JFHB Journal of Food Health and Bioenvironmental Science

Journal of Food Health and Bioenvironmental Science

Journal homepage : http://jfhb.dusit.ac.th/

Personal Variable Relationship of Cancer Patients in Thailand Using Log-linear Models

Pornpis Yimprayoon* & Sittipong Ruktamatakul

Faculty of Liberal Arts and Science, Kasetsart University, Kamphaeng Saen Campus, Nakhonpathom 73140, Thailand

Article info

Article history:
Received 10 April 2018
Revised 16 August 2018
Accepted 22 August 2018

Keywords:
Cancer, Log-Linear Model, Chi-square, Cramer's *V*-Value

Abstract

In this study, data sets of new cancer patients are used who were admitted between January 2012 and December 2013 at the National Cancer Institute Thailand. The main objective of this study is to test association between personal and cancer related variables using log-linear models. A test of independence is used to find the relationships between any two variables by chi-square tested and Cramer's *V*. The results in this study display that most paired variables of personal and cancer/clinical variables are significantly related at p-value < 0.01. For both male and female patients, the variable site of cancer is highly related to the stage of diagnosis, which provides the highest Cramer's *V* value. Moreover, the two-dimensional log-linear models are used to find estimated parameters, expected frequencies and standardized residuals obtaining the best model for the relationships between two defined variables.

Introduction

Cancer is a noninfectious disease, which is on the increase throughout the world and has become a serious problem of public health in many countries. Cancer is not only a problem for the patients but also affects their families and the community as a whole in terms of job loss, social isolation and family tension which may follow closely on the occurrence of cancer. The economic burden of cancer is most obvious in health care costs, such as for hospitals, other health services and drugs. Recent trends indicate that, without rigorous control measures, cancer can become the leading cause of death in many countries during the early part of the next century, with 300 million new cancer cases and 200 million deaths from cancer over a 25-year period. Almost two-thirds of all cancers will occur in the developing countries, which have only 5% of resources for cancer control. In the year 2000, there were approximately 10.1 million new cancer cases, 4.7 million in more developed countries and 5.4 million in less developed countries. For the year 2020, these figures are likely to reach 15.3, 6.0 and 9.3 million, respectively (World Health Organization, 2002) as shown in Table 1.

In Thailand, health problems have changed dramatically during the past two decades. Death due to communicable diseases, for example, tuberculosis and pneumonia have declined, while noncommunicable diseases such as cancer have become of greater importance. Most importantly, cancer has risen significantly to become a leading cause of death (Bureau of Policy and Strategy, 2015) as displayed in Figure 1. Moreover, this research looked at public health statistics in 2014 (Bureau of Policy and Strategy, 2015) and found

* Corresponding Author
e-mail: faasppy@ku.ac.th

Table 1 Numbers of cancer deaths and new cancer cases in the world as estimated for 2000 and predicted for 2020

Year	Region	Deaths (millions)	New cases (millions)
2000	More developed countries	2.6	4.7
	Less developed countries	3.6	5.4
	All countries	6.2	10.1
2020	More developed countries	3.5	6.0
	Less developed countries	6.3	9.3
	All countries	9.8	15.3

Remark: World Health Organization (2002)

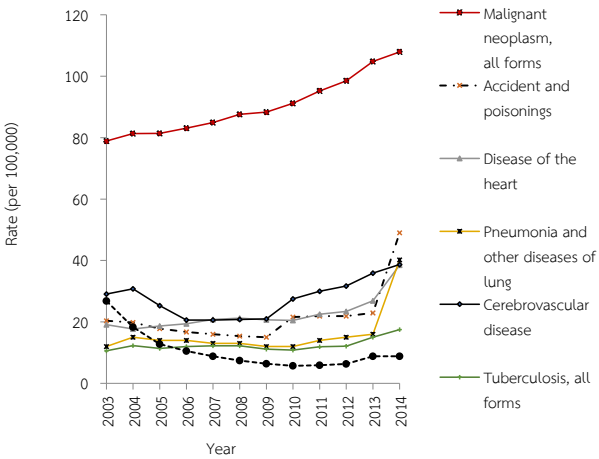

Fig. 1 Trends in the rates of mortality, Thailand, 2003-2014 (Bureau of Policy and Strategy, 2015)

that cancer that includes malignancy of various organs ranked first in cause of death as presented in Table 2.

