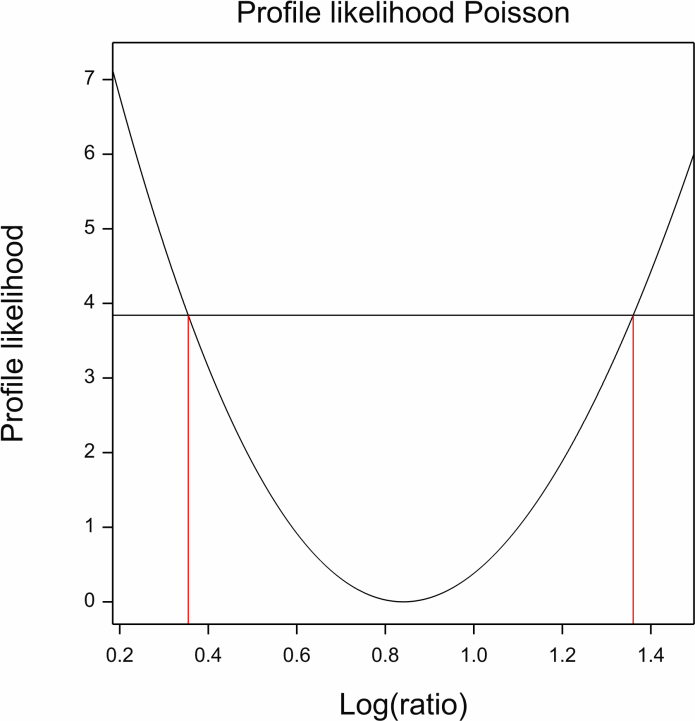
**Technical documentation Amiga Power Analysis Tool**

**Equivalence testing using the Profile likelihood** (see EquivalencePoisson.gen)

Equivalence testing in a log-linear can be performed by means of a profile likelihood confidence interval for the parameter of interest. Suppose we have a sample from two populations and it has to be tested whether the ratio of the two underlying means is outside an equivalence interval (L,U). As an example the first sample equals (3, 4, 7, 5, 3) and the second sample equals (11, 12, 9, 9, 10). Fitting a log‑linear model, without overdispersion, results in the following parameter estimates in which xx represents the difference of the two means:

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | estimate | s.e. | exp(estimate) |
| Constant | 1.482 | 0.213 | 4.400 |
| xx | 0.841 | 0.255 | 2.318 |

The deviance of this model equals 3.080. A 95% Wald interval for the difference on the log-scale is then given by 0.841 ± 1.96 x 0.255. The interval on the original scale is then given by back-transforming the interval giving (1.406, 3.821). Alternatively the profile likelihood for the difference can be constructed by repeatedly fitting the model with a fixed difference between the two samples. This can be done by using an offset. This results in the Profile likelihood in the graph to the left. Using a critical value of 3.841, which is the 95% point of the chi-squared(1) distribution, this results in a LR interval for the difference of (0.355, 1.360) or (1.426, 3.897) on the ratio scale. Note that this is only slightly different from the Wald interval.

Now suppose that the lower limit of concern equals the lower confidence limit 1.426. Fitting a log‑linear model such that the ratio of the two means equals this LOC, again employing an offset, results in a deviance of 6.921. The difference between the deviance of this restricted model and the deviance of the full model equals 6.921-3.080 = 3.841 which is the critical value used in constructing the profile likelihood interval. The same holds when the limit of concern is set to the upper confidence limit 3.897.

This implies that LR equivalence testing can be performed in the following way:

1. Fit the full model. Save the deviance as dev1, and also save the difference parameter
2. In case the difference parameter is outside the interval (log(L), log(U)) the null hypotheses of non-equivalence is not rejected and we are done.
3. In case the difference parameter is smaller than 0 the relevant limit of concern is L. Fit the model with an fixed ratio L of the two means and save the difference as dev0.
4. In case the difference parameter is larger than 0 the relevant limit of concern is U. Fit the model with an fixed ratio R of the two means and save the difference as dev0.
5. Compare the deviance difference (dev0-dev1) with the 95% point of a chi-squared(1) distribution. In case the deviance difference is larger than the critical value reject the null hypothesis of non-equivalence in favour of the alternative hypothesis of equivalence.

