# The causal analysis in the log-linear model

Gloria Gheno
Department of Statistical Sciences, University of Padua
gheno@stat.unipd.it

March 2015

**Abstract**

In this paper I try to provide a causal theory which can be applied to the log-linear models because they are devoid of complete methods to calculate the different effects. My causal theory uses odds ratios and Pearl's definitions (2001 [4],2009 [5],2012 [6]). I find that although I delete the log-linear parameter which regulates the interaction, there is still an interaction effect, which I call cell effect and which is due to the mere presence of two variables which affect a third. I then try to calculate this effect using Pearl's formulas for the direct, indirect and total effects and I find that a part of this effect is contained in Pearl's direct natural effect, this part is that related to the nonlinearity. To conclude I provide a test to see if this effect is equal to 0.

**Keywords:** additive interaction, causality, mediation, multiplicative interaction, loglinear model

## 1 Introduction

The analysis of causality is important in many fields of research, for example in economics and in social sciences, because the analyst seeks to understand the mechanisms of analyzed phenomena using the relations among the variables ( i.e. the relations cause-effect, where some variables are the causes, other variables the effects). These variables can influence directly, indirectly or in both ways other variables. The set of all effects which influence a variable is called "Total effect". The direct effect is the effect of a variable on other variable without any intervening variables, while the indirect effect is the effect of a variable on other variable considering only the effect through the intervention of other variables, called mediators. Wright (1921 [13]) defines a diagram for the causal relations, which he calls "path diagram". In the path diagram, the direct causal relation between 2 variables is represented by an arrow which goes from the variable influencing to the influenced variable . If 2 variables are not connected, then there is not direct causal relation between them. The correlation between 2 variables is represented by a double arrow. To explain better the direct, indirect and total effects then I use the path diagram represented in Figure 1: the arrow which goes from X to Y represents the direct effect of X on Y, the 2 arrows which goes from X and Z to Z and Y represent the indirect effect of X on Y through Z and the arrow which goes from Z to Y represents the direct effect of Z on Y. Then the indirect effect is the effect of X on Y mediated by Z. An analyst, then, who is interested in the variable Y, will be interested to understand what affects Y and then he will study the direct, indirect and total effects. It is possible to complicate these effects by introducing the concept of interaction. The interaction occurs when the effect of one cause-variable may depend in some way on the presence or absence of another cause-variable. In literature the interaction effect can be additive or multiplicative and in many case induces that the effect of one variable on another varies by levels of a third and vice versa. Figure 2 shows the path diagram of interaction , where X and Z influences directly Y but also their joint effect XZ influences Y. In a model both interaction effects can be present. A problem of using the log-linear models is the inability to calculate all these effects and this can be considered a limitation. In this paper I provide a method for calculating such effects in a log-linear model in order to explain in detail the mechanisms which govern the relationships among the variables.

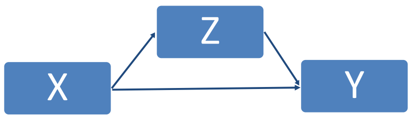


Figure 1: Simple mediation model

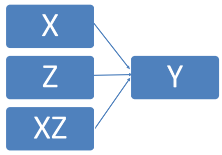


Figure 2: Simple interaction model

## 2 Causal log-linear model

Before introducing the method to calculate the effects, I explain the transition from a log-linear model to the causal log-linear model which represents a log-linear model where the variables have a causal role, i.e. for example X becomes the cause and Y the effect. Vermunt (1996 [11]), instead, distinguishes the log-linear model in these 2 models, which he calls respectively log-linear models and causal log-linear models. The log-linear model describes the observed frequencies, it doesn't distinguish between dependent and independent variables and it measures the strength of the association among variables. The causal log-linear model, introduced by Goodman (1973 [3]) and called "modified path analysis approach", is a log-linear model which considers a causal order of the variables a priori. This model, as written by Vermunt (2005 [12]), consists of specifying a "recursive" system of logit models. In this system the variable, which appears as the dependent in a particular logit equation, may appear as one of the independent variables in one of the next equations. For simplicity, I consider a model with 3 variables, X, Z and Y, which are categorical. The joint probability is

$$P(X=x,Z=z,Y=y)=\pi^{X=x,Z=z,Y=y}=\eta\mu^{X=x}\mu^{Y=y}\mu^{Z=z}\mu^{X=x,Y=y}\mu^{X=x,Z=z}\mu^{Y=y,Z=z}\mu^{X=x,Y=y,Z=z}$$

Now I suppose that X, Z and Y are binary ( 0 or 1), and I consider the dummy code, that is:

$$\mu^{Y=0}=\mu^{X=0}=\mu^{Z=0}=1$$

$$\mu^{Y=0,Z=0}=\mu^{Y=0,X=0}=\mu^{Y=1,Z=0}=\mu^{Y=1,X=0}=\mu^{Y=0,Z=1}=\mu^{Y=0,X=1}=1$$

$$\mu^{Z=0,X=0}=\mu^{Z=1,X=0}=\mu^{Z=0,X=1}=1$$

The joint probability is shown in table 1.

Now I consider the model of Figure 1, which gives the a priori information on the causal order, and I suppose that the variables $Y, X, Z$ are categorical. To consider the model of Figure 1 in log-linear terms, however, I must suppose that the three-interaction term is equal to 1 because, if it is present, it introduces the causal multiplicative interaction term of X and Z on Y (Figure 2). The presence or absence of this parameter, indeed, brings about the presence or absence of the multiplicative interaction. The multiplicative interaction is measured calculating the ratio between the odds ratio obtained by the marginal table XZ given Z=0 (table 2) and the odds ratio calculated using the marginal table XZ given Z=1 (table 3). If this ratio is equal to

| x | z | y | $\pi^{X=x,Z=z,Y=y}$ |
|---|---|---|---|
| 0 | 0 | 0 | $\eta$ |
| 0 | 0 | 1 | $\eta\mu^{Y=1}$ |
| 0 | 1 | 0 | $\eta\mu^{Z=1}$ |
| 0 | 1 | 1 | $\eta\mu^{Y=1}\mu^{Z=1}\mu^{Y=1,Z=1}$ |
| 1 | 0 | 0 | $\eta\mu^{X=1}$ |
| 1 | 0 | 1 | $\eta\mu^{X=1}\mu^{Y=1}\mu^{Y=1,X=1}$ |
| 1 | 1 | 0 | $\eta\mu^{X=1}\mu^{Z=1}\mu^{X=1,Z=1}$ |
| 1 | 1 | 1 | $\eta\mu^{X=1}\mu^{Y=1}\mu^{Z=1}\mu^{X=1,Y=1}\mu^{X=1,Z=1}\mu^{Y=1,Z=1}\mu^{X=1,Y=1,Z=1}$ |

Table 1: The joint probability

| | $Y=0$ | $Y=1$ |
|---|---|---|
| $X=0$ | $\eta$ | $\eta\mu^{Y=1}$ |
| $X=1$ | $\eta\mu^{X=1}$ | $\eta\mu^{X=1}\mu^{Y=1}\mu^{Y=1,X=1}$ |

Table 2: Marginal table $XY$ given $Z=0$

| | $Y=0$ | $Y=1$ |
|---|---|---|
| $X=0$ | $\eta\mu^{Z=1}$ | $\eta\mu^{Y=1}\mu^{Z=1}\mu^{Y=1,Z=1}$ |
| $X=1$ | $\eta\mu^{X=1}\mu^{Z=1}\mu^{X=1,Z=1}$ | $\eta\mu^{X=1}\mu^{Y=1}\mu^{Z=1}\mu^{X=1,Y=1}$ $\mu^{X=1,Z=1}\mu^{Y=1,Z=1}\mu^{Y=1,Z=X=1,Z=1}$ |

Table 3: Marginal table $XY$ given $Z=1$

1, then there is not multiplicative effect, and this occurs only if $\mu^{Y=1,X=1,Z=1}$ is equal to 1. Following the probability structure proposed by Goodman (1973 [3]), I obtain $P(X,Z,Y) = P(Y|Z,X)P(Z|X)P(X)$: the causal model is, then, a decomposition of the joint probability into conditional probabilities.

