**113 Lab Learning Objectives**

**Week 12: Information and Natural Selection lab #7**

Learning Objectives for Environmental Information and Natural Selection

*Skills*

* Use statistics to determine the level of significance of data.
* Prepare graphical representations of data and its variance.

*Cognitive*

* Employ a scientific approach to answer biological questions and test hypotheses.
* Analyze experimental data and reach logical conclusions.
* Construct a generalizable explanation linking genetically defined taste capacity to natural selection and evolution.
* Identify possible sources of environmental information that communicates the level of toxicity for plant tissues.
* Design an experiment to use model organisms to extrapolate potential toxicity of a compound or mixture.

**Week 12: Information and Evolution Lab #5**

Learning Objectives for Bacterial Evolution

*Skills*

* Pipet accurately.
* Work with bacterial cells using sterile technique.
* Make dilutions of stock solutions.

*Cognitive*

* Employ a scientific approach to answering biological questions and test hypotheses.
* Describe the big idea of evolution based on lab experiences.
* Explain how antibiotic resistant bacteria can appear quickly in the population.
* Design directed evolution process to select antibiotic resistant bacteria.
* Formulate an hypothesis how antibiotic resistant bacteria evolve outside the laboratory.
* Propose a mechanism that allows bacteria to evolve rapidly when exposed to antibiotics.

**Bio113 Week 12**

Before you come to lab

1) Look up the information about TAS2R38 human gene (three links on lab web page).

2) Read about MRSA and why evolution of drug resistance is a medical concern.

3) Answer each of these four questions in two sentences or less.

A) How widespread is TAS2R38 gene among animals?

B) What is a null hypothesis? What is the correct way to use a p value in conjunction with a H0?

C) What is MRSA and why should you be concerned about it?

D) What human behavior leads to antibiotic resistance in human pathogens?

**Week 12**

In Lab

**Evolution: Directed Evolution of Antibiotic Resistance in Bacteria (a 6 week project)**

4) Analyze your results from your last directed evolution experiment.

5) Design your final experiment to determine whether your bacterial cells have evolved antibiotic resistance. You will want to test all 4 antibiotics using the disc method. Each group will use their frozen stock of parental (prior to directed evolution) bacterial cells on LB plates without antibiotic. You will want to compare the evolved strains of bacteria with the original strain of bacteria. How well does each stain grow in the presence of all four antibiotic disks? Remember to use the negative control disk too. This is your final experiment that you will present in your final oral presentation. You will have 9 plates to use.

**Information and Evolution: Genetics of Perceiving Danger Module (an 8 week project)**

6) Discuss brine shrimp toxicity data and conduct statistical analysis of your data. Did you find any statistically significant results? How will you indicate which results are significant and which ones are not? Can you have smaller or bigger statistical differences?

7) After conducting the Chi square test using RStudio, produce a graph using Excel for your final written lab report on tasting and toxicity with regards to natural selection.

8) Optional review session once everyone has finished the two lab modules.

**Do the plant extracts contain any toxins?**

You have extracted compounds from two plant tissues and we want to know if there are any significant survival differences with the two treatments. We are going to use the Chi squared test.

Generate a electronic or paper table that looks like the one below. We have provided you with some simulated data so you can walk through the sample analysis process and know that you have done the test properly. Replace the simulated data with your own data.

|  |  |  |  |
| --- | --- | --- | --- |
| Treatment | # Alive | # Dead | Notes |
| water control |  |  | Should add to 80; obtain results of water control from other groups |
| methanol (solvent control) |  |  | Should add to 80; obtain results of methanol control from another group |
| Tissue A |  |  | = 80 (1st set of tubes #4-11 pooled) |
| Tissue B |  |  | = 80 (2nd set of tubes #4-11 pooled) |

*Null Hypothesis = the two treatments are equally toxic.*

**Written Lab Report**

Your next graded assignment is to write a paper that incorporates all the data associated with taste, genotyping and shrimp bioassay for toxicity. This paper is due as indicated in the online lab schedule. This experiment has been a group project until now, when it is time to write the paper. You may share methods and the graph, but each person turns in his or her own single-authored paper. List your lab mates as coauthors, but your name comes first. Use CATME peer evaluation to characterize student participation in the collective effort (prior to writing).

