 858$, and $\sigma_B = \sqrt{11} \times 12 = 39.7995$. Then $\text{pr}(B > 800) = 1 - \text{pr}(B \leq 800) = 0.9275$.

(c)

Variable	Mean		Standard deviation
S_{7M} = sum of wts of 7 men	7×78	$= 546$	$\sqrt{7} \times 12 = 31.74902$
S_{4W} = sum of wts of 4 women	4×68	$= 272$	$\sqrt{4} \times 12 = 24$
$C = S_{7M} + S_{4W}$	$546 + 272$	$= 818$	$\sqrt{31.74902^2 + 24^2}$ $= 39.7995$

Since $C \sim \text{Normal}(818, 39.7995)$, $\text{pr}(C > 800) = 1 - \text{pr}(C \leq 800) = 0.6745$.

(d) We need $\text{pr}(C > B) = \text{pr}(D > 0)$, where $D = C - B$. The distribution of D is Normal with mean $\mu_D = \mu_C - \mu_B = 818 - 858 = -40$ and standard deviation $\sigma_D = \sqrt{39.7995^2 + 39.7995^2} = 56.2857$. Then, $\text{pr}(D > 0) = 0.2386$.

***(e)** Assume they are all men and there are m of them with total weight T. Then, $T \sim \text{Normal}(m \times 78, \sqrt{m} \times 12)$. Using this distribution, we calculated the value of $\text{pr}(T > 800)$ for each of several values of m.

m	7	8	9	10	11
$\text{pr}(T > 800)$	6.7×10^{-16}	1.1×10^{-07}	0.0032	0.2991	0.9275

If we choose $m = 9$, there is still only a very small chance that 9 men will overload the elevator, whereas there is almost a 30% chance of an overload with 10 men. We recommend a limit of 9.

(f) University students will be lighter on average, as people tend to gain weight with age.

14. **(a)** $Y_1 = X_1 + X_2 + \ldots + X_8$, so that $\mu_1 = 8 \times 58 = 464$ and $\sigma_1 = \sqrt{8} \times 12 = 33.94$.

(b) $Y_2 = 8X_3$, so that $\mu_2 = 8 \times 58 = 464$ and $\sigma_2 = 8 \times 12 = 96$.

(c) $Y_3 = \frac{1}{8}(X_9 + X_{10} + \ldots + X_{17})$, so that $\mu_3 = 58$ and $\sigma_3 = \frac{12}{\sqrt{8}} = 4.24$

(d) We have assumed that the distribution for the whole class is well approximated by the Normal and that students observed are a random sample from the class. This latter would be reasonable if the ID-number order is a random order as far as test ability is concerned. (Can you think of any reasons why it might not be?)

15. Since the distribution of the weight of a single random ingot is $\text{Normal}(\mu_I = 500, \sigma_I = 10)$, the total weight of for $n = 100$ ingots is Normal with mean $100 \times 500\,\text{g} = 50\,\text{kg}$ and standard deviation $\sqrt{100} \times 10\,\text{g} = 0.1\,\text{kg}$, i.e., $Sum \sim \text{Normal}(50\,\text{kg}, 0.1\,\text{kg})$.

(a) $\text{pr}(49.9 < Sum < 50.1) = \text{pr}(Sum < 50.1) - \text{pr}(Sum \leq 49.9) = 0.6827$.

***(b)** Let X be the number of weighings. Then $\text{pr}(X = 1) = \text{pr}(Sum \geq 49.9) = 0.8413$, and $\text{pr}(X = 101) = \text{pr}(Sum < 49.9) = 0.1587$.
Hence $E(X) = \sum x \,\text{pr}(x) = 1 \times 0.8413 + 101 \times 0.1587 = 16.87$, or about 17.