Now I consider the relation between the log-linear model and the causal log-linear model. I calculate the conditional probabilities using the joint probability and the marginal probabilities. For example the conditional probability of Y=1 given X=1, Z=1, i.e. $\pi^{Y=1|X=1,Z=1}$, is calculated using the table 3 and constraining the three-interaction term equal to 1:

$$\pi^{Y=1|X=1,Z=1} = \frac{\pi^{Y=1,X=1|Z=1}}{\pi^{Y=1,X=1|Z=1} + \pi^{Y=0,X=1|Z=1}} = \frac{\mu^{Y=1}\mu^{Y=1,X=1}\mu^{Y=1,Z=1}}{1+\mu^{Y=1}\mu^{Y=1,X=1}\mu^{Y=1,Z=1}}$$

For simplicity, I write this conditional probability as

$$\eta^{Y|X=1,Z=1}\mu^{Y=1}\mu^{Y=1,X=1}\mu^{Y=1,Z=1}$$

which I call causal form, where

$$\eta^{Y|X=1,Z=1} = \frac{1}{1+\mu^{Y=1}\mu^{Y=1,X=1}\mu^{Y=1,Z=1}}$$

can be seen as a normalization factor. This can be proved recalling that the sum of conditional probabilities $P(Y=1|X=1,Z=1)$ and $P(Y=0|X=1,Z=1)$ is equal to 1. If I write the probabilities in causal form I have

$$\begin{cases} P(Y=1|X=1,Z=1) = \eta^{Y|X=1,Z=1}\mu^{Y=1}\ \mu^{Y=1,X=1}\mu^{Y=1,Z=1} \\ P(Y=0|X=1,Z=1) = \eta^{Y|X=1,Z=1}\mu^{Y=0}\mu^{Y=0,X=1}\mu^{Y=0,Z=1} = \eta^{Y|X=1,Z=1} \end{cases}$$

where, in this case, I do not assume particular values for $\eta^{Y|X=x,Z=z}$. The sum of conditional probabilities is equal to $\eta^{Y|X=1,Z=1}(1+\mu^{Y=1}\mu^{Y=1,X=1}\mu^{Y=1,Z=1})$. Recalling that this sum must be equal to 1, I obtain that $\eta^{Y|X=1,Z=1}$ is equal to $(1+\mu^{Y=1}\mu^{Y=1,X=1}\mu^{Y=1,Z=1})^{-1}$ which is exactly the value which I obtain rewriting the conditional probability in causal form. For this reason, $\eta^{Y|X=x,Z=z}$ can be seen as a normalization factor. The conditional probability $P(Z=z|X=x)$ is calculated using the table $XZ$. Then I write the marginal probability of X and the conditional probabilities of Z and Y in causal form:

$$\pi^{X=x} = \eta_c^X \mu_c^{X=x} \tag{1}$$

$$\pi^{Z=z|X=x} = \eta_c^{Z|X=x}\mu_c^{Z=z}\mu_c^{Z=z,X=x} \tag{2}$$

$$\pi^{Y=y|X=x,Z=z} = \eta^{Y|X=x,Z=z}\mu^{Y=y}\mu^{Y=y,X=x}\mu^{Y=y,Z=z} \tag{3}$$

where for example the ratio between the causal one-effect $\mu_c^{Z=1}$ and the "not" causal one-effect $\mu^{Z=1}$ is $\eta^{Y|X=0,Z=0}/\eta^{Y|X=0,Z=1}$ and the ratio between the causal two-effect $\mu_c^{Z=1,X=1}$ and the "not" causal two-effect $\mu^{Z=1,X=1}$ is

$$\frac{\eta^{Y|X=1,Z=0}\eta^{Y|X=0,Z=1}}{\eta^{Y|X=0,Z=0}\eta^{Y|X=1,Z=1}}$$

The causal normalization factors $\eta^{Y|X=x,Z=z}$ and $\eta_c^{Z|X=x}$ are calculated so:

| | $Y=0$ | $Y=1$ |
|---|---|---|
| $X=0$ | $\eta(1+\mu^{Z=1})$ | $\eta\mu^{Y=1}(1+\mu^{Z=1}\mu^{Y=1,Z=1})$ |
| $X=1$ | $\eta\mu^{X=1}(1+\mu^{Z=1}\mu^{Z=1,X=1})$ | $\eta\mu^{X=1}\mu^{Y=1}\mu^{Y=1,X=1}(1+\mu^{Z=1}\mu^{Z=1,X=1}\mu^{Y=1,X=1})$ |

Table 4: Marginal table $XY$

$$\begin{cases} \eta^{Y|X=0,Z=0} = \frac{1}{1+\mu^{Y=1}} \\ \eta^{Y|X=1,Z=0} = \frac{1}{1+\mu^{Y=1}\mu^{Y=1,X=1}} \\ \eta^{Y|X=0,Z=1} = \frac{1}{1+\mu^{Y=1}\mu^{Y=1,Z=1}} \\ \eta^{Y|X=1,Z=1} = \frac{1}{1+\mu^{Y=1}\mu^{Y=1,Z=1}\mu^{Y=1,X=1}} \\ \eta_c^{Z|X=0} = \frac{1}{1+\mu_c^{Z=1}} \\ \eta_c^{Z|X=1} = \frac{1}{1+\mu_c^{Z=1}\mu_c^{Z=1,X=1}} \end{cases}$$

In this section, I presented two formulations of the same model: they are founded on two different assumptions (causal model and not causal model) and are estimated with two different approaches. The parameters of causal form (i.e those with subscript c) are estimated by a causal log-linear model, the parameter without subscript are estimated by a traditional log-linear model. Only the parameters of conditional probability $\pi^{Y=y|X=x,Z=z}$ remain equal in both forms and for this reason I do not use never the subscript c for them.

# 3 Odds ratio and causal log-linear model

In log-linear model, the causal effects are considered in partial way and for this reason, a true causal analysis is not made. If I consider the causal model of the Figure 1, Bergsma et al. (2009 [2]) calculate the total effect by the marginal table $XY$ (table 4) and the direct effect by the 2 marginal tables $XY$ given $Z=z$ (tables 2 and 3) using the odds ratio. The odds ratios describe the relationship between binary variables; if the variables are categorical, it is necessary a trasformation in binary variables to use them. For example if I want analyze the relation between X and Y, which are categorical variables with 5 categories, I transform them in binary variables: the transformed X and Y are equal to 1 if their original value is 5, 0 otherwise. The relationships considered by the odd ratios can be associative or causal (Zhang, 2008 [14]): in the first type the relation is measured using the actual response variable, while in the second using the potential response. If the two types of odds ratio are different, this is due to the influence of a third variable called confounding variable (Zhang, 2008 [14]; Szumilas, 2010 [8]). This confounding variable is causally linked to the response variable but it is not related causally to other cause or it is linked causally but it is not a mediator variable (Szumilas, 2010[8]): for example if X and Z influences Y, and X and Z are correlated (link which is not of causal type), Z is a confoundering variable of the relation between X and Y. Then in a simple mediation model without confounders, the total effect (TE) and the direct effect used in the log-linear literature (LDE) are given by the following formulas:

$$OR^{TE}_{x,x'} = \frac{P(Y|X=x')}{1-P(Y|X=x')}\frac{1-P(Y|X=x)}{P(Y|X=x)} \tag{4}$$

$$OR^{LDE}_{x,x'}(Z) = \frac{P(Y|X=x',Z=z)}{1-P(Y|X=x',Z=z)}\frac{1-P(Y|X=x,Z=z)}{P(Y|X=x,Z=z)} \tag{5}$$

where the subscript $x,x'$ indicates that the odds ratios measure the effect of the variation of X from x to $x'$. I note that they coincide with the definitions of total effect and controlled direct effect proposed by Pearl (2001 [4], 2009 [5], 2012 [6]). I remember however that Pearl never uses the odds ratio to calculate the effects, but prefers to calculate them using the conditional moments. For this reason, I propose a causal analysis for the log-linear models, applying Pearl's theory to the odds ratio. Using the dummy code, the total effect is equal to

