Discrete Multivariate Analysis - 579 - HW #7

Alex Towell (atowell@siue.edu)

Problem 1

Consider an experiment on chlorophyll inheritance in maize. A genetic theory predicts the ratio of green to yellow to be 3:1. In a sample of $n = 1103$ seedlings, $n_1 = 854$ were green and $n_2 = 249$ were yellow.

Part (a)

$\boxed{\text{Compute the statistic } X^2 \text{ for testing the proposed model.}}$

The statistic is defined as

$$X^2 = \sum_{j=1}^{c} \frac{(n_j - n\pi_{j0})^2}{n\pi_{j0}}.$$

Under the null model, the odds are 3:1, or $\pi_{10} = 0.75$ and $\pi_{11} = 0.25$. We are given $n = 1103$, $n_1 = 854$, and $n_2 = 249$. Under the null model, these values have expectations given respectively by $n\pi_{10} = 827.25$ and 275.75.

The observed statistic is thus given by

$$X_0^2 = \frac{(854 - 827.25)^2}{827.25} + \frac{(249 - 275.75)^2}{275.75} = 3.46.$$

Part (b)

$\boxed{\text{Determine the upper 10th percentile for the reference distribution.}}$

The reference distribution is the chi-squared distribution with 1 degree of freedom, denoted by $\chi^2(1)$.

The upper 10th percentile given 1 degree of freedom, denoted by $\chi^2_{0.10}(1)$, is found by solving for $\chi^2_{0.10}(1)$ in the equation $\Pr(\chi^2(1) \geq \chi^2_{0.10}(1)) = 1 - 0.10 = 0.9$, which yields the result

$$\chi^2_{0.10}(1) = 2.71.$$

Part (c)

$\boxed{\text{Provide an interpretation of your result, stated in the context of the problem.}}$

We see that any observed statistic X_0^2 with df $= 1$ greater than $\chi^2_{0.10}(1) = 2.71$ is not compatible with the null model at significance level $\alpha = 0.10$.

Since the observed statistic $X_0^2 = 3.46 > 2.71$, the null model, which is the genetic theory where the ratio of green to yellow is 3:1, is not compatible with the data.

Part (d)

What are the shortcomings of hypothesis testing as a measure of evidence?

Hypothesis testing, as a dichotomous measure of evidence, does not provide as much information as a more quantitative evidence measure. For instance, it does not provide information about *effect size*.

Problem 2

Part (a)

Compute a 90% confidence interval for π_1.

The MLE of π_1 is given by $\hat{\pi}_1 = \frac{n_1}{n} = \frac{854}{1103} = 0.774$. Letting $\alpha = 0.10$ and inverting the Wald test statistic, we get the 90% confidence interval for π_1,

$$\hat{\pi}_1 \pm z_{1-alpha/2}\sigma_{\hat{\pi}} = 0.744 \pm 1.645\sqrt{\frac{0.774(1-0.774)}{1103}},$$

which may be rewritten as

$$[0.754, 0.795].$$

Part (b)

Provide an interpretation of your result, stated in the context of the problem.

Based on the observed data, we estimate that the probability π_1 of a green strain is between 0.754 and 0.795.

As expected, the null model specifies a value for π_1 (0.75) that is not contained in this confidence interval.