Design and Analysis of Ecological Data
Classical Tests

(Written by Kevin McGarigal, but borrowed heavily from Michael Crawley, The R Book (2007))

The purpose of this lab exercise is to familiarize you with a wide range of classical statistical tests for single-sample and two-sample problems in R. It is assumed that you are already familiar with the basic theory behind these tests, so the focus in this lab will be on the practical aspects of how to use R to conduct these tests. If you are not familiar with the theory behind these tests, then focus on the purpose of the test and the mechanism for conducting the test in R for now. Here is outline of what is included in this lab exercise:

1. Set up your R work session

Open R and set the current working directory to <u>your</u> local workspace, for example:

setwd('c:/work/stats/ecodata/lab/classical/')

Load the biostats library (which is not a formal library) by typing, substituting the appropriate path:

source('.../biostats.R')

2. Single sample

With a single sample, the questions we might want to answer include:
- What is the mean or median value?
- Is the mean or median value significantly different from current expectation or theory?
- What is the level of uncertainty associated with our estimate of the mean or median value?

And in order to be reasonably confident that our inferences are correct, we need to establish some facts about the distribution of the data, such as:
- Are the values normally distributed or not?
- Are there outliers in the data?

- Is there evidence for serial correlation (autocorrelation, or lack of independence among observaations)?

Non-normality, outliers and serial correlation can all invalidate inferences made by standard parametric tests like Student's *t* test. In cases involving non-normality and/or outliers, it is better to use a non-parametric technique such as Wilcoxon's signed-rank test. If there is serial correlation in the data, then you need to use alternative procedures such as times series analysis or mixed effects models which go well beyond the scope of this lab.

Let's consider the questions above for a single sample representing reproductive catastrophe rates for marbled salamanders in 13 vernal pools. First, read in the data:

catrate<-read.csv('catrate.csv',header=TRUE)
catrate

The data represent the observed reproductive success of marbled salamanders in 13 vernal pools over a period of 7 years in western Massachusetts. Catastrophe rate (cat.rate) is the proportion of years with any breeding effort (i.e., >0 breeding females) in which the number of emerging juveniles per breeding female (i.e., fecundity) was less than 1, a level that we deemed to be a reproductive "catastrophe".

2.1 Distribution plots & tests for normality

We already learned how to examine the distribution of variables and assess the occurrence of "outliers" in a previous lab, so we will not review those techniques here. Here, we will use the set of simple plots produced with uv.plots() function in the biostats library to graphically examine the distribution of cat.rate:

uv.plots(catrate,var='cat.rate')

What do the these plots reveal about the distribution?

Given the relatively small sample size (N=13), it may be difficult to clearly "see" the shape of the distribution or be certain of its shape. And while these plots are most informative, we might also want to formally test whether the data come from a normal distribution. We can use the *Shapiro-Wilk test* for this purpose, as follows:

shapiro.test(catrate$cat.rate)

The *p* value is an estimate of the probability that a particular result, or a result more extreme than the result observed, could have occurred by chance, if the null hypothesis were true. Thus, a small *p* value in this case indicates that the data is unlikely to have come from an underlying *normal* distribution.

What does the Shapiro-Wilk test result tell how us about cat.rate?

There are many other tests of normality. The nortest library provides a suite of such tests (see the corresponding help files for a description of each of these tests), including the following:

```
library(nortest)
ad.test(catrate$cat.rate)
sf.test(catrate$cat.rate)
cvm.test(catrate$cat.rate)
lillie.test(catrate$cat.rate)
pearson.test(catrate$cat.rate)
```

What do these tests say about the normality of cat.rate? Are they in substantial agreement or disagreement?