$$\mu^{Y=1,X=1}\left\{\left[\frac{\eta^{Y|X=0,Z=0}+\mu_c^{Z=1}\eta^{Y|X=0,Z=1}}{\eta^{Y|X=1,Z=0}+\mu_c^{Z=1}\mu_c^{X=1,Z=1}\eta^{C|X=1,Z=1}}\right]\left[\frac{\eta^{Y|X=0,Z=0}+\mu_c^{Z=1}\mu^{Y=1,Z=1}\eta^{Y|X=0,Z=1}}{\eta^{Y|X=1,Z=0}+\mu_c^{Z=1}\mu_c^{X=1,Z=1}\mu^{Y=1,Z=1}\eta^{Y|X=1,Z=1}}\right]^{-1}\right\}$$

and the direct effect is equal always to the causal two-effect $\mu^{Y=1,X=1}$, i.e. it is independent of Z. If in a linear-in parameters model without interaction the variable X and the variables Z influence Y but X does not influence Z, the total effect of X on Y is equal to direct effect of X on Y. This is not true in a loglinear model without interaction: I find, indeed, that when $\mu^{Y=1,Z=1,X=1}=1$, the total effect is not equal to direct effect, but there is another effect, which I call cell effect. The cell effect is present only if more variables influence the same variable, as in this case X and Z which influence Y. The cell effect formula is:

$$\begin{aligned}\text{Cell}_{x,x'}^{\text{effect}}(Z)=&\left[\frac{\sum_z P(Y|X=x',Z=z)P(Z|X=x)}{1-\sum_z P(Y|X=x',Z=z)P(Z|X=x)}\frac{1-\sum_z P(Y|X=x,Z=z)P(Z|X=x)}{\sum_z P(Y|X=x,Z=z)P(Z|X=x)}\right]\\&\left[\frac{P(Y|X=x',Z=z)}{1-P(Y|X=x',Z=z)}\frac{1-P(Y|X=x,Z=z)}{P(Y|X=x,Z=z)}\right]^{-1}\end{aligned}\tag{6}$$

In a loglinear model with dummy code, the cell effect is equal to

$$\text{Cell}_{x=0,x'=1}^{\text{effect}}=\frac{\eta^{Y|X=0,Z=0}+\eta^{Y|X=0,Z=1}\mu_c^{Z=1}}{\eta^{Y|X=0,Z=0}+\eta^{Y|X=0,Z=1}\mu_c^{Z=1}\mu^{Y=1,Z=1}}\frac{\eta^{Y|X=1,Z=0}+\eta^{Y|X=1,Z=1}\mu_c^{Z=1}\mu^{Y=1,Z=1}}{\eta^{Y|X=1,Z=0}+\eta^{Y|X=1,Z=1}\mu_c^{Z=1}}\tag{7}$$

Of course, if the $\mu^{Y=1,Z=1}$ is equal to 1 or $\mu^{Y=1,X=1}$ is equal to 1, the cell effect becomes equal to1 and the total effect is equal to direct effect of X on Y or of Z on Y. In a model without multiplicative interaction, the cell effect can be interpreted as an constant interaction effect. As seen in the introduction, indeed, the interaction effect can cause that the direct effect of one variable on another is a function of a third variable, and therefore varies as the third variable varies, while the cell effect remains constant as the third variable varies.

Because the the total effect and the direct effect used in the loglinear literature are the odds ratio versions of the total effect and the controlled direct effect proposed by Pearl (2001 [4], 2009 [5], 2012 [6]), then I propose the odds ratio version of his indirect effect:

$$OR_{x,x'}^{IE}=\frac{\sum_z P(Y|X=x,Z=z)P(Z|X=x')}{1-\sum_z P(Y|X=x,Z=z)P(Z|X=x')}\frac{1-\sum_z P(Y|X=x,Z=z)P(Z|X=x)}{\sum_z P(Y|X=x,Z=z)P(Z|X=x)}\tag{8}$$

Then the total effect is equal to

$$OR_{x,x'}^{TE}=OR_{x,x'}^{LDE}(z)\text{Cell}_{x,x'}^{\text{effect}}\frac{1}{OR_{x',x}^{IE}}\tag{9}$$

The direct effect used in the loglinear literature and the cell effect form the odds ratio version of Pearl's natural direct effect. Pearl, indeed, proposes 2 direct effects: the natural direct effect and the controlled direct effect. The first is the change of Y when X changes and Z is costant at whatever value obtained by the start value of X, while the second is the change of Y when X changes and all other factors are held fixed. The natural direct effect is:

$$OR_{x,x'}^{NDE}=OR_{x,x'}^{LDE}(Z)\text{Cell}_{x,x'}^{\text{effect}}(Z)=\frac{\sum_z P(Y|X=x',Z=z)P(Z|X=x)}{1-\sum_z P(Y|X=x',Z=z)P(Z|X=x)}\frac{1-P(Y|X=x)}{P(Y|X=x)}\tag{10}$$

Because the natural direct effect is independent of Z for Pearl's definition and the direct effect used in the loglinear model literature is also independent of Z, the cell effect is independent of Z and it is a costant which measures the the joint effect of 2 variables on another variable ( this is not true in the loglinear model with interaction, section 5)

The interpretation of the effects calculated as odd ratio is the following: a value of the effect bigger than 1 means a variation of the same direction (if X increases, Y increases) and a value of the effect smaller than 1 means a variation of different direction (if X increases, Y decreases).

If I want calculate the effects of the variation of X from x' a x, I obtain

$$OR^{TE}_{x',x} = \frac{1}{OR^{TE}_{x,x'}}$$

$$OR^{LDE}_{x',x} = \frac{1}{OR^{LDE}_{x,x'}}$$

$$OR^{NDE}_{x',x} \neq \frac{1}{OR^{NDE}_{x,x'}}$$

$$OR^{IE}_{x',x} \neq \frac{1}{OR^{IE}_{x,x'}}$$

$$\text{Cell}^{\text{effect}}_{x',x} \neq \frac{1}{\text{Cell}^{\text{effect}}_{x,x'}}$$

Now I consider the relation among the effects and the parameters. In literature, the two causal effects parameters ($\mu^{X=1,Y=1}, \mu^{Y=1,Z=1}, \mu_c^{X=1,Z=1}$) determine the presence or absence of direct link between the variables: for example if I suppose that $\mu_c^{X=1,Y=1}$ is equal to 1 (i.e there is not direct effect between X and Y) there is not a direct effect of X on Y. In terms of path diagram, the arrow which goes from X to Y is not present. If I set the causal two-effect parameter $\mu_c^{X=1,Z=1}$ equal to 1, I eliminate the direct effect of X on Z, while if I set the "not" causal two-effect parameter $\mu^{X=1,Z=1}$ equal to 1, I don't eliminate the direct effect of X on Z, this because only the causal parameters can determine the presence or absence of direct link. This can be shown using a simple example. I consider the following "not" causal parameters: $\mu^{Y=1,X=1} = 0.02$, $\mu^{Y=1,Z=1} = 0.01$, $\mu^{Y=1}$=0.2, $\mu^{Z=1,X=1} = 1$, $\mu^{Z=1} = 2$ and $\mu^{X=1} = 1.5$. The "not" causal parameter $\mu^{Z=1,X=1}$ is equal to 1, i.e. there isn't a effect between $Z$ and $X$. If I calculate the indirect effect, I find that $OR^{IE}$ is equal to 0.8894, i.e. a effect mediated by Z exists. This occurs because $\mu_c^{Z=1,X=1}$ is equal to 1.1929, i.e. because the variable X is still linked directly to Z, also if the "not" causal parameter is equal to 1. The total effect $OR^{TE}$ is equal to $OR^{LDE}$ because the cell effect is equal to the inverse of the indirect effect $OR^{IE}$ which measures the inverse change of X (from $x'$ to $x$). Now I consider a new log-linear model where the values of parameters $\mu^{XY}, \mu^{YZ}, \mu^{Y}, \mu^{Z}$ and $\mu^{X}$ remain equal to those of previuous example and the value of $\mu^{Z=1,X=1}$ becomes 0.8383. In this case $OR^{IE}$ is equal to 1 because the causal parameter $\mu_c^{Z=1,X=1}$ is equal to 1. In conclusion, if $\mu^{Z=1,X=1}$ and $\mu_c^{Z=1,X=1}$ are equal to 1, the total effect $OR^{TE}$ is equal to the direct effect used in the loglinear literature $OR^{LDE}$, but in first case there is the indirect effect, while in second case, it disappears.