The main objective of this study is to test the association between personal and cancer variables by using log-linear models. In this study, the data sets were the number of new cancer patients who were admitted between January 2012 and December 2013 at the National Cancer Institute Thailand. A test of independence was used to find the relationships between any two variables which are chi-square tested and Cramer's V (Bhattacharyya & Johnson, 1977; Marija, 1993) and two-dimensional log-linear models were applied to obtain estimated parameters, expected frequencies and standardized residuals (Nelder & Wedderburn, 1972; Nelder, 1974; Baker & Nelder, 1978; Fienberg, 1982; Aitkin et al., 1989; Dobson, 1996; Tiensuwan et al., 2000; Tiensuwan et al., 2005).

Materials and methods

1. Data collection

The subjects were new cancer patients treated at the National Cancer Institute Thailand. Researchers collected cancer data using a hospital-based cancer registry of annual report 2012-2013 from information and technology division (National Cancer Institute, 2014; 2015). The classification and coding of primary site and morphology in cancer notification form used the International Classification of Diseases for Oncology (World Health Organization, 1990). The data were the

Table 2 Death Rates per 100,000 population by leading cause of death, 2003-2014

Cause of Death	Rate											
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Malignant neoplasm, all forms	78.9	81.3	87.4	83.1	84.9	87.6	88.34	91.2	95.2	98.5	104.8	107.9
Accident and poisonings	56.9	58.9	57.6	59.8	57.5	55.1	55.63	56.1	52.8	51.6	50.2	49
Disease of the heart	27.7	26.8	28.2	28.4	29.3	29.8	28.96	28.9	31.4	32.9	38.1	38.5
Pneumonia and other diseases of lung	23.9	26.3	22.4	22	22.5	23	22.92	25.7	26.3	23.7	33.5	40.2
Hypertension and cerebrovascular disease	29.1	30.8	25.3	20.6	20.6	20.8	21	27.5	30	31.7	35.9	38.7
Tuberculosis, all forms	10.6	12.3	11.38	12	12.2	12.2	11.1	10.8	11.9	12.1	15	17.5
AIDS/HIV infection	26.8	18.3	12.8	15	8	7.4	6.4	5.7	5.9	6.3	8.8	8.8

number of new cancer patients treated between January 2012 and December 2013 at the National Cancer Institute Thailand. In this data set, there were 7,842 cancer patients, who consisted of 3,127 male patients and 4,715 female patients. The data were classified into two parts: 1) personal data and 2) cancer/clinical data as summarized in Table 3.

Table 3 Divisions within each category variable

Categories	Divisions
Personal data	
Sex	**Female; male**
Age (years)	**≥39; 40-44; 45-49; 50-54; 55-59; ≤60**
Region	**Bangkok and vicinities; sub-central; eastern; western; northern; northeastern; southern**
Cancer/clinical data	
Diagnostic evidence	**Clinical only; X-rays/endoscopy/radio diagnosis; specific biochem/immuno. test; cytology/ hematology; histology of metastasis; histology of primary**
Treatment	**Surgery; radiation; chemotherapy; surgery & radiation; surgery & chemotherapy; radiation & chemotherapy; surgery & radiation & chemotherapy; others (such as hormone, supportive, etc.)**
Stage of cancer	**Stage 1; stage 2; stage 3; stage 4; unknown**
Site of cancer (male)	**Trachea, bronchus & lung; colon & rectum; liver & intrahepatic bile ducts; lip & oral cavity; esophagus; others such as prostate gland, non-hodgkin's lymphoma, nasopharynx, oropharynx, thyroid gland, stomach, etc.)**
Site of cancer (female)	**Breast; cervix uteri; colon & rectum; trachea, bronchus & lung; liver & intrahepatic bile ducts; others (such as corpus uteri, ovary, non-hodgkin's lymphoma, nasopharynx, oropharynx, thyroid gland, stomach, etc.)**

In this study, a test of independence was applied to study the relationship between cancer variables incorporated with Cramer's *V* statistic (Bhattacharyya & Johnson, 1977; Marija, 1993) and the application of generalized linear models for use with two-dimensional log-linear models (Nelder & Wedderburn, 1972; Nelder, 1974; Baker & Nelder, 1978; Fienberg, 1982; Aitkin et al., 1989). These are employed to obtain estimates of parameter for the best models. The models indicate the associations between the personal and cancer/clinical variables.

