



MATHEMATICAL SCIENCES

Assessment of SNP-SNP interactions by using square contingency table analysis

ÖZGE KARADAĞ, GÖKÇEN ALTUN & SERPİL AKTAŞ

Abstract: The evolution of SNP-SNP interactions has become an interesting field in genetic epidemiology. Most of the studies, aimed to analyze the relationship between genetic factors and disease of interest, are focused on single SNP associations. However, for quantitative traits, influenced by the interplay of environmental and more than one genetic factors, interaction between the multi factors should be taken into consideration. In this study, symmetry models for square contingency tables are applied to the cross-classified SNP-SNP interactions data. Results from a genome-wide association analysis of blood pressure are used as a prior evidence for the interacted SNPs.

Key words: complete symmetry, marginal homogeneity, quasi-symmetry, SNP-SNP interaction analysis, square contingency tables.

INTRODUCTION

A single nucleotide polymorphism (SNP) is defined as the variation at a single position in a DNA sequence among individuals. DNA sequence is formed from a chain of four nucleotide bases, namely, A, C, G, and T. For example, the variation is classified as a SNP, when a substitution of a T for a A in the nucleotide sequence GGA**A**TCG, consequently turning out the sequence GGAT**T**CG.

SNP-SNP interaction is generally defined as the interaction between different loci and, the statistical interaction analysis is expected to explain the etiology of many complex human traits such as diabetes, hypertension and asthma. In recent years, in spite of the increased number of genome-wide association analysis, interaction analysis of these genome-wide research are still few in number. As a major drawback, in genome-wide association analysis, the main focus is usually the single SNP associations. However, the effect of one locus is masked by the effects at another locus or the joint effect two SNPs may be significant whereas they are ineffective separately. Thus, the interactions play an important role in explaining the missing heritability of complex diseases.

A common and simple way for detecting the interaction between two SNPs is to use several types of regression techniques, with straightforward implementation of interaction analysis (Hartwing 2013). Moreover, when there is more than two SNPs, classical statistical approaches may lack power due to the high dimensionality. Vaidyanathan et al. (2017) followed up a clustering approach based on pairing up the SNPs based on similarity of genetic identity and they carried on the analysis by conducting a standard case control association test using Cochran-Mantel-Haenszel test to analyze the SNP-disease association.

On the other hand, a case control study is a common study design used for testing the association by comparing the frequency of SNP alleles in cases who have been diagnosed with the disease under study and controls who known to unaffected (Lewis 2002, Clarke et al. 2011). Contingency table analysis methods allow alternative models of genetic relationship by summarizing the counts in different ways. For example, Haber (1982) performed intraclass contingency table analysis for testing the independence under the assumptions of marginal homogeneity, quasi-independence and symmetry by intercrossing maternally and paternally inherited genes. A comparative chi-square analysis was applied by Song et al. (2014) to screen the large gene expression data for conserved and differential gene interactions.

The goal of this paper is to provide an alternative approach to analysis the cross-classified SNP- SNP interactions with symmetry models. The symmetry models described in the literature are complete symmetry, quasi-symmetry and marginal homogeneity models. Using these symmetry models, we explore the most suitable symmetry structure among the SNP-SNP pairs that have been found to have a relationship in the preliminary study. These symmetry models are applied to 24 SNP-SNP pairs and the most relevant SNP-SNP-pair is obtained.

The remaining part of the paper contains the description of genetic association and its theoretical properties, symmetry models with their theoretical background, a real life data application and some concluding remarks, respectively.

GENETIC ASSOCIATION

Genetic association studies are used for testing the relationship between the phenotype of interest and the genetic variant. Phenotype can be defined as the observable properties of an organism that are produced by the interaction of the genetic and environmental variants. In genetic epidemiology, the phenotype of interest is usually obtained as a disease status or a continuous indicator. In addition to this, in most of the association studies, SNPs are considered as the genetic variants. As above-mentioned, SNP is the variation at a single position in a DNA sequence among individuals and usually coded in genotype as a combination of alleles. Considering a SNP consisting of a single bi-allelic locus with alleles a and A . Then, the SNP can be characterized by three different possible categories, aa , aA and AA .

