EDUCATION COMMITTEE

OF THE

SOCIETY OF ACTUARIES

SHORT-TERM ACTUARIAL MATHEMATICS STUDY NOTE

REPLACEMENT PAGES FOR SECTION 16.5.3 FROM

LOSS MODELS: FROM DATA TO DECISIONS, FOURTH EDITION

by

Stuart A. Klugman

STAM-21-18 **Printed in U.S.A.**

16.5.3 Score-based approaches

Some analysts might prefer an automated process for selecting a model. An easy way to do that would be to assign a score to each model and let the model with the best value win. The following scores are worth considering:

1. Kolmogorov–Smirnov test statistic - choose the model with the smallest value.
2. Anderson–Darling test statistic - choose the model with the smallest value.
3. Chi-square goodness-of-fit test statistic - choose the model with the smallest value.
4. Chi-square goodness-of-fit test - choose the model with the highest p-value.
5. Likelihood (or loglikelihood) function at its maximum - choose the model with the largest value.

All but the chi-square p-value have a deficiency with respect to parsimony. First, consider the likelihood function. When comparing, say, an exponential to a Weibull model, the Weibull model must have a likelihood value that is at least as large as the exponential model. They would only be equal in the rare case that the maximum likelihood estimate of the Weibull parameter τ is equal to 1. Thus, the Weibull model would always win over the exponential model, a clear violation of the principle of parsimony. For the three test statistics, there is no assurance that the same relationship will hold, but it seems likely that, if a more complex model is selected, the fit measure will be better. The only reason the chi-square test p-value is immune from this problem is that with more complex models, the test has fewer degrees of freedom. It is then possible that the more complex model will have a smaller p-value. There is no comparable adjustment for the first two test statistics listed.

With regard to the likelihood value, there are two ways to proceed. One is to perform the likelihood ratio test and the other is to impose a penalty for employing additional parameters. The likelihood ratio test is technically only available when one model is a special case of another (e.g., Pareto versus generalized Pareto). The concept can be turned into an algorithm by using the test at a 5% significance level. Begin with the best one-parameter model (the one with the highest loglikelihood value). Add a second parameter only if the two-parameter model with the highest loglikelihood value shows an increase of at least 1.92 (so twice the difference exceeds the critical value of 3.84). Then move to three-parameter models. If the comparison is to a two-parameter model, a 1.92 increase is again needed. If the early comparison led to keeping the one-parameter model, an increase of 3.00 is needed (because the test has two degrees of freedom). To add three parameters requires a 3.91 increase; four parameters, a 4.74 increase; and so on. In the spirit of this chapter, this algorithm can be used even when one model is not a special case of the other model. However, it would not be appropriate to claim that a likelihood ratio test was being conducted.

Aside from the issue of special cases, the likelihood ratio test has the same problem as any hypothesis test. Were the sample size to double, the loglikelihoods would also double, making it more likely that a model with a higher number of parameters will be selected, tending to defeat the parsimony principle. Conversely, it could be argued that, if we possess a lot of data, we have the right to consider and fit more complex models. A method that effects a compromise between these positions is the Schwarz Bayesian criterion (SBC) [99], which is also called the Bayesian Information Criterion (BIC). This method recommends that, when ranking models, a deduction of $(r/2)\ln n$ should be made from the loglikelihood value, where r is the number of estimated parameters and n is the sample size. Thus, adding a parameter requires an increase of $0.5\ln n$ in the loglikelihood. For larger sample sizes, a greater increase is needed, but it is not proportional to the sample size itself.

An alternative penalty is the Akaike Information Criterion (AIC).[1] This method deducts the number of parameters from the loglikelihood.[2] Section 3 of Brockett [16] promotes AIC while in a discussion of that paper Carlin provides support for SBC. The difference in the two methods is that SBC adjusts for the sample size and AIC does not. To summarize, the scores are:

$$\text{SBC/BIC:}\quad \ln L - \frac{r}{2}\ln n$$
$$\text{AIC:}\quad \ln L - r.$$

EXAMPLE 16.11

For the continuing example in this chapter, choose between the exponential and Weibull models for the data.