**Data Analysis - Making Sense of Your Results**

**by Dr. Patricia Peroni and A. Malcolm Campbell**

After using RStudio to calculate the *p* values for your brine shrimp experiment, graph your data using Excel. Be sure to include your controls in the graph. Also include the p value you calculated.

Examine the graph and note the most striking results (trends).

**From Trends to Conclusions: Hypothesis Testing**

In this experiment, and in virtually all research, you used samples from your study population, *i.e.*, your brine shrimp. Because you used samples, the results in the graphs could reflect:

real differences in brine shrimp survival and mortality among treatments,

or

chance results associated with sampling.

For example, imagine that you obtained two randomly selected samples of 25 students from the Davidson College population, and determined the sex ratio for each sample. The chances that the two samples would have sex ratios identical to each other, or to the true sex ratio of the entire student population, are very small.

Hypothesis testing allows us to estimate the probability that the results from two treatments would be at least as different as the ones obtained simply due to chance events associated with sampling. If the probability of chance being the source of the measured differences is very small (usually < 0.05), then biologists feel comfortable concluding that we have evidence that brine shrimp react differently to the two treatments.

**Hypotheses** - Hypotheses always refer to the overall populations’, not the samples’, true conditions. There are two types of hypotheses: null and research hypotheses.

The **null hypothesis (H0)** always posits that the true condition for the dependent variable (in this experiment, mortality of brine shrimp) is equal for the various treatments. For example:

**(H0): The true distributions of brine shrimp survival and mortality are equal for the methanol control and plant extract treatments.**

The null hypothesis assumes that any differences in mortality observed between the treatments resulted from chance events associated with sampling. The goal in hypothesis testing is to determine if our results from the treatments are so different from the control condition that the probability of obtaining our results when the null hypothesis is true is extremely small. In statistical parlance, we want to determine if we have sufficient evidence to reject the null hypothesis (and if we do not, we state that we "fail to reject" the null hypothesis, not that we "accept" the null hypothesis).

The **research hypothesis (H1)**, predicts that the null hypothesis is false. Research hypotheses can be non-directional or directional. For example,

**The true distributions for brine shrimp survival and mortality   
for the methanol control and tissue extract treatments differ.**

**OR**

**The true brine shrimp mortality for tissue extract A treatment   
is greater than that of the methanol control.**

You make a directional research hypothesis when common sense or past experience tells you that if the two treatments differ, it would be in one direction only. For example, if any true difference in brine shrimp survival and mortality exists between the water control and a methanol treatment, we expect the methanol treatment to cause greater morality than the water control. You will use directional research hypotheses for your analysis of the results from this experiment.

**Statistical Tests of Hypotheses**

Statistical tests are used to determine if we can reject the null hypothesis. Selection of the correct statistical test for analysis of an investigation’s results depends upon the study’s research design, the types of independent and dependent variables involved, and the sample sizes used.

You will use Chi square tests of independence (also referred to as a Chi square tests of contingency table data) to determine if you can reject any of your null hypotheses for this investigation. Several statistical tests use the Chi square distribution, so do not be confused if this test differs from one used to compare observed phenotypic ratios with Mendelian genetics or observed genotype frequencies with Hardy-Weinberg Theory expectations. The calculations will be generated when you provide the correct computer code to the R statistical software program.

Example: Water control vs methanol treatment

***H0***: The true distributions of survival and mortality of brine shrimp for the water control and the methanol treatment are equal.

***H1***: The true distribution of survival and mortality of brine shrimp is greater for the methanol treatment than for the water control.

We begin by constructing our contingency table for the two treatments.

