

## BMD Analysis: Tests of Fit

The BMDS software provides four different Tests of Fit that the user may use to determine an appropriate model for fitting their data. These Tests of Fit are based on asymptotic theories of the likelihood ratio. Without getting too technical, the likelihood ratio is just the ratio of two [likelihood](#) values, many of which are given in the BMDS output. Statistical theory proves that  $-2 \cdot \log(\text{likelihood ratio})$  converges to a Chi-Square random variable as the sample size gets large and the number of dose levels gets large. These values can in turn be used to obtain approximate probabilities to make inferences about model fit. Chi-Square tables can be found in almost any statistical book.

Each of the five models, described in the section on [likelihood](#), has a likelihood value. The BMDS program uses these values to create ratios from two models that form a meaningful test. Suppose the user wishes to test two models, A and B, for fit. One assumption that is made for these tests is that model A is “nested within” Model B, i.e., that Model B can be simplified (via restriction of some parameters in Model B) in such a way that the simplified model is Model A. This implies that Model A has fewer varying parameters. As an example, consider that the linear model is a “simpler” or “nested” model relative to the power model because the linear model has the power parameter restricted to be equal to 1.

Note The model with a higher number of parameters is always in the denominator of this ratio.

Now, using the theory,  $-2 \cdot \log\{L(A)/L(B)\}$  approaches a Chi-Square random variable. This can be simplified by using the fact that the log of a ratio is equal to the difference of the logs, or simply put,

$$-2 \cdot \log\{L(A)/L(B)\} = -2 \cdot (\log\{L(A)\} - \log\{L(B)\}) = 2 \cdot \log\{L(B)\} - 2 \cdot \log\{L(A)\}.$$

The likelihood values given by BMDS are in fact the log-likelihoods,  $\log\{L(B)\}$  and  $\log\{L(A)\}$ , so this likelihood ratio calculation becomes just a subtraction problem. This value can then in turn be compared to a Chi-Square random variable with a specified number of degrees of freedom.

As mentioned in the section on [likelihood](#), each log likelihood value has an associated number of parameters. The number of degrees of freedom for the Chi-Square test statistic is merely the difference between the two model parameter counts. In the mini-example above, suppose Model A has 5 fitted parameters, and that Model B has 8. In this case, the Chi-Square value you would compare this to would be a Chi-Square with  $8 - 5 = 3$  degrees of freedom.

In the A vs B example, what is exactly being tested? In terms of hypotheses, it would be:

H0: A models the data as well as B

H1: B models the data better than A

Keeping these tests in mind, suppose  $2 \cdot \log\{L(B)\} - 2 \cdot \log\{L(A)\} = 4.89$  based on 3 degrees of freedom. Also, suppose the rejection criteria is a Chi-Square probability of less than .05. Looking on a Chi-Square table, 4.89 has a p-value somewhere between .10 and .25. In this case,  $H_0$  would not be rejected, and it would seem to be appropriate to model the data using Model A. BMDS automatically does the “table look-up” for the user, and provides the p-value associated with the calculated log-likelihood ratio having degrees of freedom as described above.

The BMDS software provides four default tests. BMDS provides interpretation of the test results, based on p-values that have been selected by EPA. However, the computed p-values are presented so that the user is free to use any rejection criteria they want. Each of the four default tests provided for any of the continuous models is discussed in some detail below.

**Test 1 (A2 vs R):** Tests the null hypothesis that responses and variances don't differ among dose levels. If this test fails to reject the null hypothesis, there may not be a dose-response.

This test compares Model R (the simpler model) to Model A2. Model R is a simpler A2 (or nested within A2) since R can be obtained from A2 by restricting all the mean parameters to be equal to one another and restricting all the variance parameters to be equal to one another. If this test fails to reject the null hypothesis, then there may not be a dose-response, as the inference would be that the simpler model (R) is not much worse than the saturated model. The default p-value for the test (as reported in the Tests of Interest section of the output) is 0.05. A p-value less than 0.05 is associated with the statement that “There appears to be a difference between response and/or variances among the dose levels. It seems appropriate to model the data.” A p-value greater than 0.05 is associated with the statement that the data may not be suitable for dose-response modeling.

