ISSN No:-2456-2165

Comparison of Cochran Test and Loglinear Modeling

Taiwo, Abraham Soji
Academic Planning Unit
KolaDaisi University Ibadan

Adejumo Adebowale Olusola
Professor, Department of Statistics
University of Ilorin, Ilorin

Sulaimon, Ibikunle Idris
Department of Statistics
University of Ilorin, Ilorin

Abstract:- When performing a test of conditional independence or Homogeneity of associations, one of the tests that come to mind is the Cochran or Mantel Haenszel test, furthermore, the Breslow day test. The attributes of these tests are notable and it is being recommended that log linear models can be utilized to test similar theories. This research paper takes a look at the two tests in the framework of loglinear modeling. Expanding on the knowledge of both tests by delineating how theories of homogeneous associations and conditional independence can be applied along with extra speculations and introducing a speculation of the ideas of homogeneity of associations and conditional independence in higher request connections in the context of loglinear modeling.

Methods were illustrated using real life data on HIV /AIDS. The data which was classified into a $2 \times 2 \times 5$ partial table consisted of the morbidity pattern and mortality for both genders from 2010-2014. It was revealed that the odds ratio of gender by survival status for each of the year was heterogeneous. It was also discovered that the survival status for HIV/AIDS depended on gender conditionally while controlling for the effect of the years. Finally, we observed that the conditional odds ratio for male deaths is 1.612 based on the available data.

***Keywords**:- Conditional Independence, Homogeneity of Associations, Loglinear Modeling, Cochran or Mantel Haenszel Test, Breslow Day Test.*

I. INTRODUCTION

A categorical variable has an estimation scale comprising of a bunch of classes. For instance, political way of thinking is frequently estimated as liberal, moderate, or traditionalist. Conclusions with respect to bosom disease dependent on a mammogram utilize the classes ordinary, generous, most likely benevolent, dubious, and threatening. In this research paper, the statistical methods will be used to analyze gender difference in mortality of HIV/AIDS.

Human Immune-deficiency Infection contamination/Acquired Immune Deficiency Syndrome is an ailment of the human safe framework achieved by disease with Human Immune inadequacy Virus (HIV) the term HIV/AIDS addresses the whole scope of sicknesses brought about by HIV infection from early contamination to late stage manifestations. During the underlying pollution, an individual may experience a brief time frame of influenza like disorder. This is routinely followed by a drawn out period without incidental effects. As the infection propels, it intrudes more with the safe structure, making the individual essentially more inclined to pollutions.

HIV is sent fundamentally through unprotected sex (tallying butt-driven and oral sex), degraded blood holding, hypodermic needles, and from mother to kid during pregnancy, transport, or bosom taking care of. Some natural liquids, for example, salivation and tears don't communicate HIV. Counteraction of HIV contamination fundamentally through safe sex and needle trade programs is a vital methodology to control the spread of the infection. There is no fix or antibodies, be that as it may, antiretroviral treatment can hinder the illness and may prompt a close to typical future.

Hereditary exploration demonstrates that HIV started in West Central Africa during the late nineteenth or mid 20th century. Helps was first perceived by the United States Centers for Disease Control and Prevention (CDC) in 1981 and its motivation, HIV contamination was recognized in the early piece of the decade. Since its revelation, AIDS has caused an expected 56 million passing around the world (as of 2017). As of 2017, 30.3 million people are living with the illness universally. HIV/AIDS is regarded as a pandemic (a sickness episode which is available over an enormous region and is altogether spreading (Wikipedia, 2017).

It is an unavoidable truth that men partake in certain actual benefits over ladies. By and large, men are quicker, taller and more grounded and less inclined to be overweight. In any case, none of these characteristics seem to issue over the long haul. For whatever the real morals of maleness, life range isn't among them. Women as a social event live more than men. In totally made countries and most deficient with regards to ones, women outlive men, sometimes by an edge of as much as 10years. Which is one reason ladies dwarf men nine to one, around the world (Thomas, 2017).

Additionally, the destruction rates for women are lower than those for men at all ages – even before birth. Despite the fact that young men start existence with some mathematical influence around 115 guys are considered for each 100 females – their numbers are trimmed down from there on. A bigger number of young men than young ladies bite the dust in earliest stages. Also, during each ensuing year of life, death rates for guys surpass those for females, so that by age 25 ladies are in the larger part.

As far as we might be concerned, these insights bring up two issues: Why do men kick the bucket so youthful? What's more, for what reason do ladies pass on so old? From the start we might want to concede that we have no authoritative responses to these inquiries. Be that as it may, the accessible proof ensnares conduct just as natural contrasts between the genders, contrasts in the impacts of clinical innovation, just as friendly and mental elements (Jerry, 2014).

