

## **Lab #1 – Testing hypotheses and sampling procedures: Campus vehicular diversity and contributions to air pollution**

Written by Profs. Joan Morrison and Jonathan Gourley

### Important concepts for this lab:

Hypothesis testing  
Sampling  
Making assumptions  
Biodiversity and taxonomic classification  
Environmental impacts – vehicle emission  
Microsoft Excel spreadsheets

### Materials:

Your field notebook  
Lab handout  
Map of campus showing parking lots  
Laptop computer  
memory stick

---

Scientists ask questions about what they observe, in order to try and understand the natural world and systems within it. They form hypotheses that propose specific explanations for those observations. These hypotheses are then tested by collecting and analyzing data. New knowledge is gained by disproving hypotheses. For example, if your hypothesis is that A and B do not differ, but your scientific investigation determines that A and B do in fact differ, you have disproved your hypothesis and you have gained new knowledge.

In this lab we will address the question: What kinds of vehicles do people at Trinity drive, and how much do these vehicles contribute to air pollution on campus? We will examine the diversity of vehicles on campus, and we will discuss important considerations about sampling and analysis including sample size, randomness, and making assumptions about the vehicles you sample. You will also learn to classify your observations and how to organize your data to facilitate analyses and get more specific answers to your questions.

## **PROCEDURE**

In this study you will collect data by sampling the different kinds of cars on campus. You will use these data to:

- 1) develop a “taxonomic” description of the vehicle diversity on campus
- 2) collect physical data about vehicles on campus
- 3) test hypotheses about differences among vehicles and parking lots using some basic statistics
- 4) determine emissions of these vehicles using information obtained from vehicle efficiency websites and test hypotheses about these emissions - these emissions contribute to air pollution and global warming

**Work in pairs.** We will pick partners out of a beaker. I will use this method throughout the semester so that you will all get the chance to meet each other. Generally, work with your partner will end after lab. Everyone will be expected to write-up their own lab reports. However, it will be important to record who you worked with in both your notebook and report. Make sure you each take copious notes in your notebook!

Each pair forms two hypotheses to test regarding the vehicles in 2 different parking lots on campus. You may consider using the hypotheses that you formulated for the pre-lab exercise. One hypothesis should focus on comparing emissions of vehicles between your two lots and another should be a more

general hypothesis about physical differences between cars in your two lots. Each group visits two lots – they can both be faculty, student, or one of each. Be sure you keep track of which vehicles were sampled in which lot. Develop your hypotheses accordingly.

**Classification and taxonomy.** Living organisms are classified as shown below on the left and the similar “classification” of vehicles is shown on the right.

|               |   |
|---------------|---|
| ○ Kingdom     | Vehicle   |
| ○ Phylum      |   |
| ○ Class       |   |
| ○ Order       |   |
| ○ Family      | Type (small, mid, or large-sized passenger car, station wagon, pick-up truck, SUV, minivan, full-sized van) |
| ○ Genus       | Make (Toyota, Mercedes, BMW, Chevrolet, etc)  |
| ○ Species     | Model (Accord, Pathfinder, Frontier, Forrester, etc)  |
| ○ Sub-species | approximate Year (important when looking up emission data)  |

**Before heading out, set up your field notebook.** It is always easier to work in the field if you are prepared for what data you are going to collect. Start by making a heading on a blank page. **This heading should appear for each week’s lab.** Don’t forget to use pencil. The heading should include:

A descriptive title

The date

Your name

Your partner’s name

A description of the lab activity

weather

any other information that seems important

NOTE: All fieldwork must be recorded in your yellow Rite-in-Rain.

**Set up your data table.** Build a table in your field notebook in which you will enter your data. Use the vehicle “taxonomy” above as a guide. State two null hypotheses that you are interested in testing. One should relate to physical characteristics and one should relate to emissions.