# 4 Empirical examples

In this section, I apply my causal analysis to empirical results. They consider the relations between a typical product (in this case the Sauris' ham) and its festival. This analysis is in marketing field but it can be applied in many economic fields or in social sciences.

The first dataset, which I consider, is composed of 3 dummy variables (X measures the interest about Sauris' ham considering the possibility of buying Sauris' ham, Z measures the satisfaction about Sauris' festival considering the happiness which an individual has if he thinks about Sauris' festival and Y measures future behavior towards Sauris' ham considering if an individual will buy Sauris's ham more often). The results of causal log-linear model are shown in table 5. The two-effect parameters are all significant (i.e all are different from 1). According to the traditional loglinear literature, the causal two-effects parameter is the direct effect. In this case, because all causal two-effects parameters are greater than 1, then an increase of variable X produces an increase of variable Z, and same result occurs for the relation between X and Y and for that between Z and Y. Now I calculate the effects using the formulas (4), (6), (8) and (10). The

| parameter | value |
|---|---|
| $\mu^{YX}$ | 1.9240** |
| $\mu^{YZ}$ | 2.4038*** |
| $\mu^{Y}$ | 0.4881*** |
| $\mu_c^{ZX}$ | 3.3059*** |
| $\mu_c^{Z}$ | 0.4659*** |
| $\mu_c^{X}$ | 1.7132*** |

Table 5: Parameters of first dataset, dummy code Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1

| parameter | value complete model | value restricted model |
|---|---|---|
| $\mu^{YX}$ | 1.3412 | 1 (fixed) |
| $\mu^{YZ}$ | 7.5134 *** | 7.7834*** |
| $\mu^{Y}$ | 0.2770*** | 0.3290*** |
| $\mu_c^{ZX}$ | 1.9481** | 1.9481** |
| $\mu_c^{Z}$ | 0.4607*** | 0.4607 *** |
| $\mu_c^{X}$ | 1.7077 *** | 1.7077*** |

Table 6: Parameters of second dataset, dummy code Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1

total effect is equal to 2.4008, then an increase of X produces an increase of Y, the natural direct effect is equal to 1.8741, then, an increase of X produces an increase of Y. The indirect effect is equal to 1.2845: an increase of X produces , indirectly, an increase of Y. The cell effect is 0.9741: it mitigates the controlled direct effect. Because of presence of 2 variables which influence Y, there is the cell effect and then the direct effect becomes 1.8741. From this analysis, I conclude that if a customer becomes interested in Sauris' ham, then he will buy Sauris' ham more often also thanks to the happiness due to Sauris' festival. In marketing research, this means that a event linked to the product can increase its sell. However, the role of this event is minus important than the interest about the product (indirect effect/ total effect < direct effect/total effect ) and their joint effect decreases the indirect effect and the direct effect effect used in the loglinear literature (cell effect < 1).

Now I consider a second dataset. This dataset is composed of 3 dummy variables (X measures the interest about Sauris' ham considering the possibility of buying Sauris' ham, Z measures the satisfaction about Sauris' festival considering the quality of products presented during the Sauris' festival and Y measures future behavior towards Sauris' festival considering if an individual will suggest others to go to Sauris' festival) which are linked as in Figure 1 (i.e I suppose that there are not confounders). The values of parameters of a complete model (i.e where all arrows of Figure 1 are considered) are showed in first column of the table 6. The parameter which links the variables X and Y is not significant, then I estimate the restricted model whose values are showed in second column of the table 6. In this model, all parameters, which measure the relation between two variables, are significant. I consider the BIC and AIC criteria to choose the best model (between complete and restricted). I recall that, given two estimated models, the model with the lower value of BIC (or AIC) is chosen. The BIC of the complete model is 1370.6260, that of restricted model is 1366.0895. Then I choose the restricted model also because the difference between the two BIC is 4.5355, which according to Raftery' grade (1995 [7]) is positive (i.e there is a positive evidence of difference between the BIC values of the two models). The AIC goes from 1347.4442 to 1346.7713, i.e the index decreases and I choose the restricted model. Because the causal log-linear parameter $\mu^{Y=1,X=1}$ is placed equal to 1, the direct effect used in literature and the natural direct effect are equal to 1 and the cell effect is equal to 1. Then in the restricted model, X influences Y only through Z and the total effect is equal to the indirect effect (in this case $OR^{IE}_{x,x'} = 1/OR^{IE}_{x',x}$): this value is 1.3543. Because this value is bigger than 1, an increase of X produces an indirect increase of Y, in particular an increase of X produces an increase of Z ($\mu_c^{Z=1,X=1} = 1.9481 > 1$) and an increase of Z produces an increase of Y ($\mu^{Y=1,Z=1} = 7.7834 > 1$). Then an increase of the interest about Sauris' ham produces an increase of the positive future behavior towards Sauris' festival. In marketing field, this can be seen as the fact that an increase of interest about the product leads to recommend more often its festival.

# 5 Odds ratio and multiplicative interaction

In this section I consider that the three-interaction $\mu^{Y=1,X=1,Z=1}$ is different from 1. Then the path diagram of the only direct effects on Y is shown in Figure 2. The introduction of the three-interaction parameter produces a multiplicative interaction effect on Y. Then Y is influenced directly by the variables X, by variable Z and by their joint effect due to the three-interaction term. If I consider the marginal probability of X and the conditional probabilities of Z and Y, for the introduction of the interaction term I must modify only the formula (3) : now the three-interaction term is added to the conditional probability of Y given X and Z so the model becomes:

$$\pi^{X=x} = \eta_c^X \mu_c^{X=x}$$
$$\pi^{Z=z|X=x} = \eta_c^{Z|X=x} \mu_c^{Z=z} \mu_c^{Z=z,X=x}$$
$$\pi^{Y=y|X=x,Z=z} = \eta^{Y|X=x,Z=z} \mu^{Y=y} \mu^{Y=y,X=x} \mu^{Y=y,Z=z} \mu^{Y=y,Z=z,X=x}$$

Using the definition of multiplicative interaction, the direct effect of X on Y becomes a function of Z. I show this recalling that the formulas (4), (5), (8) and (10) remain valid and applying the formula (5) to a causal log-linear model with dummy code the direct effect used in the loglinear literature becomes:

$$OR^{LDE}_{x=0,x'=1}(Z) = \mu^{Y=1,X=1} \mu^{Y=1,X=1,Z=z}$$

For the same reason, also the cell effect becomes a function of Z:

$$\text{Cell}^{\text{effect}}_{x=0,x'=1}(Z) = \frac{1}{\mu^{Y=1,X=1,Z=z}} \frac{\eta^{Y|X=0,Z=0} + \eta^{Y|X=0,Z=1} \mu_c^{Z=1}}{\eta^{Y|X=0,Z=0} + \eta^{Y|X=0,Z=1} \mu_c^{Z=1} \mu^{Y=1,Z=1}} \frac{\eta^{Y|X=1,Z=0} + \eta^{Y|X=1,Z=1} \mu^{Y=1,X=1,Z=1} \mu_c^{Z=1} \mu^{Y=1,Z=1}}{\eta^{Y|X=1,Z=0} + \eta^{Y|X=1,Z=1} \mu_c^{Z=1}}$$

The natural direct , indirect and total effects, instead, do not become function of Z. The indirect effect of a model with multiplicative interaction remains equal to that of a model without multiplicative interaction when x goes from 0 to 1.