Testing genetic association is performed by using different statistical methods depending on the structure of the phenotype. For continuously measured phenotypes such as blood pressure measurements, linear models are useful tools. When the binary phenotype (0/1) is a case and control, then a logistic regression model can be used to detect the relationship between the trait and the genetic variation

In genetic epidemiology, case control studies are widely used designs since they allow contingency table analysis as a result of the categorical structure of the genotype (Slager & Schaid 2001, Velez et al. 2016). Under the null hypothesis of no association with the disease, the genotype frequencies are expected to be the same in case and controls. A contingency table can be analyzed by using standard test statistics that measure the divergence of the observed frequencies from the expected ones under the null hypothesis of no association.

For a single bi-allelic SNP with alleles *a* and *A* tested in a case control study, the data generated consist of six counts of the numbers of genotypes (*aa*, *aA*, *AA*) in cases and controls. In case of interaction of two SNPs, structure of the table transforms from 2x3 (Table I.(a)) to 3x3 (Table I.(b)) for cases and controls, separately.

Table I. (a) Contingency tables of SNP-U and SNP-V (b) Contingency tables for interaction of two SNPs for cases and controls.

SNP-U	<i>aa</i>	<i>aA</i>	<i>AA</i>	Total
Case	$n_{1.}$	$n_{2.}$	$n_{3.}$	n
Control	$m_{1.}$	$m_{2.}$	$m_{3.}$	m
Total	$n_{A.1}$	$n_{A.2}$	$n_{A.3}$	N

Cases	U \ V	<i>aa</i>	<i>aA</i>	<i>AA</i>	Total
	<i>aa</i>	n_{11}	n_{12}	n_{13}	$n_{1.}$
<i>aA</i>	n_{21}	n_{22}	n_{23}	$n_{2.}$	
<i>AA</i>	n_{31}	n_{32}	n_{33}	$n_{3.}$	
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	n	

SNP-V	<i>aa</i>	<i>aA</i>	<i>AA</i>	Total
Case	$n_{.1}$	$n_{.2}$	$n_{.3}$	n
Control	$m_{.1}$	$m_{.2}$	$m_{.3}$	m
Total	$n_{B.1}$	$n_{B.2}$	$n_{B.3}$	N

Controls	U \ V	<i>aa</i>	<i>aA</i>	<i>AA</i>	Total
	<i>aa</i>	m_{11}	m_{12}	m_{13}	$m_{1.}$
<i>aA</i>	m_{21}	m_{22}	m_{23}	$m_{2.}$	
<i>AA</i>	m_{31}	m_{32}	m_{33}	$m_{3.}$	
Total	$m_{.1}$	$m_{.2}$	$m_{.3}$	m	

In Table I, *n* is the number of cases, *m* is the number of control and $N = n + m$ is the total number of patients.

For the interaction analysis of SNPs, square contingency table which is a special case of contingency tables, could be obtained as in the above tables. Square contingency tables that arise in dependent samples where the row and column variables have same level. Some specific models should be used in the analysis of these kinds of tables. These models are mostly in the symmetrical pattern that represents the symmetric structure of tables.

In this study, three different symmetry models are considered which are employed in the case of complete symmetry, quasi symmetry and marginal homogeneity.

SYMMETRY MODELS

Let n_{ij} be the observed frequency in the cell (*i,j*) and p_{ij} denotes the probability of the corresponding cell. Then, the complete symmetry (S) model is defined by;

$$p_{ij} = p_{ji}, \quad i, j = 1, 2, \dots, R, \quad (\text{for } i \neq j) \tag{1}$$

and, the S model is based on $R(R - 1) / 2$ degrees of freedom (df), where *R* is the dimension of the square table (Goodman 1985, Bishop et al. 1975).

This model indicates that the probability that an observation will fall in cell (i,j) is equal to probability that it falls in symmetric cell (j,i). In addition, as an extended model of the S model, Quasi Symmetry (QS) model is defined by;

$$p_{ij}p_{jk}p_{ki} = p_{ji}p_{kj}p_{ik}, \quad 1 \leq i < j < k \leq R \tag{2}$$

The QS model has $(R - 1)(R - 2) / 2$ df (Yamaguchi 1990). QS model indicates that equality of odds ratio on one side of the main diagonal and the other side.

