An Introduction to Likelihood

Applied Likelihood Methods

Common Statistical Approaches:

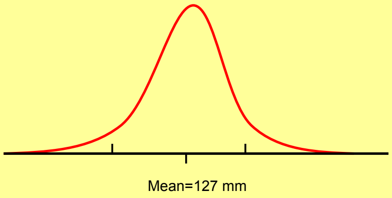
1. Frequentist (t-test, ANOVA)
2. Likelihood Based
3. Bayesian methods

Frequentists

Say we collect 30 age-0 spotted sucker and estimate a mean TL and standard error.

Confidence intervals in the frequentist sense infer that if we repeated the experiment 100 times, that 95% of the time the true population mean would be included in our confidence intervals

Key Point!!



$$CI_{\bar{x}} = \bar{x} \pm t_{2,\alpha,n-1} \times \frac{s}{\sqrt{n}}$$

Frequentists

So, the *t* distribution acts as a slightly more conservative estimate of the parameter bounds than using a standard normal distribution (*Z* distribution), depending on sample size

95% CI infer that if we repeated the experiment 100 times, 95% of those CI would contain the true value of the parameter (e.g., mean)

Frequentists

Same concept applies for frequentist confidence intervals around any parameter estimate:

L_∞
K
G, etc.

Likelihood Methods

First assume that the data follow some probability distribution that makes sense for data of the type you've collected (not a guess, based on previous evaluations)

- -normal
- -lognormal
- -binomial
- -poisson

Given this overall distribution, you can estimate the likelihood of the data given each hypothesized parameter estimate

Likelihood Methods

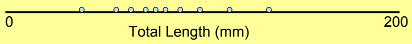
An alternate way to measure uncertainty in itself

However, likelihood methods are the base format that parametric methods use in hypothesis testing

In other words, likelihood methods using probability distributions are the common bond in statistics

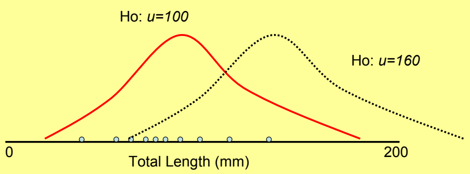
General Idea

Given a set of ten measurements of individual fish lengths,
what is the probability that the data came from a population with
a mean length of 100 mm, assuming a normal probability distribution?

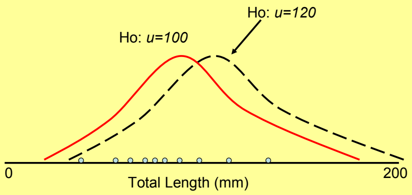


Likelihood Methods

Likelihood methods estimate the odds of the data occurring, given a probability distribution and a hypothesized value of a parameter.



Which hypothesis is more credible given the data?



Which hypothesis is more credible given the data?

Sum of Squares

- Sum of squares (SS) is the most basic way to “fit” a model to data
- SS estimators are part of some likelihoods, and the basic premise of “observed versus predicted” is common to all likelihoods

Likelihood Methods

Each datum has a likelihood of occurring given a set of parameter estimates

We obtain the total likelihood of a set of data, given some hypothesized parameter estimates

The maximum likelihood estimate (MLE) is the parameter(s) that have the highest likelihood

Likelihood Methods

Note that the odds or likelihood of any one data point, or combination of data points is very small, because the likelihood of values x1, x2 and x3 is:

$P(x1, x2, x3)=P(x1)\cdot P(x2)\cdot P(x3)$

Can limit easily reach the limit of computers with regard to decimal places

Likelihood Methods

Because of this, likelihoods are often transformed to log-likelihoods for analysis (simply a transformation to rescale the likelihoods)

Also by convention, we often seek to *minimize* the negative log likelihood

Can maximize the log likelihood as well, no difference

Likelihood Methods

Likelihoods can be estimated for any probability distribution.

Binomial likelihood is:

$$L(p \mid n, y) = \binom{n}{y} p^{y}(1-p)^{n-y}$$

Where:
p = probability of a “yes”
y = number of “yes”
n-y = number of “no”
n = number of trials

Likelihood Methods

$$L(p \mid n, y) = \binom{n}{y} p^{y}(1-p)^{n-y}$$

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}$$

Term gives the number of possible y combinations out of a total of n samples:

$$\binom{5}{2} = \frac{5!}{2!(5-2)!} = 10$$

Let’s make one in a spreadsheet

Likelihood Methods

A likelihood profile represents the relative likelihood values across a range of hypothesized parameter estimates

Can be used to evaluate uncertainty (similar to confidence intervals) in itself

Back to the spreadsheet……..

Likelihood Methods

For probability distributions with only one parameter (binomial, Poisson), a simple likelihood profile will give a good idea of the uncertainty

For other distributions (e.g., normal, negative binomial) we have to use other methods, or conduct a profile by ….

The Normal Distribution Likelihood

So, we need a probability distribution to evaluate the likelihood of hypothesized parameter estimates.

Example: normal probability distribution

$$L\{x_i \mid \mu, \sigma\} = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

Where:
x_i = data
μ = hypothesized mean
σ = hypothesized variance

Likelihood Kernel Approach

Assume that σ is equal to the MLE (e.g., SSE), and thus setting the derivative of σ to zero

Remove constants such as π, to find a simpler normal likelihood kernel:

$$L\{x_i \mid \mu\} = -\frac{n}{2} \times (SSE)$$

Where:
x_i = data
μ = hypothesized mean
-Conditioned on sigma being the MLE

-but again, likelihoods tend to be very small, so…

Log Likelihood Kernel of the Normal Distribution

If you minimize this quantity, it is the least squares fit

$$L\{x_i \mid \mu\} = -\frac{n}{2} \times \ln(SSE)$$

Where:
x_i = data
μ = hypothesized mean

And we iterate μ hypotheses to evaluate the relative Log likelihoods

Let's do one in a spreadsheet.

Likelihood Methods

Let's do a normal likelihood kernel in a spreadsheet

Optimization Routines

The algorithm used to find the MLE and the slope of the likelihood surface at the MLE

A wide range of approaches

Trade off between the efficiency of the algorithm versus the detail of the search (derivative and non-derivative methods)

Optimization Routines

The algorithm used to find the MLE and the slope of the likelihood surface at the MLE

A wide range of approaches

Trade off between the efficiency of the algorithm versus the detail of the search (derivative and non-derivative methods)

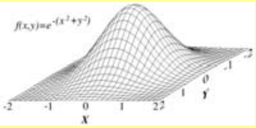
Here's a few common ones

Optimization Routines

Commonly used methods are:

Gauss-Newton – based on least squares, does not require a second derivative

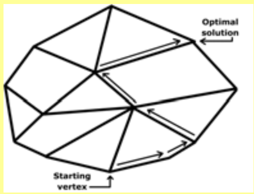
BFGS (Broyden-Fletcher-Goldfarb-Shanno) – seeks a stationary location on a likelihood surface. Starts at a random location and searches incrementally to move along a gradient.



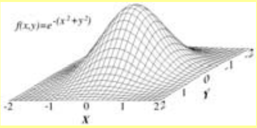
Optimization Routines

Commonly used methods are:

Simplex algorithm – takes a geometric approach to searching the entire surface.



False Minima/Maxima



All optimization routines can fail!

Change starting values, best to run the routine several times from a range of starting values!

Problems more likely with complex optimization tasks