# Homework #7

Due: Tuesday, November 16 @ 6pm

**Please remember to give R code, as well as answers, for any problems where you used R**

**Problem 1:**

A lake contains 600 fish, 80 of which have been tagged by scientists. Suppose a researcher randomly catches 15 fish from the lake: 3 are tagged and 12 are not tagged.

a. Is this a binomial random variable? Why or why not?
b. What is the probability of catching at least 3 tagged fish out of a sample of 15? (*In other words, is there an enrichment of tagged fish in the researcher's sample?*). **Solve this problem by hand (not using R functions like `dbinom()` or `dhyper()`). You may use R to help with the calculation in other ways.**
c. Now, using the `*binom()` and/or `*hyper()` functions in R, repeat (b). How does your answer compare?
d. If this problem is binomial, how could you re-write the problem to be a hypergeometric distribution? Alternatively, if this problem is hypergeometric, how could you re-write the problem to be binomial?
e. Solve this new problem from (d) using any method. How does your answer compare to (b) and (c)?

**Problem 2:**

Consider a fictitious population of mice in which each animal's coat is either black (B) or gray (G) in color and is either wavy (W) or smooth (S) is texture. Suppose a random sample of mice is selected and the coat color and texture are observed. Consider the accompanying contingency table for the data.

| | | Color | | |
|---|---|---|---|---|
| | | **B** | **G** | **Total** |
| **Texture** | **W** | 40 | 50 | **90** |
| | **S** | 20 | 100 | **120** |
| | **Total** | **60** | **150** | **210** |

a. Express the following conditional probabilities as numbers: (i) $P(B|W)$, (ii) $P(S|G)$ (iii) Smooth coats that are black, (iv) Black coats that are smooth

b. Using a hand-held calculator or computer (not R functions), calculate the expected counts for each cell under the null hypothesis that coat color is independent from texture.

c. Using a hand-held calculator or computer (not R functions), calculate the $\chi^2$ statistic to test that hypothesis.

d. Using your answer to (c) and the `*chisq()` suite of functions (*i.e.* `pchisq()`, `dchisq()`, `rchisq()`, or `qchisq()`), test your hypothesis that coat color is independent from texture. *Give your code, the p-value, and a written interpretation of your results*

e. What assumptions did you make to perform the chi-square test for independence?

f. Read the help page for the `chisq.test()` function in R and apply it to the data above. How does it compare to your answer in (d)?

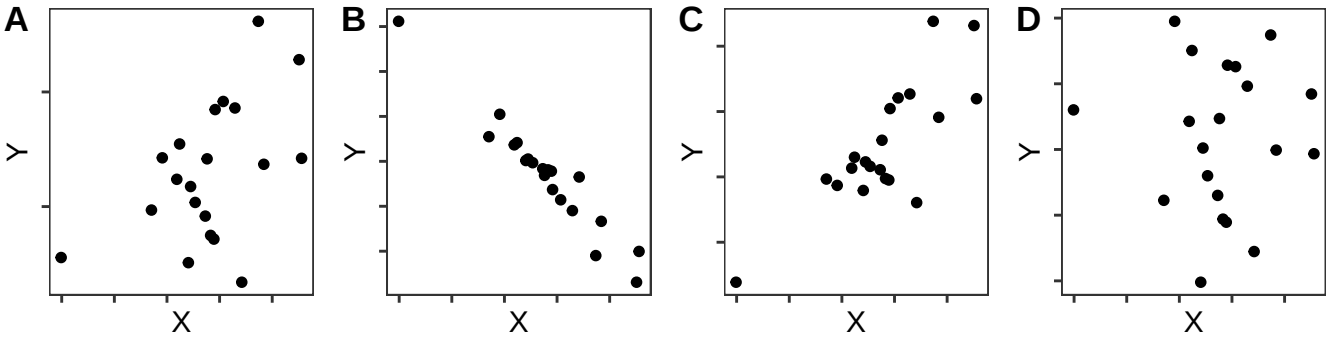
g. The Fisher's Exact Test is more accurate (and more computationally expensive) than the $\chi^2$-test for count data. Read the help page for the `fisher.test()` function and apply it here. How does it compare to your answers from (c) and (d)?

**Problem 3:**

Prior to an influenza season subjects were randomly assigned to receive either a flu vaccine or a placebo. During that season there were 28 cases of the flu among 813 vaccine recipients and 35 cases of the flu among the 325 subjects who were given the placebo.

a. Calculate the relative risk (conditional probability) of getting the flu for individuals who received the placebo versus those who received the vaccine. Write one sentence explaining this value.
b. Calculate the odds ratio for comparing flu cases among individuals who received the placebo to flue cases among individuals who received the vaccine. Write one sentence explaining this value.
c. The output from the `fisher.test()` in R also gives an odds ratio. Perform this test for this data. How does it compare to the answer from (b)?
d. Does the odds ratio give a good approximation to the relative risk for these data? Why or why not?

**Problem 4:**




a. Arrange the plots in order of their correlations (from closest to -1 to closest to +1)
b. Arrange the plots in order of their corresponding $P$-values (smallest to largest) for the test $H_0 : \rho = 0$. *Note: all of the plots display the same number of observations*

**Problem 5:**

Laetisaric acid is a compound that holds promise for control of fungus diseases in crop plants. The accompanying data show the results of growing the fungus *Pythium ultimum* in various concentrations of laetisaric acid. Each growth value is the average of four radial measurements of a *P. ultimum* colony grown in a petri dish for 24 hours; there were two petri dishes at each concentration.

| Laetisaric acid concentration (uG/mL) | 0 | 0 | 3 | 3 | 6 | 6 | 10 | 10 | 20 | 20 | 30 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fungus growth (mm) | 33.3 | 31 | 29.8 | 27.8 | 28 | 29 | 25.5 | 23.8 | 18.3 | 15.5 | 11.7 | 10 |

a. Plot the data in a way that you can visualize the relationship between laetisaric acid concentration and fungus growth. By eye, does there seem to be a linear relationship?
b. Calculate the correlation coefficient using the R function `cor()`. Give a one sentence interpretation of this value.
c. Perform a test to see whether the population correlation is zero. You may choose to use the `cor.test()` function in R. *Give your code, the p-value, and a written interpretation of your results*
d. Is this study an observational study or an experiment?
e. It is suggested that acid could be used to impede fungus growth. Could these data be used to verify this claim? If not, what could be said? Briefly explain.