MODULE 3: ANALYSIS OF CHANNEL MORPHOLOGY AND PHYSICOCHEMICAL FEATURES OF STREAMS[1]

Introduction

In this module we will begin the process of testing the hypothesis that the physical and chemical characteristics of the channel units that we visited in the last two laboratories are similar. We can test many specific predictions about the channels units (e.g., flows are similar among units); therefore we can do many statistical tests. We will begin this process in laboratory, but students will have to continue the analyses outside of class. It is probably best that students work together on the analysis in teams of two or three. We will use the data in two ways. First, as an assignment for this module, each student will write a short summary (2 - 3 single spaced pages exclusive of graphics) comparing and contrasting the physical and chemical characteristics of the channel units. This assignment is explained at the end of this module. Second, these data will then be incorporated into a larger report which each student will write individually. The overall purpose of the report will be to compare channel morphologies and physicochemical features among areas and then to compare differences in these features to differences in invertebrate assemblages which will be determined in future laboratory exercises. Moreover, you will use some of the statistical techniques presented in this module to analyze your invertebrate data. Your report will also be the place where you will interpret data. You can use information given in lecture to help with interpretation. We will do little on interpretation today. The report will be described in a future laboratory.

Exercise 1: Average characteristics of the channel

Average of characteristics other than sediment size

In this exercise we will calculate average (mean $=\bar{Y}$) characteristics of each of the channel units (Rocky Creek riffle and pool and Bear Creek). We will also calculate measures of variability (standard deviation = s and variance = s^2) of each characteristic and precision of our mean estimates (standard error of the mean = se). These calculations will be used to carry out statistical tests, called a *t*-tests, to determine whether some these features differ among the stream reaches and channel units. [A note to remember is that some of the characteristics that we measured (e.g., width of the channel) were not replicated (taking multiple measurements or samples) or were descriptive (e.g., maps of the reaches) so we will not use statistics on them; however, these characteristics could still be included in your report.]

---

[1] Last modified 8/22/2005

Formulas for Y, s and se are given below. The variance (s^2) is just the square of the standard deviation (s). You may have a calculator that calculates these or you may use a spreadsheet such as Excel or Quattro. Be sure you are using the correct calculation for s.

Mean = $\overline{Y} = (\sum y_i) / n$

Standard deviation = $s = \sqrt{\{[\sum (y_i - \overline{Y})^2] / (n - 1)\}}$

Variance = $s^2 = \{[\Sigma(y_i - \overline{Y})^2]/(n - 1)\}$

Standard error of the mean = se = $\sqrt{(s^2 / n)}$

Where y_i are the observations (e.g., each measurement of depth) for a given reach, n is the number of measurements (also called the sample size), Σ is a symbol that means sum (add up all the observations), and the others are as defined above. Calculate these values for each of the physicochemical characteristics in Data Table 1. **When calculating average stream velocity use the velocity data from the horizontal velocity profiles (not the vertical velocity profiles).**

Sediment size

To determine the distribution of sediment sizes in each reach, prepare a graph of the frequency of each sediment type in each reach. On the x-axis you should have the sediment types (boulder, cobble, pebble, gravel and sand & silt). On the y-axis you should plot the number of each sediment particles in each category. These are usually plotted as bars. You should have three graphs one for each reach. Calculate the median particle size for each reach. The median is just the particle size where ½ of the observations are above it and ½ of the observations are below it (particle size class for particles number 14 and 15 if you took 30 particles). Record the median substrate sizes in Data Table 2.

Exercise 2: Cross section of stream and calculation of discharge

For each of the channel units (Rocky Creek - pool, Rocky Creek - riffle, Bear Creek) draw the cross sectional profile of the stream as in Figure 1 on the graph paper provided. Draw them to scale. These cross-sectional areas can be presented with your maps of the sites in your report.

