Section 6.3 Order Statistics

A q–q plot of these data is shown in Figure 6.3-2. Note that the points do fall close to a straight line, so the normal probability model seems to be appropriate on the basis of these few data.

Exercises

6.3-1. Some biology students were interested in analyzing the amount of time that bees spend gathering nectar in flower patches. Thirty-nine bees visited a high-density flower patch and spent the following times (in seconds) gathering nectar:

\[
\begin{array}{cccccccccccc}
235 & 210 & 95 & 146 & 195 & 840 & 185 & 610 & 680 & 990 \\
146 & 404 & 119 & 47 & 9 & 4 & 10 & 169 & 270 & 95 \\
329 & 151 & 211 & 127 & 154 & 35 & 225 & 140 & 158 & 116 \\
46 & 113 & 149 & 420 & 120 & 45 & 10 & 18 & 105 \\
\end{array}
\]

(a) Find the order statistics.
(b) Find the median and 80th percentile of the sample.
(c) Determine the first and third quartiles (i.e., 25th and 75th percentiles) of the sample.

6.3-2. Let \( X \) equal the forced vital capacity (the volume of air a person can expel from his or her lungs) of a male freshman. Seventeen observations of \( X \), which have been ordered, are

\[
\begin{array}{cccccccc}
3.7 & 3.8 & 4.0 & 4.3 & 4.7 & 4.8 & 4.9 & 5.0 \\
5.2 & 5.4 & 5.6 & 5.6 & 5.6 & 5.7 & 6.2 & 6.8 & 7.6 \\
\end{array}
\]

(a) Find the median, the first quartile, and the third quartile.
(b) Find the 35th and 65th percentiles.

6.3-3. Let \( Y_1 < Y_2 < Y_3 < Y_4 < Y_5 \) be the order statistics of five independent observations from an exponential distribution that has a mean of \( \theta = 3 \).

(a) Find the pdf of the sample median \( Y_3 \).
(b) Compute the probability that \( Y_4 \) is less than 5.
(c) Determine \( P(Y_3 < 1) \).

6.3-4. In the expression for \( g_r(x) = G_r(x) \) in Equation 6.3-1, let \( n = 6 \), and \( r = 3 \), and write out the summations, showing that the “telescoping” suggested in the text is achieved.

6.3-5. Let \( Y_1 < Y_2 < \cdots < Y_9 \) be the order statistics of eight independent observations from a continuous-type distribution with 70th percentile \( \pi_{0.7} = 27.5 \).

(a) Determine \( P(Y_7 < 27.3) \).
(b) Find \( P(Y_5 < 27.3 < Y_8) \).

6.3-6. Let \( W_1 < W_2 < \cdots < W_n \) be the order statistics of \( n \) independent observations from a \( U(0, 1) \) distribution.

(a) Find the pdf of \( W_1 \) and that of \( W_n \).
(b) Use the results of (a) to verify that \( E(W_1) = 1/(n+1) \) and \( E(W_n) = n/(n+1) \).
(c) Show that the pdf of \( W_r \) is beta.

6.3-7. Let \( Y_1 < Y_2 < \cdots < Y_{19} \) be the order statistics of \( n = 19 \) independent observations from the exponential distribution with mean \( \theta \).
256 Chapter 6 Point Estimation

(a) What is the pdf of $Y_1$?

(b) Using integration, find the value of $E[F(Y_1)]$, where $F$ is the cdf of the exponential distribution.

6.3-8. Let $W_1 < W_2 < \ldots < W_n$ be the order statistics of $n$ independent observations from a $U(0,1)$ distribution.

(a) Show that $E(W_i^2) = r(r + 1)/(n + 1)(n + 2)$, using a technique similar to that used in determining that $E(W_i) = r/(n + 1)$.

(b) Find the variance of $W_r$.

6.3-9. Let $Y_1 < Y_2 < \ldots < Y_n$ be the order statistics of a random sample of size $n$ from an exponential distribution with pdf $f(x) = e^{-x}$, $0 < x < \infty$.

(a) Find the pdf of $Y_r$.

(b) Determine the pdf of $U = e^{-Y_r}$.