# 6 Pearl's causal effects and causal log-linear model

In the section 3, I calculated the effects transforming Pearl's theory (2001 [4], 2009 [5],2012 [6]) in odds ratio version and I start with the odds ratio to connect the causal theory already present in the log-linear model with my proposed causal theory. Now I apply Pearl's causal effects to the causal log-linear model. Pearl's causal method uses the expected values and the conditional probabilities because it can be applied to many types of models. For a simple mediation model as that of Figure 1, the formulas proposed by Pearl (2012 [6]) are the following:

$$\text{Controlled}DE_{x,x'}(Z) = E(Y|x', z) - E(Y|x, z) \tag{11}$$

$$DE_{x,x'}(Y) = \sum_z [E(Y|x', z) - E(Y|x, z)]P(z|x) \tag{12}$$

$$IE_{x,x'}(Y) = \sum_z E(Y|x, z)[P(z|x') - P(z|x)] \tag{13}$$

$$TE_{x,x'}(Y) = E(Y|x') - E(Y|x) \tag{14}$$

$$TE_{x,x'}(Y) = DE_{x,x'}(Y) - IE_{x',x}(Y)$$

In linear-in parameters models without interaction and without curvilinear effects, the relation is additive, i.e. $TE_{x,x'}(Y) = DE_{x,x'}(Y) + IE_{x,x'}(Y)$, because in this particular model $IE_{x',x}(Y)$ is equal to $-IE_{x,x'}(Y)$. Then in a general model, the ratio IE/TE is different from 1-DE/TE and, if Y is binary, the ratio $(1-IE/TE)$ measures the fraction of subjects whose responses depend on the direct effect while the ratio $(1 - DE/TE)$

measures the fraction of subjects whose responses depend on the indirect effect (Pearl, 2012 [6]). Now I apply this formulas to model of Figure 1 when the variables are dummy.

$$\text{Controlled}DE_{x=0,x'=1}(Z=0) = \mu^{Y=1}\eta^{Y|X=1,Z=0}\eta^{Y|X=0,Z=0}(\mu^{Y=1,X=1}-1) \tag{15}$$

$$\text{Controlled}DE_{x=0,x'=1}(Z=1) = \mu^{Y=1}\mu^{Y=1,Z=1}\eta^{Y|X=1,Z=1}\eta^{Y|X=0,Z=1}(\mu^{Y=1,X=1}-1) \tag{16}$$

$$\begin{aligned} NDE_{x=0,x'=1}(Y) = \left(\mu^{Y=1,X=1}-1\right)\Big[\mu^{Y=1}\eta_c^{Z|X=0}\Big(&\eta^{Y|X=1,Z=0}\eta^{Y|X=0,Z=0} \\ &+\mu_c^{Z=1}\mu^{Y=1,Z=1}\eta^{Y|X=1,Z=1}\eta^{Y|X=0,Z=1}\Big)\Big] \end{aligned} \tag{17}$$

$$\begin{aligned} IE_{x=0,x'=1}(Y) =&\mu_c^{Z=1}\mu^{Y=1}\left(\mu_c^{Z=1,X=1}-1\right)\left(\mu^{Y=1,Z=1}-1\right) \\ &\eta_c^{Z|X=1}\eta_c^{Z|X=0}\eta^{Y|X=0,Z=1}\eta^{Y|X=0,Z=0} \end{aligned} \tag{18}$$

$$\begin{aligned} TE_{x=0,x'=1}(Y) =&\mu^{Y=1}\Big[\mu^{Y=1,Z=1}\left(\mu^{Y=1,X=1}-1\right)\eta^{Y|X=1,Z=1}\eta^{Y|X=0,Z=1}+\left(\mu^{Y=1,Z=1}-1\right) \\ &\Big(\eta_c^{Z|X=0}\eta^{Y|X=0,Z=0}\eta^{Y|X=0,Z=1}-\eta_c^{Z|X=1}\mu^{Y=1,X=1}\eta_c^{X=1,Z=0}\eta^{Y|X=1,Z=1}\Big)\Big] \end{aligned} \tag{19}$$

As in the log-linear literature seen in the section 2, the causal two effects parameters ($\mu^{X=1,Y=1}$, $\mu^{Y=1,Z=1}$,$\mu_c^{X=1,Z=1}$) continue to determine the presence or absence of a direct link between the variables, but don't measure the effects: for example if I suppose that $\mu^{X=1,Y=1}$ is equal to 1 (i.e there is not direct effect between X and Y), the controlled direct effect and the natural direct effect are 0, but if $\mu^{X=1,Y=1}$ is different from 1, the controlled direct effect and the natural direct effect are different from 0 and from $\mu^{Y=1,X=1}$. The interpretation of Pearl's causal effects is the following: if the effects are positive, an increase of X produces an increase of Y, if they are negative, an increase of X produces a decreases of Y.

Now I calculate the effect which I call cell effect using the Pearl's formulas. Before doing this, I note that in the causal odds ratio the product of the cell effect and of the direct effect used in the log-linear literature is equal to the natural direct effect (NDE) obtained applying Pearl's definition of NDE to the odds ratio. In the same way, I obtain my cell effect in Pearl's theory

$$\text{Cell}_{x,x'}^{\text{effect}}(Z) = DE_{x,x'}(Y) - \text{Controlled}DE(Z) \tag{20}$$

$$\begin{aligned} \text{Cell}_{x,x'}^{\text{effect}}(Z=0) =&(\mu^{Y=1,X=1}-1)\mu^{Y=1}\eta_c^{Z|X=0}\mu_c^{Z=1} \\ &\left[\mu^{Y=1,Z=1}\eta^{Y|X=1,Z=1}\eta^{Y|X=0,Z=1}-\eta^{Y|X=1,Z=0}\eta^{Y|X=0,Z=0}\right] \\ =&-\frac{\text{Cell}_{x,x'}^{\text{effect}}(Z=1)}{\mu_c^{Z=1}} \end{aligned} \tag{21}$$

This cell effect can be seen as a measure of the interaction due to the presence of more variables which influence the same variable Y. It is important to note that this effect exists also if the three-effect parameter is not present. My cell effect is so linked to the measure of the additive interaction proposed by VanderWeele (2012 [10]):

$$\text{Cell}_{x,x'}^{\text{effect}}(Z=z) = \underbrace{\pi^{Y=1|X=1,Z=1}-\pi^{Y=1|X=0,Z=1}-\pi^{Y=1|X=1,Z=0}+\pi^{Y=1|X=0,Z=0}}_{\text{VanderWeele's additive interaction effect}}\eta_c^{Z|X=0}\mu_c^{Z=z-1}$$

If VanderWeele's measure is positive then my cell effect is positive and the interaction is said superadditive, while if the VanderWeele's measure is negative then my cell effect is negative and the interaction is said subadditive. If Vanderweele's measure is equal to 0, my cell effect is null and there is not additive interaction. Then the presence of absence of the cell effect obtained by Pearl's formula causes the presence or absence of the additive interaction. I find that the cell effect is equal to 0 in 3 different cases: in first case if the two-effect between Y ad X is equal to 1 (i.e. $\mu^{Y=1,X=1} = 1$), in second case if the two-effect between Y and Z is equal to 1 (i.e. $\mu^{Y=1,Z=1} = 1$) and in third case if the the two-effect between Y and Z is equal to $(\mu^{Y=1})^{-2}(\mu^{Y=1,X=1})^{-1}$. The first case and the last case are the same cases analyzed in the section 3 when I propose the cell effect for the odds ratio version of Pearl's theory. The second case does not give a null cell effect in the odds ratio. This is the difference between the 2 methods: the cell effect calculated using the odds ratio is equal to 1 (i.e. it is null) only if the one variable influences Y, while Pearl's cell effect is equal to 0 only if the additive interaction effect is not present. The other difference is the fact that the cell effect and the controlled direct effect in Pearl's theory depend on the value of variable Z, while the same effects in the odds ratio are independent of the value of variable Z.

