Application of Categorical Data-nested Design of Knowledge & Control Practices of HBV Infection

O. A. P. Otaru1* and P. N. Ogbonda2

1Department of Mathematical Sciences, Federal University, Lokoja, Nigeria.
2Center for Occupational Health, Safety and Environment, University of Port Harcourt, Nigeria.

Authors’ contributions

This work was carried out in collaboration between both authors. Author OAPO designed the study, wrote the protocol and performed the statistical analysis. Author PNO wrote the first draft of the manuscript, managed the literature searches and instrument design. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AIR/2020/v21i630210

Received 23 March 2020
Accepted 29 May 2020
Published 11 June 2020

ABSTRACT

In real-life, most experimental data are presented in frequencies with no underlying metric probably because of some reasons such as less susceptibility to observational errors. Unfortunately, some of these data have been erroneously analyzed resulting to either type I or type II error. The significance of main factor (University) and sub-factor (Faculty) are studied using categorical data in nested classification. The CATANOVA technique used is suitable for mixed design, having some factors crossed and others nested. The study considered frequency data involving response scores of student’s knowledge and control practices of HBV infection using a scale of good, fair and poor. Numerical results revealed that the main factor, University and the sub-factor, Faculty are not significant (p>0.05) in each case. This implies that irrespective of the different Faculties and Universities, student’s knowledge and control practices of HBV infection was not significantly influenced. More so, there was poor level of student’s knowledge and control practices of HBV infection which were also found to be significantly (p>0.05) same in Universities.

Keywords: Frequency; nested; categorical; knowledge; practices; infection; Hepatitis B; liver cirrhosis.
1. INTRODUCTION

Hepatitis B infection is the world’s most common liver infection, which is life threatening and caused by hepatitis B virus [1]. HBV is highly contagious and is 50–100 times more infectious than HIV and it is transmitted through infected blood, semen, vaginal fluid, and mucous membranes [2,3]. Also, it is transmitted most commonly by unprotected sexual intercourse, exposure to infected blood & body fluids (such as saliva, seminal, menstrual and vaginal fluids), needle stick injury, tattooing, piercing, from mother to child at birth, and among children in early childhood [1]. HBV infection is of major global health concern and is the most common blood-borne viral infection [4]. It can cause chronic infection and puts people at high risk of death from cirrhosis and liver cancer [1].

Most HBV infections do not show symptoms, especially when it is acquired newly which means that people who are infected are at a risk without knowing it [1,5]. However, some people may experience symptoms of acute illness that can persist for several weeks such as jaundice, dark urine, loss of appetite, nausea, vomiting, abdominal pain and extreme fatigue. A small percentage of persons can develop acute liver failure from the acute illness, which can lead to death. This further shows that infected young persons will most likely suffer from liver cirrhosis or liver cancer in later life, if not medically managed [6,1].

Despite the availability of HBV vaccination since 1982, which gives 90%–100% protection against HBV infection, nearly two billion people in the world have been infected with HBV, of which 350 million are chronic carriers. As a consequence of this, approximately 700,000 die every year from HBV-related liver disease or hepatocellular carcinoma in the world [7,8,9]. The increasing prevalence is a constant threat to most communities as it is an important cause of liver cirrhosis and hepatocellular carcinoma resulting in significant morbidity and mortality [10,11,12].

The prevalence of HBV fluctuates greatly in different areas of the world. Approximately three quarters of chronic carriers of HBV live in Asia and Africa [13]. The occurrence of HBV infection is 5-10% in Southeast Asia and 1% in North Europe and America. Nigeria has one of the highest prevalence of HBV infection in the world [14]. Nigeria also accounts for 8.3% of the global burden of chronic HBV infection [15]. Nigeria has HBV prevalence of 11% [15]. With a population of 170 million, this relates to approximately twenty-three million of its general populace living with HBV. HBV together with HIV are responsible for 20% mortality in Nigeria [16]. The incidence of HBV infection can be reduced by giving proper education and awareness regarding its transmission and vaccination to young people. Hence, the study aimed to assess the knowledge and control practice of HBV infection among students of tertiary institutions applying categorical data- nested design. And the objectives include, (i) assess the level of knowledge and control practices of HBV and (ii) assess the effect of universities and faculties on student’s knowledge and control practices of HBV.