**Sampling scheme.** The last thing you should do before heading out is to consider how you are going to sample vehicles in each lot. Often in nature it is impossible to sample every organism, so methods are developed to sample a representative population. How will you sample your parking lots? The first 15 cars you see? Every other vehicle? Every 5<sup>th</sup> vehicle? Note that you will need to collect information on 15 vehicles for each lot. While there is no right or wrong way to do this, it will be important to note your sampling method in both your notebook and final report.

**Population definitions.** Be sure to clearly label which lots you visited to collect your data.

Faculty lots: LSC, Austin Arts, behind Seabury, Admissions-longwalk.

Student lots: North Summit, Summit St., east of Ferris, High Rise, Allen St, Townhouse lots.

**Write out two null hypotheses** and check with your lab instructor that they are well crafted before heading out. One hypothesis should be about a physical difference between the two parking lots (e.g. the number of mini-vans). The second should focus on one of the three pieces of information that you will be looking up for each car on the fueleconomy.gov website: 1) fuel efficiency (mpg), 2) carbon footprint (tons/year) and 3) oil consumption (barrels of oil/year).

### **Field observations and data analysis**

1. **Sample 15 cars in each lot.** Each student should record the following information about all the cars that you sample: type, make, model and ***an approximate age*** (if unknown, then make an educated estimate). Record these data in a well-developed table in your lab notebook. You will find that the type may be the most difficult to determine in the field (e.g. small vs. mid-size). You can look up this information on the websites as long as you have the make and model. Please complete your sampling within 45 minutes. **If only one student records data in the field make sure your partner records the data in his or her own notebook before starting the data analysis.**
2. **Return to McCook 115 for Data and Analysis.** Each student should check out their own laptop computer and login under username: *trinlab* password: *trinity*. Create a new file in MS Excel and put your classification information into this file. **Pay special attention to page layouts, headings and labeling because this file will be part of your lab report.** Save your work to your USB drive or save it to the desktop and email it to yourself. Any work left on the desktop will be lost when the machine is shut down. Any function can be looked up under Help (e.g. Average, standard deviation, t-test). All functions begin with an equal (=) sign.

e.g. =AVERAGE (A1:A15). This functions takes the average of numbers in the A column between rows 1 and 15). For more examples use the help in Excel.

3. **Obtain vehicle emissions data from the Web.** Add additional columns to your Excel file for information on emissions and gas efficiency. Visit the following website to obtain information about emissions of your vehicles: <http://www.fueleconomy.gov>. **Record all of the following information** in your spreadsheet for each vehicle:
  - average miles per gallon (highway)
  - total carbon footprint in tons/ year
  - annual energy impact score in barrels of oil/year

Note: If the year of car is older than listed on websites, do a quick Google search to see if you can find the data. If you still can't find it use the oldest data available. Some cars may not even be listed because they are too old.

4. **Calculate Simpson's diversity index.** One way scientists describe biodiversity is to use a diversity index. This index provides information about the number of different "species" and

about the relative abundance of each species in the community or at your particular site. Simpson's diversity is one of the simplest indices to calculate because it uses proportions of species. The index value "D" is not an absolute value but is a relative index, meaning it can be used in a comparison of several communities. A larger value of "D" means a more diverse community compared to other communities. Calculate Simpson's diversity index:

$$D = \frac{1}{\sum(p_i^2)}$$

where  $p_i$  is the proportion of each "species" within your sample for each parking

lot. This index describes the diversity of your vehicle "communities". The sigma notation means to sum all the values of  $p_i^2$ .  $D$  is the inverse of this sum. Do all calculations in Excel.

What do the values of D for each population of vehicles tell you about the diversity of vehicles in the two lots? (note: answer this question directly on your excel worksheet).

The easiest way to calculate the diversity index in Excel is to create a column that records the count of each unique car "species" If you only have one Honda Accord then that would count as one. If you have two Toyota Highlanders then you would enter 2 for one of the Highlander rows but then leave the second Highlander row blank. The counts column should always add up to 15.

