6. ONE-SAMPLE INFERENCE

Confidence interval for the mean

Unfortunately, the mean calculated from a sample, $\overline{X}$, will differ from the population mean μ. The expected discrepancy between $\overline{X}$ and μ depends on the size of the sample and the variability of X. If sample size is small and X has high variance, then $\overline{X}$ may be quite far from the population mean. In contrast, if sample size is large and X has low variance, $\overline{X}$ will probably be close to μ. The confidence interval (CI) combines information on sample size and variability to put probabilistic bounds on estimates of the population mean. CI's can be calculated for any desired degree of confidence, but 95% confidence intervals are most common. If your sample is random and the population has a normal distribution, you can be "95% confident" that your confidence interval includes the population mean. More accurately, if you sample repeatedly and generate a 95% CI's each time, you can expect the CI to include the population mean in 95% of the cases, and not in the other 5% of cases. Since you usually don't know the population mean, you'll never know when this happens. If the data are not from a normal distribution, then the 95% CI will include the true mean in approximately 95% of cases <u>only</u> if sample size is large (follows from the Central Limit Theorem).

Hypothesis testing for the mean

Samples from a population may also be taken to test hypotheses about the population mean. For example, the sample may be the result of an experiment designed to test for a proposed treatment effect against the null hypothesis of "no effect":

   Ho: $\mu = 0$
   Ha: $\mu \neq 0$

If the random sample is from a normally distributed population, then the (two-tailed) one-sample t-test may be used: reject Ho if $t \geq t_{0.05(2),v}$ or $t \leq -t_{0.05(2),v}$ where

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

and v is degrees of freedom. The same procedure may be used if the data are <u>not</u> from a normal distribution <u>only</u> if n is large (follows from the Central Limit Theorem). If the data are not from a normal distribution and the sample size is <u>not</u> large, the Wilcoxon signed-rank test may be used instead. We will learn more about the Wilcoxon and other "non-parametric" tests in future lab exercises. These tests are based on ranks and do not require the assumption that the population is normally distributed. However, rank tests are generally less powerful than tests based on the normal distribution, and the latter are therefore preferred if the assumption of normality can be met.

Using the Program

JMP IN calculates the 95% confidence interval for the means automatically when you examine the **Distribution** of a variable. The values for the upper and lower limits to the interval are shown in the **Moments table**. The 95% confidence interval for the mean is also displayed in the outlier boxplot as a <u>diamond shape</u>, with the mean being the midpoint of the diamond. To obtain 99% or other confidence intervals click the red "▼" next to the variable name in the results window and

select **Confidence Interval** → **.99**. To carry out a one-sample *t*-test or Wilcoxon test click the red "▼"next to the variable name in the results window and select **Test Mean**. The results will displayed under the **Test Mean=value** heading. If you click the red "▼" symbol at the bottom of this display you will see a visualization of the *P*-value calculation (use the bottons at the bottom of the visualization to toggle between one- and two-tailed *P*-values). This visualization is especially helpful if you are carrying out a one-tailed test.

Problems

1. A *meta-analysis* involves the combining of information from available published studies that have investigated the same question using "similar" methods. The approach analyzes the magnitude of effects seen in separate experiments, and the unit of observation is the individual published study. The data file **metacompetition.jmp** (located on the shared drive S) contains a summary of available studies from a recent meta-analysis of plant competition (Gurevitch and Hedges 1993, *Meta-analysis: combining the results of independent experiment*. Pp 378-398 in S. M. Scheiner and J. Gurevitch, eds, *Design and analysis of ecological experiments*. Chapman & Hall, New York). The variable *Competition effect size* measures the difference in plant performance (e.g., growth rate) between two types of experimental plots: those in which the target species occurs alone, and those in which a competing species is also present. Effect size has been scaled relative to variability within studies, and therefore has no units. A positive number means that adding a potentially competing species reduced growth, whereas a negative number means that adding a competitor enhanced growth.

    a) Estimate the mean *Competition effect size* among studies, assuming that the 43 cases represent a random sample of effect sizes in plant competition experiments. Generate a 95% confidence interval for the mean.

    b) State the interpretation of the 95% confidence interval.

    c) Other than the assumption of a random sample, what assumption did you require when calculating the confidence interval? Is this assumption met in the present data set? Explain.

    d) State the Central Limit Theorem. Can we appeal to this theorem in our analysis in (a) ?

    e) Test whether the mean competition effect size is significantly different from zero. In light of your answer to (c), which test would you recommend?

    f) What was the probability of committing a Type I error in your analysis?

    g) A number of the measurements of competition effect size come from the same study (as indicated by the *Source* variable). Is this a problem for your analysis? Explain. What would you recommend as a solution? (If time permits, try out your solution and see if you obtain the same answer as in (e)).