16. $P_1 = 1000W \sim \text{Normal}(1000 \times 10, 1000 \times 0.5)$, i.e., $P_1 \sim \text{Normal}(10000, 500)$.
$P_2 = 500W+500V \sim \text{Normal}(500\times10+500\times10 = 10000, 500\times\sqrt{0.5^2 + 3^2} = 1520.69)$.
$P_3 = 250X+250Y+250V+250W \sim \text{Normal}(4\times250\times10, 250\times\sqrt{1 + 2^2 + 3^2 + 0.5^2})$, i.e., $\text{Normal}(10000, 943.73)$.

- **(a)** **(i)** $\text{pr}(P_1 > 12,000) = 3.2 \times 10^{-5}$.
  $\text{pr}(P_2 > 12,000) = 0.0942$.
  $\text{pr}(P_3 > 12,000) = 0.0170$.
  (ii) $\text{pr}(P_1 < 5000) \approx 0$.
  $\text{pr}(P_2 < 5000) = 0.0005$.
  $\text{pr}(P_3 < 5000) \approx 0$.
- **(b)** P_1 is the safest and P_2 is the riskiest.
- **(c)** P_2 gives the best chance of making 2500 dollars profit (i.e., achieving a total value of $12,500) because of its greater variability.

17. **(a)** **(i)** Since X is a sum from a sample of size $n = 50$ sheets,
$X \sim \text{Normal}(\mu = 50 \times .5, \sigma = \sqrt{50} \times .05)$, i.e., $\text{Normal}(25, 0.35355)$.
(ii) $\text{pr}(X > 26) = 0.00234$.

- **(b)** $Y \sim \text{Normal}(\mu = 49 \times 0.05, \sigma = 7 \times 0.02)$, i.e., $\text{Normal}(\mu = 2.45, \sigma = 0.14)$.
- **(c)** $W = X + Y \sim \text{Normal}(\mu = 25 + 2.45 = 27.45, \sigma = \sqrt{0.35355^2 + 0.14^2} = 0.3803)$ and $\text{pr}(W > 28.3) = 0.0127$.
- **(d)** **(i)** We no longer have a sum made up of independently varying thicknesses, but now have 49 copies of one random thickness, so $M \sim \text{Normal}(\mu = 49 \times .05, \sigma = 49 \times .02)$, i.e., $\text{Normal}(2.45, 0.98)$.
  (ii) $X + M \sim \text{Normal}(\mu = 27.45, \sqrt{0.35355^2 + 0.98^2} = 1.0418)$ and $\text{pr}(X + M > 28.3) = 0.2073$
- **(e)** We have essentially answered this in (d)(i). Y is the sum of 49 independently varying thicknesses, whereas M is made up of 49 copies of the same random thickness. Thus, M is much more variable than Y (see page 265 of the book).

18. **(a)** $A \sim \text{Normal}(\mu = 8.8, \sigma = 0.4)$. We need the 20th percentile for A, i.e., a so that $\text{pr}(A \leq a) = 0.2$. We find that $a = 8.463$.
$B \sim \text{Normal}(\mu = 4.1, \sigma = 0.8)$. We need the 80th percentile for B, i.e., b so that $\text{pr}(B \leq b) = 0.8$. We find that $b = 3.427$.

(b) **(i)** For A, we have:

Variable		Mean	Standard deviation	
10×Issue 1	10×6.8	$= 68.0$	10×0.2	$= 2.0$
5×Issue 2	5×5.5	$= 27.5$	5×0.2	$= 1.0$
2×Issue 3	2×8.8	$= 17.6$	2×0.4	$= 0.8$
Issue 4		$= 7.4$		$= 0.8$
Issue 5		$= 6.2$		$= 0.5$
Sum of the above	sum of means	$= 126.7$	$\sqrt{2^2 + 1^2 + 0.8^2 + 0.8^2 + 0.5^2}$	$= 2.5554$

Hence $I_A \sim \text{Normal}(126.7, 2.5554)$.

(ii) Similarly $I_B \sim \text{Normal}(\mu = 126, \sigma = 3.9256)$.