2. Test of independence

Considering two-dimensional contingency tables with two variables, *A* and *B*, which have *J* and *K* categories respectively. Using the notation for the observed frequencies in the cell of a contingency table, this is formed by the cross-classification of variable *A* and variable *B* as in Table 4. It follows from Table 4 that Y_{jk} represents the random variables for the (j,k) cell, while y_{jk} represents the observed value of Y_{jk}. We denote the totals of *j*th row and *k*th column by Y_{j} and Y_{k}, respectively, and the overall total as $Y_{..}$.

Table 4 Notation for two-dimensional contingency tables

	B_1	B_2	. . .	B_K	Total
A_1	Y_{11}	Y_{12}	. . .	Y_{1K}	$Y_{1.}$
A_2	Y_{21}	Y_{22}	. . .	Y_{2K}	$Y_{2.}$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
A_J	Y_{J1}	Y_{J2}	. . .	Y_{JK}	$Y_{J.}$
Total	$Y_{.1}$	$Y_{.2}$	. . .	$Y_{.K}$	$Y_{..} = n$

The study considers the problems in which each member of a population can be classified according to two distinct variables which is denoted as the *A* variable and *B* variable. In testing the hypotheses that *A* and *B* are independent variables and the chi-squared statistic x^2 is used to measure goodness of fit for two-dimensional contingency tables. With an approximated significance level α, and degree of freedom $(J\text{-}1)(K\text{-}1)$, if $x^2 > x^2_{\alpha,(J-1)(K-1)}$, then the test is to reject H_0 and accept H_0 otherwise.

The chi-square statistic is not a good measure of the degree of association between two variables but its widespread use in tests of independence justifies its inclusion in this study. Cramer's V is one of those measures based on the chi-square that attempts to modify the chi-square statistic to minimize the influence of sample size and degrees of freedom as well as to restrict the range of values of the measure to those between 0 to 1 (Marija, 1993). The statistic, Cramer's V (Bhattacharyya & Johnson, 1977), is given as

$$V = \sqrt{\frac{\chi^2}{n\,(\mathrm{v}-1)}}$$

where v is the lesser of the number of rows and columns and the sample size *n*. This statistic, Cramer's *V*, can attain the maximum of 1 for tables of any dimension.

3. The Generalized linear model to use with the two-dimensional log-linear model

To formulate log-linear models (Dobson, 1996), the study uses the multinomial distribution equation. Under the model of independence

$$E(Y_{jk}) = n\theta_{j.}\theta_{.k},$$

the maximum likelihood estimators are

$$\theta_{j.} = \frac{y_{j.}}{n}$$

and

$$\theta_{.k} = \frac{y_{.k}}{n}.$$

Therefore the fitted cell frequencies are

$$e_{jk} = \frac{y_{j.}y_{.k}}{n}; j = 1, 2, \ldots, J, k = 1, 2, \ldots, K.$$

The terms of log-linear model under the independent models are

$$\eta_{jk} = \log E(Y_{jk}) = \mu + \alpha_j + \beta_k$$

where μ is the grand average of the logarithms of the expected frequencies

and α_j, β_k are the main effects of the factors

with $\sum_j \alpha_j = \sum_k \beta_k = 0.$

From the estimated

$\log e_{jk} = \hat{\eta}_{jk}$,

we have

$$\log y_{j.} + \log y_{.k} - \log n = \hat{\mu} + \hat{\alpha}_j + \hat{\beta}_k.$$

Thus the corresponding estimates in our case are

$$\hat{\mu} = \frac{1}{J}\sum_j \log y_{j.} + \frac{1}{K}\sum_k \log y_{.k} - \log n$$

$$\hat{\alpha}_j = \log y_{.j} - \frac{1}{J}\sum_j \log y_{j.}$$

and $\hat{\beta}_k = \log y_{.k} - \frac{1}{K}\sum_k \log y_{.k}.$

Under the maximal model

$$E(Y_{jk}) = n\theta_{jk}$$

and the maximum likelihood estimators of the θ_{jk} are

$$\theta_{jk} = \frac{y_{jk}}{n}.$$

The estimated expected cell frequencies are

$$e_{jk} = y_{jk}$$

and the terms of the log-linear model under the maximal model are

$$\eta_{jk} = \log E(Y_{jk}) = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$$

