Multivariate Verfahren 2

contingency tables - the log-linear model

Helmut Waldl

March 7th and 8th 2011

categorical regression:
asymmetric problems (differentiation dependent - independent variables)

contingency tables:
symmetric problems, no dependent or independent variables. We only want to know if there is a connection between the variables

bivariate models
We distinguish between **sampling schemata** which affect the hypotheses and the interpretation of a survey

a product multinomial schema

b multinomial schema

c Poisson schema

contingency tables

preface

categorical regression:
asymmetric problems (differentiation dependent - independent variables)

contingency tables:
symmetric problems, no dependent or independent variables. We only want to know if there is a connection between the variables

bivariate models
We distinguish between **sampling schemata** which affect the hypotheses and the interpretation of a survey

a product multinomial schema

b multinomial schema

c Poisson schema

contingency tables

preface

categorical regression:
asymmetric problems (differentiation dependent - independent variables)

contingency tables:
symmetric problems, no dependent or independent variables. We only want to know if there is a connection between the variables

bivariate models
We distinguish between **sampling schemata** which affect the hypotheses and the interpretation of a survey

a product multinomial schema

b multinomial schema

c Poisson schema

contingency tables

preface

categorical regression:
asymmetric problems (differentiation dependent - independent variables)

contingency tables:
symmetric problems, no dependent or independent variables. We only want to know if there is a connection between the variables

bivariate models
We distinguish between **sampling schemata** which affect the hypotheses and the interpretation of a survey

a product multinomial schema

b multinomial schema

c Poisson schema

contingency tables

preface

categorical regression:
asymmetric problems (differentiation dependent - independent variables)

contingency tables:
symmetric problems, no dependent or independent variables. We only want to know if there is a connection between the variables

bivariate models
We distinguish between **sampling schemata** which affect the hypotheses and the interpretation of a survey

a product multinomial schema

b multinomial schema

c Poisson schema

contingency tables

preface

categorical regression:
asymmetric problems (differentiation dependent - independent variables)

contingency tables:
symmetric problems, no dependent or independent variables. We only want to know if there is a connection between the variables

bivariate models
We distinguish between **sampling schemata** which affect the hypotheses and the interpretation of a survey

a product multinomial schema

b multinomial schema

c Poisson schema

contingency tables

product multinomial schema

Similar to ANOVA we interpret one variate as a factor. The elementary units are randomly assigned to the different values of the factor.

Unlike ANOVA the dependent variable is categorical and not metrically scaled

		response variable				
		1	2	$\cdots$	J	
factor	1	x_{11}	x_{12}		x_{1J}	x_{1+}
	2	x_{21}	x_{22}		x_{2J}	x_{2+}
	$\vdots$					$\vdots$
	I	x_{I1}	x_{I2}		x_{IJ}	x_{I+}
		x_{+1}	x_{+2}	$\cdots$	x_{+J}	x_{++}

contingency tables

product multinomial schema

Similar to ANOVA we interpret one variate as a factor. The elementary units are randomly assigned to the different values of the factor.

Unlike ANOVA the dependent variable is categorical and not metrically scaled

		response variable				
		1	2	$\cdots$	J	
factor	1	x_{11}	x_{12}		x_{1J}	x_{1+}
	2	x_{21}	x_{22}		x_{2J}	x_{2+}
	$\vdots$					$\vdots$
	I	x_{I1}	x_{I2}		x_{IJ}	x_{I+}
		x_{+1}	x_{+2}	$\cdots$	x_{+J}	x_{++}

contingency tables

product multinomial schema

Similar to ANOVA we interpret one variate as a factor. The elementary units are randomly assigned to the different values of the factor.

Unlike ANOVA the dependent variable is categorical and not metrically scaled

		response variable				
		1	2	$\cdots$	J	
factor	1	x_{11}	x_{12}		x_{1J}	x_{1+}
	2	x_{21}	x_{22}		x_{2J}	x_{2+}
	$\vdots$					$\vdots$
	I	x_{I1}	x_{I2}		x_{IJ}	x_{I+}
		x_{+1}	x_{+2}	$\cdots$	x_{+J}	x_{++}

contingency tables

product multinomial schema

The investigator fixes the number of experimental units for each value of the factor variable (i.e. $x_{i+}, \quad i = 1, \dots, I$ and with it also x_{++})

The response variable R has a **multinomial distribution** for each value of the factor variable. These I distributions are now compared to each other.

Let $p_{ij} =$
$P(R \text{ is in category } j | \text{ experimental unit comes from factor value } i)$

We now test H_0: "the factor level does not affect the response variable" or expressed mathematically:

$$H_0 : \; p_{1j} = p_{2j} = \cdots = p_{Ij} \text{ for } j = 1, \dots, J$$

contingency tables

product multinomial schema

The investigator fixes the number of experimental units for each value of the factor variable (i.e. $x_{i+}, \quad i = 1, \dots, I$ and with it also x_{++})

The response variable R has a **multinomial distribution** for each value of the factor variable. These I distributions are now compared to each other.