Graphs were constructed in the various examples and exercises. Table 16.14 summarizes the numerical measures. For the truncated version of Data Set B, the SBC is calculated for a sample size of 19, while for the version censored at 1,000, there are 20 observations. For both versions of Data Set B, while the Weibull offers some improvement, it is not convincing. In particular, none of the likelihood ratio test, SBC, or AIC indicate value in the second parameter. For Data Set C, it is clear that the Weibull model is superior and provides an excellent fit. □

EXAMPLE 16.12

In Example 7.8 an ad hoc method was used to demonstrate that the Poisson–ETNB distribution provided a good fit. Use the methods of this chapter to determine a good model.

The data set is very large and, as a result, requires a very close correspondence of the model to the data. The results are given in Table 16.15.

From Table 16.15, it is seen that the negative binomial distribution does not fit well, while the fit of the Poisson–inverse Gaussian is marginal at best ($p = 2.88\%$). The Poisson–inverse Gaussian is a special case ($r = -0.5$) of the Poisson–ETNB. Hence, a likelihood ratio test can be formally applied to determine if the additional parameter r is justified. Because the loglikelihood increases by 5, which is more than 1.92, the three-parameter model is a significantly better fit. The chi-square test shows that the

[1] Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19 (6), 716–723.

[2] When using software that computes SBC and AIC it is important to note how they are being determined as there are alternative formulas available. For example, when applying AIC, the software may start with twice the loglikelihood and then subtract twice the number of parameters. It is also common to change the sign, in which case smaller values are preferred. For this text and Exam C, The AIC and SBC/BIC values will be calculated as shown on this page.

Table 16.14 Results for Example 16.11.

	B truncated at 50		B censored at 1,000	
Criterion	Exponential	Weibull	Exponential	Weibull
K–S*	0.1340	0.0887	0.0991	0.0991
A–D*	0.4292	0.1631	0.1713	0.1712
χ^2	1.4034	0.3615	0.5951	0.5947
p-value	0.8436	0.9481	0.8976	0.7428
Loglikelihood	−146.063	−145.683	−113.647	−113.647
SBC	−147.535	−148.628	−115.145	−116.643
AIC	−147.063	−147.683	−114.647	−115.647
	C			
χ^2	61.913	0.3698		
p-value	10^{-12}	0.9464		
Loglikelihood	−214.924	−202.077		
SBC	−217.350	−206.929		
AIC	−215.924	−204.077		

*K–S and A–D refer to the Kolmogorov–Smirnov and Anderson–Darling test statistics, respectively.

Poisson–ETNB provides an adequate fit. In contrast, the SBC, but not the AIC, favors the Poisson–inverse Gaussian distribution. This illustrates that with large sample sizes using the SBC makes it harder to add a parameter. Given the improved fit in the tail for the three-parameter model, it seems to be the best choice. □

EXAMPLE 16.13

This example is taken from Douglas [27, p. 253]. An insurance company's records for one year show the number of accidents per day that resulted in a claim to the insurance company for a particular insurance coverage. The results are in Table 16.16. Determine if a Poisson model is appropriate.

A Poisson model is fitted to these data. The method of moments and the maximum likelihood method both lead to the estimate of the mean,

$$\hat{\lambda} = \frac{742}{365} = 2.0329.$$

The results of a chi-square goodness-of-fit test are in Table 16.17. Any time such a table is made, the expected count for the last group is

$$E_{k+} = n\hat{p}_{k+} = n(1 - \hat{p}_0 - \cdots - \hat{p}_{k-1}).$$

The last three groups are combined to ensure an expected count of at least one for each row. The test statistic is 9.93 with six degrees of freedom. The critical value at a 5% significance level is 12.59 and the p-value is 0.1277. By this test, the Poisson distribution is an acceptable model; however, it should be noted that the fit is poorest at the large values, and with the model understating the observed values, this may be a risky choice. □

Table 16.15 Results for Example 16.12

No. of claims	Observed frequency	Negative binomial	Poisson–inverse Gaussian	Poisson–ETNB
		Fitted distributions		
0	565,664	565,708.1	565,712.4	565,661.2
1	68,714	68,570.0	68,575.6	68,721.2
2	5,177	5,317.2	5,295.9	5,171.7
3	365	334.9	344.0	362.9
4	24	18.7	20.8	29.6
5	6	1.0	1.2	3.0
6+	0	0.0	0.1	0.4
Parameters		$\beta = 0.0350662$	$\lambda = 0.123304$	$\lambda = 0.123395$
		$r = 3.57784$	$\beta = 0.0712027$	$\beta = 0.233862$
				$r = -0.846872$
Chi square		12.13	7.09	0.29
Degrees of freedom		2	2	1
p-value		$<$1%	2.88%	58.9%
$-$Loglikelihood		251,117	251,114	251,109
SBC		$-$251,130	$-$251,127	$-$251,129
AIC		$-$251,119	$-$251,116	$-$251,112

