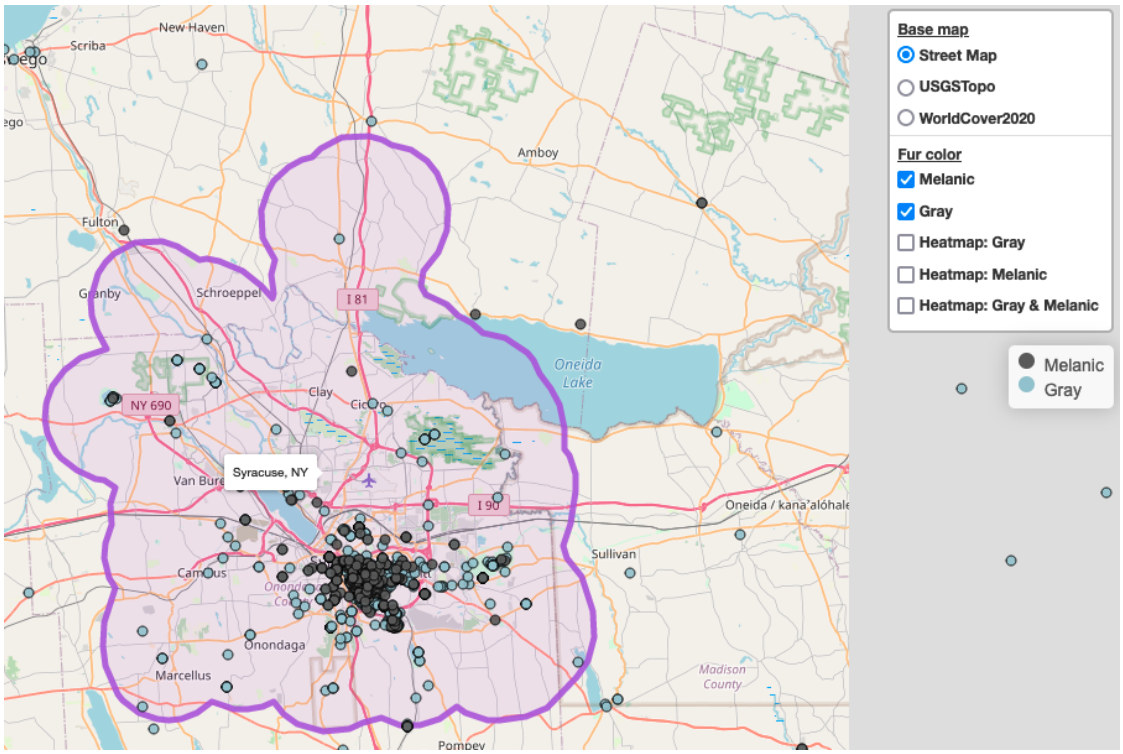
Chi-Squared Analysis Tutorial

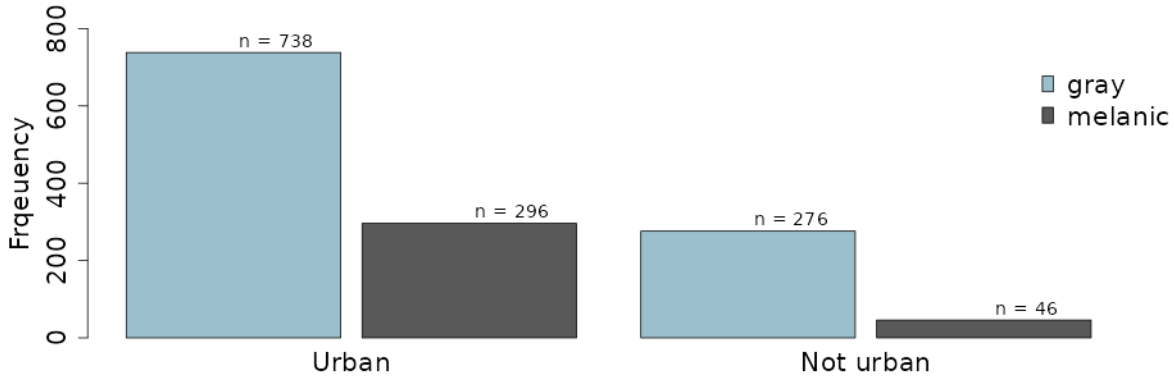
Instructions to compare urban-rural morph proportions with a Chi-squared contingency analysis with a single sample

Getting the data

1) Navigate to SquirrelMapper
2) Select an area. For this tutorial I will select Syracuse, NY under “Choose a Story”.



3) Scroll to the bottom of the page and select “Morph totals by urban land use” to see the breakdown of melanic and gray squirrels by urban and rural land use categories. The sample sizes are reported on the top of the bars on the chart:



4) Make a contingency table of the "observed" data in a spreadsheet. I'll use this example spreadsheet.

