

MATH 204 - ASSIGNMENT 3

Please Hand in Assignment in the Lecture on Friday 30th March.

For this assignment, all calculations can be done by hand with a calculator. However, you may use SPSS or other statistics packages.

- The following data relate to a study of the relationship between chronic coughing and cigarette smoking in a cohort of $n = 2847$ twelve year old boys

Reference: Effect of children's and parents' smoking on respiratory symptoms. M Bland, BR Bewley, V Pollard and MH Banks, *Archives of Disease in Childhood*, Vol 53, 100-105, 1978.

	Smoking Status			Total
	Non-smoker	Occasional	Regular	
Cough	266	395	80	741
No cough	1037	977	92	2106
Total	1303	1372	172	2847

For this $r \times c$ table ($r = 2, c = 3$), we wish to test the null hypothesis of **independence** between the row and column factors.

- Form the table of **expected values** under the null hypothesis with entries \hat{n}_{ij} given by the formula

$$\hat{n}_{ij} = \frac{n_{i.}n_{.j}}{n} \quad i = 1, 2, j = 1, 2, 3.$$

where

$n_{i.}$ is the row total for row i
 $n_{.j}$ is the column total for column j .

2 Marks

- Compute the Chi-squared statistic

$$X^2 = \sum_{i=1}^2 \sum_{j=1}^3 \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$$

2 Marks

- Complete the test at the $\alpha = 0.05$ significance level of the null hypothesis, recalling that if the independence hypothesis is true, $X^2 \sim \text{Chi-squared}((r - 1)(c - 1))$.

2 Marks

- An alternative test statistic is the **Likelihood Ratio** test statistic, LR , given by

$$LR = 2 \sum_{i=1}^2 \sum_{j=1}^3 n_{ij} \log(n_{ij}/\hat{n}_{ij})$$

(where log is natural log, or ln). Under the null hypothesis of independence, this test statistic also has an approximate Chi-squared $((r - 1)(c - 1))$ distribution.

Report the result of the test of independence using LR . Test at the $\alpha = 0.05$ significance level.

4 Marks

2. In the following case-control study, the relationship between the single factor (tonsillectomy surgery) and disease status (Hodgkin's disease case or healthy control) was investigated.

Reference: Tonsillectomy history in Hodgkin's disease, S.K. Johnson and R.E. Johnson, *New England Journal of Medicine*, Vol 30;287(22), pp 1122-51972.

Tonsillectomy	Hodgkin's Disease		Total
	Yes	No	
Yes	90	165	255
No	84	307	391
Total	174	472	646

Using the log odds ratio and its standard error

$$\log \hat{\psi} = \log \left(\frac{n_{11} n_{22}}{n_{12} n_{21}} \right) \quad \text{s.e.}(\log \hat{\psi}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

and the test statistic Z

$$Z = \frac{\log \hat{\psi}}{\text{s.e.}(\log \hat{\psi})}$$

test for an association between the factor and disease status. Use the result that under the hypothesis of no association, $Z \sim N(0, 1)$ (see p 894, McClave and Sincich for Normal tables).

4 Marks

3. **The Mann-Whitney-Wilcoxon (MWW)** two-sample test is the non-parametric equivalent of the two-sample t -test. It is used to test the equality of the population medians for the two populations from which samples are drawn.

The test proceeds as follows: suppose that samples from populations 1 and 2 are available. Let the sample sizes be n_1 and n_2 , and the individual samples be x_1, x_2, \dots, x_{n_1} and y_1, y_2, \dots, y_{n_2} .

- (i) List all the data in ascending order, noting which population each value is drawn from.
- (ii) Assign numbers (termed **ranks**) $1, 2, \dots, n_1 + n_2$ to the ordered sampled values, and compute the quantity R_2 , the **sum of the ranks** for sample values from **population 2**.
- (iii) Form the test statistic Z by first computing

$$U = R_2 - \frac{n_2}{2}(n_2 + 1)$$

then computing

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

If the null hypothesis of equal population medians is **true**, $Z_1 \sim N(0, 1)$.

Use the MWW procedure to test for equality of medians for the following two samples: the data are measurements of permeability constants of a placental membrane at full term (Pop 1.) and between 12 to 26 weeks of pregnancy (Pop 2.) Test the hypothesis at the $\alpha = 0.05$ level.

Pop. 1 : Term 0.80, 0.83, 1.89, 1.04, 1.45, 1.38, 1.91, 1.64, 0.73, 1.46

Pop. 2 : 12-26 Weeks 1.15, 0.88, 0.90, 0.74, 1.21

Note that this test is available in SPSS under the menus

Analyze → Nonparametric tests → 2 Independent samples

6 Marks