**Table 1.** Number of brine shrimp that survived and died when exposed to water or methanol treatment.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Treatment** |  | **Totals** |
| **Brine shrimp** | **Water Control** | **Methanol** |  |
| # alive | 76 | 12 | 88 |
| # dead | 4 | 68 | 72 |
| **Totals** | 80 | 80 | **160** |

The data are used to calculate the Chi square test statistic using this equation:



c = the number of classes of the independent variable;

r = the number of attributes (alive or dead) or outcomes for the dependent variable;

oij = the number of individuals from the jth group (control or methanol treatment)   
 observed for the ith attribute (alive or dead);

eij = the number of individuals from the jth group expected to show the ith attribute.

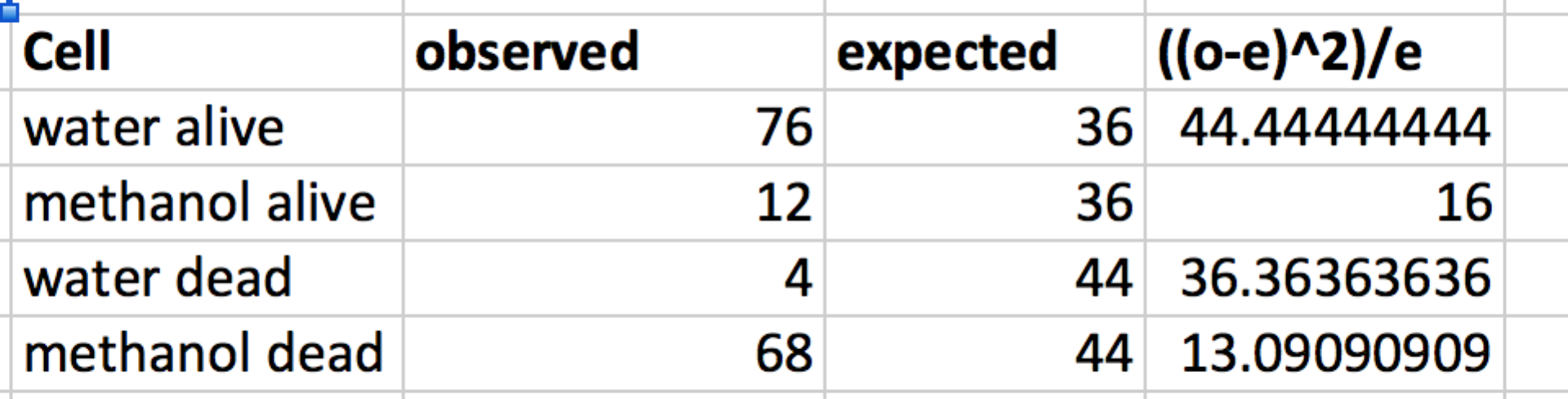
As the magnitude of differences increase between the observed and expected values for each combination of the independent and dependent variable, so do the values of Chi square.

Expected values are calculated by pooling data from the two treatments, which yields the results expected for each sample if the null hypothesis is true.

The R statistical software you will use automatically calculates the expected values. A full account of exactly how the expected values are calculated is included in an appendix to this document. If we were conducting the Chi square test by hand we would calculate

for each cell in our table.

**Table 2.** Calculations for the Chi square test.



**Table 3.** Observed (**bold** font) and expected numbers (*italics*) of live and dead brine shrimp exposed to a water or a methanol treatment.

