

Log Likelihood Ratios for Common Statistical Tests Using the `likelihoodR` Package

by Peter Cahusac

Abstract The `likelihoodR` package has been developed to allow users to obtain statistics according to the likelihood approach to statistical inference. Commonly used tests are available in the package, such as: t tests, ANOVA, correlation, regression and a range of categorical analyses. In addition, there is a sample size calculator for t tests, based upon the concepts of strength of evidence, and the probabilities of misleading and weak evidence.

1 Introduction

Maximum likelihood estimation (MLE) is well-understood and widely used throughout statistics. In contrast, the use of the likelihood function as a basis for inference is much less understood and even confused with MLE. As Edwards wrote in his excellent book: "At one recent international conference at which I laboured for three-quarters of an hour to make clear the advantages of likelihood inference, the chairman thanked me for my lecture on the Method of Maximum Likelihood" (Edwards, 1992) p 101. In the Epilogue of this book on p 212, Edwards says that MLE is "a red herring". To clarify: MLE is used to estimate a parameter value according to a supposed probability distribution, while likelihood inference is used to compare two hypotheses through the ratio of their values on a likelihood function.

All the main statistical approaches to scientific inference (frequentist, Bayesian and information criterion) are based upon calculated probabilities. The frequentist approach, for example, typically uses a sampling distribution centred on a null hypothesis and calculates the probability of obtaining the observed value or values more extreme from the null. The likelihood approach, also known as the evidential approach, differs in that it is simply based upon the evidence provided by the observed data (not including more extreme values) and represented by the likelihood function. The ratio of the heights under the likelihood function according to the hypothesis values tested provides the likelihood ratio. To create a linear scale of evidence, the natural logarithm of this is taken to give us the log likelihood ratio, also known as the support, a term that was first defined by Harold Jeffreys (Jeffreys, 1936).

The likelihood approach is not subject to the same criticisms often leveled at the frequentist and Bayesian approaches (Goodman and Royall, 1988; Edwards, 1992; Royall, 1997; Goodman, 1999; Dixon, 2003; Dienes, 2008; Wasserstein and Lazar, 2016; Lakens, 2021). Likelihood ratios provide objective measures of evidence between competing hypotheses, unaffected by the intentions of the investigator. Log likelihood ratio (*support*, see below) values are proportional to the quantity of data, representing the weight of evidence. This means that support values from independent studies can simply be added together, e.g. for meta-analysis. Unlike other approaches based on probabilities, likelihood ratios are unaffected by transformations of variables.

Despite the apparent advantages of the evidential approach there is a dearth of available resources in statistical computing. None of the major commercial packages (e.g. SPSS, SAS, Minitab) provide likelihood ratios or support values. This should not be confused with the wide availability of likelihood ratio tests (also known as G-tests), which ultimately provide p values according to the frequentist approach. Because virtually no software is available for analysis, this impacts on the use of the evidential approach in scientific reporting. This also impacts on the teaching of the evidential approach, which then negatively feeds back to reduced scientific reporting. The `likelihoodR` package for R (Cahusac, 2021) is an attempt to address this situation and calculations are based upon the recent book (Cahusac, 2020b). R (Ihaka and Gentleman, 1996) is a widely used statistical platform with a huge variety of packages.

2 Support

The package always reports the support (log likelihood ratio). This relative evidence scale ranges from $-\infty$ to $+\infty$, with zero representing no evidence either way. A working interpretation for this scale has been offered by (Goodman and Royall, 1988), see Table 1. No threshold need be applied, and S values are given to just one decimal place, e.g. 2.3.

S	Interpretation of H_1 vs H_2
0	No evidence either way
1	Weak evidence
2	Moderate evidence
3	Strong evidence
4	Extremely strong evidence

Table 1: Interpretation for values of S , the support, calculated as the natural logarithm of the likelihood ratio. Negative values would represent support for hypothesis values H_2 vs H_1 . Typically, it is sufficient to give S to one decimal place.

The support (S) values reported in the package are distinct from the surprise/surprisal S -values described by (Palm, 2012) based on Shannon information theory, and that by (Greenland, 2019) produced by simply taking the negative log base 2 of the p value.

