

Lab 11. Crayfish Population Ecology - Analysis

Notes:

- You may want to bring materials from previous Labs to use as a reference for RStudio.
 - Make sure to download the file titled "crayfishdata-spring2024.csv" from the Biology 200 Lab website.
-

Objectives:

1. Practice basic graphing and statistical analysis in R.
 2. Articulate hypotheses and testable predictions of those hypotheses.
 3. Evaluate the fit between your data and your predictions.
 4. Identify alternative explanations for the predicted patterns.
-

KEY WORDS: null hypothesis; *P*-value; statistical significance; continuous variable; categorical variable; linear regression; correlation; dependent variable; independent variable

Contents

1	Background	1
2	Considerations of Data Analysis	2
3	Data Graphing and Analysis	4
4	No Pre-lab Exercise.	6
5	Post-Lab Assignment	6

1 Background

Last time, we learned about studying crayfish burrows in their natural habitats. Before that, you were introduced to using RStudio to graph and analyze data. Today, we will investigate relationships between active crayfish burrows and their environment using the data we collected last week. Our primary hypothesis was that vegetation type and soil properties will impact crayfish habitat selection and, thus, the number of crayfish burrows observed.

We have several potential mechanisms that could explain the expected trends and we made the following predictions:

1. Soil moisture, pH, and temperature influence burrow density. Soil moisture will be positively correlated with number of burrows. For soil pH, burrow density will be highest at neutral pH and decrease with increasing acidity. We would expect that too high a temperature would be related to reduced burrow density as crayfish are ectotherms and need to stay moist to survive.
2. Vegetation type influences burrow density. Based on previously published studies, we might expect crayfish to prefer more disturbed, early successional vegetation such as the mowed areas of the field.
3. Soil nutrient availability could influence vegetation type and thus also serves as a factor shaping plant community composition. We predict that soil nutrients (N, P, K) will vary with vegetation type

to influence active burrow density.

4. Finally, as crayfish excavate burrows, the texture of the soil may be particularly important to them. Soil texture varies from loose sandy soils to heavily clay soils. We predict crayfish to prefer soils with substantial clay as those soils may be best for maintaining burrow structure as well as retaining water.

Now that we've collected data we can evaluate our predictions. We will conduct several different statistical analyses and graph our data. Having been introduced to R previously, this week we'll use R to visualize and statistically analyze a new dataset. You may wish to consult your handouts and notes from the last few R labs as you complete this exercise.

As a reminder, we are testing for evidence towards any of the above mechanisms by measuring the correlations between soil and vegetation parameters with the number of active burrows. As water and water-associated vegetation are each known habitats for crayfish, we would therefore expect positive correlations between each and the number of active burrows. In terms of vegetation type, we can compare the average number of burrows by each type classified with box plots. We will use graphical techniques to visualize our results and statistical analyses to test the significance of these relationships.

2 Considerations of Data Analysis

In the previous R-related lab work, we encountered both categorical and continuous explanatory variables. You may recall that we use different types of graphs and analyses depending on the types of variables. Most of our variables today are continuous, meaning we will mostly be using linear regression for analysis and scatterplots for graphing. Remember that for categorical explanatory variables (like vegetation type), we will still use ANOVA to analyze the data and boxplots for graphing.

To test a prediction, we will perform a simple linear regression analysis to create a predictive model for the number of burrows as a function of soil temperature. We can then determine whether the correlation between variables is statistically significantly different from zero and quantitatively describe how the number of burrows changes with soil moisture in our dataset.

Simple linear regression analysis allows us to quantify the relationship between active burrow numbers and relative soil moisture using the following:

- **Regression equation:** This is an equation for the line describing burrow number (y) as a function of soil temperature (x), of the form $y = b + mx$ (where b is the y -intercept and m is the slope of the line).
- **P -value:** As in the chi-square analysis we used in an earlier lab exercise, a probability of 5% or lower will be considered statistically significant. Here, our null hypothesis will be 'there is no association between active burrows and soil temperature.' Thus, the P -value gives us the probability that the slope of the line is not different from zero (which would imply zero relationship between the variables).
- **R^2 value:** The R -squared value measures how much of the variation in active burrow numbers is explained by variation in soil temperature. This value ranges between zero (no variation explained) and one (100% of the variation is explained).

