Solutions to Problems in Chapter 6 of
Simulation Modeling and Analysis, 5th ed., 2015, McGraw-Hill, New York
by Averill M. Law

6.1. By the central limit theorem, $\overline{X}(n)$ and $\sum_{i=1}^{n} X_i$ are approximately normally distributed if n is "large."

6.2. Let $X \sim \text{Weibull}(\alpha, \beta)$. From Sec. 8.3.5 the generator for X is given by

$$X = \beta\left(-\ln U\right)^{1/\alpha}$$

If we raise each side of this equation to the α power, then we get

$$X^{\alpha} = -\beta^{\alpha} \ln U$$

which is the generator for an exponential distribution with scale parameter β^{α} (see Sec. 8.3.2).

6.4. Let $X \sim \text{Pareto}(c, \alpha_2)$. Then it can be shown that

$$F(x) = 1 - \left(\frac{c}{x}\right)^{\alpha_2}$$

Setting $U = F(X)$ and solving for X (see Sec. 8.2.1), we get

$$X = \frac{c}{(1-U)^{1/\alpha_2}}$$

Thus,

$$\ln X = \ln c - \frac{1}{\alpha_2}\ln(1-U)$$

which is the generator for an exponential distribution with location parameter $\ln c$ and scale parameter $1/\alpha_2$ (see Secs. 6.2.1 and 8.3.2).

6.5. Assume for definiteness that X is nonnegative. Suppose that $F(x)$ is defined by

$$F(x)=\begin{cases}0 & \text{if } x < X_{(1)} \\ i/n & \text{if } X_{(i)} \le x < X_{(i+1)} \quad \text{for } i = 1,2,\ldots,n-1 \\ 1 & \text{if } X_{(n)} \le x\end{cases}$$

Then, in particular, $P(X = X_{(1)}) = 1/n$, which is not reasonable for a continuous distribution. For $0 \le x < X_{(1)}$, suppose that we try to correct this problem by defining $F(x)$ as

$$F(x)=\begin{cases}0 & \text{if } 0 \le x < c \\ (x-c)/\left[n\left(X_{(1)}-c\right)\right] & \text{if } c \le x < X_{(1)}\end{cases}$$

for some c such that $0 \le c < X_{(1)}$. Then how should c be chosen? In general, it is not known whether x values smaller than $X_{(1)}$ are possible.

6.6. Each of the $n - 1$ subintervals of length $1/(n - 1)$ along the $F(x)$ axis has a probability of $1/(n - 1)$ of a random number U falling in it. If U falls in the subinterval $[(i - 1)/(n - 1), i/(n - 1))$ for $i = 1, 2, \ldots, n - 1$, then the mean value of X generated will be $[X_{(i)} + X_{(i+1)}]/2$. Therefore,

$$
\begin{aligned}
E(X) &= \frac{1}{n-1}\sum_{i=1}^{n-1}\left[X_{(i)} + X_{(i+1)}\right]/2 \\
&= \left[X_{(1)}/2 + X_{(2)} + X_{(3)} + \cdots + X_{(n-1)} + X_{(n)}/2\right]/(n-1) \\
&\neq \bar{X}(n) \quad \text{in general}
\end{aligned}
$$

6.7. $h_j = \sum_{i=1}^{n} Y_i \big/ n$, so $E(h_j) = E(Y_i) = (1)P(X_i = x_j) + (0)P(X_i \neq x_j) = p(x_j)$.

6.8. The empirical distribution function $F(x)$ defined in Sec. 6.2.4 can be used regardless of the histogram shape.

6.9. The average number of failures before the first success is $\overline{X}(n)$. Thus, on average one out of every $\overline{X}(n) + 1$ trials should be a success.

6.10. (*a*) $L(b) = 1/b^n$, and b must satisfy $0 \le X_i \le b$ for $i = 1, 2, \ldots, n$. Thus, $L(b)$ is maximized by setting b as small as possible, i.e., $\hat{b} = X_{(n)}$, the largest X_i.

(*b*) $L(a) = 1/(-a)^n$, and we must have $a \le X_i \le 0$ for all i. Thus, $L(a)$ is maximized by setting a to be as large as it can be (i.e., as close to 0 as possible since $a < 0$), giving $\hat{a} = X_{(1)}$, the smallest X_i.

(*c*) $L(a, b) = 1/(b-a)^n$, where $a \le X_i \le b$ for each i. Again by inspection $L(a, b)$ is maximized by setting $\hat{a} = X_{(1)}$ and $\hat{b} = X_{(n)}$.

(*d*) $L(\mu, \theta) = (2\pi)^{-n/2}\sigma^{-n}\exp\left[-\sum_{i=1}^n (X_i-\mu)^2/(2\sigma^2)\right]$ and $l(\mu, \sigma) = (-n/2)\ln(2\pi) - n\ln\sigma - [1/(2\sigma^2)]\sum_{i=1}^n (X_i-\mu)^2$. Taking partial derivatives of l, we get $\partial l/\partial\mu = \sum_{i=1}^n (X_i-\mu)/\sigma^2$ and $\partial l/\partial\sigma = -n/\sigma + \sum_{i=1}^n (X_i-\mu)^2/\sigma^3$. Setting $\partial l/\partial\mu = 0$, we get $\hat{\mu} = \overline{X}(n)$, immediately. Substituting this $\hat{\mu}$ for μ into $\partial l/\partial\sigma$ and setting this partial to 0 yields $\hat{\sigma} = \sqrt{\sum_{i=1}^n \left[X_i - \overline{X}(n)\right]^2/n}$.

(*e*) Since $X \sim \mathrm{LN}(\mu, \sigma^2)$ if and only if $\ln X \sim \mathrm{N}(\mu, \sigma^2)$, $\hat{\mu} = \sum_{i=1}^n \ln X_i/n$ and $\hat{\sigma} = \sqrt{\sum_{i=1}^n (\ln X_i - \hat{\mu})^2/n}$.

(*f*) $L(p) = p^A(1-p)^{n-A}$, where $A = \sum_{i=1}^n X_i$, so $l(p) = A\ln p + (n-A)\ln(1-p)$. Thus, $dl/dp = A/p - (n-A)/(1-p)$, which equals 0 if and only if $p = \hat{p} = \overline{X}(n)$. This is a maximum, since $d^2l/dp^2 = -A/p^2 - (n-A)/(1-p)^2 < 0$.

(*g*) $L(i, j) = 1/(j-i+1)^n$, where $i \le X_k \le j$ for $k = 1, 2, \ldots, n$. $L(i, j)$ is maximized when j is as small as possible and i is as large as possible, i.e., $\hat{i} = X_{(1)}$ and $\hat{j} = X_{(n)}$.

(*h*) $L(p) = \binom{t}{X_1}\binom{t}{X_2}\cdots\binom{t}{X_n}p^A(1-p)^{nt-A}$, where $A = \sum_{i=1}^n X_i$, so $l(p) = \sum_{i=1}^n \ln\binom{t}{X_i} + A\ln p + (nt-A)\ln(1-p)$. Thus, $dl/dp = A/p - (nt-A)/(1-p)$, which equals 0 if and only if $p = \hat{p} = \overline{X}(n)/t$. This is a maximum, since $L(0) = L(1) = 0 < L(\hat{p})$.

(*i*) $L(p) = \binom{s+X_1-1}{X_1}\binom{s+X_2-1}{X_2}\cdots\binom{s+X_n-1}{X_n}p^{ns}(1-p)^A$, where $A = \sum_{i=1}^n X_i$, so $l(p) = \sum_{i=1}^n \ln\binom{s+X_i-1}{X_i} + ns\ln p + A\ln(1-p)$. Thus, $dl/dp = ns/p - A/(1-p)$, which equals 0 if and only if $p = \hat{p} = s/\left[\overline{X}(n)+s\right]$. As in part (*h*), this is a maximum.

(*j*) $L(\theta) = \begin{cases} 1 & \text{if } \theta - 0.5 \le X_i \le \theta + 0.5 \quad \text{for all } i \\ 0 & \text{otherwise} \end{cases}$ since the width of the interval is 1 regardless of θ. Thus, L is maximized for *any* choice of θ satisfying the condition $\theta - 0.5 \le X_i \le \theta + 0.5$ for all i; this condition is equivalent to $X_{(n)} - 0.5 \le \theta \le X_{(1)} + 0.5$. If this condition can be met (i.e., if $X_{(1)} + 0.5 \ge X_{(n)} - 0.5$, i.e., if $X_{(n)} - X_{(1)} \le 1$), then any θ between $X_{(n)} - 0.5$ and $X_{(1)} + 0.5$ maximizes L. Thus, the MLE of θ in this case is *not* unique.

6.11. The likelihood function is

$$L(\lambda) = \frac{e^{-\lambda n} \lambda^{\sum_{i=1}^{n} X_i}}{\prod_{i=1}^{n} X_i !}$$

so $l(\lambda) = -\lambda n + \sum_{i=1}^{n} X_i \ln \lambda - \sum_{i=1}^{n} \ln X_i !$. Thus, $dl/d\lambda = -n + \sum_{i=1}^{n} X_i \big/ \lambda$ and $d^2l/dl^2 = -\sum_{i=1}^{n} X_i \big/ \lambda^2$. Since the X_i's are IID Poisson with parameter λ, $E(X_i) = \lambda$ and so $E(d^2l/dl^2) = -n\lambda/\lambda^2 = -n/\lambda$. Therefore, $\delta(\theta) = \lambda$ and an approximate $100(1 - \alpha)$ percent confidence interval for λ is $\hat{\lambda} \pm z_{1-\alpha/2}\sqrt{\hat{\lambda}/n}$, where $\hat{\lambda} = \overline{X}(n)$.

6.12. The likelihood function is $L(\gamma, \beta) = \beta^{-n} \exp[-\sum_{i=1}^{n}(X_i - \gamma)/\beta]$, with $\gamma \le X_i$ for all i, so $l(\gamma, \beta) = -n \ln \beta - \sum_{i=1}^{n}(X_i - \gamma)/\beta = -n \ln \beta - (1/\beta)\sum_{i=1}^{n} X_i + n\gamma/\beta$. Since $\beta > 0$, we want to maximize the final term, $n\gamma/\beta$, by choosing γ as large as possible, i.e., $\hat{\gamma} = X_{(1)}$. With this choice of γ, we want to maximize the function $g(\beta) = -n \ln \beta - (1/\beta)\sum_{i=1}^{n}\left[X_i - X_{(1)}\right]$ over all $\beta > 0$. The solution is $\hat{\beta} = \overline{X}(n) - X_{(1)}$ by manipulations similar to those in Example 6.6.

6.13. Let $\hat{f}(x)$ be the fitted density function and let

$$Y_i = \begin{cases} 1 & \text{if } X_i \in \left[b_{j-1}, b_j\right) \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, n$. Then $\overline{Y}(n)$ is the proportion of the n observations in $[b_{j-1}, b_j)$ and $E[\overline{Y}(n)] = E(Y_i) = (1)P(Y_i = 1) + (0)P(Y_i = 0) = P\{X_i \in [b_{j-1}, b_j)\} = \int_{b_{j-1}}^{b_j} \hat{f}(x)\,dx$ if X_i has density $\hat{f}(x)$; this integral is equal to r_j.

6.14. If we let $q_i = i/n$ for $i = 1, 2, \ldots, n$, then $q_n = 1$ and $\hat{F}^{-1}(q_n) = \infty$ for many distributions. It is, of course, difficult to plot ∞.

6.15. There is not, in general, a value of x such that $F(x) = q$ for $0 < q < 1$.

6.16. The U(0, 1) distribution.

6.17. $M_j \sim \text{bin}(n, p_j)$, and $E(M_j) = np_j$.

6.19. Let

$$I_i = \begin{cases} 1 & \text{if } X_i \le x \\ 0 & \text{otherwise} \end{cases}$$

Then

$$F_n(x) = \frac{\text{number of } X_i\text{'s} \le x}{n} = \frac{\sum_{i=1}^{n} I_i}{n} = \overline{I}(n)$$

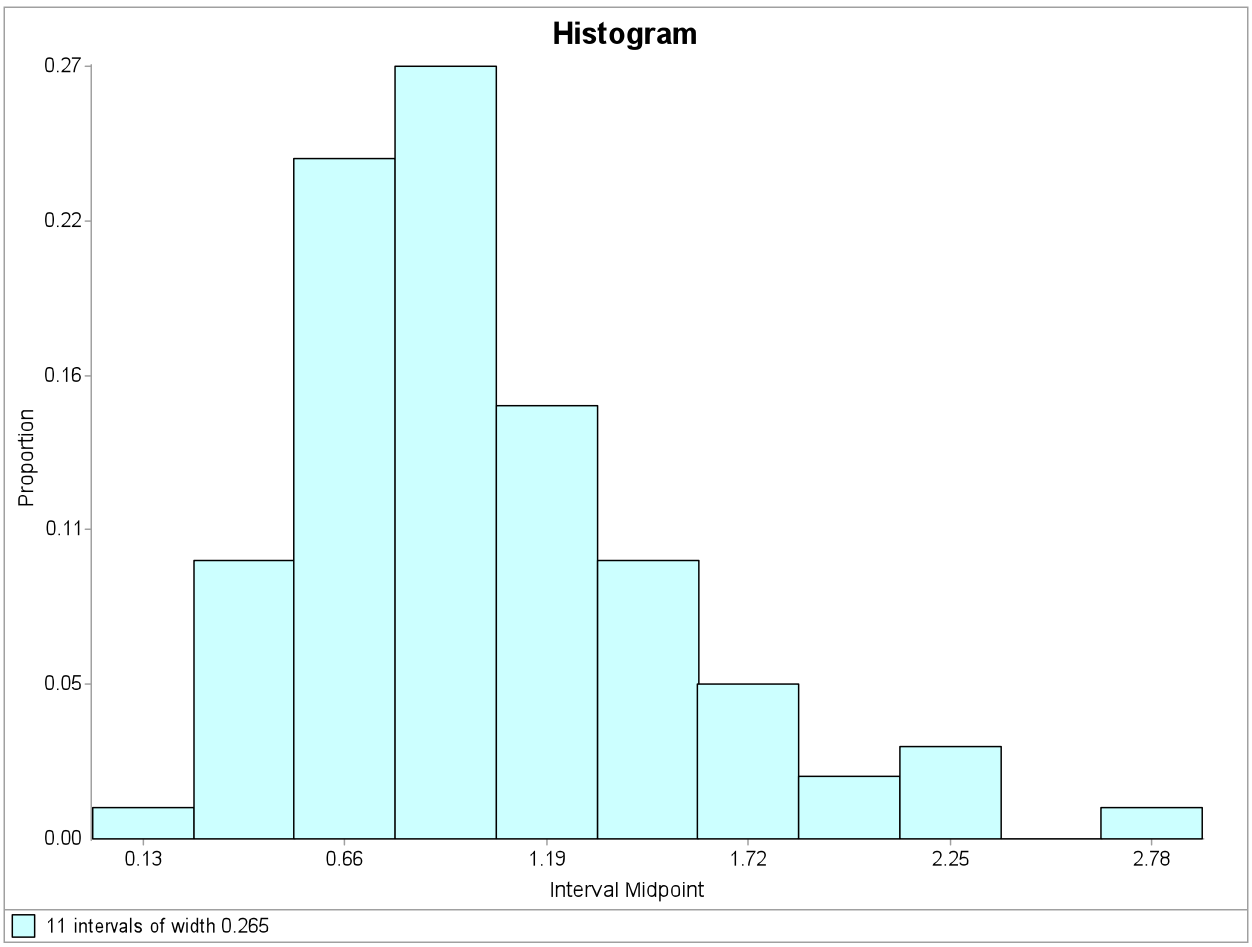
By the strong law of large numbers (Section 4.6), $\overline{I}(n) \to E(I)$ w.p. 1 as $n \to \infty$. Furthermore, $E(I) = 1 \cdot P(X \le x) + 0 \cdot P(X > x) = P(X \le x) = F(x)$. Therefore,

$$F_n(x) \to F(x) \text{ as } n \to \infty \text{ (w.p. 1) for all } x$$

6.20. (*a*) 91 laboratory-processing times

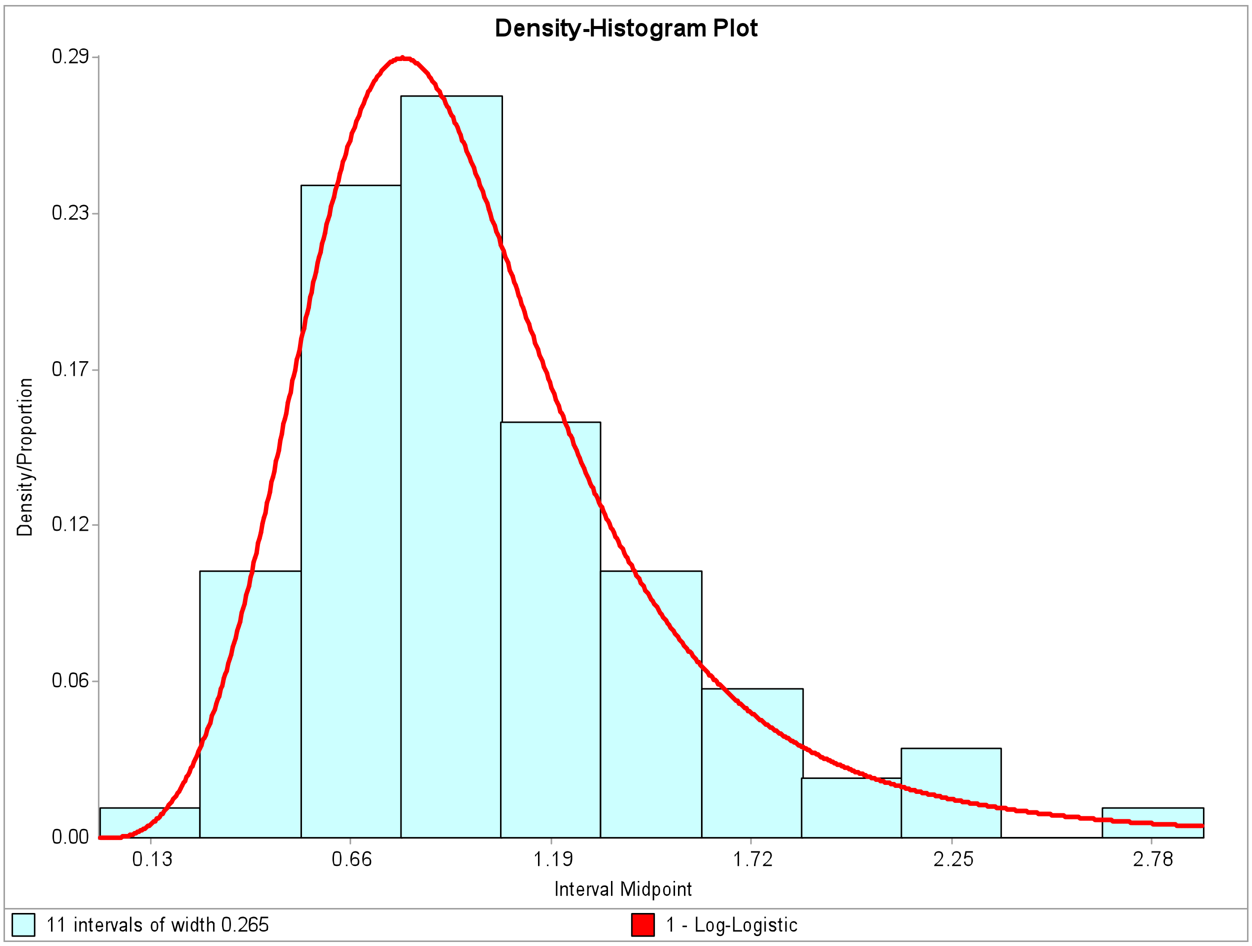
Data Characteristic	Value
Source file	PROB620a
Observation type	Real valued
Number of observations	91
Minimum observation	0.067
Maximum observation	2.733
Mean	1.019
Median	0.917
Variance	0.227
Coefficient of variation	0.467
Skewness	1.121

Since the mean is larger than the median and the skewness is positive, this suggests that the underlying density function is skewed to the right.