## 7 The cell effect obtained by Pearl's causal formulas is a measure of linearity of model

In this section I analyze the relation among the causal log-linear model and the linearity. As seen in the section 2, the causal log-linear model analyzes the relation among the variables using the cell frequencies, then the causal relation can be analyzed only using the conditional probabilities, but the relation among the variables can be expressed by any function, for example $Y = f(X)$. A simple linear model requires that the causal relation among the variables is linear, i.e $Y = \beta_0 + \gamma_1 X$. Of course the world is not perfect: it is necessary to introduce an error term, then the relation between Y and X becomes $Y = \beta_0 + \gamma_1 X + \zeta$. In its simpler formulation, the linear model considers the variables $X, Y$ and $\zeta$ continuous and normally distributed. Now I analyze what occurs in a log-linear model if the relation between X and Y is linear. In a first step, I consider a perfect word, i.e. where Y is perfectly given by the relation $\beta_0 + \gamma_1 X$ and X and Y are continuous variables with a generic joint distribution $P(X, Y)$. To analyze the same variables with a causal log-linear model, I must discretize the continuous variables. Now I transform X and Y in two binary variables $X^*$ and $Y^*$ so: the the values of X (or Y) which are smaller than the mean become 0, the values of X (or Y) which are bigger than the mean become 1. This particular transformation is made in order that the linearity is inserted in causal log-linear model. The marginal probabilities of the new variables $X^*$ and $Y^*$ are:

$$P(X^*) = \begin{cases} P(X < E(X)) & X^* = 0 \\ P(X \geq E(X)) & X^* = 1 \end{cases} \tag{22}$$

$$P(Y^*) = \begin{cases} P(Y < E(Y)) & Y^* = 0 \\ P(Y \geq E(Y)) & Y^* = 1 \end{cases} \tag{23}$$

Now, using the linear relation between X and Y, I obtain that :

$$Y < E(Y) => \beta_0 + \gamma_1 X < \beta_0 + \gamma_1 E(X)$$

I simplify and obtain that

$$Y < E(Y) \quad \text{is equal to} \quad \gamma_1 X < \gamma_1 E(X)$$

i.e

$$P(Y^* = 0) = P(Y < E(Y)) = \begin{cases} P(X < E(X)) = P(X^* = 0) & \text{if} \quad \gamma_1 > 0 \\ P(X > E(X)) & \text{if} \quad \gamma_1 < 0 \end{cases} \tag{24}$$

To analyze the relation between $X^*$ and $Y^*$, I consider the variables T and W, which are so built:

| | | W | | |
|---|---|---|---|---|
| | | 1 | 0 | |
| T | 1 | 1 | 0 | 1 |
| | 0 | 0 | 0 | 0 |
| | | 1 | 0 | 1 |

Table 7: The joint probability with $\gamma_1 > 0$, without error term

| | | W | | |
|---|---|---|---|---|
| | | 1 | 0 | |
| T | 1 | $\pi_{11}$ | $\pi_{10}$ | $\pi_{1+}$ |
| | 0 | $\pi_{01}$ | $\pi_{00}$ | $\pi_{0+}$ |
| | | $\pi_{+1}$ | $\pi_{+0}$ | 1 |

Table 8: The joint probability with $\gamma_1 > 0$ and error term

$$W = \begin{cases} 1 & \text{if} \quad P(Y^* = 1 | X^* = 1) \\ 0 & \text{if} \quad P(Y^* = 0 | X^* = 1) \end{cases}$$

$$T = \begin{cases} 1 & \text{if} \quad P(Y^* = 0 | X^* = 0) \\ 0 & \text{if} \quad P(Y^* = 1 | X^* = 0) \end{cases}$$

If $\gamma_1$ is positive, the joint distribution of T and W is showed in table 7: without error, the probability of $Y^*$ equal to 1 given $X^*$ equal to 1 is 1, i.e. it is the certain event. The event "$Y^*$ equal to 0 given $X^*$ equal to 0" is the certain event.

Because X is a continuous variable, the sign of equality in the inequality is not important, then the formula (24) can be written so:

$$P(Y^* = 0) = P(Y < E(Y)) = \begin{cases} P(X < E(X)) = P(X^* = 0) & \text{if} \quad \gamma_1 > 0 \\ P(X > E(X)) = P(X^* = 1) & \text{if} \quad \gamma_1 < 0 \end{cases} \tag{25}$$

Unfortunately, the world is not perfect and the relation between X and Y contains an error term, which has zero mean. Then I obtain:

$$Y < E(Y) => \beta_0 + \gamma_1 X + \zeta < \beta_0 + \gamma_1 E(X)$$

I simplify and obtain that

$$Y < E(Y) = \gamma_1 E(X) \quad \text{is equal to} \quad \gamma_1 [X - E(X)] < -\zeta$$

With error term and $\gamma_1$ bigger than 0, the joint probability of variables T and W is showed in table 8: there is not the certain event as the case without error term because the presence of the error term produces the existence of discordant events (i.e. "$Y^*$ equal to 0 given $X^*$ equal to 1" or "$Y^*$ equal to 1 given $X^*$ equal to 0"). I follow the Tutz's method (2011 [9]) for the repeated measurements for binary variables. The repeated measurements occur when the researcher measures the same variables at different time or under different conditions. To analyze if the distribution changes over times or conditions, he considers the joint distribution of the repeated measurements and controls if the marginal homogeneity holds. The marginal homogeneity can be seen in the table 7: it holds if $\pi_{+1}$ is equal to $\pi_{1+}$. In the perfect world, the marginal homogeneity calculated for the joint distribution of the binary variables T and W holds: $\pi_{+1} = 1 = \pi_{1+} = 1$. In the imperfect world the homogeneity holds iff $\pi_{+1}$ is equal to $\pi_{1+}$. I consider the log linear model showed in table 9, where I use the dummy code. Then the marginal homogeneity condition becomes:

$$P(Y^* = 1 | X^* = 1) = \frac{\mu^{Y^*=1} \mu^{X^*=1} \mu^{Y^*=1, X^*=1}}{\mu^{X^*=1} [1 + \mu^{Y^*=1} \mu^{Y^*=1, X^*=1}]} = P(Y^* = 0 | X^* = 0) = \frac{1}{1 + \mu^{Y^*=1}}$$

i.e. the two "not" causal parameter $\mu^{Y^*=1, X^*=1}$ is equal to reciprocal of the squared "not" causal parameter $\mu^{Y=1}$ (i.e. $\mu^{Y^*=1, X^*=1} = 1/[\mu^{Y^*=1}]^2$).

Now I consider a mediation linear model in a perfect world, where X influences linearly Y and Z, which influences in turn linearly Y. This model is so:

$$Z = \alpha_0 + \alpha_1 X$$

| $x^*$ | $y^*$ | $\pi^{X^*=x^*,Y^*=y^*}$ |
|---|---|---|
| 0 | 0 | $\eta$ |
| 0 | 1 | $\eta\mu^{Y^*=1}$ |
| 1 | 0 | $\eta\mu^{X^*=1}$ |
| 1 | 1 | $\eta\mu^{Y^*=1}\mu^{X^*=1}\mu^{Y^*=1,X^*=1}$ |

Table 9: The joint probability of simple linear model

| $x^*$ | $z^*$ | $y^*$ | $\pi^{X^*=x^*,Z^*=z^*,Y^*=y^*}$ |
|---|---|---|---|
| 0 | 0 | 0 | $\eta$ |
| 0 | 0 | 1 | $\eta\mu^{Y^*=1}$ |
| 0 | 1 | 0 | $\eta\mu^{Z^*=1}$ |
| 0 | 1 | 1 | $\eta\mu^{Y^*=1}\mu^{Z^*=1}\mu^{Y^*=1,Z^*=1}$ |
| 1 | 0 | 0 | $\eta\mu^{X^*=1}$ |
| 1 | 0 | 1 | $\eta\mu^{X^*=1}\mu^{Y^*=1}\mu^{Y^*=1,X^*=1}$ |
| 1 | 1 | 0 | $\eta\mu^{X^*=1}\mu^{Z^*=1}\mu^{X^*=1,Z^*=1}$ |
| 1 | 1 | 1 | $\eta\mu^{X^*=1}\mu^{Y^*=1}\mu^{Z^*=1}\mu^{X^*=1,Y^*=1}$ |
| | | | $\mu^{X^*=1,Z^*=1}\mu^{Y^*=1,Z=1}\mu^{X^*=1,Y^*=1,Z^*=1}$ |