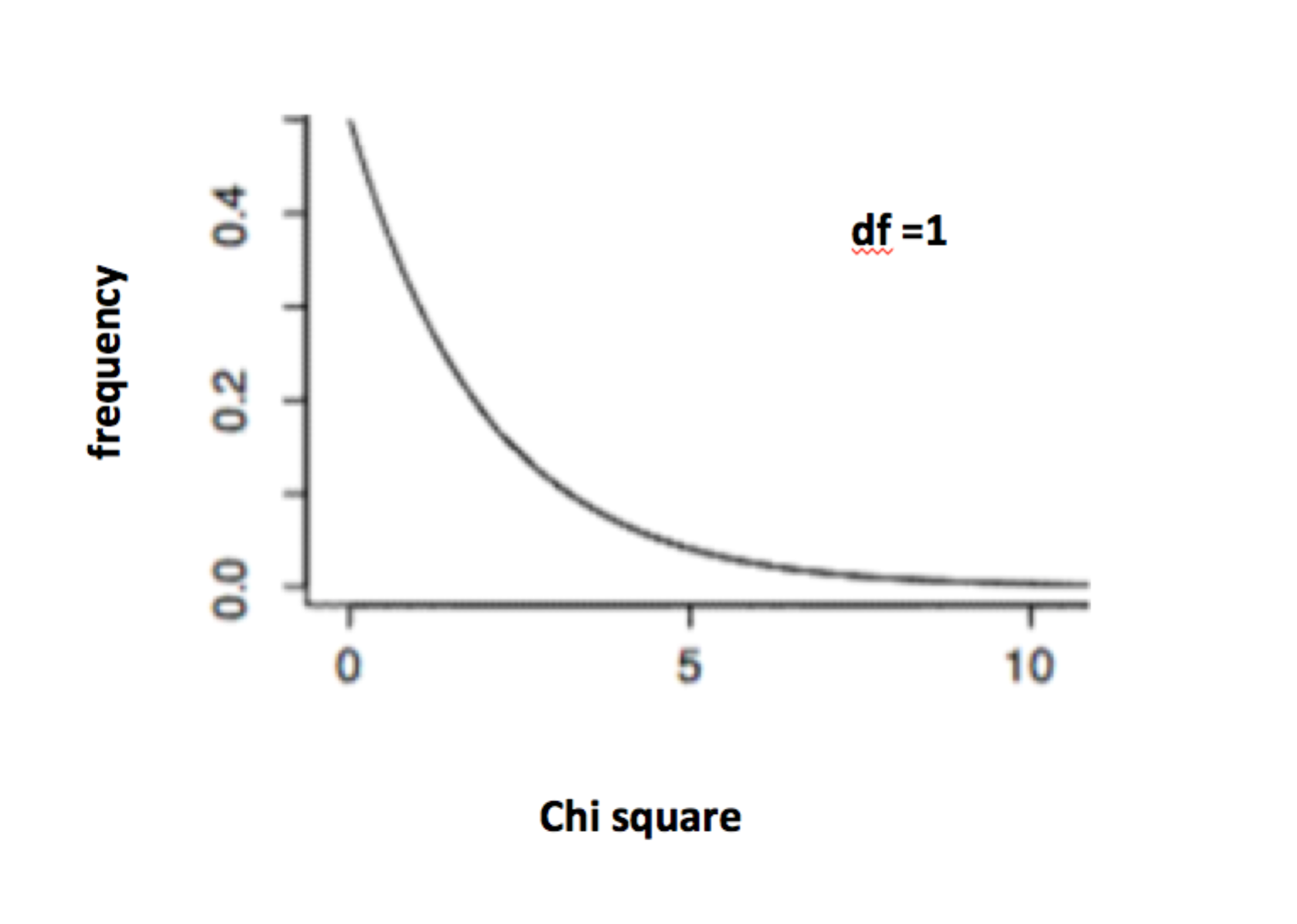
|  |  |  |  |
| --- | --- | --- | --- |
|  | **Treatment** |  | **Totals** |
| **Brine shrimp** | **Water Control** | **Methanol** |  |
| # alive | **76 (observed)**  *36 (expected)* | **12**  *36* | **88** |
| # dead | **4**  *44* | **68**  *44* | **72** |
| **Totals** | 80 | 80 | **160** |

**2 = [((o-e)2)/e], which in this case = 109.9**

Now we can calculate the probability (*p*) of obtaining a data set that would yield a 2 test statistic > 109.9 if the null hypothesis is true. Recall that if the observed and expected values for each cell were equal, **2** would equal 0.

To calculate the *p* value, we need to consult the correct Chi square distribution. Chi square distributions are defined by the degrees of freedom, df, they have. In this example, we need to use a Chi square distribution with df =1 (Figure 1). [For this Chi square test of independence, you calculate the df as: (# states of independent variable -1) x (# states of the dependent variable-1).] The graph shows how often we would expect to obtain samples with certain ranges of **2** values from a population for which the null hypothesis is true if we repeated this experiment many, many times. We calculate the probability of obtaining a **2** test statistic greater than a certain value by determining the proportion of the area under the curve that fits that description.