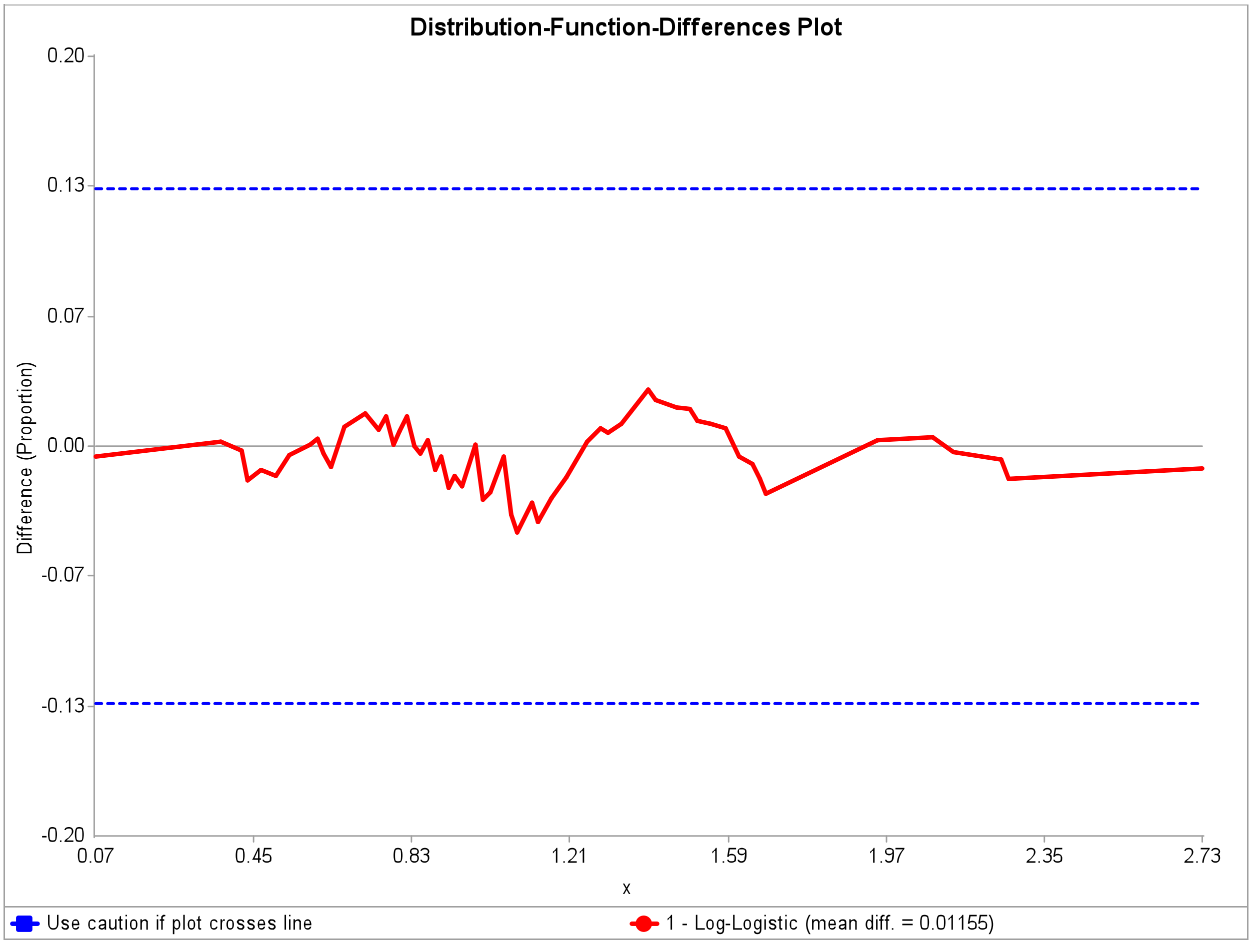


The histogram starts at 0 and is based on 11 intervals of width 0.265, which was determined by trial and error. The shape of the histogram confirms that the underlying density function is skewed to the left.

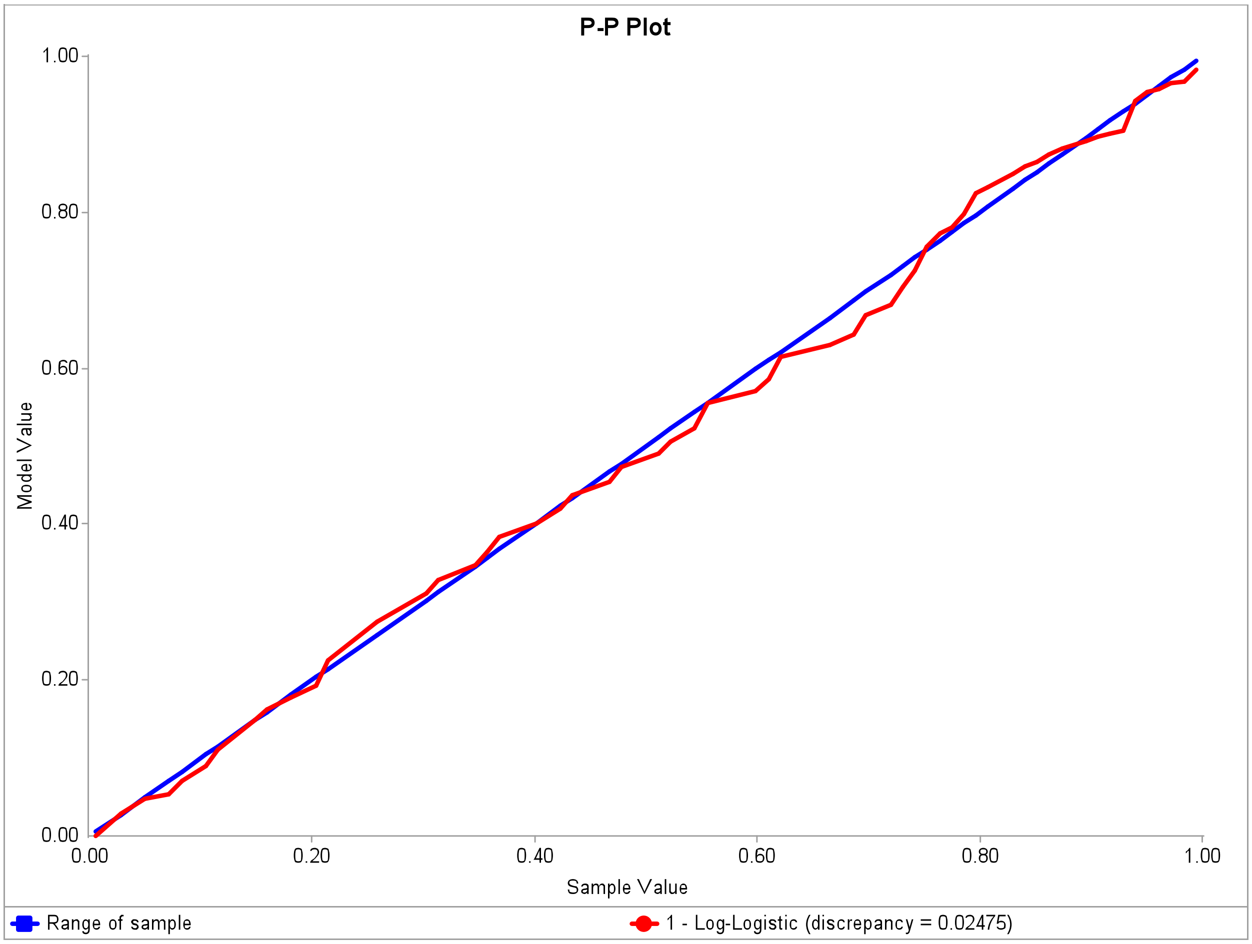
The best-fitting distribution is the log-logistic distribution (location = 0, scale = 0.927, shape = 3.765) with "Relative Score" = 96.43 and "Absolute Evaluation" = "Good." (See the context-sensitive "Help.") Actually, the log-Laplace distribution also had Relative Score = 96.43, but is not considered here because it is not in the book.



The density function of the log-logistic distribution fits over the histogram fairly well, considering that the histogram is not very "smooth."



The distribution-function-differences plot indicates that the log-logistic distribution fits the data well.



The *P-P* plot indicates that the log-logistic distribution fits the data well.

For the chi-square test, we used the default 18 equiprobable intervals, each of which has an expected (model) count greater than the recommended value of 5 (see p. 349 in the book). Since the test-statistic value, 4.934, is less than the critical value (point) of 27.587 for 17 df and $\alpha = 0.05$, we fail to reject the null hypothesis that the data have a log-logistic distribution.

For the K-S test, the test statistic is $D_{91} = 0.056$ (see Case 5 on p. 355 in the book) and the modified (adjusted) test statistic is $\sqrt{91} \times D_{91} = 0.530$. Since the value 0.530 is less than the modified critical value of 0.770 (n = 50) to 0.780 ($n = \infty$) for $\alpha = 0.05$, we again fail to reject the log-logistic distribution.

For the A-D test, the test statistic is $A_{91}^2 = 0.248$. Since this value is less than the critical value of 0.658 for $\alpha = 0.05$ given by ExpertFit, we once again fail to reject the log-logistic distribution. Note that ExpertFit actually rejects the null hypothesis in this case if

$$A_{91}^2 > \frac{0.660}{1 + (0.25/91)} = 0.658$$

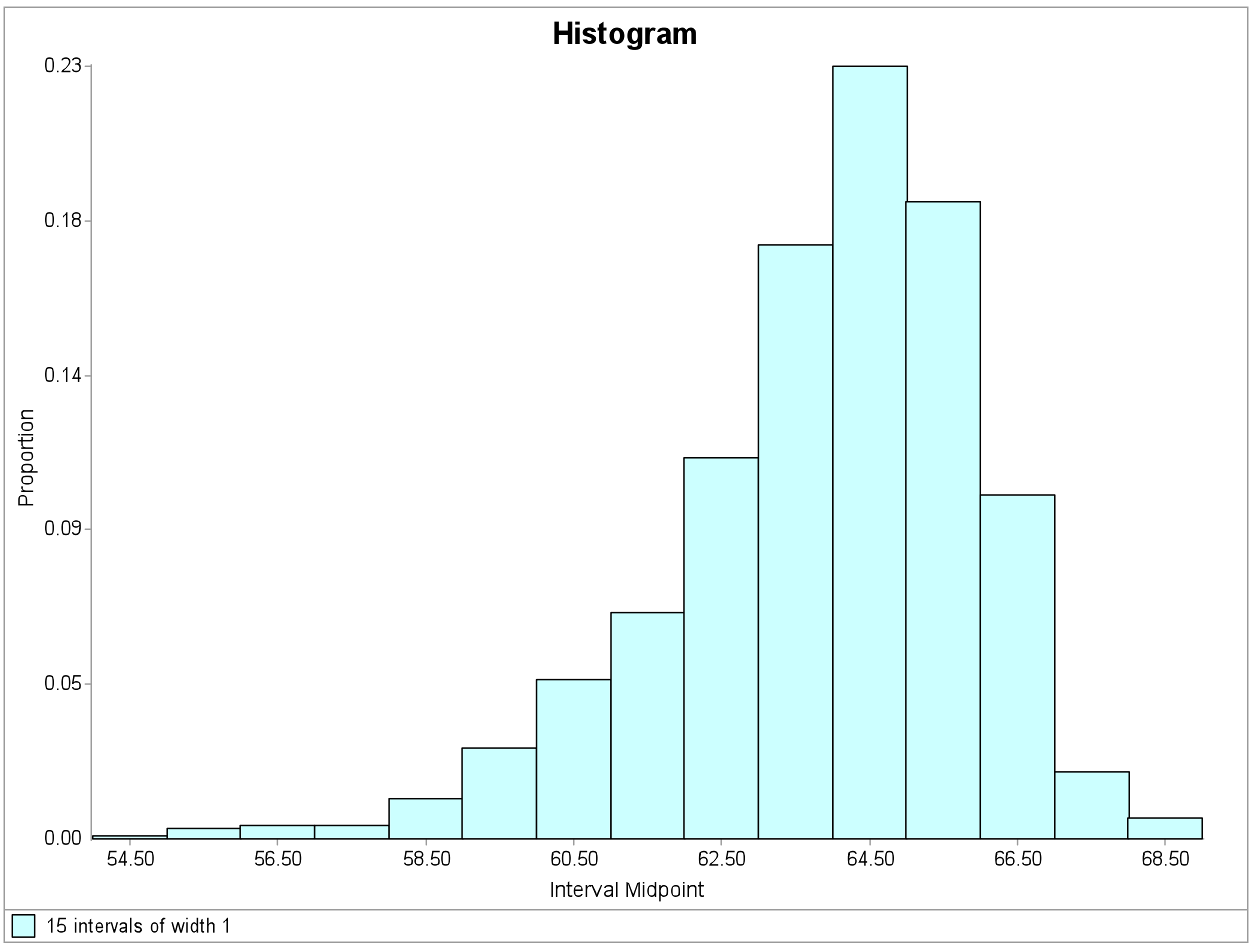
[The formula $1+(0.25/\sqrt{n})$ in the last row of Table 6.18 should actually be $1+(0.25/n)$, as stated in the errata for the book (see www.mhhe.com/law).]

Based on the Absolution Evaluation of Good, the quality of the plots, and the results from the three goodness-of-fit tests, it appears that the log-logistic distribution provides a good representation for the laboratory-processing times.

(*b*) 1000 paper-roll yardages

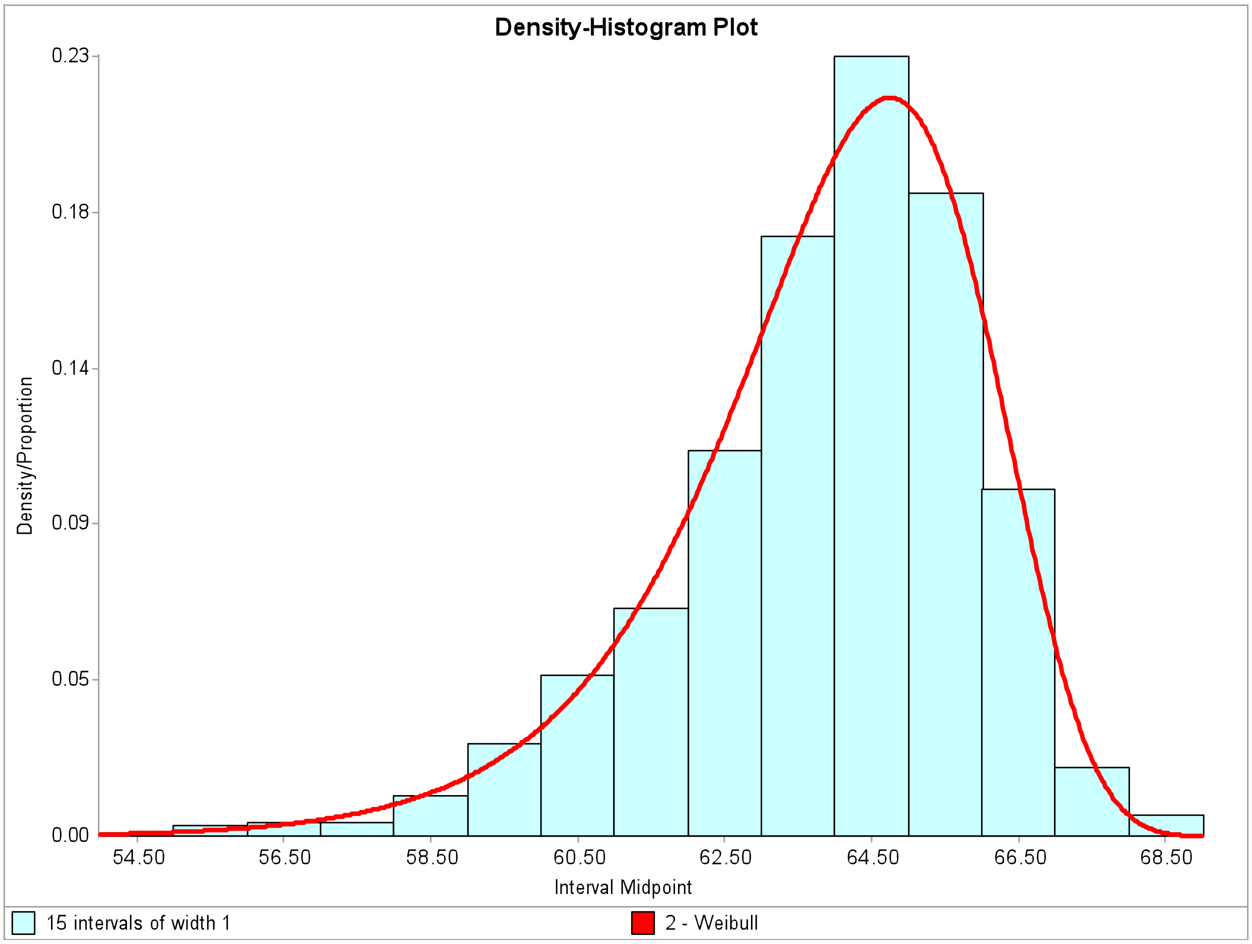
Data Characteristic	Value
Source file	PROB620b
Observation type	Real valued
Number of observations	1,000
Minimum observation	54.816
Maximum observation	68.652
Mean	63.876
Median	64.224
Variance	4.258
Coefficient of variation	0.032
Skewness	-0.898

Since the mean is smaller than the median and the skewness is negative, this suggests that the underlying density function is skewed to the left.

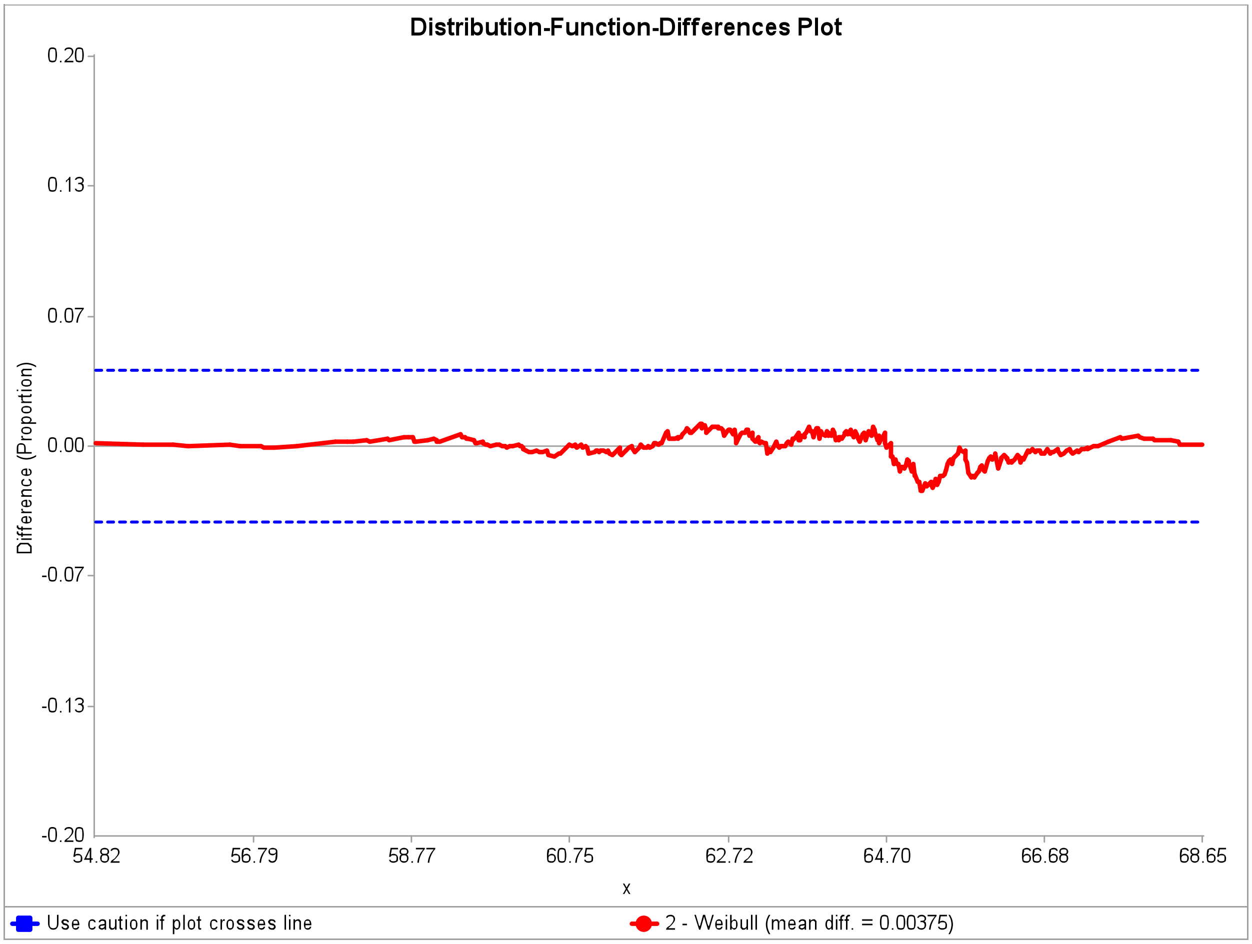


The histogram starts at 54 and is based on 15 intervals of width 1, which was determined by trial and error. The shape of the histogram confirms that the underlying density function is skewed to the left.