Table 10: The joint probability of mediation linear model

$$Y = \omega_0 + \omega_1 X + \omega_2 Z$$

This model can be rewritten in reduced form, i.e:

$$Y = (\omega_0 + \alpha_0) + (\omega_1 + \omega_2\alpha_1)X = \beta_0 + \gamma_1 X$$

which is equal to the relation between X and Y analyzed until now. Now I transform X, Z and Y in binary variables $X^*$, $Z^*$ and $Y^*$ ( 0 if the value of variable is smaller than its mean, 1 if the value of variable is bigger than its mean). As in the simple linear model if $\alpha_1$ is positive and there is not error term , the probability $P(Z^* = 0)$ is equal to probability $P(X^* = 0)$. Now If $\omega_1$ , $\alpha_1$ and $\omega_2$ are positive, also the probability $P(Y^* = 0)$ is equal to probability $P(X^* = 0)$, because in this case in reduce form $\gamma_1$ is positive. Then the variables W and T becomes:

$$W = \begin{cases} 1 & \text{if} \quad P(Y^* = 1|X^* = 1, Z^* = 1) \\ 0 & \text{if} \quad P(Y^* = 0|X^* = 1; Z^* = 1) \end{cases}$$

$$T = \begin{cases} 1 & \text{if} \quad P(Y^* = 0|X^* = 0, Z^* = 0) \\ 0 & \text{if} \quad P(Y^* = 1|X^* = 0, Z^* = 0) \end{cases}$$

In this case, the conditional probabilities $Y^*$ given $Z^*$ and $X^*$ are all equal to 0 in a perfect world when $Y^* = X^* = Z^* = 1$ and $Y^* = X^* = Z^* = 0$ do not occur. The log-linear model is showed in table 10. If I introduce the error terms and I use Tutz's method (Tutz, 2011 [9]) the marginal homogeneity condition, when $\mu^{Y=1,X=1,Z=1}$ is equal to 1, becomes:

$$\frac{\mu^{Y^*=1}\mu^{Y^*=1,Z^*=1}\mu^{Y^*=1,X^*=1}}{1 + \mu^{Y^*=1}\mu^{Y^*=1,Z^*=1}\mu^{Y^*=1,X^*=1}} = \frac{1}{1 + \mu^{Y^*=1}}$$

i.e. the "two" not causal parameter $\mu^{Y^*=1,X^*=1}$ is equal to product between the two not causal parameter $\mu^{Y^*=1,Z^*=1}$ and reciprocal of the squared "not" causal parameter $\mu^{Y=1}$ (i.e $\mu^{Y^*=1,X^*=1} = 1/\{[\mu^{Y^*=1}]^2 \mu^{Y^*=1,Z^*=1}\}$. This condition doesn't imply the condition $P(Y^* = 1|X^* = 1) = P(Y^* = 0|X^* = 0)$: indeed the error term in the relation between the variable $X^*$ and $Z^*$ causes the inequality $P(Y^* = 1|X^* = 1) \neq P(Y^* = 0|X^* = 0)$. Then I must consider also the relation between the variable $X^*$ and $Z^*$. The marginal homogeneity condition holds iff $\mu_c^{Z^*=1,X^*=1} = 1/(\mu_c^{Z^*=1})^2$, where c defines that the parameters are those of a causal log-linear model. As seen in section 2, the causal parameters can be always transformed in not causal log-linear parameters. Then the mediation linear model implies that:

$$\begin{aligned} \mu^{Y^*=1,X^*=1} &= \frac{1}{[\mu^{Y^*=1}]^2\mu^{Y^*=1,Z^*=1}} \\ \mu_c^{Z^*=1,X^*=1} &= \frac{1}{(\mu_c^{Z^*=1})^2} \end{aligned} \tag{26}$$

If these two conditions are satisfied, the equivalence $P(Y^* = 1|X^* = 1) = P(Y^* = 0|X^* = 0)$ is true. Then I conclude that if I suppose that the variables X,Z and Y are linearly linked, then the relative parameters of the causal model must satisfy the bonds (26). This is important because the first bond of (26) is the same found in section 6, i.e. the conditions for which the cell effect is equal to 0. Then I can affirm that the cell effect in Pearl's formula is due to the nonlinearity of the relationship among the variables, while the cell effect in the odds ratio is due only to the effects of more variables on the same variables. Then it can occur that Pearl's cell effect is equal to 0, but the cell effect obtained by odds ratio is not equal to 1 because there are 2 interaction effects: additive interaction effect and that I call true cell effect. Pearl's cell effect depends on the linearity (additive interaction), the true cell effect depends on the number of the variables which influence another variable.

# 8 Test on the presence of the cell effect obtained by Pearl's causal formulas

In this section I find a test to analyze the presence of the cell effect obtained by Pearl's causal formula. For simplicity, I consider the log transformation of the parameters [1]. Agresti (2002 [1] ) shows that the "not" causal linear model without two-effect parameter for a 2x2 table (i.e. a contingency table for 2 variable, X and Y) can be so written:

$$\log(\boldsymbol{m}) = \begin{bmatrix} \log m^{X=0,Y=0} \\ \log m^{X=0,Y=1} \\ \log m^{X=1,Y=0} \\ \log m^{X=1,Y=1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \log(\eta) \\ \log(\mu^X) \\ \log(\mu^Y) \end{bmatrix} = \boldsymbol{D\lambda} \tag{27}$$

where $\boldsymbol{m}$ denotes the column vector of expected counts of the contingency table and $\boldsymbol{\lambda}$ is the vector of the additive "not" causal linear parameters. The formula (27) can be extent to a "not" causal log-linear model with all interactions for a nxn contigency table. Then, in a general "not" causal log-linear model, using the maximum likelihood method, the variance-covariance matrix for the estimated additive "not" causal parameters is

$$Cov(\log(\hat{\boldsymbol{\mu}})) = Cov(\hat{\boldsymbol{\lambda}}) = [\boldsymbol{D}' diag(\hat{\boldsymbol{m}})\boldsymbol{D}]^{-1} \tag{28}$$

Now I consider the particular case where the cell effect is equal to 0 also if 2 variables influence the variable Y . This occurs, as showed in section 6, if $\mu^{Y=1,Z=1} = [(\mu^{Y=1})^2\mu^{Y=1,X=1}]^{-1}$. Because these parameters remain equal both in causal log-linear model and in "not" causal log-linear model (see section 2), this relation can be tested both in causal log-linear model and in "not" causal log-linear model. I test this relation in "not" causal log-linear model. For simplicity, I consider the additive parameter, then the relation becomes:

$$\log(\mu^{Y=1,Z=1}) + 2\log(\mu^{Y=1}) + \log(\mu^{Y=1,X=1}) = \lambda^{Y=1,Z=1} + 2\lambda^{Y=1} + \lambda^{Y=1,X=1} = 0 \tag{29}$$

Now I propose a z-test. Because the vector of estimated lambda are distributed as a multivariate normal, the left-side (29) is a variable normally distributed with the mean equal to

$$E(\hat{\lambda}^{Y=1,Z=1} + 2\hat{\lambda}^{Y=1} + \hat{\lambda}^{Y=1,X=1}) = E(\hat{\beta}) \quad \text{and} \quad \lambda^{Y=1,Z=1} + 2\lambda^{Y=1} + \lambda^{Y=1,X=1} = \beta$$

and the variance equal to

[1] The log-linear model and the causal log-linear model can be expressed in multiplicative form, i.e. that seen in section 2, or in additive form (Vermunt, 1996 [11]; Bergsma et al., 2009 [2]). The parameters of the additive form are equal to the log transformation of the parameters of multiplicative form

$$Var(\hat{\beta}) = Var(\hat{\lambda}^{Y=1,Z=1}) + 4Var(\hat{\lambda}^{Y=1}) + Var(\hat{\lambda}^{Y=1,X=1}) + 4Cov(\hat{\lambda}^{Y=1,Z=1}, \hat{\lambda}^{Y=1})$$
$$+ 2Cov(\hat{\lambda}^{Y=1,Z=1}, \hat{\lambda}^{Y=1,X=1}) + 4Cov(\hat{\lambda}^{Y=1,X=1}, \hat{\lambda}^{Y=1})$$

