1.5 Solutions

1. $X = 38$ Heads in $n = 100$ tosses corresponds to a $p$-value = .021, which is less than $\alpha = .05$; hence in this case we are able to reject the null hypothesis, and conclude that the coin is not fair, at this significance level. However, $p = .021$ is greater than $\alpha = .01$; hence we are unable to reject the null hypothesis of fairness, at this significance level. We tentatively “accept” – or, at least, not outright reject – that the coin is fair, at this level. (The coin may indeed be biased, but this empirical evidence is not sufficient to show it.) Thus, lowering the significance level $\alpha$ at the outset means that based on the sample data, we will be able to reject the null hypothesis less often on average, resulting in a more conservative test.

2. (a) If the coin is known to be fair, then all $2^{10}$ outcomes are equally likely; the probability of any one of them occurring is the same (namely, $1/2^{10}$)!

(b) However, if the coin is not known to be fair, then Outcomes 1, 2, and 3 – each with $X = 5$ Heads and $n - X = 5$ Tails, regardless of the order in which they occur – all provide the best possible evidence in support of the hypothesis that the coin is unbiased. Outcome 4, with $X = 7$ Heads, is next. And finally, Outcome 5, with all $X = 10$ Heads, provides the worst possible evidence that the coin is fair.

3. The issue here is one of sample size, and statistical power – the ability to detect a significant difference from the expected value, if one exists. In this case, a total of $X = 18$ Heads out of $n = 50$ tosses yields a $p$-value = 0.0649, which is just above the $\alpha = .05$ significance level. Hence, the evidence in support of the hypothesis that the coin is fair is somewhat borderline. This suggests that perhaps the sample size of $n = 50$ may not be large enough to detect a genuine difference, even if there is one. If so, then a larger sample size might generate more statistical power. In this experiment, obtaining $X = 36$ Heads out of $n = 100$ tosses is indeed sufficient evidence to reject the hypothesis that the coin is fair.
4. **R exercise**

(a) If the population ages are uniformly distributed between 0 and 100 years, then via **symmetry**, the mean age would correspond to the midpoint, or **50 years**.

(b) The provided R code generates a random sample of \( n = 500 \) ages from a population between 0 and 100 years old. The R command `mean(my.sample)` should typically give a value fairly close to the population mean of 50 (but see part (d)).

(c) The histogram below is typical. The frequencies indicate the **number** of individuals in each age group of the sample, and correspond to the heights of the rectangles. In this sample, there are:

- 94 individuals between 0 and 20 years old, i.e., 18.8%,
- 98 individuals between 20 and 40 years old, i.e., 19.6%,
- 105 individuals between 40 and 60 years old, i.e., 21.0%,
- 100 individuals between 60 and 80 years old, i.e., 20.0%,
- 103 individuals between 80 and 100 years old, i.e. 20.6%.

**If the population is uniformly distributed, we would expect the sample frequencies to be about the same in each of the five intervals, and indeed, that is the case; we can see that each interval contains about one-hundred individuals (i.e., 20%).**

![Histogram of my.sample](image-url)
(d) **Most results should be generally similar to (b) and (c) – in particular, the sample means fairly close to the population mean of 50 – but there is a certain nontrivial amount of variability, due to the presence of “outliers.”** For example, if by chance a particular sample should consist of unusually many older individuals, it is quite possible that the mean age would be shifted to a value that is noticeably larger than 50. This is known as “skewed to the right” or “positive skew.” Similarly, a sample containing many younger individuals might be “skewed to the left” or “negatively skewed.”

(e) The histogram below displays a simulated distribution of the means of many (in this case, 2000) samples, each sample having \( n = 500 \) ages. Notice how much “tighter” (i.e., less variability) the graph is around 50, than any of those in (c). The reason is that it is much more common for a random sample to contain a relatively small number of outliers – whose contribution is “damped out” when all the ages are averaged – than for a random sample to contain a relatively large number of outliers – whose contribution is sizeable enough to skew the average. Thus, the histogram is rather “bell-shaped”; highly peaked around 50, but with “tails” that taper off left and right.
5. The following is typical output (“copy-and-paste”) directly from R. Comments are in blue.