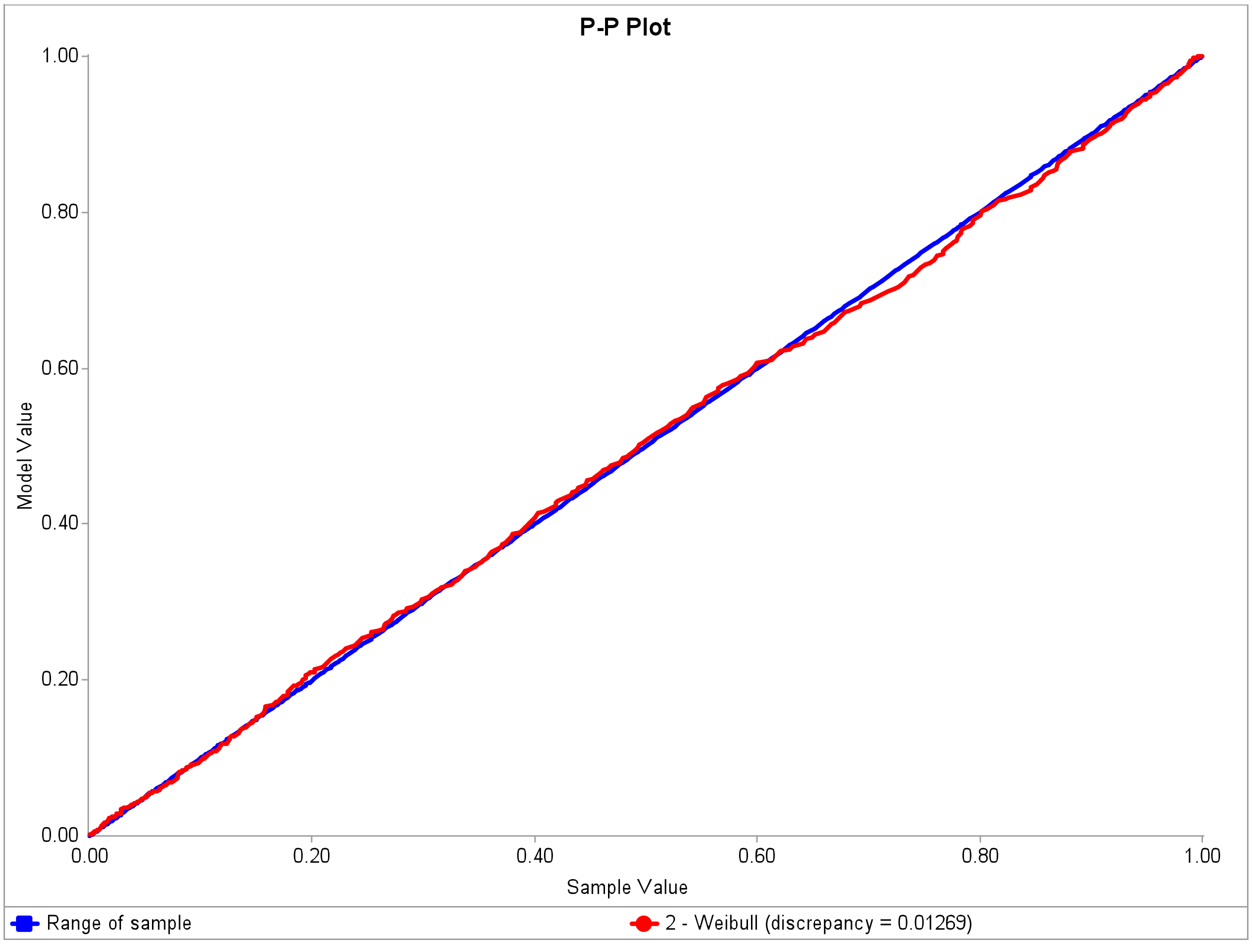
The second-best-fitting distribution is the Weibull distribution (location = 0, scale = 64.800, shape = 38.212) with "Relative Score" = 97.58 and "Absolute Evaluation" = "Good." We do not consider the best-fitting distribution (also a Weibull) here, since it has an *estimated* location parameter and exact critical values for the tests are *not* available.



The density function of the log-logistic distribution fits over the histogram extremely well (except for $64 \le x \le 65$), suggesting that the Weibull distribution is a good representation for the data.



The distribution-function-differences plot indicates that the Weibull distribution fits the data well.



The *P-P* plot indicates that the Weibull distribution fits the data very well.

For the chi-square test, we used the default 40 equiprobable intervals, each of which has an expected (model) count greater than the recommended value of 5 (see p. 349 in the book). Since the test-statistic value, 34.72, is less than the critical value of 54.572 for 39 df and $\alpha = 0.05$, we fail to reject the null hypothesis that the data have a Weibull distribution.

For the K-S test, the test statistic is $D_{1000} = 0.024$ (see Case 4 on p. 355 in the book) and the modified test statistic is $\sqrt{1000} \times D_{1000} = 0.745$. Since the value 0.745 is less than the modified critical value of 0.856 (n = 50) to 0.874 ($n = \infty$) for $\alpha = 0.05$, we again fail to reject the Weibull distribution.

For the A-D test, the test statistic is $A^2_{1000} = 0.412$. Since this value is less than the critical value of 0.752 for $\alpha = 0.05$ given by ExpertFit, we once again fail to reject the Weibull distribution. Note that ExpertFit rejects the null hypothesis in this case if

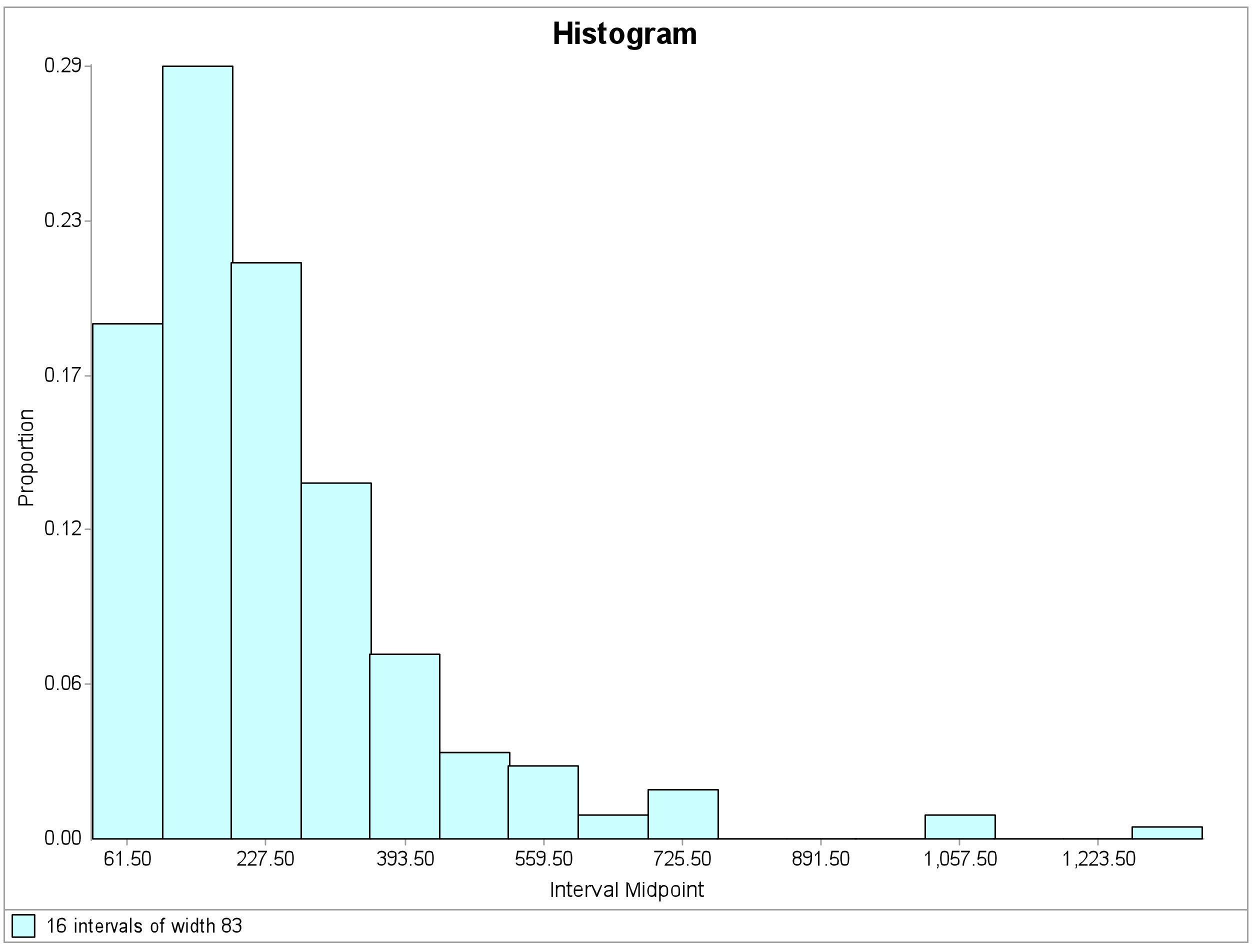
$$A^2_{1000} > \frac{0.757}{1 + (0.2 / \sqrt{1000})} = 0.752$$

Based on the Absolution Evaluation of Good, the quality of the plots, and the results from the three goodness-of-fit tests, it appears that the Weibull distribution provides a good representation for the paper-roll yardages.

(*c*) 218 post-office service times

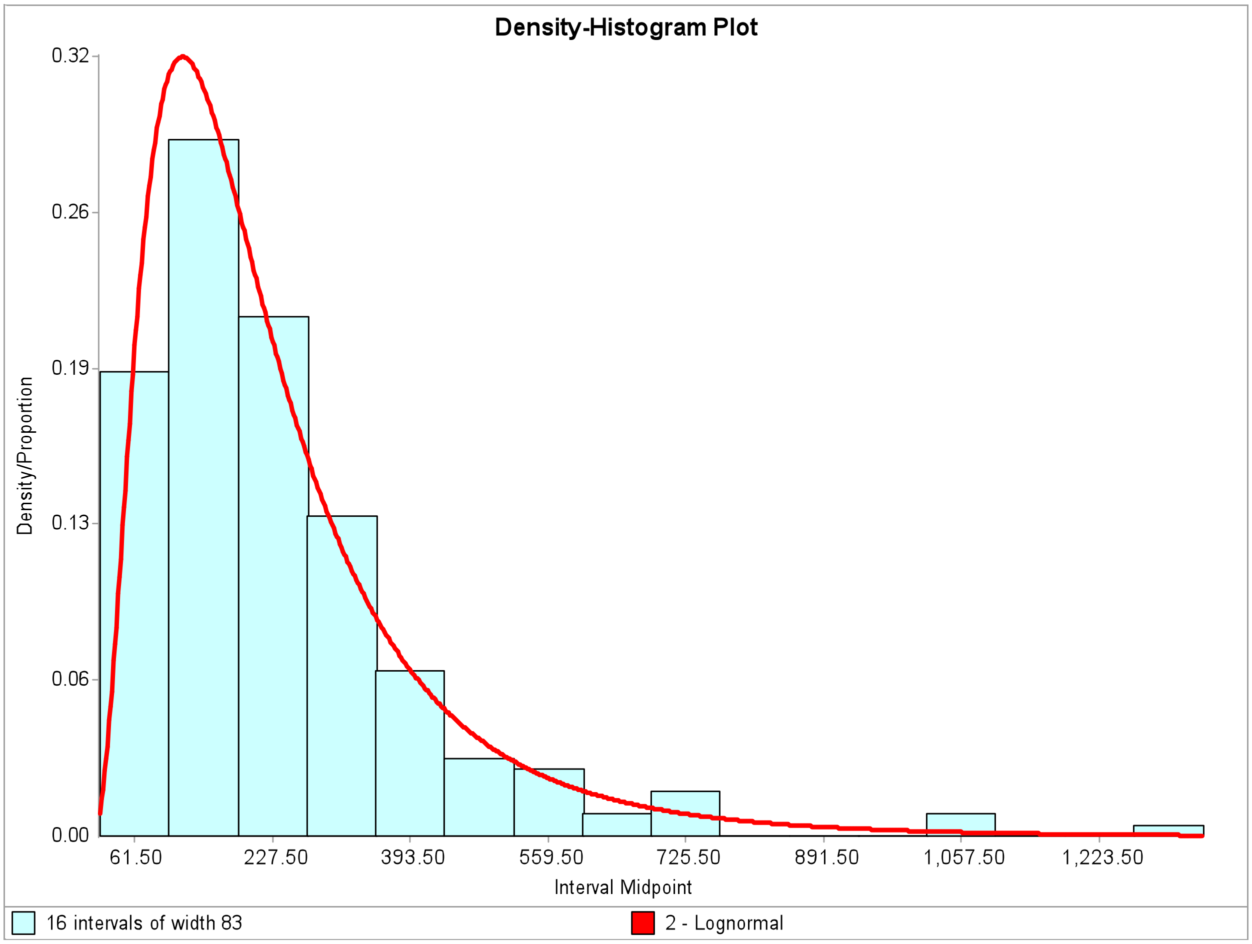
Data Characteristic	Value
Source file	PROB620c
Observation type	Real valued
Number of observations	218
Minimum observation	26.000
Maximum observation	1284.200
Mean	238.693
Median	195.500
Variance	32,421.908
Coefficient of variation	0.754
Skewness	2.332

Since the mean is larger than the median and the skewness is positive, this suggests that the underlying density function is skewed to the right.

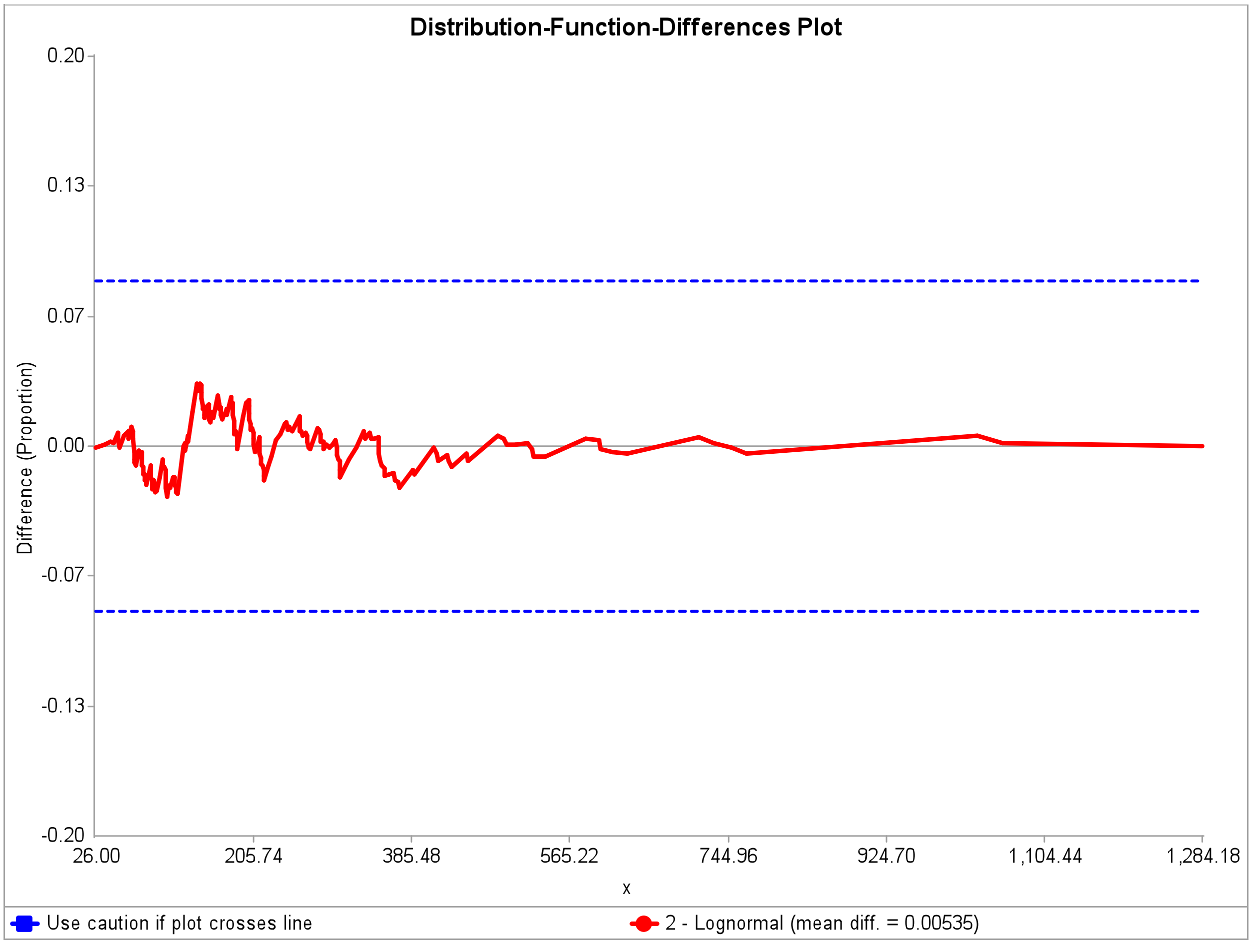


The histogram starts at 20 and is based on 16 intervals of width 83, which was determined by trial and error. The shape of the histogram confirms that the underlying density function is skewed to the right.

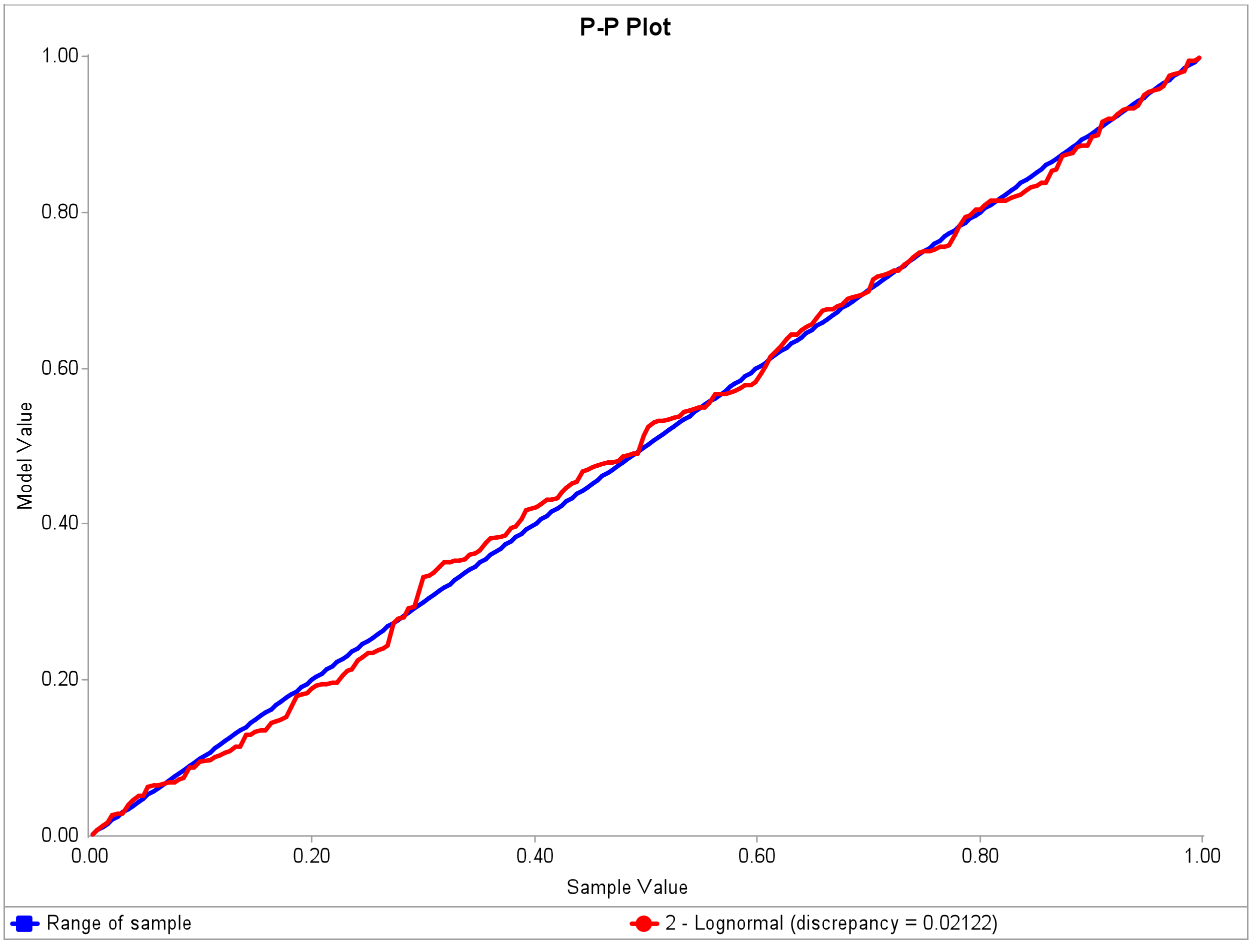
The second-best-fitting distribution is the lognormal distribution (location = 0, scale = 189.862, shape = 0.678) with "Relative Score" = 92.00 and "Absolute Evaluation" = "Good." We do not consider the best-fitting distribution (Pearson type VI) here, since exact critical values for the tests are *not* available.



