Discovering Patterns in the Growth of *Pinus strobus*: Data Analysis Using Minitab

If you would like to have access to a portion of this data set contact P.A. Heithaus at heithausp@kenyon.edu

Now that you have explored the pine plantation you should have some ideas about factors that may explain the variation in the height of the trees in the front and back sections as well as within sections. This week you will be working with a large data set that has been collected over a 10 year period by former students in the introductory laboratory. For simplicity you will be given data on how the height of the trees has changed since they were planted. It is your job to explore factors that might explain this variation in height. You will also be given data on the trees that produced cones in 1999 so you can explore factors that might be related to the chance an individual tree will produce cones.

History

Before beginning your analysis there are a few historical facts that you should be aware of.

- All trees were planted on Earth Day, 1990 in two blocks.
- In one block trees were spaced 10 feet apart and in the other they were 15 feet apart.
- Half of the trees in each spacing were fertilized from 1992-1996.
- Shrubby vegetation growing around the trees was removed in the summer of 1996 and all subsequent years.

The Data Set

You will be working primarily with the dependent variable, height, which was measured in centimeters. Height data from 1990 to 1999 has been entered into labeled Minitab columns. The data for each year has been entered in separate columns labeled "Hgt" plus the last two digits of the year the data was collected (eg. Hgt99). When trees died or were not measured an "*" was entered.

If you are interested in growth rate (change in height from one year to the next) you will have to use the **Calc** function offered by Minitab to calculate annual growth rate for the years of interest and store it in labeled columns.

The data set also contains a number of **code columns** that contain categorical information about each tree. These independent variables may help explain some of the variability in height that you observed when you visited the plantation. The independent variables and their columns are as follows:

1. Fertilization (C11-T)
    - unfert = unfertilized trees
    - fert = trees fertilized annually from 1992 to 1996

2. Spacing (C12-T)
    - 10 = trees spaced 10 ft apart
    - 15 = trees spaced 15 ft apart

3. Leader (C13-T) - any leader loss from the time of planting to 1999
    - normal = no leader damage from 1990 -1999
    - lost = leader lost at least once from1990 - 1999

4. Timing (C14-T) - timing of leader loss
    - early = prior to 1995 and most likely caused by deer browsing
    - late = 1995 to 1999 most likely due to insect damage weakening the leader
    - none = no leader damage recorded between 1990 and 1999

5. Cover96 (C15-T) proportion of tree circumference covered with shrubby vegetation in
    - none = no shrubby vegetation on tree in 1996
    - moderate = from minor cover up to 2/3 of the tree surrounded by shrubby vegetation in 1996
    - heavy = more than 2/3 of the tree surrounded by shrubby vegetation

6. Browsed99 (C16-T) - buds on lateral branches missing due to deer browsing. This column could be used to see if browsing and spacing are independent.
    - yes = browsed in 1999
    - no = not browsed in 1999

7. Cones98 (C17-T) - tree is producing large female cones
    - present = cones present in 1998
    - absent = no cone production in 1998

8. Hgtcode (C18) - the heights of trees in 1998 were divided into three categories. This column could be used to determine if the frequency of cone production in 1998 and the frequency of tree heights were independent.
    - 1 = 125-300 cm
    - 2 = 301-476 cm
    - 3 = 477-653 cm

Exploring patterns in a data set.

You are now invited to look for patterns in the data set. A good way to begin is to use descriptive statistics to see how mean tree height changed from 1990 - 1999. The technique described below organizes the data so it can be analyzed to create new columns of data that can be transferred to a graphics program.

Organize the height data into a single column with a corresponding code column:

Manip ➤ Stack/Unstack ➤ Stack Columns ➤ Stack Columns (*Block all the height data columns and press select*) ➤ Store stacked data in (*name the column* ***heights***) ➤ Store subscripts in (*name the column* ***age***) ➤ **Ok**

Describe patterns in the data:

Stats ➤ Basic Statistics ➤ Store Descriptive Stats ➤ Variables (*enter the **heights** column you just created*) ➤ By Variable (*enter the **age** column you just created*) ➤ Statistics ➤ ✓ Mean, ✓ Standard Error, ✓ N nonmissing ➤ OK

The above analysis summarizes the variability in tree heights each year by calculating the mean and the standard error of the mean based on a known sample size.

