Assignment 4

1. In animals without paternal care, the number of offspring sired by a male increases as the number of females he mates with increases. This fact has driven the evolution of multiple matings in the males of many species. It is less obvious why females mate multiple times, because it would seem that the number of offspring that a female has would be limited by her resources and not by the number of her mates, as long as she has at least one mate. To look for advantages of multiple mating, a study of the Gunnison's prairie dog followed females to find out how many times they mated (Hoogland 1998). They then followed the same females to discover whether or not they had given birth later. The results are compiled in the following table:

Number of times female mated:	**1**	**2**	**3**	**4**	**5**
Number who gave birth:	81	85	61	17	5
Number who didn't give birth:	6	8	0	0	0

Did the number of times that a female mated affect her probability of giving birth?
~~a. Calculate expected frequencies for a contingency test.~~
~~b. Examine the expected frequencies. Do they meet the assumptions of the χ^2 contingency test? If not, what steps could you take to meet the assumptions and make a test?~~
~~c. An appropriate test shows that the number~~ of mates of the female prairie dogs is associated ~~with giving birth. Does this mean that mating with more males increases the probability of giving birth? Can you think of an alternative explanation?~~

R exercise
~~d. Load the data *PrairieDogMultipleMating.csv* and Perform the χ^2 contingency test suggested in b. Check the warning message R gives.~~
~~fe Assuming you can't do the analysis think of a way you can change to data to by able to do a Fisher exact test (*fisher.test*()). Does provide the same biological insights as d?~~
~~f. Advanced: Make the table of the expected frequencies using R.~~
~~A smart way to get the sum of each row and column is to use *apply*(a, b, c). a takes the data where the apply should work with, b is if it should apply the function (defined in c) on each column or row. Example: *apply*(table.with.data, 1, sum) takes the sum of each row table.with.data. Get the *sum*() of each new vector and turn them into probabilities by dividing each element by the sum. This you can use to calculate the predicted frequencies.~~

2. Daycare centers expose children to a wider variety of germs than the children would be exposed to if they stayed at home more often. This has the obvious downside of frequent colds, but it also serves to challenge the immune system of children at a critical stage in their development. A study by Gilham et al. (2005) tested whether social activity outside the house in young children affected their probability of later developing the white blood cell disease acute lymphoblastic leukemia (ALL), the most common cancer of children. They compared 1272 children with ALL to 6238 children without ALL. Of the ALL kids, 1020 had significant social activity outside the home (including daycare) when young. Of the kids without ALL, 5343 had significant social activity outside the home. The rest of both groups lacked regular contact with children who were not in their immediate families.

~~a. Is this an experimental or observational study?~~

~~b. What are the proportions of children with significant social activity in children with and without ALL?~~

~~c. What is the odds ratio for ALL, comparing the groups with and without significant social activity?~~

~~d. What is the 95% confidence interval for this odds ratio?~~

~~e. Does this confidence interval indicate that amount of social activity is associated with ALL? If so, did the children with more social activity have a higher or lower occurrence of having ALL?~~

~~f. The researchers interpreted their study in terms of the differing immune system exposure of the children, but gave several alternative explanations for the pattern. Can you think of any possible confounding variables?~~

3. It is common wisdom that death of a spouse can lead to health deterioration of the partner left behind. Is common wisdom right or wrong in this case? To investigate, Maddison and Viola (1968) measured the degree of health deterioration of 132 widows in the Boston area, all of whose husbands had died at the age of 45-60 within a fixed six-month period before the study. A total of 28 of the 132 widows had experienced a marked deterioration in health, 47 had seen a moderate deterioration and 57 had seen no deterioration in health. Of 98 control women with similar characteristics who had not lost their husbands, 7 saw a marked deterioration in health over the same time period, 31 experienced a moderate deterioration of health, and 60 saw no deterioration. Test whether the pattern of health deterioration was different between the two groups of women. Give the *P*-values as precisely as possible from the statistical tables, and interpret your result in words.

R exercise

~~a. Load the data *WidowHealth.csv* into R and turn it into a table using *table*(). Perform a χ^2 contingency test and check your results above.~~

b. Does it matter what you put on the rows or the columns for the test? Check if it does by transforming the table you made by turning rows into columns and vice versa. Try first to figure out how this can be done and if you struggle check hint[3] at the end of the page.

4. In Europe 53% of flowers of the rewardless orchid, *Dactylorhiza sambucina*, are yellow, whereas the remaining flowers are purple (Gigord et al. 2001). For the following, you may use the normal approximation only if it is appropriate to do so.