Other extended model of the S model, the Marginal Homogeneity (MH) model is defined by;

$$p_{i.} = p_{.i}, \quad i, j = 1, 2, \dots, R \tag{3}$$

where $p_{i.} = \sum_{t=1}^R p_{it}$, $p_{.i} = \sum_{s=1}^R p_{si}$. The model has $(R - 1)$ df (Stuart 1955, Tahata et al. 2008). This model indicates that the row marginal distribution is identical to the column marginal distribution.

In the genetic field, for testing the interaction between two separate bi-allelic SNPs, symmetry models can be used and it can be interpreted as the similar genotype distribution occurs in SNP-1 and SNP-2.

Let $p_{ij} = n_{ij}/n$ denotes the probability of an individual having i^{th} genotype level for the SNP-1 and j^{th} genotype level for the SNP-2, $i = 1, 2, 3$. In terms of the S model, the null hypothesis states that there are no differences between $p_{21} = p_{12}$ for genotypes "aa" and "aA", $p_{31} = p_{13}$ for genotypes "AA" and "aA" and $p_{32} = p_{23}$ for genotypes "aa" and "AA". For the QS model, the null hypothesis of no difference is the statement of $p_{12}p_{23}p_{31} = p_{21}p_{32}p_{13}$.

Let $p_{i.} = n_{i.}/n$ denotes the probability of i^{th} genotype level for the SNP-1 and $p_{.j} = n_{.j}/n$ denotes the probability of j^{th} genotype level for the SNP-2. The null hypothesis for the MH model tests the differences between, namely $p_{.1} = p_{1.}$ for genotype "aa", $p_{.2} = p_{2.}$ for genotype "aA" and $p_{.3} = p_{3.}$ for genotype "AA".

The Maximum Likelihood estimates of expected values e_{ij} under S model is

$$\hat{e}_{ij} = \begin{cases} \frac{n_{ij} + n_{ji}}{2} & i \neq j \\ n_{ij} & i = j \end{cases} \tag{4}$$

The likelihood equations for the QS model are defined as;

$$\begin{aligned} \hat{e}_{i.} &= n_{i.} \quad \text{and} \quad \hat{e}_{.j} = n_{.j} \quad i = 1, 2, 3 \\ \hat{e}_{ij} + \hat{e}_{ji} &= n_{ij} + n_{ji} \quad i \neq j \end{aligned} \tag{5}$$

Note that marginal homogeneity is not equivalent to a log linear model and for studying marginal homogeneity (Agresti 2002). The MH model assumes that summation of marginal is symmetric whereas the structure of table is non-symmetric. When $\alpha = (p_{12}p_{23}p_{31}) / (p_{21}p_{32}p_{13})$ equals to 1, QS is equivalent to Caussinus (1965) showed that S is equivalent to QS and MH holding simultaneously. Thus, the distribution of a SNP-SNP interaction that satisfies both QS and MH, also satisfies S. Vice versa also holds, $S \equiv QS \cap MH$.

The cell distribution of the parameters under symmetry models can be represented in a matrix format. Let S_{ij} denotes the element of a design matrix S in row i and column j :

$$S_{ij} = \begin{cases} (k + 1) - (i + 1) \left(\frac{1}{2}i + 1\right) + (R + 3)(i + 1) - 3 - 2l & i \leq j \\ (k + 1) - (j + 1) \left(\frac{1}{2}j + 1\right) + (R + 3)(j + 1) - 3 - 2l & i > j \end{cases} \quad (6)$$

where, $k = |i - j|$ (Lawal & Sundheim 2002, Efendioğlu 2015).

Thus, for a bi-allelic SNP-SNP interaction, the S matrix and the corresponding R and C matrices are given below,

$$S = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix} \quad R = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix} \quad (7)$$

where, R is the row matrix, C is the column matrix.