2.2 Tests for difference from expectation

One of the simplest one-sample tests is whether the observed mean (or median) is significantly different from a specified value. The default null hypothesis is that the mean is equal to zero, but this is not a very meaningful null hypothesis in many cases. We might want to test the hypothesis that the observed mean cat.rate (0.54) is significantly different from an expected value of 0.28, where the expected value is based on the null hypothesis that reproductive catastrophes are caused solely by late pond-filling in the fall – which causes the dessication and/or freezing of the eggs prior to inundation. Under this causal model, 2 of the 7 (2/7=0.28) years of the study between 1999-2005 experienced late pond filling. Is the mean observed catastrophe rate significantly different from expected?

If the data are normally distributed, we can use the *Student's t test*, as follows:

```
t.test(catrate$cat.rate,mu=2/7)
```

What does the Student's *t* test say about the probability that the observed catastrophe rate comes from a population in which catastrophes are solely the result of late pond-filling in the fall?

Note, the default *t* test is a two-sided alternative; i.e., the alternative hypothesis is that the observed mean is not equal to the expected mean, which can mean less than *or* greater than the expected mean. In some cases, it may be more meaningful to specify a one-sided alternative; more specifically, that the observed mean is less than the expected mean or that the observed mean is greater than the expected mean. We can conduct the latter one-sided test as follows:

```
t.test(catrate$cat.rate,mu=2/7,alternative='greater')
```

Does this change the conclusion about the significance of the observed catastrophe rate?

Note, by the Central Limit Theorem, means of a sample from a population with finite variance approach a normal distribution regardless of the distribution of the population (i.e., regardless of the distribution of the data itself). Rules of thumb say that the sample means are basically normally distributed as long as the sample size is at least 20 or 30. For a t-test to be valid on a sample of

smaller size, the population distribution would have to be approximately normal. Thus, the t-test is invalid for small samples from non-normal distributions, but it is valid for large samples from non-normal distributions. In the case of small samples from a non-normal distribution, we can use the *Wilcoxon's signed rank test* (also known as the *Mann-Whitney test*), as follows:

wilcox.test(catrate$cat.rate,mu=2/7,conf.int=TRUE)

Does the Wilcoxon's signed rank test agree with the Student's *t* test result? Which is more trustworthy in this case?

Note, the default test is again a two-sided alternative; in this case, the alternative hypothesis is that the observed median differs from the expected median (specified as *mu* in this function, but not to be confused with the *mu*=mean in the t.test). A one-sided alternative hypothesis that the observed median is greater than the expected median can be tested as follows:

wilcox.test(catrate$cat.rate,mu=2/7,alternative='greater',conf.int=TRUE)

Does this change the result?

In this example, we treated the data as continuous, but in fact the data are discrete, taking on integer values ranging from 0 to 7. An alternative test is the *binomial test*, which is suitable for data consisting of a single trial in which each observation can take on one of two values; e.g., success or fail, present or absent, live or die. We might ask the question, what is the evidence that reproductive success is more likely than reproductive failure? The answer comes from a two-sided binomial test. We observed a total of 33 successes out of 61 years when summed across all 14 ponds. How likely is a response of 33/61 if the populations are actually the same (i.e., p=0.5)? We use a binomial test for this, specifying the number of successes (33) and the total sample size (61), as follows:

success<-sum(catrate$success)
years<-sum(catrate$years)
binom.test(success,years)

Note, we had to first compute the number of successes and the total number of years pooled across all 14 ponds.

Instead of testing the null hypothesis of equal probability of success and failure (i.e., p=0.5), we might instead ask the question, what is the evidence that reproductive success is more or less frequent than expected under the causal model described above, in which 5/7years are expected to be successful? Conducting this test is as simple as specifying the expected probability of success, as follows:

binom.test(success,years,p=5/7)

Note, we could have just as easily constructed this test for failures instead of successes. In addition, note again that the default test is a two-sided alternative. We might instead prefer the one-sided alternative hypothesis that the observed success rate is less than the expected success rate, which can

be tested as follows:

binom.test(success,years,p=5/7,alternative='less')

Note, instead of treating each pond-year as a separate observation with a binary outcome (i.e. a Bernoulli trial), we might instead treat each pond an observation and the dependent variable is the frequency of successful years out of the total number of years. You should recognize this as classic proportion data in which the trial size is the total number of years and each year has binary outcome. If we had some independent variable (or predictor) that we wanted to use to explain or predict the frequency of success, then we would want to use a logistic regression analysis, which goes beyond the scope of this lab exercises.

