INTEGRATED TRAINING AREA MANAGEMENT

ITAM Learning Module

LCTA Scenario

Species Diversity

Recommended Reading

ITAM Technical Reference Manual:
  Chapter 8: *Structured Query Language (SQL)*
  Chapter 11: *Data Analysis and Interpretation*
ITAM Learning Module Notes:
  Creating SQL Statement in Quest
  Using the t-Test: Paired Two Sample for Means in Microsoft Excel
  Generating Graphs in Microsoft Excel
  Creating a Pivot Table in Microsoft Excel
  Calculating Confidence Intervals

Marurran, A. E. 1988. *Ecological diversity and its measurement*. Princeton University Press, Princeton. 179 pp.

  This book discusses the differences and similarities among diversity indices and provides examples of how to calculate the indices.

Elzinga, C.L., D.W. Salzer, and J.W. Willoughby. 1998. *Measuring and Monitoring Plant Populations*. BLM Technical Reference 1730-1, USDI Bureau of Land Management, National Applied Resource Sciences Center, Denver, CO.

Questions

Is there a significant change in species diversity at Fort USA from 1997 to 1999 on the open woodland community type (where training is common)?

Is there a significant change in species diversity at Fort USA by training area for the same time period and community type?

There are a number of indices that express species diversity; however, there is little consensus among ecologist to which is the best measure. Species diversity consists of two components: 1) species richness, or the number of species present, and 2) species evenness, or the species abundance. The following analysis examines Species Richness, perhaps the easiest index to calculate and interpret.

Data set

Data from the AerCover table in the LCTA database are used. This group of data includes information on all life forms present in the sample areas. Similar analysis procedures could be followed using data from quadrat sampling or other forms of sampling where species and abundance data are collected. Diversity of woody plants could be examined using the belt transect data.

Data from years with long-term monitoring were used (1997 and 1999) because short-term monitoring does not record data by species. While some plots during the 1998 survey used long-term methods, the majority of plots were monitored using short-term methods.

The analyses presented in this scenario are not limited to data from the AerCover table. Below is a table describing the data extracted from the database. If you would like to use other sources of data to perform these analyses you will need to gather the same data outlined below. To see the SQL statements that were used for this example see the Accessing Data section at the end of this document.

Variable Name	**Description**
Plot	Plot number
Yr	Year
VegID	Species Code
VegType	Plant Community Types
Train	Training Areas

Notice that species abundance is not needed because it is not used in richness calculations. The table below shows a portion of the data set once it has been imported into Microsoft Excel.

PlotID	Yr	VegID	VegType	Train
1	1997	AMBR2	DENSE WOODLAND	TA2
1	1997	AMELA	DENSE WOODLAND	TA2
1	1997	ANQU	DENSE WOODLAND	TA2
1	1997	APAN2	DENSE WOODLAND	TA2
1	1997	AQCA	DENSE WOODLAND	TA2
1	1997	ASMA2	DENSE WOODLAND	TA2
1	1997	CAPE6	DENSE WOODLAND	TA2

PlotID	Yr	VegID	VegType	Train
1	1997	COAM3	DENSE WOODLAND	TA2
1	1997	COFOR	DENSE WOODLAND	TA2
1	1997	CRATA	DENSE WOODLAND	TA2
1	1997	DASP2	DENSE WOODLAND	TA2
1	1997	FRVE	DENSE WOODLAND	TA2
1	1997	GABOS	DENSE WOODLAND	TA2
…	…	…	…	…

Question Number 1

Is there a significant change in species diversity at Fort USA from 1997 to 1999 on the open woodland community type (where training is common)?

Determine the number of species present on each plot and community type for each year. The Microsoft Excel pivot table tool is used here. Create a pivot table with VegType and PlotID as the designation for rows, Yr for columns and VegID for data. The count summary was used for the VegID data. Also, 1998 data was excluded from the species (VegID) count and all vegetation types except open woodland were excluded from VegType. The resulting table contains the vegetation type, plot and number of unique species per plot in 1997 and 1999 (see below). Refer to the ITAM Learning Module Note, *Creating a Pivot Table in Microsoft Excel*, for help on creating Pivot Tables.