Calculate the 95% CI: Since all your sample sizes are greater than 120 you can multiply the standard error by 1.960 to calculate what you would need to add or subtract from the mean to get the 95% CI.

Calc ➤ Store result in variable (use the column that currently has the SEMean, your calculation will replace the data currently in the column). ➤ Expression (SEMean * 1.960) ➤ OK ➤**Rename the column 95%CI**

The pattern and variability of the data can be expressed by graphing the means ± 95%CI.

Create an Excel graph: You can now copy the columns of data to Excel and proceed to make your figure. For help refer to **Quick Guide to Excel Graphics** (pg 43).

When adding error bars:

- **select custom**
- **block the column containing the 95%CI to move the information into both dialog boxes.** This will take two steps ➤ OK

Testing Hypotheses about sources of variability in the data set: Data Analysis

The graph you created suggests that not all of the trees on the pine plantation are the same height and that variability increases as the trees get older. How can we explain this pattern?

Perhaps some of the independent variables listed earlier might help clarify the pattern. Below are four possible hypotheses to explain differences in tree height. "

H_1 Fertilized trees are taller than unfertilized trees. (Your choice of year)
H_2 Trees covered by shrubby vegetation are shorter than those not covered with shrubby vegetation.
H_3 In 1998 taller trees were more apt to produce cones than shorter trees. (This is distribution data)
H_4 The trees grew more after cover was removed. (You will need to use the ***Calc menu*** to determine how much the trees grew in 1995 with shrubby cover and how much they grew in 1996 after cover was removed. This information should be stored in two new columns. You might name them growth95 and growth96)

Select the appropriate statistical test to evaluate the above hypotheses (see "What Statistic Should I Use?", pg 17 or the Biology 109 online Resource page). Read about the test if you are

not familiar with it. Then follow the instructions for conducting the test in Minitab. Check with your instructor if you need help interpreting the output.

It is a good practice to write the hypothesis being tested in the Minitab session window before running the test. This makes it easier to keep track of the hypothesis associated with each test. Examine the output of the test and record your conclusion in the session window.

In addition to testing the above hypotheses you are expected to generate 3 additional hypotheses and evaluate them using the appropriate statistical test. Record your hypotheses in the session window.

- Two Sample t-Test - look at the effect of one independent variable on height
- Analysis of Variance - explore the effect of a variable with more than 2 levels
- Chi Square - explore the independence of two types of events (distributions)

The test(s) used will depend on the hypothesis you are testing. Please submit a copy of your session window at the end of the laboratory session.

Ask for help if you are having difficulty getting started

Assignment: Write a brief results section that summarizes your findings during this laboratory. Refer to "Preparation of Laboratory Papers" for guidelines on writing results. Begin by organizing your ideas in a logical pattern. Focus on the biology and include the results of the statistical tests parenthetically. Begin with the pattern of height from 1990 to1999. Include the Excel graph that summarizes these data. Then select **at least three** of the hypotheses you tested and develop the rest of the results section.

- Present the results from a hypothesis that was evaluated using analysis of variance. Include a figure or table that summarizes these results. In the text point out the major trend shown in the figure or table.
- Results based on either of the two types of t-tests can be reported in text form.
- Results based on a Chi Square test can be summarized in the text. Inclusion of a figure or table is optional.

This is not intended to be a lengthy assignment, but rather an opportunity to improve your ability to clearly present information. I would anticipate one page, excluding the figures will be adequate to cover your findings.

Field Lab Next Week - Rain or Shine - come prepared - long pants and sturdy shoes are required. Do Tutorial 3 and email the questions before coming to lab.