Homework assignment # 11

Note: Please circle your answers when appropriate!

1) Come up with a 2 x 2 contingency table for which the row and column totals are not the same and $\chi^{2^*} = 0$. Now calculate $\hat{p}_1$ and $\hat{p}_2$. Are you surprised? ***Why or why not?***
(Hint and comment: if you have trouble, reverse the problem - come up with a table for which $\hat{p}_1$ and $\hat{p}_2$ are equal, and then calculate χ^{2^*}).

2) You want to find out if there is a food preference between two species of bats. You collect the following data. You suspect ***species B prefers Beetles***:

	Moths	Beetles
Bat species A	13	9
Bat species B	11	23

Does species B prefer beetles? Make sure you clearly state H_0, H_1, etc. Use $\alpha = 0.05$. Also clearly write out your decision and what it means.

3) You want to examine competition in barnacles. You examine two species and count the number of each species at the low and high water marks (many barnacles live in tidal areas). You get the following results:

	Low water	High water
Species A	22	7
Species B	9	18

(a) Are the two species of barnacle independent of water level? Make sure you clearly state H_0, H_1, etc. Use $\alpha = 0.05$. Also clearly write out your decision and what it means.

(b) Which species prefers low water? Which prefers high water?
(You will need to calculate some proportions to answer this question).

4) There shouldn't be a blood group difference between men and women as blood group type is not dependent on sex. However, you want to determine if there is any evidence to doubt this. You go out and collect the following data on blood types:

		Sex	
		Male	Female
Blood type	A	72	85
	B	15	21
	AB	9	14
	O	76	87

Can you find a significant difference in blood types based on sex? Make sure you clearly state H_0, H_1, etc. Also clearly write out your decision and what it means.

5) Here are some data loosely based on tree counts collected in ecology (BIOL 308).

Is there a difference in tree composition based on landscape position? Make sure you clearly state H_0, H_1, etc. Also clearly write out your decision and what it means.

You can either do this by hand, or if you prefer, use R (it'll be *much* quicker!). See the R instructions at the end.

	Species					
Landscape Position	**Oaks**	**Maple**	**Beech**	**Pine**	**Tulip Poplar**	**Other**
Ridge Top	70	15	70	20	60	35
Mid-Slope	80	20	120	5	20	55
Valley Bottom	15	5	85	5	10	30
Disturbed area	10	12	45	40	20	32

R instructions:

There are actually several ways of doing a contingency table in R, but assuming you are given data that is already arranged into a table (like problem 5 above), here's how to proceed (we'll use an old text example looking at blood types and ulcers):

We want to find out if the proportion of blood types is different between ulcer patients and controls:

Blood type	Ulcer patients	Controls	TOTAL
O	911	4578	5489
A	579	4219	4798
B	124	890	1014
AB	41	313	354
TOTAL	1655	10000	11655

Step 1: read the data into R

```
patients <- scan(nlines = 1)
911 4578 579 4219 124 890 41 313
```

Notice that we do NOT input the totals, just the table. Also note the order (by row) in which we enter the values.

Step 2: tell R to arrange this into a matrix (table):

```
patients <- matrix(patients,nrow = 4, byrow = TRUE)
```

Okay, now we have to explain a few things...

`matrix` tells R to convert the data into a matrix. We can think of this as a table for now.

(If you've had matrix or linear algebra, then you should know that this really is a matrix, and you can do matrix operations on it!)

`nrow = 4` tells R how many rows we have (we have four rows and two columns).

`byrow = TRUE` tells R to arrange the data by row. In other words, 911, then 4578.

(If you read in 911 followed by 579, that would be by column).

Step 3: look at your matrix to make sure it looks right (optional):

```
patients
```

This should return the following:

```
     [,1] [,2]
[1,]  911 4578
[2,]  579 4219
[3,]  124  890
[4,]   41  313
```

Step 4: do your chi-square test:

```
chisq.test(patients, correct = FALSE)
```

This part is mostly self explanatory. Note that R gives you the χ^{2*} value as well as the *p*-value.

> The `correct = FALSE` part is needed to make sure the results from R match what we've learned. R automatically applies a correction we didn't talk about to 2 x 2 tables, so it's just a good habit to get into (even for bigger tables).

Step 5: if you wanted to get the expected values, you could do:

```
chisq.test(patients)$expected
```

(You have to run this separately - you could combine steps steps 4 and 5, but it's easier this way).

Due in recitation the week of November 18th.