II. METHODOLOGY

The variables involved in this study are gender, survival status, and years. We wish to take a systematic approach to analyzing the dependence between gender and survival status keeping the years constant.

Firstly, we consider the principles for the classification of this data, definitions such as; what partial tables are homogeneous association, conditional independence and other related definitions.

We take a look at the Breslow day test for the Homogeneity of chances proportion. We need this to test for the homogeneity of the chances proportion of gender by survival status between each level of years.

We will use Cochran test and Mantel Haenszel test for repeated test of independence. . There are three ostensible factors we wish to know whether two of the factors are free of one another keeping the third factor which distinguishes the rehashes steady. The most widely recognized circumstance is that you have different tables of autonomy. There are renditions of Cochran and Mantel Haenszel test for quite a few lines and segments that are currently being worked on. Lastly, the loglinear analysis; this will be used to carry out tests of association carried out by the previous tests.

Partial Association in Stratified 2×2 Tables

A significant piece of most examinations, particularly observational investigations, is the decision of control factors. In contemplating the impact of X on Y, one should control any covariate that can impact that relationship. This includes utilizing some component to hold the covariate consistent. Something else, a noticed impact of X on Y may really reflect impacts of that covariate on both X and Y. The connection among X and Y then, at that point shows frustrating. Trial studies can eliminate impacts of confounding covariates by arbitrarily doling out subjects to various degrees of X, yet this is preposterous with observational examinations. (Simpson 1951, Yule1903).

Partial Tables

Two-way cross-sectional cuts of the three-way table cross group X and Y at isolated degrees of Z. These cross segments are called partial tables. They show the XY relationship at fixed degrees of Z, subsequently showing the impact of X on Y while controlling for Z. The halfway tables eliminate the impact of Z by holding its worth steady.

The two-way possibility table that outcomes from consolidating the fractional tables is known as the XY peripheral table. Every cell include in it is an amount of checks from a similar cell area in the incomplete tables. The negligible table contains no data about Z, so instead of controlling Z, it overlooks Z. It's anything but a two-way table relating X and Y. Strategies for two-way tables don't consider impacts of different factors.

The relationship in incomplete tables are called contingent affiliations, since they allude with the impact of X on Y restrictive on fixing Z at some level. Restrictive relationship in fractional tables can be very not the same as relationship in peripheral tables, as the following model shows.

Conditional Independence versus Marginal Independence

On the off chance that X and Y are autonomous in every partial table, X and Y are supposed to be conditionally independent, given Z; all contingent chances proportions among X and Y then, at that point equivalent 1. Restrictive autonomy of X and Y, given Z, doesn't suggest minor freedom of X and Y. That is, when chances proportions among X and Y equivalent 1 at each degree of Z, the minimal chances proportion may vary from 1

Homogeneous Association via the Breslow Day Test

A $2\times2\times k$ table has homogeneous XY association when

$$\theta_{xy(1)} = \theta_{xy(2)} = \theta_{xy(k)} \quad (1)$$

Then, at that point the impact of X on Y is something similar at every class of Z.

Conditional Independence of X and Y is the uncommon case where each conditional odds ratio equals 1.

When there isn't homogeneous affiliation, the contingent odd proportion for any pair of factors changes across levels of the third factor. For X=smoking (yes or no), Y= Lung malignant growth (yes or no), and Z= Age (<45, 45-64, >65). Assume , and , then, at that point smoking weakly affects cellular breakdown in the lungs for youngsters, however the impact fortifies significantly with age.

Two tests that can be used to test for the homogeneity of odds ratio in a $2\times2\times k$ contingency table are the Breslow day test and the Tarone test. Also the Mantel Haenszel common odds ratio estimate gives the odd that an outcome in variable Y given a particular category of X compared to the odds of that outcome in Y occurring giving the other category of X keeping the third category Z constant.

Breslow and Day proposed the test statistic;

$$BD = \frac{\sum\left(n_{11k} - \mathrm{E}\left(n_{11k}|OR_{MH}\right)\right)^2}{V\left(n_{11k}|OR_{MH}\right)} \quad (2)$$

Where OR_{MH} is the Mantel Haenszel estimator of common odds ratio,

$\mathrm{E}\left(n_{11k}|OR_{MH}\right)$ and $V\left(n_{11k}|OR_{MH}\right)$ are the expected and variance of n_{11k} under the null hypothesis of homogeneity of odds ratio. Under the supposition of huge sample size in each 2×2 table, BD has surmised chi square distribution with $k-1$ degree of freedom.