The density function of the lognormal distribution fits over the histogram moderately well, suggesting that the lognormal distribution is a reasonable representation for the data.



The distribution-function-differences plot indicates that the lognormal distribution fits the data well.



The *P-P* plot indicates that the lognormal distribution fits the data well.

For the chi-square test, we used the default 40 equiprobable intervals, each of which has an expected (model) count greater than the recommended value of 5 (see p. 349 in the book). Since the test-statistic value, 41.817, is less than the critical value of 54.572 for 39 df and $\alpha = 0.05$, we fail to reject the null hypothesis that the data have a lognormal distribution.

For the K-S test (see Case 2 on p. 354 in the book), the test statistic is $D_{218} = 0.034$ and the modified test statistic is

$$\left(\sqrt{218} - 0.01 + \frac{0.85}{\sqrt{218}}\right) D_{218} = 0.506$$

Since the value 0.506 is less than the modified critical value of 0.892 for $\alpha = 0.05$, we again fail to reject the lognormal distribution.

For the A-D test, the test statistic is $A_{218}^2 = 0.229$. Since this value is less than the critical value of 0.749 for $\alpha = 0.05$ given by ExpertFit, we once again fail to reject the lognormal distribution. Note that ExpertFit actually rejects the null hypothesis in this case if

$$A_{218}^2 > \frac{0.752}{1 + (0.75/218) + (2.25/218^2)} = 0.749$$

[see D'Agostino and Stephens (1986, p. 123) for a discussion of this form of the test].

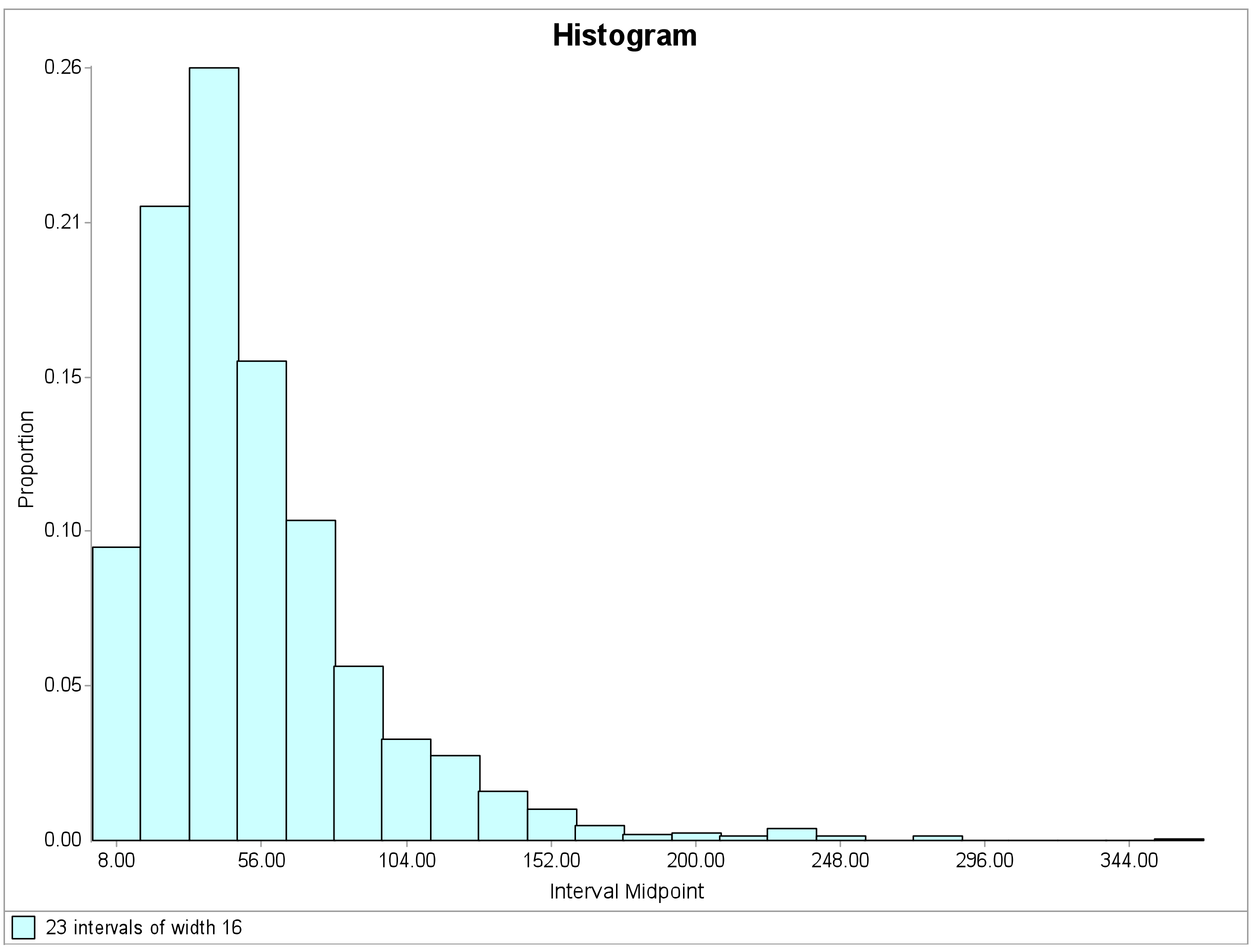
Based on the Absolution Evaluation of Good, the quality of the plots, and the results from the three goodness-of-fit tests, it appears that the lognormal distribution provides a good representation for the post-office service times.

(*d*) 1592 times at an ATM machine

Consider the data to be real valued.

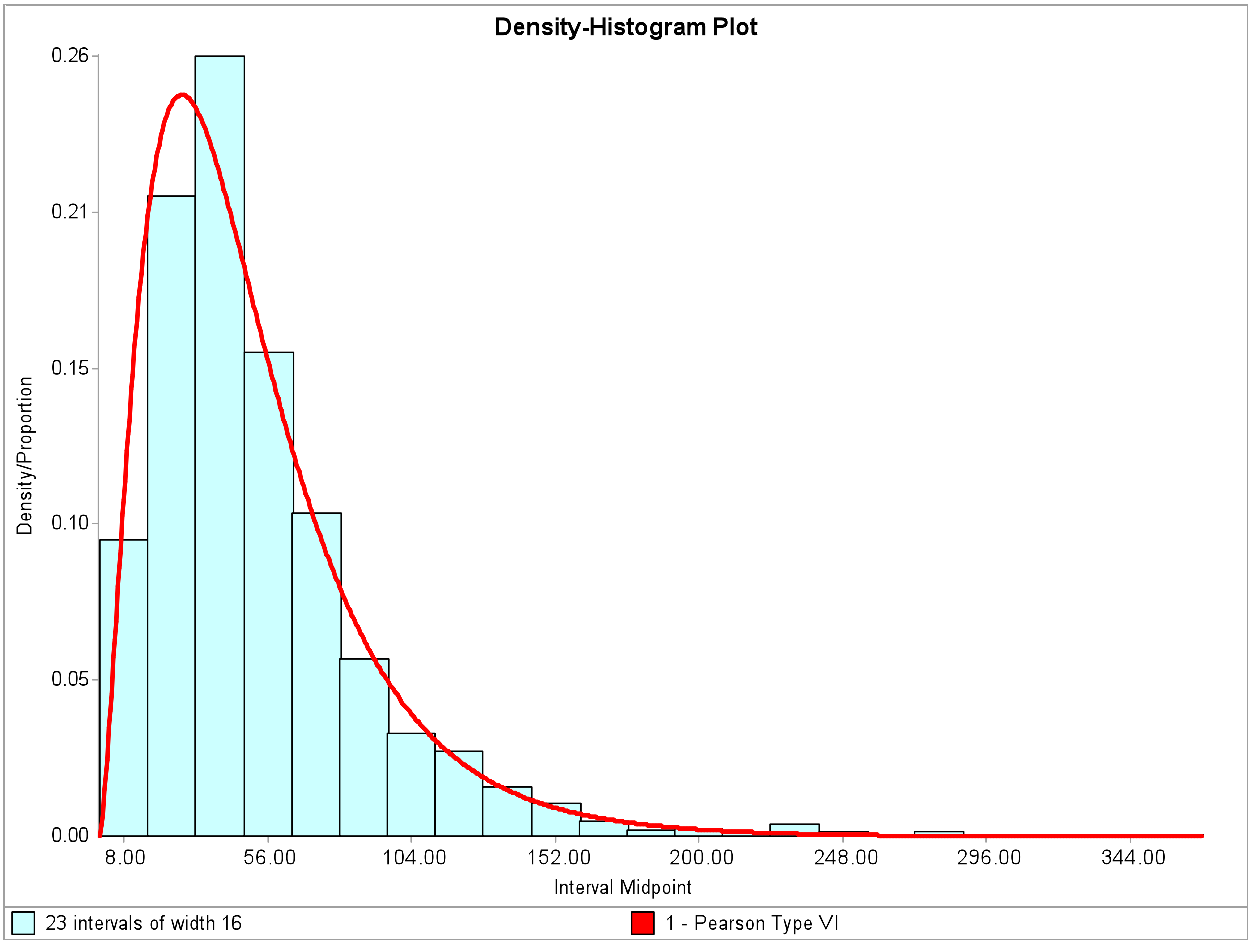
Data Characteristic	Value
Source file	PROB620d
Observation type	Real valued
Number of observations	1592
Minimum observation	5.000
Maximum observation	353.000
Mean	52.002
Median	42.000
Variance	1416.603
Coefficient of variation	0.724
Skewness	2.125

Since the mean is larger than the median and the skewness is positive, this suggests that the underlying density function is skewed to the right.

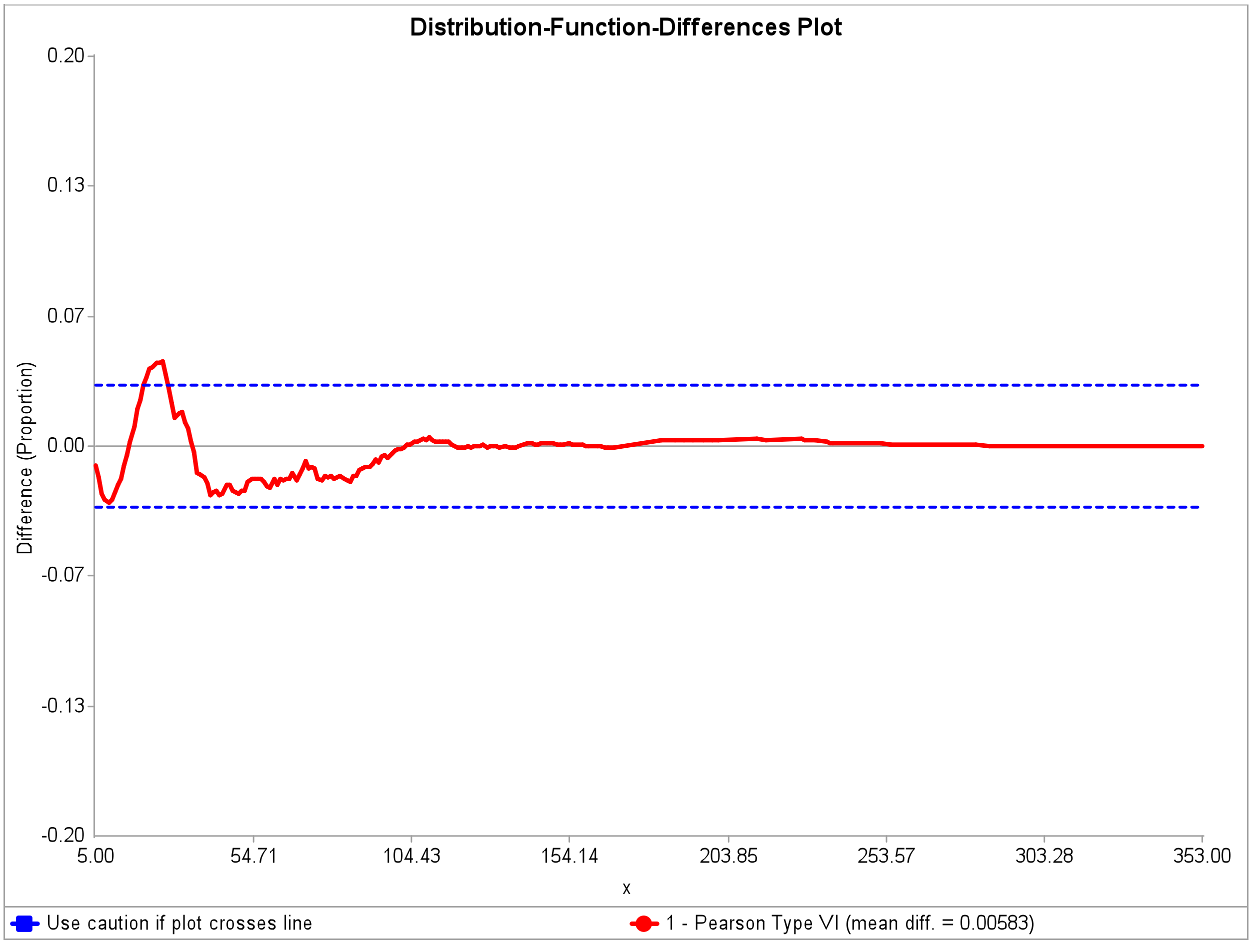


The histogram starts at 0 and is based on 23 intervals of width 16, which was determined by trial and error. The shape of the histogram confirms that the underlying density function is skewed to the right.

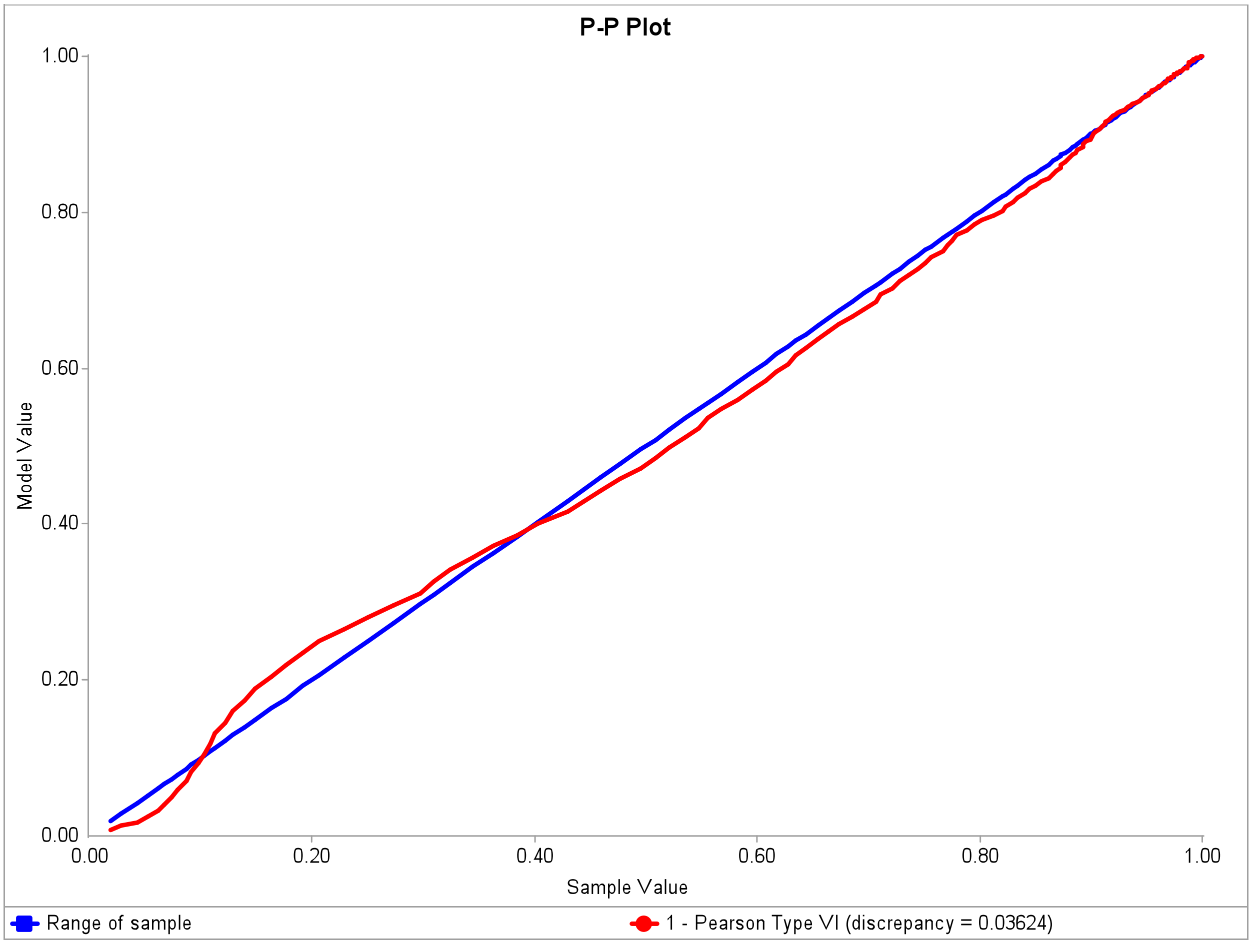
The best-fitting distribution is the Pearson type VI distribution (location = 0, scale = 250.338, shape 1 = 2.599, shape 2 = 13.519) with "Relative Score" = 91.25 and "Absolute Evaluation" = "Bad." The fact that the Absolute Evaluation is Bad strongly suggests that the Pearson type VI distribution is *not* an acceptable representation of the data.



The density function of the Pearson type VI distribution fits over the histogram reasonably well except in the range 16 to 48.



The distribution-function-differences plot strongly indicates that the Pearson type VI distribution does not fit the data well, since the plot *crosses* the dotted blue line.



The *P-P* plot also indicates that the Pearson type VI distribution does not fit the data well.

For the chi-square test, we used the default 40 equiprobable intervals, each of which has an expected (model) count greater than the recommended value of 5 (see p. 349 in the book). Since the test-statistic value, 226.291, is greater than the critical value of 54.572 for 39 df and $\alpha = 0.05$, we *reject* the null hypothesis that the data have a Pearson type VI distribution.