The likelihood intervals are reported wherever possible and these are given in terms of support rather than likelihood ratio. A typical likelihood interval (support interval) is the S -2 interval, which numerically often closely corresponds to the frequentist 95% confidence interval (Cahusac, 2020b). The S -2 interval represents an interval based upon $e^{-2} = 0.135 = 1 / 7.40$ likelihood ratio interval. All the values within the interval would have a likelihood ratio of no less than $1 / 7.40$, or an $S = -2$. The S -3 interval would be $e^{-3} = 0.05 = 1 / 20.09$, and so on. Analyses also report other relevant statistics, such as t , F and χ^2 , as well as the corresponding frequentist p value. Where possible the likelihood function is given, decorated with hypothesis parameter values shown as coloured lines. Currently there are 14 different statistical tests implemented by the package, the summary for each are given in Table 2 (listed alphabetically).

Function	Description
L_1way_ANOVA	Independent samples one-way ANOVA
L_1way_cat	One-way categorical data analysis for binomial and multinomial
L_1way_RM_ANOVA	One-way repeated measures ANOVA
L_2S_ttest	Independent samples t test
L_2way_cat	Two-way categorical data analysis
L_2way_Factorial_ANOVA	Two-way independent samples factorial ANOVA
L_corr	Bivariate normal correlation
L_efficacy	Efficacy analysis for binomial categorical data
L_logistic_regress	Multiple logistic regression
L_OR	Odds ratio
L_regress	Bivariate regression for linear, quadratic and cubic comparisons
L_RR	Relative risk
L_ttest	One sample and related samples t test
L_t_test_sample_size	Sample size calculation using the evidential approach for t tests

Table 2: A summary of the functions available in the `likelihoodR` package

3 Continuous data

A range of tests for differences of continuous variables is available. One function `L_ttest` performs a one sample and related samples test. As well as specifying the null value, two alternative hypothesis values can be specified, one in terms of the measurements used and the other in terms of Cohen's d .

```
L_ttest(data1, data2, null=0, d=0.5, alt.2=NULL, L.int=2, verb=TRUE)
```

Where the arguments are:

```
data1  a (non-empty) numeric vector of data values
data2  a (non-empty) numeric vector of data values for related sample, default = NULL
null   value for the null hypothesis, default = 0
d      Cohen's effect size, default = 0.5
alt.2  value for an alternative hypothesis, in units used for data, default = NULL
L.int  likelihood interval given for a given support value, e.g. 2 or 3, default = 2
verb   show output, default = TRUE
```

As an example:

```
> # one sample Gosset's original additional hours of sleep data Cahusac (2020) p 29
> mysample <- c(0.7, -1.6, -0.2, -1.2, -0.1, 3.4, 3.7, 0.8, 0.0, 2.0)
> a1=L_ttest(mysample, d=0.5, alt.2=2, L.int=2)
```

```
Maximum support for the observed mean 0.75 (dashed line) against the null 0 (black line) = 0.892
Support for d of 0.5 (0.8945048, blue line) versus null = 0.856
Support for d versus 2nd alt Hypothesis 2 (green line) = 2.131
Support for 2nd alt Hypothesis versus null = -1.275
```

```
S-2 likelihood interval (red line) is from -0.44025 to 1.94025
```

```
t(9) = 1.326, p = 0.2175978, d = 0.419
```

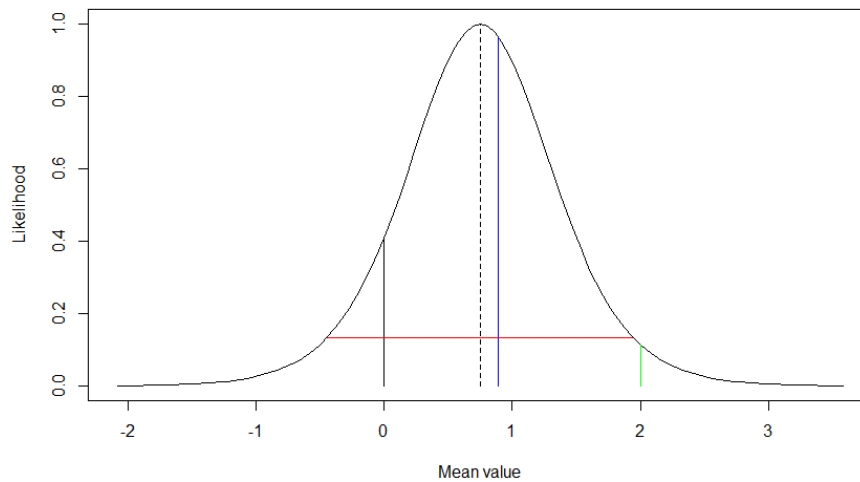


Figure 1: Graphical output from running the one-sample test function `L_ttest`. The likelihood function for the mean value given the data, with the MLE (sample mean) represented by the vertical dashed line. The black line is the specified null value (here at 0), the blue line is for a specified alternative hypothesis (for an effect size of 0.5), and the green line represents a second alternative hypothesis (for 2 hours). The red horizontal line shows the S-2 likelihood interval.

