ISyE 4803 - Advanced Simulation
Spring 2018
Homework 4 - due 3/13/18
Charu Thomas - 903126680

1. $f(x;p) = p(1-x)^{p-1}, \ 0 < x < 1, \ p > 0$ $\quad u = 1-x \quad du = -dx \quad u(0) = 1 \quad u(1) = 0$

(a) $E[X] = \int_0^1 x\, p(1-x)^{p-1}\, dx = \int_0^1 (1-u)p(u)^{p-1}\, du$

$$= p\int_0^1 (1-u)u^{p-1}\,du = p\int_0^1 (u^{p-1} - u^p)\,du$$

$$= p\left[\frac{u^p}{p} - \frac{u^{p+1}}{p+1}\right]_{u=0}^{u=1} = p\left[\frac{1}{p} - \frac{1}{p+1}\right] = p\left[\frac{(p+1)-p}{p(p+1)}\right] = \frac{1}{p+1}$$

$$\bar{x} = \frac{1}{\hat{p}+1} \rightarrow \hat{p}+1 = \frac{1}{\bar{x}} \rightarrow \hat{p} = \frac{1}{\bar{x}} - 1$$

(b) $L(x_i;p) = \prod_{i=1}^{n} p(1-x_i)^{p-1}$

$$\ell(x_i;p) = \ln L(x_i;p) = \sum_{i=1}^{n} \ln\left(p(1-x_i)^{p-1}\right)$$

$$= n\ln(p) + (p-1)\sum_{i=1}^{n} \ln(1-x_i)$$

$$\frac{\partial \ell}{\partial p} = \frac{n}{\hat{p}} + \sum_{i=1}^{n} \ln(1-x_i) = 0$$

$$\hookrightarrow \frac{n}{\hat{p}} = -\sum_{i=1}^{n} \ln(1-x_i)$$

$$\rightarrow \hat{p} = \frac{n}{-\sum_{i=1}^{n} \ln(1-x_i)}$$

(c)

$x_{(i)}$	.08	.09	.10	.11	.24	.26	.35	.37	.39	.47
$\hat{F}(x_{(i)})$	2.5392	2.484	2.43	2.376	1.733	1.643	1.268	1.1907	1.116	0.84
$\frac{i}{10} - \hat{F}(x_{(i)})$	-2.439	-2.284	-2.13	-1.976	-1.233	-1.043	-0.568	-0.391	-0.216	0.157
$\hat{F}(x_{(i)}) - \frac{i-1}{10}$	2.5392	2.384	2.23	2.076	1.333	1.143	0.668	0.491	0.316	-0.05

$D_{10} = 2.5392$

Modified Test Statistic $= \left(\sqrt{10} + .12 + \frac{.11}{\sqrt{10}}\right)(2.5392) = 8.4227$

B/c $8.4227 > c_{1-\alpha}$ for $\alpha \le 0.1$, we reject H_0 and conclude the beta dist'n w/ $p=3$ is not a good fit.

2. (a) $L(\mu,\sigma^2) = \prod_{i=1}^{n} \left((2\pi\sigma^2)^{-1/2} x_i^{-1} \exp\left[-\frac{(\ln(x_i)-\mu)^2}{2\sigma^2}\right]\right)$

$$= (2\pi\sigma^2)^{-n/2} \prod_{i=1}^{n} \left(x_i^{-1} \exp\left[\frac{-(\ln(x_i)-\mu)^2}{2\sigma^2}\right]\right)$$

$$\ln L(\mu,\sigma^2) = \ell(\mu,\sigma^2) = \ln((2\pi\sigma^2)^{-n/2}) - \sum_{i=1}^{n} \ln(x_i) - \frac{\sum_{i=1}^{n}(\ln(x_i)-\mu)^2}{2\sigma^2}$$

$$= -\frac{n}{2}\ln(2\pi\sigma^2) - \sum_{i=1}^{n}\ln(x_i) - \frac{\sum_{i=1}^{n}\ln(x_i)^2}{2\sigma^2} + \frac{\sum_{i=1}^{n}\ln(x_i)\mu}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}$$

$$\frac{\partial \ell}{\partial \mu} = -\frac{n\hat{\mu}}{\sigma^2} + \frac{\sum_{i=1}^{n}\ln(x_i)}{\sigma^2} = 0 \rightarrow \frac{n\hat{\mu}}{\sigma^2} = \frac{\sum_{i=1}^{n}\ln(x_i)}{\sigma^2}$$

$$\boxed{\hat{\mu} = \frac{\sum_{i=1}^{n}\ln(x_i)}{n}}$$

$$\frac{\partial \ell}{\partial \sigma^2} = \frac{-n}{2\hat{\sigma}^2} + \frac{\sum_{i=1}^{n}(\ln(x_i)-\hat{\mu})^2}{2\hat{\sigma}^4} = 0 \rightarrow \frac{n}{2\hat{\sigma}^2} = \frac{\sum_{i=1}^{n}(\ln(x_i)-\hat{\mu})^2}{2\hat{\sigma}^4}$$

$$\hookrightarrow n\hat{\sigma}^2 = \sum_{i=1}^{n}(\ln(x_i)-\hat{\mu})^2 \rightarrow \boxed{\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(\ln(x_i) - \frac{\sum_{i=1}^{n}\ln(x_i)}{n}\right)^2}{n}}$$

From MOM, $\tilde{\mu} = -\frac{\ln(\sum_{i=1}^{n} x_i^2)}{2} + 2\ln(\sum_{i=1}^{n} x_i) - \frac{3}{2}\ln(n)$

$$\tilde{\sigma}^2 = \ln\left(\sum_{i=1}^{n} x_i^2\right) - 2\ln\left(\sum_{i=1}^{n} x_i\right) + \ln(n)$$

compared to MOM estimators, the MLE estimators are asymptotically normal.

(b) $L(t,p) = \prod_{i=1}^{t} p^{x_i}(1-p)^{t-x_i}$ ← B/c for particular x_i, $\left(\frac{t!}{x_i!(t-x_i)!}\right)$ is the same and doesn't impact likelihood

$$\ln L(t,p) = \ell(t,p) = \sum_{i=1}^{t} \ln\left(p^{x_i}(1-p)^{t-x_i}\right)$$

$$= \sum_{i=1}^{t}\left[x_i \ln(p)\right] + \sum_{i=1}^{t}\left[(t-x_i)\ln(1-p)\right]$$

$$\frac{d\ell}{dp} = \frac{\sum_{i=1}^{t} x_i}{\hat{p}} - \frac{t - \sum_{i=1}^{t} x_i}{1-\hat{p}} = 0 \rightarrow \frac{t - \sum_{i=1}^{t} x_i}{1-\hat{p}} = \frac{\sum_{i=1}^{t} x_i}{\hat{p}}$$

$$\hookrightarrow \hat{p}t - \hat{p}\sum_{i=1}^{t} x_i = \sum_{i=1}^{t} x_i - \hat{p}\sum_{i=1}^{t} x_i \rightarrow \boxed{\hat{p} = \frac{\sum_{i=1}^{t} x_i}{t}}$$

For MOM, $\tilde{p} = \frac{\bar{x}}{t}$

MOM has an extra factor of $1/t$ when compared to MLE.

(c) $L(p,s) = \prod_{i=1}^{n} (1-p)^s p^{x_i}$ ← combination term doesn't impact likelihood, so we omit without loss of generality

$$\ln L(p,s) = \ell(p,s) = sn\ln(1-p) + \ln(p)\sum_{i=1}^{n} x_i$$

$$\frac{d\ell}{dp} = \frac{-sn}{1-\hat{p}} + \frac{\sum_{i=1}^{n} x_i}{\hat{p}} = 0 \rightarrow \frac{\sum_{i=1}^{n} x_i}{\hat{p}} = \frac{sn}{1-\hat{p}}$$

$$\hookrightarrow \sum_{i=1}^{n} x_i - \hat{p}\sum_{i=1}^{n} x_i = \hat{p}sn \rightarrow \boxed{\hat{p} = \frac{\sum_{i=1}^{n} x_i}{(\sum_{i=1}^{n} x_i + sn)}}$$

For MOM, $\tilde{p} = \frac{\bar{x}}{(s+\bar{x})}$

Multiply MLE estimator by $\frac{n}{n} = 1$.