(a) > prob = 0.5
    > tosses = rbinom(100, 1, prob)  This returns a random sequence of 100 single tosses.*
    > tosses # view the sequence
       [1] 1 1 1 1 0 1 1 0 1 0 1 1 1 1 0 0 0 1 0 0 0 1 0 1 0 0 0 1 0 1 1 0 1 1 1 0 1 0 1 1 1 0 1 0 1 1 0
       [38] 1 1 1 1 0 1 1 1 0 0 0 1 1 0 1 0 1 1 1 1 0 1 1 0 1 0 0 1 1 1 0 0 1 0 1 1 0
       [75] 1 1 1 1 0 0 0 1 1 1 1 0 1 1 0 0 1 1 1 1 0 1 0 1 1 0
    > sum(tosses) # count the number of Heads
       [1] 58

* Note: rbinom(1, 100, prob) just generates the number of Heads (not the actual sequence) in 1 run of 100 random tosses, in this case, 58.

This simulation of 100 random tosses of a fair coin produced 58 Heads. According to the chart on page 1-4, the corresponding \( p\) -value = 0.1332. That is, if the coin is fair (as here), then in 100 tosses, there is an expected 13.32% probability of obtaining 8 (or more) Heads away from 50. This is above the 5% significance level, hence consistent with the coin being fair. Had it been below (i.e., rarer than) 5%, it would have been inconsistent with the coin being fair, and we would be forced to conclude that the coin is indeed biased.  
Alas, in multiple runs, this would eventually happen just by chance!  
(See the outliers in the graphs below.)

(b) > X = rbinom(500, 100, prob)
This command generates the number of Heads in each of 500 runs of 100 tosses, as stated.
> sort(X) This command sorts the 500 numbers just found in increasing order (not shown).
> table(X)
Produces a frequency table for \( X = \# \) Heads, i.e., 35 Heads occurred twice, 36 twice, etc.
X
35 36 37 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 66
2  2  2  2  8  7  9 13 15 24 23 30 38 41 35 41 41 33 27 21 31 16  8 10  9  1  6  3  1  2  1

> summary(X) This is often referred to as the “five number summary”:
    Min.  1st Qu.  Median  Mean   3rd Qu.  Max.
    35.00  46.00  50.00  49.53  53.00  66.00

Notice that the mean \( \approx \) median (suggesting that this may be close to a more-or-less symmetric distribution; see page 2-14 in the notes) \( \approx 50 \), both of which you might expect to see in 100 tosses of an unbiased coin, as confirmed in the three graphs below.
(c) The sample proportions obtained from this experiment are quite close to the theoretical \( p \)-values we expect to see, if the coin is fair.

<table>
<thead>
<tr>
<th>lower</th>
<th>upper</th>
<th>prop</th>
<th>( p )-values (from chart)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>49</td>
<td>51</td>
<td>0.918</td>
</tr>
<tr>
<td>[2,]</td>
<td>48</td>
<td>52</td>
<td>0.766</td>
</tr>
<tr>
<td>[3,]</td>
<td>47</td>
<td>53</td>
<td>0.618</td>
</tr>
<tr>
<td>[4,]</td>
<td>46</td>
<td>54</td>
<td>0.488</td>
</tr>
<tr>
<td>[5,]</td>
<td>45</td>
<td>55</td>
<td>0.386</td>
</tr>
<tr>
<td>[6,]</td>
<td>44</td>
<td>56</td>
<td>0.278</td>
</tr>
<tr>
<td>[7,]</td>
<td>43</td>
<td>57</td>
<td>0.198</td>
</tr>
<tr>
<td>[8,]</td>
<td>42</td>
<td>58</td>
<td>0.152</td>
</tr>
<tr>
<td>[9,]</td>
<td>41</td>
<td>59</td>
<td>0.106</td>
</tr>
<tr>
<td>[10,]</td>
<td>40</td>
<td>60</td>
<td>0.070</td>
</tr>
<tr>
<td>[11,]</td>
<td>39</td>
<td>61</td>
<td>0.054</td>
</tr>
<tr>
<td>[12,]</td>
<td>38</td>
<td>62</td>
<td>0.026</td>
</tr>
<tr>
<td>[13,]</td>
<td>37</td>
<td>63</td>
<td>0.020</td>
</tr>
<tr>
<td>[14,]</td>
<td>36</td>
<td>64</td>
<td>0.014</td>
</tr>
<tr>
<td>[15,]</td>
<td>35</td>
<td>65</td>
<td>0.006</td>
</tr>
<tr>
<td>[16,]</td>
<td>34</td>
<td>66</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Since these values are comparable, it seems that we have reasonably strong confirmation that the coin is indeed unbiased.