**Comparisons, interaction factors and modifiers**

The default comparison is between the GMO and the comparator (CMP) averaged over all levels of all other factors. This implies that the other factors provide effective replications.

The default comparison can be changed in the Design tab (for all endpoints), in the Interactions tab (making the comparison endpoint specific) and finally in the Comparisons tab to make it specific for comparisons. Suppose we have two additional factors F and G with 3 and 2 levels respectively. Further suppose that we would like to make the following comparison which is averaged over all levels of G:

|  |  |  |
| --- | --- | --- |
|  | GMO | CMP |
| F1 | 1 | 0 |
| F2 | 0 | 1 |
| F3 | 0 | 1 |

F is called an interaction factor. It is then assumed that the mean count of the CMP for F2 and F3 are both equal to the general mean (as given in the Endpoints data tab), and the mean count of the GMO for F1 equals in which is some factor which is defined by the limits of concern in the Endpoints tab. The dummies which are used in fitting the associated model for F are given below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variety | F | G | Comparison | Constant | Dum[1] | Dum[2] | Dum[3] | Dum[4] |
| GMO | F1 | G1 | 1 | 1 | 1 | 0 | 0 | 0 |
| GMO | F1 | G2 | 1 | 1 | 1 | 0 | 1 | 0 |
| GMO | F2 | G1 | 0 | 1 | 0 | 1 | 0 | 0 |
| GMO | F2 | G2 | 0 | 1 | 0 | 1 | 0 | 0 |
| GMO | F3 | G1 | 0 | 1 | 0 | 0 | 1 | 0 |
| GMO | F3 | G2 | 0 | 1 | 0 | 0 | 1 | 0 |
| CMP | F1 | G1 | 0 | 1 | 0 | 0 | 0 | 1 |
| CMP | F1 | G2 | 0 | 1 | 0 | 0 | 0 | 1 |
| CMP | F2 | G1 | -1 | 1 | 0 | 0 | 0 | 0 |
| CMP | F2 | G2 | -1 | 1 | 0 | 0 | 0 | 0 |
| CMP | F3 | G1 | -1 | 1 | 0 | 0 | 0 | 0 |
| CMP | F3 | G2 | -1 | 1 | 0 | 0 | 0 | 0 |

Note that although the full interaction table Variety\*F contains 6 cells, there are only 5 parameters associated with the table (including the constant parameter). This is because it is assumed that F2 and F3 for CMP have equal means. The vector Dum[1] in the model Constant + Dum[1,2,3,4] is associated with the statistical test of the comparison. More specifically a statistical test can be constructed by comparing the fit of the full model “Constant + Dum[1,2,3,4]” with the fit of the restricted model “Constant + Dum[2,3,4]”.

The remaining factor G in this example can be specified as a modifier. This will modify the mean count of G1 and G2 in the same way for all levels of the factors Variety and F. This implies that a modifier is an additive term in the model, similarly to (random) block effects. Also a modifier provides extra levels of replication for the comparison.

As an example suppose the general mean for CMP equals =10. Further suppose that the other means in the Variety\*F table are specified as below. Note that in the Comparions tab the user will not be able to modify the mean count for cells that are part of the comparison, and also that is set to 1 to specify the mean counts for the levels which are associated with the GMO comparison (in this case cell GMO-F1).

|  |  |  |
| --- | --- | --- |
|  | GMO | CMP |
| F1 | 10 | 8 |
| F2 | 6 | 10 |
| F3 | 4 | 10 |

Further suppose that factor G is a modifier with multiplicative values 0.5 for G1 and 1 for G2. The basic means which are used in the simulation are then given by

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variety | F | G | Comparison | Mean |
| GMO | F1 | G1 | 1 | 5 |
| GMO | F1 | G2 | 1 | 10 |
| GMO | F2 | G1 | 0 | 3 |
| GMO | F2 | G2 | 0 | 6 |
| GMO | F3 | G1 | 0 | 2 |
| GMO | F3 | G2 | 0 | 4 |
| CMP | F1 | G1 | 0 | 4 |
| CMP | F1 | G2 | 0 | 8 |
| CMP | F2 | G1 | -1 | 5 |
| CMP | F2 | G2 | -1 | 10 |
| CMP | F3 | G1 | -1 | 5 |
| CMP | F3 | G2 | -1 | 10 |

In case there are extra varieties, in addition to GMO and CMP, the mean counts for these are set to the general mean , irrespective of any interaction factors, and modified according to the modifiers. However in fitting the model it is assumed that these extra varieties do interact with the interaction factors. In the example above, with F as interaction factor, this results in three extra dummies for every extra variety.