To calculate discharge, the cross-sectional area of rectangular subsections around each site where velocity and depth measurements are made is multiplied by the velocity at the site. The results are summed. Assume that the subsections are rectangular, defined by the depth at each vertical and divisions halfway between each vertical, as shown in Figure 1. The "lost" discharge in the triangular areas at the edges is usually assumed negligible. Discharge (Q) is then calculated as:

$$Q = w_1D_1v_1 + w_2D_2v_2 + \ldots\ldots + w_nD_nv_n$$

where w is the width in **meters**, D is the depth of the vertical in **meters**, v is the velocity at each vertical (**meters/sec**) and **n** indicates the number of velocity measurements made across the stream. Then the units of Q are in m^3/s. Record discharge in Data Table 3. Check your calculations with others.

Exercise 3: Froude number, Reynolds number and shear stress in pools and riffles

Using the average velocities (U) and depths (D) for pools and riffles of the Rocky Creek and Bear Creek (from Data Table 1, horizontal velocity profiles) calculate Froude number and bulk flow Reynolds number for each. The equation for Froude number is

$$Fr = U/\sqrt{gD}$$

where $g = 9.8\ m/s^2$. The equation for Reynolds number is

$$Re = UD/v$$

where v is kinematic viscosity ($1.004 \times 10^{-6}\ m^2/s$). Record these data in Data Table 3. Follow the units in your calculations to insure that these numbers are dimensionless.

Plot the vertical velocity profiles of Rocky Creek riffle and pool on the graph paper provided. Plot these as distance above the stream bed (depth in m) on the y-axis and velocity (m/s) on the x-axis (as in Figure 2a). Construct a log-linear plot such as in Figure 2b or Figure 3c. Figure 2b uses log (base 10) scale of depth on the x axis and velocity on the y-axis. To construct one like Figure 3c (which is really what you are doing the regression on) take the log (base 10) of the depth and plot that on the x-axis. Plot velocity on the y-axis. Note that because our streams are shallow < 1 m and the depth measurement is in meters (e.g., 0.2 m) the values for the x-axis will mostly (if not all) be negative. Don't let this worry you. If this plot is approximately a straight line then shear stress can be calculated from the slope of the profile. The equation for shear stress is

$$\tau = \rho\,(V_*)^2$$

where ρ is the density of water ($1000.0\ kg/m^3$) and V_* along a hydraulically rough stream bed is obtained from the following equation

$$V_* = b_{y \cdot x} / 5.75$$

where $b_{y \cdot x}$ is the slope of the logarithmic velocity profile. Units of V_* are m/s.

To get the slope you must calculate a linear regression for each line. Calculate the linear regression for each line using the procedures in Table 1. $b_{y \cdot x}$ is the slope of the regression line. What are the units of shear stress?

Exercise 4: Statistical tests

Once you have calculated $\overline{Y}$, s, s^2, and se for the data in Exercise 1, we can do some statistical tests. One commonly used statistical test is called the *t*-test. This tests the hypothesis that two means come from the same population (or in commonly used terminology, the two means do not differ from each other). Although many people think that statistics and hypothesis testing are just exercises in futility, they really provide a way for a scientist (and others, e.g., in court) to determine how confident they are that their data are "real" (e.g., two things really do differ from each other at a "given level of confidence"). In a good hypothesis test, the statistician or biologist decides how confident he or she wants to be before tests are done and then sticks to that level of confidence. The level of confidence that the biologist sets is related to the α-level. In one way, the α-level determines how stringent the test is. For example, if a biologist only wanted a 5% chance of being "wrong" if she said that the abundance of two stoneflies differed between two streams then she would choose an α-level of 0.05. In this case "wrong" means that she would say the stonefly abundances differed when they really did not . A more stringent test occurs if the $\alpha = 0.01$ (only a 1% chance of being "wrong") and a less stringent test would occur at $\alpha = 0.1$ (a 10% chance of being "wrong"). Please note there are other ways of being "wrong" (e.g., what happens if two means really do differ and the biologist says they don't?). α says nothing about the level of confidence of this problem.