From this point on, all proportions are 0.
(d) > prob = runif(1, min = 0, max = 1)  \text{This selects a random probability for Heads.}\n> \text{tosses <- rbinom(100, 1, prob)}\n> \text{sum(tosses) \# count the number of Heads}
[1] 62

This simulation of 100 random tosses of a fair coin produced 62 Heads, which corresponds to a \textit{p-value} = .021 < .05. Hence, based on this sample evidence, we may \textbf{reject} the hypothesis that the coin is fair; the result is \textbf{statistically significant} at the \(\alpha = .05\) level. Graphs are similar to above, centered about the mean (see below).

> \text{table(X)}

\begin{verbatim}
X
46 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 77 78 79 80
1 2 1 1 6 6 6 5 12 16 22 27 28 44 50 42 45 39 29 18 14 16 15 4 4 11 2 1 1 1
\end{verbatim}

> \text{summary(X)}

\begin{verbatim}
Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
46.00 61.00 64.00 64.31 67.00 80.00
\end{verbatim}

According to these data, the mean number of Heads is 64.31 out of 100 tosses; hence the \textit{estimated} probability of Heads is 0.6431. The actual probability that R used here is

> prob
[1] \textbf{0.6412175}
6. (a) 

<table>
<thead>
<tr>
<th>$X$ = Sum</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$1/36 = 0.02778$</td>
</tr>
<tr>
<td>3</td>
<td>$2/36 = 0.05556$</td>
</tr>
<tr>
<td>4</td>
<td>$3/36 = 0.08333$</td>
</tr>
<tr>
<td>5</td>
<td>$4/36 = 0.01111$</td>
</tr>
<tr>
<td>6</td>
<td>$5/36 = 0.01389$</td>
</tr>
<tr>
<td>7</td>
<td>$6/36 = 0.01667$</td>
</tr>
<tr>
<td>8</td>
<td>$5/36 = 0.01389$</td>
</tr>
<tr>
<td>9</td>
<td>$4/36 = 0.01111$</td>
</tr>
<tr>
<td>10</td>
<td>$3/36 = 0.08333$</td>
</tr>
<tr>
<td>11</td>
<td>$2/36 = 0.05556$</td>
</tr>
<tr>
<td>12</td>
<td>$1/36 = 0.02778$</td>
</tr>
</tbody>
</table>

(b)  
- $P(2 \leq X \leq 12) = 1$, because the event $2 \leq X \leq 12$ comprises the entire sample space.
- $P(2 \leq X \leq 6 \text{ or } 8 \leq X \leq 12) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 8) + P(X = 9) + P(X = 10) + P(X = 11) + P(X = 12),$  
  or, $1 - P(X = 7) = 1 - 6/36 = 30/36 = 0.83333$  
  Likewise,  
- $P(2 \leq X \leq 5 \text{ or } 9 \leq X \leq 12) = 20/36 = 0.55556$  
- $P(2 \leq X \leq 4 \text{ or } 10 \leq X \leq 12) = 12/36 = 0.33333$  
- $P(2 \leq X \leq 3 \text{ or } 11 \leq X \leq 12) = 6/36 = 0.16667$  
- $P(X \leq 2 \text{ or } X \geq 12) = 2/36 = 0.05556$  
- $P(X \leq 1 \text{ or } X \geq 13) = 0$, because neither the event $X \leq 1$ nor $X \geq 13$ can occur.
7. Absolutely not. That both sets of measurements average to 50.0 grams indicate that they have the same accuracy, but \textbf{Scale A has much less variability} in its readings that Scale B, so it has \textbf{much greater precision}. This experiment suggests that if many more measurements were taken, those of A would show a much higher density of them centered around 50 g than B, whose distribution of values would show much more spread around 50 g. Variability determines reliability, a major factor in quality control of services and manufactured products.

![Graph showing distribution of measurements for A and B](image)

- Measurements obtained from the A distribution are much more tightly clustered around their center, than those of the B distribution.