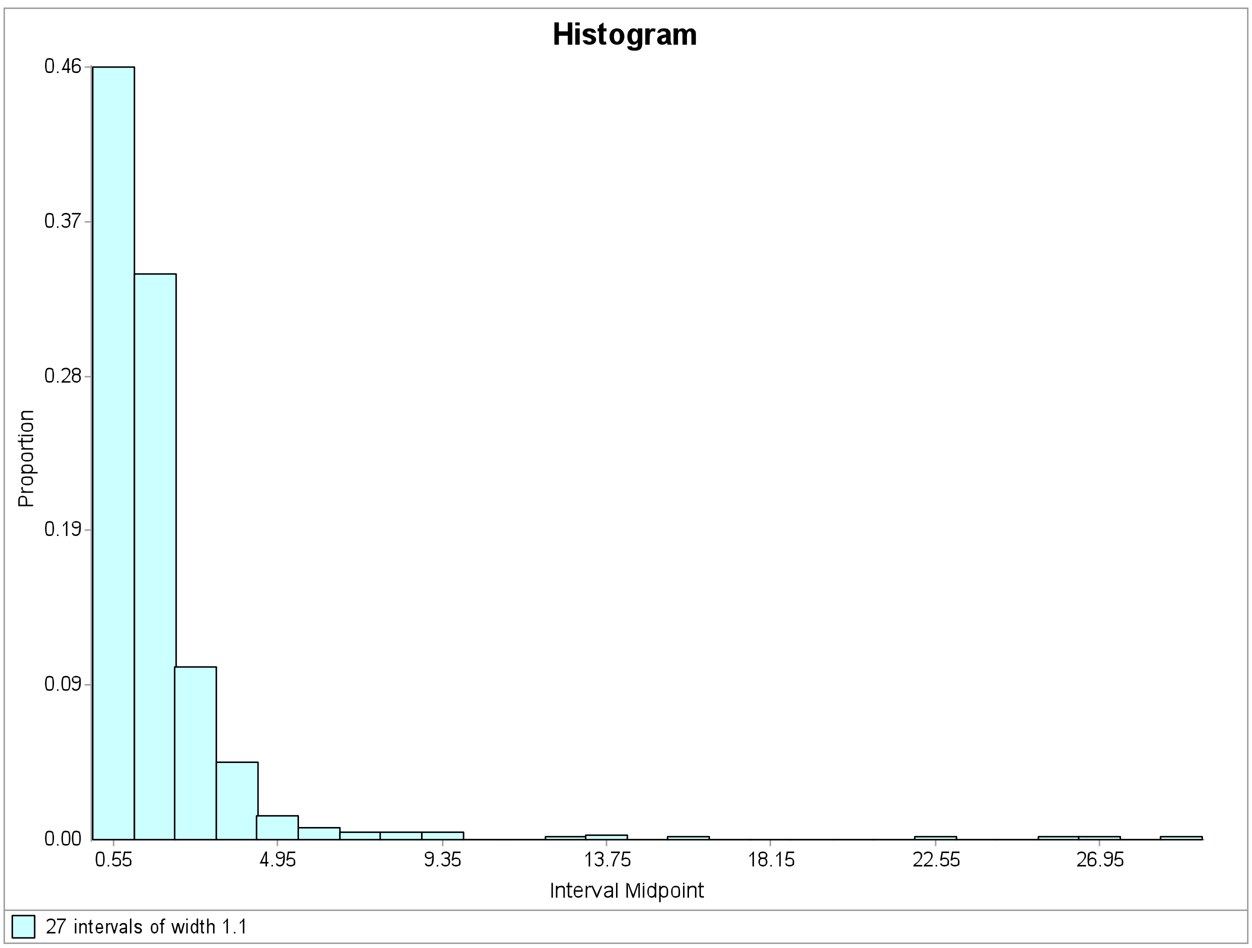
For the K-S test, there are no critical values available for the Pearson type VI distribution, so the test cannot be done in a statistically valid way. For the A-D test, there are also no critical values available for the Pearson type VI distribution, so the test cannot be performed correctly.

Based on the Absolute Evaluation, the distribution-function-differences and *P-P* plots, and the results from the chi-square test, we conclude that the Pearson type VI distribution is *not* a good representation for the ATM data.

(*e*) 694 machine-repair times

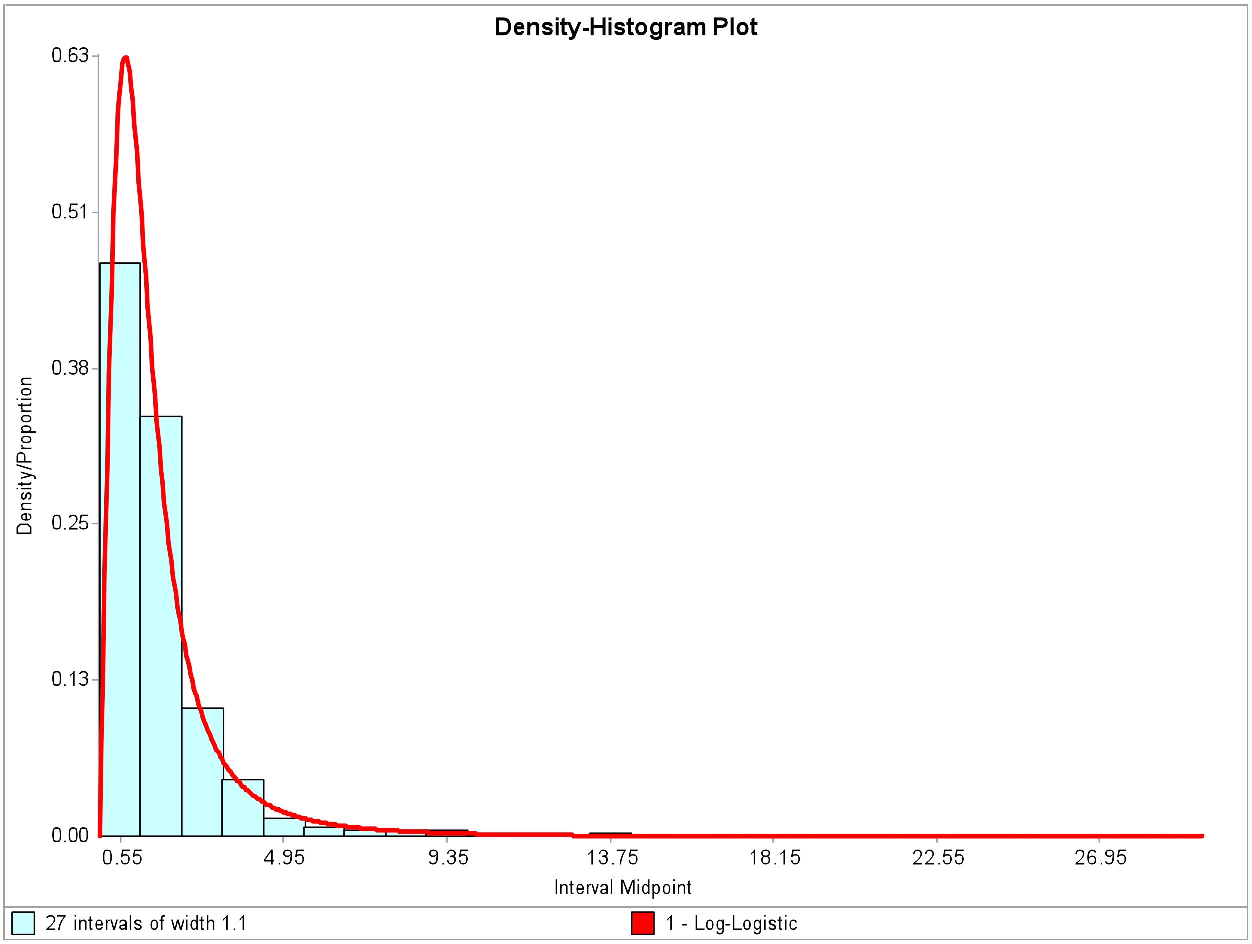
Data Characteristic	Value
Source file	PROB620e
Observation type	Real valued
Number of observations	694
Minimum observation	0.033
Maximum observation	28.767
Mean	1.717
Median	1.150
Variance	5.982
Coefficient of variation	1.424
Skewness	6.735

Since the mean is larger than the median and the skewness is (very) positive, this suggests that the underlying density function is skewed to the right.

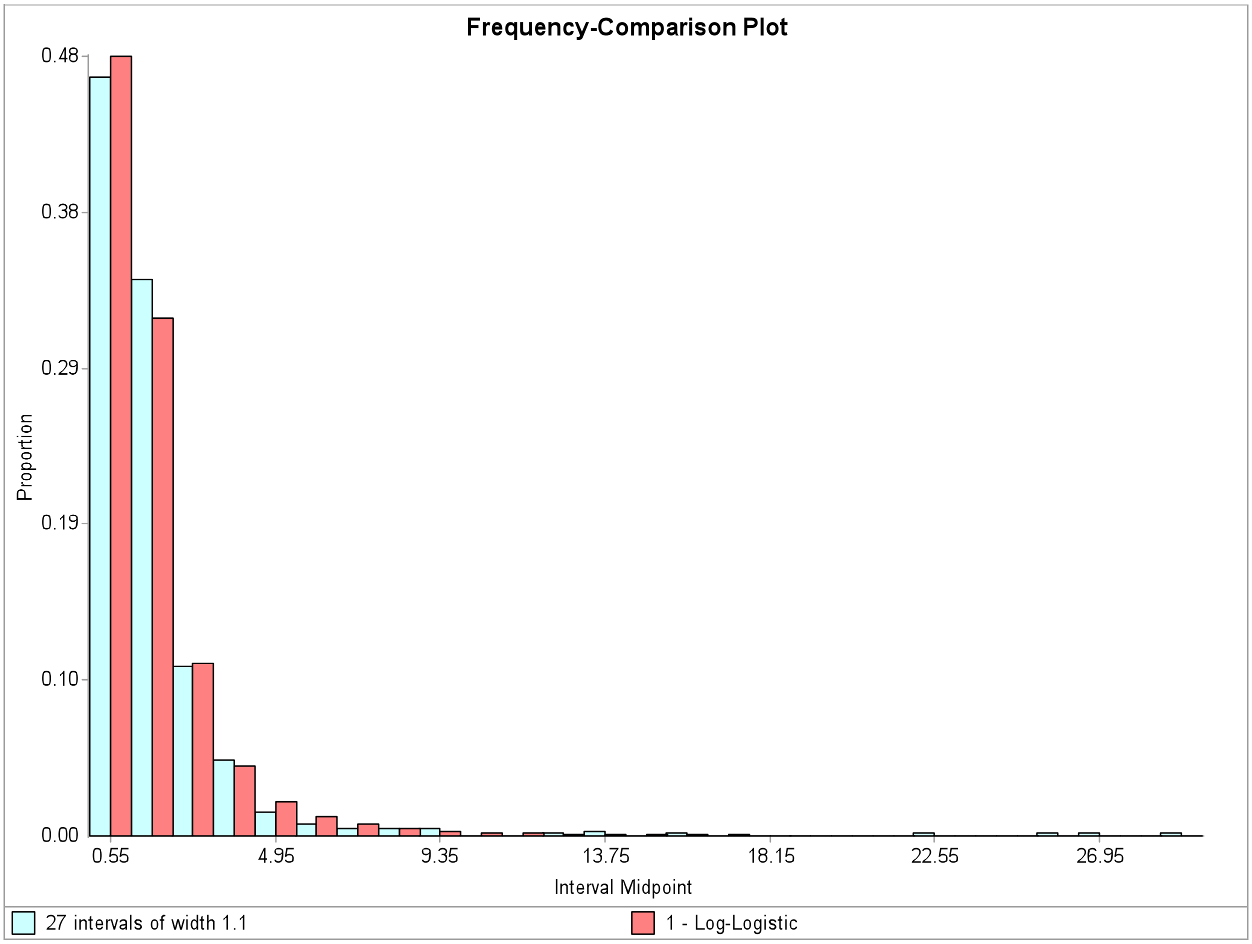


The histogram starts at 0 and is based on 27 intervals of width 1.1, which was determined by trial and error. There are 686 (out of 694) observations with values less than 10, but three observations have values of at least 25, suggesting that the underlying density function has a long right tail. (See the "View/Modify Data" option in the "Data" tab for details.)

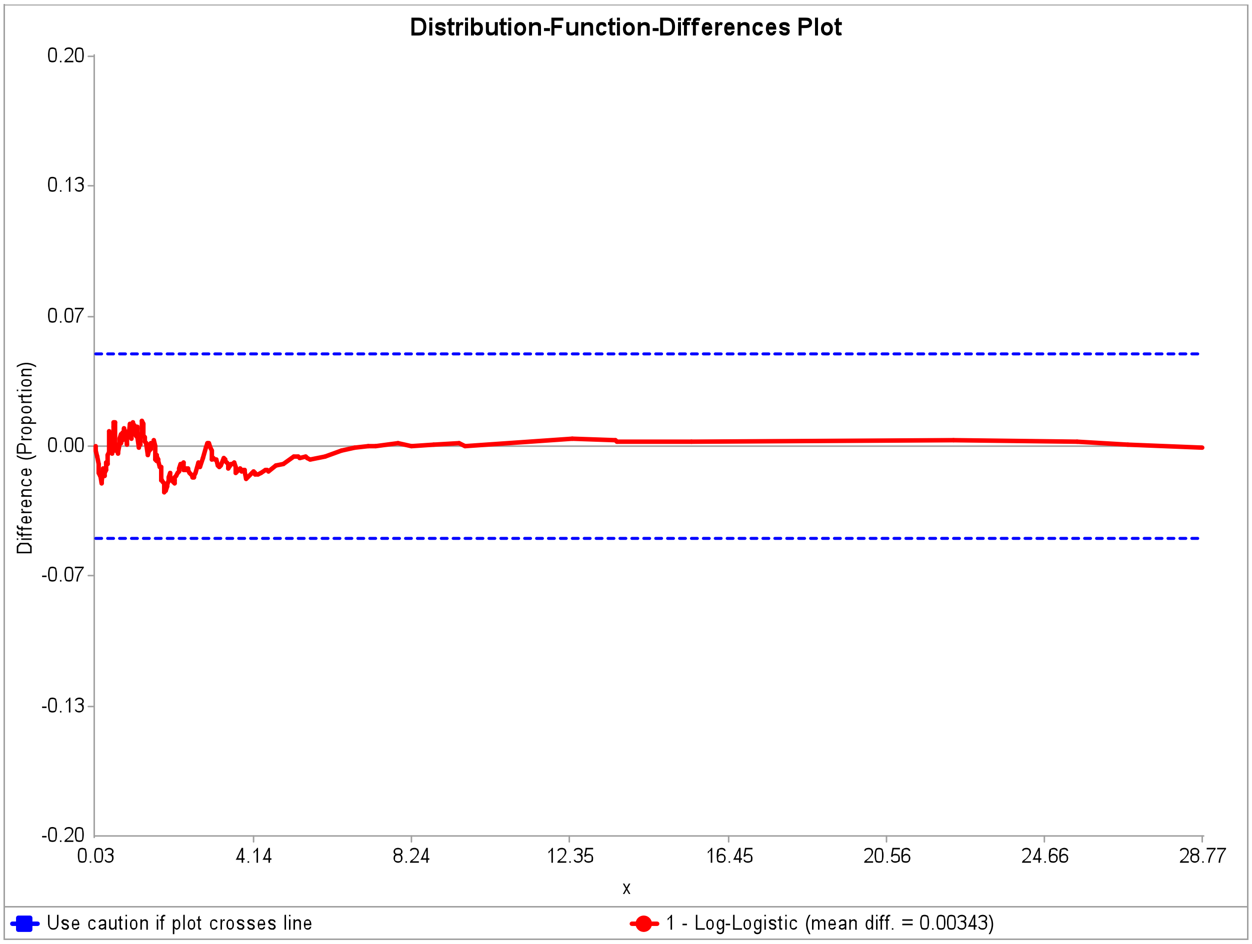
The best-fitting distribution is the log-logistic distribution (location = 0, scale = 1.151, shape = 2.074) with "Relative Score" = 100.00 and "Absolute Evaluation" = "Good." The fact that the Absolute Evaluation is Good suggests that the log-logistic distribution is probably a good representation of the data, but this should be confirmed using the recommended plots and goodness-of-fit tests.

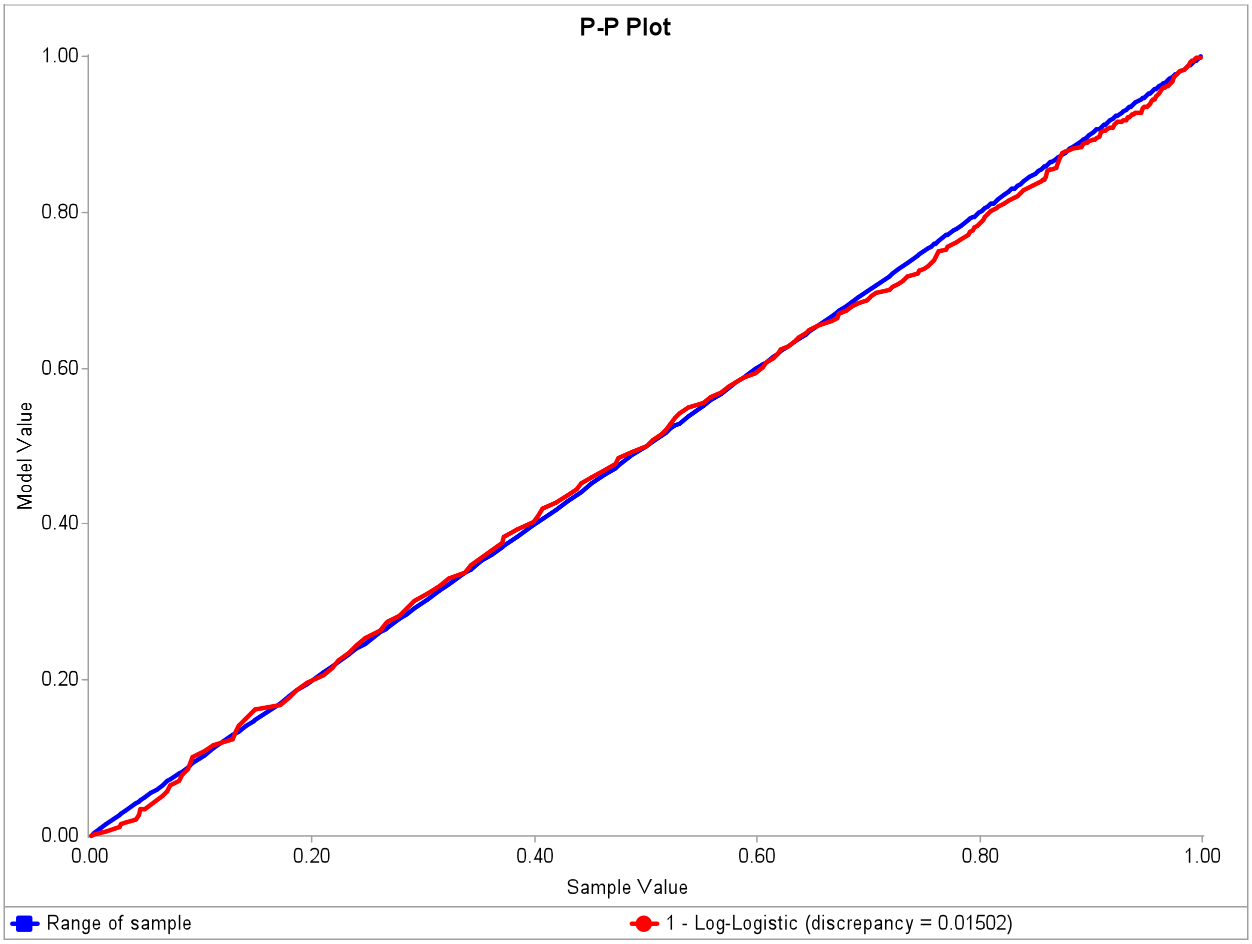


The density-histogram plot suggests that the log-logistic distribution is not a good representation of the data. However, if the maximum value of the true underlying density function is close to 0, then the shape of the histogram might be quite different for an alternative interval width. (For example, try an interval width of 0.5 and 39 intervals.) Therefore, we present below a frequency-comparison plot (see p. 335 in the book), which does, in fact, indicate that the log-logistic distribution provides a reasonably good representation of the data.



Distribution-function-differences and *P*-*P* plots are given below, neither of which gives us any particular reason for concern.