Typically, *t* is calculated for a given level of confidence (α) based on certain formulae. This calculated value of *t* (t_{calc}) is then compared to a critical value of *t* (t_{crit})in a table for a certain number of "degrees of freedom" which are related to the number of replicates. If the absolute value of t_{calc} is > than t_{crit} then the two means are significantly different at that particular α-level. If the absolute value of t_{calc} is < t_{crit} the biologist can't really say anything about the data (e.g., failed to detect a difference, not that the two means did not differ). At the time statistics was born (when R.A. Fisher developed hypothesis testing for agriculture, another way biology contributed to applied mathematics), we didn't have computers. That is why tables are used. However, if you do statistical tests on computers with specialized statistical programs (SAS, SPSS, Systat) the output will often give you the exact α-value for the test statistic (e.g., *t*) which is often called the p-value (probability value).

Just a word of caution. Biologists also have to determine what is "biologically" significant. For example, if in the example above, the average abundance of stoneflies differed significantly (in the statistical sense) between streams by 2 individuals, the biologist must decided whether this is really biologically significant (i.e., that so little difference in stonefly abundance is important).

Each statistical test has certain underlying assumptions (often normality and equal standard deviations or variances) which should be examine before any tests are done. In this example you will use the F-statistic to test the assumption that the variances (standard deviations) are equal before you do the *t*-test. Tables 2, 3 and 4 contain the methods to do both the test of variances and appropriate *t*-tests. These tables are taken from Sokal and Rohlf (1981). The table of the critical values of F and *t* will be made available in class. If you wish to do other statistical tests for your report, statistical tables are available in almost any introductory statistics text.

Choose one set of comparisons to make (e.g., current velocity differs between riffles and pools of Rocky Creek), test for equal variances and then do the appropriate *t*-test. It would be best if each pair of students work on a different comparison and then share the results. Present your results in Data Table 3. This calculation will provide an example. You will do other calculations for other comparisons that you would like to use in your report.

Literature cited

Gordon, N.D., McMahon, T.A. and B.L. Finlayson. 1992. Stream hydrology. John Wiley and Sons, NY.

Horne, A. J. and Goldman, C.R. 1994. Limnology. McGraw-Hill, Inc., N.Y.

Sokal, R.R. and Rohlf, F.J. 1981 Biometry. Freeman and Co., New York.

Table 1. Computation of Regression Statistics, Single Value of y for each value of x.

Example data

Weight loss in mg (y_i)	8.98	8.14	6.67	6.08	5.90	5.83	4.68	4.20	3.72
Percent relative humidity (x_i)	0	12.0	29.5	43.0	53.0	62.5	75.5	85.0	93.0

Basic computations

1. Compute sample size and sums.

   $n = 9$ $\Sigma x_i = 453.5$ $\Sigma y_i = 54.20$

2. The means, sums of squares, and sums of products are

   mean of $x_i = \overline{X} = \Sigma x_i/n = 50.389$ mean of $y_i = \overline{Y} = \Sigma y_i/n = 6.022$

   sum of squares of $x_i = \Sigma\,(x - \overline{X})^2 = 8301.3889$

   sum of squares of $y_i = \Sigma\,(y - \overline{Y})^2 = 24.1306$

   sum of products $= \Sigma(x_i - \overline{X})(y_i - \overline{Y}) = -\,441.8178$

3. The regression coefficient (slope) of the line is

   slope $= b_{y \cdot x} = [\Sigma(x_i - \overline{X})(y_i - \overline{Y})]/[\Sigma(x - \overline{X})^2] = -441.8178/\ 8301.3889 = -0.05322$

4. The Y-intercept of the line is

   intercept $= a = Y - b_{y \cdot x}\overline{X} = 6.022 - (-\,0.05322)(50.389) = 8.7037$

5. The equation of the regression line is

   $y = a + b_{y \cdot x}x$

This example is modified from Sokal and Rohlf (1981).