**Test 2 (A1 vs A2):** Tests the null hypothesis that variances are homogeneous. If this test fails to reject the null hypothesis, the simpler constant variance model may be appropriate.

This test compares A1 (the simpler model) to Model A2. Model A1 is a simpler A2 (or nested within A2) since A1 can be obtained from A2 by restricting all the variance parameters to be equal to one another. If this test rejects the null hypothesis, the inference is that the constant variance assumption is incorrect and a modeled variance is necessary to adequately represent the data. The default p-value for the test (as reported in the Tests of Interest section of the output) is 0.1. A p-value less than 0.1 is associated with the statement that the user should “Consider running a non-homogeneous variance model. A p-value greater than 0.1 is associated with the statement that a constant variance assumption is suitable for the dose-response modeling.

**Test 3 (A3 vs A2):** Tests the null hypothesis that the variances are adequately modeled. If this test fails to reject the null hypothesis, it may be inferred that the variances have been modeled appropriately.

Here, the test is one to see whether or not the user-specified variance model, is appropriate. If the user-specified variance model is “constant variance,” then Models A1 and A3 are identical; this test is the same as Test 2, with the same interpretation. If the user-specified variance model is nonconstant ( $\sigma(i)^2 = \alpha \cdot \mu(i)^\rho$ ), this test determines if that particular equation

appears adequate to describe the variance across dose groups. Model A3 is the simpler version of Model A2 obtained by constraining the variances to fit the nonconstant variance equation. The default p-value for the test (as reported in the Tests of Interest section of the output) is 0.1. A p-value less than 0.1 is associated with the statement that “You may want to consider a different variance model.” [Unfortunately BMDS has no further way to model variance. Look for different variance models in future releases of BMDS.] A p-value greater than 0.1 is associated with the statement that the modeled variance appears to be suitable for the dose-response modeling.

**Test 4 (Fitted vs A3):** Tests the null hypothesis that the model for the mean fits the data. If this tests fails to reject the null hypothesis, the user has support for the selected model.

This test compares the Fitted Model to Model A3. The Fitted Model is as simpler Model A3 (or nested within Model A3) because it can be obtained by restricting the means (unrestricted in A3) to be described by the dose-response function under consideration. If this test fails to reject the null hypothesis, the inference is that the fitted model is adequate to describe the dose-related changes in the means (conditional on the form of the variance model; the form of the variance model is the same for the Fitted Model and Model A3). Failure to reject the null hypothesis is associated with the inference that the restriction of the means to the shape of the dose-response function under consideration is adequate. The default p-value for the test (as reported in the Tests of Interest section of the output) is 0.1. A p-value less than 0.1 is associated with the statement that “You may want to try a different model.” I.e., the fit of the Fitted Model is not good enough. A p-value greater than 0.1 is associated with a statement that the Fitted Model appears to be suitable for dose-response modeling. of Model A2 obtained by constraining the variances to fit the nonconstant variance equation. The default p-value for the test (as reported in the Tests of Interest section of the output) is 0.1. A p-value less than 0.1 is associated with the statement that “You may want to consider a different variance model.” [Unfortunately BMDS has no further way to model variance. Look for different variance models in future releases of BMDS.] A p-value greater than 0.1 is associated with the statement that the modeled variance appears to be suitable for the dose-response modeling.

**Test 4 (Fitted vs A3):** Tests the null hypothesis that the model for the mean fits the data. If this tests fails to reject the null hypothesis, the user has support for the selected model.

This test compares the Fitted Model to Model A3. The Fitted Model is as simpler Model A3 (or nested within Model A3) because it can be obtained by restricting the means (unrestricted in A3) to be described by the dose-response function under consideration. If this test fails to reject the null hypothesis, the inference is that the fitted model is adequate to describe the dose-related changes in the means (conditional on the form of the variance model; the form of the variance model is the same for the Fitted Model and Model A3). Failure to reject the null hypothesis is associated with the inference that the restriction of the means to the shape of the dose-response function under consideration is adequate. The default p-value for the test (as reported in the Tests of Interest section of the output) is 0.1. A p-value less than 0.1 is associated with the statement that “You may want to try a different model.” I.e., the fit of the Fitted Model is not

good enough. A p-value greater than 0.1 is associated with a statement that the Fitted Model appears to be suitable for dose-response modeling.