2. MATERIALS AND METHODS

2.1 Study Design

The study applied a cross sectional study among students of three universities (University of Port Harcourt, Rivers State University and Ignetus Ajaro University of Education). Three faculties were considered in each university (University of Port Harcourt: Medical Sciences, Sciences and Management Sciences; Rivers State University: Environmental Sciences, Engineering and Law; Ignetus Ajaro University of Education: Humanities, Education and Social Sciences).

2.2 Inclusion and Exclusion Criteria

The recruited respondents were only students from the faculties considered in the study. Also students with studentship less than two years and students who were not available at the time of the study were excluded from the study.

2.3 Sample Size

The sample size (n) for each University is 143. The sample was obtained using the formula given as:

\[ \text{Sample Size (n)} = \frac{Z_{1-\alpha}^2 \cdot P(1-P)}{d^2} \]

Where,

\[ Z_{1-\alpha}^2 \] is standard normal variate (at 5% type I error = 1.96; P is the expected proportion in
population based on previous studies or pilot studies =0.896 [17] and d is the absolute error or precision =0.05.

2.4 Sampling Method

Multi-stage sampling method was adopted. The students were grouped by departments and then selections were made from each group by simple random sampling. All the students who gave consent for the study were administered the preformed structured questionnaire.

2.5 Instrument

The study tool used was a pre-formed self-administered structured questionnaire. The questionnaire included questions on various aspects of HBV infection. The study instrument was validated and a reliability of 0.83 using test-retest of 20 students, who were not part of the study sample. The coefficient was obtained using Pearson Product Moment of Correlation coefficient.

2.6 Method of Data Entry and Analysis

The collected data were entered and analyzed; with the aid of Microsoft Excel sheet version 2010. Response scores were represented in a scale of 1-10. Scores ≥ 8 was rated good, 5-7 was rated fair and ≤ 4 was rated poor. These scores were presented in frequencies and percentages.

2.7 Two-way CATANOVA Nested Modelling

Categorical data analysis deals with data that classify an observation into one or more categories [18]. Nominal data are inherently less informative than quantitative data, it does not convey information about the magnitude of differences and it is not also clear on how to deal with missing data [19].

A categorical variable sometimes called nominal variable is one that has two or more categories, but there is no intrinsic ordering to the categories [20]. These variables are summarized in the form of a contingency table. In real-life, most experimental data are presented in frequencies with no underlying metric probably because of some reasons such as less susceptibility to observational errors [21]. Unfortunately, some of these data have been erroneously analyzed resulting to either type I or type II error. The nesting of factors is of great importance when an experimental situation requires that unique levels of one factor occurring within each level of a second factor. Researchers are increasing faced with the problem of developing useful methods for analyzing categorical data. Although methods for analyzing quantitative continuous variables are vast, the assumptions are completely different when the data structure is non-quantitative or categorical [20]. Literature has shown the existence of several techniques for analyzing data in contingency table. Some researchers deal with categorical data in their original format while others transform categorical data to enable analysis with existing methods for quantitative data [22-26]. There are analytical tools for mixed design when some factors may be crossed while other are nested [21]. Applying these techniques, frequency data are analyzed in their original format, without the need for any transformation.

A two-way categorical analysis of variance (CATANOVA) in a nested arrangement due to [21] is adopted for this study. We assume no loss in generality using the method, for unequal levels of sub-factor that do not differ significantly. However, for computational ease, the average of the sub-factor levels shall be employed. The layout for the two-way CATANOVA nested classification is presented in Appendix Table 1.

Appendix Table 1 shows that the main factor A (1, I) and sub-factor B (1, J) have k=1, k quanta responses per unit.