Count of VEGID		Yr	
VegType	PlotID	1997	1999
OPEN WOODLAND	31	21	21
	32	40	21
	33	25	19
	34	18	11
	35	25	24
	36	32	30
	37	20	15
	…	…	…

Calculate Statistics

We will use the statistical capabilities of Excel to answer this question. Because the data was collected on permanent plots, the paired two sample t-Test in Microsoft Excel was used.

Use the columns from the pivot table that represent the number of unique species per plot in 1997 and 1999. Include the column headings but exclude the sum of columns at the bottom of the table. If any of the cells used as the input for the t-Test are blank, you must enter zeros before performing the t-Test. All cells contain values for this example.

t-Test: Paired Two Sample for Means
Alpha = 0.05

	1997	*1999*
Mean	23	20.53
Variance	43.241	42.189
Observations	30	30
Pearson Correlation	0.614	
Hypothesized Mean Difference	0	
Df	29	
t Stat	2.354761	
P(T<=t) one-tail	0.012787	
t Critical one-tail	1.699127	
P(T<=t) two-tail	0.025573	
t Critical two-tail	2.045231	

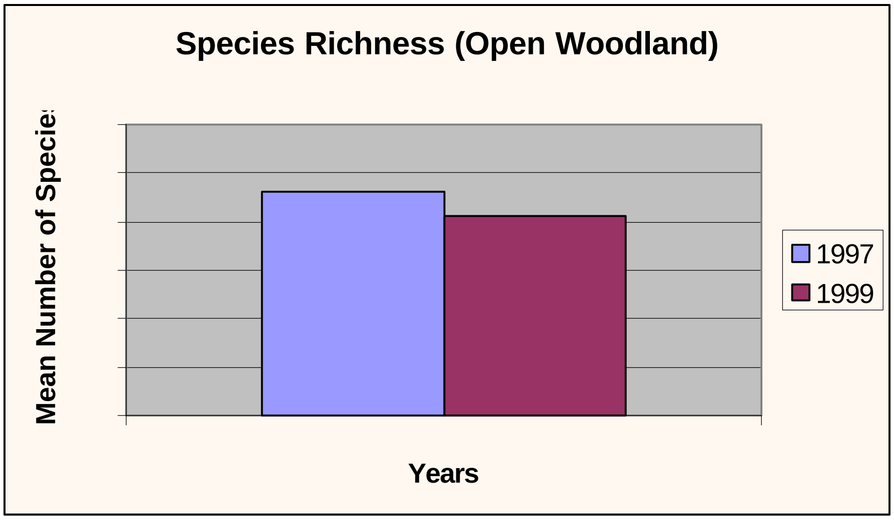
The results of the t-Test are shown above. Note the Groups 1997 and 1999 (years), Observations (sample size), Mean (the average number of species per plot), Variance (how variable each sample is), and P-value (if greater than 0.05 than no significant difference, or $P>0.05$; if smaller than 0.05, then a significant difference, or $P<0.05$).

Given the P-value is 0.025573 (two tailed) is there a statistical difference between the years? Yes! A typical statement would be:

Based on point intercept data, there is a significant difference in species richness between 1997 and 1999 (P<0.05) for open woodland plots.

Displaying the data

Create a histogram chart displaying the species richness for 1997 and 1999. Refer to the ITAM Learning Module Note, *Generating Graphs in Microsoft Excel*, for information on generating graphs.



Notice that we have not included confidence intervals. Recall that the data are paired (the same plots were measured in both years). Graphical presentation of paired data is not as straight forward as with independent samples. The proper procedure is to construct a point graph of the mean difference in mean number of species from 1997 to 1999. Then a 95% confidence interval around this mean difference is plotted. Refer to the ITAM Learning Module Note*: Calculating Confidence Intervals* for information confidence intervals.

First the difference in the number of species between 1997 and 1999 is found for each plot in the open woodlands (shown below).

Count of VEGID		Yr		
VegType	PlotID	1997	1999	Difference
OPEN WOODLAND	31	21	21	0
	32	40	21	-19
	33	25	19	-6
	34	18	11	-7
	…	…	…	…

Next the mean difference is calculated. This is done using the =AVERAGE() function in Excel for the difference values (i.e. the average mean difference). Given that the range of cells for the differences is E3 to E32, the mean difference equation is entered as:

=AVERAGE(E3:E32)

We also calculate the variance of the differences by using the =VAR() function.