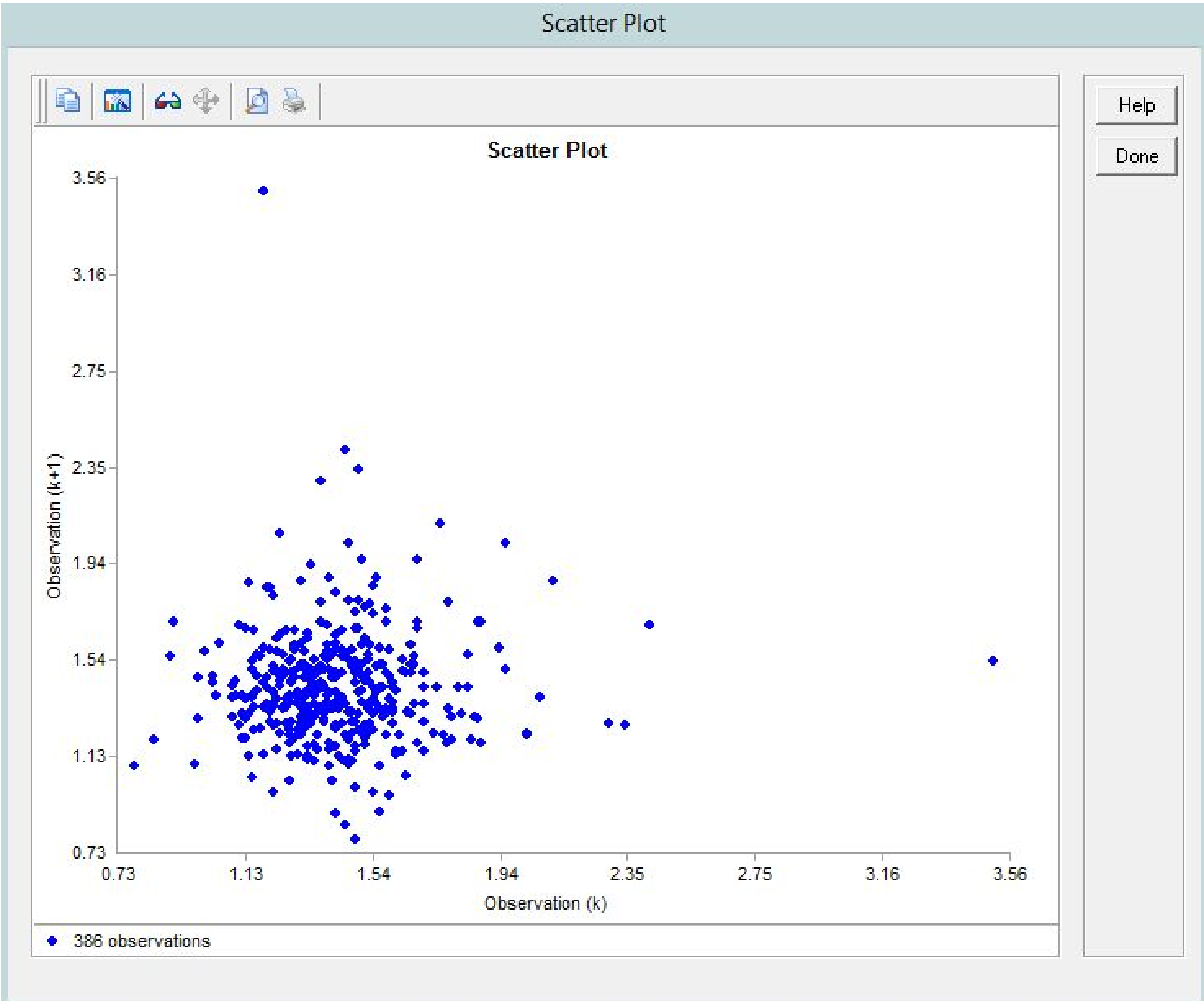
3. $L(\gamma,\beta) = \prod_{i=1}^{n} \frac{1}{\beta}\exp(-(x_i-\gamma)/\beta) = \frac{1}{\beta^n}\prod_{i=1}^{n}\exp(-(x_i-\gamma)/\beta)$

$\ln L(\gamma,\beta) = \ell(\gamma,\beta) = \ln(\frac{1}{\beta}) - \frac{1}{\beta}\sum_{i=1}^{n}(x_i-\gamma)$

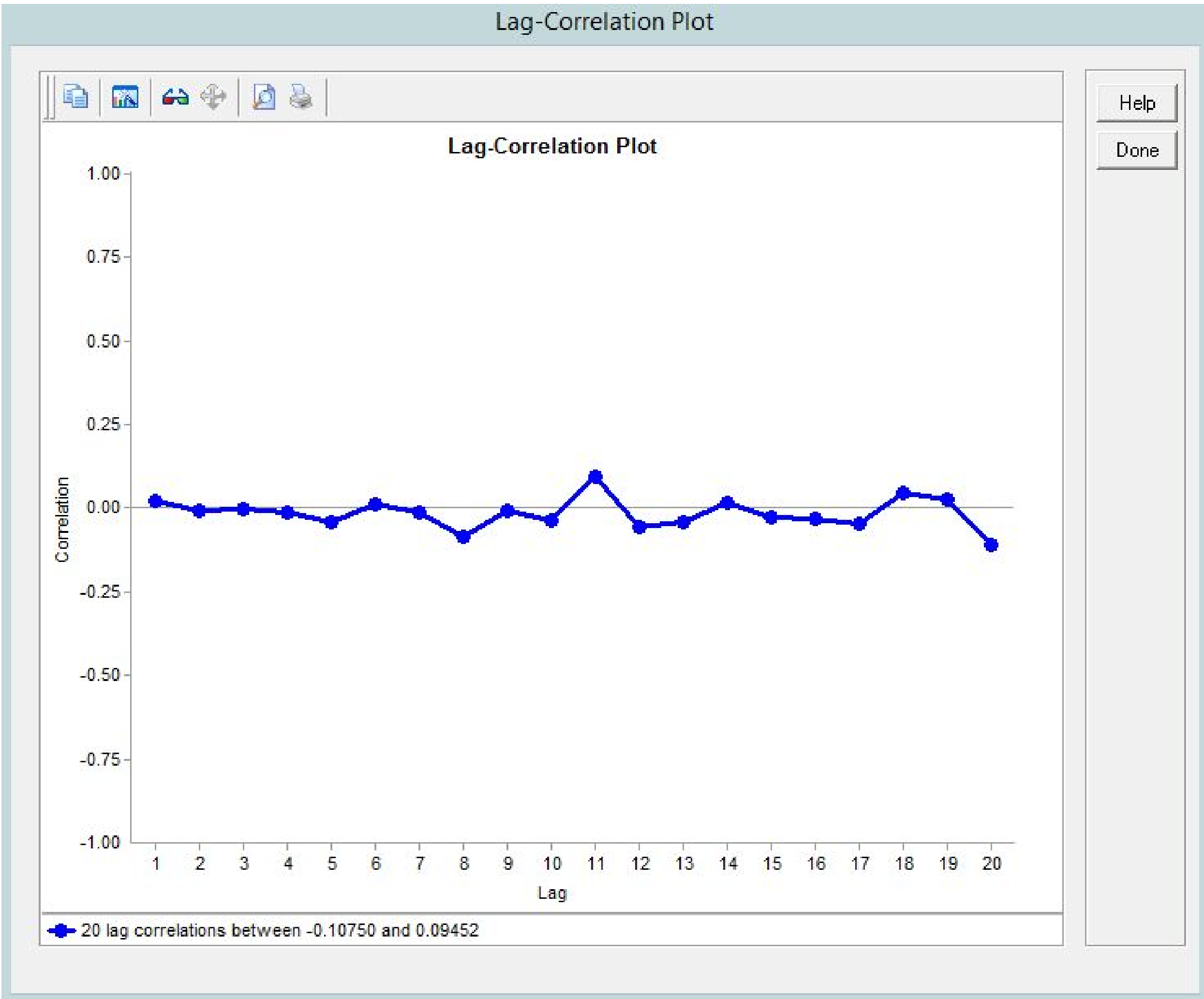
$\frac{\partial\ell}{\partial\gamma} \Rightarrow \hat{\gamma} = \min(x_i)$

$\frac{\partial\ell}{\partial\beta} \Rightarrow \hat{\beta} = \frac{\sum_{i=1}^{n} x_i}{n} - \min(x_i)$

4. Original:



Scatter Plot
Help
Done
Scatter Plot
3.56
3.16
2.75
2.35
1.94
1.54
1.13
0.73
Observation (k+1)
0.73
1.13
1.54
1.94
2.35
2.75
3.16
3.56
Observation (k)
386 observations


Lag-Correlation Plot
Help
Done
Lag-Correlation Plot
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
Correlation
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Lag
20 lag correlations between -0.10750 and 0.09452

Data-Summary Table

Data Characteristic	Value
Source file	<edited>
Observation type	Real valued
Number of observations	386
Minimum observation	0.78000
Maximum observation	3.51000
Mean	1.42448
Median	1.40000
Variance	0.05980
Coefficient of variation	0.17167
Skewness	2.15606

Copy

Print

Help

Done

Automated-Fitting Results

Copy

Print

Help

Done

Relative Evaluation of Candidate Models

Model	Relative Score	Parameters	
1 - Log-Logistic(E)	99.14	Location	0.26079
		Scale	1.14009
		Shape	9.58568
2 - Log-Logistic	97.41	Location	0.00000
		Scale	1.40234
		Shape	11.79421
3 - Pearson Type V	85.34	Location	0.00000
		Scale	56.03732
		Shape	40.35672

30 models are defined with scores between 1.72 and 99.14

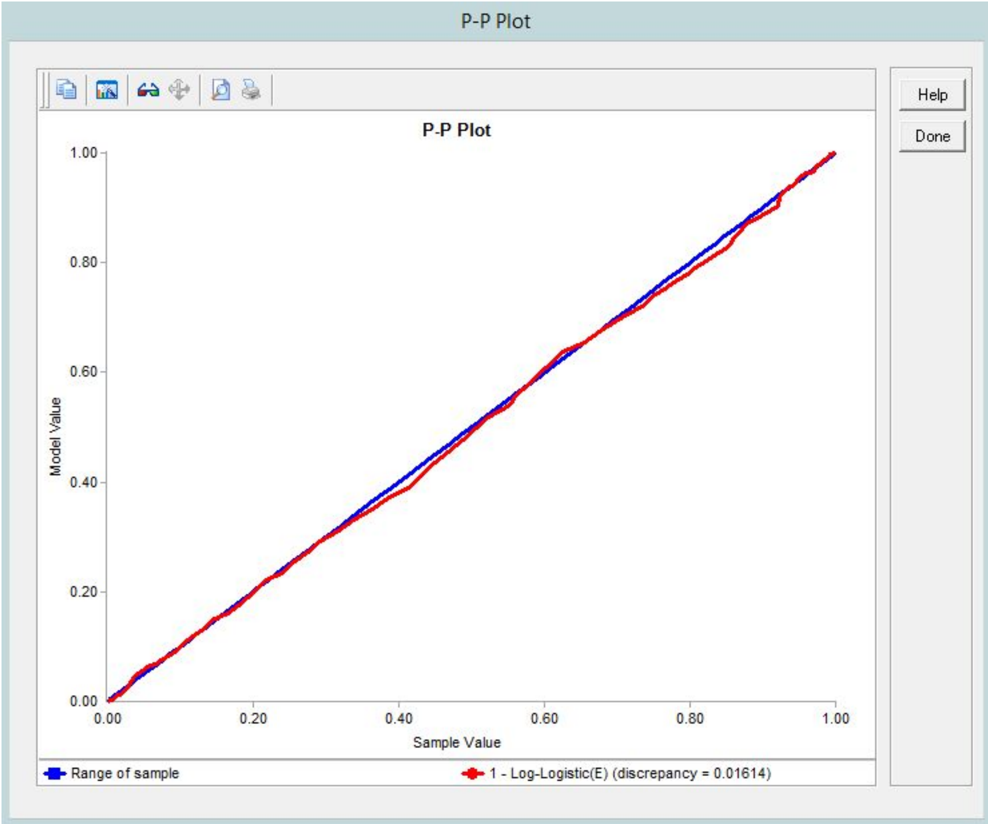
Absolute Evaluation of Model 1 - Log-Logistic(E)