Table 16.16 Data for Example 16.13

No. of claims/day	Observed no. of days
0	47
1	97
2	109
3	62
4	25
5	16
6	4
7	3
8	2
9+	0

EXAMPLE 16.14

The data in Table 14.17 come from Beard et al. [12] and are analyzed in Example 14.7. Determine a model that adequately describes the data.

Parameter estimates from fitting four models are in Table 14.7. Various fit measures are given in Table 16.18. Only the zero-modified geometric distribution passes the goodness-of-fit test. It is also clearly superior according to the SBC and AIC. A likeli-

Table 16.17 Chi-square goodness-of-fit test for Example 16.13.

Claims/day	Observed	Expected	Chi square
0	47	47.8	0.01
1	97	97.2	0.00
2	109	98.8	1.06
3	62	66.9	0.36
4	25	34.0	2.39
5	16	13.8	0.34
6	4	4.7	0.10
7+	5	1.8	5.66
Totals	365	365	9.93

Table 16.18 Test results for Example 16.14.

	Poisson	Geometric	ZM Poisson	ZM geometric
Chi square	543.0	643.4	64.8	0.58
Degrees of freedom	2	4	2	2
p-value	$< 1\%$	$< 1\%$	$< 1\%$	74.9%
Loglikelihood	−171,373	−171,479	−171,160	−171,133
SBC	−171,379.5	−171,485.5	−171,173	−171,146
AIC	−171,374	−171,480	−171,162	−171,135

hood ratio test against the geometric has a test statistic of $2(171{,}479-171{,}133) = 692$, which with one degree of freedom is clearly significant. This result confirms the qualitative conclusion in Example 14.7. □

EXAMPLE 16.15

The data in Table 16.19, from Simon [103], represent the observed number of claims per contract for 298 contracts. Determine an appropriate model.

The Poisson, negative binomial, and Polya–Aeppli distributions are fitted to the data. The Polya–Aeppli and the negative binomial are both plausible distributions. The p-value of the chi-square statistic and the loglikelihood both indicate that the Polya–Aeppli is slightly better than the negative binomial. The SBC and AIC verify that both models are superior to the Poisson distribution. The ultimate choice may depend on familiarity, prior use, and computational convenience of the negative binomial versus the Polya–Aeppli model. □

EXAMPLE 16.16

Consider the data in Table 16.20 on automobile liability policies in Switzerland taken from Bühlmann [18]. Determine an appropriate model.

Table 16.19 Fit of Simon data for Example 16.15.

Number of claims/contract	Number of contracts	Fitted distributions: Poisson	Negative binomial	Polya–Aeppli
0	99	54.0	95.9	98.7
1	65	92.2	75.8	70.6
2	57	78.8	50.4	50.2
3	35	44.9	31.3	32.6
4	20	19.2	18.8	20.0
5	10	6.5	11.0	11.7
6	4	1.9	6.4	6.6
7	0	0.5	3.7	3.6
8	3	0.1	2.1	2.0
9	4	0.0	1.2	1.0
10	0	0.0	0.7	0.5
11	1	0.0	0.4	0.3
12+	0	0.0	0.5	0.3
Parameters		$\lambda = 1.70805$	$\beta = 1.15907$	$\lambda = 1.10551$
			$r = 1.47364$	$\beta = 0.545039$
Chi square		72.64	4.06	2.84
Degrees of freedom		4	5	5
p-Value		<1%	54.05%	72.39%
Loglikelihood		−577.0	−528.8	−528.5
SBC		−579.8	−534.5	−534.2
AIC		−578.0	−530.8	−530.5

Three models are considered in Table 16.20. The Poisson distribution is a very bad fit. Its tail is far too light compared with the actual experience. The negative binomial distribution appears to be much better but cannot be accepted because the p-value of the chi-square statistic is very small. The large sample size requires a better fit. The Poisson–inverse Gaussian distribution provides an almost perfect fit (p-value is large). Note that the Poisson–inverse Gaussian has two parameters, like the negative binomial. The SBC and AIC also favor this choice. This example shows that the Poisson–inverse Gaussian can have a much heavier right-hand tail than the negative binomial. □

EXAMPLE 16.17

Comprehensive medical claims were studied by Bevan [14] in 1963. Male (955 payments) and female (1,291 payments) claims were studied separately. The data appear in Table 16.21 where there was a deductible of 25. Can a common model be used?