Keep in mind that for each of our analyses, we're only going to be working with a subset of all of the data that we collected; we're only going to be comparing one type of interaction in each individual test. So, we will be conducting a statistical test (regression) and creating a graph for each comparison. Remember, these are the comparisons we can make with our dataset:

- Number of active burrows as a function of vegetation type.
- Number of active burrows as a function of the relative soil moisture level.
- Number of active burrows as a function of the soil pH.
- Number of active burrows as a function of soil temperature.
- Number of active burrows as a function of soil nutrients (N, P, and K).

2A Importing the Data

Now let's import that dataset and get analyzing!

1. First, download the data sheet from the lab website to your computer. Make sure you remember where you saved the file on your computer!
2. Now open your browser, navigate to rstudio.oberlin.edu and log in.
3. Before going any further, you should open a new R script file (go to File→New File→R script) so you can keep track of all the commands you run in lab today.
4. You should also probably clear out your workspace, to avoid confusion of today's data with any work from prior labs or courses. Go to Session → Clear Workspace → Yes.
5. Next, you need to upload the data file into RStudio before you can do anything with it. Just like last time, go to the Files pane and click Upload. Navigate to the file on your computer, select it, and click OK to import the file into RStudio. Now, we want to tell R to read the file and save it to memory as an object we'll call `craydata`. Type the following into your new script and then execute it in the Console:

```
library(tidyverse)
```

```
craydata <- read_csv("crayfishdata-spring2024.csv")
```

6. Now you can check to make sure the importing worked correctly by looking at your dataset via the new object `craydata`. In your source window, type `craydata` and click Control+Enter (PC) or Command+Enter (Mac) to execute the code in the Console. The first 10 lines of the dataset will appear in the Console pane.
7. You can see what R knows about each of the variables in the dataset by typing `str(craydata)` into your script (and then executing it, like you did with the previous command). Once executed, in the Console window you can see that categorical variables are called 'chr', or character, by R. You also have some 'numeric' variables. R is able to treat different types of variables differently when you are using functions so it's important to make sure your variables are interpreted correctly. Tidyverse also assigns columns as 'col_double()' when they are numeric.
8. Make sure you are familiar with the names of the variables, you'll need to use some of them later. Remember, you can always find them again by executing `names(craydata)`.

3 Data Graphing and Analysis

3A Number of active burrows as a function of soil temperature

You will work with the soil moisture data first, examining how soil moisture in a quadrat affects the number of the active crayfish burrows in that quadrat. Recollect that our quadrats were 1 m², so that the number of active burrows is per square meter (which is burrow density).

Graphing

9. Before we proceed with any analysis, let's make a graph to see what the data set looks like. Before you type in any code, do you remember what relationship you predicted for soil moisture and number of burrows?

Once you recall your prediction, enter the following at the prompt to make the graph. **Does it look like the data show the pattern you expected?**

```
craydata %>%
  ggplot() +
  geom_jitter(mapping=aes(x=soil.moisture, y=number.active.burrows)) +
  xlab("Soil moisture (0-10)") +
  ylab("Burrows per square meter") +
  ggtitle("Active burrows as a function of soil moisture")
```

10. Now, that was just considering one variable at a time. We might wonder whether the different vegetation types differed in soil moisture. We can easily adjust our code to add that information on the graph:

```
craydata %>%
  ggplot() +
  geom_jitter(mapping=aes(x=soil.moisture, y=number.active.burrows, size=1)) +
  guides(size="none") +
  facet_wrap(~ vegetation.type) +
  xlab("Soil moisture (0-10)") +
  ylab("Burrows per square meter") +
  ggtitle("Active burrows as a function of soil moisture")
```

11. Does it look like vegetation types differ for soil temperature? You can think about the variable of vegetation type further for your post-lab exercise.

Regression Analysis

12. Time to run a simple linear regression and view a summary of the results. Execute the following code to create the regression, summarize the results, and save the results as an object (`soil.temp_mod`):

```
soil.moisture_mod <- summary(lm(number.active.burrows ~ soil.moisture, data=craydata))
```

13. To view the results summary, including *P*-values for the intercept and slope (the Estimate for 'soil.moisture') and the R-squared value, enter at the prompt: `soil.moisture_mod`

Lab 11. Crayfish Population Ecology - Analysis

Your results should look something like the image below. From the output, we should notice a few things.

- The Estimate column contains the value of the y-intercept, (Intercept) = 2.37578 burrows per square meter
- The slope is the Estimate value for soil.moisture. -0.08579 burrows for every unit increase in soil moisture.
- For the R-squared (R^2) value, we use the Adjusted R-squared = 0.007311 or approx. 0.7 percent
- The *P*-value for our slope is listed both in the Coefficients table under Pr(>|t|) AND on the last line of the output, after the F-statistic. *P* = 0.236.