Evaluation: Good

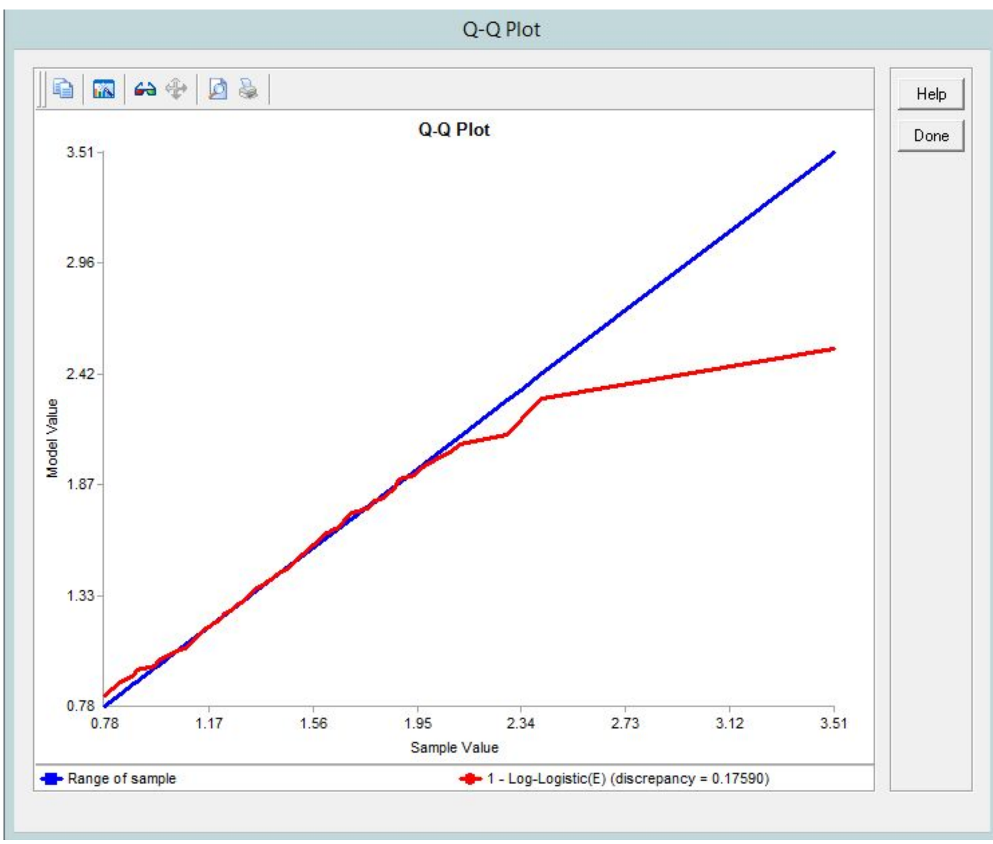
Suggestion: Additional evaluations using Comparisons Tab might be informative.
See Help for more information.

Additional Information about Model 1 - Log-Logistic(E)

Results of the Anderson-Darling goodness-of-fit test at level 0.1	Not applicable
"Error" in the model mean relative to the sample mean	0.00293 = 0.21%



P-P Plot
Help
Done
P-P Plot
Model Value
Sample Value
1.00
0.80
0.60
0.40
0.20
0.00
0.00
0.20
0.40
0.60
0.80
1.00
Range of sample
1 - Log-Logistic(E) (discrepancy = 0.01614)



Q-Q Plot
Help
Done
Q-Q Plot
Model Value
Sample Value
3.51
2.96
2.42
1.87
1.33
0.78
0.78
1.17
1.56
1.95
2.34
2.73
3.12
3.51
Range of sample
1 - Log-Logistic(E) (discrepancy = 0.17590)

Anderson-Darling Test

Anderson-Darling Test with Model 1 - Log-Logistic(E)

Sample size 386
Test statistic 0.25175

Note: No critical values exist for this special case.
The following critical values are for the case where
all parameters are known, and are conservative.

	Critical Values for Level of Significance (alpha)					
Sample Size	0.250	0.100	0.050	0.025	0.010	0.005
386	1.248	1.933	2.492	3.070	3.857	4.500
Reject?	No					

Copy
Print
Help
Done

Kolmogorov-Smirnov Test

Kolmogorov-Smirnov Test with Model 1 - Log-Logistic(E)

Sample size	386
Normal test statistic	0.02852
Modified test statistic	0.56034

Note: No critical values exist for this special case. The following critical values are for the case where all parameters are known, and are conservative.

Sample Size	Critical Values for Level of Significance (alpha) 0.150	0.100	0.050	0.025	0.010
386	1.131	1.216	1.349	1.471	1.618
Reject?	No				

Copy

Print

Help

Done

Equal-Probable Chi-Square Test

Equal-Probable Chi-Square Test with Model 1 - Log-Logistic(E)

Number of intervals	40
Expected (model) count	9.65
Test statistic	62.29016

Warning: The test may not be statistically valid because a method other than maximum likelihood was used to estimate parameters.

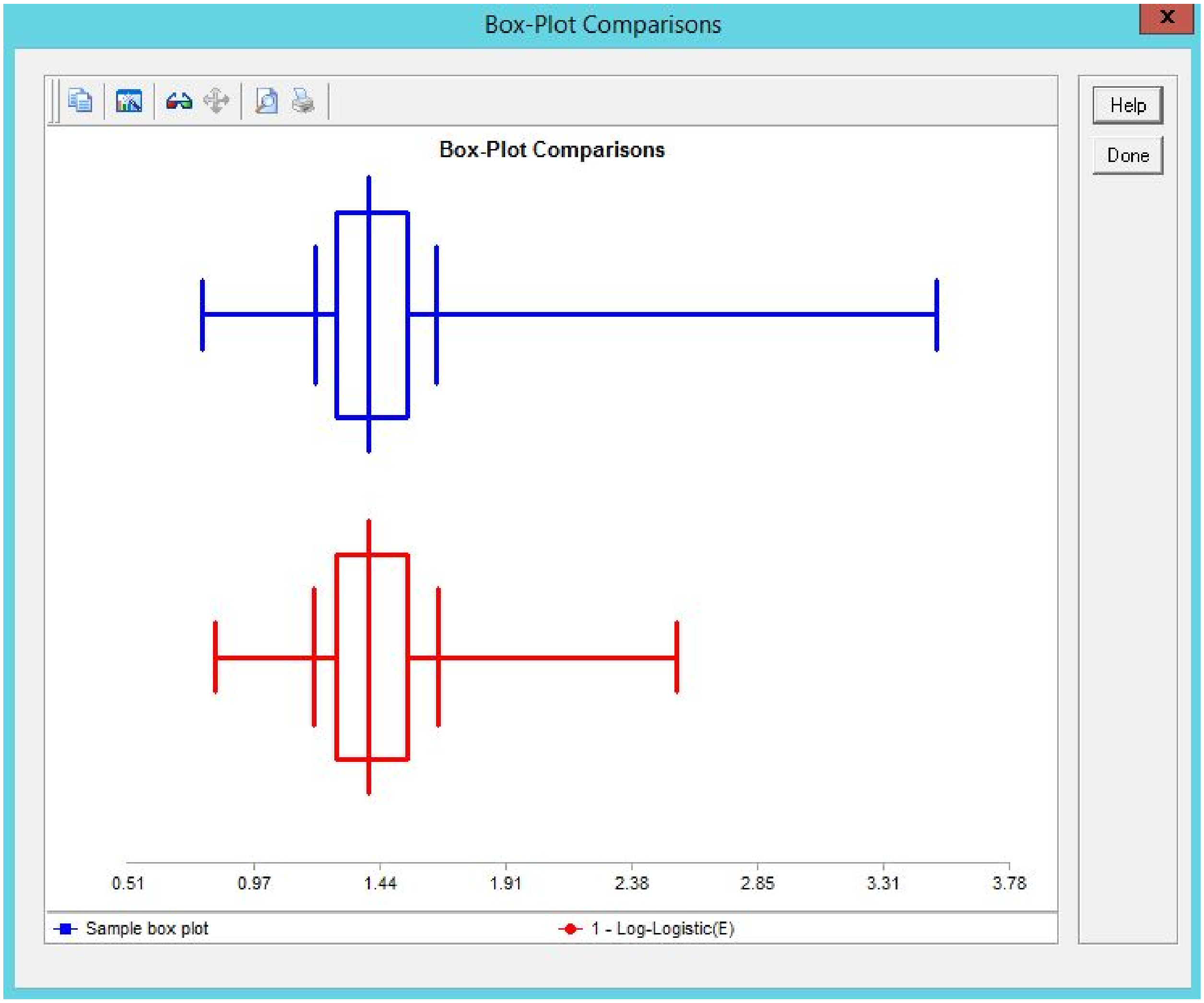
Degrees of Freedom	Observed Level of Significance	Critical Values for Level of Significance (alpha) 0.25	0.15	0.10	0.05	0.01
36	0.004	41.304	44.764	47.212	50.998	58.619
39	0.010	44.539	48.126	50.660	54.572	62.428
	Reject?	Yes				No

Copy

Print

Help

Done



Simio Expression: Random.LogLogistics(9.58568, 1.14009)

This model seems appropriate. The only issue is the model does not handle the right tail of the distribution well. As shown in the Box Plot and on the upper edges of the QQ Plot, the right tail is significantly longer in the observed distribution than the fitted distribution. Correlation plot of lag looks good, fairly centered around 0. The scatter plot shows a few points that may be outliers and may explain the poor fitting at the tails of the distribution.

With a lower bound of 0.5: We get the same results as before with same plots.