Mantel and Haenszel (1959) proposed a non-model based trial of H_0: conditional independence in $2\times 2\times k$ tables. Zeroing in on review investigations of infections, they treated response (column) marginal totals as fixed. This, in each partial table k of cell counts $\{n_{1jk}\}$, their analysis conditions on both the predictor totals (n_{i+k}, n_{2+k}) and the response outcome totals (n_{+1k}, n_{+2k}), the usual sampling schemes then yield a hyper-geometric distribution for the first cell count n_{i1k} in each dissemination for the partial table. That count decides $\{n_{12k}, n_{21k}, n_{22k}\}$, given the marginal aggregates.

Under H_0, the hypergeometric mean and variance of n_{11k} are;

$$\mu_{11k} = \mathrm{E}\left(n_{11k}\right) = \frac{n_{1+k}n_{+1k}}{n_{++k}}$$

$$Var\left(n_{11k}\right) = \frac{n_{1+k}n_{2+k}n_{+1k}n_{+2k}}{n^2_{++k}\left(n_{++k}-1\right)}$$

Cell counts from the diverse partial tables are independent. The test statistic includes every partial table or $\theta_{XY(k)} < 1$ in each table, $\Sigma_k\left(n_{11k} - \mu_{11k}\right)$ tends to be fairly large in total value. This test works best when the XY association is similar in each partial table.

Cochran (1954) proposed a similar statistic. He treated the rows in each 2×2 table as two independent binomials rather than hypergeometric. Cochran's statistic is with $Var\left(n_{11k}\right)$ replaced by

$$Var\left(n_{11k}\right) = \frac{n_{1+k}n_{2+k}n_{+1k}n_{+2k}}{n^3_{++k}}$$

The Mantel and Haenszel approach utilizing the hypergeometric is broader in that it likewise applies to certain cases wherein the columns are autonomous binomial examples from two populaces. Models are review examines and randomized clinical preliminaries with the variable subjects haphazardly dispensed to the two medicines. In the primary case the section aggregates are normally fixed. In the second, under the invalid theory, the section edges are the equivalent paying little heed to how the subjects were relegated to medicines, and randomization contention lead to the hypergeometric in each 2×2 table.

Mantel and Haenszel (1959) proposed inclusion of a continuity correction. The p-value from the test then better approximates an exact conditional test but it tends to be conservative.

Loglinear models for three-way tables

With three-way possibility tables, loglinear models can address different freedom and affiliation designs. Two-factor affiliation terms portray the contingent chances proportions between factors.

For cell expected frequencies $\{m_{ijk}\}$, consider loglinear model

$$\log\left(m_{ijk}\right) = \mu + \mu_{i(x)} + \mu_{j(y)} + \mu_{k(z)} + \mu_{ik(XZ)} + \mu_{jk(YZ)}$$

Models that erase extra affiliation terms are too easy to even think about fitting most informational collections well. For example, the model that contains just single-factor terms, signified by (X, Y, Z), is known as the common independence model. It regards each pair of factors as autonomous, both conditionally and marginally. At the point when factors are picked astutely for an examination, this model is infrequently suitable.

$$\log\left(m_{ijk}\right) = \mu + \mu_{i(x)} + \mu_{j(y)} + \mu_{k(z)} + \mu_{ik(XZ)} + \mu_{jk(YZ)}$$

This loglinear model is called the homogeneous association model and is represented by (XY, XZ, YZ).

The utmost universal loglinear model for three-way tables is

$$\log\left(m_{ijk}\right) = \mu + \mu_{i(x)} + \mu_{j(y)} + \mu_{k(z)} + \mu_{ik(XZ)} + \mu_{jk(YZ)} + \mu_{ijk(XY}$$

Key Concepts

$$\ln\left(m_{ij}\right) = \mu + \mu_{i(A)} + \mu_{j(B)} + \mu_{ij(AB)}$$

$\ln\left(m_{ij}\right)$ is the log of the expected cell frequency of the cell ij in the table.

μ is the overall mean of the natural log of the expected frequencies.

$\mu_{i(A)}$ or $\mu_{j(B)}$ terms each represent effect which the variables have cell frequencies.

A and B are the variables

i and j refer to the categories within the variables.