What do these bionimal tests say about the success (or catastrophe) rate? Do these results agree with the Student's *t* test and/or Wilcoxon's signed rank test? Which test do think is most appropriate?

3. Two samples

The classical tests for two samples include:
- comparing two variances
- comparing two sample means or medians
- tests for paired samples
- correlating two variables
- comparing two distributions
- comparing two (or more) proportions
- testing for independence of two variables in a contingency table

3.1 Comparing two variances

Before we can carry out a test to compare two means (see below), we need to test whether the sample variances are significantly different. The simplest test is called *Fisher's F test*. It is based on the idea that if the variances of the two samples are the same, then the ratio of the variances will be 1. In order to be significantly different, the ratio will need to be significantly smaller or larger than 1, depending on whether the smaller variance is in the numerator or denominator.

Let's consider an example. The data come from a field experiment designed to assess tree seedling response to understory vegetation treatments. Specifically, four treatments (including a control) were randomly assigned to 32 plots in a randomized block design. First, read in the data:

veg<-read.csv('vegdata.csv',header=TRUE)
veg

For our immediate purposes, let's consider just the number of pine seedlings (pine) under the various treatments and ignore the block design of the experiment. Let's visualize the data using box plots, as follows:

box.plots(veg,var='pine',by='treatment')

It is apparent from the adjacent box plots that the means (and medians) differ among the treatments, but also that the variances differ as well. Let's test whether the variance in pine seedling count differs between two treatments: control (do nothing) versus clipped (continuous clipping of fern fronds), as follows:

var.test(pine~treatment,data=veg,subset=treatment %in% c('control','clipped'))

What does Fisher's *F* test say about equal variances for these two samples?

Fisher's *F* test for unequal variances assumes that the data are normally distributed. Visual inspection of the box plots indicates that this may be the case, but we might want to test for normality using one of the tests described previously, for example:

shapiro.test(veg$pine[veg$treatment=='control'])
shapiro.test(veg$pine[veg$treatment=='clipped'])

Note, because the Shapiro-Wilk test is a one-sample test, we had to select the records for each treatment and conduct separate tests. If the results indicate that the data are non-normal, then we should use a non-parametric test of homogeneity of variances, such as the *Fligner-Killeen test*, as follows:

fligner.test(pine~treatment,data=veg,subset=treatment %in% c('control','clipped'))

What does the Fligner-Killen test say about homogeneity of variances for these two samples?

Thus far, we have been concerned with comparing the variances of *two* samples. However, there are roughly equivalent tests for *k*-sample problems; i.e., when there are more than two groups. The *k*-sample parametric test is called *Bartlett's test*, which we can use to test for homogeneity of variances among all four treatment levels as follows:

bartlett.test(pine~treatment,data=veg)

What does Bartlett's test say about homogeneity of variances among treatments?

However, Bartlett's test, like Fisher's F test is highly sensitive to non-normality and the presence of outliers. The non-parametric alternative test which is largely preferred by many statisticians is called the *Fligner-Killeen test*, which we can use as follows:

fligner.test(pine~treatment,data=veg)

Do the results agree with Bartlett's test?

3.2 Comparing two sample means (or medians)

Given what we know about the variation from replicate to replicate within each sample (the within-sample variance), how likely is it that our sample means were drawn from populations with the same

mean? If it is highly likely, then we shall say that our two sample means are not significantly different. If it is rather unlikely, then we shall say that our sample means are significantly different. But perhaps a better way to proceed is to work out the probability that the two samples were indeed drawn from populations with the same mean. If this probability is very low (say, less than 5% or less than 1%), then we can be reasonably certain (95% or 99% in these two examples) that the means really are different from one another.