Table 2. Testing the significance of differences between two variances.

Survival in days of the cockroach *Blattela vaga* when kept without food or water.

Females	$n_1 = 10$	$\bar{Y}_1 = 8.5$ days	$s_1^2 = 3.6$
Males	$n_2 = 10$	$\bar{Y}_2 = 4.8$ days	$s_2^2 = 0.9$

The alternative hypothesis is that the two variances are unequal. We have no reason to suppose that one sex should be more variable than the other. When we test variances like this it is what we call a two-tailed test because the variance of males > females or the variance of females > males. We calculate F_s as the ratio of the greater variance over the lesser one:

$$F_s = s_1^2/s_2^2 = 3.6/0.9 = 4.00$$

Because the test is two-tailed, we use $\alpha/2$ as our probability. We look up the critical value of $F_{\alpha/2(v1,v2)}$ where α is the type I error accepted (0.05 in this case), $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$, the degrees of freedom (d.f.) of the numerator and denominate variances, respectively. If in your calculations v_2 is in the numerator then look up $F_{\alpha/2(v2,v1)}$.

From the F-Table we find that $F_{.025(9,9)} = 4.03$ and $F_{.05(9,9)} = 3.18$. The F-value of 4.03 represents a probability of $\alpha = 0.05$ and the F-value of 3.18 represents a probability of $\alpha = 0.1$. Since our value of F_s is between 4.03 and 3.18, the probability that the variances are unequal is between 0.10 and 0.05. Since the probability > 0.05, technically the sample variances are not significantly different using $\alpha = 0.05$ as our cutoff.

After you have tested for the variances continue to Table 3 if the variances are equal and Table 4 if the variances are not equal.

This example is modified from Sokal and Rohlf (1981).

Table 3. Calculation of *t-test* when variances (test in Table 2) and samples sizes are equal.

This test assumes that the variances in the populations from which the two samples were taken are identical. If the variances of the two samples are highly significantly different do not employ this test, but carry out the test in Table 4.

For the present data, since sample sizes are equal, we choose the following expression:

$$t_{calc} = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n}\left(s_1^2 + s_2^2\right)}}$$

where Y_1 and Y_2 are the two means you are testing. μ_1 and μ_2 are the population means. Since we are testing the hypothesis that the estimated means Y_1 and Y_2 are equal, the null hypothesis is that $\mu_1 - \mu_2 = 0$; therefore, we replace his quantity by zero in the example. Since sample sizes are equal $n_1 = n_2 = n$.

In the cockroach example (Table 2)

$$t_{calc} = (8.5 - 4.8)/\sqrt{[(3.6 + 0.9)/10]} = 3.7/\sqrt{(4.5/10)} = 3.7/0.671 = 5.514.$$

Because the sample sizes are equal the degrees of freedom (d.f.) for this example are 2(n - 1) or 18. Since $Y_1 > Y_2$ or $Y_2 > Y_1$ this is a two sided test; therefore, to have an overall α-level of 0.05, we chose an α of 0.025. The critical value of t, $t_{crit} = t_{\alpha/2[d.f.]} = t_{.025[18]} = 2.101$. Since the absolute value of our observed t_{calc} is more than t_{crit}, the means are significantly different.

Table 4. Calculation of *t-test* when variances (test in Table 2) are unequal and samples sizes are equal.

When samples sizes are equal, but variances are unequal, we use the same expression as in Table 3 to calculate t_{calc}. However, with unequal variances the appropriate degrees of freedom = n -1 or 9 in this case. The critical value of t, $t_{crit} = t_{\alpha/2[9]} = t_{0.025[9]} = 2.262$. If in our cockroach example the variances were unequal, our test would still be significant since the absolute value of $t_{calc} > t_{crit}$.

These examples are modified from Sokal and Rohlf (1981).