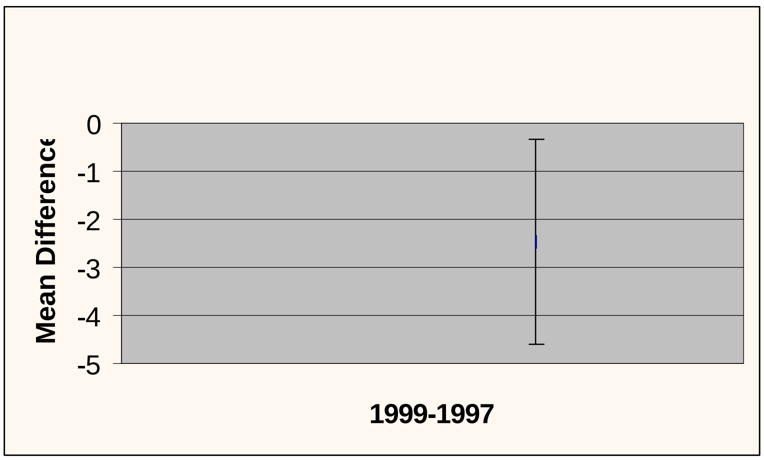
=VAR(E3:E32)

Finally, the confidence interval is calculated using the variance of the differences. The following Excel equation gives the confidence interval:

=TINV(0.05, 29) * SQRT(E34/30)

where:

> TINV(0.05, 2) gives us the two-tailed critical t value for a confidence level of 95% and 29 degrees of freedom
>
> SQRT(E34/30) is the standard error (E34 is the cell reference of the calculated variance of differences and 30 is the number of observations (n))



Notice that the mean difference is negative. Because the interval does not include zero (which would indicate the possibility of no change), this difference is significantly different from zero at the 95% confidence level (i.e. P < 0.05).

Question Number 2

Is there a significant change in species diversity at Fort USA by training area for the same time period and community type?

Again, create a pivot table from the original data set. The only difference here from the previous pivot table is the addition of training area. Use the count summary for VegID and exclude 1998 data. In addition, we will only look at training area TA8 and TA9. Also exclude all vegetation types except Open Woodland. The new table looks similar to the last one with the exception of the division of the data into groupings by training areas. A portion of the resulting pivot table is shown below.

Count of VegID			Yr	
VegType	PlotID	Train	1997	1999
OPEN WOODLAND	31	TA8	21	21
	32	TA8	40	21
	33	TA8	25	19
	34	TA8	18	11
	35	TA8	25	24
	36	TA8	32	30
	37	TA8	20	15
	38	TA8	24	23
	39	TA8	23	24
	40	TA8	20	17
	41	TA9	20	31
	…	…	…	…

Run the paired t-test for each training area (TA8 and TA9) in the open woodlands. The results of the t-Test are shown below.

t-Test: Paired Two Sample for Means
Training Area TA8
Alpha=0.05

	1997	*1999*
Mean	24.8	20.5
Variance	43.73333	28.5
Observations	10	10
Pearson Correlation	0.547618	
Hypothesized Mean Difference	0	
Df	9	
t Stat	2.347007	
P(T<=t) one-tail	0.021759	
t Critical one-tail	1.833114	
P(T<=t) two-tail	0.043518	
t Critical two-tail	2.262159	