where μ is the grand average of the logarithms of the expected frequencies,

α_j, β_k are the main effects of the factors,

and $(\alpha\beta)_{jk}$ is the interaction of two effects subject to the constraints

$$\sum_j \alpha_j = \sum_k \beta_k = \sum_j \sum_k (\alpha\beta)_{jk} = 0.$$

From the estimate

$$\log e_{jk} = \hat{\eta}_{jk},$$

we have

$$\log y_{jk} = \hat{\mu} + \hat{\alpha}_j + \hat{\beta}_k + \hat{(\alpha\beta)}_{jk}.$$

So the estimated parameters in this model are

$$\hat{\mu} = \frac{1}{JK}\sum_j \sum_k \log y_{jk}$$

$$\hat{\alpha}_j = \frac{1}{K}\sum_k (\log y_{jk}) - \hat{\mu}$$

$$\hat{\beta}_k = \frac{1}{J}\sum_j (\log y_{jk}) - \hat{\mu}$$

and $\hat{(\alpha\beta)}_{jk} = \log y_{jk} - \hat{\mu} - \hat{\alpha}_j - \hat{\beta}_k.$

4. Goodness of fit

We can test the hypothesis to measure goodness of fit for two-dimensional contingency tables. The testing hypotheses are

H_0 : The model $\eta_{jk} = \mu + \alpha_j + \beta_k$ fits the data well.

H_1 : The model $\eta_{jk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$ fits the data well.

The statistics which are used to measure goodness of fit (Dobson, 1996) are:

(i) log-likelihood ratio statistic or deviance (Nelder & Wedderburn, 1972),

$$D = 2\sum_j \sum_k \left(y_{jk} \log\left(\frac{y_{jk}}{e_{jk}} \right) \right)$$

with degree of freedom (J-1)(K-1). An approximated significance level α, we reject H_0 if $D > \chi^2_{\alpha,(J-1)(K-1)}$ and accept H_0 otherwise;

(ii) chi-square statistic,

$$\chi^2 = \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\left(y_{jk} - e_{jk}\right)^2}{e_{jk}}$$

with degree of freedom (*J*-1)(*K*-1). The test is to reject H_0 if H_0 if $\chi^2 > \chi^2_{\alpha,(J-1)(K-1)}$ and accept otherwise.

Results and discussion

During the 24 months observation, January 2012-December 2013, there were 7,842 new cancer patients that consisted of 3,127 male patients and 4,715 female patients in the National Cancer Institute of Thailand. In this study, the statistical packages, SPSS for Windows and Microsoft EXCEL were used to analyze the cancer data and to estimate parameters, fitted and residual values of random variables of two-dimensional log-linear models. The number and percentage of variables in personal data and cancer/clinical data are presented in Table 5. In this section, the results from testing associations between two distinct cancer variables and cancer/clinical variables are presented. The chi-square values, degree of freedom, Cramer's V and deviance between these variables are presented in Table 6.

As indicated in Table 6, at significance level 0.01, it finds that sex is related to age, diagnostic evidence and treatment. The variable age has a relationship with stage of cancer while site of cancer is associated to age, region and stage of cancer for both male and female patients. Moreover, Table 6 shows that the site of cancer

Table 5 The number and percentage of personal data and cancer/clinical data classify by sex