When using the combined data set, the lognormal distribution is the best two-parameter model. Its negative loglikelihood (NLL) is 4,580.20. This value is 19.09

Table 16.20 Fit of Buhlmann data for Example 16.16.

No. of accidents	Observed frequency	Fitted distributions: Poisson	Negative binomial	P.–i.G.*
0	103,704	102,629.6	103,723.6	103,710.0
1	14,075	15,922.0	13,989.9	14,054.7
2	1,766	1,235.1	1,857.1	1,784.9
3	255	63.9	245.2	254.5
4	45	2.5	32.3	40.4
5	6	0.1	4.2	6.9
6	2	0.0	0.6	1.3
7+	0	0.0	0.1	0.3
Parameters		$\lambda = 0.155140$	$\beta = 0.150232$	$\lambda = 0.144667$
			$r = 1.03267$	$\beta = 0.310536$
Chi square		1,332.3	12.12	0.78
Degrees of freedom		2	2	3
p-Values		<1%	<1%	85.5%
Loglikelihood		−55,108.5	−54,615.3	−54,609.8
SBC		−55,114.3	−54,627.0	−54,621.5
AIC		−55,109.5	−54,617.3	−54,611.8

*P.–i.G. stands for Poisson–inverse Gaussian.

better than the one-parameter inverse exponential model and 0.13 worse than the three-parameter Burr model. Because none of these models is a special case of the other, the likelihood ratio test (LRT) cannot be used, but it is clear that, using the 1.92 difference as a standard, the lognormal is preferred. The SBC requires an improvement of $0.5\ln(2{,}246) = 3.86$, while the AIC requires 1.00, and again the lognormal is preferred. The parameters are $\mu = 4.5237$ and $\sigma = 1.4950$. When separate lognormal models are fit to males ($\mu = 3.9686$ and $\sigma = 1.8432$) and females ($\mu = 4.7713$ and $\sigma = 1.2848$), the respective NLLs are 1,977.25 and 2,583.82 for a total of 4,561.07. This result is an improvement of 19.13 over a common lognormal model, which is significant by the LRT (3.00 needed), the SBC (7.72 needed), and the AIC (2.00 needed). Sometimes it is useful to be able to use the same nonscale parameter in both models. When a common value of σ is used, the NLL is 4,579.77, which is significantly worse than using separate models. □

EXAMPLE 16.18

In 1958 Longley-Cook [70] examined employment patterns of casualty actuaries. One of his tables listed the number of members of the Casualty Actuarial Society employed by casualty companies in 1949 (55 actuaries) and 1957 (78 actuaries). Using the data in Table 16.22, determine a model for the number of actuaries per company that employs at least one actuary and find out whether the distribution has changed over the eight-year period.

Table 16.21 Comprehensive medical losses for Example 16.17.

Loss	Male	Female
25–50	184	199
50–100	270	310
100–200	160	262
200–300	88	163
300–400	63	103
400–500	47	69
500–1,000	61	124
1,000–2,000	35	40
2,000–3,000	18	12
3,000–4,000	13	4
4,000–5,000	2	1
5,000–6,667	5	2
6,667–7,500	3	1
7,500–10,000	6	1

Table 16.22 Number of actuaries per company for Example 16.18.

Number of actuaries	Number of companies—1949	Number of companies—1957
1	17	23
2	7	7
3–4	3	3
5–9	2	3
10+	0	1

Because a value of zero is impossible, only zero-truncated distributions should be considered. In all three cases (1949 data only, 1957 data only, combined data), the ZT logarithmic and ZT (extended) negative binomial distributions have acceptable goodness-of-fit test values. The improvement in NLL is 0.52, 0.02, and 0.94. The LRT can be applied (except that the ZT logarithmic distribution is a limiting case of the ZT negative binomial distribution with $r \to 0$), and the improvement is not significant in any of the cases. The same conclusions apply if the SBC or AIC are used. The parameter estimates (where β is the only parameter) are 2.0227, 2.8114, and 2.4479, respectively. The NLL for the combined data set is 74.35, while the total for the two separate models is 74.15. The improvement is only 0.20, which is not significant (there is one degree of freedom). Even though the estimated mean has increased from $2.0227/\ln(3.0227) = 1.8286$ to $2.8114/\ln(3.8114) = 2.1012$, there is not enough data to make a convincing case that the true mean has increased. □

Table 16.23 Data for Exercise 16.22.

