Model Based Statistics in Biology.
Part II. Quantifying Uncertainty.
Chapter 7.5   Confidence Limits

ReCap.   Part I (Chapters 1,2,3,4)
ReCap   Part II (Ch 5, 6)
7.0 Inferential Statistics
7.1 The Logic of Hypothesis Testing
7.2 Hypothesis Testing with an Empirical Distribution
7.3 Hypothesis Testing with Cumulative Distribution Functions
7.4 Parameter Estimates
7.5 Confidence Limits

The truth is out there.
We're going to surround it.

Anonymous student. 2011.

on chalk board

ReCap Part I (Chapters 1,2,3,4)
Quantitative reasoning: Example of scallops, which combined
models (what is the relation of scallop density to substrate?)
with statistics (how certain can we be?)
ReCap (Ch5)
Data equations summarize pattern in data as a series of parameters (means, slopes).
ReCap (Ch 6)
Frequency distributions are a key concept in statistics.
They are used to quantify uncertainty.
Empirical distributions are constructed from data
Theoretical distributions are models of data.
ReCap (Ch 7)
Inferential statistics are a logical procedure for making decisions when there is
uncertainty due to variable outcomes.
Hypothesis testing is concerned with making a decision about an unknown population
parameter.
Estimation is concerned with the specific value of an unknown population parameter.

Today: Cumulative distributions to compute confidence limits on estimates

Wrap-up
We used cumulative distributions to compute confidence limits, a measure of the
reliability of an estimate.

Confidence Limits

We will use confidence limits to evaluate the uncertainty on an estimate made from data.

Definition. A Confidence Limit consists of two values that bracket the true value of a statistic, at some specified level of confidence (say, 95%).

Example -- Brook trout lengths

We measure a sample of 16 0-group trout from Cat Arm lake on the Great Northern Peninsula, in 1982, prior to flooding to create a reservoir.

0-group are fish less than 1 year old. First year fish were of interest in 1982 because one potential impact of flooding of Cat Arm lake was reduction in numbers or size of fish recruiting to the population. If there was an effect on size or numbers, then Newfoundland Hydro was committed to build a hatchery to mitigate the effects of flooding on this and other fish in the lake. The hatchery would be built by Newfoundland Hydro, at the expense of those who pay Hydro for electricity. Size was measured prior to flooding to establish a baseline for comparison to first year fish after flooding.

Quantity is fork length $Y = $ mm

Sample size is 16
Total population in the lake is *ca* 700 trout
Sampling is haphazard.
Sampling fraction is 16/700, or approximately 2%

The sample mean is *mean*$(Y) = 53.8$ mm
This is an estimate of the true mean E(Y), which is unknown.

How reliable is our estimate of the mean? We'd like to know whether our estimate is close to the true value, the average length in the population of approximately 700 fish. We can't know the true mean, but we can make a statement about the reliability of our estimate, relative to the true value of the mean.

To make a statement about the reliability of our estimate, we compute a range that includes the true value a high percentage of the time. This is called the confidence limit. Here is a generic recipe for calculating confidence limit on any estimate.

**Table 7.5a** Generic recipe for calculating a confidence limit.
_____
1. State population; state the statistic of interest.
2. Calculate an estimate of the statistic from data
3. Determine the distribution of the estimate.
4. State tolerance for Type I error.
5. Write a probability statement about the estimate or statistic.
6. Plug values into the statement to obtain confidence limits.
7. Make a statement about the probability that the line
    (or limits) include the true value.
    This statement is not about the statistic or estimate.
_____

Confidence Limits – Computational Flow

To compute the confidence limit, we fix some probability that we can live with, then make a probability statement about a line that includes the true value at a pre-stated level of confidence.
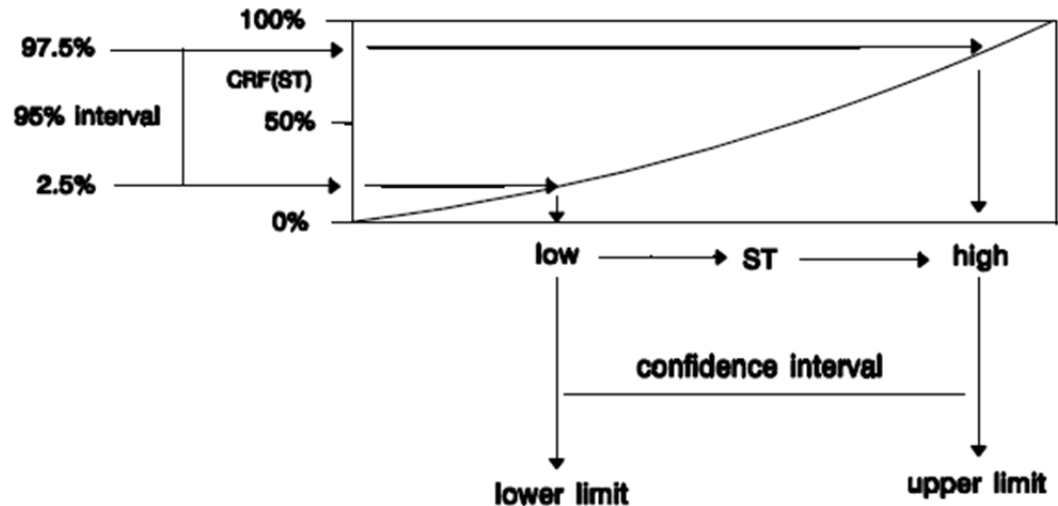
To compute a confidence limit, we will:
  obtain a statistic from data.
  obtain the distribution of this statistic
      (this is not the same as the distribution of the data).
  use the distribution of the statistic to compute confidence limits
      around the estimate.