**Figure 1.** The **2** distribution for df = 1.

In Figure 1, we see that if we repeated this experiment many, many times in a population for which the null hypothesis is true, we would expect most **2** test statistics to be < 10. The **2** test statistic calculated from our example data was 109.9 which is way bigger than 10. The actual probability of obtaining a **2** test statistic > 109.9 from a population for which the null hypothesis is true = 1.03 X 10-25 , which is our *p* value (calculations not shown). We reject the null hypothesis whenever the *p* value is < 0.05. Therefore, we would reject our null hypothesis and conclude that methanol is toxic to brine shrimp.

R is an open source computing tool that has grown in popularity very quickly. In Biology 113 lab, you will learn some basic commands to execute your statistical analysis using R. The following directions were written for the Macintosh operating system. R can be run in Windows and Unix, but we will use Macintosh.

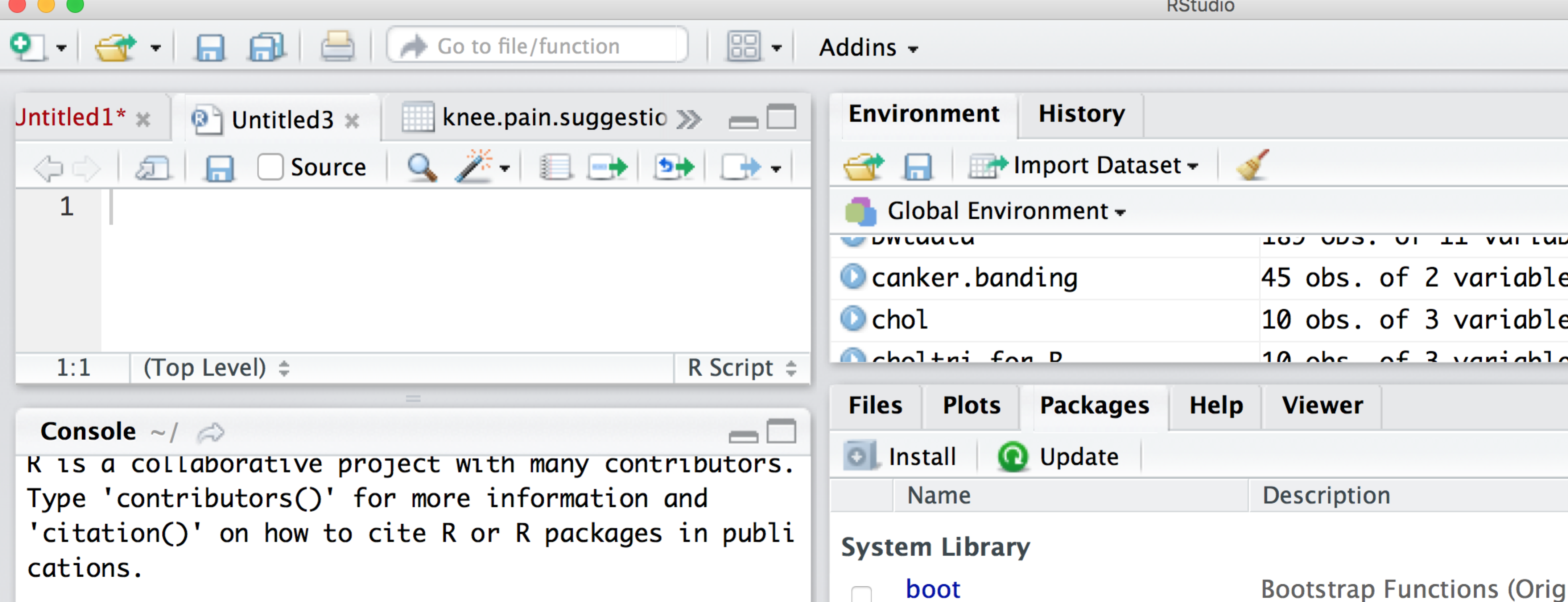
**Using R Software for a Chi square Test of Independence**

**Open** **RStudio.** Your instructor will show you where to find the RStudio icon on the lab computer desktop. Click on the File menu, choose New File 🡪 R script to produce an empty script window.