6.3-10. Use the heuristic argument to show that the joint pdf of the two order statistics $Y_i < Y_j$ is

$$g(y_i, y_j) = \frac{n!}{(i - 1)!(j - i - 1)!(n - j)!} \times [F(y_i)]^{i-1} [F(y_j) - F(y_i)]^{j-1} \times [1 - F(y_i)]^{n-j} f(y_i)f(y_j), \quad -\infty < y_i < y_j < \infty.$$  

6.3-11. Use the result of Exercise 6.3-10.

(a) Find the joint pdf of $Y_1$ and $Y_n$, the first and the $n$th order statistics of a random sample of size $n$ from the $U(0,1)$ distribution.

(b) Find the joint and the marginal pdfs of $W_1 = Y_1/Y_n$ and $W_2 = Y_n$.

(c) Are $W_1$ and $W_2$ independent?

(d) Use simulation to confirm your theoretical results.

6.3-12. Nine measurements are taken on the strength of a certain metal. In order, they are 7.2, 8.9, 9.7, 10.5, 10.9, 11.7, 12.9, 13.9, 15.3, and these values correspond to the 10th, 20th, \ldots, 90th percentiles of this sample. Construct a $q-q$ plot of the measurements against the same percentiles of $N(0,1)$. Does it seem reasonable that the underlying distribution of strengths could be normal?

6.3-13. Some measurements (in mm) were made on specimens of the spider *Sosippus floridanus*, which is native to Florida. Here are the lengths of nine female spiders and nine male spiders.

<table>
<thead>
<tr>
<th>Female spiders</th>
<th>11.06</th>
<th>13.87</th>
<th>12.93</th>
<th>15.08</th>
<th>17.82</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14.14</td>
<td>12.26</td>
<td>17.82</td>
<td>20.17</td>
<td></td>
</tr>
<tr>
<td>Male spiders</td>
<td>12.26</td>
<td>11.66</td>
<td>12.53</td>
<td>13.00</td>
<td>11.79</td>
</tr>
<tr>
<td></td>
<td>12.46</td>
<td>10.65</td>
<td>10.39</td>
<td>12.26</td>
<td></td>
</tr>
</tbody>
</table>

(a) Construct a $q-q$ plot of the female spider lengths. Do they appear to be normally distributed?

(b) Construct a $q-q$ plot of the male spider lengths. Do they appear to be normally distributed?

6.3-14. An interior automotive supplier places several electrical wires in a harness. A pull test measures the force required to pull spliced wires apart. A customer requires that each wire that is spliced into the harness withstand a pull force of 20 pounds. Let $X$ equal the pull force required to pull a spliced wire apart. The following data give the values of a random sample of $n = 20$ observations of $X$:

| 28.8 | 24.4 | 30.1 | 25.6 | 26.4 | 23.9 | 22.1 | 22.5 | 27.6 | 28.1 |
| 20.8 | 27.7 | 24.4 | 25.1 | 24.6 | 26.3 | 28.2 | 22.2 | 26.3 | 24.4 |

(a) Construct a $q-q$ plot, using the ordered array and the corresponding quantiles of $N(0,1)$.

(b) Does $X$ appear to have a normal distribution?

6.4 MAXIMUM LIKELIHOOD ESTIMATION

In earlier chapters, we alluded to estimating characteristics of the distribution from the corresponding ones of the sample, hoping that the latter would be reasonably close to the former. For example, the sample mean $\bar{x}$ can be thought of as an estimate of the distribution mean $\mu$, and the sample variance $s^2$ can be used as an estimate of the distribution variance $\sigma^2$. Even the relative frequency histogram associated with a sample can be taken as an estimate of the pdf of the underlying distribution. But how good are these estimates? What makes an estimate good? Can we say anything about the closeness of an estimate to an unknown parameter?

In this section, we consider random variables for which the functional form of the pmf or pdf is known, but the distribution depends on an unknown parameter (say, $\theta$) that may have any value in a set (say, $\Omega$) called the parameter space. For example, perhaps it is known that $f(x; \theta) = (1/\theta) e^{-x/\theta}$, $0 < x < \infty$, and that $\theta \in \Omega = \{\theta : 0 < \theta < \infty\}$. In certain instances, it might be necessary for the experimenter to