Under the null hypothesis that no interaction exists between the SNPs, test statistic follows a chi-square distribution with associated df. The likelihood ratio test statistic equals

$$G^2 = \sum_{i=1}^R \sum_{j=1}^R n_{ij} \log \left(\frac{n_{ij}}{\hat{e}_{ij}} \right) \quad (8)$$

Several models may fit to the data in the square contingency table. In such cases, the model selection process refers to the selection of the best fitting model among the models. For model selection, ranking information criteria is a common way. The well-known information criteria is the Akaike's Information Criterion (AIC) that might be used for the model selection:

$$AIC = G^2 - 2df \quad (9)$$

The model having the smallest AIC value gives the best fitting model (Akaike 1974).

REAL DATA APPLICATION

Hypertension and relatedly, the abnormal levels of blood pressure, are the cardiovascular risk factors. In this paper, to evaluate the performance of the symmetry models in testing SNP-SNP interaction, as a prior knowledge, the results of a genome wide analysis are analysed (Karadağ & Aktaş 2018). The 24 top associated SNPs are detected by using a multilevel latent class modelling approach considering familial and serial correlations.

Table II includes the genetic position of the variants with a chromosome number (Chr) and a chromosomal position. The number of individuals having recoded genotypes levels of 1, 2, 3 are n_1, n_2 and n_3 , respectively. The association results are summarized in Table II, where b denotes the effect of variant and $se(b)$ is the standard error of parameter. More detailed information about the association analysis can be found in Karadağ & Aktaş 2018.

For the interaction analysis of SNPs, 3x3 square tables are generated by using number of individuals, n_1, n_2 and n_3 given in Table II. The possible number of two-way interactions is 276 for 24

Table II. Results for top associated 25 variants (listed according to chromosomal position).

No	Chr	Position	n ₁	n ₂	n ₃	b	se(b)	χ ²	p-value
1	1	87705031	447	340	62	0.293	0.063	21.995	2.73x10 ⁻⁰⁶
2	1	94479725	566	252	31	-0.332	0.073	20.717	5.32x10 ⁻⁰⁶
3	3	60483876	618	218	13	0.361	0.081	19.988	7.79x10 ⁻⁰⁶
4	5	110501261	630	194	25	-0.366	0.082	19.918	8.08x10 ⁻⁰⁶
5	5	110506386	630	194	25	-0.366	0.082	19.918	8.08x10 ⁻⁰⁶
6	5	110515510	628	196	25	-0.374	0.082	20.839	4.99x10 ⁻⁰⁶
7	5	110529043	610	213	26	-0.361	0.079	20.728	5.29x10 ⁻⁰⁶
8	5	110543925	624	200	25	-0.374	0.081	21.244	4.04x10 ⁻⁰⁶
9	7	11766844	208	449	192	-0.263	0.057	21.390	3.74x10 ⁻⁰⁶
10	7	11766964	208	448	193	-0.264	0.057	21.614	3.33x10 ⁻⁰⁶
11	7	52257371	226	436	187	-0.267	0.058	21.232	4.06x10 ⁻⁰⁶
12	7	144982124	733	115	1	0.524	0.114	21.142	4.26x10 ⁻⁰⁶
13	9	10749312	368	376	105	-0.268	0.059	20.909	4.81x10 ⁻⁰⁶
14	11	129152516	313	421	115	-0.272	0.058	21.929	2.82x10 ⁻⁰⁶
15	11	129155346	311	420	118	-0.264	0.058	20.838	4.99x10 ⁻⁰⁶
16	13	105385655	745	102	2	0.522	0.116	20.337	6.49x10 ⁻⁰⁶
17	17	15291162	258	437	154	-0.259	0.057	20.917	4.79x10 ⁻⁰⁶
18	19	47351966	621	213	15	-0.382	0.079	22.895	1.71x10 ⁻⁰⁶
19	19	48295716	686	159	4	-0.455	0.099	21.214	4.10x10 ⁻⁰⁶
20	19	48296296	686	159	4	-0.455	0.099	21.214	4.10x10 ⁻⁰⁶
21	19	48296505	687	158	4	-0.465	0.099	21.990	2.73x10 ⁻⁰⁶
22	19	48297228	684	161	4	-0.447	0.099	20.557	5.78x10 ⁻⁰⁶
23	21	40292951	450	346	53	-0.303	0.066	21.323	3.88x10 ⁻⁰⁶
24	21	40295176	453	339	57	-0.318	0.065	24.132	8.99x10 ⁻⁰⁷

variants. In Table III, likelihood ratio test statistic G^2 and p-values are summarized for only significant pairs under every three symmetry models.