(iii) $\text{pr}(I_A > I_B) = \text{pr}(D > 0)$ where $D = I_A - I_B$ has a Normal distribution with $\mu_D = 126.7 - 126 = 0.7$ and $\sigma_D = \sqrt{2.5554^2 + 3.9256^2}) = 4.684$. Then, $\text{pr}(D > 0) = 0.5594$.

(c) Normality is questionable as each issue has only 10 possible scores. Issues are not likely be independent as they are each rated by the same person.
For checking Normality, you can use a bar graph (better methods are discussed in Chapter 10). Plot one issue versus another for each person to see whether the ratings are independent or seem to be related.

19. (a) Let C be the weight of a randomly chosen carton. Then $C \sim \text{Normal}(\mu_C = 100, \sigma_C = 1.25)$. A carton is strapped with probability $\text{pr}(C > 101) = 0.2119$.
If there are 64 cartons, then the number to be strapped is $\text{Binomial}(n = 64, p = 0.2119)$.

(b) Let W be the weight of a pallet. Then $W \sim \text{Normal}(\mu_W = 150, \sigma_W = 3)$.

Let S_C be the total weight (a sum) for 64 independent cartons. Then the distribution of S_C is Normal with mean $\mu_S = 64 \times 100 = 6400$ and $\sigma_S = \sqrt{64} \times 1.25 = 10$.

Let T be the total weight of a loaded pallet, then the distribution of $T = W + S_C$ is Normal with $\mu_T = 150 + 6400 = 6550$, and standard deviation
$\sigma_T = \sqrt{3^2 + 10^2} = 10.4403$.

We require the 5th percentile of the distribution of T, i.e., t such that $\text{pr}(T < t) = 0.05$. We find that $t = 6533\,\text{kg}$.

***20. (a)** $X_i \sim \text{Binomial}(n = 60, p)$.

(b) $\text{E}(X_i) = 60p$, $\text{sd}(X_i) = \sqrt{60p(1-p)}$.

(c) $\text{E}(Y) = 4 \times (\sum_{i=1}^{17} \text{E}(X_i) = 4 \times 17 \times 60p = 4080p$.
Yes, this is a reasonable measure of the underlying idleness of the car.

(d) By independence
$\text{sd}(Y) = 4 \times \text{sd}(\sum X_i) = 4 \times \sqrt{\sum \text{sd}(X_i)^2} = 4 \times \sqrt{17} \times \sqrt{60p(1-p)}$
$= 127.75\sqrt{p(1-p)}$.

(e) The natural estimate of p is $\frac{824}{17 \times 240} = 0.20196$.

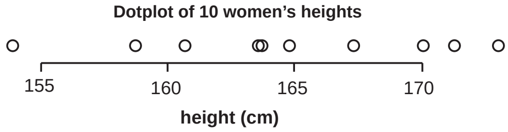
(f) An estimate of $\text{sd}(Y)$ is $127.75 \times \sqrt{0.20196 \times 0.79804}$, or 51.29.

(g) Let U be the number of cars that are idle on any day. Then $U \sim \text{Binomial}(n = 17, p \approx 0.202)$. Hence $\text{pr}(U \geq 1) = 1 - \text{pr}(U = 0) = 0.98$ and $\text{pr}(U \geq 2) = 1 - \text{pr}(U \leq 1) = 0.89$. We might then sell off either 1 or 2 cars, as the above probabilities are close to 1.

