

Likelihood worksheet

Don't worry if you can't complete everything. Much of the material here is adapted from Chapter 11 of my book:

Baguley, T. (2012). *Serious stats: A guide to advanced statistics for the behavioral sciences*. Palgrave Macmillan.

Alternatively there is similar coverage in:

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan.

Working with likelihoods in R (binomial example)

Imagine that someone is given a two-alternative-forced-choice recognition test with 25 trials. The chance of getting any item right by chance alone is .50. A participant gets 19 correct? What is the evidence that their memory is above chance (they are not guessing)?

There are several ways to calculate likelihoods in R. If you have an equation for the likelihood function, you can use this directly¹ - the likelihood for 19 successes and 6 fails from a binomial distribution if the probability of success (P) is .76 is:

$$.76^{19} * .24^6$$

The corresponding likelihood for $P = .5$ is:

$$.5^{19} * .5^6$$

The likelihood ratio is therefore:

$$(.76^{19} * .24^6) / (.5^{19} * .5^6)$$

Q1 Use R to work out the equivalent likelihood ratio for 7 successes on a four-alternative choice test with 12 items (thus the guessing probability is .25). *What is the LR?*

¹ A quick refresher may be helpful. The probability of a x successes out of n trials from a binomial distribution is equal to the binomial coefficient times $P^x * (1 - P)^{n-x}$. The binomial coefficient is a constant for any particular set of data, so the likelihood is proportional to this simple function of P and the number of 'successes' (x) or 'fails' ($n-x$).

Any quantity proportional to this is also likelihood function, meaning that *pmf* and *pdf* functions can be used. Here you could use the `dbinom(x, size, prob)` function:

```
dbinom(x=19, size=25, prob=.76)
dbinom(19, 25, .50)
dbinom(19, 25, .76)/dbinom(19, 25, .50)
```

Q2 These function calls are identical except for the third term (.76 or .50).

What does that term represent and why is the only one to vary?

Although the likelihoods in the preceding equations don't match, the ratios do. The key thing to remember is that a typical call to these functions involves keeping N and P fixed and allowing the number of successes to vary. In a basic likelihood function, the data (here the number of trials and successes) are fixed while the parameter (P in this case) varies.

Q3 Ten Nottinghamshire constituency parties select prospective parliamentary candidates (PPC) with no overlap in applicants. In each case 7 equally qualified candidates (3 men and 4 women) are shortlisted. Eight out of ten PPCs selected are male.

a) Use R to calculate a LR to assess the evidence supporting the hypothesis that selection is gender biased versus the hypothesis that selection is unbiased.

b) What is the probability that selection is fair (assuming each hypothesis is equally likely a priori)?

Likelihood intervals in R

Calculating or plotting likelihood intervals is trickier than likelihood ratios. Baguley (2012) includes functions for calculating likelihood ratios and intervals (including graphs) for various distributions.² For example, the `binom.lik()` function for a binomial proportion is reasonably versatile and called by

```
binom.lik(successes=19, trials=25)
```

it returns the $MLE(\hat{P})$ and the intervals. If you add `plot=TRUE` it also plots the intervals: `binom.lik(19, 25, plot=TRUE)`. Other functions do the same for the Poisson and t distributions. The `pois.lik(count)` and `t.lik.int(mean, SE, df, independent=TRUE)` for an independent design or `t.lik.int(mean, SE, df, independent=FALSE)` for a paired design.

The function `LR.t(mu1, mu2, mu.obs, SE, df, independent=TRUE)` also provides likelihood ratios based on the profile likelihood for normal means when the variance is unknown. In this case the first two arguments are the hypothesized values of the parameter, `mu.obs` is the observed parameter (either a mean or difference in means).

Q4 Use R to obtain the 1/8 likelihood interval for the proportion of male PPCs selected in Q3.

² The functions are available at

<http://www2.ntupsychology.net/seriousstats/SeriousStatsAllfunctions.txt>

You can also use load the functions directly into R using the `source()` function:

```
> source('http://www2.ntupsychology.net/seriousstats/SeriousStatsAllfunctions.txt')
```

a) *What is the interval?*

b) *What does the 1/8 interval suggest about the probability of selecting a female PPC?*

(Optional) If you have time ...

Q5 *Use the `LR.t()` function to decide between two competing hypotheses for some of your own data.*

If you don't have appropriate data try the following input (from a later example):

`mu1 = 0, mu2 = 0.06, mu.obs = 0.142, SE = 0.0703, df = 183`