Then the statistic z, which is equal to $(\hat{\beta} - \beta)(Var(\hat{\beta}))^{-2}$, is normally distributed with mean equal to 0 and variance equal to 1. Now I has a statistic to test when the cell effect is equal to 0. The equality (29) requires that $\beta$ is equal to 0, then testing the equality condition is equal to testing that $\beta$ is equal to 0. Now I return to first dataset, which I considered in section 4. The causal log-linear parameters are showed in table 4. The direct effect is equal to 0.1553, the indirect effect is equal to 0.0611 and the total effect is equal to 0.2155. According to Pearl's mediation formulas, all effects are positive and this brings to the same results obtained by the odds ratio. The ratio $1 - DE/TE$ is equal to 0.2794, then I conclude that 27.94% of those with positive behavior owe their positive position to a lot of satisfaction and to a lot of interest. The ratio $1 - IE/TE$ is equal to 0.7164, then I conclude that 71.64% of those with positive behavior owe their positive position only to a lot of interest. Finally, I conclude that the direct effect is the most important cause of positive decision. Now I apply this test and I find that z is equal to 0.4174 (the p-value is 0.6764) then I accept the hypothesis that the cell effect is equal to 0: the relation between Y, Z and X is linear and for this reason the presence of two variables, which influence the same variable, does not produce an additive interaction effect. The positive behavior is due only to the direct effect and to the indirect effect of the interest about Sauris' ham, but it is not due to the joint effect of the interest and of the satisfaction. This is an example where the fact that 2 variables influence the same variable produces different results according to the 2 causal theories: the additive interaction effect is equal to 0 because the relation between X and Z is linear, but X and Z influence Y and for this reason the cell effect obtained by the odds ratio is different to 1. The additive interaction effect is equal to 0, but the true cell effect is still present.

# 9 Pearl's causal effects and causal log-linear model with interaction

Now I apply the Pearl's effects to a causal log-linear model with interaction. The obtained effects are:

$$\text{Controlled}_{x=0,x'=1}DE(Z=0) = \mu^{Y=1}\eta^{Y|X=1,Z=0}\eta^{Y|X=0,Z=0}(\mu^{Y=1,X=1}\mu^{Y=1,X=1,Z=1} - 1)$$

$$\text{Controlled}_{x=0,x'=1}DE(Z=1) = \mu^{Y=1}\mu^{Y=1,Z=1}\eta^{Y|X=1,Z=1}\eta^{Y|X=0,Z=1}(\mu^{Y=1,X=1} - 1)$$

$$\begin{aligned} DE_{x=0,x'=1}(Y) = \quad & \eta^{Z|X=0}\mu^{Y=1}\Big\{\mu^{Y=1,Z=1}\mu_c^{Z=1}\eta^{Y|X=1,Z=1}\eta^{Y|X=0,Z=1} \\ & \left[\mu^{Y=1,X=1}\mu^{Y=1,X=1,Z=1} - 1\right] \\ & +\eta^{Y|X=1,Z=0}\eta^{Y|X=0,Z=0}\left[\mu^{Y=1,X=1} - 1\right]\Big\} \end{aligned}$$

$$\begin{aligned} IE_{x=0,x'=1}(Y) = \quad & \mu^{Y=1}\mu_c^{Z=1}\left(\mu_c^{Z=1,X=1} - 1\right)\left(\mu^{Y=1,Z=1} - 1\right) \\ & \eta_c^{Z|X=1}\eta_c^{Z|X=0}\eta^{Y|Z=1,X=0}\eta^{Y|Z=0,X=0} \end{aligned}$$

The indirect effect of x which goes from 0 to 1 does not change when I introduce the multiplicative interaction.

# 10 My R package: efflog

My package efflog provides fuctions to calculate directly the effects using the odds ratios. The commands for the effects of a loglinear model without multiplicative interaction are:

```
cell_effect_or(x,y,z,w)
ndirect_effect_or(x,y,z,w,t)
indirect_effect_or(x,y,z,w,t)
total_effect_or(x,y,z,w,t)
```

where $x = \mu^{Y=y}$, $y = \mu^{Y=y,X=x}, z = \mu^{Y=y,Z=z}, w = \mu_c^{Z=z}, t = \mu_c^{Z=z,X=x}$. The commands for the effects of a loglinear model with multiplicative interaction are:

```
cell_effect_mult_or(x,y,z,w,q)
ndirect_effect_mult_or(x,y,z,w,t,q)
total_effect_mult_or(x,y,z,w,t,q)
```

where $x = \mu^{Y=y}$, $y = \mu^{Y=y,X=x}, z = \mu^{Y=y,Z=z}, w = \mu_c^{Z=z}, t = \mu_c^{Z=z,X=x}, q = \mu^{Y=y,X=x,Z=z}$. The package

```
efflog_1.0.tar.gz
```

is in the zipped file `http://www.stat.unipd.it/sites/default/files/efflog_1.0.tar__1.zip`.

# 11 Conclusion

When a researcher analyzes the data, he is interested in understanding the mechanisms which govern the changes of the variables. To understand these mechanisms he uses the causal effects. Unfortunately, when the researcher uses the log-linear models to study the data, he has not available a causal theory, but only few comments on various papers using odds ratios. For this reason, using the causal concepts provided by Pearl, I provide a causal theory for the log-linear models but using odds ratios so that the parameters have the same interpretation given by the log-linear literature. Making so I find a new effect never studied in the literature: I call it cell effect and it can be interpreted as an interaction effect which occurs whenever I consider two variables affecting a third. Then the researcher, who studies his data with my causal theory will have the traditional effects (direct, indirect and total) plus an interaction effect. In the second part of the paper I apply Pearl's causal theory to the log-linear models. It requires that the parameters are interpreted in a different way from that given in the log-linear literature. I find that the cell effect is a part of the natural direct effect but has a different interpretation respect to that found for the odds ratios. In this case the interaction effect of the 2 variables on a third varibile is due to the nonlinearity (additive interaction effect) and therefore if the values of the conditional probabilities are consistent with those of a linear model, the presence of 2 variables which influence a third variable, does not cause the cell effect. The researcher can apply to causal log-linear model both my causal theory and that proposed by Pearl, recalling the diffent interpretation and that the odds ratios analyze the true cell effect while the Pearl's formula only the additive interaction effect.Then I would recommend the use of the odds ratio in the case of dichotomous or categorical variables, while the use of Pearl's theory for ordered categorical variables.

# References


[1] Alan Agresti. *Categorical data analysis.* John Wiley & Sons, New York, second edition, 2002.

[2] Wicher Bergsma, Marcel A Croon, and Jacques A Hagenaars. *Marginal models: For dependent, clustered, and longitudinal categorical data.* Springer-Velag, New York, 2009.

[3] Leo A Goodman. The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. *Biometrika*, 60(1):179–192, 1973.

[4] Judea Pearl. Direct and indirect effects. In *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*, pages 411–420. Morgan Kaufmann Publishers Inc., 2001.

[5] Judea Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146, 2009.

[6] Judea Pearl. The mediation formula: a guide to the assessment of causal pathways in nonlinear models. In Carlo Berzuini, Phillip Dawid, and Luisa Bernardinelli, editors, *Causality: Statistical Perspectives and Applications*, pages 151–179. John Wiley & Sons, 2012.

[7] Adrian E Raftery. Bayesian model selection in social research. *Sociological methodology*, 25:111–164, 1995.

[8] Magdalena Szumilas. Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(3):227–229, 2010.

[9] Gerhard Tutz. *Regression for categorical data*, volume 34. Cambridge University Press, 2011.

[10] Tyler J VanderWeele. Sample size and power calculations for additive interactions. *Epidemiologic methods*, 1(1):159–188, 2012.

[11] Jeroen K Vermunt. Causal log-linear modeling with latent variables and missing data. pages 35–60. Walter de Gruyter, Berlin/New York, 1996.

[12] Jeroen K Vermunt. Log-linear models. In B. Everitt and D. Howell, editors, *Encyclopedia of Statistics in Behavioral Science*, pages 1082–1093. Wiley, Chichester, UK, 2005.

[13] Sewall Wright. Correlation and causation. *Journal of agricultural research*, 20(7):557–585, 1921.

[14] Zhiwei Zhang. Estimating a marginal causal odds ratio subject to confounding. *Communications in Statistics-Theory and Methods*, 38(3):309–321, 2008.