**Acquaint yourself with the window.** The RStudio window has four sections.

The top left is the Script section. You will type your commands into the script section.

The bottom left is the Console section where your program will run. You will not need to use the two right hand sections for Bio113 lab.



**SCRIPT**

**CONSOLE**

***What are the hashtags in R commands?***

A “#” symbol at the beginning of a line of R code indicates that the material following the hashtag is a comment intended to be read by a human, and not a command to be executed by the software.

***Commands to type into R Script section of RStudio***

R is picky about spacing and punctuation. You can cut and paste the comments and commands below into the R Script window. Just make sure to change the values for numbers alive and dead to those **you actually collected**. The code below uses the hypothetical data in Table 3 above.

1) Cut and paste the following. See how RStudio changes the colors to distinguish data from comments.

# sample Chi square test of independence for Bio 113 brine shrimp experiment

# comparing distributions of living and dead brine shrimp for two treatments:

# first step: provide R with the data

# to do this, give each treatment a vector (c)

# with the number of living first and the number of dead second (alive,dead)

water <-c(76,4)

methanol <-c(12,68)

tissue\_A <-c(62,18)

tissue\_B <-c(42,38)

2) Put your cursor in front of the word “water” and click on the “Run” button near the top right of your Script section. Do the same for each line of data you want to enter into RStudio. You will know this worked when the top right section shows the entered data.

3) Cut and paste the following. Execute the command by placing your cursor at the beginning of the line and clicking the “Run” button.

# start by comparing the water control with the methanol treatment

# make a data frame that turns the vectors into a contingency table.

# In this example, I named the data frame "wVmeth"

wVmeth=as.data.frame(rbind(water,methanol))

4) Cut and paste the following. Execute the command as usual.

# make a vector to label the two states of the dependent variables, Live or Dead

names(wVmeth)=c("live","dead")

5) Cut and paste the following. Execute the command as usual.

# tell R to conduct a Chi square test on the contingency table "wVmeth"

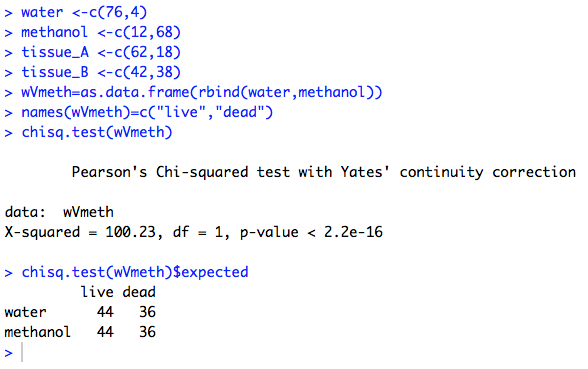
chisq.test(wVmeth)

6) Cut and paste the following. Execute the command as usual.

# Tell R to make a table of the expected values predicted by null hypothesis

chisq.test(wVmeth)$expected

You should see the commands and results of the test appear in the **Console** section (below the R script section). For the sample data in Table 3, your outcomes should match those shown below:

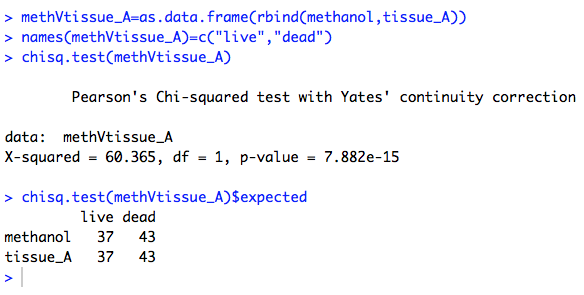


The **2** test statistic = 100.23, df=1, *p* value < 2.2 X 10-16. The table below the Chi square test results provides the values predicted by the null hypothesis. You should reject the null hypothesis. The test statistic does not exactly equal the one we calculated by hand using these data because R adds in some extra calculations to make the test a bit more reliable. Add the *p* value to your Excel graph as appropriate.