According to the results in Table III, as an example of a significant interaction, we can say that the pair 1-24 fits to whole symmetry structures depending on the p-values, 0.970, 0.900 and 0.892, respectively ($p > 0.05$). The distribution of observed counts is given by Table IV.(a) and the expected counts under S, QS and MH are given by Table IV.(b), Table IV.(c) and Table IV.(d), respectively.

Table III. Results under symmetry models.

SNP Pair	Complete Symmetry		Quasi Symmetry		Marginal Homogeneity	
	G ²	p-value	G ²	p-value	G ²	p-value
1 \ 23	0.765	0.580	0.027	0.869	0.740	0.691
1 \ 24	0.245	0.970	0.016	0.900	0.230	0.892
2 \ 7	6.415	0.093	0.320	0.571	6.120	0.050
3 \ 4	6.785	0.079	1.477	0.224	5,270	0.072
3 \ 5	6.785	0.079	1.477	0.224	5.270	0.072
3 \ 6	6.480	0.090	1.437	0.231	5.000	0.082
3 \ 7	5.702	0.127	1.268	0.260	4.370	0.112
3 \ 8	7.298	0.063	2.728	0.099	4.530	0.104
3 \ 18	0.704	0.872	0.469	0.481	0.210	0.901
4 \ 18	3.525	0.317	0.176	0.675	3.330	0.189
5 \ 18	3.525	0.317	0.176	0.675	3.330	0.189
6 \ 18	3.306	0.347	0.158	0.691	3.130	0.209
7 \ 18	3.356	0.340	0.147	0.702	3.180	0.204
8 \ 18	2.953	0.399	0.126	0.723	2.810	0.246
9 \ 11	1.328	0.723	0.301	0.583	1.030	0.598
10 \ 11	1.369	0.713	0.337	0.562	1.030	0.597
12 \ 16	4.813	0.816	3.653	0.056	1.150	0.362
14 \ 15	6.931	0.074	0.000	0.988	5.030	0.081
23 \ 24	7.473	0.058	0.000	0.998	5.840	0.054

(a)

	24	AA	AG	GG
1				
AA		207	201	39
AG		204	119	17
GG		42	19	1

Table IV. (a) Observed counts, (b) expected counts under S model, (c) expected counts under QS model, (d) expected counts under MH model.

(b)

207	202.2	40.8
202.8	119	17.7
40.2	18.3	1

(c)

207	201.3	38.7
203.7	119	17.3
42.3	18.7	11

(d)

207	202.5	40.5
202.5	119	18
40.5	18	1

Table V. Model comparison results.

SNP Pair	Complete Symmetry			Quasi Symmetry			Marginal Homogeneity		
	df	G ²	AIC	df	G ²	AIC	df	G ²	AIC
1 \ 23	3	0.765	-5.235	1	0.027	-1.973	2	0.740	-3.260
1 \ 24	3	0.245	-5.755	1	0.016	-1.984	2	0.230	-3.770
2 \ 7	3	6.415	0.415	1	0.320	-1.68	2	6.120	2.120
3 \ 4	3	6.785	0.785	1	1.477	-0.523	2	5.270	1.270
3 \ 5	3	6.785	0.785	1	1.477	-0.523	2	5.270	1.270
3 \ 6	3	6.480	0.48	1	1.437	-0.563	2	5.000	1.000
3 \ 7	3	5.702	-0.298	1	1.268	-0.732	2	4.370	0.370
3 \ 8	3	7.298	1.298	1	2.728	0.728	2	4.530	0.530
3 \ 18	3	0.704	-5.296	1	0.469	-1.531	2	0.210	-3.790
4 \ 18	3	3.525	-2.475	1	0.176	-1.824	2	3.330	-0.670
5 \ 18	3	3.525	-2.475	1	0.176	-1.824	2	3.330	-0.670
6 \ 18	3	3.306	-2.694	1	0.158	-1.842	2	3.130	-0.870
7 \ 18	3	3.356	-2.644	1	0.147	-1.853	2	3.180	-0.820
8 \ 18	3	2.953	-3.047	1	0.126	-1.874	2	2.810	-1.190
9 \ 11	3	1.328	-4.672	1	0.301	-1.699	2	1.030	-2.970
10 \ 11	3	1.369	-4.631	1	0.337	-1.663	2	1.030	-2.970
12 \ 16	3	4.813	-1.187	1	3.653	1.653	2	1.150	-2.850
14 \ 15	3	6.931	0.931	1	0.000	-2.000	2	5.030	1.030
23 \ 24	3	7.473	1.473	1	0.000	-2.000	2	5.840	1.840