Automated-Fitting Results

Copy

Print

Help

Done

Relative Evaluation of Candidate Models

Model	Relative Score	Parameters	
1 - Log-Logistic(E)	98.86	Location	0.26079
		Scale	1.14009
		Shape	9.58568
2 - Log-Logistic	97.73	Location	0.00000
		Scale	1.40234
		Shape	11.79421
3 - Log-Logistic(K)	96.59	Location	0.50000
		Scale	0.89879
		Shape	7.52493

45 models are defined with scores between 1.70 and 98.86

Absolute Evaluation of Model 1 - Log-Logistic(E)

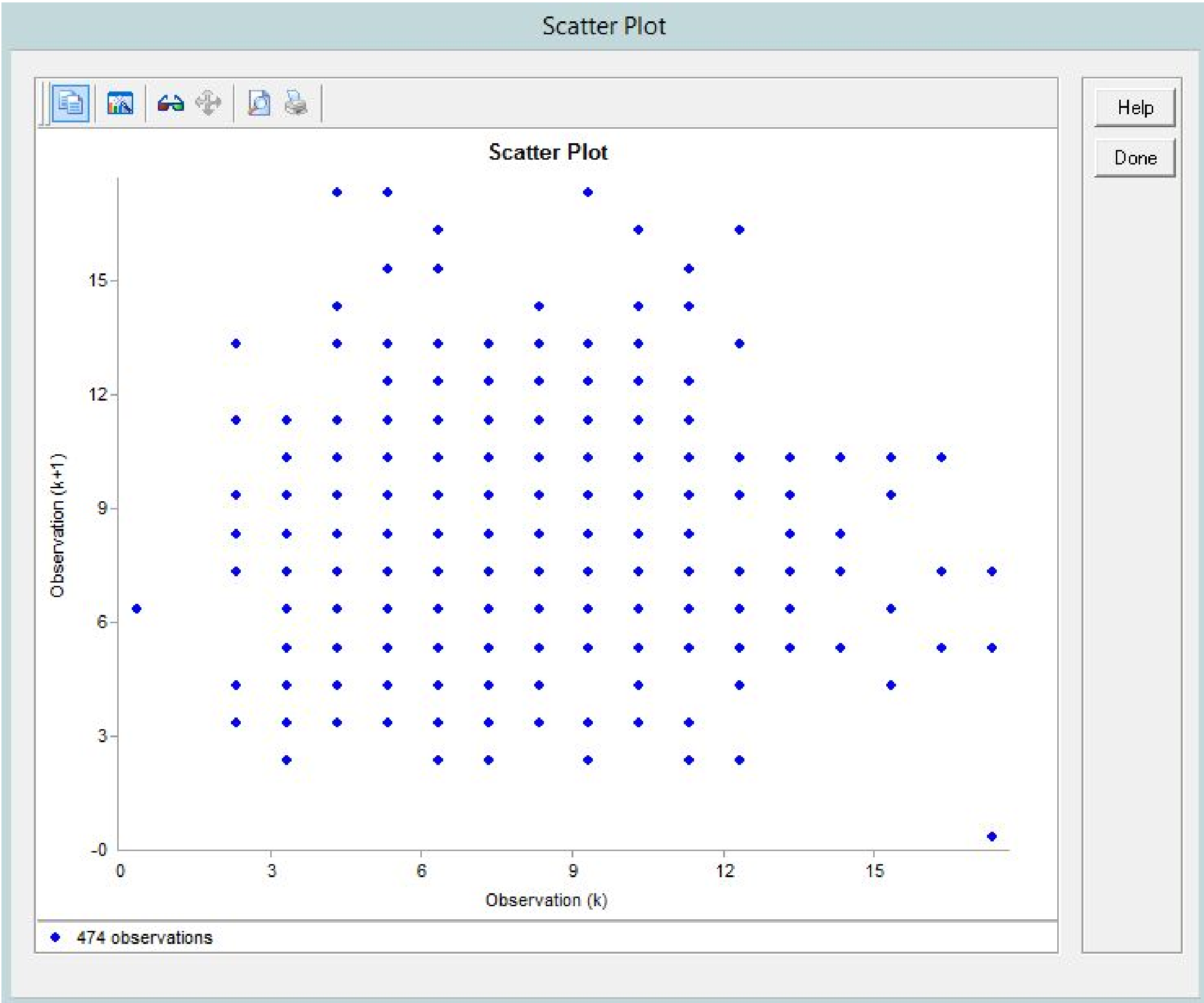
Evaluation: Good

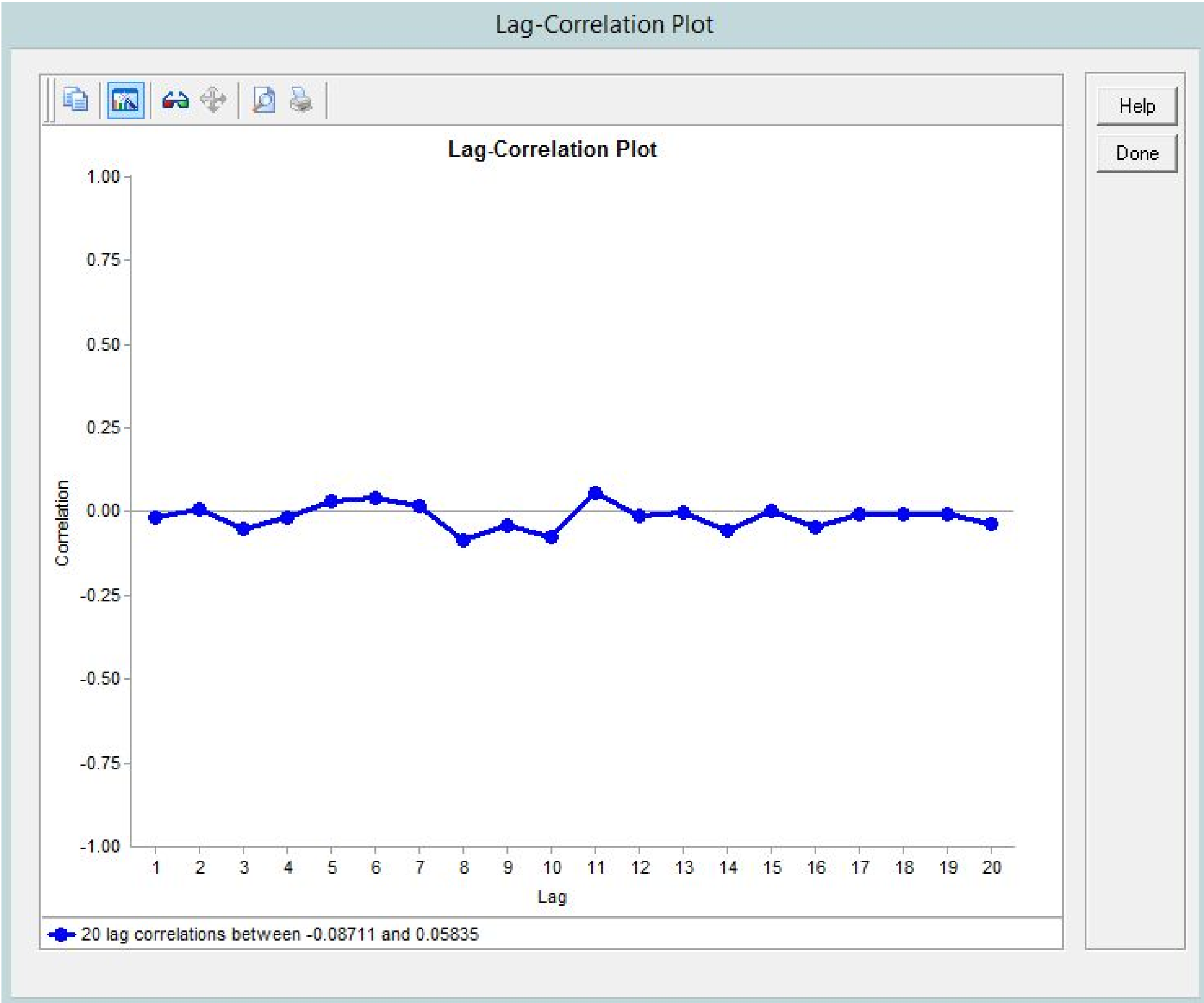
Suggestion: Additional evaluations using Comparisons Tab might be informative.
See Help for more information.

Additional Information about Model 1 - Log-Logistic(E)

Results of the Anderson-Darling goodness-of-fit test at level 0.1	Not applicable
"Error" in the model mean relative to the sample mean	0.00293 = 0.21%

5.


Scatter Plot
Help
Done
Scatter Plot
Observation (k+1)
Observation (k)
474 observations


Lag-Correlation Plot
Help
Done
Lag-Correlation Plot
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
Correlation
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Lag
20 lag correlations between -0.08711 and 0.05835

Automated-Fitting Results

Copy

Print

Help

Done

Relative Evaluation of Candidate Models

Model	Relative Score	Parameters	
1 - Negative Binomial	100.00	Probability	0.85966
		Success	47
2 - Poisson	66.67	Lambda	7.67300
3 - Discrete Uniform	33.33	Lower endpoint	0
		Upper endpoint	17

4 models are defined with scores between 0.00 and 100.00

Absolute Evaluation of Model 1 - Negative Binomial

Evaluation: Good

Suggestion: Additional evaluations using Comparisons Tab might be informative.
See Help for more information.