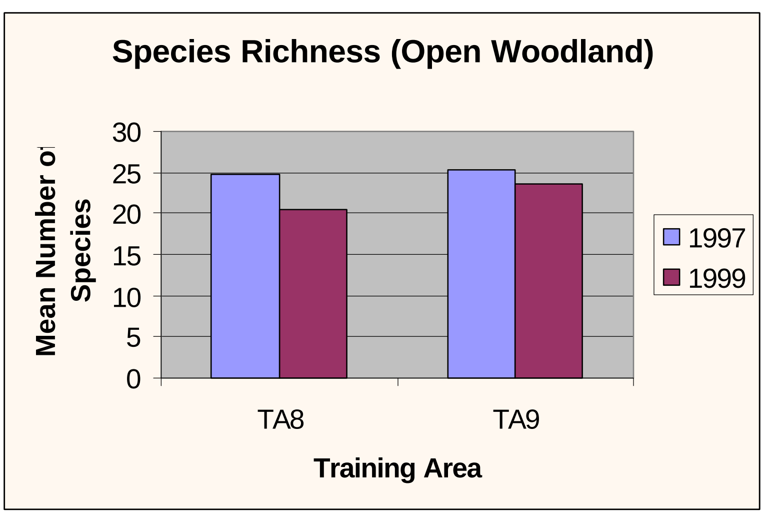
t-Test: Paired Two Sample for Means
Training Area TA9
Alpha=0.05

	1997	*1999*
Mean	25.3	23.5
Variance	56.01111	51.61111
Observations	10	10
Pearson Correlation	0.618936	
Hypothesized Mean Difference	0	
Df	9	
t Stat	0.888235	
P(T<=t) one-tail	0.198766	
t Critical one-tail	1.833114	
P(T<=t) two-tail	0.397532	
t Critical two-tail	2.262159	

Training Area TA8 shows a significant difference between years for open woodland plots (P < 0.05) while training area TA9 does not (P > 0.05).

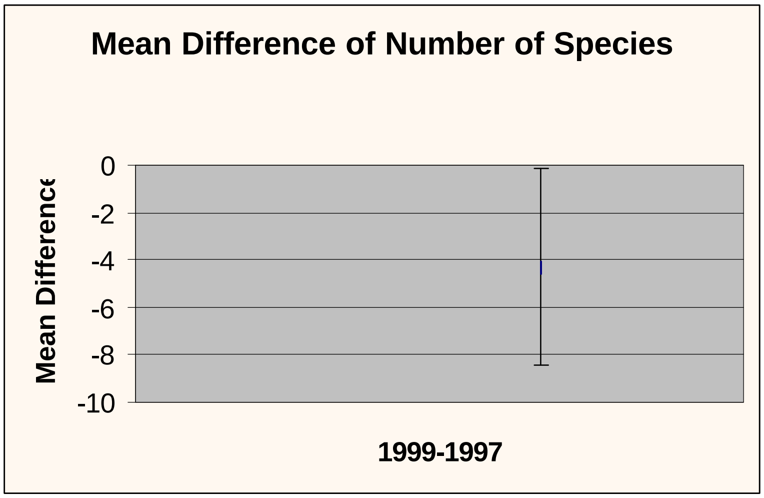
Displaying the data

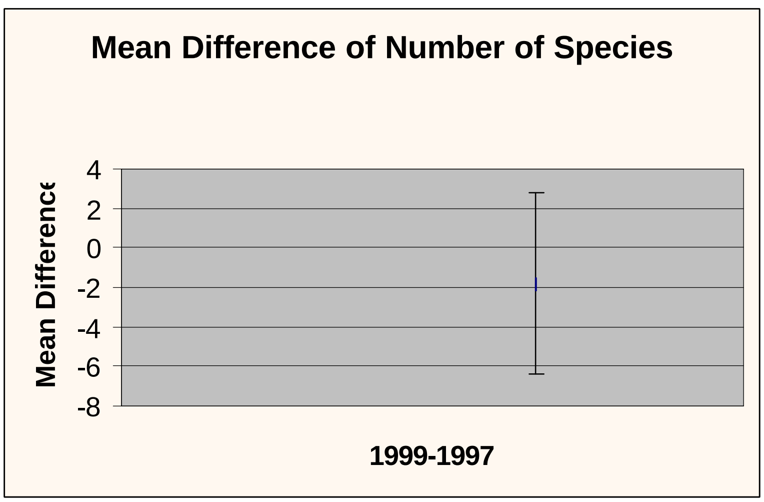
Create a histogram chart displaying the species richness for 1997 and 1999 (shown below).