**Simulation distributions**

The mean count for CMP and its associated coefficient variation , as specified in the Endpoints data tab, define the dispersion parameter of the associated distribution. This dispersion parameter is then used for **all** other mean counts. For the power model, with variance function , the parameter is first derived as . This value of is the used for all mean counts levels. However since there is no true distribution associated with the power model, for every mean count level separately, a corresponding value of the negative binomial dispersion parameter is calculated by equating the variance functions of the power model and the negative binomial distribution. The negative binomial distribution, with a unit specific dispersion parameter, is then used to simulate data.

In the Modifiers tab the user can specify a coefficient of variation for block effects. In that case the basic model for the mean counts is in which represents all the factors in the model and is a random normally distributed block effect. For a lognormal distribution the coefficient of variation is given by . and this equation is used to calculate the variance of the random block effect. Then for every simulated dataset random block effects are generated and applied to the mean counts. Note that block effects are always part of the fitted model, even in case blocks are not considered to be modifiers.

**LN and SQ models for analysis**

The LN and SQ models simply transform the simulated data and then perform a linear regression on the transformed scale. The two-sided difference test for the comparison is obtained by fitting the full model and then calculating the squared t-value for dummy Dum[1]. This is then compared with a critical F‑value with appropriate degrees of freedom.

The equivalence test uses the generalized confidence interval (GCI) approach as described in one of the AMIGA reports. This requires estimated means for the CMP and the GMO. In case there are no modifiers these are given simply by the estimate for the Constant in the model (for CMP) and by the sum of the estimates for the Constant and Dum[1] (for GMO). However when there are modifiers it is necessary to average over the modifiers. In GenStat this can be done by using the PREDICT directive with dum[1] set to 0 (for CMP) and 1 (for GMO) and all other dummies to 0. GCI also requires the effective level of replication of both estimates. This is calculated as where equals the standard errors of the two predicted means.

**OP model for analysis**

A deviance based difference test is obtained by fitting the full model and the restricted model. The deviance difference is scaled by Pearsons statistic and the test statistic is compared to an F distribution.

The equivalence test is based on comparing the deviance of the model in which the ratio of the GMO and CMP is set to the limit of concern, employing an offset in the model, with the deviance of the full model. This is only necessary when the estimate of Dum[1] is inside the interval set by the limits of concern, since when it is outside this interval the equivalence hypothesis will not be rejected. In case the estimate of Dum[1] is smaller than zero the relevant null-hypothesis concerns the lower limit of concern, while when the estimate is larger than zero the upper limit of concern is of interest. Again Pearsons statistic is used to scale the deviance difference and the test statistic is compared with the F statistic. This approach is checked by means of the GenStat program EquivalencePoisson.gen.

**NB model for analysis**

The approach is largely the same as for the OP model. The only difference is that the deviance difference is not scaled and that the test statistics are compared with a critical value based on the Chi-squared distribution.

**Timing in GenStat**

Fitting the NB model is notably slow since this uses a robust bisection algorithm to estimate the dispersion parameter. On the other hand fitting the LN and SQ models is fast although the GCI approach can be time consuming when the number of samples is large.

The following settings were used for timing of the example with factors F and G in the example above

1. Using a grid of 5 values for the difference between limits of concern 0.5 and 2.0
2. 6 difference replications 2,3,4,5 , 6 and 7
3. Blocking with a 10% coefficient of variation
4. 10000 samples for the GCI approach for LN and SQ
5. simulation by means of the negative binomial distribution
6. 10 datasets were simulated for each 5x6 = 30 settings

|  |  |
| --- | --- |
| Fitted models | Seconds |
| LN SQ | 24 |
| LN SQ OP | 28 |
| LN SQ OP NB | 191 |

Hopefully multiple instances of Rdotnet are possible so that multiple processors can be used simultaneously for different parts of the simulation.