The assigned object `a1` contains values of interest which can be assessed in the usual way with `a1$...`, but entering `a1` on its own gives all the values:

\$obs.mean	the observed mean or difference in mean for related samples
\$df	degrees of freedom
\$alt.H1	mean value according to specified d
\$alt.H2	specified second hypothesis value
\$S_max	maximum support for observed mean against the null
\$S_10	support for d versus null
\$S_12	support for d versus specified second hypothesis
\$S_20	support for second hypothesis versus the null
\$like.int	likelihood interval
\$L.int.spec	specified likelihood interval in units of support
\$null.value	null value
\$t.val	t value for test against null
\$p.val	p value for test against null
\$d.obs	observed effect size

For the related sample test a second vector of values of equal length to data1 (and paired) is entered for data2. The independent samples test function is `L_2S_ttest`. The data is entered as the first argument followed by the vector of the same length coding groups. Similar arguments to the one sample/related sample test can be specified and similar output, but this time showing the difference in means.

The package includes one-way ANOVA (`L_1way_ANOVA`), one-way repeated measures ANOVA (`L_1way_RM_ANOVA`) and two-way between-participants factorial ANOVA (`L_2way_Factorial_ANOVA`). All these functions use contrasts, employing the model comparison approach espoused by Glover and Dixon (Dixon, 2003; Glover and Dixon, 2004; Dixon, 2013). For the one-way analyses if the arguments for the contrasts are not specified then they default to testing a linear and a quadratic contrast. In the factorial ANOVA, no contrast comparisons are made if the contrasts are left unspecified. If the first contrast is specified then this is compared to the main effects, and if the second contrast is specified then this is compared to the first contrast. Such contrast comparisons are easy to do in this approach, and are difficult or impossible to do using frequentist *p* values. The support values for the between-participants analyses are adjusted using Akaike's correction (Hurvich and Tsai, 1989). We will look at `L_2way_Factorial_ANOVA`. The first line of output gives the S for comparing the full model of main effects and interaction with the null model. Using data from Cahusac (2020b) p 91

```
> time <- c(6.4, 4.6, 6.4, 5.6, 5.9, 6.1, 6.3, 4.5, 4.8, 6.6, 7, 9.3, 7.9, 9.4, 8.2,
           4.4, 4.2, 5, 6.9, 4.5, 4, 4.3, 6.9, 5.5, 5.8, 4.4, 4.2, 5.1, 6.9, 4.5)
> Treatment = gl(3,5,30, labels=c("T1", "T2", "T3"))
> Health = gl(2,15,30, labels=c("Hemophiliac", "Normal"))
> contrast1 <- c(-1, -1, 5, -1, -1, -1) # interaction Hemo T3 higher than others
> contrast2 <- c(-1, -1, -1, 1, 1, 1) # main effect of health status (Hemo higher)
> m1=L_2way_Factorial_ANOVA(time, Treatment, Health, contrast1, contrast2, verb=TRUE)
```

```
Support for full model (including interaction) versus null = 7.036
Support for full model versus main effects = 2.968
Support for contrast 1 versus main effects = 7.587
Support for contrast 1 versus contrast 2 = 8.462
```

```
First factor main effect F(2,24) = 5.039, p = 0.01488452, partial eta-squared = 0.296
Second factor main effect F(1,24) = 15.992, p = 0.0005281828, partial eta-squared = 0.4
Interaction F(2,24) = 6.219, p = 0.006664992, partial eta-squared = 0.341
Contrast 1 F(1,24) = 36.048, p = 3.373928e-06
```

As before, the assigned object `m1` contains values of interest which can be assessed in the usual way with `m1$...`, but entering `m1` on its own gives all the values.

