Visualisation of 2-variable dependence

a hypothetical example

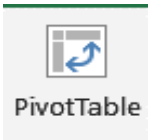
smoking status	women	men
non-smoker	10	6
light smoker	8	8
heavy smoker	4	8

Data about teacher genders were collected at high schools on 4 subjects.
Gender question: Is the subject taught gender conditioned?

gender	subject
man	Gym
man	Gym
man	Gym
man	Gym
man	Gym
man	Gym
man	Gym
man	Bio
man	Bio
man	Bio
man	Bio
man	Math
man	Math
man	Math
man	Math
man	Math
man	Math
man	Math
man	Math
man	Math
man	Math
man	Math
man	Hist
man	Hist
man	Hist
man	Hist
man	Hist
man	Hist
man	Hist
man	Hist
woman	Gym
woman	Gym
woman	Gym
woman	Gym
woman	Gym
woman	Gym
woman	Gym
woman	Gym
woman	Gym
woman	Bio
woman	Bio
woman	Bio
woman	Bio
woman	Bio
woman	Bio
woman	Bio
woman	Bio
woman	Bio
woman	Bio
woman	Bio
woman	Bio
woman	Bio

woman	Bio
woman	Bio
woman	Math
woman	Math
woman	Math
woman	Math
woman	Math
woman	Math
woman	Math
woman	Math
woman	Math
woman	Hist
woman	Hist
woman	Hist
woman	Hist
woman	Hist
woman	Hist
woman	Hist
woman	Hist
woman	Hist
woman	Hist
woman	Hist

contingency table



$$\chi^2 = \sum\sum \frac{(o-e)^2}{e}$$

brand	model	mileage	price	color	
Porsche	911	326,000	22,000 Kč	Yellow	Q1
Porsche	911	326,000	22,000 Kč	Silver	
Fiat	Croma	318,500	27,000 Kč	Silver	
Fiat	Croma	318,500	27,000 Kč	Red	
Skoda	Favorit	200,000	30,000 Kč	Red	
Fiat	Coupe	308,500	34,000 Kč	Silver	
Fiat	Coupe	308,500	34,000 Kč	Yellow	
Ford	Focus	302,000	39,000 Kč	Blue	
Ford	Focus	302,000	39,000 Kč	Silver	
Skoda	Favorit	200,000	40,000 Kč	Blue	
Citroen	Pluriel	294,500	44,000 Kč	Silver	
Citroen	Pluriel	294,500	44,000 Kč	Silver	
Citroen	Saxo	284,500	51,000 Kč	Silver	
Citroen	Saxo	284,500	51,000 Kč	Blue	
Skoda	Octavia	278,000	56,000 Kč	Silver	Q2
Skoda	Octavia	278,000	56,000 Kč	Silver	
Skoda	Fabia	270,500	61,000 Kč	Blue	
Skoda	Fabia	270,500	61,000 Kč	Silver	
Skoda	Fabia	260,500	68,000 Kč	Silver	
Skoda	Fabia	260,500	68,000 Kč	Blue	
Fiat	Coupe	254,000	73,000 Kč	Red	
Fiat	Coupe	254,000	73,000 Kč	Silver	
Alfa Romeo	Spider	242,500	80,000 Kč	Silver	
Alfa Romeo	Spider	242,500	80,000 Kč	Silver	
Skoda	Fabia	56,500	87,000 Kč	Silver	
Citroen	Saxo	100,500	87,000 Kč	Silver	
Citroen	Pluriel	230,000	89,000 Kč	Silver	
Citroen	Pluriel	230,000	89,000 Kč	Blue	
Alfa Romeo	146	242,500	90,000 Kč	Yellow	
Alfa Romeo	146	242,500	90,000 Kč	Silver	
Skoda	Octavia	62,500	98,000 Kč	Yellow	
Opel	Corsa	84,000	98,000 Kč	Yellow	
Skoda	Fabia	206,000	106,000 Kč	Blue	
Skoda	Fabia	206,000	106,000 Kč	Silver	
Chrysler	Neon	55,000	109,000 Kč	Silver	
Alfa Romeo	Spider	90,500	109,000 Kč	Silver	
Renault	Laguna	189,500	118,000 Kč	Silver	
Renault	Laguna	189,500	118,000 Kč	Red	
Peugeot	307	186,500	120,000 Kč	Silver	
Peugeot	307	186,500	120,000 Kč	Silver	
Chrysler	Neon	42,500	120,000 Kč	Silver	
Fiat	Coupe	80,000	120,000 Kč	Silver	
Peugeot	307	186,500	120,000 Kč	Yellow	
Peugeot	307	186,500	120,000 Kč	Silver	
Chrysler	Neon	42,500	120,000 Kč	Blue	
Dodge	Viper	77,000	123,000 Kč	Silver	
Fiat	Coupe	175,000	128,000 Kč	Blue	
Fiat	Coupe	175,000	128,000 Kč	Silver	
Opel	Corsa	165,500	135,000 Kč	Red	