Observed	Urban	Rural
Melanic	296	46
Gray	738	276

Conducting the contingency analysis

1) State the statistical hypotheses.
    a) Null hypothesis (H_0): Morph proportions are independent of land use (urban vs. rural).
    b) Alternative hypothesis (H_A): Morph proportions differ between urban and rural land use.

2) Use the observed data in the contingency table to quantify expected proportions of each coat color in urban and rural environments.
    a) Calculate the proportion of melanic and gray squirrels overall:
        i) Melanic: (296 + 46) / (296 + 46 + 738 + 276) = 0.25
        ii) Gray: (738 + 276) / (296 + 46 + 738 + 276) = 0.75
    b) Calculate the proportion of urban and rural squirrels overall:
        i) Urban: (296 + 738) / (296 + 46 + 738 + 276) = 0.76
        ii) Rural: (46 + 276) / (296 + 46 + 738 + 276) = 0.24
    c) Use the multiplicative probability rule to quantify the expected probabilities of each coat color in urban and rural environments. The key here is that the multiplicative rule assumes that coat color and land use are independent, so these probabilities provide the expected data for the null hypothesis.
        i) P(melanic and urban) = 0.25 * 0.76 = 0.19
        ii) P(gray and urban) = 0.75 * 0.76 = 0.57
        iii) P(melanic and rural) = 0.25 * 0.24 = 0.06

iv) P(gray and rural) = 0.75 * 0.24 = 0.18

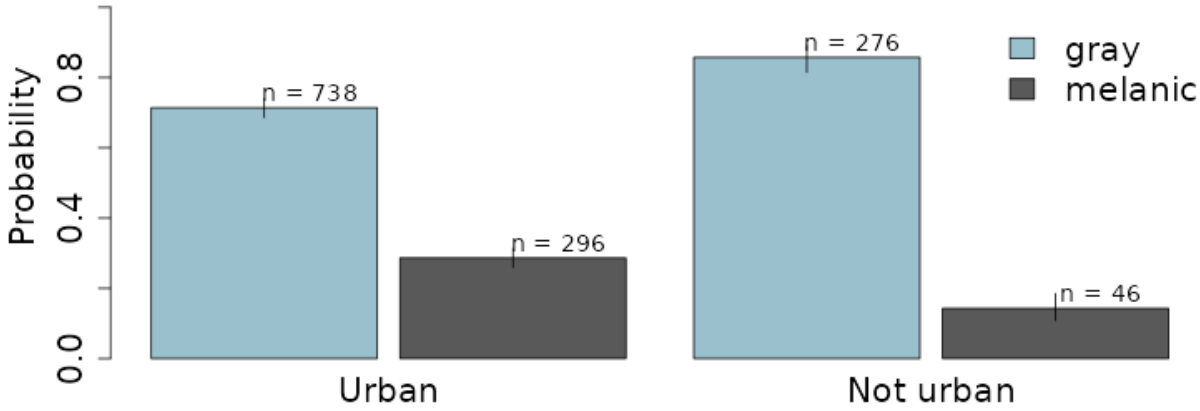
3) Chi-squared analyses are done on counts of individuals, not proportions (i.e., probabilities, so you have to convert your expected probabilities to expected counts. Simply multiple each probability by the total sample size of squirrels (296 + 46 + 738 + 276 = 1356). The values will have decimals. That's OK because these are the expectations on average.
    a) Expected melanic and urban = 0.19 * 1356 = 260.8
    b) Expected gray and urban = 0.57 * 1356 = 773.2
    c) Expected melanic and rural = 0.06 * 1356 = 81.2
    d) Expected gray and rural = 0.18 * 1356 = 240.8

4) Calculate the Chi-squared test statistic and P-value.

$$\chi^2 = \sum_i \frac{(Observed_i - Expected_i)^2}{Expected_i}$$

    a) X^2 = (296 - 260.8)2/260.8 + (738-773.2)2/773.2 + (46-81.2)2/81.2 + (276-240.8)2/240.8 = 26.78.
    b) The degrees of freedom for a contingency analysis with two categories for both variables is 1.
    c) The P-value is quantified with spreadsheet software using the =CHISQ.DIST.RT function. In this case the P-value is 0.0000002, far below the typical significance value of 0.05.

5) Interpret
    a) If the P-value is lower than the significance value of 0.05, we usually reject the null hypothesis. In this case the P-value is VERY low (P<0.001), so we conclude that morph proportions do vary between urban and rural land use categories.
    b) Conclusion statement: The probability of melanic squirrels was greater in urban than rural land use (X^2 = 26.8, df = 1, P < 0.001).
    c) The figure reported on SquirrelMapper shows the observed probabilities of each morph by land use category, supporting the analysis here.
    d) Note that the SquirrelMapper app reports a Fisher's Exact Test, which is a slightly different type of test but is used for the same purpose of comparing morph proportions between land use categories. The conclusion from the contingency analysis and exact test will usually be the same.


Probability
0.0
0.4
0.8
n = 738
n = 296
n = 276
n = 46
gray
melanic
Urban
Not urban