Creativity3.r: Explanation

Goals of code:

- Your workspace and R Studio projects
- Calculate two-sample t-test

Note: I only describe what is new. The code file includes preparatory steps (this week, reading the data file) to make the code standalone. Explanations of reading the data file are in the creativity1 and creativity2 documents.

Saving and loading your workspace (both R and R Studio)
`rm(list=ls()); load('.Rdata');`
When you quit R, you will have noticed that it asks if you want to save your workspace. If you say yes, then all objects (variables, data frames, ...) in your workspace are saved in a file named .Rdata in your current working directory. This means that your progress is saved when you stop work. To resume, all you need to do load that saved workspace. I illustrate what I do.

`setwd('stat 401');` Use the specified folder as the working directory. You will probably need to change the folder name to something appropriate for you. Or, use File / Change dir, or setwd(choose.dir()) to set the appropriate working directory.

`rm(list=ls())` Remove any automatically loaded objects from the workspace. There may be a .Rdata file in the initial working directory. This clears away those. This step is not essential.

`load('.Rdata')` Load the .Rdata file from your new working directory. If you see an error message, that means there isn't anything saved. Not a problem if you starting something new. Is a problem if you expected to have something saved.

At this point, your workspace is restored to its status when you last quit R. If you have loaded any R packages (discussed later), you will need to load any libraries that your code needs.

R Studio projects
R Studio projects simplify saving and loading previous work, especially if you have many simultaneous projects. A project is a special file with a .Rproj extension that

identifies a working directory and set of libraries that you want for an analysis (one data set or one course, or even everything you do).

This only applies to R Studio. The equivalent in plain R is saving and loading your workspace.

Creating a project: File / New Project
The popup dialog allows you to create a project in a new working directory, or in a currently existing directory. If you have already done some work in a folder, you want the currently existing directory option. You will have to enter or browse to that working directory. If you are starting something new, choose new directory, then select empty project and provide a name for the new folder (directory) and where to put it in your folder system. R Studio will create an empty folder for your project. (You do not want the new option if you already have stuff in a folder).

You will see various options that may interest you in the future. R Studio makes it easy to link to material stored on GitHub (but not CyBox, yet) and create new R packages and web applications. We will not discuss any of these options.

Using a project:
By default, R Studio saves the last project name and restores that project (working directory, saved workspace, and loaded libraries) when you restart R Studio.

If you want to change projects, File / Open Project or File / Open Project in new session will do that. The new session choice opens a second instance of R Studio.

Two-sample t-tests: `t.test(score~treatment, data=creativity, var.equal=T)`
A formula is a convenient way to specify the response and grouping variables. The response variable is on the left-hand side (score, here) of the $\sim$ and the grouping variable is on the right-hand side (treatment, here). The `data=creativity` specifies the data frame with the variables names in the formula.

The `var.equal=T` argument specifies that the function should assume that the two groups have the same variance and to use the pooled standard deviation in the t-test. This is the customary practice, at least in the areas PMD works in. The default for t.test is the unequal variance (Welch or Welch-Satterthwaite) t-test.

The output from the t-test function includes:
two-sided p-value: p-value =
(You know it is two-sided because of the next line, which tells you that the alter-

native is “not equal to 0”.)
with error df and the t-statistic.
95% confidence interval for the difference of means
the sample averages (estimated means) for each group.

It is good practice to check that the degrees of freedom (top of output, df=) match what you expect. If not, the data weren't read or interpreted correctly.

To get something other than a 95% confidence interval, specify the desired coverage as a decimal fraction in the `conf.level=` argument (next command).