There are two simple tests for comparing two sample means (or medians): the Student's *t* test and Wilcoxon's rank-sum test.

The *Student's t test* is appropriate when the samples are independent, the variances constant, and the errors normally distributed, and is implemented as follows for the pine seedling data in the control and clipped treatment plots:

t.test(pine~treatment,data=veg,subset=treatment %in% c('control','clipped'), conf.int=TRUE)

Because we asked for a confidence interval (conf.int=TRUE), the output includes a 95% (by default) confidence interval on the difference between sample means. A confidence interval that includes 0 indicates that the sample means are not significantly different.

What does the two-sample *t* test say about the differences in pine seedling counts between the control and clipped treatments? Despite the rather large differences in the sample means (17.9 vs. 1.9), can we say with confidence that the sample means are different?

Of course, the validity of Student's *t* test depends on whether the assumptions have been met. In this case, we know that variances are significantly different between treatments (see above). In addition, we know that the "clipped" sample is somewhat non-normally distributed. Hence, there is good reason to be suspect of the test result.

The *Wilcoxon's rank-sum test* is appropriate when the samples are independent but the errors are not normally distributed, and is implemented as follows:

wilcox.test(pine~treatment,data=veg,subset=treatment %in% c('control','clipped'), conf.int=TRUE)

Do the results agree with the Student's *t* test? Which test result to you have more confidence in?

3.3 Tests for paired samples

Sometimes, two-sample data come from paired observations. In this case, we might expect a correlation between the two measurements, either because they were made on the same individual, or were taken from the same location. A positive covariance between the two samples reduces the variance of the difference between means, which makes it easier to detect significant differences between the means. Pairing is not always effective, because the correlation between samples may be weak. In general, however, if you can do a paired *t* test, then you should always do the paired test. It can never do any harm, and sometimes it can do a huge amount of good.

Let's take the pine seedling data. In the two-sample *t* test above, the 8 replicates in the control and 8 replicates in the clipped treatments were assumed to be independent samples, but in fact the treatments were implemented in a randomized block design. Specifically, the experimental units were grouped together into blocks representing different forest stands. It is reasonable to suspect that the stands differed somewhat in ecological conditions, perhaps in ways that we were not able to observe directly, but in ways that indirectly affect pine seedling response to the understory treatments. If this is the case, then the samples will exhibit a positive covariance and it will be to our advantage to account for this covariance by using a paired *t* test.

First, we need to create separate vectors for the "control" observations and "clipped" observations because the t.test() doesn't accept formula's (as above) for the paired option, as follows:

control<-veg$pine[veg$treatment=='control']
clipped<-veg$pine[veg$treatment=='clipped']

Note, since the experiment consists of 4 blocks, not 8, the individual samples are not really paired (i.e., one control and one clipped per block), but we will ignore this minor detail for our purposes here.

Now we can use the t.test() function as before, but with the added argument paired=TRUE, as follows:

t.test(control,clipped,paired=TRUE)

Do the results differ from the unpaired t test?

Recall that we rejected the Student's *t* test for this data set because of failure to meet the underlying assumptions of constant variance and normally distributed data. As before, we can use the Wilcoxon's rank-sum test when the samples are independent but the errors are not normally distributed, as follows:

wilcox.test(control,clipped,paired=TRUE,conf.int=TRUE)

Do the results agree with the unpaired Wilcoxon's rank-sum test? Which test result to you have more confidence in?

3.4 Correlating two variables

In the paired two-sample problem above, we suggested that a strong correlation between the two measurement variables will affect the test for significant differences between means or medians. More generally, with any two continuous variables, x and y, the question naturally arises as to whether their values are correlated with each other.