**The method of Lyles using a synthetic dataset**

A synthetic dataset for a multiplicative effect and number of replication is constructed as follows:

1. Construct the dummies associated with the variety factor and possibly with the interaction between the variety factor and other treatment factors. Set up factors representing additive effects (called modifiers above), and create the vector of means with equal to the desired value. This results in a design matrix.
2. Expand the design matrix according to the number of blocks or replicates.
3. Create a block factor if required. Derive the expected effect of blocking by means of Blom normal scores. More specifically a) calculate the between block variance as . Calculate multiplicative block effects as in which the Blom scores are given by . Finally multiply the expanded means with the block effects.
4. For every element in the thus acquired vector of means obtain the vector of possible outcomes and associated probabilities according to the simulation distribution. This should be such that the sum of the vector of probabilities is close to 1.
5. Stack the vectors and into single vectors and and also stack the expanded design matrix to give a new design matrix. The synthetic dataset is now ready.
6. Do a weighted regression of on the design matrix with weights . Obtain the deviance of the full model and of the restricted model. The non-centrality parameter for the difference test is given by the deviance difference scaled by an estimate of the variance (see below). The power of the difference test is then approximated by P( where denotes the non-central Chi-squared distribution and denotes the 1- percentage point of the (central) chi-squared distribution.
7. The power of the equivalence test is obtained in a similar way by fitting the restricted model in which the difference between the GMO and CMP equals log(L) or Log(U) depending on whether is smaller or greater than 1.

The non-centrality parameter for the normal case (i.e. LN and SQ transformations) and for the overdispersed Poisson case require an estimate of the dispersion. For the LN and SQ case such an estimate is obtained by the mean variance of the transformed data. For the LN case, for every element in step 4 above, the weighted mean of the transformed data is calculate as and the variance of the transformed data by . An estimate of the dispersion (or residual variance) is then given by the mean of the values. The variance for the SQ transformation is calculated similarly. For the OP model the dispersion parameter equals the ratio between the variance and the mean. This ratio is calculated for every element in step 4 and the mean of these values is used as an estimate of the dispersion. Moreover for these distributions the F distribution with appropriate denominator degrees of freedom is used instead of the distribution.

Note that Lyles method cannot be used to approximate the power of the LN and SQ equivalence tests.

In case there are no blocks the non-centrality parameter is exactly proportional to the number of replications. This is because the expanded dataset for e.g. 4 replications is exactly equal to 2 replicates of the expanded dataset for 2 replications. The deviances for 4 replicates are therefore twice the deviances for 2 replicates. When there are blocks there is no exact proportionality due to the Blom normal scores block effects which are different for different number of blocks. An alternative would be to “estimate” the (linear) relationship between noncentrality and number of replications for each level of the Ratio separately and then interpolate and extrapolate for other numbers of replication. extrapoconstant by doing the calculations for say =2 and =8 and then linearly interpolate and extrapolate. This was tested for a dataset with small means (see Appendix A), the overdispersed Poisson distribution, CVcomparator=150, CVblock=10, lowerLOC=0.25, upperLOC=4.0, a grid of 9 Ratio points in between the LOCs equidistant on the Log-scale, number of blocks 2,4,8,16,32,64. The maximal difference between power values estimated from a pair of blocks and the true values is given in the table below (basic results in ToolLyles-SmallMeans1.xlsx). relDiffNc is the maximal relative difference in the non-centrality parameter and DiffPow the maximal difference in the power itself.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Pair (2,4) | | Pair (2,8) | | Pair (2,16) | |
| Test | relDiffNc | DiffPower | relDiffNc | DiffPower | relDiffNc | DiffPower |
| LN diff | 9.56E-04 | 0.0004 | 4.56E-04 | 0.0002 | 2.74E-04 | 0.0001 |
| SQ diff | 1.20E-03 | 0.0005 | 4.54E-04 | 0.0002 | 3.37E-04 | 0.0001 |
| PO diff | 2.89E-03 | 0.0012 | 1.11E-03 | 0.0004 | 1.20E-03 | 0.0003 |
| OP diff | 2.89E-03 | 0.0011 | 1.11E-03 | 0.0004 | 7.31E-04 | 0.0002 |
| NB diff | 3.11E-03 | 0.0013 | 1.03E-03 | 0.0004 | 7.04E-04 | 0.0003 |
| PO equiv | 4.32E-03 | 0.0015 | 1.62E-03 | 0.0003 | 2.07E-03 | 0.0004 |
| OP equiv | 4.32E-03 | 0.0017 | 1.18E-03 | 0.0005 | 1.01E-03 | 0.0003 |
| NBequiv | 1.47E-01 | 0.0016 | 1.47E-01 | 0.0004 | 1.47E-01 | 0.0002 |