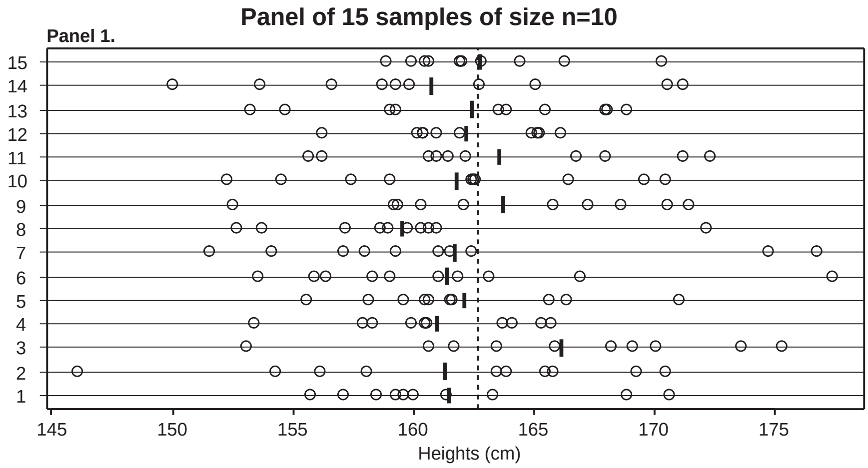
21. **(a)** The "heights of ten women" randomly sampled from a Normal distribution with $\mu = 162.7\,\text{cm}$ and $\sigma = 6.2\,\text{cm}$ are given below.

158.7860	171.3006	164.8145	170.0762	163.7403
173.0414	153.9394	163.6225	160.6812	167.3418

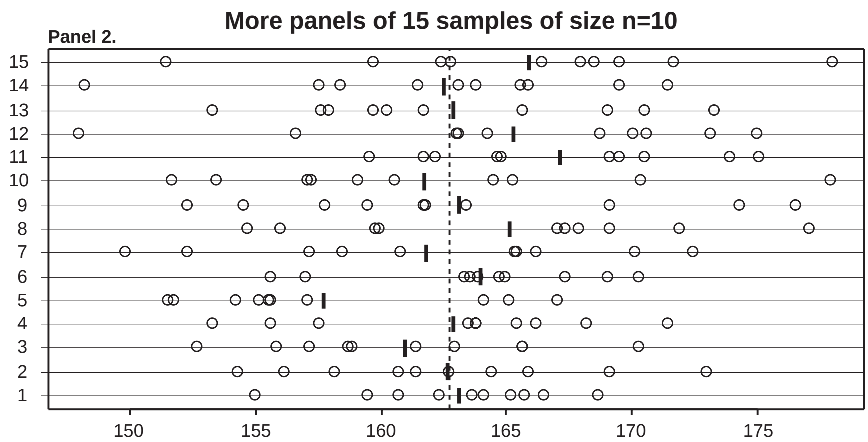
The corresponding dot plot of the 10 observations follows.