Therefore

$\mu_{i(A)}$ = main effect for A

$\mu_{j(B)}$ = main effect for B

$\mu_{ij(AB)}$ = the interaction effect for A and B

The above model is a saturated model since it incorporates all conceivable one way and two way impacts. Given that the soaked model has similar measure of cells in the possibility table as it does impacts, the normal cell frequencies will in every case precisely match the noticed frequencies without any levels of opportunity remaining (Knoke and Burke, 1980). For instance, in a 2×2 table there are four cells and in a saturated model including two variables, there are four impacts; $\mu, \mu_{i(A)}, \mu_{j(B)}, \mu_{ij(AB)}$, along these lines the normal cell frequencies will precisely coordinate with the noticed frequencies. Hence, to track down a more parsimonious model that will segregate the impacts best showing the information designs, a non-saturated model should be looked for. This can be accomplished by setting a portion of the impact boundaries to nothing. For instance, if we set the effects parameter $\mu_{ij(AB)}$ to zero (for example we accept that factors A has no impact on factor B or the other way around) we are left with the unsaturated model.

$$\ln(m_{ij}) = \mu + \mu_{i(A)} + \mu_{j(B)} \quad (5)$$

This specific unsaturated model is named the independence model since it does not have an association impact boundary among A and B undifferentiated from the chi square investigation, testing theory of autonomy.

Parameter Estimates

Once estimates of the expected frequencies for the given model are gotten, these numbers are gone into suitable recipes to deliver the outcome parameter estimates (μ's) for the factors and their connection (Knoke and Burke, 1980). The impact boundary gauges are identified with chances and chances proportions. Chances are depicted as the proportion between the recurrence of being in one class and the recurrence of not being in that classification. The chances proportion is one contingent chances separated by another briefly factor.

The Pearson chi square statistic or the likelihood ratio (G^2) can be used to test a model's fit. The formula for the G^2 statistic is as follows:

$$G^2 = 2\sum_{ijk} n_{ijk} \ln\left(\frac{n_{ijk}}{m_{ijk}}\right) \quad (6)$$

Where

n_{ijk} is the observed frequency for cell ijk

m_{ijk} is the expected frequency for cell ijk

It is normal tracked down that more than one model gives a sufficient fit to the information as shown by the non-significance of the probability proportion. Now, the probability proportion can be utilized to think about a general model inside a more modest, nested model (contrasting a saturated model and one interaction or principle impact dropped to survey the significance of that term). The condition is as per the following:

$$G^2_{comparison} = G^2_{model1} + G^2_{model2}$$

with degree of freedom, $df = df_{model1} - df_{model2}$

Model 1 is the model nested within model 2. If the $G^2_{comparison}$ statistic is not significant, then the nested model 1 is not significantly worse than the saturated model 2. Therefore, choose the more parsimonious (nested) model.

Loglinear representation of the Cochran and Breslow Day Test

Here we see how loglinear modeling can be used to carry out both the test of homogeneous association and test of conditional independence.

Loglinear representation of Homogeneous Association

Homogeneous association infers that the conditional connection between any pair of factors given the third one is something similar at each level of the third factor. The model is otherwise called a no three factor associations model or no second interactions model.

There is actually no graphical portrayal for this model, however the loglinear documentation is (AB, BC, AC), demonstrating that on the off chance that we know each of the two way tables, we have adequate data to process the normal tallies under this model. In the loglinear documentation, the soaked model or three factor association models is (ABC). The homogeneous affiliation model is middle in intricacy, between the contingent autonomy model (AC, BC) and the soaked model (ABC).

The saturated model (ABC) allows the AB odds at each level of C to be arbitrary. The conditional independence model (AC, BC) involves the AB odds ratios at each level of C to be equal to 1The homogeneous association model (AB, AC, BC) requires the AB odds ratios at every level of C to be the same, but not certainly equal to 1

Loglinear Test of Homogeneous Association

Here we illustrate how to use the loglinear approach to testing for homogeneous associations. Assuming we have three variables A, B and C. To test the hypothesis that odds ratios of A and B across the categories of C is homogeneous, we analyze two hierarchical loglinear models. The first model is the saturated model, which is the model (XYZ). The second model excludes the ABC interaction. The Likelihood ratio estimates are obtained for both models and the difference (subtracting the saturated model from that without the ABC interaction) between both likelihood ratio estimates is calculated. If this difference is significant with a chi square distribution having a degree of freedom which equals the difference in the degrees of freedom from both likelihood ratio estimates for the models, then we can say that homogeneous association between the odds ratios of A and B at each categories of C does not hold. A very similar result will be obtained if the Breslow Day test statistics is used.

Loglinear representation of Conditional Independence

The idea of conditional independence is vital and it is the reason for some measurable models (for example dormant class models, factor investigation, thing reaction models, graphical models, and so forth) There are three potential conditional independence models with three arbitrary factors, (AB, AC), (AB, BC), (AC, BC). Think about the model (AC, BC), which implies that A and B are conditionally autonomous given C. In numerical terms, the model (AC, BC) implies that the conditional likelihood of An and B given C equivalents the result of conditional probabilities of A given C and B given C.

As far as chances proportions, the model suggests that on the off chance that we take a gander at the halfway tables, that are AXB tables at each degree of C that the chances proportions in these tables ought not essentially vary from 1. Tying this back to two way tables, we can test in every one of the halfway AXB tables at each degree of C to check whether independence holds.

