Advanced Statistics

Exercice 1: Is there a correlation between consumed oil and cardiovascular problems?

1.1 We will perform a χ^2 test for independance between cardiovascular problem and consumed oil. We are given information for 200 persons.

Let X be a random variable modelling the consumed oil. It takes value in { Olive, Groundnut }

Let Y be a random variable modelling the cardivascular status. It takes value in { HasProblem, NoProblem }

The contingence table is given below : $N_{ij}{}_{1 \le i \le 2}{}_{1 \le j \le 2}$

y \ X	Olive	Groundnut	
HasProblem	20 (given) $= N_{11}$	10 $= N_{21}$	$\rightarrow N_{\cdot 1} = N_{11} + N_{21} = 30$
NoProblem	100 $= N_{12}$	70 (given) $= N_{22}$	$\rightarrow N_{\cdot 2} = N_{12} + N_{22} = 170$

$N_{1\cdot} = 200 - N_{2\cdot} = 120$ $\quad$ $N_{2\cdot} = 80$ (given)

1.2. The hypothesis of our χ^2 test are below:

$H_0 = \{$ X and Y are independant $\}$

$H_1 = \{$ X and Y are not independant $\}$

Under H_0, the statistic $\xi_n = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left(N_{ij} - \frac{N_{i\cdot} N_{\cdot j}}{n}\right)^2}{\frac{N_{i\cdot} N_{\cdot j}}{n}}$

converges in distribution towards a $\chi^2(1)$

The test is defined by $W_n = \{ \xi_{200} > q_1(1-\alpha) \}$

with $q_1(1-\alpha)$ the quantile of order $1-\alpha$ of $\chi^2(1)$