Bivariate normal correlation uses the `L_corr` function, and is demonstrated by using data from the heptathlon (Cahusac, 2020b) p 104

```
> m200 <- c(22.6, 23.7, 23.1, 23.6, 23.6, 23.6, 25.5, 23.9, 24.5, 23.9, 24.9, 24.8, 24.7,
           25.0, 24.6, 24.9, 25.0, 25.6, 24.8, 25.5, 25.7, 24.9, 26.6, 25.2, 26.2)
> m800 <- c(128.5, 126.1, 124.2, 132.5, 134.7, 132.5, 138.5, 127.9, 133.7, 132.2, 136.1, 142.8,
           125.8, 131.5, 137.1, 134.9, 146.7, 133.9, 146.4, 144.0, 133.4, 138.0, 139.2, 137.3, 163.4)
> m2=L_corr(m200, m800, null=0, exp.r=0.5, L.int=3, alpha=.05, verb=TRUE)
```

```
Support for observed correlation 0.6198 (dashed line) versus null of 0 (black line) = 5.776
Support for specified correlation of 0.5 (blue line) versus observed r = -0.338
```

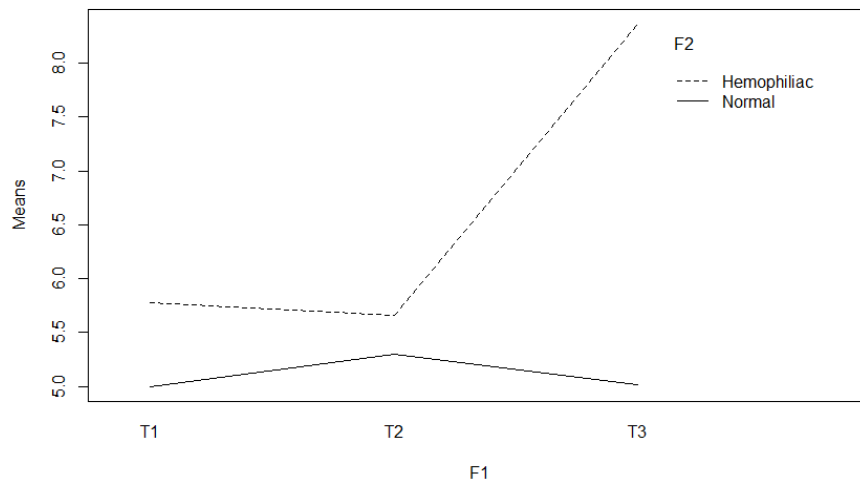


Figure 2: Graphical output from running the `L_2way_Factorial_ANOVA` function. The interaction plot shows the means for the 2 groups of patients (dashed line is hemophiliac, solid line is normal). The horizontal axis represents the 3 different treatments. An interaction is apparent from the plot, where the 3rd treatment shows a clear difference between the hemophiliac and normal patients.

Support for specified correlation versus null = 5.438
S-3 likelihood interval (red line) is from 0.19968 to 0.8474

P value = 0.00095
N = 25

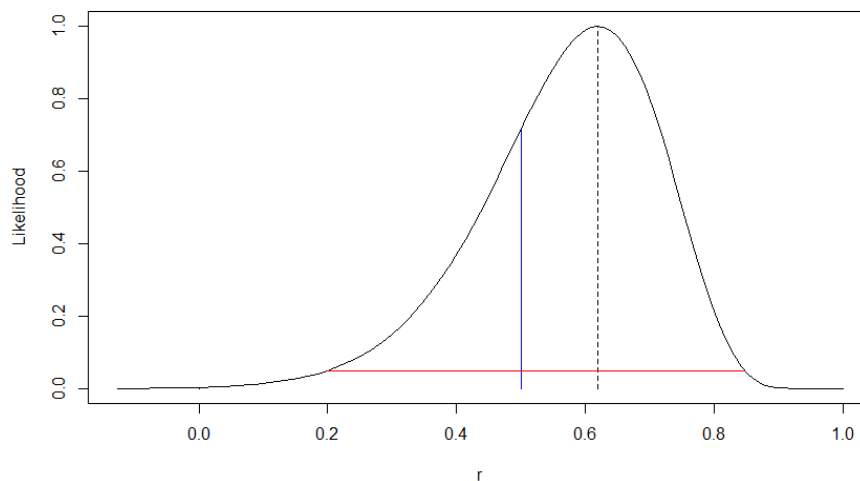


Figure 3: Graphical output from running the `L_corr` function. This is the likelihood function for the correlation coefficient, given the data. As before, the vertical dashed line is the MLE (the sample r), the blue line is a specified alternative value (0.5), and red horizontal line is the S-3 likelihood interval. The null value at 0 is outside of this interval and the evidence against it is extremely strong.

The assigned object `m2` contains values of interest which can be assessed in the usual way with `m2$...`, but entering `m2` on its own gives all the values.

