

Homework assignment # 11

Note: Please circle your answers when appropriate!

Remember: use R only if the problem states you can use R.

1) You have the following made up data:

		Factor A	
		Level 1	Level2
Factor B	Level 1	14	26
	Level 2	7	13

- Perform a χ^2 contingency table test (two sided is fine). Give $\hat{p}_1$, $\hat{p}_2$, and χ^{2*} .
- Now multiply the first row by 2 (so you have 28 and 52) and repeat (a). Are you surprised?
- Now use the original data given above and multiply the first column by 4 (so you have 56 and 28) and repeat (a). Are you surprised?

2) The non-venomous scarlet king snake (*Lampropeltis elapsoides*) is thought to mimic the venomous coral snake (*Micrurus fulvius*). You want to find out if the scarlet king snake is more or less common if the coral snake is present. You examine 78 plots in a wooded area in Georgia with the following results:

		Coral snake	
		present	absent
King snake	present	26	5
	absent	9	38

- is there a significant relationship between King snakes and Coral snakes?
- if there is, are there more King snakes when Coral snakes are present or absent?

3) You want to examine the relationship between two species of salamander. You suspect species A prefers moister environments. You find the following numbers of salamanders:

	Moist	Dry
Species A	15	7
Species B	8	17

Does species A prefer moister environments? Make sure you clearly state H_0 , H_1 , etc. Use $\alpha = 0.05$. Also clearly define $\hat{p}_1$, $\hat{p}_2$, and write out your decision and what it means.

4) You may have heard that people with blood type O are just a little less susceptible to Covid. You record the blood types of 83 people who do not have Covid, and then the blood types of 93 people who have Covid.

Presence of Covid			
Blood type	No Covid		Covid
A	33		54
B	9		11
AB	5		8
O	37		19

Can you find a significant difference in blood types based on Covid? Make sure you clearly state H_0 , H_1 , etc. Also clearly write out your decision and what it means. Use $\alpha = 0.01$.

(We haven't learned how to analyze $r \times k$ table in more detail, but if you examine the proportion of people with blood type O for No Covid vs. Covid, that should give you a clue).

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.

Landscape Position	Species					
	Oaks	Maple	Beech	Pine	Tulip Poplar	Other
Ridge Top	75	19	68	18	57	33
Mid-Slope	81	36	123	6	14	51
Valley Bottom	12	7	81	3	11	30
Disturbed area	9	12	31	41	16	22

Due in recitation Wednesday, July 8th.

Remember: Copying answers from other students, web sites, previous classes or using AI is a violation of the honor code!!

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).