Let's take the marbled salamander dispersal data introduced in a previous lab. Recall that the data represent the dispersal of first-time breeders (ftb) from their natal ponds, representing juvenile dispersal, and the dispersal of experienced breeders (eb) from their established breeding ponds,

representing adult dispersal. The data set includes three variables: (1) dist.class = distance class, based on 100 m intervals; (2) disp.rate.ftb = standardized dispersal rate for first-time breeders, which can be interpreted as a relative dispersal probability; and (3) disp.rate.eb = standardized dispersal rate for experienced breeders, which can be interpreted as a relative dispersal probability. The question arises as to whether the dispersal rates for first-time breeders and experienced breeders are correlated.

First, we need to read in the data and check it, as follows:

disp<-read.csv('dispersal.csv',header=TRUE)
disp

Next, we can test the significance of the correlation using the cor.test() function, as follows:

cor.test(disp$disp.rate.ftb,disp$disp.rate.eb,use='complete.obs')

Not, we needed to specify the use='complete.obs' argument to address the missing values for the 700 m distance class (for which there are no ponds in this particular distance interval).

What does the correlation test say about the significance of the correlation between juvenile and adult dispersal-distance functions?

The default correlation test statistic is based on Pearson's product-moment correlation coefficient (*r*) cor(x,y) which follows a *t* distribution with length(x)-2 degrees of freedom if the samples follow independent normal distributions. If the data are non-normal, then a non-parametric rank-based measure of association is more appropriate. If method is "kendall" or "spearman", Kendall's *tau* or Spearman's *rho* statistic is used to estimate a rank-based measure of association. These tests may be used if the data do not necessarily come from a bivariate normal distribution. Let's try a test of the Spearman's rank correlation:

cor.test(disp$disp.rate.ftb,disp$disp.rate.eb,use='complete.obs',method='spearman')

What does this test say about the correlation between dispersal-distance functions? Does it agree with the parametric test? Which do you trust more with this data set?

3.5 Comparing two distributions

Another way to compare two samples, whether they be paired or not, is to compare the empirical cumulative distributions of the samples. This test is known for its wonderful name, the *Kolmogorov-Smirnov test*, which is an extremely simple test for asking one of two different questions:

- Are two sample distributions the same, or are they significantly different from one another in one or more (unspecified) ways?
- Does a particular sample distribution arise from a particular hypothesized theoretical distribution?

The two-sample problem is the one most often used and the one we will concern ourselves with here. The apparently simple question is actually very broad. It is obvious that two distributions could be different because their means were different – this was the subject of the Student's *t* test and Wilcoxon's rank sum test above. But two distributions with exactly the same mean could be significantly different if they differed in variance, or in skew or kurtosis, or both.

The Kolmogorov-Smirnov test works on *empirical cumulative distribution functions* (ecdf). Recall that these give the probability that a randomly selected value of X is less than or equal to x. Let's see what the ecdf for the sample of juvenile dispersal rate looks like:

plot(ecdf(disp$disp.rate.ftb),verticals=TRUE)

Now let's add the ecdf for the adult dispersal rate, but change the line type (lty) so that we can distinguish it from the ecdf for the juvenile dispersal rate.

plot(ecdf(disp$disp.rate.eb),verticals=TRUE,lty=3,add=TRUE)

Are these two distributions different? We can use the Kolmogorov-Smirnov test (ks.test) to determine if they differ significantly in any aspect. The test statistic is the maximum difference in value of the cumulative distribution functions; i.e., maximum vertical difference in the curves for a given value of X.

ks.test(disp$disp.rate.ftb,disp$disp.rate.eb)

Is there enough evidence to suggest that the dispersal-distance relationship differs between first-time breeders and experienced breeders?

3.6 Comparing two (or more) proportions

Suppose that only 4 female salamanders were killed crossing the road, compared to 16 males. Is this an example of strong sex-linked risk of road mortality? Before we can judge, of course, we need to know about the number of male and female candidates. It turns out that 16 males were killed out of 250 male individuals crossing the road, compared to 4 deaths out of only 40 crossings for females. Now, if anything, it looks like the females did worse than males in successfully crossing the road (10% mortality for females versus 6% for males).