Graph shows this. Start with probability range to compute range of statistic.

```
As example, draw Chisquare distribution,
(labelled df = 4)
The is the distribution for the variance.
(Draw pdf)
Then draw the cdf above or below the pdf.
Then show Minitab computations for text
example p 155 (Sokal and Rohlf 1995)

MTB > invcdf 0.025;
SUBC> chisquare 4.
    0.025    0.484
MTB > invcdf 0.975;
SUBC> chisquare 4.
    0.975    11.1433
```

Confidence Limits –Key to choice of distribution for computing limit.

Table 7.2. Key for choosing the frequency distribution of a statistic.

Statistic is the population mean
    If data are normal or cluster around a central value
        If sample is large (n > 300) .... Normal distribution
        If sample is small (n < 300 .... t distribution
    If data are Poisson .... Poisson distribution
    If data are Binomial .... Binomial distribution
    If data do not cluster around central value, examine residuals (deviations
        from the mean)
    If residuals are normal or cluster around a central value
        If sample is large (n > 300) .... Normal distribution
        If sample is small (n < 300) .... t distribution
    If residuals are not normal .... Empirical (bootstrap)

Statistic is the population variance
    If data are normal or cluster around a central value .... Chi-square
    If data do not cluster around central value
        If sample is large (n > 30) .... Chi-square
        If sample is small (n < 30 .... Empirical (bootstrap)

Statistic is the ratio of two variances (ANOVA tables)
    If data are normal or cluster around a central value .... F-distribution
    If data do not cluster around a central value, calculate residuals
    If residuals are normal or cluster around a central value .... F-distribution
    If residuals do not cluster around central values
        If sample is large (n > 30) .... F-distribution
        If sample is small (n < 30) .... Empirical

Statistic is none of the above
    Search statistical literature for appropriate distribution
        or confer with statistician
    If not in literature or cannot be found .... Empirical

Empirical distributions are generated by taking all permutations, by sampling permutations, or by subsampling (bootstrap methods).

Confidence Limits from Generic Recipe -- Brook trout fork lengths

1. The population is all brook trout less than 1 year of age in Cat Arm Lake in 1982
The statistic of interest is the population mean.

2. The mean for a sample of 16 fish was
$$mean(Y) = 53.8 \text{ mm}$$

3. The distribution of an estimate such as a mean will not be exactly the same as the distribution of data so we use a key (Table 7.2, above).

   The statistic is the population mean,
   the data cluster around a central value,
   and the sample size is small (16),
   so the appropriate distribution is the t-distribution.

4. The tolerance of Type I error will be set at 10%.

5. Now that we have a distribution, and a stated tolerance for error, we can write a probability statement.

   First in verbal form. "The probability that a line from L_1 to L_2 includes the true mean fork length μ_Y of CatArm brook trout is equal to 90%"

   The probability statement is about $1 - \alpha$

   Next in graphical form (refer back to figure, limits on y axis projected to x-axis)
   The frequency distribution is used to go from probability to outcome.
   This is the opposite direction from that used in hypothesis testing.
   The frequency distribution is used to go from outcome to probability in hypothesis testing.

   Now, the same thing in symbolic form.
   $$P\{L_1 \leq \mu_Y \leq L_2\} = 1 - \alpha$$
   $$P\{\overline{Y} - s_{\overline{Y}} \cdot |t(\alpha/2, n-1)| \leq \mu_Y \leq \overline{Y} + s_{\overline{Y}} \cdot |t(\alpha/2, n-1)|\} = 1 - \alpha$$

   We'll examine each of these components in detail.

   For now, we note $|t(\alpha/2, n-1)|$ is the absolute value of the t from the cdf

   We subtract this quantity from $\overline{\overline{Y}}$ to obtain the lower limit L_1

1 add it to $\overline{\overline{Y}}$ to obtain the upper limit L_2

Confidence limits -- Brook trout fork lengths (continued)

5.   Write probability statement (continued)

$$P\left\{\overline{Y} - s_{\overline{Y}} \cdot \left| t(\alpha/2, n-1) \right| \le \mu_Y \le \overline{Y} + s_{\overline{Y}} \cdot t(\alpha/2, n-1) \right\} = 1 - \alpha$$