With that information, we can write the regression equation! For the above output that looks like (using 2 significant digits and rounding): $Y = 2.37 - 0.08X$

```
Call:
lm(formula = number.active.burrows ~ soil.moisture, data = craydata)

Residuals:
    Min       1Q   Median       3Q      Max
-2.2042 -1.5179 -0.5608  1.3534  6.4821

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.37578    0.56218   4.226 0.0000852 ***
soil.moisture -0.08579    0.07163  -1.198   0.236
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.856 on 58 degrees of freedom
Multiple R-squared:  0.02414, Adjusted R-squared:  0.007311
F-statistic: 1.435 on 1 and 58 DF, p-value: 0.2359
```

Is the slope significantly different from zero in this example?

Updating your graph with analysis results

14. We can add the regression line, equation, *P*-value, and R-squared to a plot by executing the following several lines of code:

```
craydata %>%
  ggplot(mapping=aes(x=soil.moisture, y=number.active.burrows)) +
  geom_jitter() +
  xlab('Soil moisture (0-10)') +
  ylab('Active burrows per square meter') +
  ggtitle('Active burrows as a function of soil moisture') +
  geom_abline(slope=-0.08579, intercept=2.37578) +
  annotate('text', label='Y = 2.37 - 0.08X, R-squared = 0.007, P = 0.23',
         x = 5.5, y = 7, size = 3)
```

REMEMBER TO CHANGE THE INFO FOR THE SPECIFIC DATA THAT YOU ARE GRAPHING!
And round all your values to 2 decimal places. Note that for the annotate line, you need to choose values for x and y that make sense on your actual graph.

15. Once you've got the plot the way you want it, you can export it from the Plots pane in RStudio, as you've done in a previous lab.

Other Analyses

Using the above instructions as an example, write the code to perform two more analyses, each using a different explanatory variable. You may choose which explanatory variables to use but of course your response variable will be number of burrows for both analyses.

For each of the two analyses you will need to:

1. Make the plots to visualize the data.
2. Create the models, using a new name for each analysis to differentiate which variables are used in each.
3. Remake the plots with the addition of the statistical information from the models.
4. Export the finalized plot, so you can include it in the assignment you will submit.

Refer to the predictions listed earlier in the handout for a reminder of your options for explanatory variables.

Remember that for a categorical explanatory variable you will use `geom_boxplot` instead of `geom_jitter`. On the graph, you can report the *P*-value alone for the ANOVA (you do not need to run the Kruskal Wallis test), you do not need to add additional text. And of course you do not need the `geom_abline` for a boxplot either. See your earlier R labs for examples of boxplot code.

4 No Pre-lab Exercise.

5 Post-Lab Assignment

Write up your responses to the following questions. Submit a single document for your group – be sure to include everyone's names on the submitted file. **Save the file as a PDF.** Submit the file via the 200 Lab website as usual.

1. Did you find data bearing on how aspects of the environment might relate to crayfish habitat preferences? Respond to this question by composing a “Results” section discussing your findings, similar to this section in a published research paper. Begin by restating your hypothesis for this study and the mechanism(s) you focused on.

Then clearly state the predictions tested in each of the three analyses you ran. For each prediction tested: (a) briefly discuss the results including the relevant statistical data and (b) insert pictures of the appropriate graphs with informative captions.

One by one, address whether each test supported the hypothesis and mechanisms, and if so, how strong the evidence was.

Make your graphs as professional as you can – no typos, include labeled axes with units, a caption, include the regression equation and line, R-squared value, and P-value. Values should be to two decimal

Lab 11. Crayfish Population Ecology - Analysis

places.

2. Be critical of the ecology: Provide one well-developed alternative hypothesis mechanism (i.e., other than that stated) that could also produce the results we found. Be sure to describe how your alternative hypothesis would explain the results found.
3. Be critical of the study design: What flaws did you find with our study? Should the experimental design or methods used to collect our data be changed? If so, how? Are our analyses actually answering the questions we are trying to address? Are there any variables or problems we did not account for? How could we do so? Answer these questions in critique.
4. Would you expect the methods we used to work equally well across different seasons? Why or why not?

Make sure all group members sign the Honor Code (or type their names in). This will indicate your affirmation that everyone individually completed the R analysis and that everyone contributed meaningfully to the group discussions today and to completion of the assignment.

If you will not be able to submit the assignment before leaving lab today, speak with your instructor.