The regression function `L_regress` only accommodates one predictor, while the logistic regression function `L_logistic_regress` allows up to 6 predictors, which need to be dummy coded for nominal data with more than 2 levels.

4 Categorical data

There are 5 different tests included in the package (excluding logistic regression mentioned above). The simplest is the one-way categorical data analysis using the function `L_1way_cat`. Two categories represents the binomial (giving the likelihood function plot), while multiple categories represents the multinomial distribution. The two-way categorical analysis uses the `L_2way_cat` function. For these two functions an additional evidence-based statistic S for the variance is calculated. This uses the formula given by Cahusac (2020b) p 158 and derived from Edwards (1992) p 187:

$$S = \frac{df}{2} \left(\log \frac{df}{\chi_{df}^2} \right) - \frac{1}{2}(df - \chi_{df}^2), \quad (1)$$

where df is the degrees of freedom. This is most useful to test the variance in the model, specifically whether data are "too good to be true", i.e. the data fit a particular hypothesis closer than we would expect by chance (Edwards, 1986). Using just 2 categories in the one-way analysis can be demonstrated:

```
> obs <- c(18,5); exp.p <- c(0.7, 0.3) # observed and expected values
> m3 <- L_1way_cat(obs, exp.p, verb = TRUE)
```

```
Binomial support for difference of MLE 0.7826087 (dashed line)
from 0.7 (blue line) with 1 df = 0.398
Support for variance differing more than expected = 0.019
```

```
S-2 likelihood interval (red line) from 0.5853 to 0.91794
```

```
Chi-square(1) = 0.747, p = 0.3872968
Likelihood ratio test G(1) = 0.795, p = 0.37258, N = 23
Likelihood-based 95% confidence interval from 0.58958 to 0.91597
```

As previously, the assigned object `m3` contains values of interest which can be assessed in the usual way with `m3$...`, but entering `m3` on its own gives all the values.

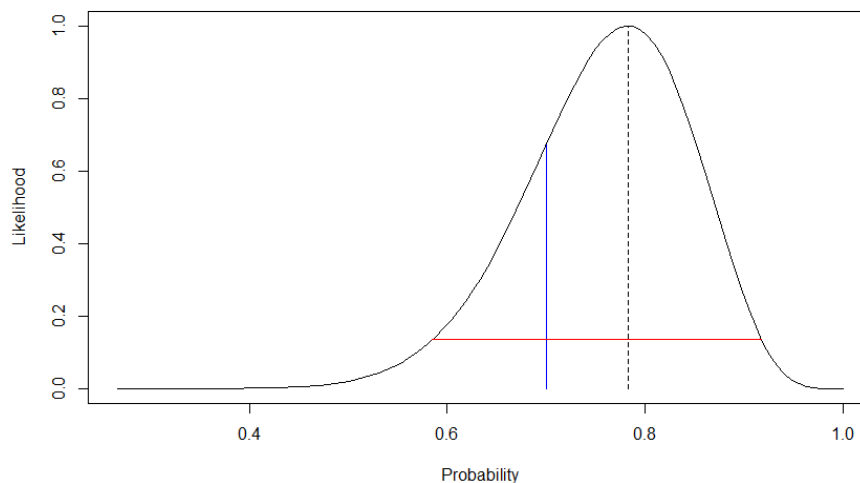


Figure 4: Graphical output from running the `L_1way_cat` function. This shows the likelihood function for the proportion, given the data. The vertical dashed line is the MLE, the blue line is the alternative hypothesis value (0.7), and the red horizontal line is the S-2 likelihood interval.

Some of the functions provide likelihood-based % confidence interval (Aitkin et al., 1989). Finally, there are functions to calculate the odds ratio `L_OR`, relative risk `L_RR` and binomial efficacy `L_efficacy`.

5 Sample size calculations

The main challenge faced by the evidential researcher is of obtaining sufficiently strong evidence for or against one of two specified hypotheses. Like the probability of a Type II error, this probability is large with a small sample size and decreases as sample size increases. The function `L_t_test_sample_size` can be used to calculate the pre-study sample size for all the t tests (Cahusac and Mansour, 2022). The combined misleading and weak probability (Royall, 1997, 2000, 2004) is entered, with a default of 0.05, together with the strength of evidence desired (default = 3). For a paired samples test, where we wish to calculate sample size with a combined .2 probability of obtaining misleading or weak evidence, strength of evidence $S = 2$ and effect size 0.5, we would obtain a value of 38 by using the following:

```
> L_t_test_sample_size(MW = 0.2, sd = 1, d = 0.5, S = 2, paired = TRUE)
```

```
For 1 sample, or related samples, t test with M1 + W1 probability of 0.2
Strength of evidence required is 2, and effect size of 0.5
Required sample size = 38
```

The somewhat comparable calculation for Type II error of 0.2, two-sided alpha = 0.05 and same effect size of 0.5 produces a sample size of 34 (using `stats::power.t.test(power = .80, delta = 0.5, sig.level=0.05, type="paired")`).