Additional Information about Model 1 - Negative Binomial

"Error" in the model mean
relative to the sample mean -2.6645e-15 = 0.00%

Data-Summary Table

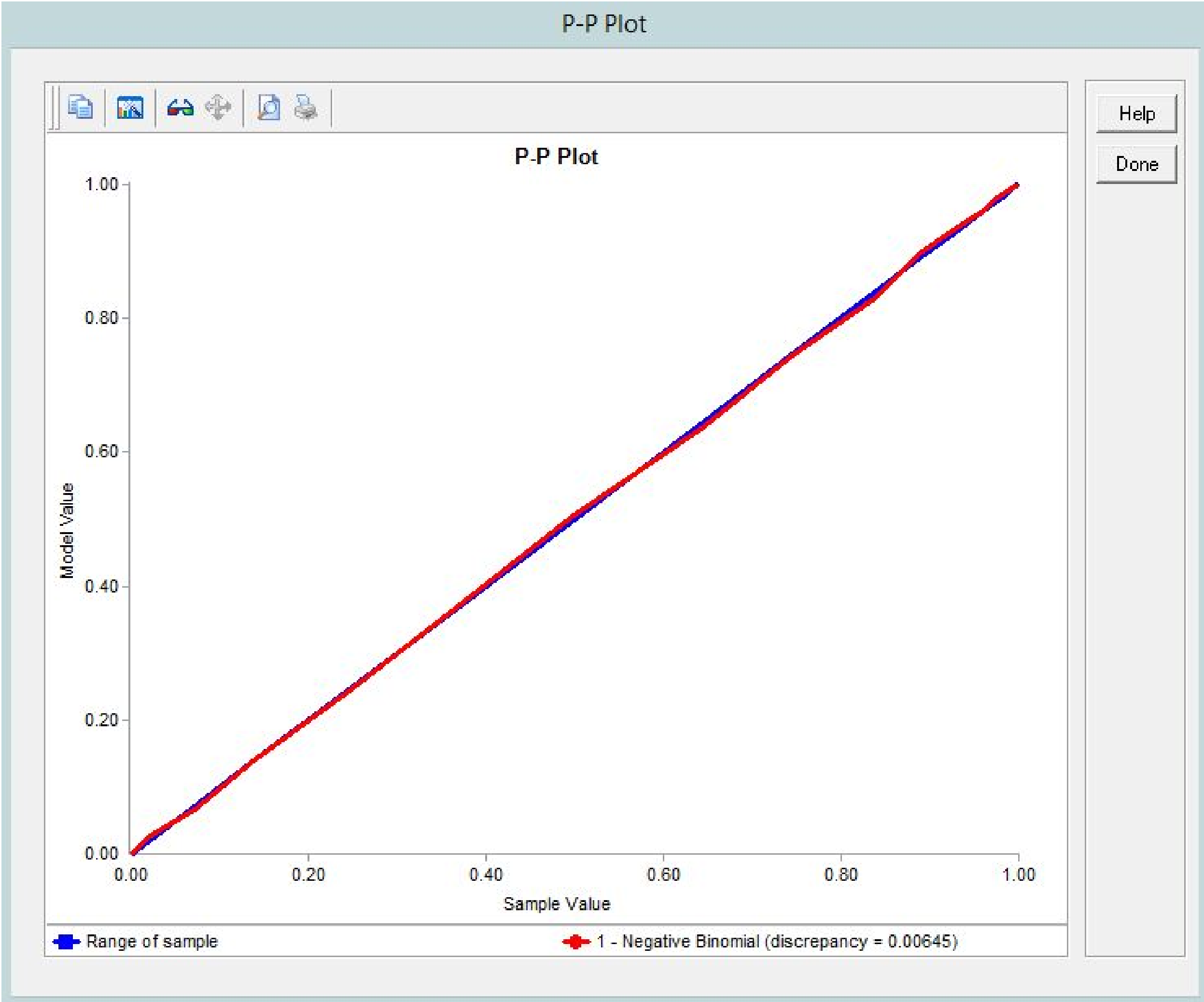
Data Characteristic	Value
Source file	<edited>
Observation type	Integer valued
Number of observations	474
Minimum observation	0
Maximum observation	17
Mean	7.67300
Median	7.50000
Variance	8.93935
Lexis ratio (var./mean)	1.16504
Skewness	0.46054

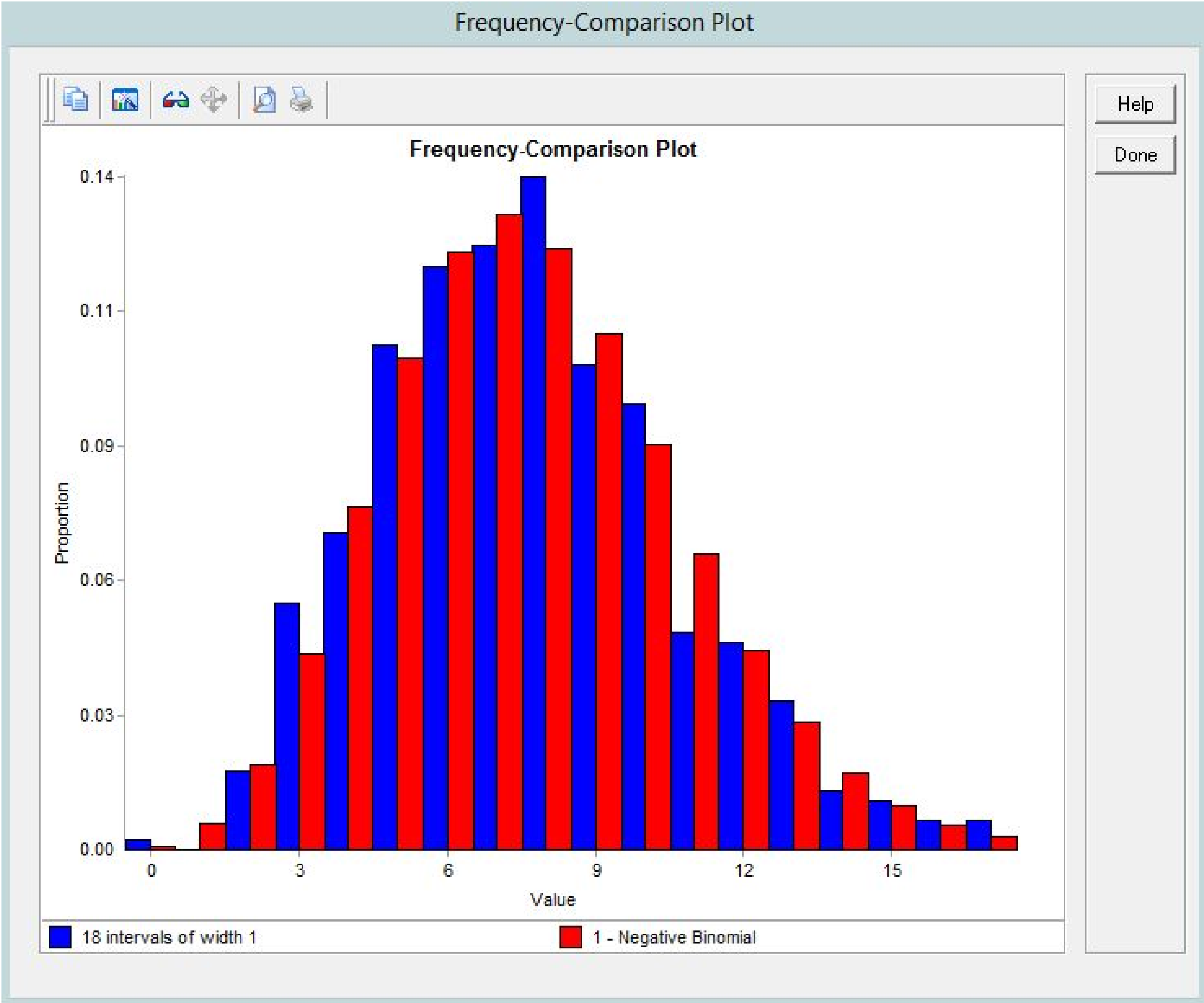
Copy

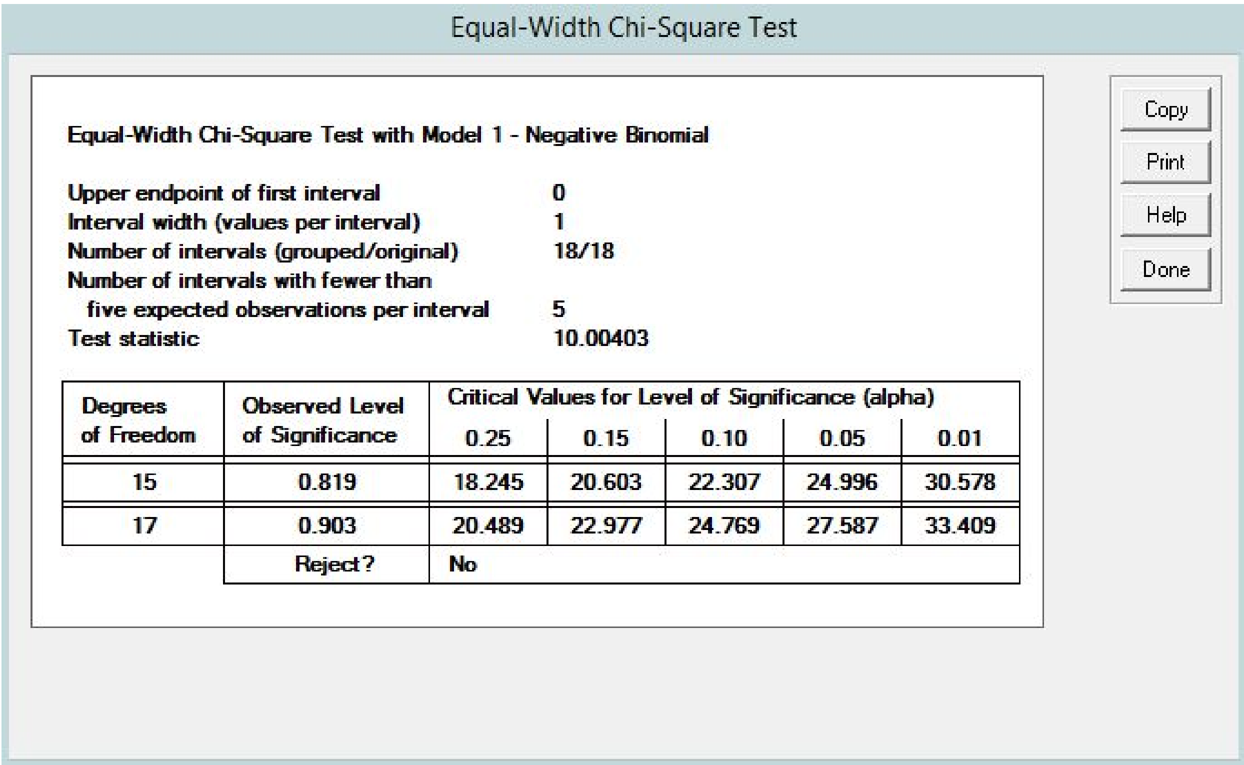
Print

Help

Done


P-P Plot
Help
Done
P-P Plot
1.00
0.80
0.60
0.40
0.20
0.00
Model Value
0.00
0.20
0.40
0.60
0.80
1.00
Sample Value
Range of sample
1 - Negative Binomial (discrepancy = 0.00645)


