Due: midnight, Oct 4 (Thursday labs); midnight, Oct 5 (Friday labs)

Lab 6: χ^2 test of goodness of fit

Objectives

1) Use χ^2 tests to assess a probability distribution's goodness of fit to data.
2) Use the R functions `chisq.test` and `pchisq`.

Exercise

1. A study of injuries to cats falling from high-rise buildings in Manhattan noted the month in which 119 randomly sampled falls occurred. The data are in file "cats.csv".
   a. Read in these data and make a bar plot of the results.
   b. Carry out a test of the hypothesis that falls are randomly distributed across months of the year. Explicitly perform the five basic steps of hypothesis testing. Clearly state the conclusion of the test. Do this test "by hand" in R.
   c. Repeat the test using R's `chisq.test` function.
2. In snapdragons, variation in flower color is determined by a single gene. RR individuals are red, Rr (heterozygous) individuals are pink, and rr individuals are white. In a cross between heterozygous individuals, the expected ratio of red:pink:white offspring is 1:2:1.
   a. The results of such a cross were 10 red-, 21 pink-, and 9 white-flowered offspring. Make a bar plot that shows both these observed values and the corresponding expected values.
   b. Test the hypothesis that the true ratio is 1:2:1. Do this test "by hand" in R. You do not need to explicitly perform the five basic steps of hypothesis testing, but you do need to clearly state the conclusion of the test.
   c. Repeat the test using R's `chisq.test` function.
   d. In another, larger experiment, you count 100 times as many flowers as in the experiment above, and get 1000 red, 2100 pink, and 900 white offspring. Repeat the hypothesis test with these data, using `chisq.test`.
   e. Do the proportions observed in the two experiments differ? Did the results of the two hypothesis tests differ? Briefly explain why or why not.
3. When they reproduce, female nine-banded armadillos always give birth to quadruplets, which they raise in an underground burrow for two to three weeks. Researchers examined all the offspring in a random sample of armadillo burrows. For each armadillo pup, they determined whether it was infested with a mite that specializes on armadillos. The following table lists the numbers of parasitized pups observed in 96 burrows, each of which contained exactly four pups:

Number of parasitized armadillo pups	**Number of burrows**
0	12
1	25
2	23
3	19
4	17

a. Estimate the proportion of parasitized pups in this population of armadillos.

b. Consider the hypothesis that the number of parasitized pups in each burrow follows a binomial distribution. Calculate how many burrows of each type are expected if this hypothesis is true.

c. Make a bar plot that shows both the observed values and the corresponding expected values that you just calculated.

d. Use the expected values you calculated to test the hypothesis that the number of parasitized pups follows a binomial distribution. Carry out an appropriate hypothesis test "by hand" in R. You do not need to explicitly perform the five basic steps of hypothesis testing, but you do need to clearly state the conclusion of the test.

e. If the number of parasitized pups is not binomially distributed, suggest a possible explanation (i.e., what assumption of the binomial distribution is violated and why)?

Assignment: Turn in a single Microsoft Word document with a heading that includes your name, the lab number, and your section (day and time). Create a subheading for each exercise. For each exercise, include (1) your annotated R script, (2) your annotated output, and (3) your answer to any specific questions in each exercise. In addition to the word document, you must also submit a separate .R file with all of your R scripts.