Again, notice that we have not included any confidence intervals. Shown below is the point graph of the mean difference in mean number of species from 1997 to 1999 for training areas TA8 and

TA9. Also plotted is the 95% confidence interval around these mean differences. Because the interval for TA8 does not include 0 (which would indicate the possibility of no change), this difference is significant at the 95% confidence level (i.e. $P < 0.05$). The interval for TA9 does include 0 suggesting the possibility of no change, thus the difference is not significant at the 95% confidence level (i.e. $P > 0.05$).





Keep in mind that that the results of this comparison may not be the consequence of training activities. As noted in the beginning, environmental factors also affect species richness. To correctly evaluate differences requires determining the species richness in areas without training impacts (control samples) over the same time period and to compare that information to changes in the training areas. Keep this in mind when evaluating the results, because there are no "control" data in the Fort USA data set.

In this example we did not differentiate between native and introduced species. Species richness can be high, but of poor quality. This might occur if there is a large number of introduced undesirable species.

Accessing the data

These SQL statements were executed with the goal of obtaining required data for this scenario. If you are not using the LCTA database you may wish to skip this section.

The following series of statements return information from the LCTA database and are specific to the SQLBase and Microsoft Access databases. Statements for both SQLBase and Access are given. Unfortunately, no single statement will satisfy all of the data needs.

The SQLBase statements create tables. These tables are not automatically updated when data from the AerCover and PlntList tables is changed. You will need to delete the working tables and execute all of the SQLBase statements again when data is changed. The Access statements create queries, which will reflect changes made to the base data tables.

SQLBase Syntax

```
create table Spp
(PlotID integer,
Yr integer,
Spp Char(8));

insert into Spp (plotid,yr,Spp)
select distinct plotid, @yearno(recdate), @upper(vegid)
from aercover
where @upper(aercover.vegid) in (select @upper(plntlist.vegid) from plntlist);
```

> The statements above create a table (Spp) and loads the table with unique plot number (PlotID), year (Yr), and species code (Spp) from the table AerCover. Because only plants are considered in the estimation of species richness, other entries such as litter, duff, deadwood, etc. are excluded. Stipulating that VegIDs must be comparable between the tables AerCover and PlntList ensures this.
>
> SQLBase, the database engine used by Quest, is case sensitive for all text data. To ensure VegIDs match between AerCover and Plntlist, the @Upper function is used.

```
select Spp.plotid as PlotID, Spp.yr as Yr, Spp.spp as VegID, vegtype, train
from Spp, plotsurv
where Spp.plotid = plotsurv.plotid
and Spp.yr= @yearno(plotsurv.recdate)
order by 1,2,3;
```

PlotID and Yr join data from the tables Spp and PlotSurv.

The data are then exported to a dBase file. Select Edit from the main Quest menu, then Copy To from the submenu. Enter a file name, select the file location, and file format of dBase.

From a spread sheet program, such as Microsoft Excel, open the file by selecting the File and Open menu options. Pick the subdirectory the file was saved to and change the file type of listed files to dBase.

Microsoft Access Syntax

```
select distinct plotid, year(recdate) as yr, aercover.vegid
from aercover, plntlist
where aercover.vegid = plntlist.vegid;
```

This statement creates a query consisting of unique values of plot number (PlotID), year (Yr), and species code (VegID). Because only plants are considered in the estimation of species richness, other entries such as litter, duff, deadwood, etc. are excluded. Stipulating that VegIDs must be comparable between the tables AerCover and PlntList ensures this.

Create the query above in Access and save it with the name Spp.

```
select Spp.plotid as PlotID, Spp.yr as Yr, Spp.vegid as VegID, vegtype, train
from Spp, plotsurv
where Spp.plotid = plotsurv.plotid
and Spp.yr= year(plotsurv.recdate);
```

PlotID and Yr join data from the queries/tables Spp and PlotSurv.

The data are then exported to a dBase file. From a spread sheet program, such as Microsoft Excel, open the file by selecting the File and Open menu options. Pick the subdirectory the file was saved to and change the file type of listed files to dBase.

References

Elzinga, C.L., D.W. Salzer, and J.W. Willoughby. 1998. *Measuring and Monitoring Plant Populations*. BLM Technical Reference 1730-1, USDI Bureau of Land Management, National Applied Resource Sciences Center, Denver, CO.