The question arises as to whether the apparent positively biased mortality rate for females is statistically significant, or whether this sort of difference could arise through chance alone. This is a simple *binomial proportions test*, which we can easily do in R by specifying two vectors: the first containing the number of mortalities for females and males c(4,16) and the second containing the total number of female and male candidates c(40,250):

prop.test(c(4,16),c(40,250))

A significant *p*-value indicates that the proportions are different between samples; i.e., that the proportions observed were unlikely to have been drawn from the same underlying population.

Conversely, an insignificant *p*-value means that there is insufficient evidence to reject the null hypothesis and we would conclude that the proportions are not different (statistically).

What does the binomial test indicate? Do females suffer a higher mortality rate crossing the road or is the observed difference merely a sampling artifact? How would the result change (p-value) if the sample size was doubled but the proportions killed remained the same (i.e., 8 out of 80 females and 32 out of 500 males)?

3.7 Testing for independence of two variables in a contingency table

A great deal of data in ecology and conservation comes in the form of counts (whole numbers or integers); the number of animals that died, the number of branches on a tree, the number of days of below-freezing days, etc.. With count data, the number 0 is often the value of the response variable; in other words, there are often observations that receive a count of 0. With contingency tables, counts are cross-classified according to one or more categorical contingent variables, where the contingencies are all the events of interest that could possibly happen. A contingency table shows the counts of how many times each of the contingencies actually happened in a particular sample. In a contingency table, each observation is cross-classified according to each of the categorical contingent variables; i.e., each observation is placed into one bin representing a unique categorical level of each contingent variable.

Consider a sample of 40 forest stands where a survey determined that barred owls were either 'present' or 'absent' and where each stand was classified as either 'young' (<80 years) or 'old' (>80 years). Thus, each of the 40 observations could be cross-classified into one of the four contingencies: present-old, present-young, absent-old, and absent-young. We have the following 2x2 contingency table (with expected values shown in parentheses):

owls<-matrix(c(16,9,4,11),nrow=2)
owls

	present	absent
old	16 (12.5)	4 (7.5)
young	9 (12.5)	11 (7.5)

We would like to know whether the observed counts differ from what we would expect if presence/absence was independent of stand age. Do you remember how to calculate the expected values in a contingency table? If not, look this up. It is clear that the observed frequencies and the expected frequencies are different. But in sampling, everything always varies, so this is no surprise. The important question is whether the expected frequencies are *significantly* different from the observed frequencies.

We can assess the significance of the differences between observed and expected frequencies in a variety of ways. The usual way is with *Pearson's chi-squared test* (generalized linear models are an alternative). Do you remember how to calculate the chi-square test statistic? If not, look it up. In R

we can compute the chi-squared test as follows:

chisq.test(owls)

What does the result indicate? Are barred owls present more frequently than expected in old forest stands?

You might recall that Pearson's chi-squared test expects the expected cell values to be large, generally greater than 4 or 5. As you can seen in the table above, we have meet this rule of thumb with the barred owl data, so the Pearson's chi-squared test is appropriate. However, when one or more of the expected frequencies is less than 4 (or 5), then it is wrong to use Pearson's chi-squared test for your contingency table. This is because small expected values inflate the value of the test statistic, and it can no longer be assumed to follow the chi-square distribution. In this case, an alternative test called *Fisher's exact test* is more appropriate. We use the function in the same way as before:

fisher.test(owls)

Does the conclusion change?

Note, the procedures described in this section can all be used with contingency tables much bigger than 2x2.

4. Exercise

The purpose of the following exercise is to give you additional experience working with classical statistical tests. You can work individually or in teams to complete the exercise.