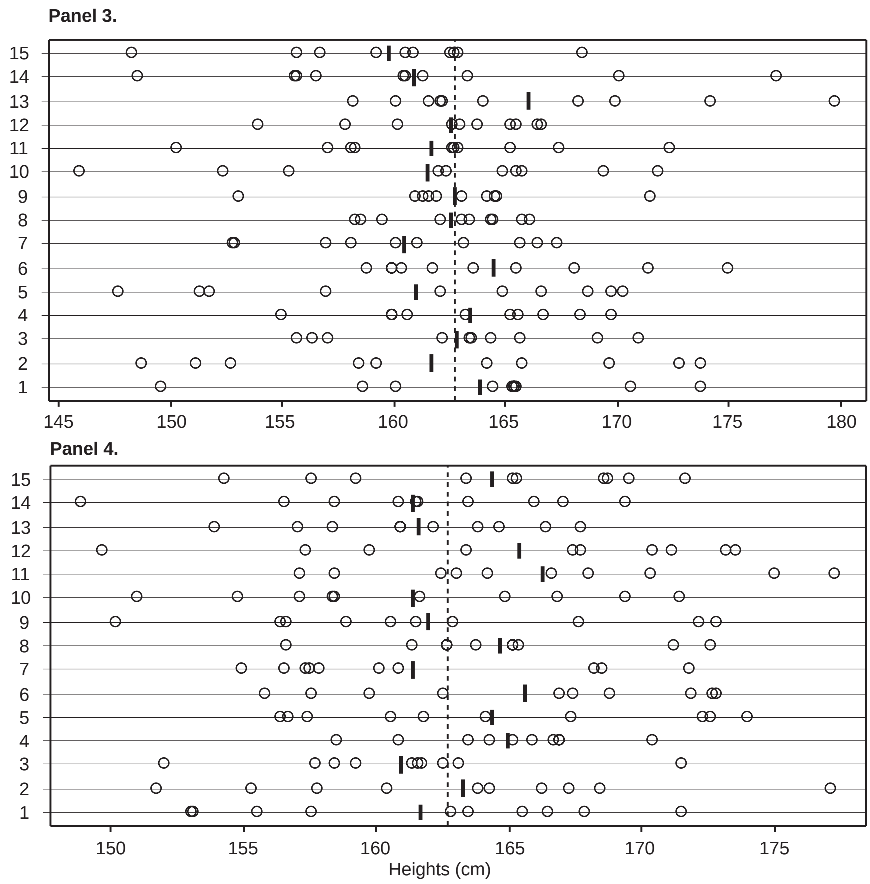


(b) A dot plot of fifteen samples each of ten women is given below. The dotted vertical line represents the population mean of 162.7cm and the sample means for each of the fifteen samples are represented by small vertical bars. A further three sets of 15 follow. Each line in a plot corresponds to a single sample.