6 Conclusions

The functions described for the likelihoodR package may be useful for those researchers and statisticians who wish to use the evidential approach for their data analysis (Cahusac, 2020a). In addition to the advantages mentioned earlier in the introduction there are other desirable features. First, categorical data analyses are not restricted by normality assumptions, and support values for independent components of cross-tabulated data sum precisely and algebraically (unlike such calculations in chi-square analyses). Second, in categorical and measurement analyses it is possible to show that the data fit the null (or other) hypothesis too well (e.g. for detection of data fraud). Third, analyses are versatile with unlimited complexity for model comparisons within a dataset, for example in ANOVA (Glover and Dixon, 2004).

As far as the author is aware, no other packages are available in R or other platforms. Currently users calculate likelihood ratios manually. This package addresses this shortcoming and hopefully will encourage more users to express their results in terms of log likelihood ratios.

One of the products of the likelihoodR package is that a module has been developed for jamovi (The jamovi project, 2021), named **jeva**, which includes many of these functions. Hopefully this will encourage further interest in the likelihood approach and facilitate teaching and research. The equivalent jamovi analysis is given earlier for the `L_ttest` produces output given in Figure 5. The output produced by jamovi is identical to that produced earlier by the package, although simpler in that it lacks the option of comparing an effect size (d , illustrated by the blue line in package output). The null versus observed (1st line of Support output) is -0.892 , while in the package it is 0.892 , the positive value being due to comparing observed versus the null. Other outputs match apart from decimal rounding, which can be selected in jamovi. The t , degrees of freedom (df) and p are the same for the null versus observed, although the jamovi output includes another line giving these statistics for the alternative hypothesis versus the observed. The jamovi output includes group descriptive statistics (although these are available from the assigned object in the package, e.g. `$obs.mean`). The jamovi output also includes the option of a descriptives plot (not shown) which displays the mean with specified likelihood interval.

The likelihoodR package described in this article provides a large number of functions, currently many more than envisaged for the jamovi module. As such, it will provide a major reference package for users interested in the likelihood approach.