No. of accidents	No. of policies
0	100
1	267
2	311
3	208
4	87
5	23
6	4
Total	1,000

Table 16.24 Results for Exercise 16.25.

Model	No. of parameters	Negative loglikelihood
Generalized Pareto	3	219.1
Burr	3	219.2
Pareto	2	221.2
Lognormal	2	221.4
Inverse exponential	1	224.3

16.5.4 Exercises

16.22 (*) One thousand policies were sampled and the number of accidents for each recorded. The results are in Table 16.23. Without doing any formal tests, determine which of the following five models is most appropriate: binomial, Poisson, negative binomial, normal, gamma.

16.23 For Example 16.1, determine if a transformed gamma model is more appropriate than either the exponential model or the Weibull model for each of the three data sets.

16.24 (*) From the data in Exercise 16.11, the maximum likelihood estimates are $\hat{\lambda} = 0.60$ for the Poisson distribution and $\hat{r} = 2.9$ and $\hat{\beta} = 0.21$ for the negative binomial distribution. Conduct the likelihood ratio test for choosing between these two models.

16.25 (*) From a sample of size 100, five models are fit with the results given in Table 16.24. Use the SBC and then the AIC to select the best model.

16.26 Refer to Exercise 13.41. Use the likelihood ratio test (at a 5% significance level), the SBC, and the AIC to decide if Sylvia's claim is true.

16.27 (*) Five models were fit to a sample of 260 observations. The following are the number of parameters in the model followed by the loglikelihood value: $1, -414$, $2, -412$, $3, -411$, $4, -409$, $6, -409$. According to the SBC, which model (identified by the number of parameters) should be selected? Does the decision change if the AIC is used?

16.28 Using results from Exercises 14.3 and 16.17, use the chi-square goodness-of-fit test, the likelihood ratio test, the SBC, and the AIC to determine the best model from the members of the $(a, b, 0)$ class.

Table 16.25 Data for Exercise 16.31.

No. of medical claims	No. of accidents
0	529
1	146
2	169
3	137
4	99
5	87
6	41
7	25
8+	0

16.29 Using results from Exercises 14.5 and 16.18, use the chi-square goodness-of-fit test, the likelihood ratio test, the SBC, and the AIC to determine the best model from the members of the $(a, b, 0)$ class.

16.30 Using results from Exercises 14.6 and 16.19, use the chi-square goodness-of-fit test, the likelihood ratio test, the SBC, and the AIC to determine the best model from the members of the $(a, b, 0)$ class.

16.31 Table 16.25 gives the number of medical claims per reported automobile accident.

(a) Construct a plot similar to Figure 6.1. Does it appear that a member of the $(a, b, 0)$ class will provide a good model? If so, which one?

(b) Determine the maximum likelihood estimates of the parameters for each member of the $(a, b, 0)$ class.

(c) Based on the chi-square goodness-of-fit test, the likelihood ratio test, the SBC, and the AIC, which member of the $(a, b, 0)$ class provides the best fit? Is this model acceptable?

16.32 For the four data sets introduced in Exercises 14.3, 14.5, 14.6, and 16.31, you have determined the best model from among members of the $(a, b, 0)$ class. For each data set, determine the maximum likelihood estimates of the zero-modified Poisson, geometric, logarithmic, and negative binomial distributions. Use the chi-square goodness-of-fit test and likelihood ratio tests to determine the best of the eight models considered and state whether the selected model is acceptable.

16.33 A frequency model that has not been mentioned to this point is the **zeta distribution**. It is a zero-truncated distribution with $p_k^T = k^{-(\rho+1)}/\zeta(\rho + 1),\ k = 1, 2, \ldots, \rho > 0$. The denominator is the zeta function, which must be evaluated numerically as $\zeta(\rho + 1) = \sum_{k=1}^{\infty} k^{-(\rho+1)}$. The zero-modified zeta distribution can be formed in the usual way. More information can be found in Luong and Doray [72].