Let $p_{ij} =$
$P(R \text{ is in category } j| \text{ experimental unit comes from factor value } i)$

We now test H_0: "the factor level does not affect the response variable" or expressed mathematically:

$$H_0: \; p_{1j} = p_{2j} = \dots = p_{Ij} \text{ for } j = 1, \dots, J$$

contingency tables

product multinomial schema

The investigator fixes the number of experimental units for each value of the factor variable (i.e. $x_{i+}, \quad i = 1, \ldots, I$ and with it also x_{++})

The response variable R has a **multinomial distribution** for each value of the factor variable. These I distributions are now compared to each other.

Let $p_{ij} =$
$P(R \text{ is in category } j | \text{ experimental unit comes from factor value } i)$

We now test H_0: "the factor level does not affect the response variable" or expressed mathematically:

$$H_0 : \; p_{1j} = p_{2j} = \cdots = p_{Ij} \text{ for } j = 1, \ldots, J$$

contingency tables

product multinomial schema

The investigator fixes the number of experimental units for each value of the factor variable (i.e. $x_{i+}, \quad i = 1, \dots, I$ and with it also x_{++})

The response variable R has a **multinomial distribution** for each value of the factor variable. These I distributions are now compared to each other.

Let $p_{ij} =$
$P(R \text{ is in category } j | \text{ experimental unit comes from factor value } i)$

We now test H_0: "the factor level does not affect the response variable" or expressed mathematically:

$$H_0 : \ p_{1j} = p_{2j} = \dots = p_{Ij} \text{ for } j = 1, \dots, J$$

contingency tables

product multinomial schema

The investigator fixes the number of experimental units for each value of the factor variable (i.e. $x_{i+}, \quad i = 1, \dots, I$ and with it also x_{++})

The response variable R has a **multinomial distribution** for each value of the factor variable. These I distributions are now compared to each other.

Let $p_{ij} =$
$P(R \text{ is in category } j | \text{ experimental unit comes from factor value } i)$

We now test H_0: "the factor level does not affect the response variable" or expressed mathematically:

$$H_0 : \; p_{1j} = p_{2j} = \cdots = p_{Ij} \text{ for } j = 1, \dots, J$$

contingency tables

product multinomial schema

We may express the null hypothesis also with expected frequencies:

Let x_{ij} be the number of experimental units with factor value i that are in category j. Then the joint distribution of all $I \cdot J$ frequencies is the **product-multinomial distribution**:

$$P(x_{11} \dots x_{IJ}) = \prod_{i=1}^{I} \frac{x_{i+}!}{x_{i1}! \cdots x_{iJ}!} p_{i1}^{x_{i1}} \cdots p_{iJ}^{x_{iJ}}$$

The expected frequency of cell (i, j), $m_{i,j}$ then is $E(x_{ij}) = m_{ij} = x_{i+} \cdot p_{ij}$.

Together with $m_{i+} = \sum_j x_{i+} p_{ij}$ and $m_{+j} = \sum_i x_{i+} p_{ij}$ the null hypothesis now may be written as:

$$H_0 : \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}} \qquad (m_{++} = N)$$

contingency tables

product multinomial schema

We may express the null hypothesis also with expected frequencies:

Let x_{ij} be the number of experimental units with factor value i that are in category j. Then the joint distribution of all $I \cdot J$ frequencies is the **product-multinomial distribution**:

$$P(x_{11} \dots x_{IJ}) = \prod_{i=1}^{I} \frac{x_{i+}!}{x_{i1}! \cdots x_{iJ}!} p_{i1}^{x_{i1}} \cdots p_{iJ}^{x_{iJ}}$$

The expected frequency of cell (i, j), $m_{i,j}$ then is $E(x_{ij}) = m_{ij} = x_{i+} \cdot p_{ij}$.

Together with $m_{i+} = \sum_j x_{i+} p_{ij}$ and $m_{+j} = \sum_i x_{i+} p_{ij}$ the null hypothesis now may be written as:

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}} \qquad (m_{++} = N)$$

contingency tables

product multinomial schema

We may express the null hypothesis also with expected frequencies:

Let x_{ij} be the number of experimental units with factor value i that are in category j. Then the joint distribution of all $I \cdot J$ frequencies is the **product-multinomial distribution**:

$$P(x_{11} \dots x_{IJ}) = \prod_{i=1}^{I} \frac{x_{i+}!}{x_{i1}! \cdots x_{iJ}!} p_{i1}^{x_{i1}} \cdots p_{iJ}^{x_{iJ}}$$

The expected frequency of cell (i, j), $m_{i,j}$ then is $E(x_{ij}) = m_{ij} = x_{i+} \cdot p_{ij}$.