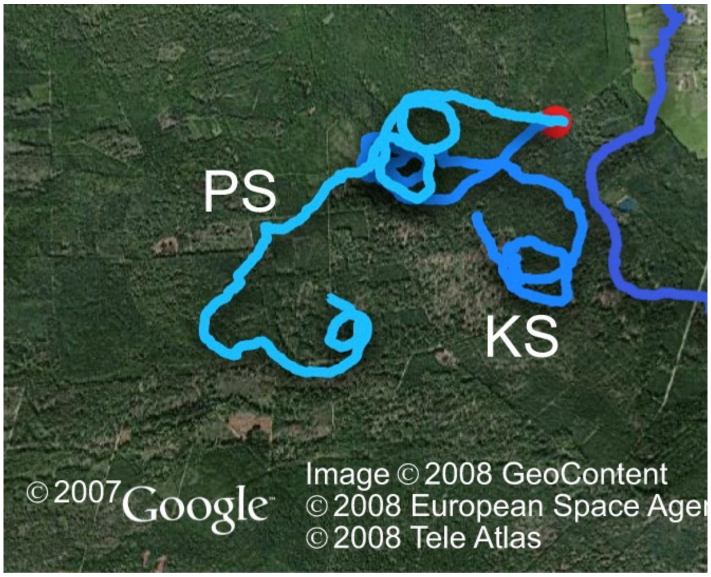
~~a. If we took a random sample of a single individual from this population, what is the probability that it would be purple?~~
~~b. If we took a random sample of 5 individuals, what is the probability that at least 3 are yellow?~~
~~c. If we took many samples of $n = 5$ individuals, what is the expected standard deviation of the sampling distribution for the proportion of yellow flowers?~~
~~d. If we took a random sample of 263 individuals, what is the probability that no more than 150 are yellow?~~

~~**5**. The proportion of traffic fatalities for each U.S. state resulting from drivers with high alcohol blood levels in 1982 was approximately normally distributed, with mean 0.569 and standard deviation 0.068 (U.S. Department of Transportation Traffic Safety Facts 1999).~~
~~a. What proportion of states would you expect to have more than 65% of their traffic fatalities from drunk driving?~~
~~b. What proportion of deaths due to drunk driving would you expect to be at the 25th percentile of this distribution?~~

~~**6**. Two different researchers measured the weight of two separate samples of ruby-throated hummingbirds from the same population. Each calculated a 95% confidence interval for the mean weight of these birds. Researcher 1 found the 95% confidence interval to be 3.12 g < μ < 3.48 g, while Researcher 2 found the 95% confidence interval to be 3.05 g < μ < 3.62 g.~~
~~a. Why would the two researchers get different answers?~~
~~b. Which researcher most likely had the larger sample?~~
~~c. Can you be certain about your answer in part (b)? Why or why not?~~

7. Without external cues such as the sun, people attempting to walk in a straight line tend to walk in circles (the accompanying image shows the paths of two participants, PS and KS, attempting to walk in a straight line in an unfamiliar forest on a cloudy day). One idea is that most individuals have a tendency to turn in one direction because of internal physiological asymmetries, or because of differences between legs in length or strength. Souman et al. (2009) tested for a directional tendency by blindfolding 15 participants in a large field and asking them to walk in a straight line. The numbers below are the median change in direction (turning angle) of each of the 15 participants measured in degrees per second. A negative angle refers to a left turn, whereas a positive number indicates a right turn.

-5.19, -1.20, -0.50, -0.33, -0.15, -0.15, -0.15,
-0.07, 0.02, 0.02, 0.28, 0.37, 0.45, 1.76, 2.80



~~a. Draw a graph showing the frequency distribution of the data. Is a trend in the mean angle suggested?~~
~~b. Do people tend to turn in one direction (e.g., left) more on average than the other direction (e.g., right)? Test whether the mean angle differs from zero.~~
~~c. Based on your results in (b) is the following statement justified: “People do not have a tendency to turn more in one direction, on average, than the other direction”? Explain.~~

R exercise
~~d. Load the data *WalkingInCircles.csv* into R and make a histogram. Change the bin width to check if this changes your interpretation of the data.~~
~~e. Calculate the mean, mode, median, standard deviation and standard error of the data. Try no to look back at earlier code and remember it (or try out). The functions in R quite often have descriptive names. Note the mode is trickier!~~

Hints:

[3] Use *t*().