Distribution-Function-Differences Plot
Difference (Proportion)
0.20
0.13
0.07
0.00
-0.07
-0.13
-0.20
0.03
4.14
8.24
12.35
16.45
20.56
24.66
28.77
x
Use caution if plot crosses line
1 - Log-Logistic (mean diff. = 0.00343)



For the chi-square test, different outcomes can be obtained for different values of the number of intervals, k. For example, the chi-square test *rejects* the log-logistic distribution for $k = 40$, but *fails to reject* for $k = 20$. This shows the potential danger of relying on the chi-square test, because there is no definite prescription for choosing the value of k.

For the K-S test, the test statistic is $D_{694} = 0.024$ and the modified test statistic is $\sqrt{694} \times D_{694} = 0.645$. Since the latter value 0.645 is less than the modified critical value of 0.770 ($n = 50$) to 0.780 ($n = \infty$) for $\alpha = 0.05$, the K-S test fails to reject the log-logistic distribution.

For the A-D test, the test statistic is $A_{694}^2 = 0.725$. Since this value is greater than the critical value of 0.660 for $\alpha = 0.05$ given by ExpertFit, we reject the log-logistic distribution. Note that ExpertFit actually rejects the null hypothesis in this case if

$$A_{694}^2 > \frac{0.660}{1 + (0.25/694)} = 0.660$$

[The formula $1 + (0.25/\sqrt{n})$ in the last row of Table 6.18 should actually be $1 + (0.25/n)$, as stated in the errata for the book (see www.mhhe.com/law).] Since the A-D test is the most powerful of the three goodness-of-fit tests, it will often reject for “large” n unless the hypothesized distribution is essentially a perfect fit. Thus, one should not be too concerned that the A-D test rejects the log-logistic distribution for the large sample size of $n = 694$.

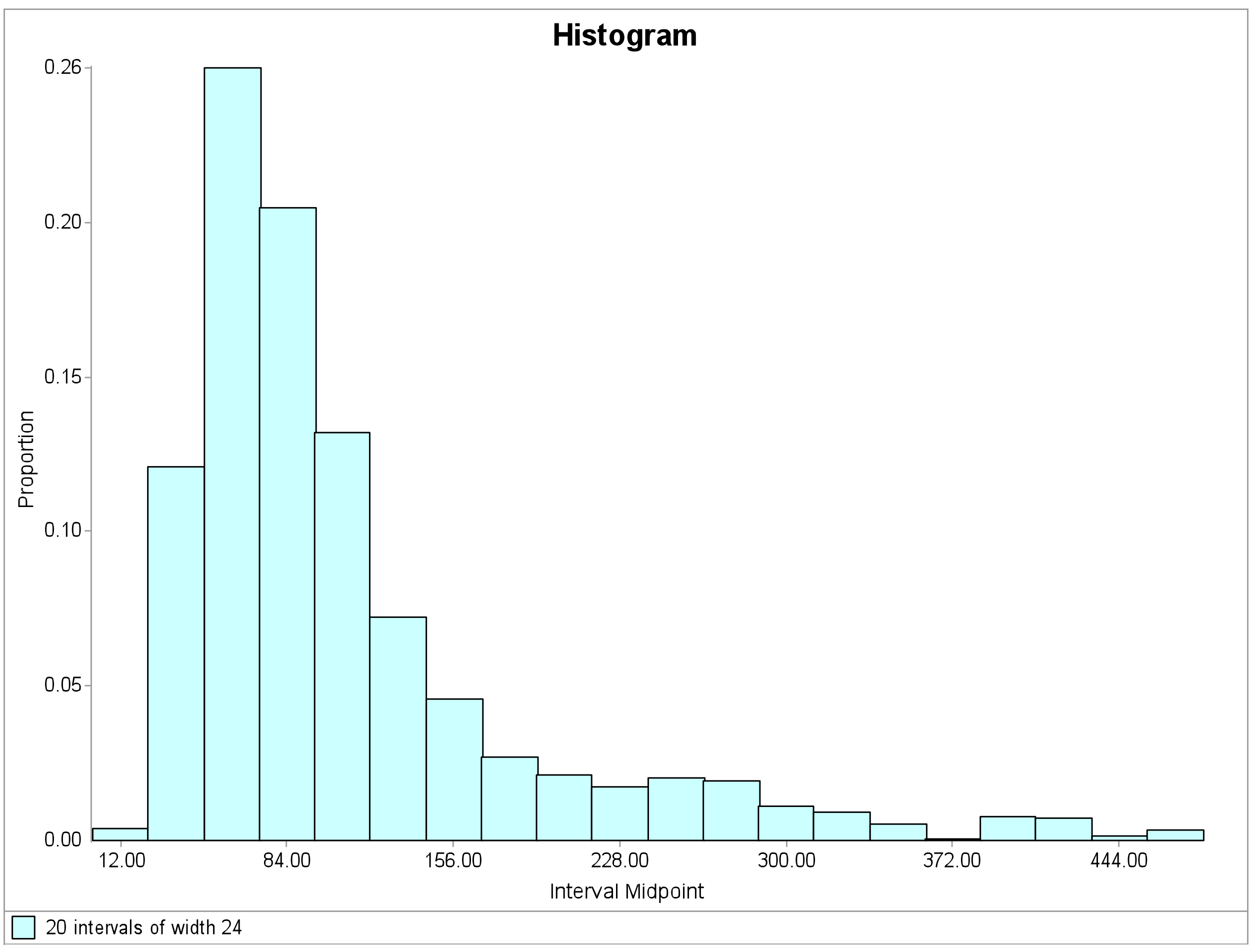
Overall, the log-logistic distribution provides a good representation of the machine-repair times and is acceptable to use in a simulation model.

(*f*) 3035 post-anesthesia recovery times

Consider the data to be real valued.

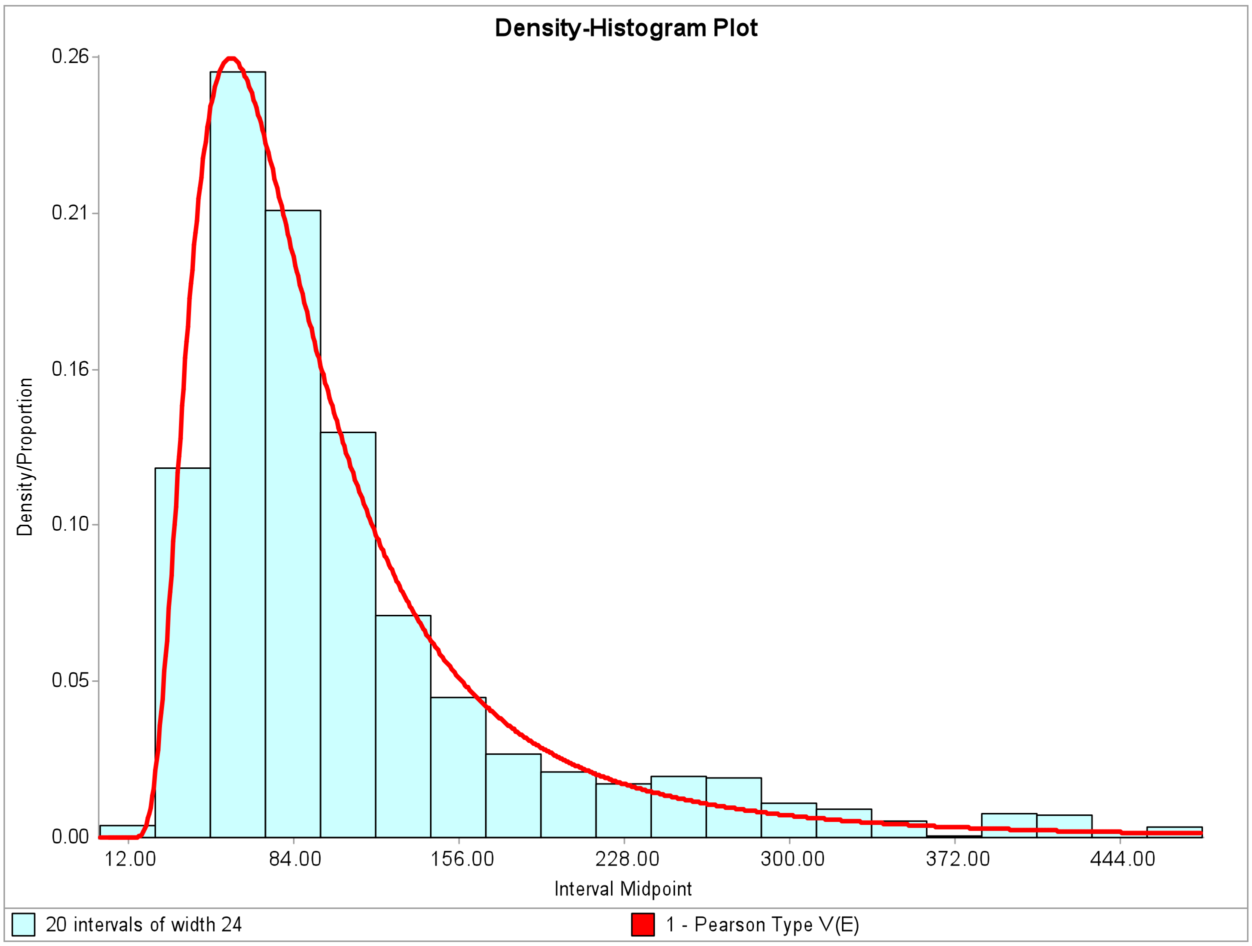
Data Characteristic	Value
Source file	PROB620f
Observation type	Real valued
Number of observations	3035
Minimum observation	21.000
Maximum observation	477.000
Mean	109.197
Median	84.000
Variance	6162.663
Coefficient of variation	0.719
Skewness	1.993

Since the mean is larger than the median and the skewness is positive, this suggests that the underlying density function is skewed to the right.

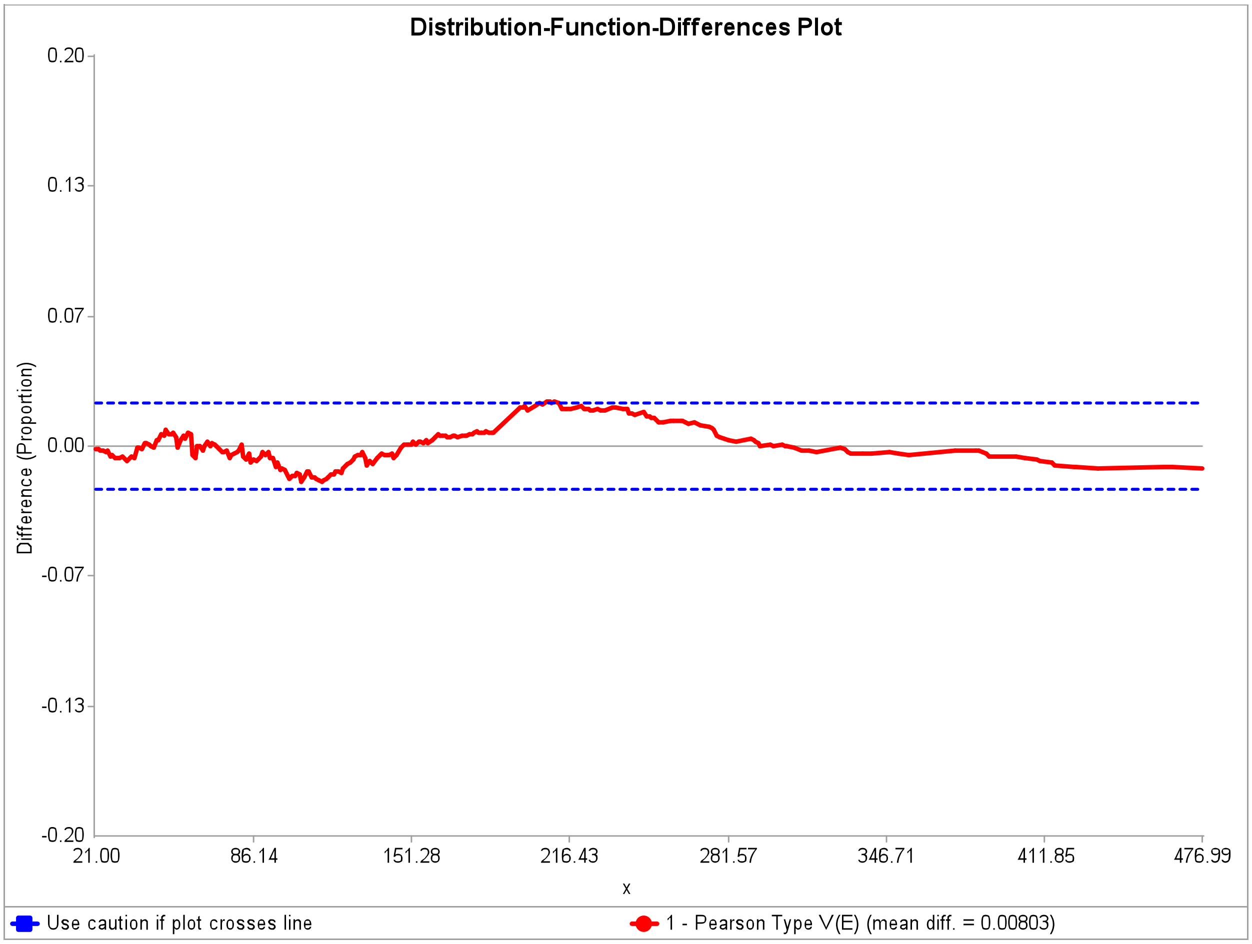


The histogram starts at 0 and is based on 20 intervals of width 24, which was determined by trial and error. Note that the histogram has a long and thick right tail.

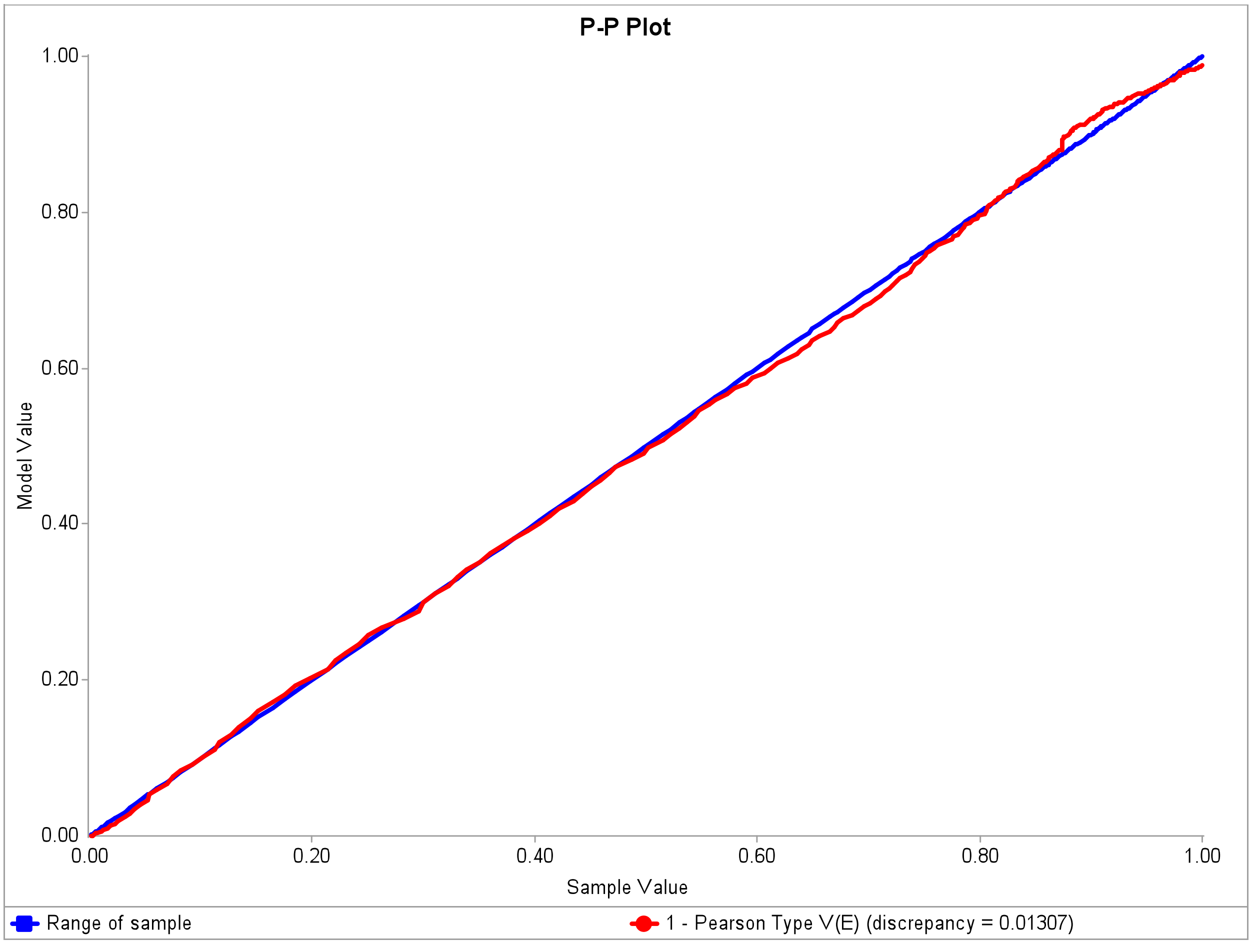
The best-fitting distribution is the Pearson type V distribution (location = 3.445, scale = 213.108, shape = 2.967) with "Relative Score" = 97.22 and "Absolute Evaluation" = "Intermediate." (Note that this distribution has an *estimated* location parameter, which is denoted by "E" in ExpertFit.) The fact that the Absolute Evaluation is "Intermediate" means that the Pearson type V distribution should not be used in a simulation model unless the recommended plots and goodness-of-fit tests (available in the "Comparisons" tab) show that it provides a reasonable representation of the data.



The density-histogram plot indicates that the Pearson type V distribution provides a *fair* representation of the data, with the main problem being the fit in the right tail.



The distribution-function-differences plot indicates that the Pearson type V distribution is a marginal representation of the data, since the plot touches the upper blue dotted line.



The *P-P* plot gives us little reason to be concerned.

The chi-square test *cannot* be applied in a statistically valid manner to this particular Pearson type V distribution, since the location parameter was not estimated using the method of maximum likelihood. (It is problematic to simultaneously estimate all three parameters by maximum likelihood, as discussed on p. 365 in the book.) Also, because of the estimated location parameter, the K-S and A-D tests cannot be applied in a statistically correct manner, although some software packages ignore this limitation and apply the tests anyway, resulting in tests with *low power*.