A second column can then be created that calculates the proportion of each count. Assuming you have counted 15 cars in each lot then each of the count values should be divided by 15 to determine it's proportion. **DO NOT DO THIS BY HAND.** Create a formula in the first row (remember all formulas start with an = sign) that references the first count value. Assuming the first count is in cell F3 the formula would be:

=F3/15

Then use Fill Down to repeat the formula for all subsequent rows. Don't worry about the rows that you left blank. They will return as zeros and will not factor in at all.

Next, create a column that calculates the squares of previous column.

Finally sum the squares (best to do this in the cell beneath the last entry) and take the reciprocal. This can be done in one step by: = 1/(SUM(G1:G15)) **Note** your actual cell definition may be different I used G1:G15 to show how it would look if your data were in rows G1 through G15.

5. Repeat for the second lot and compare the two D values. The larger of the two is the more diverse population.
6. **Test your hypothesis about vehicle emissions between parking lots using basic statistics.** Calculate using Excel. Set up your spreadsheet so that it can be included in your lab report. Make sure you label your tables and results clearly. **Answer these questions directly in your Excel spreadsheet using Excel functions.** Calculate the following:
  - a. Total sum weight in tons/year of the carbon footprint of your sample set of 15 vehicles. You should have two numbers to compare. One sum from each lot.
  - b. Total annual energy impact in each of your 2 parking lots in barrels of oil/year.
  - c. The mean and standard deviation of highway mpg for all vehicles in each of your 2 parking lots.
  - d. What were the vehicle makes and models in each parking lot with the lowest mpg? The highest? What were these values?

- e. Consider your hypothesis about the emissions from cars in your 2 parking lots. Perform a *t-test* to test this hypothesis. You can do this test easily in Excel. The t-test is found within the Analysis ToolPak and directions on how to add this can be found easily by typing in the word “analysis” in the Excel Help window. When the ToolPak is loaded it will be found under the Data tab on your Excel toolbar. Run the t-test that assumes unequal variance. The resulting table should be included in your final presentation of data. The important value to look at in the t-test results is the *P value for the two tail test*. If this P value is below .05 then the null hypothesis should be rejected. A rejected null hypothesis indicates that there is a significant difference between your two populations.
- f. Consider your hypothesis about the physical characteristics you selected. How do the two lots differ (if at all)? You will not be able to compare using a t-test but make some statement about your results.

**What is due for this lab?** (Due by your lab session next week)

Each student is to turn in his/her own lab report and spreadsheets. While you and your partner will have similar data your write-ups and spreadsheets should be individual. **These must be typed and printed out.** Electronic copies will not be accepted. The lab report must include these four parts:

- 1) Heading – Make sure to include your name and your partner’s name in the heading.
- 2) Introduction - The introduction establishes the framework of your study. In the Introduction, you clearly state the purpose and objectives of your study. You also present some background information on what is already known, other issues that relate to your study, and anything else that puts your study into context. You should also be sure to state the hypotheses that you tested in your study. The introduction should convince anyone reading your report that you understand what you did in the study and why your study was important. You should also state any assumptions you made and how you obtained your data (sampling decisions). Finally you should briefly summarize your conclusions. This should cover approximately two pages (double spaced).
- 3) Data and Analysis -Your data and statistical analysis (Diversity Index, sums, means, t-test) should follow the introduction and must be constructed in Excel. Answers to the questions in part 5 can be answered directly in the spreadsheet but should be very clear to the reader. Each student should construct their own excel spreadsheet. **Pay close attention to format and labeling of the information** so that it is clear and concise. Your ability to present data clearly and effectively will be graded. This includes making sure your tables are easy to read and final answers are highlighted clearly. Centering data under column headings is always a good idea.
 

You can view what it would look like as printed pages by changing the View. Go to View -> Page Break Preview. In this view you can manipulate page breaks. Make sure you label your tables clearly to guide the reader to the important information that helps support your conclusions.
- 4) References - Cite in proper format of the website(s) you used for this lab. Proper website citation format can be found in the document “How to cite websites properly” in the Course Documents section of the ENVS 149 lab website.