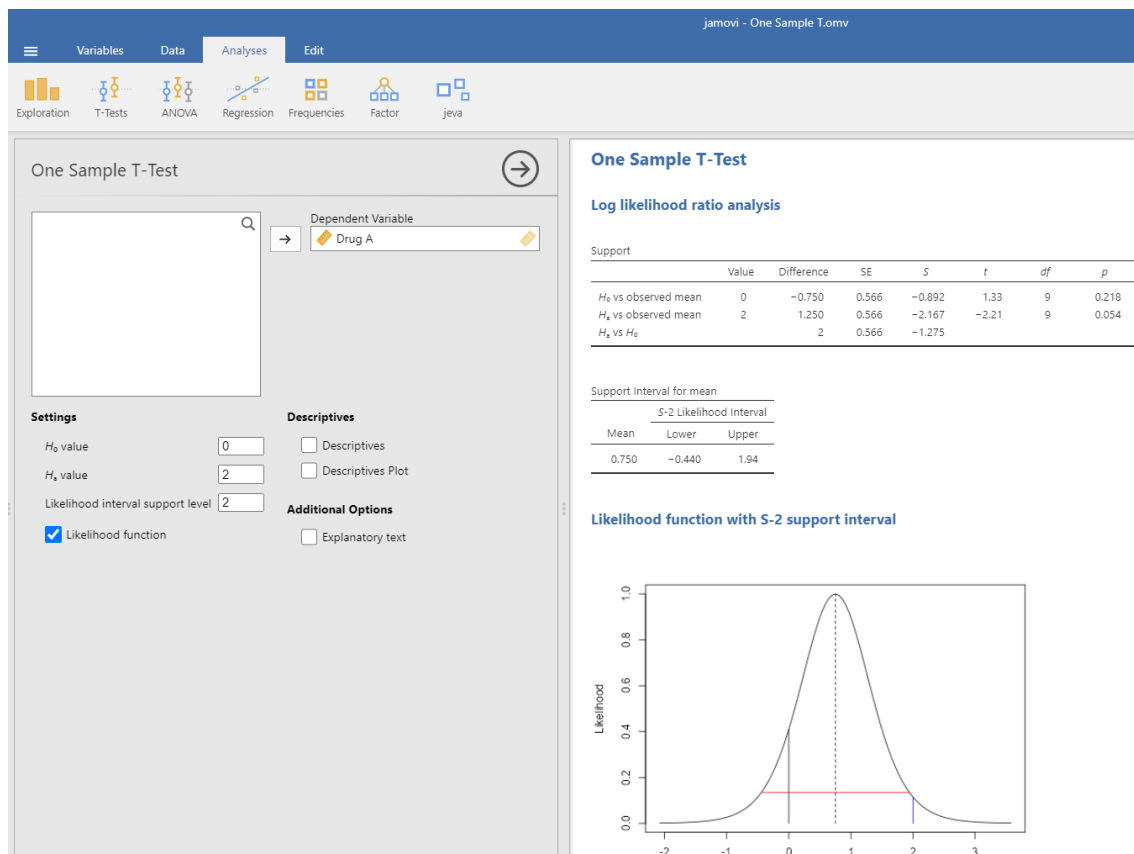


Figure 5: Graphical output from running the one-sample t test in the jamovi module **jeva** (created from the `L_ttest` function). The screenshot should be compared directly with Figure 1. On the left side is the dialog box where a variable can be selected (Drug A), and various settings chosen, including the null and alternative hypothesis values, and to display the likelihood function. An additional option gives an explanation about the obtained support values and likelihood interval, and their interpretation. On the right side is a summary of the analysis, at the top showing the S values for hypothesis comparisons (see text for the interpretation of the tabulated values). Below this is shown the support interval, and finally below that the likelihood function. The line colours for the likelihood function are the same as those given in Figure 1, although there is no option for a specified effect size (no green line).

Bibliography

- M. Aitkin, D. Anderson, B. Francis, and J. Hinde. *Statistical modelling in GLIM*. Oxford Statistical Science Series, Oxford, UK, 1989. ISBN 978-0198522041. URL <https://www.amazon.co.uk/Statistical-Modelling-GLIM-Oxford-Science/dp/0198522045/>. [p209]
- P. Cahusac. Data as evidence. *Experimental Physiology*, 105(7):1071–1080, 2020a. ISSN 0958-0670. doi: 10.1113/ep088664. URL <https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/EP088664>. [p209]
- P. M. B. Cahusac. *Evidence-Based Statistics: An Introduction to the Evidential Approach – from Likelihood Principle to Statistical Practice*. John Wiley and Sons, New Jersey, 2020b. URL <https://onlinelibrary.wiley.com/doi/book/10.1002/9781119549833>. [p203, 204, 206, 208]
- P. M. B. Cahusac. *Package 'likelihoodR'*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL <https://cran.r-project.org/web/packages/likelihoodR/index.html>. [p203]
- P. M. B. Cahusac and S. E. Mansour. Estimating sample sizes for evidential t tests. *Research in Mathematics*, 9(1):1–12, 2022. ISSN 2768-4830. doi: 10.1080/27684830.2022.2089373. URL <https://www.tandfonline.com/doi/full/10.1080/27684830.2022.2089373>. [p209]
- Z. Dienes. *Understanding Psychology as a Science: An introduction to scientific and statistical inference*. Palgrave MacMillan, 2008. ISBN 978-0-230-54230-3. URL <https://www.amazon.co.uk/Understanding-Psychology-Science-Introduction-Statistical/dp/023054231X>. [p203]
- P. Dixon. The p-value fallacy and how to avoid it. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie experimentale*, 57(3):189–202, 2003. ISSN 1878-7290(Electronic),1196-1961(Print). doi: 10.1037/h0087425. URL <https://pubmed.ncbi.nlm.nih.gov/14596477/>. [p203, 206]
- P. Dixon. The effective number of parameters in post hoc models. *Behavior research methods*, 45(3): 604–612, 2013. ISSN 1554-3528. URL <https://link.springer.com/article/10.3758/s13428-013-0373-7>. [p206]
- A. W. F. Edwards. Are mendel's results really too close? *Biological Reviews*, 61(4):295–312, 1986. ISSN 1464-7931. doi: 10.1111/j.1469-185X.1986.tb00656.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-185X.1986.tb00656.x>. [p208]
- A. W. F. Edwards. *Likelihood*, volume 2nd edition. John Hopkins University Press, Baltimore, 1992. ISBN 0801844436. URL <https://www.amazon.co.uk/Likelihood-Dr-W-F-Edwards/dp/0801844452>. [p203, 208]
- S. Glover and P. Dixon. Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic bulletin and review*, 11(5):791–806, 2004. ISSN 1069-9384. doi: 10.3758/BF03196706. URL <https://link.springer.com/article/10.3758/BF03196706>. [p206, 209]
- S. N. Goodman. Toward evidence-based medical statistics. 1: The p value fallacy. *Annals of Internal Medicine*, 130(12):995–1004, 1999. ISSN 0003-4819. doi: 10.7326/0003-4819-130-12-199906150-00008. URL <http://dx.doi.org/10.7326/0003-4819-130-12-199906150-00008>. [p203]
- S. N. Goodman and R. M. Royall. Evidence and scientific research. *American Journal of Public Health*, 78(12):1568–1574, 1988. ISSN 0090-0036. doi: 10.2105/AJPH.78.12.1568. URL <https://ajph.aphapublications.org/doi/10.2105/AJPH.78.12.1568>. [p203, 204]
- S. Greenland. Valid p-values behave exactly as they should: Some misleading criticisms of p-values and their resolution with s-values. *The American Statistician*, 73(1):106–114, 2019. ISSN 0003-1305. doi: 10.1080/00031305.2018.1529625. URL <https://cogentoa.tandfonline.com/doi/full/10.1080/00031305.2018.1529625>. [p204]
- C. M. Hurvich and C.-L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989. ISSN 0006-3444. doi: 10.1093/biomet/76.2.297. URL <https://doi.org/10.1093/biomet/76.2.297>. [p206]
- R. Ihaka and R. Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314, 1996. URL <https://doi.org/10.1080/10618600.1996.10474713>. [p203]