2. The mean specific activity of the enzyme Na+-K+-ATPase in gills of most freshwater teleost fishes at 15°C is known to be 3.33 micromoles of phosphate / milligram of protein / hour. The specific activity of this enzyme in the gills of marine fishes is expected to be <u>higher</u> than in freshwater fishes due to the greater salinity of their environment. To test this, the specific activity of Na+-K+-ATPase in was measured in gills of a sample of marine-dwelling hagfish (*Eptatretus stouti*) (units are micromoles of phosphate / mg of protein / hour). The data are stored in the file called **hagfish.jmp**.

   a) Is the specific activity of Na+-K+-ATPase in gills of the hagfish <u>greater than</u> in gills of most freshwater teleost fishes? Use the most powerful test available. Show all steps taken in testing the hypotheses.

   b) In addition to the assumption of a random sample, what assumption did your test in (a) require? Test this assumption. Was your assumption valid?

3. Vertebrates and other animals frequently lose mass during periods of low food supply, but their structural body size generally continues growing or remains the same. The data in **iguana.jmp** are measurements of changes in body length of male Galápagos marine iguanas from Santa Fe island, Galápagos, during the El-Niño event of 1997/1998. (Wikelski, M. and C. Thom. 2000. *Nature*. 403:37–38; partial data set kindly provided by M. Wikelski).

   a) Test whether mean body length changed during the El-Niño event. Justify the method you used by also testing its assumptions.

   b) If you rejected the null hypothesis in (a), do the results imply that conditions during the El-Niño caused shrinkage of marine iguanas?

   c) What was the mean % change in length? Comment on the difference between the magnitude of the *P*-value in your test in (a) and the magnitude of the effect. Are the *P*-value and the magnitude of effect expected to be related in general?

   d) Observe the 95% confidence interval for change in body length. Are the confidence limits likely to be accurate? Explain.

   e) Plot the histograms for *Length before* and for *Change in length* side-by-side. In the histogram for *Length before* click on the bars corresponding to individuals of greater than average size in the population. Observe the values of *Change in length* for these individuals (indicated by dark shading in the corresponding histogram). Repeat for individual iguanas of smaller than average size. Do you notice a pattern? Use the remaining two variables in the data table (*Size category* and *Direction of change*) to test whether change in body length during the El-Niño depended upon initial body length.

4. With a computer we can carry out multiple random samples from a population and compute confidence intervals each time. This will aid the interpretation of the confidence interval based on only a single sample (the usual case, in real life). Click on the link below to visit the confidence limits page of the **Rice Virtual Lab in Statistics**. Once you are there, click **Begin** to start the Java Applet and then click the Back button on your browser so that you return to this page: http://www.ruf.rice.edu/~lane/stat_sim/conf_interval/index.html (if this did not work, paste this web address into the location bar of your browser and hit the Enter key on your keyboard). The Java Applet allows you to take multiple random samples from a normal population with mean $\mu = 50$ and standard deviation $\mu=10$, and compute 95% and 99% confidence intervals for each sample. In the Applet window specify the sample size (*n*) and click **Sample** to generate confidence limits for 100 random samples. Observe the number of confidence intervals that contained the true population mean $\mu = 50$ and the number that did not. Click **Sample** to add another 100 samples; click **Clear** to begin again.

a) What percent of the 95% confidence intervals would you expect to contain the population mean $\mu = 50$? What percent of the 99% confidence intervals would you expect to contain the population mean?

b) Which is wider, 95% or 99% confidence intervals? Why?

c) How does sample size affect the number of intervals that contain the population mean? Explain.

d) How does sample size affect the width of the intervals?

e) The widths of the intervals vary somewhat even for a given sample size. Why?