The same case, but now with CVcomparator=400 and CVblock=50 yields (ToolLyles-SmallMeans2.xlsx)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Pair (2,4) | | Pair (2,8) | | Pair (2,16) | |
| Test | relDiffNc | DiffPower | relDiffNc | DiffPower | relDiffNc | DiffPower |
| LN diff | 1.52E-02 | 0.0057 | 9.69E-03 | 0.0035 | 5.17E-03 | 0.0019 |
| SQ diff | 1.44E-02 | 0.0057 | 9.04E-03 | 0.0032 | 4.84E-03 | 0.0018 |
| PO diff | 1.58E-02 | 0.0056 | 9.86E-03 | 0.0023 | 5.96E-03 | 0.0026 |
| OP diff | 1.58E-02 | 0.0062 | 9.52E-03 | 0.0039 | 5.62E-03 | 0.0021 |
| NB diff | 1.58E-02 | 0.0069 | 8.69E-03 | 0.0032 | 4.30E-03 | 0.0016 |
| PO equiv | 2.52E-01 | 0.0067 | 2.48E-01 | 0.0082 | 3.50E-01 | 0.0257 |
| OP equiv | 2.52E-01 | 0.0866 | 2.48E-01 | 0.0852 | 3.50E-01 | 0.1266 |
| NBequiv | 1.49E-02 | 0.0057 | 8.57E-03 | 0.0037 | 4.21E-03 | 0.0018 |

It is evident that, in this particular case, the differences are very small and that it is save to use a pair of Blocks. There is only one exception and that is the OP equivalence test for a Ratio=1 and Nreps=16. The non-centrality parameter for that case does not fit in with the rest. It is unclear why.

Assuming a relationship between non-centrality and number of replications will speed up the calculations considerably because it is then only necessary to construct the synthetic dataset for a small number of replications. This is especially the case when the mean counts are large in combination with a large coefficient of variation because then the synthetic dataset can be quite extensive. For example for the dataset discussed in the “Comparison” section in combination with a CV of 200 for the comparator and a CV of 50 for block effects, the synthetic datasets has over 38.000 units for 16 blocks.

**Important note**: The “approximated” type I error of the Lyles method is exactly equal to the significance level of the test. However we know from a previous simulation study that this is not always the case. So a simulation approach to estimate the type I error remains necessary.

**Appendix A: Dataset with small means**

interaction between variety and Ranking; additive Spraying effect

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variety | Ranking | Spraying | Mean | Comparison |
| GMO | 1 | 1 | 1 | IncludeGMO |
| GMO | 1 | 2 | 2 | IncludeGMO |
| GMO | 2 | 1 | 0.5 | Exclude |
| GMO | 2 | 2 | 1 | Exclude |
| GMO | 3 | 1 | 0.25 | Exclude |
| GMO | 3 | 2 | 0.5 | Exclude |
| Comparator | 1 | 1 | 1.5 | Exclude |
| Comparator | 1 | 2 | 3 | Exclude |
| Comparator | 2 | 1 | 1 | IncludeComparator |
| Comparator | 2 | 2 | 2 | IncludeComparator |
| Comparator | 3 | 1 | 1 | IncludeComparator |
| Comparator | 3 | 2 | 2 | IncludeComparator |