Together with $m_{i+} = \sum_j x_{i+} p_{ij}$ and $m_{+j} = \sum_i x_{i+} p_{ij}$ the null hypothesis now may be written as:

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}} \qquad (m_{++} = N)$$

contingency tables

product multinomial schema

We may express the null hypothesis also with expected frequencies:

Let x_{ij} be the number of experimental units with factor value i that are in category j. Then the joint distribution of all $I \cdot J$ frequencies is the **product-multinomial distribution**:

$$P(x_{11} \ldots x_{IJ}) = \prod_{i=1}^{I} \frac{x_{i+}!}{x_{i1}! \cdots x_{iJ}!} p_{i1}^{x_{i1}} \cdots p_{iJ}^{x_{iJ}}$$

The expected frequency of cell (i, j), $m_{i,j}$ then is $E(x_{ij}) = m_{ij} = x_{i+} \cdot p_{ij}$.

Together with $m_{i+} = \sum_j x_{i+} p_{ij}$ and $m_{+j} = \sum_i x_{i+} p_{ij}$ the null hypothesis now may be written as:

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}} \qquad (m_{++} = N)$$

contingency tables

product multinomial schema

We may express the null hypothesis also with expected frequencies:

Let x_{ij} be the number of experimental units with factor value i that are in category j. Then the joint distribution of all $I \cdot J$ frequencies is the **product-multinomial distribution**:

$$P(x_{11} \ldots x_{IJ}) = \prod_{i=1}^{I} \frac{x_{i+}!}{x_{i1}! \cdots x_{iJ}!} p_{i1}^{x_{i1}} \cdots p_{iJ}^{x_{iJ}}$$

The expected frequency of cell (i,j), $m_{i,j}$ then is $E(x_{ij}) = m_{ij} = x_{i+} \cdot p_{ij}$.

Together with $m_{i+} = \sum_j x_{i+} p_{ij}$ and $m_{+j} = \sum_i x_{i+} p_{ij}$ the null hypothesis now may be written as:

$$H_0: \ m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}} \qquad (m_{++} = N)$$

contingency tables

product multinomial schema

We may express the null hypothesis also with expected frequencies:

Let x_{ij} be the number of experimental units with factor value i that are in category j. Then the joint distribution of all $I \cdot J$ frequencies is the **product-multinomial distribution**:

$$P(x_{11} \dots x_{IJ}) = \prod_{i=1}^{I} \frac{x_{i+}!}{x_{i1}! \cdots x_{iJ}!} p_{i1}^{x_{i1}} \cdots p_{iJ}^{x_{iJ}}$$

The expected frequency of cell (i, j), $m_{i,j}$ then is $E(x_{ij}) = m_{ij} = x_{i+} \cdot p_{ij}$.

Together with $m_{i+} = \sum_j x_{i+} p_{ij}$ and $m_{+j} = \sum_i x_{i+} p_{ij}$ the null hypothesis now may be written as:

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}} \qquad (m_{++} = N)$$

contingency tables

product multinomial schema

We may express the null hypothesis also with expected frequencies:

Let x_{ij} be the number of experimental units with factor value i that are in category j. Then the joint distribution of all $I \cdot J$ frequencies is the **product-multinomial distribution**:

$$P(x_{11} \ldots x_{IJ}) = \prod_{i=1}^{I} \frac{x_{i+}!}{x_{i1}! \cdots x_{iJ}!} p_{i1}^{x_{i1}} \cdots p_{iJ}^{x_{iJ}}$$

The expected frequency of cell (i, j), $m_{i,j}$ then is $E(x_{ij}) = m_{ij} = x_{i+} \cdot p_{ij}$.

Together with $m_{i+} = \sum_j x_{i+} p_{ij}$ and $m_{+j} = \sum_i x_{i+} p_{ij}$ the null hypothesis now may be written as:

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}} \qquad (m_{++} = N)$$

contingency tables

multinomial schema

A sample of fixed size $N = x_{++}$ is drawn from a population, all variates are random. Also the marginal sums x_{i+} are not fixed but random.

Let A and B be the first and second classification variate respectively. Then we get the joint probability

$$p_{ij} = P(A \text{ has value } A_i,\ B \text{ has value } B_j)$$

We now want to test the independence of the variates A and B

$$H_0 :\ P(A_i \cap B_j) = P(A_i) \cdot P(B_j)$$

or equivalent $H_0 :\ p_{ij} = p_{i+} \cdot p_{+j}$

contingency tables

multinomial schema

A sample of fixed size $N = x_{++}$ is drawn from a population, all variates are random. Also the marginal sums x_{i+} are not fixed but random.