The first data set represents tree seedling counts of two species, black birch (*Betula nigra*) and eastern white pine (*Pinus strobus*) following several understory treatments in western Massachusetts. For the purpose of this exercise, we are interested in knowing whether there is a difference in the mean or median between tree species following the "clipping" treatment in which all fern fronds were clipped to ground level on a regular basis to remove competition for light.

1. Read in the raw data and create a subset containing only the records corresponding to the 'clipped' treatment, as follows:

    ```
    veg<-read.csv('vegdata.csv',header=TRUE)
    clipped<-veg[veg$treatment=='clipped',]
    clipped
    ```

2. Is there a significant difference between the mean or median seedling counts for the two tree species under the clipping treatment? In other words, what is the probability that our sample means or medians were drawn from the same population? Be sure to defend your choice of test statistic.

In order to answer the above question, you need to determine the correct test to use to compare the two groups. One of the potential tests involves the assumption of not only normally distributed values (or residuals) in each group, but also homogeneity of variances between groups. To use the appropriate tests for homogeneity of variances (see previous lab work), the data set needs to be in the "long" format rather than the "wide" format. Specifically, the groups (species) cannot be in separate columns as in the "clipped" dataframe, but instead need to be in a single column. Use the following reshape() function to convert from the "wide" to "long" format (be sure to look at the result to see what happened):

```
clipped.long<-reshape(clipped,varying=list(5:6),v.names='count',
  timevar='species',times=c('birch','pine'),direction='long')
clipped.long$species<-as.factor(clipped.long$species)
```

The second data set represents counts of invasive earthworms and counts of middens of the largest species, *Lumbricus terrestris* (the common large nightcrawler) found in 98 sample plots randomly distributed across the Deerfield watershed in western Massachusetts in mature upland forested stands. Counting earthworms can be quite labor intensive, but counting surface middens of *L. terrestris* is relatively quick and easy, so it would be useful to know whether the counts of middens can be used as an index of the number of earthworms.

3. Examine the data set (worms.csv). Plot the relationship between midden counts (middens.count) and earthworm counts (worms.count). Hint, start with a histogram of each variable and then follow this with a sunflower plot (see ?sunflower for interpretation):

```
hist(worms$middens.count)
hist(worms$worms.count)
sunflowerplot(worms$middens.count,worms$worms.count)
```

4. What is the probability that the two variables come from a normal distribution?

5. What is the correlation between middens.count and worms.count? Is it significant? What method did you use and why?

The final data set represents standardized breeding bird counts (bird) and a variety of habitat variables (hab) across 1046 sample plots in the Oregon Coast Range. See birdhab.meta.pdf for a complete description of the data sets. Here, we are interested in knowing whether the presence/absence of brown creepers varies between the interior and edge of forest stands.

6. Read in the raw data and create a 2x2 contingency table consisting of counts of brown creeper presence (1) versus (0) absence and forest edge (E) versus interior (I):

```
birds<-read.csv('bird.sta.csv',header=TRUE)
hab<-read.csv('hab.sta.csv',header=TRUE)
birdhab<-merge(birds,hab,by=c('basin','sub','sta'))
bin<-data.trans(birdhab,method='power',var='AMCR:YRWA',exp=0,plot=FALSE)
temp1<-table(bin$s.edge,bin$BRCR)
```

temp2<-temp1[,c(2,1)]
temp2

Make sure you understand the script above. First, we read in the bird and habitat data and merged them into a single file based on the common fields. Then we did a binary transformation of the bird counts using the data.trans() in the biostats library. Then we used the table() function to compute the cross-classified counts. The next step simply switched the order of the columns so that the present counts were in the first column and the absent counts were in the second column, as this is the expected order in some functions (e.g, prop.test()).

7. Are brown creepers present more or less frequently than expected in forest interiors versus forest edges and is the difference significant? An alternative way of asking this is: Are brown creepers present in a greater proportion in forest interiors or forest edges? Note, there are different ways to frame this problem and test this hypothesis, but the results will probably be the same.