The comparison of three symmetry models for significant interactions, are evaluated by the AIC values. G^2 , df and AIC values are summarized for each model in Table V. The smallest values of G^2 and AIC indicate the SNP pair that best fits the model.

SNP-SNP interactions are represented in 3x3 square contingency tables for case-control. The row and column variables have labels as “aa”, “aA” and “AA” indicating the SNP characteristics.

The S, QS and MH models are applied to the 24 top associated SNP pairs which are structured in the form of square contingency tables. Considering that the S model rarely fits data very well, for all the SNP pairs, excluding the pair 14-15 and the pair 23-24, we see that we provide the $S \equiv QS \cap MH$. It means that the test statistic for goodness-of-fit of the S model is asymptotically equivalent to the sum of the QS and MH models. G^2 values for the pairs 14-15 and 23-24 are calculated as zero due to the non-diagonal elements of the contingency tables consist of zero.

For the interactions, data that fitted to the S model indicate that $p_{ij} = p_{ji}$ holds for $i, j = 1, 2, 3$. We could say that, for instance, $p_{aa,Aa} = p_{Aa,aa}$, $p_{aa,AA} = p_{AA,aa}$ and $p_{Aa,AA} = p_{AA,Aa}$ over alleles

a and A . As a consequence, SNPs including at least one allele “ A ” might be at risk for disease. The pairs fitted to the MH model show that $p_{i.} = p_{.j}$ holds for $i, j = 1, 2, 3$. For instance, for the SNP pair 1-24 it can be noted that the marginal genotype distribution of two SNPs is equal to each other. Thus, the marginal probabilities can be represented as: $P(SNP_1 = aa) = P(SNP_{24} = aa)$, $P(SNP_1 = aA) = P(SNP_{24} = aA)$ and $P(SNP_1 = AA) = P(SNP_{24} = AA)$.

The QS model holds for the pair 1-24. It can be concluded that the ratio of expected frequencies under the QS model can be calculated as $(\hat{e}_{12}\hat{e}_{23}\hat{e}_{31})/(\hat{e}_{21}\hat{e}_{32}\hat{e}_{13}) = 0.999$ which is approximately equal to 1, hence we can say that the QS is equivalent to S.

CONCLUSIONS

In this study, complete symmetry, quasi-symmetry and marginal homogeneity models representing the symmetry structure in square contingency tables were applied to 24 cross-classified SNP pairs. Our results show that by using symmetry models, SNP-SNP interactions can be evaluated when the study design is a case control study. We hope that the results given in this paper will be helpful to researchers studying on gene interaction and association analysis and also other kind of interaction analysis on various sciences.

Author contributions

ÖK, GA and SA idealized and supervised the research. ÖK and GA analyzed the data, prepared tables, authored or reviewed drafts of the paper. SA contributed to discussion and approved the final draft; all authors actively contributed to manuscript preparation.