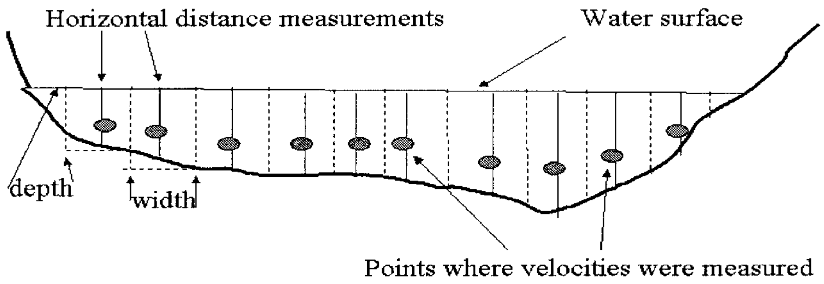


Figure 1. Definition of terms used in computing discharge from current velocity measurements. Note the variable spacing of points where velocity is measured.

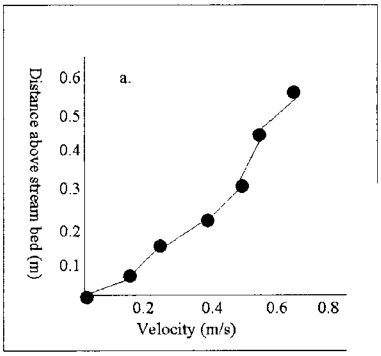


Figure 2a. Velocity profile as measured in the stream.

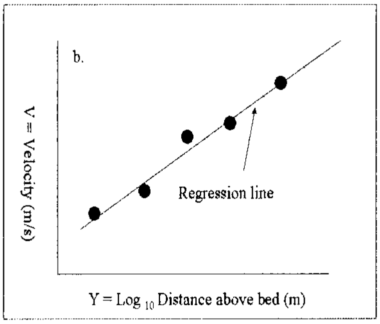
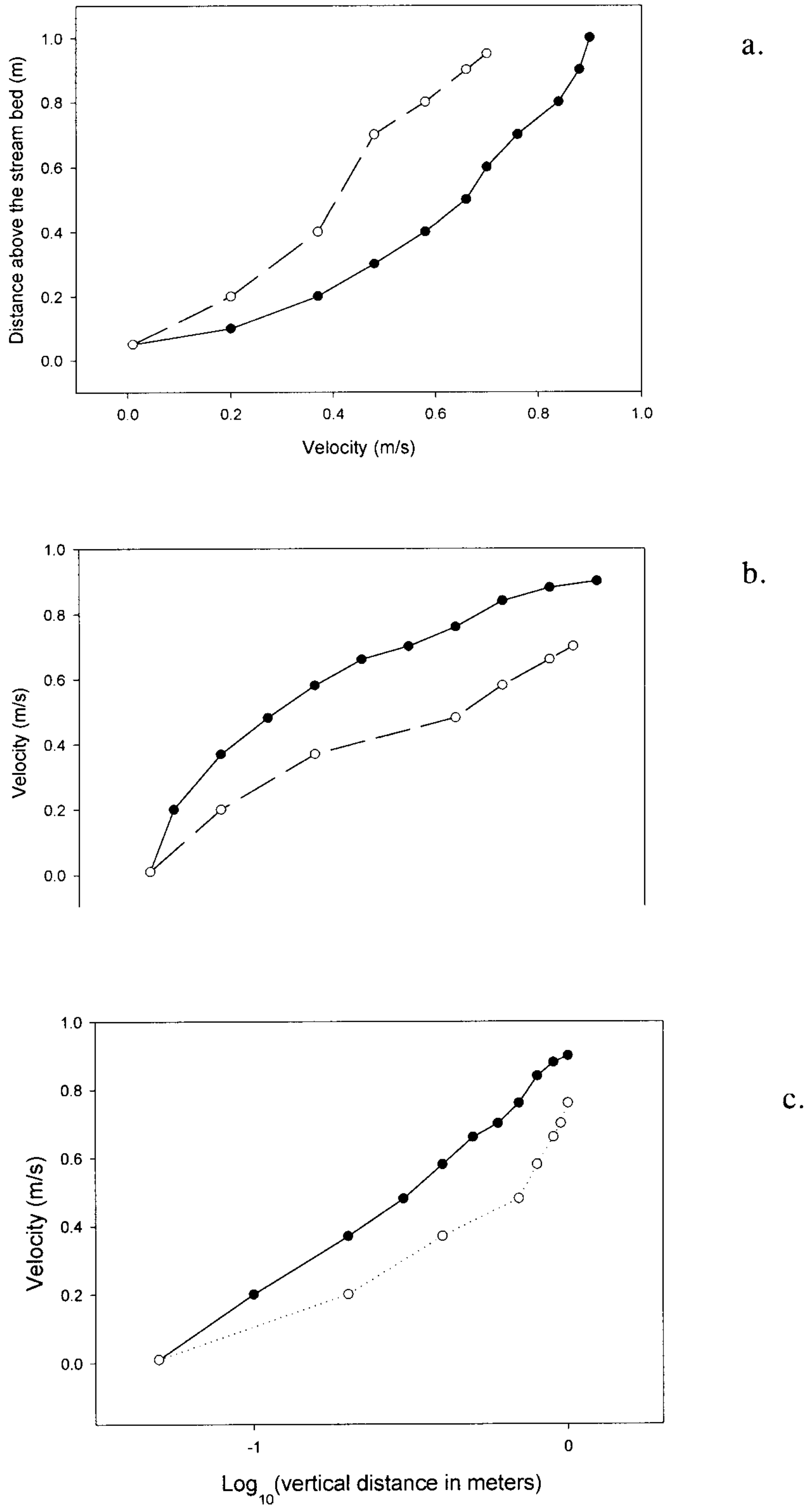