Let A and B be the first and second classification variate respectively. Then we get the joint probability

$$p_{ij} = P(A \text{ has value } A_i,\ B \text{ has value } B_j)$$

We now want to test the independence of the variates A and B

$$H_0 :\ P(A_i \cap B_j) = P(A_i) \cdot P(B_j)$$

or equivalent $H_0 :\ p_{ij} = p_{i+} \cdot p_{+j}$

contingency tables

multinomial schema

A sample of fixed size $N = x_{++}$ is drawn from a population, all variates are random. Also the marginal sums x_{i+} are not fixed but random.

Let A and B be the first and second classification variate respectively. Then we get the joint probability

$$p_{ij} = P(A \text{ has value } A_i,\ B \text{ has value } B_j)$$

We now want to test the independence of the variates A and B

$$H_0 :\ P(A_i \cap B_j) = P(A_i) \cdot P(B_j)$$

or equivalent $H_0 :\ p_{ij} = p_{i+} \cdot p_{+j}$

contingency tables

multinomial schema

A sample of fixed size $N = x_{++}$ is drawn from a population, all variates are random. Also the marginal sums x_{i+} are not fixed but random.

Let A and B be the first and second classification variate respectively. Then we get the joint probability

$$p_{ij} = P(A \text{ has value } A_i,\ B \text{ has value } B_j)$$

We now want to test the independence of the variates A and B

$$H_0 :\ P(A_i \cap B_j) = P(A_i) \cdot P(B_j)$$

or equivalent $H_0 :\ p_{ij} = p_{i+} \cdot p_{+j}$

contingency tables

multinomial schema

A sample of fixed size $N = x_{++}$ is drawn from a population, all variates are random. Also the marginal sums x_{i+} are not fixed but random.

Let A and B be the first and second classification variate respectively. Then we get the joint probability

$$p_{ij} = P(A \text{ has value } A_i,\ B \text{ has value } B_j)$$

We now want to test the independence of the variates A and B

$$H_0:\ P(A_i \cap B_j) = P(A_i) \cdot P(B_j)$$

or equivalent $H_0:\ p_{ij} = p_{i+} \cdot p_{+j}$

contingency tables

multinomial schema

A sample of fixed size $N = x_{++}$ is drawn from a population, all variates are random. Also the marginal sums x_{i+} are not fixed but random.

Let A and B be the first and second classification variate respectively. Then we get the joint probability

$$p_{ij} = P(A \text{ has value } A_i,\ B \text{ has value } B_j)$$

We now want to test the independence of the variates A and B

$$H_0 :\ P(A_i \cap B_j) = P(A_i) \cdot P(B_j)$$

or equivalent $H_0 :\ p_{ij} = p_{i+} \cdot p_{+j}$

contingency tables

multinomial schema

The joint distribution of the observed frequencies is **multinomial**:

$$P(x_{11} \dots x_{IJ}) = \frac{N!}{x_{11}! \cdots x_{IJ}!} p_{11}^{x_{11}} \cdots p_{IJ}^{x_{IJ}}$$

With expected frequencies we may express the null hypothesis as follows:

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Exercises: Show $p_{ij} = p_{i+} \cdot p_{+j} \quad \Leftrightarrow \quad m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$

contingency tables

multinomial schema

The joint distribution of the observed frequencies is **multinomial**:

$$P(x_{11} \dots x_{IJ}) = \frac{N!}{x_{11}! \cdots x_{IJ}!} p_{11}^{x_{11}} \cdots p_{IJ}^{x_{IJ}}$$

With expected frequencies we may express the null hypothesis as follows:

$$H_0 : \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Exercises: Show $p_{ij} = p_{i+} \cdot p_{+j} \quad \Leftrightarrow \quad m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$

contingency tables

multinomial schema

The joint distribution of the observed frequencies is **multinomial**:

$$P(x_{11} \dots x_{IJ}) = \frac{N!}{x_{11}! \cdots x_{IJ}!} p_{11}^{x_{11}} \cdots p_{IJ}^{x_{IJ}}$$

With expected frequencies we may express the null hypothesis as follows:

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Exercises: Show $p_{ij} = p_{i+} \cdot p_{+j} \quad \Leftrightarrow \quad m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$

contingency tables

multinomial schema

The joint distribution of the observed frequencies is **multinomial**:

$$P(x_{11} \dots x_{IJ}) = \frac{N!}{x_{11}! \cdots x_{IJ}!} p_{11}^{x_{11}} \cdots p_{IJ}^{x_{IJ}}$$