Frequency-Comparison Plot
Help
Done
Frequency-Comparison Plot
0.14
0.11
0.09
0.06
0.03
0.00
Proportion
0
3
6
9
12
15
Value
18 intervals of width 1
1 - Negative Binomial


Equal-Width Chi-Square Test

Equal-Width Chi-Square Test with Model 1 - Negative Binomial

Upper endpoint of first interval	0
Interval width (values per interval)	1
Number of intervals (grouped/original)	18/18
Number of intervals with fewer than five expected observations per interval	5
Test statistic	10.00403

Degrees of Freedom	Observed Level of Significance	Critical Values for Level of Significance (alpha) 0.25	0.15	0.10	0.05	0.01
15	0.819	18.245	20.603	22.307	24.996	30.578
17	0.903	20.489	22.977	24.769	27.587	33.409
	Reject?	No				

Copy

Print

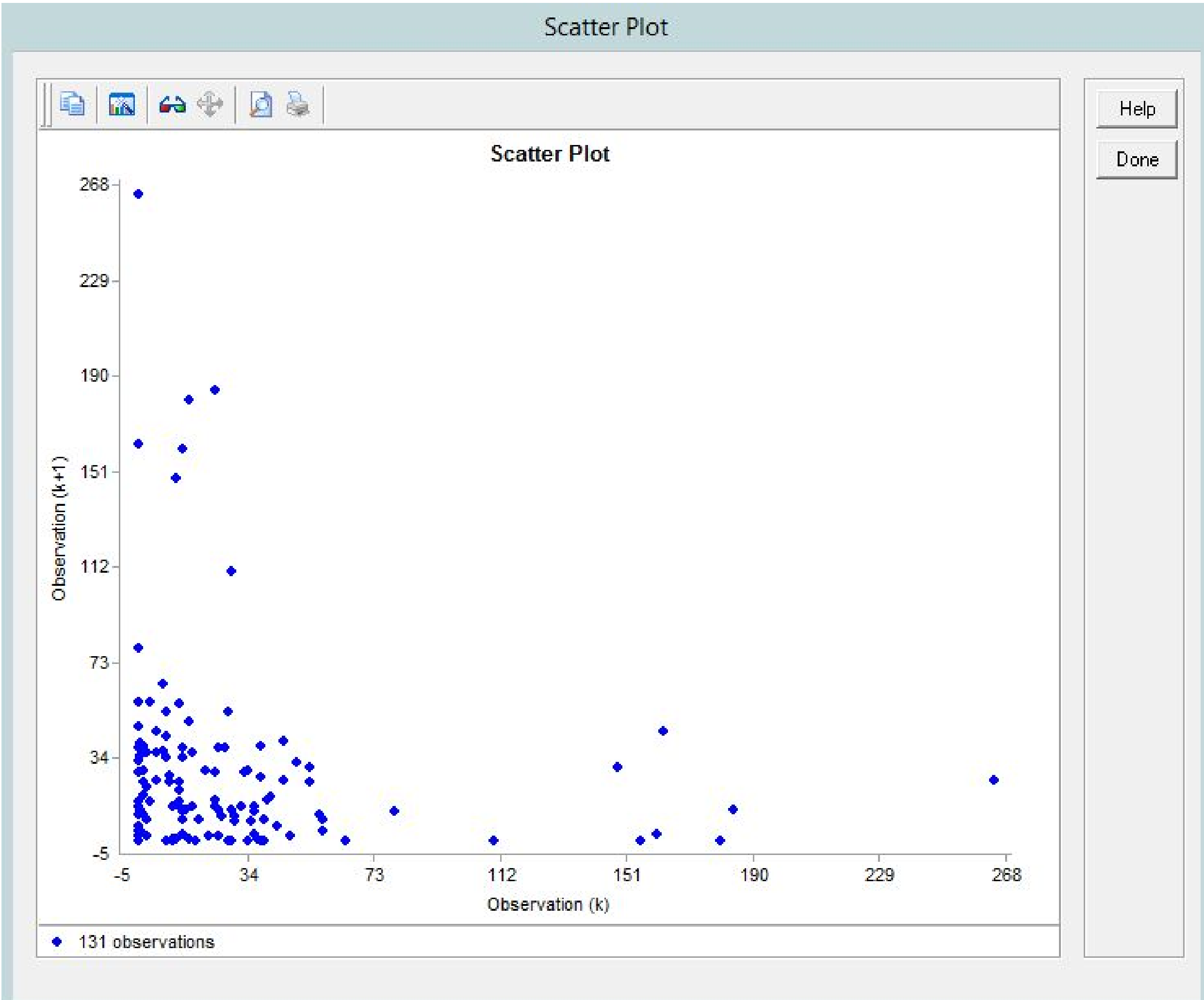
Help

Done

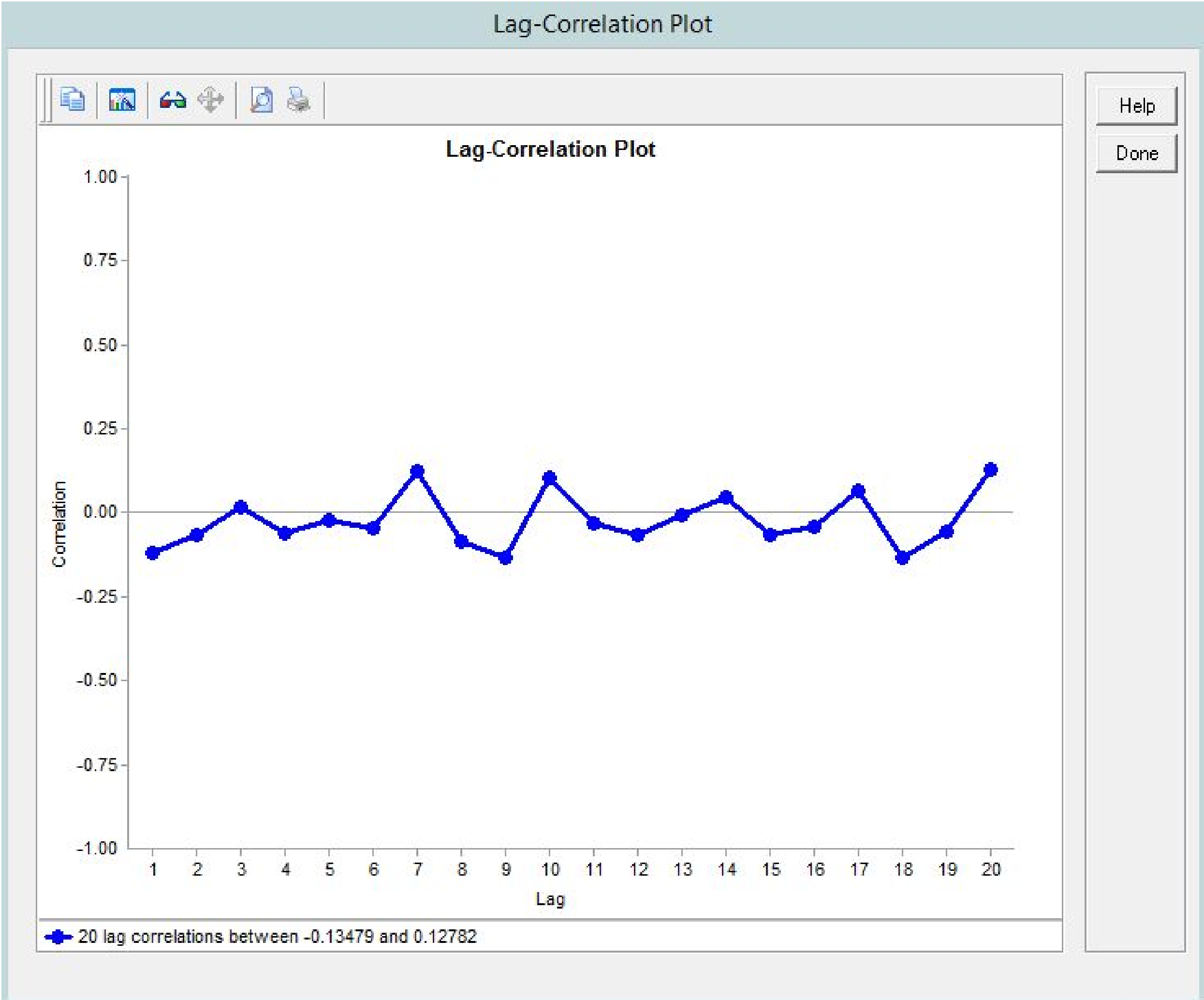
Simio Expression: Random.NegativeBinomial(0.85966, 47)

This model is very appropriate. The data is very minimally correlated and the Negative Binomial model fits very well. Even with the Chi-Squared test, the test statistic is far away from the critical values. If you increase and decrease the number of bins to extreme values, we still fail to reject the distribution fit. Even if you try to “cheat” with Chi-Squared, it looks like a pretty strong fit and hard to reject.

6.



Scatter Plot
Help
Done
Scatter Plot
268
229
190
151
112
73
34
-5
Observation (k+1)
-5
34
73
112
151
190
229
268
Observation (k)
131 observations


Lag-Correlation Plot
Help
Done
Lag-Correlation Plot
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
Correlation
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Lag
20 lag correlations between -0.13479 and 0.12782

Automated-Fitting Results

Copy

Print

Help

Done

Relative Evaluation of Candidate Models

Model	Relative Score	Parameters	
1 - Geometric	83.33	Probability	0.03478
2 - Negative Binomial	83.33	Probability	0.03478
		Success	1
3 - Poisson	22.22	Lambda	27.75573

4 models are defined with scores between 11.11 and 83.33

Absolute Evaluation of Model 1 - Geometric

Evaluation: Indeterminate

Suggestion: Additional evaluations using Comparisons Tab are strongly recommended. See Help for more information.

