

Homework Assignment 4 (Due Monday, March 1, 2010)

The final project will be a critique of a paper, in particular with regards to the statistical approaches used, and the conclusions derived. I will hand out specific details, but need to know how many people plan to hand in the final project. Please send me an email (ingo@jhu.edu) if you plan on doing the final. Thanks!

1. In a recent study, 12 out of 118 patients with autism had de novo gene copy number variations (dn-CNVs), while only 2 out of 196 normal controls had dn-CNVs. We are interested in testing whether autism is associated with dn-CNVs.
 - (a) Arrange your data in a 2×2 table, and calculate the expected numbers under the null hypothesis of no association between autism and dn-CNVs.
 - (b) Calculate the test statistic for the chi-square test.
 - (c) In testing your null hypothesis of no association between autism and dn-CNVs, are your results non-significant, border-line significant, or highly significant? Explain.

2. Consider the data in the table below:

1	2	3	4	5
17	19	22	25	17

Do these data look like they follow a multinomial distribution with $n=100$ and $p_i = \frac{1}{5}$ for all digits i ? In other words, do the outcomes 1–5 look equally likely? Use a χ^2 and a likelihood ratio test to answer that question.

3. We are interested in estimating the concentration of a biomarker on the basis of measurements of a number of technical replicates. Suppose measurements of such replicates will be approximately normally distributed with unknown mean (the true concentration) and *known* $SD = 0.75$ units. How many replicates should we measure if we wish our 95% confidence interval for the true concentration to have width < 1 units?