REFERENCES

- AGRESTI A. 2002. Categorical Data Analysis. Second Edition, John Wiley & Sons, Inc., New Jersey.
- AKAIKE H. 1974. A new look at the statistical model identification. IEEE Trans Autom Contr 19: 716-723.
- BISHOP YM, FIENBERG S & HOLLAND PW. 1975. Discrete Multivariate Analysis. Theory and Practice, MIT Press, London.
- CAUSSINUS H. 1965. Contributions a l'analyse Statistique des Tableaux de Correlation. Ann Fac Sci Univ Toulouse 29: 77-182.
- CLARKE MC, ANDERSON CA, PETTERSSON FH, CARDON LR, MORRIS AP & ZONDERVAN KT. 2011. Basic statistical analysis in genetic case-control studies. Nature Protocols 6: 121-133.
- EFENDİOĞLU G. 2015. Karesel Olumsuzluk Tablolarında Simetrik Olmayan Modeller. Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- GOODMAN LA. 1985. The Analysis of Cross-Classified Data Ordered and/or Unordered Categories: Association Models, Correlation Models, and Asymmetry Models for Contingency Tables with or without Missing Entries. Ann Statistics 13: 10-69.
- HABER M. 1982. Testing for Independence in Intra-class Contingency Tables. Biometrics 38: 93-103.
- HARTWING FP. 2013. SNP-SNP Interactions: Focusing on Variable Coding for Complex Models of Epistasis. J Genet Syndr Gene Ther 4: 189.
- KARADAĞ Ö & AKTAŞ S. 2018. A generalized, multi-stage adjusted, latent class linear mixed model for testing genetic association. Comm Statist Simulation Comput 48: 2301-2312.
- LAWAL H & SUNDHEIM R. 2002. Generating Factor Variables for Asymmetry, Non-independence and Skew-symmetry Models in Square Contingency Tables using SAS. Journal Statist Soft 7: 1-23.
- LEWIS CM. 2002. Genetic Association Studies: Design, analysis and interpretation. Briefings in Bioinformatics 3: 146-153.
- SLAGER SL & SCHAID DJ. 2001. Evaluation of Candidate Genes in Case- Control Studies: A Statistical Method to Account for Related Subjects. Am J Human Genet 68: 1457-1462.
- SONG M, ZHANG Y, KATZAROFF AJ, EDGAR BA & BUTTITTA L. 2014. Hunting complex differential gene interaction patterns across molecular contexts. Nucleic Acids Research 42.
- STUART A. 1955. A Test for Homogeneity of the Marginal Distributions in a Two-way Classification. Biometrika 42: 412-441.

TAHATA K, IWASHITA T & TOMIZAWA S. 2008. Measure of Departure from Conditional Marginal Homogeneity for Square Contingency Tables with Ordered Categories. *Statistics* 42: 453-466.

VAIDYANATHAN V, NAIDU V, KARUNASINGHE N, JABED A, PALLATI R, MARLOW G & FERGUSON RL. 2017. SNP-SNP interactions as risk factors for aggressive prostate cancer. *F1000Research* 6: 621.

VELEZ JI, MARMOLEJO-RAMOS F & CORREA JC. 2016. A Graphical Diagnostic Test for Two-Way Contingency Tables. *Rev Colomb Estadística* 9: 97-108.

YAMAGUCHI K. 1990. Some models for the analysis of asymmetric association in square contingency tables with ordered categories. *Sociological Methodology*, p. 181-212.

How to cite

KARADAĞ Ö, ALTUN G & AKTAŞ S. 2020. Assessment of SNP-SNP interactions by using square contingency table analysis. *An Acad Bras Cienc* 92: e20190465. DOI 10.1590/0001-37652020190465.

ÖZGE KARADAĞ¹

<https://orcid.org/0000-0002-2650-1458>

GÖKÇEN ALTUN²

<https://orcid.org/0000-0003-4311-6508>

SERPİL AKTAŞ¹

<https://orcid.org/0000-0003-3364-6388>

¹Hacettepe University, Department of Statistics, 06800, Beytepe, Ankara, Turkey

²Bartın University, Department of Computer Technology and Information Systems, 74100, Bartın, Turkey

Correspondence to: **Özge Karadağ**

E-mail: ozgekaradag@hacettepe.edu.tr

*Manuscript received on April 19, 2019;
accepted for publication on October 22, 2019*