Variables	Male		Female	
	Number	Percent	Number	Percent
Age(years)				
≤ 39	280	8.95	583	12.36
40-44	200	6.40	491	10.41
45-49	287	9.18	689	14.61
50-54	409	13.08	709	15.04
55-59	460	14.71	686	14.55
≥ 60	1491	47.68	1557	33.02
Region				
Bangkok and vicinities	1528	48.86	2309	48.97
Sub-central	147	4.70	204	4.33
Eastern	203	6.49	310	6.57
Western	602	19.25	843	17.88
Northern	252	8.06	426	9.03
Northeastern	275	8.79	428	9.08
Southern	120	3.84	195	4.14
Diagnostic evidence				
Clinical only	42	1.34	105	2.23
X-rays/endoscopy/radio diagnosis	485	15.51	317	6.72
Specific biochem/immuno. Test	28	0.90	19	0.40
Cytology/hematology	69	2.21	51	1.08
Histology of metastasis	193	6.17	183	3.88
Histology of primary	2310	73.87	4040	85.68
Treatment				
Surgery	153	4.89	328	6.96
Rradiation	275	8.79	350	7.42
Chemotherapy	391	12.50	714	15.14
Surgery & radiation	69	2.21	126	2.67
Surgery & chemotherapy	43	1.38	89	1.89
Radiation & chemotherapy	99	3.17	355	7.53
Surgery & radiation & chemotherapy	327	10.46	433	9.18
Others	1770	56.60	2320	49.20
Stage of cancer				
Stage 1	152	4.86	664	14.08
Stage 2	384	12.28	1223	25.94
Stage 3	729	23.31	1214	25.75
Stage 4	1451	46.40	1147	24.33
Unknown	411	13.14	467	9.90
Site of cancer				
Breast	6	0.19	1863	39.51
Cervix uteri	0.00	0.00	696	14.76
Trachea, bronchus & lung	542	17.33	315	6.68
Colon & rectum	445	14.23	398	8.44
Liver & intrahepatic bile ducts	436	13.94	172	3.65
Lip & oral cavity	195	6.24	132	2.80
Esophagus	189	6.04	20	0.42
Others	1314	42.02	1119	23.73

for female patients and stage of cancer have a strong relationship at the significance level 0.01. So for these relationships we obtained 10 models of the form:

$$\hat{\eta}_{jk} = \hat{\mu} + \hat{\alpha}_j + \hat{\beta}_k + \widehat{(\alpha\beta)}_{jk}.$$

Table 6 Summary of the relationships between personal variables and cancer/ clinical variables

Variable	Chi-square	d.f.	Cramer's *V*	Deviance	Conclusion
Sex					
Age	**207.97**	**5**	**0.16**	**209.74**	**Reject H_0**
Region	**5.19**	**6**	**0.03**	**5.20**	**Accept H_0**
Diagnostic evidence	**225.90**	**5**	**0.17**	**221.58**	**Reject H_0**
Treatment	**116.06**	**7**	**0.12**	**121.66**	**Reject H_0**
Age					
Stage of cancer	**191.95**	**20**	**0.08**	**179.44**	**Reject H_0**
Site of cancer (male)					
Age	**168.98**	**25**	**0.10**	**67.78**	**Reject H_0**
Region	**97.56**	**30**	**0.08**	**93.10**	**Reject H_0**
Stage of cancer	**347.02**	**20**	**0.17**	**350.04**	**Reject H_0**
Site of cancer (female)					
Age	**302.38**	**25**	**0.11**	**309.08**	**Reject H_0**
Region	**74.87**	**30**	**0.06**	**73.86**	**Reject H_0**
Stage of cancer	**1181.08**	**20**	**0.25**	**1181.90**	**Reject H_0**

From Table 6, this study chose and presents only one association between two variables site of cancer for female patients and stage of cancer (Table 7) which gave the maximum Cramer's V value to formulate the two-dimensional log-linear model (maximal model) $\hat{\eta}_{jk} = \hat{\mu} + \hat{\alpha}_j + \hat{\beta}_k + (\hat{\alpha\beta})_{jk}$, $j = 1, 2, \ldots, 6$; $k = 1, 2, \ldots, 5$. The estimates of the main effects of interaction terms of the maximal model are shown in Table 8. The other models were obtained in the same manner. Whereas the expected frequency and the observed frequency are the same, since the best model obtained is the maximal model.

Further, from Table 6 for the relationship between region and sex, we accept H0 at significance level 0.01. So that the best model is the model of independence: $\hat{\eta}_{jk} = \hat{\mu} + \hat{\alpha}_j + \hat{\beta}_k$, $j = 1, 2, \ldots, 7$; $k = 1, 2$. Table 9 presents the two-dimensional contingency table for this association. The estimates of parameters are as indicated in Table 10, while observed frequency, expected frequency and standardized residual are shown in Table 11.