With expected frequencies we may express the null hypothesis as follows:

$$H_0 : \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Exercises: Show $p_{ij} = p_{i+} \cdot p_{+j} \quad \Leftrightarrow \quad m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$

contingency tables

multinomial schema

The joint distribution of the observed frequencies is **multinomial**:

$$P(x_{11} \ldots x_{IJ}) = \frac{N!}{x_{11}! \cdots x_{IJ}!} p_{11}^{x_{11}} \cdots p_{IJ}^{x_{IJ}}$$

With expected frequencies we may express the null hypothesis as follows:

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Exercises: Show $p_{ij} = p_{i+} \cdot p_{+j} \quad \Leftrightarrow \quad m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$

contingency tables

Poisson schema

Now also the sample size N is random (e.g. if you stop observing after a fixed time and not after a fixed observation number)

Assumption: the observed cell frequencies are realisations from **independent Poisson processes**.

Then the class frequencies x_{ij} are Poisson distributed with parameters μ_{ij} and the joint distribution of all observed cell frequencies is:

$$P(x_{11} \dots x_{IJ}) = \prod_{i,j} \frac{\mu_{ij}^{x_{ij}}}{x_{ij}!} \exp(-\mu_{ij})$$

We may test the connection between the cell frequencies via

$$H_0 : \ \mu_{ij} = \frac{\mu_{i+} \cdot \mu_{+j}}{\mu_{++}}$$

contingency tables

Poisson schema

Now also the sample size N is random (e.g. if you stop observing after a fixed time and not after a fixed observation number)

Assumption: the observed cell frequencies are realisations from **independent Poisson processes**.

Then the class frequencies x_{ij} are Poisson distributed with parameters μ_{ij} and the joint distribution of all observed cell frequencies is:

$$P(x_{11} \dots x_{IJ}) = \prod_{i,j} \frac{\mu_{ij}^{x_{ij}}}{x_{ij}!} \exp(-\mu_{ij})$$

We may test the connection between the cell frequencies via

$$H_0: \ \mu_{ij} = \frac{\mu_{i+} \cdot \mu_{+j}}{\mu_{++}}$$

contingency tables

Poisson schema

Now also the sample size N is random (e.g. if you stop observing after a fixed time and not after a fixed observation number)

Assumption: the observed cell frequencies are realisations from **independent Poisson processes**.

Then the class frequencies x_{ij} are Poisson distributed with parameters μ_{ij} and the joint distribution of all observed cell frequencies is:

$$P(x_{11} \dots x_{IJ}) = \prod_{i,j} \frac{\mu_{ij}^{x_{ij}}}{x_{ij}!} \exp(-\mu_{ij})$$

We may test the connection between the cell frequencies via

$$H_0: \ \mu_{ij} = \frac{\mu_{i+} \cdot \mu_{+j}}{\mu_{++}}$$

contingency tables

Poisson schema

Now also the sample size N is random (e.g. if you stop observing after a fixed time and not after a fixed observation number)

Assumption: the observed cell frequencies are realisations from **independent Poisson processes**.

Then the class frequencies x_{ij} are Poisson distributed with parameters μ_{ij} and the joint distribution of all observed cell frequencies is:

$$P(x_{11} \ldots x_{IJ}) = \prod_{i,j} \frac{\mu_{ij}^{x_{ij}}}{x_{ij}!} \exp(-\mu_{ij})$$

We may test the connection between the cell frequencies via

$$H_0: \; \mu_{ij} = \frac{\mu_{i+} \cdot \mu_{+j}}{\mu_{++}}$$

contingency tables

Poisson schema

Now also the sample size N is random (e.g. if you stop observing after a fixed time and not after a fixed observation number)

Assumption: the observed cell frequencies are realisations from **independent Poisson processes**.

Then the class frequencies x_{ij} are Poisson distributed with parameters μ_{ij} and the joint distribution of all observed cell frequencies is:

$$P(x_{11} \dots x_{IJ}) = \prod_{i,j} \frac{\mu_{ij}^{x_{ij}}}{x_{ij}!} \exp(-\mu_{ij})$$

We may test the connection between the cell frequencies via

$$H_0: \ \mu_{ij} = \frac{\mu_{i+} \cdot \mu_{+j}}{\mu_{++}}$$

contingency tables

Poisson schema

This hypothesis is called **multiplicative**, we have a **multiplicative Poisson process**.

With $m_{ij} = E(x_{ij})$ this null hypothesis is equivalent to the null hypotheses of the other sampling schemata.

Let $\tilde{x}_{ij} = x_{ij} | x_{++}$. Then we may express the null hypothesis again with

$$H_0: \ \tilde{m}_{ij} = \frac{\tilde{m}_{i+} \cdot \tilde{m}_{+j}}{\tilde{m}_{++}}$$

Exercises: Show $\tilde{m}_{ij} = E(\tilde{x}_{ij})$

contingency tables

Poisson schema

This hypothesis is called **multiplicative**, we have a **multiplicative Poisson process**.