We use the sample mean $\overline{Y}$ to estimate of the true mean μ_Y

For the *t*-statistic we use the *t*-distribution, which is symmetrical around zero.
We use the *t*-distribution with df $= n-1$
We assign half of our tolerance for uncertainty ($\alpha/2$) to each tail (5% to each tail)
We use the inverse cumulative distribution function cdf to obtain the lower tail
      probability ($\alpha/2$) and the upper tail probability ($1-\alpha/2$) values for
      the *t*-distribution.
$s_{\overline{Y}}$ is the standard error of the mean

    The standard error is estimated from the sample standard deviation of the data s_Y

$$s_{\overline{Y}} = \frac{s_Y}{\sqrt{n}} = \sqrt{\frac{\text{var}(Y)}{n}}$$

6.   Plug values into probability statement

$$\overline{Y} = 53.8 \text{ mm} \quad s_Y = 5.8 \text{ mm} \quad \alpha = 10\% \quad \alpha/2 = 5\% \quad t(0.05,15) = -1.753$$

```
MTB> invcdf 0.05;
   SUBC> t 15.
         0.05    -1.753

   MTB> invcdf 0.95;
   SUBC> t 15.
         0.95    +1.753
```

```
Minitab reports each tail.
Spreadsheets and tables for t
distribution are usually two-
tailed, showing both tails for
positive values only.
```

P{ 53.8 − 1.753*(5.8 / sqrt(16)) $\le \mu_y \le$ 53.8 + 1.753*(5.8 / sqrt(16)) } = 90%

P{ 53.8 − 2.54 $\le \mu_y \le$ 53.8 + 2.54} = 90%

P{ 51.26 $\le \mu_y \le$ 56.34} = 90%

7.   The limits 51.26 mm to 56.3 mm enclose the true population mean 90% of the time
      It is a statement about limits that enclose the true value of the mean.
      The probability statement is not a statement about the sample mean $\overline{Y}$.

Comments on confidence limits

How do we narrow $L_2 - L_1$?
   increase α This increase tolerance of Type I error
   increase n This reduces the standard error of the estimate
   decrease σ This accomplished by eliminating sources of error or unexplained variance.

Confidence limits use the relation of outcome to probability in a different fashion than with hypothesis testing.

Hypothesis testing: go from outcome to probability

```
MTB> cdf 1.753;
SUBC> t 15.
     1.753      0.95
```

Confidence limits: go from probability to outcome

```
MTB> invcdf 0.95;
SUBC> t 15.
      0.95 1.753
```

Notice that Minitab gives outcomes for both tails, while statistical tables show for just the upper tail (positive) values of *t*, with probability in both tails. This saves space in a printed table, because the *t*-distribution is symmetrical.

Confidence limits are particularly useful in excluding hypotheses other than just H_o

A confidence limit is not of much value if the sample is not representative.

We use the t distribution for the t-statstic, the F-distribution for F-ratios, and χ^2 distribution for statistics known be distributed as chi-square. For other statistics, we use an empirical distribution generated by randomization (*cf* Table 7.2). Or we can use bootstrap methnods (Lab 11).

```
Elementary statistics courses for biologists tend to
lead to the use of a stereotyped set of tests:
1 without critical attention to the underlying model
involved;
2 without due regard to the precise distribution of
sampling errors;
3 with little concern for the scale of measurement;
4 careless of dimensional homogeneity;
5 without considering the ideal transformation;
6 without any attempt at model simplification;
7 with too much emphasis on hypothesis testing and too
little emphasis on parameter estimation.

M.J. Crawley. 1993.  GLIM for Ecologists.  (London,
Blackwell)
```

Binomial Confidence Limits -- partridge berry example

```
Binomial limits asymmetrical.
This example illustrates very low success
(infection) rate.
```

```
Ask students for examples from their work,
where showing reliability (confidence limits)
would be useful.
```

A sample of 250 partridge berries was collected to estimate worm infestation rate.
In this sample one worm was found.
How reliable is this as an estimate of the true rate of infestation ?
(i.e., what are the confidence limits for this estimate?)

1. Population = all partridge berries above the Logy Bay lab in Sept 1989
    Statistic is infection rate

2. Estimate of infection rate is $p = 1/250 = 0.004$

3. Frequency distribution is binomial with $n = 250$ trials, $p = 0.004$

4. Tolerance of Type I error: $\alpha = 10\%$

5. Frequency is one infection, for which Minitab notation is $K = 1$
    $P\{ L_1 \leq K \leq L_2 \} = 1 - \alpha$

6. Plug in $\alpha/2$,
    lower limit

    ```
    MTB> invcdf 0.05;
    SUBC> binomial n=250   p=0.004.
         K   P(X LESS OR = K) K    P(X LESS OR = K)
         0           0.000               0       0.3671
    ```

    Can't set a lower confidence limit. Not enough information.

6. Plug in $1-\alpha/2$,
    upper limit

    ```
    MTB> invcdf 0.95;
    SUBC> Binomial 250 0.004.
         K   P(X LESS OR = K) K    P(X LESS OR = K)
         2           0.9201              3       0.9813
    ```

7. $P\{ K \leq 3 \} = 0.9813$

    The true value of the infection rate lies at or below 2/250, with a Type I error rate
    of $1-0.92 = 8\%$