Table 8 Estimates of parameters of the maximum model for site of cancer for female patients and stage of cancer

Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
μ	4.410	$(\alpha\beta)_{13}$	-0.008	$(\alpha\beta)_{42}$	-0.502
α_1	1.312	$(\alpha\beta)_{14}$	-0.965	$(\alpha\beta)_{43}$	-0.068
α_2	0.269	$(\alpha\beta)_{15}$	-0.494	$(\alpha\beta)_{44}$	1.240
α_3	-0.438	$(\alpha\beta)_{21}$	1.099	$(\alpha\beta)_{45}$	-0.018
α_4	-0.947	$(\alpha\beta)_{22}$	0.455	$(\alpha\beta)_{51}$	-0.877
α_5	-1.281	$(\alpha\beta)_{23}$	0.361	$(\alpha\beta)_{52}$	-0.259
α_6	1.085	$(\alpha\beta)_{24}$	-1.103	$(\alpha\beta)_{53}$	-0.090
β_1	-0.865	$(\alpha\beta)_{25}$	-0.811	$(\alpha\beta)_{54}$	0.676
β_2	0.175	$(\alpha\beta)_{31}$	-1.161	$(\alpha\beta)_{55}$	0.550
β_3	0.518	$(\alpha\beta)_{32}$	0.184	$(\alpha\beta)_{61}$	0.835
β_4	0.673	$(\alpha\beta)_{33}$	0.363	$(\alpha\beta)_{62}$	-0.588
β_5	-0.500	$(\alpha\beta)_{34}$	0.325	$(\alpha\beta)_{63}$	-0.556
$(\alpha\beta)_{11}$	0.757	$(\alpha\beta)_{35}$	0.290	$(\alpha\beta)_{64}$	-0.173
$(\alpha\beta)_{12}$	0.710	$(\alpha\beta)_{41}$	-0.652	$(\alpha\beta)_{65}$	0.483

Table 9 The frequencies (percent) of region and sex

Region	Sex		Total
	Male	Female	
Bangkok and vicinities	**1528 (19.48)**	**2309 (29.44)**	**3837 (48.93)**
Sub-central	**147 (1.87)**	**204 (2.60)**	**351 (4.48)**
Eastern	**203 (2.59)**	**310 (3.95)**	**513 (6.54)**
Western	**602 (7.68)**	**843 (10.75)**	**1445 (18.43)**
Northern	**252 (3.21)**	**426 (5.43)**	**678 (8.65)**
Northeastern	**275 (3.51)**	**428 (5.46)**	**703 (8.96)**
Southern	**120 (1.53)**	**195 (2.49)**	**315 (4.02)**
Total	**3127 (39.88)**	**4715 (60.12)**	**7842 (100.00)**

Table 7 The frequencies (percent) of site of cancer for female patients and stage of cancer

Site of cancer (female)	Stage of cancer					Total
	1	2	3	4	Unknown	
Breast	**274 (5.81)**	**740 (15.69)**	**508 (10.77)**	**228 (4.84)**	**113 (2.40)**	**1863 (39.51)**
Cervix uteri	**136 (2.88)**	**202 (4.28)**	**259 (5.49)**	**70 (1.48)**	**29 (0.62)**	**696 (14.76)**
Colon & rectum	**7 (0.15)**	**76 (1.61)**	**128 (2.71)**	**144 (3.05)**	**43 (0.91)**	**398 (8.44)**
Trachea, bronchus & lung	**7 (0.15)**	**23 (0.49)**	**50 (1.06)**	**216 (4.58)**	**19 (0.40)**	**315 (6.68)**
Liver & intrahepatic bile ducts	**4 (0.08)**	**21 (0.45)**	**35 (0.74)**	**88 (1.87)**	**24 (0.51)**	**172 (3.65)**
Others*	**236 (5.01)**	**161 (3.41)**	**234 (4.96)**	**401 (8.50)**	**239 (5.07)**	**1271 (26.96)**
Total	**664 (14.08)**	**1223 (25.94)**	**1214 (25.75)**	**1147 (24.33)**	**467 (9.90)**	**4715 (100.00)**

Remark: *such as corpus uteri, ovary, non-hodgkin's lymphoma, etc.