(a) Determine the maximum likelihood estimates of the parameters of the zero-modified zeta distribution for the data in Example 14.7.

(b) Is the zero-modified zeta distribution acceptable?

Table 16.26 Data for Excercise 16.35(a).

No. of claims	No. of policies
0	96,978
1	9,240
2	704
3	43
4	9
5+	0

Table 16.27 Data for Excercise 16.35(b).

No. of deaths	No. of corps
0	109
1	65
2	22
3	3
4	1
5+	0

16.34 In Exercise 16.32, the best model from among the members of the $(a, b, 0)$ and $(a, b, 1)$ classes was selected for the data sets in Exercises 14.3, 14.5, 14.6, and 16.31. Fit the Poisson–Poisson, Polya–Aeppli, Poisson–inverse Gaussian, and Poisson–ETNB distributions to these data and determine if any of these distributions should replace the one selected in Exercise 16.32. Is the current best model acceptable?

16.35 The five data sets presented in this problem are all taken from Lemaire [67]. For each data set, compute the first three moments and then use the ideas in Section 7.2 to make a guess at an appropriate model from among the compound Poisson collection (Poisson, geometric, negative binomial, Poisson–binomial [with $m = 2$ and $m = 3$], Polya–Aeppli, Neyman Type A, Poisson–inverse Gaussian, and Poisson–ETNB). From the selected model (if any) and members of the $(a, b, 0)$ and $(a, b, 1)$ classes, determine the best model.

(a) The data in Table 16.26 represent counts from third-party automobile liability coverage in Belgium.

(b) The data in Table 16.27 represent the number of deaths due to horse kicks in the Prussian army between 1875 and 1894. The counts are the number of deaths in a corps (there were 10 of them) in a given year, and thus there are 200 observations. This data set is often cited as the inspiration for the Poisson distribution. For using any of our models, what additional assumption about the data must be made?

(c) The data in Table 16.28 represent the number of major international wars per year from 1500 through 1931.

(d) The data in Table 16.29 represent the number of runs scored in each half-inning of World Series baseball games played from 1947 through 1960.

(e) The data in Table 16.30 represent the number of goals per game per team in the 1966–1967 season of the National Hockey League.

Table 16.28 Data for Excercise 16.35(c).

No. of wars	No. of years
0	223
1	142
2	48
3	15
4	4
5+	0

Table 16.29 Data for Excercise 16.35(d).

No. of runs	No. of half innings
0	1,023
1	222
2	87
3	32
4	18
5	11
6	6
7+	3

Table 16.30 Data for Excercise 16.35(e).

No. of goals	No. of games
0	29
1	71
2	82
3	89
4	65
5	45
6	24
7	7
8	4
9	1
10+	3

16.36 Verify that the estimates presented in Example 7.14 are the maximum likelihood estimates. (Because only two decimals are presented, it is probably sufficient to observe that the likelihood function takes on smaller values at each of the nearby points.) The negative binomial distribution was fit to these data in Example 14.5. Which of these two models is preferable?

Solutions to the AIC portions of the exercises

The solutions preented below relate only to the additional work due to adding use of the AIC to some of the exercises.

16.23: For Data Set B truncated at 50 the transformed gaama improves the loglikelihood versus Wiebull by $0.022 < 1$ and so the AIC supports not adding this third parameter. For Data Set C there is no improvement to three decimal places and again the AIC supports staying with the Weibull distribution.

16.25: Subtracting the number of parameters from the loglikelihood values gives: generalized Pareto -222.1, Burr -222.2, Pareto -223.3, lognormal -223.4, and inverse exponential -225.3. The largest value is for the generalized Pareto model. This is a different result from using the SBC.

16.26: The difference in loglikelihood values is $0.26 < 1$ and so the AIC favors the null hypothesis.

16.27: After subtracting the number of parameters, the values are -415, -414, -414, -413, and -415. The largest value is for the model with four parameters. This differs from the SBC choice of the one parameter model.

16.28: The improvement using the negative binomial model is $0.01 < 1$ and so the AIC favors the Poisson model.

16.29: The improvement using the negative binomial model is $0.71 < 1$ and so the AIC favors the geometric model.

16.30: The improvement using the negative binomial model is $33.61 > 1$ and so this model is preferred by the AIC.

16.31: The improvement using the negative binomial model is $1.64 > 1$ and so this model is preferred by the AIC.