- H. Jeffreys. Further significance tests. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 32, pages 416–445. Cambridge University Press, 1936. ISBN 1469-8064. URL <https://www.cambridge.org/core/journals/mathematical-proceedings-of-the-cambridge-philosophical-society/article/abs/further-significance-tests/B2FA761E09518833D0F8FE96B715F932>. [p203]
- D. Lakens. Why p-values should be interpreted as p-values and not as measures of evidence. 2021. URL <https://daniellakens.blogspot.com/2021/11/why-p-values-should-be-interpreted-as-p.html>. [p203]
- G. Palm. *Novelty, Information and Surprise*. Springer Berlin, Heidelberg, 2012. ISBN 978-3-642-43583-6. doi: 10.1007/978-3-642-29075-6. URL <https://link.springer.com/book/10.1007/978-3-642-29075-6>. [p204]
- R. Royall. *The Likelihood paradigm for statistical evidence*. University of Chicago, Chicago, 2004. URL <https://press.uchicago.edu/ucp/books/book/chicago/N/bo3635206.html>. [p209]
- R. M. Royall. *Statistical Evidence: a Likelihood Paradigm*. Monographs on statistics and applied probability. Chapman and Hall, London, 1997. ISBN 0-412-04411-0. URL <https://www.routledge.com/Statistical-Evidence-A-Likelihood-Paradigm/Royall/p/book/9780412044113>. [p203, 209]
- R. M. Royall. On the probability of observing misleading statistical evidence. *Journal of the American Statistical Association*, 95(451):760–768, 2000. ISSN 0162-1459. doi: 10.1080/01621459.2000.10474264. URL <https://www.tandfonline.com/doi/abs/10.1080/01621459.2000.10474264>. [p209]
- The jamovi project. *The jamovi project*. jamovi (Version 2.3.18.0) [Computer Software], Sydney, Australia, 2021. URL <https://www.jamovi.org>. [p210]
- R. L. Wasserstein and N. A. Lazar. The ASA’s statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2):129–133, 2016. ISSN 0003-1305. doi: 10.1080/00031305.2016.1154108. URL <https://doi.org/10.1080/00031305.2016.1154108>. [p203]

Peter Cahusac
College of Medicine, Alfaisal University, Riyadh
Department of Pharmacology & Biostatistics
And Department of Comparative Medicine
King Faisal Specialist Hospital & Research Centre, Riyadh
Kingdom of Saudi Arabia
ORCID 0000-0003-4976-2834
pcahusac@alfaisal.edu