Lab 5: Binomial confidence intervals and hypothesis tests

Objectives

1) Calculate and interpret confidence intervals for the probability parameter of a binomial random variable.

2) Test hypotheses about the probability parameter of a binomial random variable, both by hand and using the R command binom.test.

Exercises

1. *Binomial distribution review:* A geneticist is planning to grow 24 pea plants from a cross between two plants. Each plant has a probability of ¼ of having wrinkled peas, and a probability of ¾ of having smooth peas. Use R's binomial probability distribution functions to calculate the following values:

   a. The probability that exactly 8 plants will have wrinkled peas.

   b. The probability that 8 or fewer plants will have wrinkled peas.

   c. The probability that 12 or more plants will have wrinkled peas.

   d. The 0.025 quantile of the number of plants with wrinkled peas.

2. Under some circumstances, sex ratio theory predicts that ant colonies should produce reproductive offspring in a 1:1 ratio (i.e., equal numbers of male and female offspring). To test whether this prediction holds up in the ant *Formica horribilis*, you gather a random sample of 25 reproductive offspring and find 15 females and 10 males.

   a. Use the data to estimate the probability p that a reproductive offspring is female.

   b. Use the Agresti-Coull method to calculate the 95% confidence interval for p.

   c. Based on the confidence interval you calculated, briefly explain whether it is plausible that the true sex ratio is 1:1 (that is, $p = 0.5$).

   d. Now carry out a formal test of the hypothesis that the sex ratio is 1:1. Perform this test by hand in R, explicitly carrying out the five steps of hypothesis testing: 1) State the null and alternative hypotheses. 2) State the significance level. 3) Define the test statistic. 4) Calculate the test statistic from the data. 5) Calculate the P-value and use it to make a decision about the hypothesis under test.

   e. Perform the same test using the R command binom.test. Show the complete output of the test.

f. Explain whether the result of your hypothesis test consistent with the confidence interval that you calculated in part b).

3. In a blind taste test, do people prefer dog food or human food? To find out, scientists presented 18 people with unlabeled samples of dog food and four foods meant for humans: duck liver mousse, pork liver paté, liverwurst, and Spam. Each participant was asked to choose their favorite among the five items. Two of the 18 participants chose the dog food as their favorite, whereas the other 16 chose one of the other items.

   a. What proportion of participants chose dog food as their favorite? What proportion are expected to choose dog food, assuming no preference among the five items (i.e., they choose a favorite randomly)?

   b. Use binom.test to test the hypothesis that people have a preference between human food and dog food. Show the complete output of the test. Also include a complete statement of your conclusions, along the following lines: "We [*do or do not*] reject the null hypothesis that [*state hypothesis here*], at a significance level of [*state significance level here*] (*Name of test* : *p value, sample size*)."

   c. binom.test reports a confidence interval. Briefly describe what this interval means, and state whether it is consistent with the results of your hypothesis test.

4. In a certain fly species, 1/4 of genes are found on the X chromosome. A theory predicts that genes involved in spermatogenesis should occur disproportionately often on the X chromosome. A study observes that 11 of 24 spermatogenesis genes are found on the X chromosome.

   a. Use these data and binom.test to test the theory that spermatogenesis genes are found disproportionately on the X chromosome. Report your results as described for exercise 3b).

   b. Briefly explain whether the confidence interval reported by binom.test is consistent with the results of your hypothesis test.

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.