Loglinear Test of Conditional Independence

Here we illustrate how to use the loglinear approach to testing conditional independence. Also assuming we have three variables A, B and C where the A and B variables has only two categories each, while the C variable has k categories. To test the hypothesis that A and C are conditionally independent given Z, we analyze these using two hierarchical loglinear models. The first model includes all three first order interactions, which is the model (AB, AC, and BC). The second model excludes the AB interaction. Neither model contains the ABC interactions. The Likelihood ratio estimates are obtained for both models and the difference (subtracting the model with all three first order interaction from that without the AB interaction) between both likelihood ratio estimates is calculated. If this difference is significant with a chi square distribution having a degree of freedom which equals the difference in the degrees of freedom from both likelihood ratio estimates for the models, then we can say that conditional independence between A and B given C does not hold. A very similar result will be obtained if the Cochran and Mantel Haenszel test statistics is used.

Generalization of loglinear test of homogeneous association and conditional independence

So far in this exploration work and in the writing accessible to us, the ideas of homogeneity of affiliations and conditional independence have been predominantly talked about at the degree of two way interactions. We currently address a characteristic speculation of the two ideas. We sum up the possibility that a homogeneous affiliations or conditional independence holds in the present of extra factors to such an extent that it incorporates interactions of higher than first request. The subsequent methodology is a trial of the homogeneity or conditional affiliation that can be deciphered as a summed up Cochran or Breslow Day tests.

Consider the situation where a specialist breaks down the four factors W, X, Y and Z. the ideas examined hitherto include a communication among X and Y and determine a speculation concerning this collaboration for the classes of a third factor Z. Consider for instance, the three way communication among the factors W, X and Y. then, at that point, the relationship among these factors can be considered homogeneous at the degree of at various classifications of variable Z. All in all, the present circumstance can be portrayed as absence of the four way association between W, X, Y and Z. Additionally, the conditional independence between three factors W, X and Y given the variable Z can likewise be tried.

To test for this two concepts, we continue in a manner undifferentiated from the one lined out for the three variable case, that is we test and think about two arrangements of hierarchical models. For the test homogeneous association, we compare the saturated model and the model not including the four variable interactions but containing all three variable interactions. For the test conditional independence, we compare the model of all three way interaction excluding the WXY interaction and the model of all the three way interaction (Note: this model excludes the four variable interactions).

The conditional independence between variables W and X could also be tested in the presence of variables Y and Z could also be tested. This could be done by subtracting the Likelihood ratio estimates of the model of only all two variable interactions from the model of only all two variable interactions excluding the WX interaction (Von Eye and Spiel, 1996).

III. RESULTS AND DISCUSSION

The data as seen in table 1has been arranged into $2\times2\times5$ table i.e. Gender by Outcome by Year. The data in table 1 was obtained from Lagos State Secretariat Alausa, which was then arranged by year and gender.

Table 1: Cases of HIV/AIDS in Lagos State by year and gender

Year	Number of infected Males	Number of infected Females	Total number of infected persons	Number of Male Deaths	Number of Female Deaths	Total Number of Deaths
2010	2706	4587	7293	149	173	322
2011	3471	5638	9109	252	297	549
2012	1096	2268	3364	103	138	241
2013	1420	3099	4519	81	127	208
2014	2592	5594	8186	238	255	493

Note: Data for cases of HIV/AIDs in Lagos State from the Secretariat Alausa Ikeja (2019)

The data as seen in table 1 arranged into $2 \times 2 \times 5$ table i.e. Gender by Outcome by Year, was used in calculating the survival status as presented in table 2; the partial table.

Table 2:The partial table of gender by survival status by year

Gender	2010		2011		2012		2013		2014	
	Survival status		Survival status		Survival status		Survival status		Survival status	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Male	149	2557	252	3219	103	993	81	1339	238	2354
Female	173	4414	297	5341	138	2130	127	2972	255	5339

Note: Partial table computation from Author (2020)

It was revealed in table 2 that survival status of respondents with HIV/AIDs varies according to year.

Test of Homogeneity of odds ratio

Breslow day test and Tarone test; the null hypothesis states that the odds ratio are homogenous against the alternative hypothesis which stated that the odds ratio are not homogenous.

The Breslow and Tarone test was obtained and presented in table 3

Table 3:Breslow Day and Tarone test results

	Chi-Squared	Df	Asymp. Sig. (2-sided)
Breslow-Day	12.179	4	.016
Tarone's	12.178	4	.016

Note: Breslow Day and Tarone test results from Author (2020)

Breslow Day and Tarone test results as presented in table 3, shows that the odds ratio is not homogenous, therefore, there is need to carry out the conditional independence test.