Table 10 Estimates of main effect parameters of region and sex

Parameter	Estimate	Parameter	Estimate
μ	5.923	α_5	-0.117
α_1	1.616	α_6	-0.081
α_2	-0.776	α_7	-0.884
α_3	-0.396	β_1	-0.205
α_4	0.639	β_2	0.205

Table 11 Observed frequency, expected frequency and standardized residual of region and sex per cell

Cell		Observed (Expected under H_1)	Expected under H_0)	Standardized residual
i	j			
1	1	1528	1530.00	-0.05
1	2	2309	2307.00	0.04
1	3	147	139.96	0.60
1	4	204	211.04	-0.48
1	5	203	204.56	-0.11
1	6	310	308.44	0.09
1	7	602	576.19	1.08
2	1	843	868.81	-0.88
2	2	252	270.35	-1.12
2	3	426	407.65	0.91
2	4	275	280.32	-0.32
2	5	428	422.68	0.26
2	6	120	125.61	-0.50
2	7	195	189.39	0.41

Conclusion

According to the results, it is shown that the test of independence can only show whether any two variables are related and cannot find the expected frequency of each parameter. This can be found from the estimates of parameters under the independence model (model under H_0) and the maximal model (model under H_1) by using the two-dimensional log-linear models. The expected frequencies of the maximal model are the observed values. In fact, any pair of relative or non-relative variables can be formulated in the log-linear model. For this study, only relative variables are presented which gives the maximum Cramer's V value in each table of relationship.

The results of the general data shows that most patients in this study are female than male. Most patients are greater than 59 years old, live in Bangkok and vicinities of Thailand and are diagnosed by using the histology of primary. A large number of female patients are diagnosed with breast cancer. In contrast, trachea, bronchus and lung cancer had the greatest number of male patients. Most male patients are in stage 4, while most female patients are in stage 2. Most cancer patients are treated by using chemotherapy.

From the test of independence, there are several variables which have a relationship at significance level 0.01, that is, the variable sex has a relationship with age, diagnostic evidence and treatment. Age is also associated to stage of cancer. For both male and female patients, the variable site of cancer is highly related to stage of cancer, which provides the highest Cramer's V value. Moreover, it was found that age and region are related to site of cancer for both sexes.

Finally, from the numerical results of the two-dimensional log-linear models for each related variable, we reject H_0 at significance level 0.01. This lead to the best model as the maximal model: $\hat{n}_{jk} = \hat{\mu} + \hat{\alpha}_j + \hat{\beta}_k + (\hat{\alpha\beta})_{jk}$, which shows interaction terms of variables that have a relationship.

Acknowledgements

This research received funding from Faculty of Liberal Arts and Science, Kasetsart University, Kamphaeng Saen Campus Research

References

Aitkin, M., Anderson, D., Frances, B., & Hinde, J. (1989). *Statistical Modelling in GLIM.* Oxford: Oxford Science Publication.

Baker, R.J., & Nelder, J.A. (1978). *GLIM Manual (Release 3).* Oxford: Oxford University Press.

Bhattacharyya, G.K., & Johnson, R. (1977). *Statistical Concepts and Methods.* New York: Wiley.

Bureau of Policy and Strategy. (2015). *Public Health Statistics AD 2014.* Bangkok: Ministry of Public Health.

Dobson, A.J. (1996). *An Introduction to Generalized Linear Models.* London: Chapman and Hall.

Fienberg, S.E. (1982). *The Analysis of Cross-Classified Categorical Data.* Massachusetts: MIT Press.

Marija, J. (1993). *SPSS for Windows Base System User's Guide Release 7.5.* Chicago: SPSS.

National Cancer Institute. (2014). *Annual Report of Hospital-Based Cancer Registry AD 2012.* Bangkok: Ministry of Public Health.

National Cancer Institute. (2015). *Annual Report of Hospital-Based Cancer Registry AD 2013*. Bangkok: Ministry of Public Health.

Nelder, J.A., & Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society (Series A), 135,* 370-84.

Nelder, J.A. (1974). Log-linear models for contingency tables: a generalization of classical least squares. *Applied Statistics, 23,* 323-29.

Tiensuwan, M., Lertprapai, S., Sirichaisinthop, J., & Lawmepol, A. (2000). Application of log-linear models to malaria patients in Thailand. *Statistics in Medicine, 19*(14), 1931-1945.

Tiensuwan, M., Yimprayoon, P., & Lenbury, Y. (2005). Application of Log-linear Models to Cancer Patients: A Case Study of Data from the National Cancer Institute of Thailand. *Southeast Asian Journal of Tropical Medicine and Public Health, 36*(5), 1283-1291.

World Health Organization. (1990). *International Classification of Diseases for Oncology.* Geneva: WHO.

World Health Organization. (2002). *National Cancer Control Programmes.* Geneva: WHO.