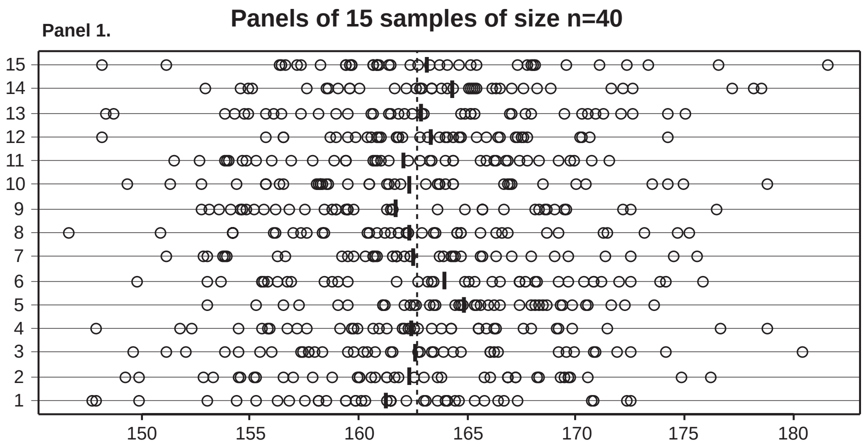


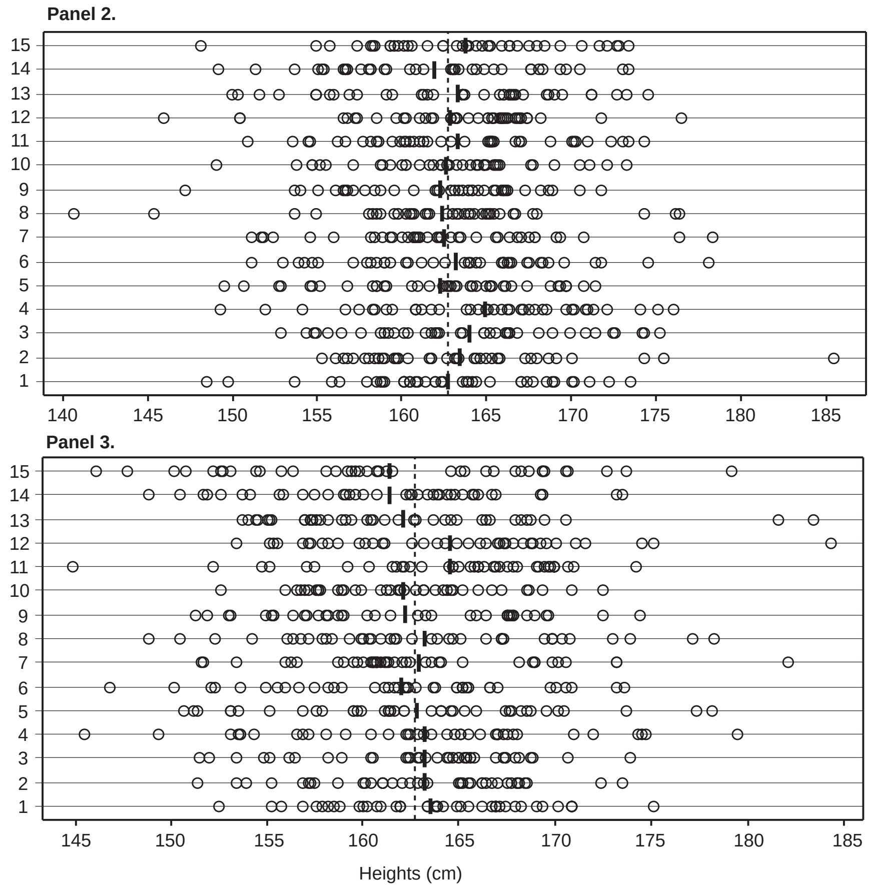
Differences in center are very easy to see as the centers are marked by the vertical bars and they vary greatly. For other features, we have picked out some of the more extreme examples. For extreme differences in spread, compare sample 1 and sample 14 or 15 in panel 2, or sample 8 in panel 3 with almost any other sample. There are many cases of extreme outliers, e.g. see on the right of sample 8 of panel 1. Sample 5 in panel 2 is one of the more extreme examples of a gap in the middle of a data set. Sample 13 in panel 3 looks rather right skewed while sample 12 looks quite left skewed. Remember what we are seeing is "perfect data" from a Normal random number generator. There are no mistakes here.



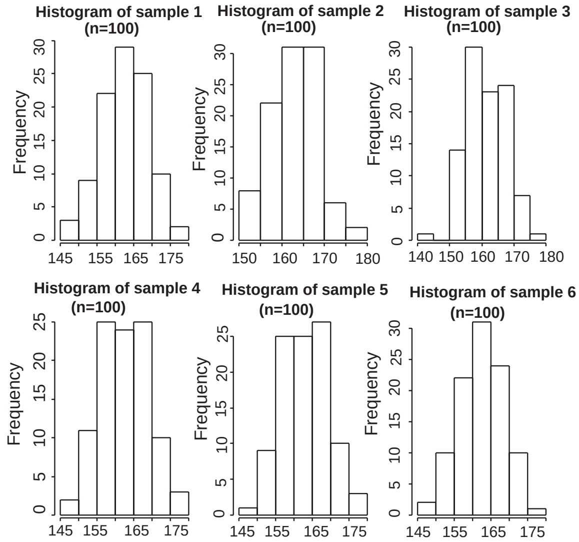


(c) Dot plots for three sets of fifteen samples each of 40 women follow. The variation in means is clearly much smaller. The "outliers" are seldom as extreme, there is less variation in spreads from sample to sample, and the gaps tend to be narrower.





(d) Histograms for 6 samples each of size 100 is given below. Athough the bulk of the observations are always falling within very much the same range and all are unimodal, at a more detailed level the shapes of the histograms vary considerably. The two top right ones even suggest skewness.



(e) Histograms for 6 samples each of size 1000 is given below. When we have batches of 1000 observations the histograms are more bell shaped and there is a little less variation in histogram shape from sample to sample, but the variation is still noticeable. We find later that if we take enormous samples, e.g. millions, we get reliably bell-shaped histograms that vary very little.