With $m_{ij} = E(x_{ij})$ this null hypothesis is equivalent to the null hypotheses of the other sampling schemata.

Let $\tilde{x}_{ij} = x_{ij} | x_{++}$. Then we may express the null hypothesis again with

$$H_0: \ \tilde{m}_{ij} = \frac{\tilde{m}_{i+} \cdot \tilde{m}_{+j}}{\tilde{m}_{++}}$$

Exercises: Show $\tilde{m}_{ij} = E(\tilde{x}_{ij})$

contingency tables

Poisson schema

This hypothesis is called **multiplicative**, we have a **multiplicative Poisson process**.

With $m_{ij} = E(x_{ij})$ this null hypothesis is equivalent to the null hypotheses of the other sampling schemata.

Let $\tilde{x}_{ij} = x_{ij} | x_{++}$. Then we may express the null hypothesis again with

$$H_0: \ \tilde{m}_{ij} = \frac{\tilde{m}_{i+} \cdot \tilde{m}_{+j}}{\tilde{m}_{++}}$$

Exercises: Show $\tilde{m}_{ij} = E(\tilde{x}_{ij})$

contingency tables

Poisson schema

This hypothesis is called **multiplicative**, we have a **multiplicative Poisson process**.

With $m_{ij} = E(x_{ij})$ this null hypothesis is equivalent to the null hypotheses of the other sampling schemata.

Let $\tilde{x}_{ij} = x_{ij} | x_{++}$. Then we may express the null hypothesis again with

$$H_0 : \ \tilde{m}_{ij} = \frac{\tilde{m}_{i+} \cdot \tilde{m}_{+j}}{\tilde{m}_{++}}$$

Exercises: Show $\tilde{m}_{ij} = E(\tilde{x}_{ij})$

This hypothesis is called **multiplicative**, we have a **multiplicative Poisson process**.

With $m_{ij} = E(x_{ij})$ this null hypothesis is equivalent to the null hypotheses of the other sampling schemata.

Let $\tilde{x}_{ij} = x_{ij} | x_{++}$. Then we may express the null hypothesis again with

$$H_0 : \ \tilde{m}_{ij} = \frac{\tilde{m}_{i+} \cdot \tilde{m}_{+j}}{\tilde{m}_{++}}$$

Exercises: Show $\tilde{m}_{ij} = E(\tilde{x}_{ij})$

the log-linear model

independence model

For all sampling schemata (product-multinomial, multinomial and Poisson) the null hypothesis may be written as

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Taking the logarithm of the above equation we get

$$\ln m_{ij} = \ln m_{i+} + \ln m_{+j} - \ln m_{++}$$

i.e. assuming H_0 the expected cell frequency may be expressed as sum of three terms:

- a term only depending on the row of the table
- a term only depending on the column of the table
- a term only depending on the observation number

the log-linear model

independence model

For all sampling schemata (product-multinomial, multinomial and Poisson) the null hypothesis may be written as

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Taking the logarithm of the above equation we get

$$\ln m_{ij} = \ln m_{i+} + \ln m_{+j} - \ln m_{++}$$

i.e. assuming H_0 the expected cell frequency may be expressed as sum of three terms:

- a term only depending on the row of the table
- a term only depending on the column of the table
- a term only depending on the observation number

the log-linear model

independence model

For all sampling schemata (product-multinomial, multinomial and Poisson) the null hypothesis may be written as

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Taking the logarithm of the above equation we get

$$\ln m_{ij} = \ln m_{i+} + \ln m_{+j} - \ln m_{++}$$

i.e. assuming H_0 the expected cell frequency may be expressed as sum of three terms:

- a term only depending on the row of the table
- a term only depending on the column of the table
- a term only depending on the observation number

the log-linear model

independence model

For all sampling schemata (product-multinomial, multinomial and Poisson) the null hypothesis may be written as

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Taking the logarithm of the above equation we get

$$\ln m_{ij} = \ln m_{i+} + \ln m_{+j} - \ln m_{++}$$

i.e. assuming H_0 the expected cell frequency may be expressed as sum of three terms:

- a term only depending on the row of the table
- a term only depending on the column of the table
- a term only depending on the observation number

the log-linear model

independence model

For all sampling schemata (product-multinomial, multinomial and Poisson) the null hypothesis may be written as

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Taking the logarithm of the above equation we get

$$\ln m_{ij} = \ln m_{i+} + \ln m_{+j} - \ln m_{++}$$

i.e. assuming H_0 the expected cell frequency may be expressed as sum of three terms:

- a term only depending on the row of the table
- a term only depending on the column of the table
- a term only depending on the observation number

the log-linear model

independence model

For all sampling schemata (product-multinomial, multinomial and Poisson) the null hypothesis may be written as

$$H_0: \ m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Taking the logarithm of the above equation we get

$$\ln m_{ij} = \ln m_{i+} + \ln m_{+j} - \ln m_{++}$$

i.e. assuming H_0 the expected cell frequency may be expressed as sum of three terms:

- a term only depending on the row of the table
- a term only depending on the column of the table
- a term only depending on the observation number

the log-linear model

independence model

For all sampling schemata (product-multinomial, multinomial and Poisson) the null hypothesis may be written as

$$H_0: \; m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Taking the logarithm of the above equation we get

$$\ln m_{ij} = \ln m_{i+} + \ln m_{+j} - \ln m_{++}$$

i.e. assuming H_0 the expected cell frequency may be expressed as sum of three terms:

- a term only depending on the row of the table
- a term only depending on the column of the table
- a term only depending on the observation number

the log-linear model

independence model

For all sampling schemata (product-multinomial, multinomial and Poisson) the null hypothesis may be written as

$$H_0: \ m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$$

Taking the logarithm of the above equation we get

$$\ln m_{ij} = \ln m_{i+} + \ln m_{+j} - \ln m_{++}$$

i.e. assuming H_0 the expected cell frequency may be expressed as sum of three terms:

- a term only depending on the row of the table
- a term only depending on the column of the table
- a term only depending on the observation number

the log-linear model

independence model

The general structure of the logarithmised expected frequencies is

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

where $\mu_{A(i)}$ only depends on i i.e. the first variate and $\mu_{B(j)}$ only depends on j i.e. the second variate. For an easier representation also in multivariate models we choose the following **restrictions** on the model parameters

$$\sum_i \mu_{A(i)} = \sum_j \mu_{B(j)} = 0$$

In summary the model equation of the **log-linear independence model** is given by

$$\ln m_{ij} = \mu + \mu_{A(i)} + \mu_{B(j)} \qquad \text{with} \qquad \sum_i \mu_{A(i)} = \sum_j \mu_{B(j)} = 0$$

the log-linear model

independence model

The general structure of the logarithmised expected frequencies is

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

where $\mu_{A(i)}$ only depends on i i.e. the first variate and $\mu_{B(j)}$ only depends on j i.e. the second variate. For an easier representation also in multivariate models we choose the following **restrictions** on the model parameters

$$\sum_i \mu_{A(i)} = \sum_j \mu_{B(j)} = 0$$

In summary the model equation of the **log-linear independence model** is given by

$$\ln m_{ij} = \mu + \mu_{A(i)} + \mu_{B(j)} \qquad \text{with} \qquad \sum_i \mu_{A(i)} = \sum_j \mu_{B(j)} = 0$$

the log-linear model

independence model

The general structure of the logarithmised expected frequencies is

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

where $\mu_{A(i)}$ only depends on i i.e. the first variate and $\mu_{B(j)}$ only depends on j i.e. the second variate. For an easier representation also in multivariate models we choose the following **restrictions** on the model parameters

$$\sum_i \mu_{A(i)} = \sum_j \mu_{B(j)} = 0$$

In summary the model equation of the **log-linear independence model** is given by

$$\ln m_{ij} = \mu + \mu_{A(i)} + \mu_{B(j)} \qquad \text{with} \qquad \sum_i \mu_{A(i)} = \sum_j \mu_{B(j)} = 0$$

the log-linear model

independence model

The general structure of the logarithmised expected frequencies is

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

where $\mu_{A(i)}$ only depends on i i.e. the first variate and $\mu_{B(j)}$ only depends on j i.e. the second variate. For an easier representation also in multivariate models we choose the following **restrictions** on the model parameters

$$\sum_i \mu_{A(i)} = \sum_j \mu_{B(j)} = 0$$

In summary the model equation of the **log-linear independence model** is given by

$$\ln m_{ij} = \mu + \mu_{A(i)} + \mu_{B(j)} \qquad \text{with} \qquad \sum_i \mu_{A(i)} = \sum_j \mu_{B(j)} = 0$$

the log-linear model

independence model

Theorem:
The hypothesis $m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$ is equivalent to the log-linear independence model for the

a Poisson schema

b multinomial schema with side condition

$$x_{++} = \sum_{i,j} \exp(\mu + \mu_{A(i)} + \mu_{B(j)})$$

c product multinomial schema with side conditions

$$x_{i+} = \sum_{j} \exp(\mu + \mu_{A(i)} + \mu_{B(j)}) \qquad i = 1 \ldots, I$$

Proof: see exercises

the log-linear model

independence model

Theorem:
The hypothesis $m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$ is equivalent to the log-linear independence model for the