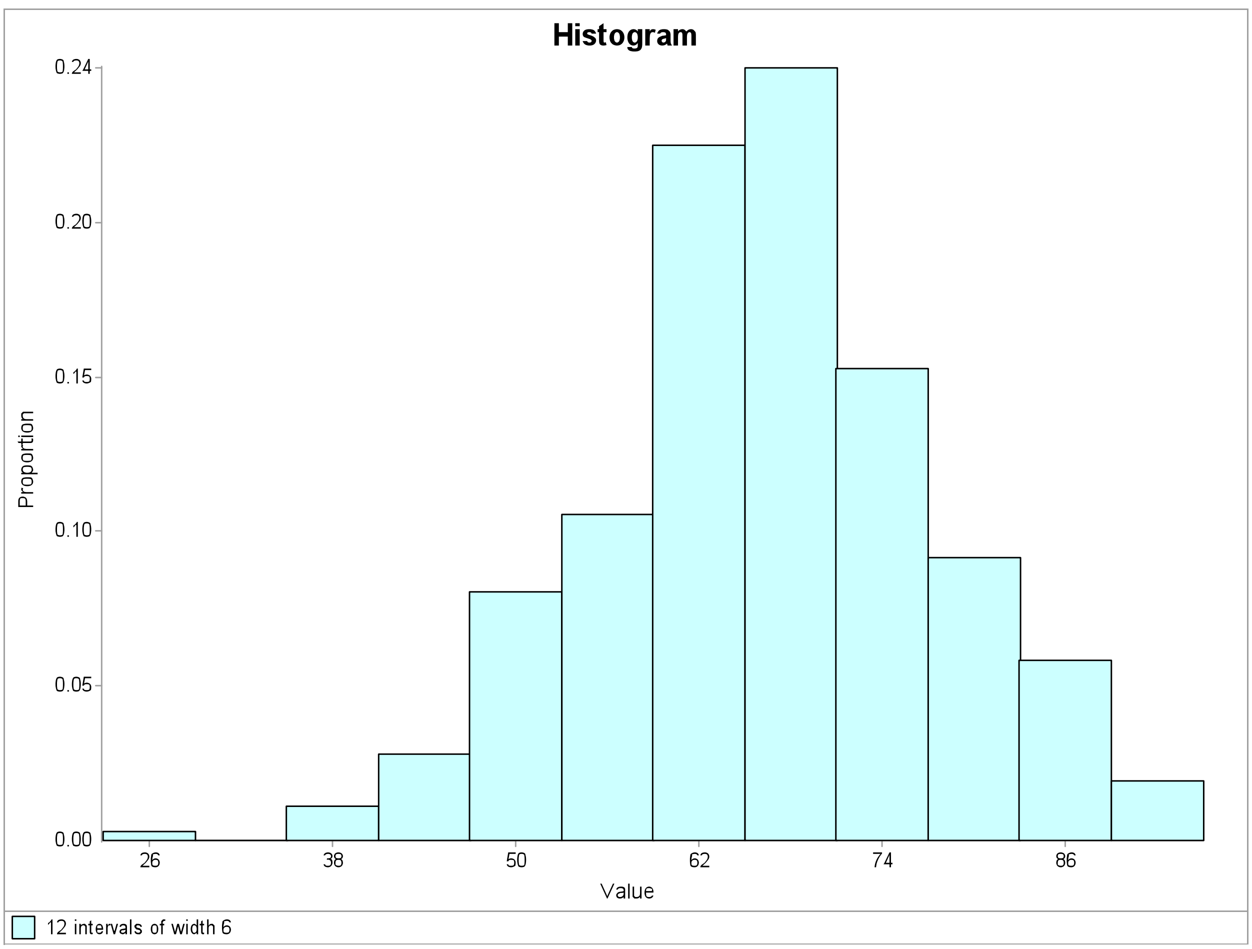
The quality of fit provided by the Pearson type V distribution is fair, as indicated by the density-histogram and distribution-function-differences plots. An alternative approach for representing post-amnesia recovery times in a simulation model is to use an empirical distribution (see p. 305 in the book), which should work well in this case because of the large sample size. On the other hand, the use of the empirical distribution would require 6070 numbers to be entered into the computer (3035 x and distribution function values).

6.21. (*a*) 369 university test scores

Consider the data to be integer valued.

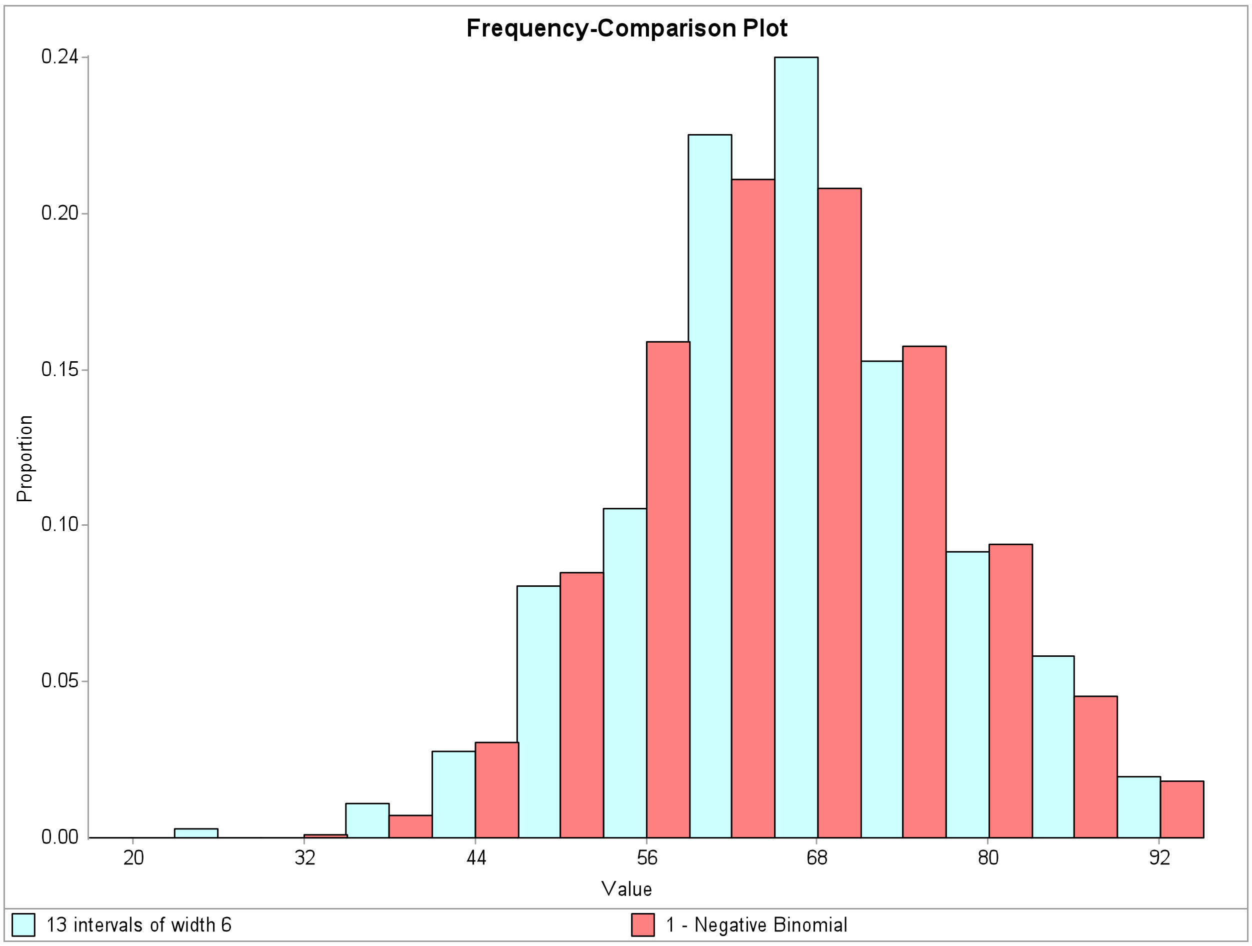
Data Characteristic	Value
Source file	PROB621a
Observation type	Integer valued
Number of observations	369
Minimum observation	26
Maximum observation	96
Mean	68.466
Median	68
Variance	124.500
Lexis ratio (var./mean)	1.818
Skewness	-0.137

Since the lexis ratio (see p. 322 in the book) is greater than 1, this suggests that negative binomial distribution will be a better representation of the data than the binomial or Poisson distributions. Note also that the data are slightly skewed to the left.

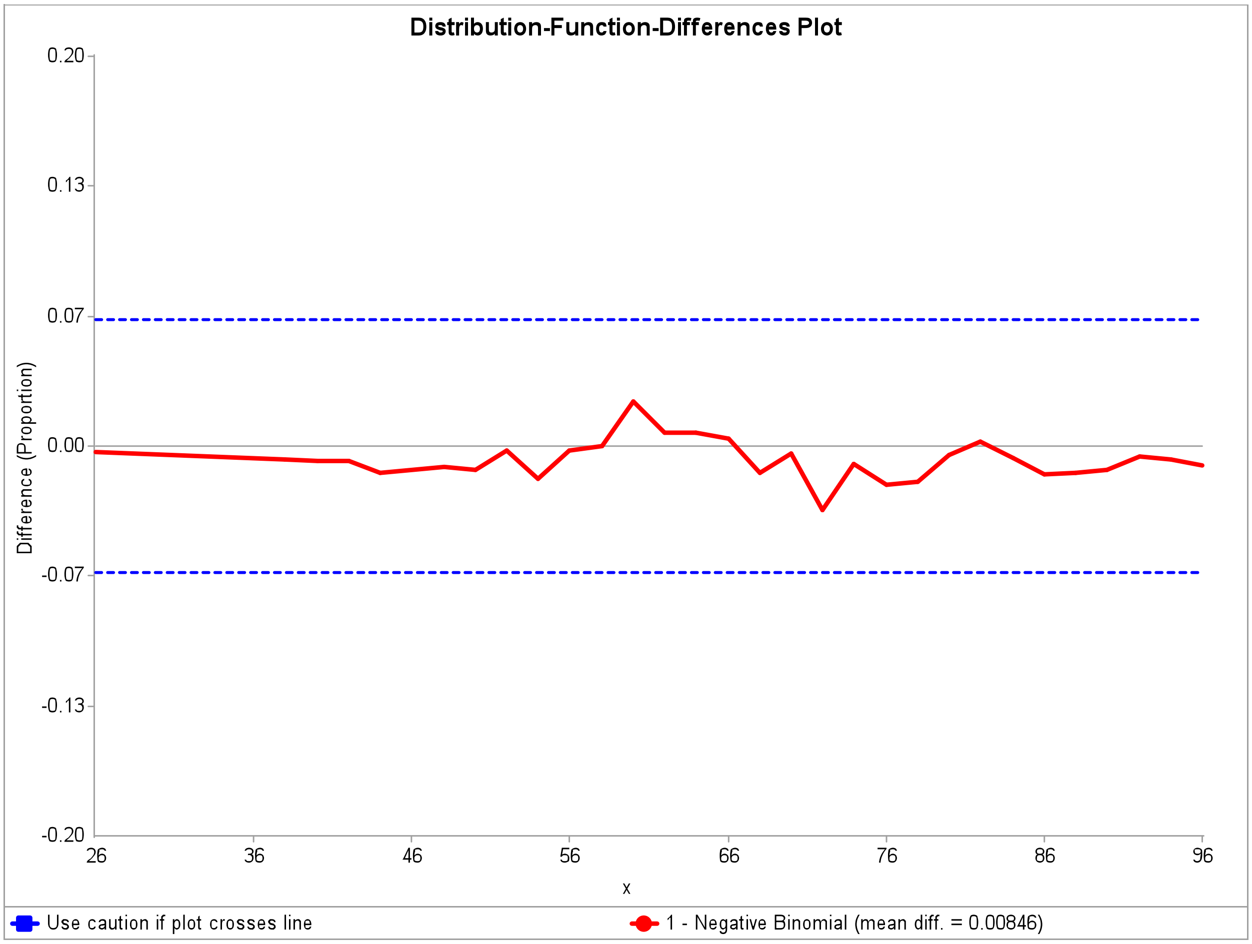


The histogram starts at 20 and is based on 13 intervals of width 6, which was determined by trial and error. Note that the histogram is quite symmetric, which is consistent with the sample skewness.

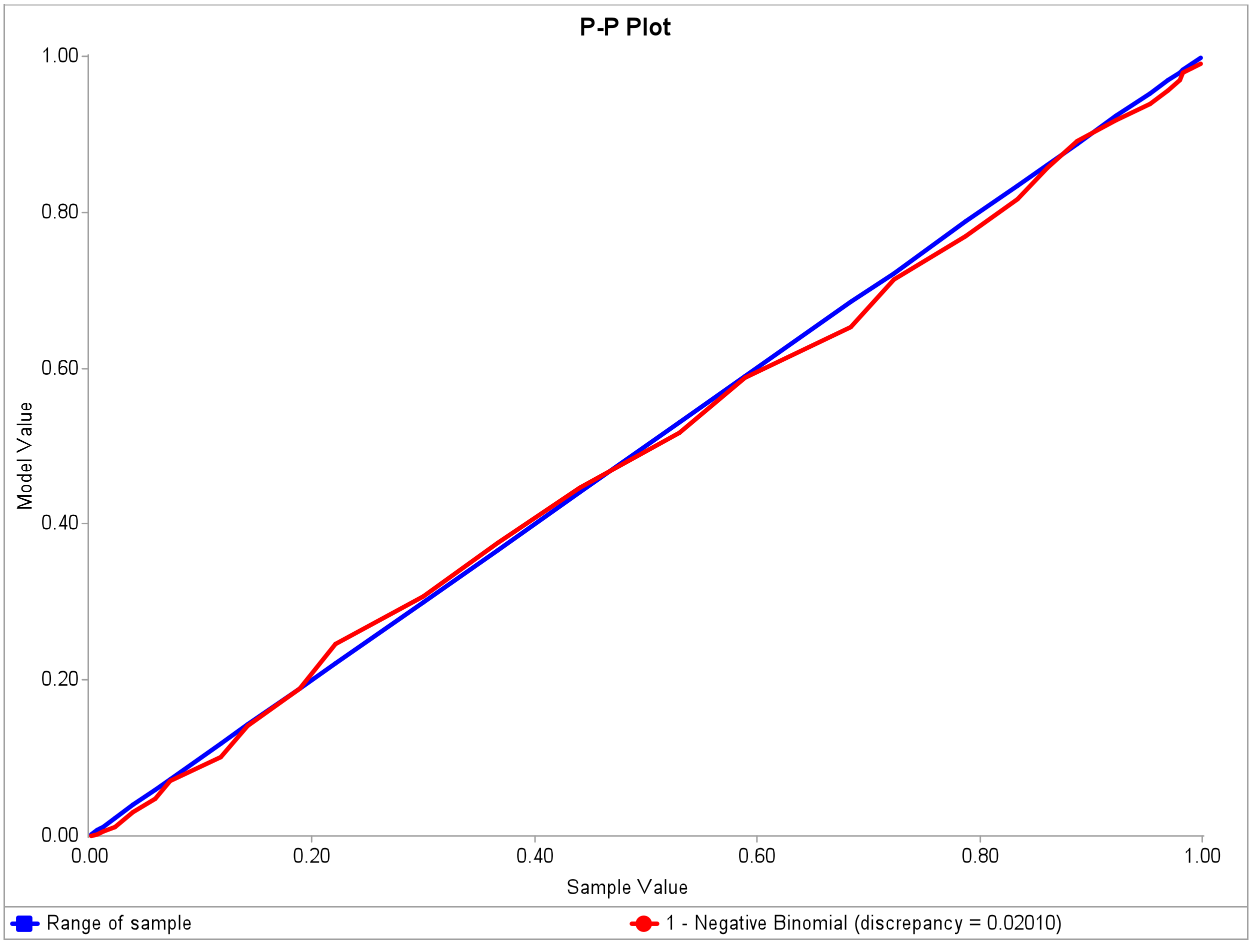
The best-fitting distribution is indeed the negative binomial distribution (probability = 0.542, success = 82) with "Relative Score" = 100.00 and "Absolute Evaluation" = "Intermediate." However, the Absolute Evaluation for discrete distributions is *conservative* in giving an evaluation of "Good" or "Bad." These evaluations will only be given if there is strong supporting evidence. (See the "Help" for the "Automated-Fitting Results" screen.) A distribution might provide a good fit for a discrete data set, but still get an evaluation of "Indeterminate." The recommended plots and the chi-square test (available in the "Comparisons" tab) should be used to show that it provides a reasonable representation for the data.



The frequency-comparison plot indicates that the negative binomial distribution fits the data reasonably well, except perhaps for x between 65 and 71.



The distribution-function-differences plot gives no reason to be concerned about the quality of representation provided by the negative binomial distribution.



The distribution-function-differences plot gives no reason to be concerned about the quality of representation provided by the negative binomial distribution.

For the chi-square test, we used the intervals {1, 2, …, 6}, {7}, {8}, {9}, {10}, and {11, 12, 13}, each of which has an expected (model) count greater than the recommended value of 5 (see p. 349 in the book); the expected counts are also "roughly" equal in size. Since the test-statistic value, 9.926, is less than the critical value (point) of 11.070 for 5 df and $\alpha = 0.05$, we fail to reject the null hypothesis that the data have a negative binomial distribution.

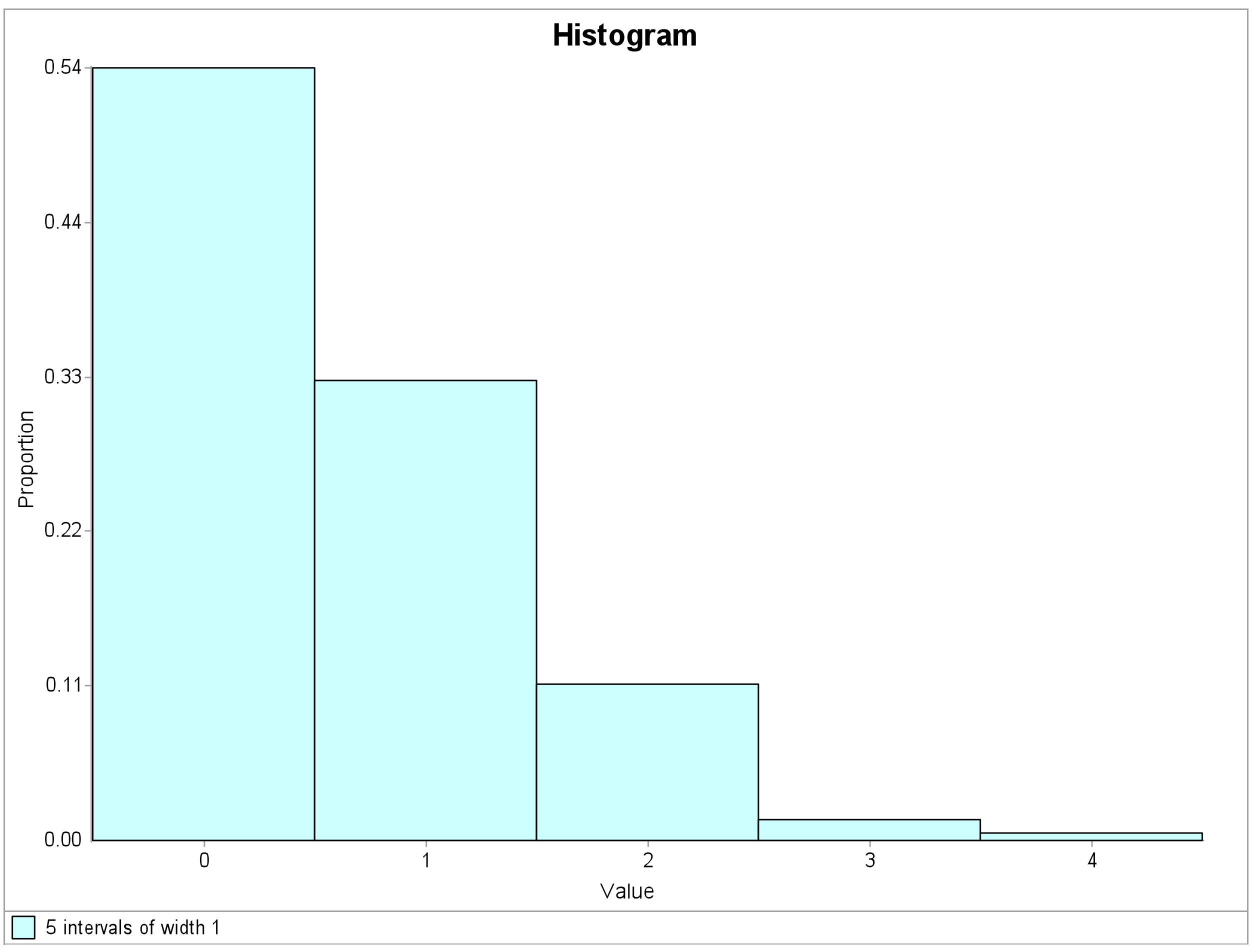
In summary, it appears that the negative binomial distribution provides a good representation for the university test scores.

(a) 200 deaths per corps per year by horse kick in the Prussian army in the late 1800s

Consider the data to be integer valued.