To test for the significance of the main factor A and the sub-factor B, the model is given as:

\[ P_{ijk} = P_k + (P_{ik} - P_k) + (P_{ijk} - P_{ik}) \] (1)

Where,

\[ P_k \] is a constant for the kth response, \( P_{a} - P_{b} \) is the effect of the ith level of factor A, and \( P_{ijk} - P_{ik} \) is the effect of the jth level of factor B within the ith level of factor A. Also,

\[ P_{ijk} = \frac{n_{ijk}}{n_{ij}} \] ; \[ P_{ik} = \frac{n_{ik}}{n_j} \] ; \[ P_k = \frac{n_k}{n} \] ;

\[ n = \sum_{ij} n_{ij} = \sum_i n_i = \sum_k n_k \] ; \[ n_{ijk} = \sum_j n_{ijk} \] ;

\[ n_{ij} = \sum_k n_{ijk} \] ; \[ n_k = \sum_{ij} n_{ijk} \] ; \[ n = \sum_{jk} n_{ijk} \]
The fundamental assumptions in CATANOVA are associated with statistical distributions, independence and constant variance [27-29]. Thus, the categorical data for the study follows:

a) Binomial Distribution: \( n_{ijk} \) is
\[
b(n_{ijk}; n_{ij}, \pi_{ijk}, \pi_{ijk}(1-\pi_{ijk}))
\]
b) Covariance:
\[
(n_{ijk}, n_{ijk}') = \begin{cases} 
-n_{ij} \pi_{ijk} \pi_{ijk}; i = i', j = j', k \neq k' \\
0 & \text{elsewhere}
\end{cases}
\]
c) Variance:
\[
\text{var} = (n_{ijk}) = n_{ij} \pi_{ijk}(1 - \pi_{ijk}) \text{ is not a constant.}
\]

The two hypotheses considered in the study are

\( H_{OA} : \pi_{ik} = \pi_k , \forall i \) (There is no main factor effect)
\( H_{OB} : \pi_{ijk} = \pi_{ik} , \forall j, i \) (There is no sub-factor effect)

Relying on equation (1), the sum of squares [20,21] is given as:
\[
\text{TSS} = n - \frac{\sum \limits_{ijk} n_{jk}^2}{n} = n - C_k ; \text{USS} = n - \frac{\sum \limits_{ijk} n_{ik}^2}{n_{ij}} ;
\]
\[
\text{SSA} = \sum \limits_{ijk} n_{ij} (P_{ik} - P_k)^2 = C_{ik} - C_k
\]
\[
\text{SSB(A)} = \sum \limits_{ij} n_{ik} (P_{ijk} - P_k)^2 = C_{ijk} - C_k
\]

Where,
\[
C_k = \frac{\sum \limits_{i} n_{ik}^2}{n} ; C_{ik} = \frac{\sum \limits_{j} n_{ik}^2}{n_{ij}} ; C_{ijk} = \frac{\sum \limits_{j} n_{ijk}^2}{n_{ijk}}
\]

SSA is the sum of square of the main factor A, SSB (A) is the sum of square of the sub-factor B, USS is the unit sum of square and TSS is the total sum of square.

Thus, the chi-square test ratio for significance of treatments is given as:
\[
\chi^2_A = \frac{\text{SSA} (n-1)(K-1)}{\text{TSS}}
\]
\[
\chi^2_{B(A)} = \frac{\text{SSB(A)} (n-1)(K-1)}{\text{TSS}}
\]

Decision: Reject \( A_i \) and \( B_{ij} \) respectively at specified \( \alpha \)-level of error if \( \chi^2_A \geq \chi^2_{(K-1)(I-1)} \) and \( \chi^2_{B(A)} \geq \chi^2_{(K-1)(J-1)} \) respectively. This implies that as \( n \rightarrow \infty \) SSA and SSB (A) is approximately Chi-Square distribution with \( (K-1)(I-1) \) and \( (K-1)(J-1) \) degree of freedom respectively provided TSS is independent of SSA and SSB (A) [21]. These results are summarized in the Appendix Table 2.

3. RESULTS

\[
C_k = \frac{66^2 + 107^2 + 261^2}{434} = 193.377
\]
\[
C_{ik} = \frac{34^2 + 34^2 + 84^2 + 11^2 + 35^2 + 92^2 + 21^2 + 38^2 + 85^2}{152 + 138 + 144} = 195.979
\]
\[
C_{ijk} = \frac{15^2 + 11^2 + 39^2 + 8^2 + 13^2 + 16^2 + 21^2 + 10^2 + 29^2 + \cdots + 8^2 + 13^2 + 33^2}{65 + 37 + 50 + \cdots + 54} = 200.948
\]
\[
\text{TSS} = 434 - 193.377 = 240.623
\]
The summary of results is presented in Appendix A (see Table 3).