a Poisson schema

b multinomial schema with side condition

$$x_{++} = \sum_{i,j} \exp(\mu + \mu_{A(i)} + \mu_{B(j)})$$

c product multinomial schema with side conditions

$$x_{i+} = \sum_{j} \exp(\mu + \mu_{A(i)} + \mu_{B(j)}) \qquad i = 1 \ldots, I$$

Proof: see exercises

the log-linear model

independence model

Theorem:
The hypothesis $m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$ is equivalent to the log-linear independence model for the

a Poisson schema

b multinomial schema with side condition

$$x_{++} = \sum_{i,j} \exp(\mu + \mu_{A(i)} + \mu_{B(j)})$$

c product multinomial schema with side conditions

$$x_{i+} = \sum_{j} \exp(\mu + \mu_{A(i)} + \mu_{B(j)}) \qquad i = 1 \ldots, I$$

Proof: see exercises

the log-linear model

independence model

Theorem:
The hypothesis $m_{ij} = \frac{m_{i+} \cdot m_{+j}}{m_{++}}$ is equivalent to the log-linear independence model for the

a Poisson schema

b multinomial schema with side condition

$$x_{++} = \sum_{i,j} \exp(\mu + \mu_{A(i)} + \mu_{B(j)})$$

c product multinomial schema with side conditions

$$x_{i+} = \sum_{j} \exp(\mu + \mu_{A(i)} + \mu_{B(j)}) \qquad i = 1 \ldots, I$$

Proof: see exercises

the log-linear model

independence model

The parameters of the log-linear independence model result in

$$\mu = \frac{1}{I \cdot J} \sum_{i,j} \ln m_{ij}$$

$$\mu_{A(i)} = \frac{1}{J} \sum_{j} \ln m_{ij} - \mu$$

$$\mu_{B(j)} = \frac{1}{I} \sum_{i} \ln m_{ij} - \mu$$

The log-linear independence model corresponds to the independence of the two variates. We may expand the model such that also dependencies between variates can be modeled.
$\rightarrow$ **saturated model**

the log-linear model

independence model

The parameters of the log-linear independence model result in

$$\mu = \frac{1}{I \cdot J} \sum_{i,j} \ln m_{ij}$$

$$\mu_{A(i)} = \frac{1}{J} \sum_{j} \ln m_{ij} - \mu$$

$$\mu_{B(j)} = \frac{1}{I} \sum_{i} \ln m_{ij} - \mu$$

The log-linear independence model corresponds to the independence of the two variates. We may expand the model such that also dependencies between variates can be modeled.
$\rightarrow$ **saturated model**

the log-linear model

independence model

The parameters of the log-linear independence model result in

$$\mu = \frac{1}{I \cdot J} \sum_{i,j} \ln m_{ij}$$

$$\mu_{A(i)} = \frac{1}{J} \sum_{j} \ln m_{ij} - \mu$$

$$\mu_{B(j)} = \frac{1}{I} \sum_{i} \ln m_{ij} - \mu$$

The log-linear independence model corresponds to the independence of the two variates. We may expand the model such that also dependencies between variates can be modeled.
→ **saturated model**

the log-linear model

independence model

The parameters of the log-linear independence model result in

$$\mu = \frac{1}{I \cdot J} \sum_{i,j} \ln m_{ij}$$

$$\mu_{A(i)} = \frac{1}{J} \sum_{j} \ln m_{ij} - \mu$$

$$\mu_{B(j)} = \frac{1}{I} \sum_{i} \ln m_{ij} - \mu$$

The log-linear independence model corresponds to the independence of the two variates. We may expand the model such that also dependencies between variates can be modeled.
$\rightarrow$ **saturated model**

the log-linear model

independence model

The parameters of the log-linear independence model result in

$$\mu = \frac{1}{I \cdot J} \sum_{i,j} \ln m_{ij}$$

$$\mu_{A(i)} = \frac{1}{J} \sum_{j} \ln m_{ij} - \mu$$

$$\mu_{B(j)} = \frac{1}{I} \sum_{i} \ln m_{ij} - \mu$$

The log-linear independence model corresponds to the independence of the two variates. We may expand the model such that also dependencies between variates can be modeled.

$\rightarrow$ **saturated model**

the log-linear model

independence model

The parameters of the log-linear independence model result in

$$\mu = \frac{1}{I \cdot J} \sum_{i,j} \ln m_{ij}$$

$$\mu_{A(i)} = \frac{1}{J} \sum_{j} \ln m_{ij} - \mu$$

$$\mu_{B(j)} = \frac{1}{I} \sum_{i} \ln m_{ij} - \mu$$

The log-linear independence model corresponds to the independence of the two variates. We may expand the model such that also dependencies between variates can be modeled.
$\rightarrow$ **saturated model**