Figure2b. Velocity profile needed to calculate slope of the regression line used to calculate shear stress.

Figure 3. Some real data.



Data Table 1

Charact.	**Measure**	**Rocky Creek- Riffle**	**Rocky Creek - Pool**	**Bear Creek**
Width (m)				
Depth (m) Horizontal data only	$\bar{Y}$			
	s			
	s^2			
	se			
	n			
Velocity (m/s) Horizontal data only	$\bar{Y}$			
	s			
	s^2			
	se			
	n			
Temper. (°C)	$\bar{Y}$			
	s			
	s^2			
	se			
	n			
Cond. (mS/cm)	$\bar{Y}$			
	s			
	s^2			
	se			
	n			

DO (mg/l)	$\bar{Y}$			
	s			
	s^2			
	se			
	n			
pH	$\bar{Y}$			
	s			
	s^2			
	se			
	n			
CPOM	$\bar{Y}$			
	s			
	s^2			
	se			
	n			

Data Table 2

	Rocky Creek- Riffle	**Rocky Creek - Pool**	**Bear Creek**
Median Sediment Size			

Data Table 3

	Rocky Creek- Riffle	**Rocky Creek - Pool**	**Bear Creek**
Discharge (m^3/s)			
Fr			
Re			
Slope of regression ($b_{y \cdot x}$)			
Shear Stress (τ)			

Data Table 4

What hypothesis are you testing?
Are the variances equal? Show your work.
What is the formula of the appropriate *t*-test based on the variance test above and the equality of the number of replicates (See Table 2)?
What is the calculated value of *t*?
What is the critical value of *t* for what level of confidence?
How do you interpret your results?

Assignment:

In two weeks (specific date selected in lab):

Each student should turn in completed Data Tables 1 to 4.

Each student should write a short summary (2-3 single spaced pages) that describes differences in physical and chemical among the channel units. Provide supporting statistical tests. The information from this paper will be included in a more formal manner in a paper in the future.