**Running More Chi square tests**

You need to edit the commands in the R script (upper left section) to conduct Chi square tests to conduct other pairwise comparisons: methanol vs tissue A and methanol vs tissue B. For each comparison, you should change the object name from “wVmeth” to a name that describes the treatments you’ll compare in all commands that include that object name. For example, you might rename it “methVtissue\_A”. Once you make those edits, execute the commands so that R can analyze your data.

# You will want to edit the commands to perform a comparison of methanol alone to

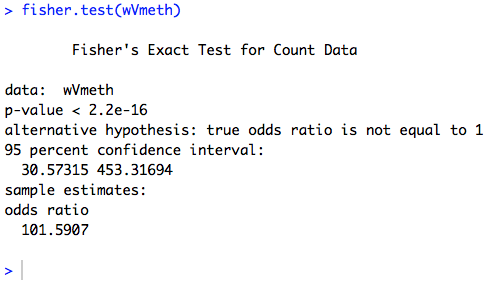
# extract from tissue A.

**The Rule of Five for Chi Square Tests of Independence**

Results from Chi square tests of independence are only reliable when less than 20% of the contingency table cells have expected values < 5. Since your tables will only have 4 cells, that means that all of the expected values in your tables will need to have expected numbers > 5. This is why we always examine the table of expected values for each Chi square test.

**What if one or more cell has an expected number < 5?**

If one of your expect values is less than 5, you must perform the Fisher Exact test. We only use the Fisher Exact test when we cannot use the Chi square test, because the Fisher Exact is less powerful. The R command for a Fisher Exact test is:

****

# fisher.test(dataframe name)

fisher.test(wVmeth)

An example of the R output for a Fisher Exact test is provide here. This test does not have a test statistic similar to **2**. It only provides a *p* value, and you reject the null hypothesis if the *p* value is < 0.05.

**Appendix**

**Calculating the Expected Values for a 2 Test of Independence**

We have the observed results for the numbers of live and dead brine shrimp in each treatment, but how does R calculate the expected number in each category?

If the two treatments have the same distribution for survival and mortality, then we can consider the data from the two treatments as independent samples taken from the same population. This means that we can estimate the number of live and dead brine shrimp expected from each treatment if the null hypothesis is true by pooling the data from both treatments and calculating the fractions of all live and dead brine shrimp in your data table. Since the data set contains a total of 160 shrimp, and 88 of them survived, the fraction of live brine shrimp in the data set equals 88/160 = 0.55. So, if the null hypothesis is correct, we expect 55% of the brine shrimp in each treatment to be alive by the end of the investigation. Since the **2** test statistic requires us to use the numbersobserved and expected, we simply multiply 0.55 by our sample size for each treatment (80 per treatment), so we expect 44 of the 80 brine shrimp in each treatment to survive if the null hypothesis is true.

We use the same strategy to calculate the number of brine shrimp expected to have died by the end of the study if the null hypothesis is true. A total of 72 brine shrimp out of the initial 160 died, so the fraction of brine shrimp expected to die = 72/160 = 0.45. To obtain the expected number of dead brine shrimp the null hypothesis predicts for each treatment, we multiply 0.45 by the sample size for each treatment (80 in this case). So, we expect 36 of the brine shrimp in each treatment to have survived. We now can produce a table with the numbers observed and expected for each combination of independent and dependent variable.

Observed (**bold font**) and expected (*italics*) numbers of live and dead brine shrimp exposed to a water control and a low concentration methanol treatment.

|  |  |  |  |
| --- | --- | --- | --- |
| **Treatment** | | | **Totals** |
| **Brine shrimp** | **Water Control** | **Methanol** |  |
| # alive | **76 (observed)**  *36 (expected)* | **12**  *36* | 88 |
| # dead | **4**  *44* | **68**  *44* | 72 |
| **Totals** | 80 | 80 | **160** |