Test of conditional independence in a $2 \times 2 \times k$ table

Cochran's test will be deployed. The null hypothesis which states that the survival status of HIV patients is independent of gender across the period

The Cochran's test statistics is given as:

$$M^2 = \frac{\left(\sum_k n_{11k} - \sum_k m_{11k}\right)^2}{\sum_k v(n_{11k})}$$

$$m_{11k} = \frac{n_{1.k} n_{.1k}}{n_{..k}}$$

$$V(n_{11k}) = \frac{n_{1.k} n_{.1k} n_{2.k} n_{.2k}}{n_{..k}^3}$$

Where

n_{11k} is the observed cell count in the first row and first column of stratum k

m_{11k} is the expected cell count in the first row and first column of stratum k

$v(n_{11k})$ is the variance of the n_{11k}

for k = 1,2,3,4,5

The Mantel Haenszel test statistics is given as:

$$M^2 = \frac{\left(\sum_k n_{iik} - \sum_k m_{iik}\right)^2}{\sum_k v(n_{iik})}$$

$$m_{11k} = \frac{n_{1.k} n_{.1k}}{n_{..k}}, \quad v(n_{iik}) = \frac{n_{1.k} n_{.1k} n_{2.k} n_{.2k}}{\left(n^2_{..k}\left(n_{..k} - 1\right)\right)}$$

Where

n_{11k} is the observed cell count in the first row and first column of stratum k

m_{11k} is the expected cell count in the first row and first column of stratum k

$v(n_{11k})$ is the variance of the n_{11k}

for k = 1,2,3,4,5

The computation (7) m_{11k}

$$m_{11k} = 119.48$$

$$v(n_{11k}) = \frac{2706 \times 4587 \times 322 \times 6971}{7293^3} = 71.83$$ for Cochran

$$v(n_{11k}) = \frac{2706 \times 4587 \times 322 \times 6971}{\left(7293^2\left(7293 - 1\right)\right)} = 71.84$$ for mantel haenszel

The Cochran and Mantel Haenszel test of conditional independence is obtained as seen in table 4;

Table 4:Cochran and Mantel Haenzel test results

	Chi-squared	df	Asymp. Sig. (2 sided)
Cochran's	97.938	1	.000
Mantel-Haenzel	97.420	1	.000

Note. Cochran and Mantel Haenzel test results from Author (2020)

Under the conditional independence presumption, Cochran's measurement is asymptotically appropriated as a 1 level of opportunity (df) chi-squared dispersion as found in table 4, just if the quantity of layers is fixed, while the Mantel-Haenszel measurement is in every case asymptotically disseminated as a 1 level of opportunity (df) chi-squared circulation. Note that the coherence revision is eliminated from the Mantel-Haenszel measurement when the amount of the contrasts between the noticed and the normal is zero (0). Therefore, the survival status is dependent on the gender across the periods.

Loglinear Analysis

Model 1:
$$\ln\left(m_{ijk}\right)=\mu+\mu_{1(i)}+\mu_{2(j)}+\mu_{3(k)}+\mu_{13(ik)}+\mu_{23(jk)}$$

Table 5:Likelihood ratio and Pearson Chi-square results for model 1

	Chi-squared	df	Sig. (p-value)
Likelihood Ratio	106.584	5	.000
Pearson Chi-Square	112.014	5	.000

Note. Likelihood ratio and Pearson Chi-square results from Author (2020)

a. Model: Poisson
b. Design: Constant + Gender + Outcome + Year + Gender*Year + Outcome*Year

The loglinear analysis for model 1 was established in table 5, with a Poisson model with the design under the 'b' note in table 5

Model 2:
$$\ln\left(m_{ijk}\right)=\mu+\mu_{1(i)}+\mu_{2(j)}+\mu_{3(k)}+\mu_{12(ij)}+\mu_{13(ik)}+\mu_{23(jk)}$$

Table 6:Likelihood ratio and Pearson Chi-square results for model 2

	Chi-squared	df	Sig. (p-value)
Likelihood Ratio	12.102	4	.017
Pearson Chi-Square	12.176	4	.016

Note. Likelihood ratio and Pearson Chi-square results from Author (2020)

a. Model: Poisson
b. Design: Constant + Gender + Outcome + Year + Gender*Outcome + Gender*Year + Outcome*Year

The loglinear analysis for model 2 was established in table 6, with a Poisson model with the design under the 'b' note in table 6

Test of joint independence

Test of conditional independence of survival status on gender while controlling for the years

Likelihood ratio test comparison of model 1 and 2

G2 comparison $= G^2$ model 1 – G^2 model 2 with $(df_1 - df_2)df$

$= 106.584 - 12.102$

$= 94.482$

$\chi^2_{1(0.95)} = 3.847$, The difference is significant, which confirms earlier result from the Cochran and Mantel Haenszel of conditional dependence between gender and survival status while keeping the effect of year constant.

