

Math 308 Assignment 5

Salk Vaccine Trial Hypothesis Testing

Nakul Joshi

16th February 2014

1 Hypergeometric Distribution

Under the null hypothesis, we can assume that everyone who developed the disease after the trial would have done so regardless of which group they were put into. Thus, the number of cases x in the treatment group follows a hypergeometric distribution where:

Population size $N = 200000 + 200000 = 400000$

Cases in population $K = 56 + 141$

Size of treatment group $n = 200000$

Then, the pmf is:

$$Pr_H(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

The requisite p -value is then the probability of 56 or fewer of the cases being chosen into the control group:

$$p_H = \sum_{i=0}^{56} Pr_H(X = i) \\ = 5.98 \times 10^{-10*}$$

This value is certainly statistically significant, and so we reject the null hypothesis that the Salk vaccine is ineffective.

2 Binomial Approximation

Even though the trial involves sampling with replacements, the large sample size means that, under the

*This result was obtained via the *Mathematica* software package.

null, the the number of cases in the treatment group can be modelled by the binomial distribution:

Proportion of cases $p = K/N = \frac{197}{400000}$

The pmf is then:

$$Pr_B(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

The p -value is:

$$p_B = \sum_{i=0}^{56} Pr_N(X = i) \\ = 2.26 \times 10^{-6}$$

This is four orders higher than p_H , but is still extremely low.

3 Normal Approximation

We can further approximate x with normally distributed variable y having $\mu = \frac{nK}{N}$ and $\sigma = \frac{\sqrt{nK(N-K)}}{N}$, giving the pdf:

$$Pr_N(Y = y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

Applying the continuity correction:

$$p_N = \int_0^{56+0.5} Pr_N(y) dy \\ = 1.00 \times 10^{-15}$$

Or five orders lower than p_H .