Additional Information about Model 1 - Geometric

"Error" in the model mean relative to the sample mean 3.5527e-15 = 0.00%

Data-Summary Table

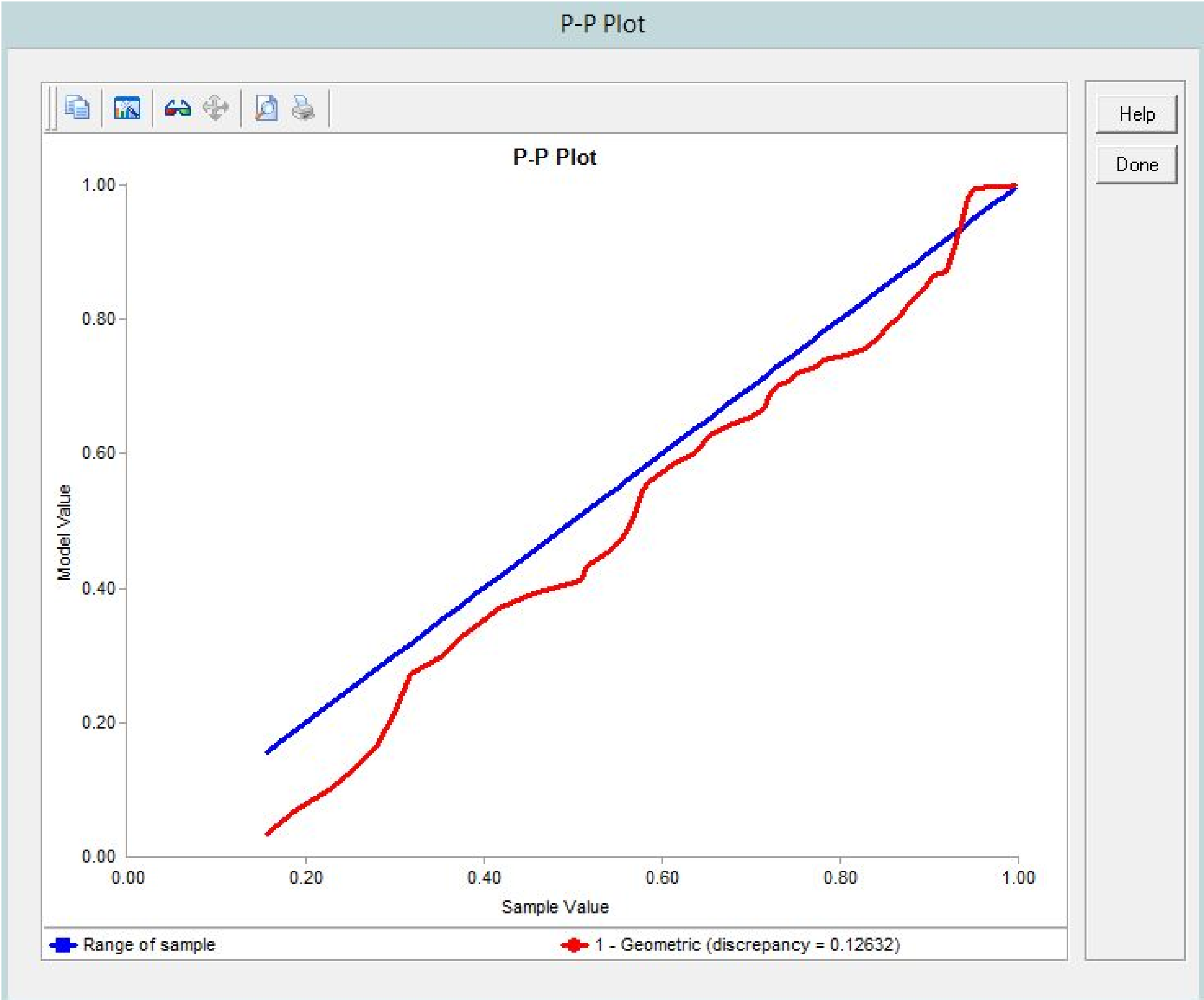
Data Characteristic	Value
Source file	<edited>
Observation type	Integer valued
Number of observations	131
Minimum observation	0
Maximum observation	264
Mean	27.75573
Median	14.00000
Variance	1,718.53987
Lexis ratio (var./mean)	61.91659
Skewness	3.18780

Copy

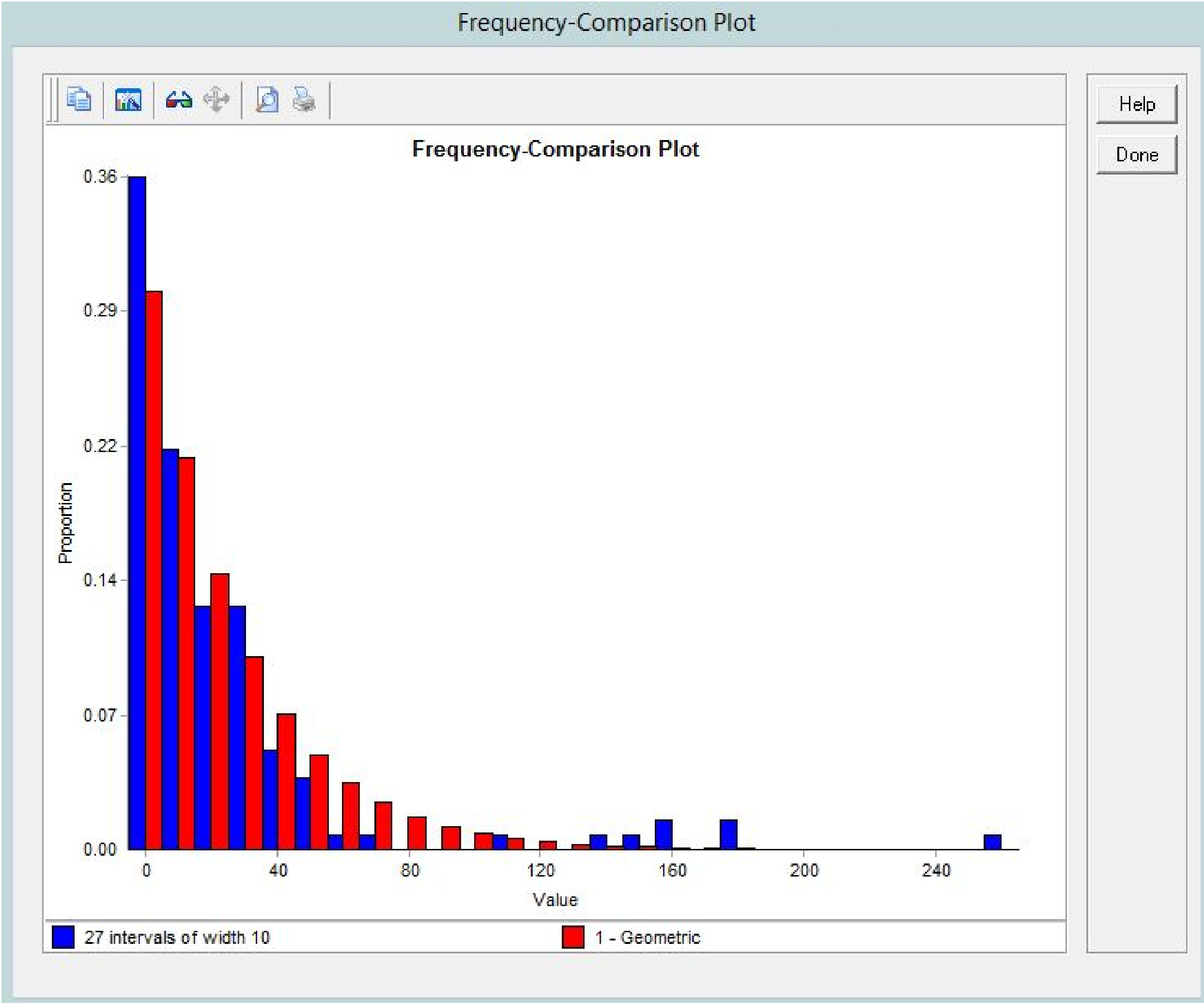
Print

Help

Done



P-P Plot
Help
Done
P-P Plot
1.00
0.80
0.60
0.40
0.20
0.00
Model Value
0.00
0.20
0.40
0.60
0.80
1.00
Sample Value
Range of sample
1 - Geometric (discrepancy = 0.12632)


Frequency-Comparison Plot
Help
Done
Frequency-Comparison Plot
0.36
0.29
0.22
0.14
0.07
0.00
Proportion
0
40
80
120
160
200
240
Value
27 intervals of width 10
1 - Geometric

Equal-Width Chi-Square Test

Equal-Width Chi-Square Test with Model 1 - Geometric

Upper endpoint of first interval	9
Interval width (values per interval)	10
Number of intervals (grouped/original)	27/27
Number of intervals with fewer than five expected observations per interval	21
Test statistic	175.25787

Degrees of Freedom	Observed Level of Significance	Critical Values for Level of Significance (alpha)				
		0.25	0.15	0.10	0.05	0.01
25	0.000	29.339	32.282	34.382	37.652	44.314
26	0.000	30.435	33.429	35.563	38.885	45.642
	Reject?	Yes				

Copy
Print
Help
Done

Simio Expression: Random.Geometric(0.03478)

This model does not seem appropriate for the given data. The PP Plot is very alarming and the theoretical probabilities do not match the theoretical probabilities well. With the default number of bins for the Chi-Square test, we reject the theoretical distribution, but if the number of bins changes to a small value, i.e. 3, we fail to reject. However, at this small level of granularity, the theoretical distribution is still a stretch when checking the test statistic and critical values.

7.