4. DISCUSSION

Categorical analysis of variance (CATANOVA) for data in nested structure has been applied in studying the significance of student’s knowledge and control practices of HBV infection. The nesting of factors is necessary when an experimental situation requires that unique levels of one factor occur within each level of a second factor. Two-way nesting due to [21] was considered where University served as the main factor and faculty served as the sub-factor. The knowledge and control practices of HBV infection score was viewed as the “response” factor, having three levels. Three levels of University were considered with equal level of faculty within each University for a balance sub-factor levels, because it is assumed that for large sample size, the CATANOVA technique employed is robust for balance structure of sub-factor levels. The study revealed that student’s knowledge and control practices of HBV infection is poor within the University (main factor) and the faculty (sub-factor). This finding is in concordance with [30,31], who reported that the knowledge, awareness and control practices of HBV infection was lacking among university students. Also, [32] reported that the overall awareness regarding HBV disease was found to be lacking among the university students. Similarly, a study among medical students at Syrian Private University revealed the weakness of general knowledge about hepatitis B among junior medical students compared to those in the fifth year [33]. The categorical data analysis revealed insignificant difference in student’s knowledge and control practices of HBV infection within the University (main factor) and the faculty (sub-factor). The p-value for each test revealed that the test is not significant (p>0.05) at 5% level of significance. More so, there is no significant effect of University (main factor) and faculty (sub-factor) on student’s knowledge and control practices of HBV infection. Thus, the null hypothesis is accepted. This implies that students’ knowledge and control practices of HBV infection have not been affected significantly by the University community.

5. CONCLUSION

Categorical or nominal (frequency) data are frequently encountered in real-life situations in all environment of human endeavors. The analysis of these data have erroneously been analyzed and used for decisions without the consciousness that there is no underlying metric with such data. Recently, researchers have propagated the need for methods of handling categorical data. Thus, the study applied categorical analysis of variance for data in nested design, in the aim of establishing the significance of main effect and sub-effect. Results revealed the efficiency of the method, thus, the study recommends massive education of students on the knowledge and control practices of HBV infection by the University community and the government.

CONSENT AND ETHICAL APPROVAL

Ethical approval to conduct the study was obtained from the research ethics committee of the University of Port Harcourt, Rivers State University and Ignetus Ajaro University of Education before commencement of the study. Informed consent was obtained from the prospective students before recruitment into the study.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES


24. Dyke GV, Patterson HD. Analysis of factorial arrangements when the data are proportions. Biometrics. 1952;8:1-12.


26. Light RJ, Margolin BH. An analysis of variance for categorical data. Journal of
APPENDIX - A

Table 1. Layout for a 2-way CATANOVA nested classification

<table>
<thead>
<tr>
<th></th>
<th>a₁</th>
<th>a₂</th>
<th>a₃</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b₁</td>
<td>b₂</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n₁₁</td>
<td>n₁₂</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n₂₁</td>
<td>n₂₂</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n₃₁</td>
<td>n₃₂</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Two-way CATANOVA nested classification

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>SS</th>
<th>MS</th>
<th>Test ratio</th>
<th>Critical Value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main factor Aᵢ</td>
<td>I-1</td>
<td>SSA</td>
<td>MSA</td>
<td>χ²</td>
<td>2.602</td>
<td>1.301</td>
</tr>
<tr>
<td>Sub-factor Bⱼ(i)</td>
<td>I (J-1)</td>
<td>SSB (A)</td>
<td>MSB(A)</td>
<td>χ²</td>
<td>4.969</td>
<td>0.828</td>
</tr>
<tr>
<td>Unit Uijk</td>
<td>n-IJ</td>
<td>USS</td>
<td>UMS</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>n-1</td>
<td>TSS</td>
<td>TMS</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3. Results for two-way CATANOVA nested classification