Table 7:Hierarchical Log linear Analysis of K-way and Higher order effects

	K	df	Likelihood Ratio Chi-Square	Likelihood Ratio Sig.	Pearson Chi-Square	Pearson Sig.	Number of Iterations
K-way and Higher Order Effects[a]	1	19	38323.629	.000	38658.705	.000	0
	2	13	283.146	.000	288.179	.000	2
	3	4	12.102	.017	12.176	.016	3
K-way Effects[b]	1	6	38040.483	.000	38370.527	.000	0
	2	9	271.044	.000	276.003	.000	0
	3	4	12.102	.017	12.176	.016	0

Note. Hierarchical Log linear Analysis of K-way results from Author (2020)

a. Tests that k-way and higher order effects are zero
b. Tests that k-way effects are zero

This shows that the joint interaction is significance as seen in table 7, therefore cannot be excluded. The saturated model is the resulting model of choice;
$$\ln\left(m_{ijk}\right)=\mu+\mu_{1(i)}+\mu_{2(j)}+\mu_{3(k)}+\mu_{12(ij)}+\mu_{13(ik)}+\mu_{23(jk)}+\mu_{123(ijk)}$$

Therefore, the variables are jointly dependent.

The conditional odds ratio showing the percentages of those who died of HIV is presented in table 8, shows the difference in the odds ratios of gender and survival status across the categories. This can also be further illustrated in figure 1

Table 8:Conditional Odds Ratio showing the percentages of those who died of HIV

Year	Gender	Percentage of dead
2010	Male	5.5%
	Female	3.7%
2011	Male	7.3%
	Female	5.3%
2012	Male	9.4%
	Female	6.1%
2013	Male	5.7%
	Female	4.1%
2014	Male	9.2%
	Female	4.6%

Note. Data for the Conditional Odds Ratio was computed bythe Author i.e. Taiwo (2020)

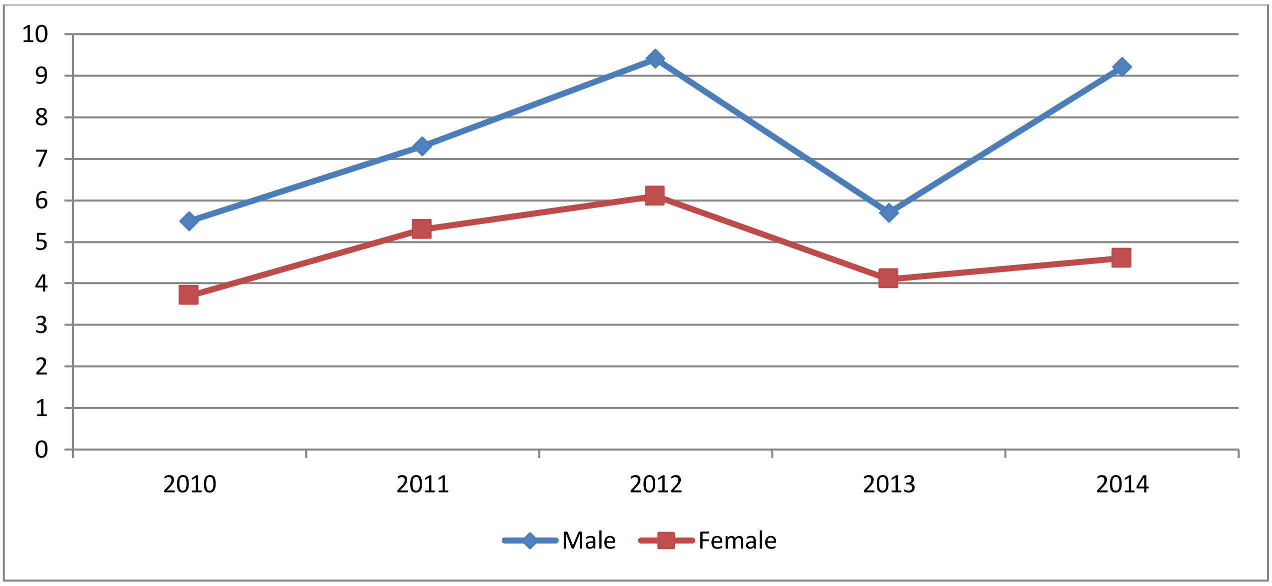

Figure 1: Line graph showing the proportion of male and female deaths across the years. Data for the proportion from the Author i.e. Taiwo (2020)

Table 9: Mantel Haenszel common odds ratio estimate

Estimate	1.612
ln(Estimate)	.478
Std. Error of ln(Estimate)	.049
Asymp. Sig. (2-sided)	.000
Asymp. 95% Common Odds Ratio Lower Bound	1.466
Confidence Interval Upper Bound	1.774

Note. Result was obtain from the analysis as computed by the Author (2020)

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.00 assumptions, also true for the natural log of the estimate as presented in table 9.