Parameter	Estimated Value	Standard Deviation
size	0.51956510122101	0.0634944866704382
mu	27.7557251985778	3.39565939739202

$\mu = r\frac{(1-p)}{p}$ where $\mu =$ 27.7557 and r = 0.519565. If you know μ and r and solve for p, you get

$$\widehat{p} = \frac{1}{(\frac{\mu}{r}+1)}$$

This implies that $\widehat{p}$ = 0.018375. The Var $= r\frac{(1-p)}{p^2}$. This implies the variance is 1510.5347. These estimates are fairly close to the sample estimates. The variance of the sample 1718.540 which is higher than our estimate.

8. (a)

Automated-Fitting Results

Copy | Print | Help | Done

Relative Evaluation of Candidate Models

Model	Relative Score	Parameters	
1 - Beta	100.00	Lower endpoint	1.91698 e -4
		Upper endpoint	161.80296
		Shape #1	1.30509
		Shape #2	4.07965
2 - Weibull	93.75	Location	0.00000
		Scale	43.12127
		Shape	1.42821
3 - Gamma	84.38	Location	0.00000
		Scale	24.20804
		Shape	1.62635

17 models are defined with scores between 4.69 and 100.00

Absolute Evaluation of Model 1 - Beta

Evaluation: Bad

Suggestion: Use an empirical distribution.
See Help for more information.

Additional Information about Model 1 - Beta

"Error" in the model mean
relative to the sample mean 0.15468 = 0.39%

Anderson-Darling Test

Copy | Print | Help | Done

Anderson-Darling Test with Model 1 - Beta

Sample size 2,590
Test statistic 8.78285

Note: No critical values exist for this special case.
The following critical values are for the case where
all parameters are known, and are conservative.

Sample Size	Critical Values for Level of Significance (alpha)					
	0.250	0.100	0.050	0.025	0.010	0.005
2,590	1.248	1.933	2.492	3.070	3.857	4.500
Reject?	Yes					

Kolmogorov-Smirnov Test

Copy | Print | Help | Done

Kolmogorov-Smirnov Test with Model 1 - Beta

Sample size 2,590
Normal test statistic 0.04720
Modified test statistic 2.40201

Note: No critical values exist for this special case.
The following critical values are for the case where
all parameters are known, and are conservative.

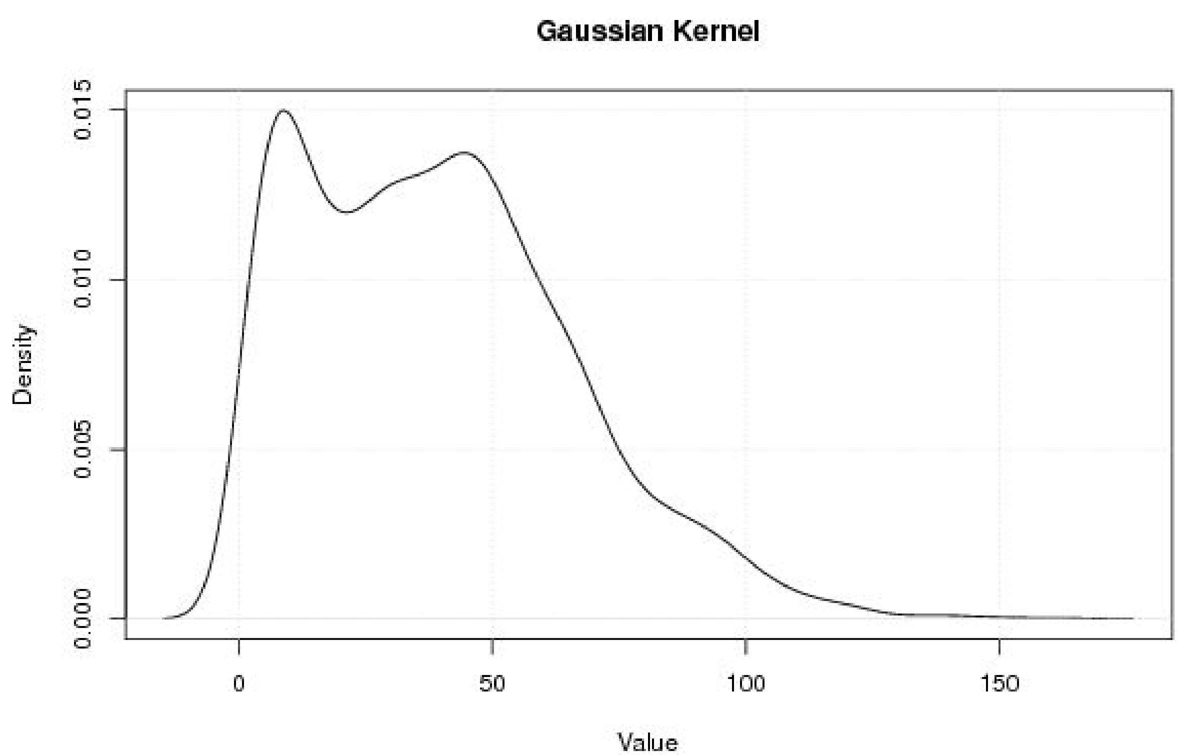
Sample Size	Critical Values for Level of Significance (alpha)				
	0.150	0.100	0.050	0.025	0.010
2,590	1.135	1.221	1.355	1.476	1.624
Reject?	Yes				

ExpertFit failed to give a good model. This is likely because there were so many data points and the tests were sensitive.

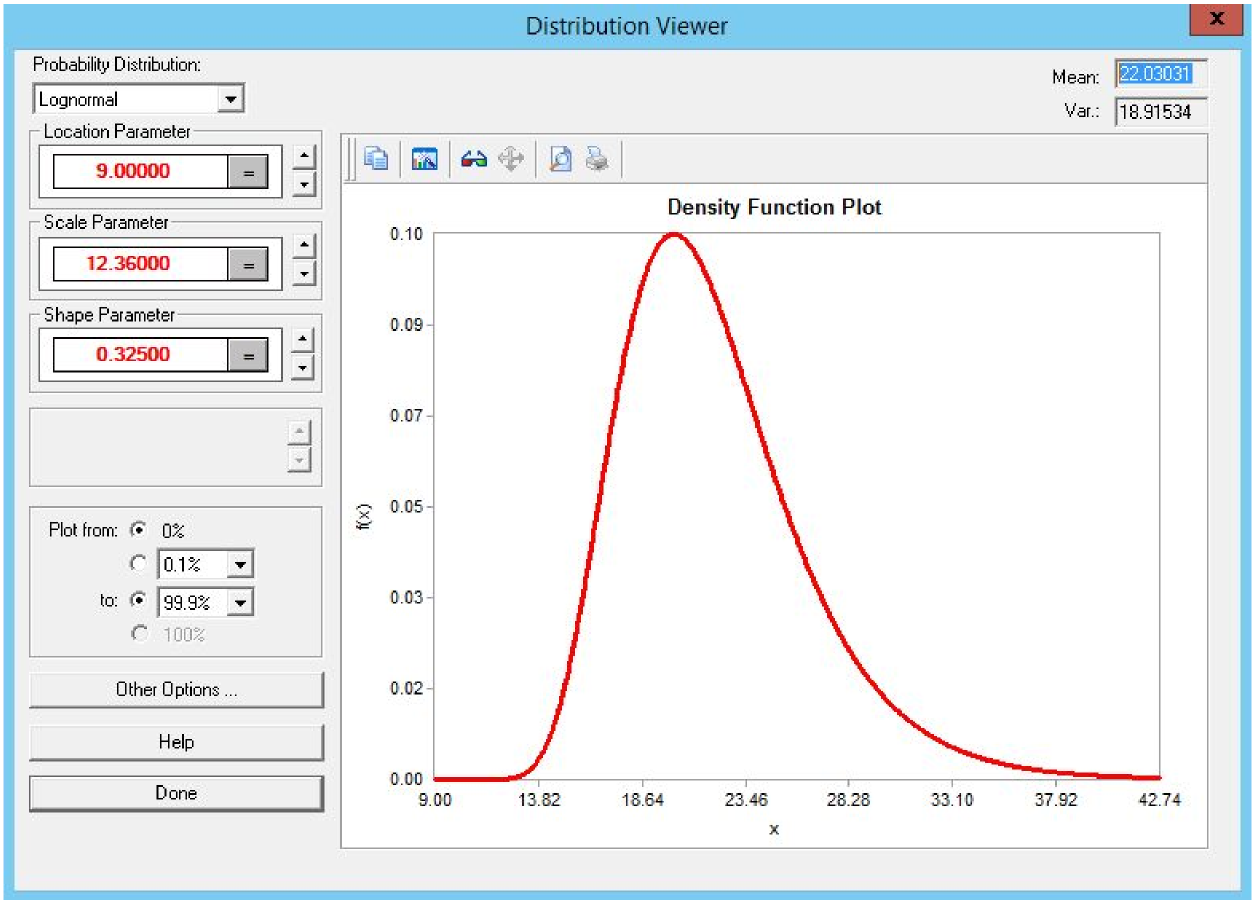
(b) KDE = 8.812508 using Gaussian and h = 4.954466 (see below)

Properties of Density Trace	
Bandwidth	4.95446595386128
#Observations	2590

Maximum Density Values		
Kernel	x-value	max. density
Gaussian	8.81250846045052	0.0149838238117316
Epanechnikov	9.93517066626168	0.0148811196669157
Rectangular	9.93517066626168	0.0153894154716978
Triangular	9.18672919572091	0.0148003349125171
Biweight	9.5609499309913	0.0148611031225658
Cosine	9.5609499309913	0.0148635896455241
Optcosine	9.93517066626168	0.0148611305289252



9. $\widehat{\mu} = 22.03031$, $\widehat{\sigma} = \sqrt{18.91534}$


Distribution Viewer
Probability Distribution:
Lognormal
Location Parameter
9.00000
Scale Parameter
12.36000
Shape Parameter
0.32500
Plot from: 0%
0.1%
to: 99.9%
100%
Other Options ...
Help
Done
Mean: 22.03031
Var.: 18.91534
Density Function Plot
0.10
0.09
0.07
0.05
0.03
0.02
0.00
f(x)
9.00
13.82
18.64
23.46
28.28
33.10
37.92
42.74
x