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f</th>
<th>SS</th>
<th>MS</th>
<th>Test Ratio</th>
<th>Critical Value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main factor Aᵢ</td>
<td>2</td>
<td>2.602</td>
<td>1.301</td>
<td>2.602</td>
<td>1.301</td>
<td>9.364</td>
</tr>
<tr>
<td>Sub-factor Bⱼ(i)</td>
<td>6</td>
<td>4.969</td>
<td>0.828</td>
<td>4.969</td>
<td>0.828</td>
<td>17.883</td>
</tr>
<tr>
<td>Unit Uijk</td>
<td>425</td>
<td>233.052</td>
<td>0.548</td>
<td>233.052</td>
<td>0.548</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>433</td>
<td>240.623</td>
<td>0.555</td>
<td>240.623</td>
<td>0.555</td>
<td>-</td>
</tr>
</tbody>
</table>

*Mean significant at 5% level of error
### Table 4. Student knowledge and control practices of HBV infection in three universities

<table>
<thead>
<tr>
<th>Response</th>
<th>University of Port Harcourt</th>
<th>Rivers State University</th>
<th>Ignatius Ajuru University of Education</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Faculty j(i)</td>
<td>Faculty j(ii)</td>
<td>Faculty j(iii)</td>
<td></td>
</tr>
<tr>
<td>Medical sciences</td>
<td>Sciences</td>
<td>Management sciences</td>
<td>n _ik</td>
<td>Environmental sciences</td>
</tr>
<tr>
<td>Good</td>
<td>15</td>
<td>8</td>
<td>11</td>
<td>34</td>
</tr>
<tr>
<td>Fair</td>
<td>11</td>
<td>13</td>
<td>10</td>
<td>34</td>
</tr>
<tr>
<td>Poor</td>
<td>39</td>
<td>16</td>
<td>29</td>
<td>84</td>
</tr>
<tr>
<td>n _i</td>
<td>65</td>
<td>37</td>
<td>50</td>
<td>152</td>
</tr>
</tbody>
</table>


Questionnaire Title

“Knowledge and control practices of HBV infection”

Bio-Data:

Age............  Sex.........  Department.........................

Please Tick (✓) appropriate box.

1. Have you heard of HBV? Yes ☐ No ☐
2. If yes, what medium did you hear of HBV? ........................................
3. What is the meaning of HBV? ........................................
4. HBV is life threatening? Yes ☐ No ☐
5. HBV is .......... infection? (a) Viral (b) Bacterial (c) Parasitic
6. Have you heard of other types of hepatitis infection? Yes ☐ No ☐
7. Is HBV preventable? Yes ☐ No ☐
8. Is screening of blood donors for HBV mandatory for safe transfusion? Yes ☐ No ☐
9. HBV patients can be completely cured by drugs? Yes ☐ No ☐
10. HBV affects which organ in the body? ...Liver.....................
11. Chronic HBV can lead to .............. (a) Cirrhosis (b) Carcinoma liver (c) Kidney disease (d) Heart disease (e) Death.
12. Have shared needles, while engaging in intravenous drug use? Yes ☐ No ☐
13. Have engaged in unprotected sexual intercourse without knowing the HBV status of my partner? Yes ☐ No ☐
14. Have shared shape objects, while engaging in body piercing & tattoo? Yes ☐ No ☐
15. Is there hepatitis B vaccination available? Yes ☐ No ☐
16. Is HBV an infectious / transmissible disease? Yes ☐ No ☐
17. Have you been immunized with HBV vaccine? Yes ☐ No ☐
18. If yes, did you complete the three dosages? Yes ☐ No ☐
19. If No to question 15, why (please give reasons)? ....................
20. If No to question 14, why (please give reasons)? ....................
21. HBV is spread by............. (a) Blood transfusion (b) Contaminated shape objects (c) Exposures to infected body fluids (d) Food prepared by infected persons (e) Shaking hands with infected persons (f) Kissing. (g) Vertical transmission from mother to child (h) Unprotected sexual intercourse (i) Coughing and sneezing (j) Sharing toilet seat with infected person (k) Sharing eating utensils with infected person.

Thank You

© 2020 Otaru and Ogbonda: This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/57198