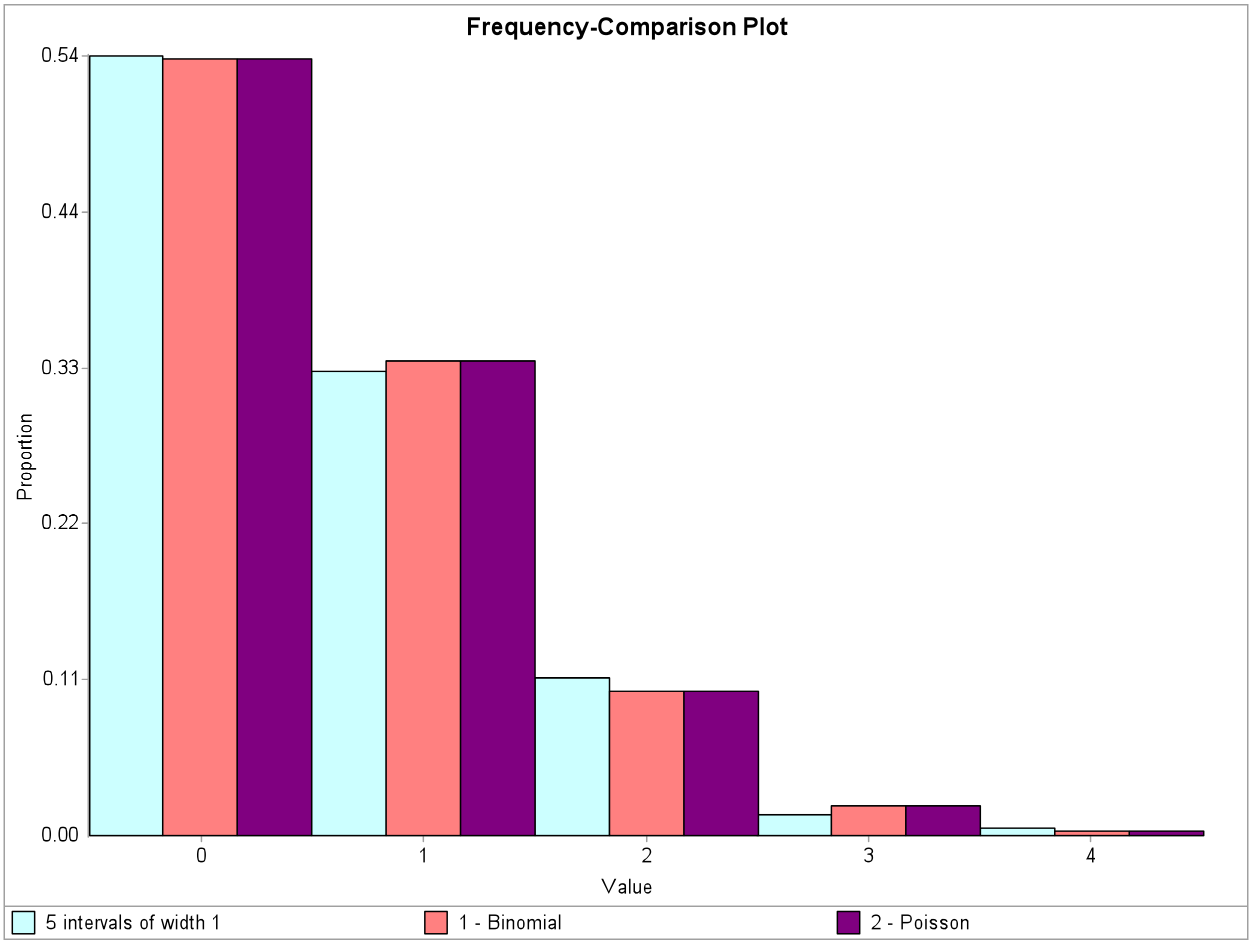
Data Characteristic	Value
Source file	PROB621b
Observation type	Integer valued
Number of observations	200
Minimum observation	0
Maximum observation	4
Mean	0.610
Median	0
Variance	0.611
Lexis ratio (var./mean)	1.002
Skewness	1.255

Since the lexis ratio (see p. 322 in the book) is extremely close to 1, this suggests that Poisson distribution will be a good representation of the data. Note also that the data are skewed to the right.

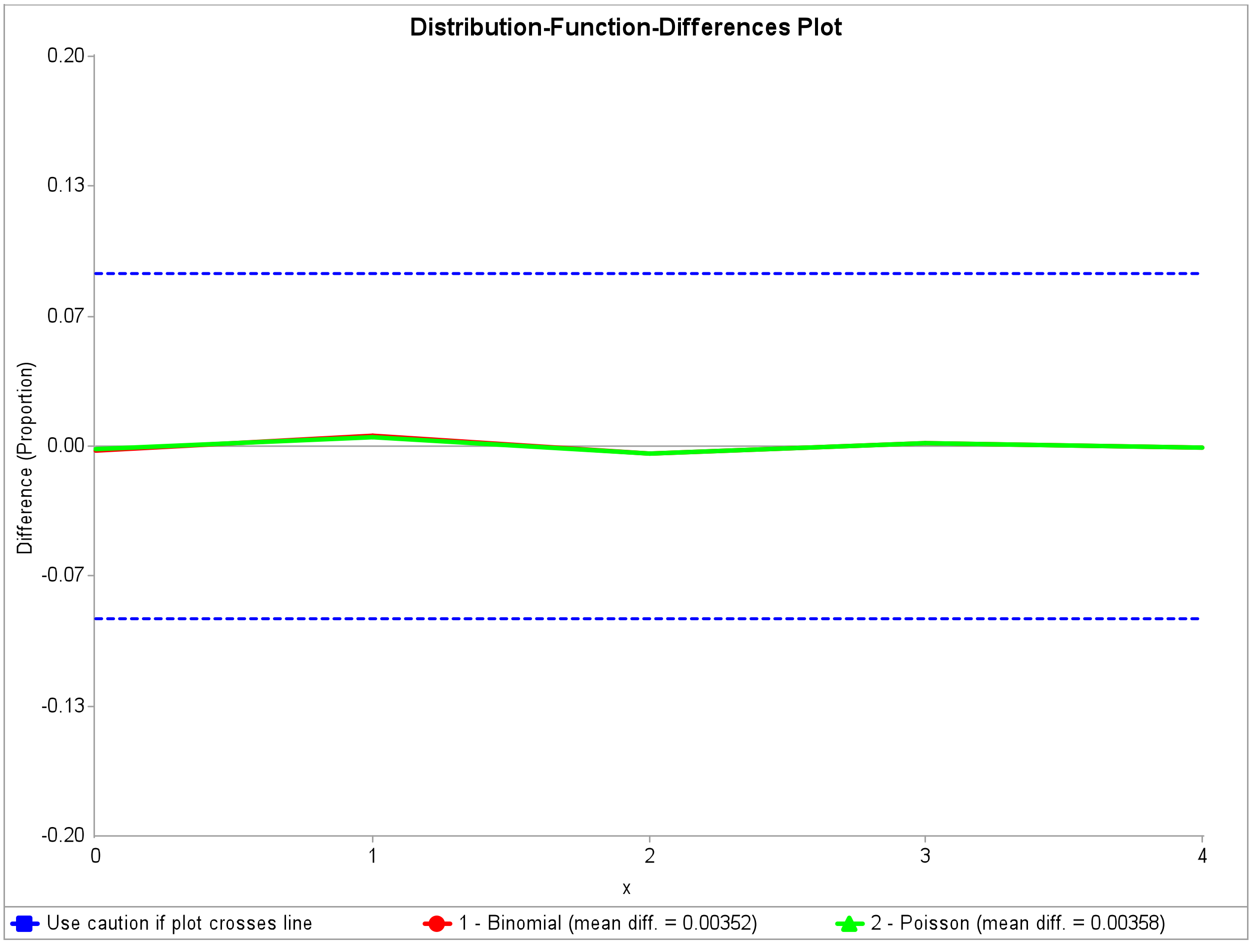


The histogram starts at 0 and is based on 5 intervals of width 1.

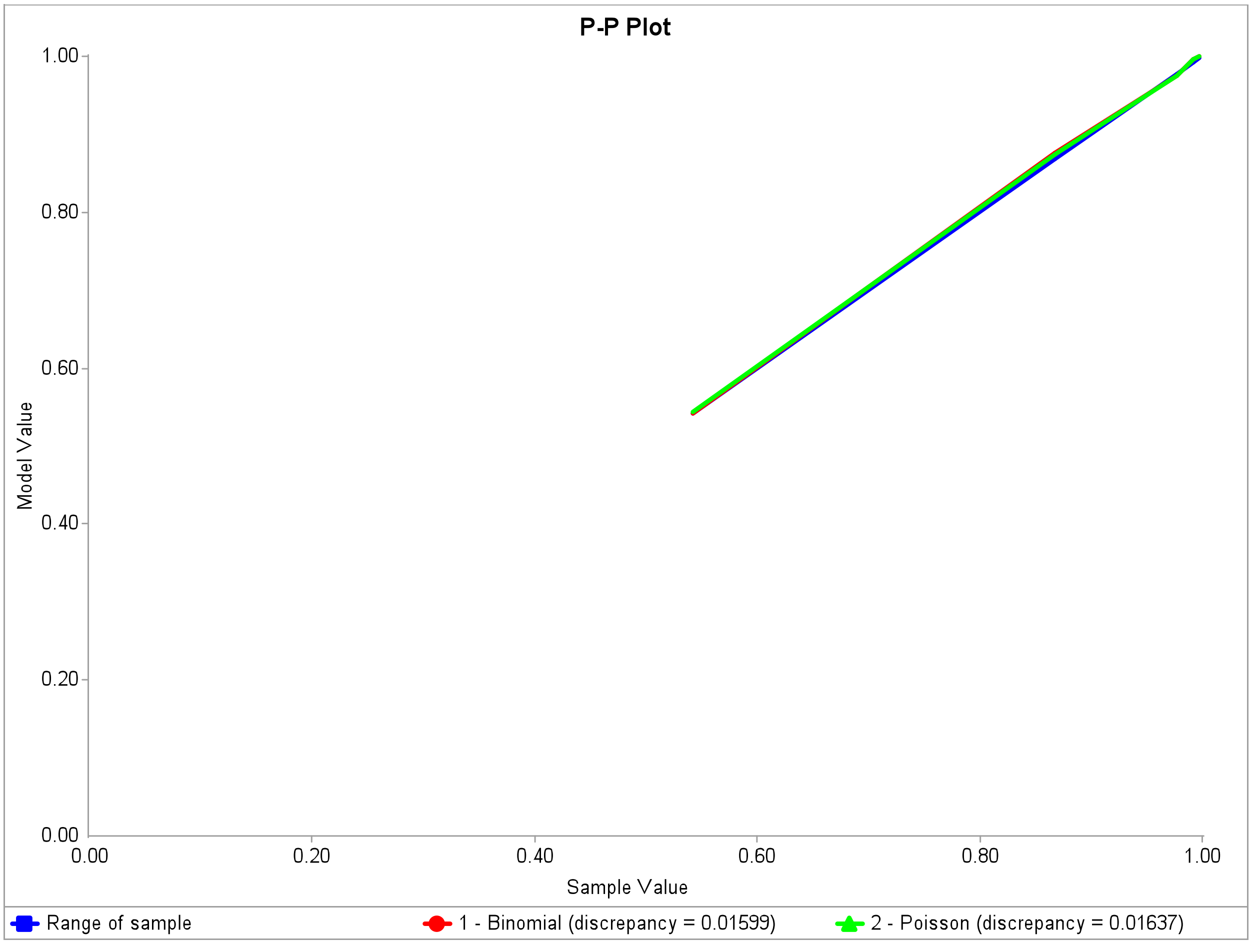
The best-fitting distribution is the binomial distribution (probability = 0.004, trials = 173) with “Relative Score” = 88.89 and “Absolute Evaluation” = “Good.” An Absolute Evaluation of Good will only be given for a discrete distribution if there is strong supporting evidence. (See the “Help” for the “Automated-Fitting Results” screen.) However, we will analyze both the fitted binomial and Poisson distributions since they are almost identical models.



The frequency-comparison plot indicates that the two distributions fit the data equally well.



The distribution-function-differences plot indicates that the two distributions fit the data equally well.



The *P-P* plots indicate that the two distributions fit the data equally well.

For the chi-square test in the case of the binomial distribution, we used the intervals {1}, {2}, {3}, and {4, 5}, since four intervals is the smallest number allowed by ExpertFit. (If k is the number of intervals and m is the number of estimated parameters, then ExpertFit requires that $k - m - 1 = k - 3 \geq 1$; see p. 348 in the book.) Since the test-statistic value, 0.315, is less than the critical value of 7.815 for 3 df and $\alpha = 0.05$, we fail to reject the null hypothesis that the data have a binomial distribution.

For the chi-square test in the case of the Poisson distribution, we used the intervals {1}, {2}, and {3, 4, 5}, since the expected counts for these three intervals are more similar than for the four intervals above. (ExpertFit requires that $k - m - 1 = k - 2 \geq 1$; see p. 348 in the book.) Since the test-statistic value, 0.063, is less than the critical value of 5.991 for 2 df and $\alpha = 0.05$, we fail to reject the null hypothesis that the data have a Poisson distribution.

Both the binomial and Poisson distributions fit the horse-kick data very well. For an explanation of why the two distributions are essentially identical for these data, see Devore (2008, p. 123).

6.22. Enter each of the data sets PROB622a and PROB622b into an element of the "Project Directory." Open the first data set and select "Advanced" from the "Mode" menu at the top of the screen. Select the "Homogeneity Tests" option. Highlight the second data set and click on "Apply." The Kruskal-Wallis test statistic is 26.786, which is larger than the critical value, 3.841, for 1 df and $\alpha = 0.05$. Therefore, we *reject* the null hypothesis that both population distribution functions are the same.

6.23. Let "iff" mean "if and only if." Then

$$
\begin{aligned}
\hat{\gamma} &< X_{(1)} && \text{iff} \\
\frac{X_{(1)}X_{(n)} - X_{(k)}^2}{X_{(1)} + X_{(n)} - 2X_{(k)}} &< \frac{X_{(1)}^2 + X_{(1)}X_{(n)} - 2X_{(1)}X_{(k)}}{X_{(1)} + X_{(n)} - 2X_{(k)}} && \text{iff} \\
0 &< \frac{X_{(1)}^2 - 2X_{(1)}X_{(k)} + X_{(k)}^2}{X_{(1)} + X_{(n)} - 2X_{(k)}} && \text{iff} \\
0 &< \frac{\left[X_{(1)} - X_{(k)}\right]^2}{X_{(1)} + X_{(n)} - 2X_{(k)}} && \text{iff} \\
0 &< X_{(1)} + X_{(n)} - 2X_{(k)} && \text{iff} \\
X_{(k)} &< \left[X_{(1)} + X_{(n)}\right]/2
\end{aligned}
$$

6.24. (*a*) First assume that $X \sim \text{LN}(\gamma, \mu, \sigma^2)$. Then for any $x > 0$,

$$P(X - \gamma \le x) = P(X \le x + \gamma)$$
$$= \int_{\gamma}^{x+\gamma} \exp\left\{-[\ln(t-\gamma) - \mu]^2 \Big/ \left(2\sigma^2\right)\right\} \Big/ \left[(t-\gamma)\sqrt{2\pi\sigma^2}\right] dt$$
$$= \int_0^x \exp\left[-(\ln y - \mu)^2 \Big/ \left(2\sigma^2\right)\right] \Big/ \left(y\sqrt{2\pi\sigma^2}\right) dy$$

where the integrand is the $\text{LN}(\mu, \sigma^2)$ density. (The last equality is obtained by using the change of variable $y = t - \gamma$ in the integral.) Now assume that $X - \gamma \sim \text{LN}(\mu, \sigma^2)$. Then for $x > \gamma$,

$$P(X \le x) = P(X - \gamma \le x - \gamma)$$
$$= \int_0^{x-\gamma} \exp\left[-(\ln t - \mu)^2 \Big/ \left(2\sigma^2\right)\right] \Big/ \left(t\sqrt{2\pi\sigma^2}\right) dt$$
$$= \int_{\gamma}^{x} \exp\left\{-[\ln(y-\gamma) - \mu]^2 \Big/ \left(2\sigma^2\right)\right\} \Big/ \left[(y-\gamma)\sqrt{2\pi\sigma^2}\right] dy$$

where the integrand is the $\text{LN}(\gamma, \mu, \sigma^2)$ density. (This time the change of variable $y = t + \gamma$ was used in the integral.)

(*b*) If $X_1, X_2, \ldots, X_n$ are IID $\text{LN}(\gamma, \mu, \sigma^2)$ random variables with γ known, then the $(X_i - \gamma)$'s are IID $\text{LN}(\mu, \sigma^2)$ by part (*a*). Thus, $\hat{\mu}$ and $\hat{\sigma}$ are as given by part (*e*) of Prob. 6.10, with X_i replaced by $X_i - \gamma$.

6.25. The average arrival rate over the interval $(t, t + s]$ is $\bar{\lambda}(t,s) = \int_t^{t+s} \lambda(y)\, dy \big/ s$. Therefore, the expected number of arrivals in this interval, $b(t, s)$, should be the interval length s times $\bar{\lambda}(t,s)$, or $\int_t^{t+s} \lambda(y)\, dy$.

6.26. (*a*) P(getting a value of X near X_i) = $P(X_i - \varepsilon < X < X_i + \varepsilon) = \int_{X_i-\varepsilon}^{X_i+\varepsilon} f_\theta(x)\,dx \approx 2\varepsilon f_\theta(X_i)$, where the final approximate equality is by the mean value theorem, since X_i is within the interval of integration, which is of width 2ε.

(*b*) P(getting a sample of n IID values of X near the observed data)
= P(getting a value of X near X_1, getting a value of X near X_2, ..., getting a value of X near X_n)
= P(getting a value of X near X_1)P(getting a value of X near X_2)$\cdots P$(getting a value of X near X_n)
(since the values of X are independent)
$\approx (2\varepsilon)^n f_\theta(X_1) f_\theta(X_2) \cdots f_\theta(X_n)$ (by applying part (*a*) to each factor)
$= (2\varepsilon)^n L(\theta)$
Since $(2\varepsilon)^n$ is a constant, P(getting a sample of n IID values of X near the observed data) is proportional to the likelihood function.

(*c*) By definition, the MLE $\hat{\theta}$ maximizes $L(\theta)$, so it will also maximize the approximate value of P(getting a sample of n IID values of X near the observed data) since the latter quantity is approximately proportional to $L(\theta)$, from (*b*).

6.27. For Example 6.1 we have that the arrival rate is $\lambda = 1$. Furthermore, from Sec. 1B.3 we know that $Q = \lambda d$, where Q is the steady-state average number in queue and d is the steady-state average delay in queue. Therefore, for a "sufficiently long" simulation run, the average delay in queue should be approximately equal to the average number in queue.

6.28.

$$\begin{aligned} P(X \le \gamma + e^{\mu + z_q\sigma}) &= P(e^Y \le e^{\mu + z_q\sigma}) \\ &= P(Y \le \mu + z_q\sigma) \\ &= P\left(\frac{Y-\mu}{\sigma} \le z_q\right) \\ &= P(X \le x_q) \end{aligned}$$

Therefore,

$$x_q = \gamma + e^{\mu + z_q\sigma}$$

6.29. Newton's method is a recursive procedure for finding the root of the equation $f(\alpha) = 0$. Let $a = (m-\gamma)/(x_q - \gamma)$ and $b = \ln[1/(1-q)]$. In our case, $f(\alpha)$ is given by

$$f(\alpha) = a - \left(\frac{\alpha - 1}{b\alpha}\right)^{1/\alpha}$$

Furthermore,

$$f'(\alpha) = \frac{-1}{\alpha}\left(\frac{\alpha-1}{b\alpha}\right)^{(1/\alpha)-1}\left(\frac{b}{\alpha^2}\right) = \frac{-b}{\alpha^3}\left(\frac{\alpha-1}{b\alpha}\right)^{(1/\alpha)-1}$$

Then the general recursive step for Newton's method is

$$\hat{\alpha}_{k+1} = \hat{\alpha}_k - \frac{f(\hat{\alpha}_k)}{f'(\hat{\alpha}_k)}$$

6.30. Assume $q > 0.5$, so that $x_q > m$. Then

$$q = F(x_q) = 1 - \frac{(b - x_q)^2}{(b-a)(b-m)}$$

or

$$(b - x_q)^2 = (1-q)(b-a)(b-m)$$

or

$$qb^2 + [(1-q)(a+m) - 2x_q]b + [x_q^2 - (1-q)am] = 0$$

which is a quadratic equation that can be solved for the maximum value *b*.

6.31. No. For example, let $a = 0$, $b = 1$, and $\mu = 0.9$. Since the mean of a triangular distribution is $\mu = (a + b + m)/3$, this implies that $m = 1.7$, which makes no sense. Since $a \le m$, this implies that $(2a + b)/3 \le \mu$, with the left side of the inequality being one-third of the way from a to b. Since $m \le b$, this implies that $\mu \le (a + 2b)/3$, with the right side of the inequality being two-thirds of the way from a to b. Thus, in general, we must have

$$\frac{2a+b}{3} \le \mu \le \frac{a+2b}{3}$$