From the confidence interval, we can say that the odd of a male dying differs from 1. This indicates that the odd of a male dying of HIV across the periods are 0.612 times more likely than a female death.

IV. CONCLUSION

From the Breslow Day and Tarone tests, we discovered that the odds in each category differ significantly, which indicates a difference in the odds ratios of gender and survival status across the categories.

The Cochran and Mantel Haenszel tests indicated a dependence of survival status of HIV on gender while controlling for the effect of the years. This means that the odd of a person dying differs from one gender to the other in at least one of the categories.

The loglinear analysis which was used as a confirmatory analysis to that of the Cochran and Mantel Haenszel test is a superior test in the sense that it can be used to analyze data of any dimension unlike the Cochran and Mantel Haenszel tests that are only for $2 \times 2 \times k$ dimension. Also while the Cochran's test can only be used to investigate conditional independence, the loglinear analysis can be used investigate both conditional and joint independence across the variables.

The Cochran and Mantel Haenszel tests is a more robust test than the loglinear analysis in that they provide us with more accurate estimates using simple computations.

The significance of the Mantel Haenszel common odd ratio estimates indicates that the odds of a male dying are significantly different from 1. Therefore we can conclude from the data that the odd of a male dying of HIV is 0.612 times higher than that of a female.

REFERENCES

[1]. Agresti, A. (1990):***Categorical data analysis.*** Wiley, New York, new York.
[2]. Agresti, A. (1996):***An introduction to categorical data analysis, theory and practice.*** Wiley, New York, New York.
[3]. Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975):***Discrete multivariate analysis.***MIT press, Cabridge, MA.
[4]. Breslow, N.E. and Day, N.E. (1980). ***Statistical Methods in cancer research, 1. The france analysis of case control studies.***International Agency for research on cancer, Lyon,France.
[5]. Claus, H. (2014): Loglinear representation of the Mantel Haenszel and Breslow Day test. URL:www.google.com.ng/url?q=http://www.dgps.de/fachgruppen/methoden/mpr-online/issue12/art2/vonEye.pdf&sa=U&ei=EJsJVO_6OZChyACACA&ved=0CB8QFjAA&usg=AFQjCNFsG9AyFWmrKbkZs8BGwKuhhWXdvA (accessed on 9/4/2014).

[6]. Cochran, W.G, (1954): Some methods for strengthening the common chi square tests. International Biometrics society. Biometrics, 10 (4), 417-451.

[7]. Fidalgo, A.M., Mellengergh, G.J. and Muniz, J. (1998): Comparacion del procedimiento Mantel Haenszelfrente a los modeloslogline ales en la detecciondelfunctionamentodifferencial de los items. Psicothema, 10: 209-218.

[8]. Gardner, J., Pathnak, D. and Indurkya, A. (1999): Power calculations for detecting interactions in stratified 2X2 tables. Statistics and probability letters, 41: 267-275.

[9]. Goodman, L. A. (1970): The multivariate analysis of qualitative data: Interaction among multipleclassifications. J. Amer. Statist. Assoc.,65: 226-256.

[10]. Goodman, L. A. (1971a): The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications.

[11]. Technometrics,13: 330-1.

[12]. Goodman, L. A. (1971b): The partitioning of chi-square, the analysis of marginal contingencytables, and the estimation of expected frequencies in multidimensional contingency tables.J. Amer. Statist. Assoc.,66: 339-344.

[13]. Haberman, S. J. (1978, 1979):Analysis of Qualitative Data. AcademicPress, New York, New York.

[14]. Tabachnick, B. G. and Fidell, L.S. (1996):**Using multivariate statistics, 3rd edition.** Harper Collins, new York, New York, New York.

[15]. UN women (2014): Women more vulnerable to HIV/AIDS infection than men. URL:www.genderandaids.org/index.php?option=com_content&view=article&id=642:Women-more-vulnerable-to-HIV/AIDS-infection-than-men&catid=70:headlines&Itemid=83 (accessed on 9/4/2014).

[16]. Von Eye, A. and Spiel, C. (1996): Standard and non-standard loglinear symmetry models for measuring change in categorical variables. The American Statisticians, 50: 300-305.