Opel	Corsa	165,500	135,000 Kč	Blue
Opel	Corsa	162,500	137,000 Kč	Yellow
Opel	Corsa	162,500	137,000 Kč	Silver
Opel	Corsa	162,500	137,000 Kč	Silver
Opel	Corsa	162,500	137,000 Kč	Silver
Porsche	911	159,000	139,000 Kč	Silver
Porsche	911	159,000	139,000 Kč	Silver
Citroen	Saxo	151,000	145,000 Kč	Blue
Citroen	Saxo	151,000	145,000 Kč	Blue
Peugeot	307	150,000	146,000 Kč	Silver
Peugeot	307	150,000	146,000 Kč	Blue
Opel	Corsa	86,000	147,000 Kč	Silver
Peugeot	307	142,000	151,000 Kč	Silver
Skoda	Octavia	80,000	151,000 Kč	Red
Peugeot	307	142,000	151,000 Kč	Silver
Fiat	Coupe	76,500	155,000 Kč	Yellow
Ford	Escort	135,000	156,000 Kč	Silver
Ford	Escort	135,000	156,000 Kč	Silver
Fiat	Bravo	76,500	157,000 Kč	Silver
Renault	Kangoo	76,500	157,000 Kč	Red
Skoda	Fabia	68,500	157,000 Kč	Yellow
Skoda	Felicia	71,500	158,000 Kč	Blue
Renault	Laguna	45,500	158,000 Kč	Red
Skoda	Felicia	127,000	162,000 Kč	Blue
Opel	Zafira	126,000	162,000 Kč	Silver
Opel	Zafira	71,500	162,000 Kč	Silver
Peugeot	307	41,000	162,000 Kč	Yellow
Skoda	Felicia	127,000	162,000 Kč	Blue
Opel	Zafira	126,000	162,000 Kč	Silver
Skoda	Fabia	48,500	163,000 Kč	Silver
Skoda	Octavia	67,500	163,000 Kč	Silver
Chrysler	Neon	67,500	164,000 Kč	Silver
Skoda	Octavia	65,000	164,000 Kč	Blue
Ford	Scorpio	76,500	165,000 Kč	Silver
Skoda	Octavia	66,000	166,000 Kč	Silver
Peugeot	307	36,000	166,000 Kč	Silver
Renault	Laguna	56,500	167,000 Kč	Blue
Peugeot	307	100,500	167,000 Kč	Blue
Alfa Romeo	156	76,500	167,000 Kč	Silver
Opel	Corsa	118,000	168,000 Kč	Red
Citroen	Pluriel	60,000	168,000 Kč	Silver
Citroen	Saxo	56,500	168,000 Kč	Blue
Skoda	Octavia	56,500	168,000 Kč	Silver
Fiat	Coupe	28,500	168,000 Kč	Silver
Opel	Corsa	17,000	168,000 Kč	Blue
Opel	Corsa	132,000	168,000 Kč	Silver
Opel	Corsa	118,000	168,000 Kč	Silver
Fiat	Coupe	79,000	169,000 Kč	Silver
Renault	Laguna	138,000	169,000 Kč	Blue
Skoda	Octavia	111,000	173,000 Kč	Yellow

Opel	Corsa	90,500	173,000 Kč	Silver
Fiat	Coupe	134,500	173,000 Kč	Red
Skoda	Octavia	111,000	173,000 Kč	Silver
Fiat	Coupe	110,000	174,000 Kč	Silver
Skoda	Fabia	144,000	174,000 Kč	Silver
Skoda	Fabia	56,500	174,000 Kč	Red
Opel	Corsa	124,000	174,000 Kč	Silver
Fiat	Coupe	110,000	174,000 Kč	Blue
Skoda	Fabia	48,500	175,000 Kč	Yellow
Fiat	Coupe	47,500	175,000 Kč	Silver
Porsche	911	117,000	175,000 Kč	Silver
Citroen	Saxo	116,000	175,000 Kč	Blue
Alfa Romeo	146	47,500	176,000 Kč	Silver
Opel	Corsa	110,000	176,000 Kč	Red
Opel	Corsa	90,500	177,000 Kč	Blue
Alfa Romeo	146	134,500	177,000 Kč	Yellow
Fiat	Bravo	104,000	178,000 Kč	Silver
Fiat	Coupe	51,000	178,000 Kč	Silver
Porsche	911	80,500	178,000 Kč	Silver
Alfa Romeo	Spider	45,000	178,000 Kč	Blue
Opel	Corsa	104,000	178,000 Kč	Yellow
Fiat	Bravo	104,000	178,000 Kč	Silver
Opel	Corsa	90,500	179,000 Kč	Red
Porsche	911	82,500	179,000 Kč	Silver
Citroen	Saxo	81,500	179,000 Kč	Blue
Alfa Romeo	Spider	134,500	179,000 Kč	Silver
Citroen	Pluriel	126,500	179,000 Kč	Silver
Skoda	Fabia	126,000	179,000 Kč	Blue
Peugeot	307	61,500	180,000 Kč	Blue
Citroen	Saxo	130,000	180,000 Kč	Blue
Renault	Kangoo	98,000	182,000 Kč	Silver
Renault	Kangoo	98,000	182,000 Kč	Red
Citroen	Pluriel	56,500	183,000 Kč	Silver
Porsche	911	100,500	183,000 Kč	Silver
Peugeot	206	94,500	184,000 Kč	Blue
Peugeot	206	94,500	184,000 Kč	Red
Peugeot	307	94,500	184,000 Kč	Yellow
Peugeot	307	61,500	184,000 Kč	Silver
Skoda	Octavia	124,000	184,000 Kč	Silver
Peugeot	206	94,500	184,000 Kč	Yellow
Peugeot	206	94,500	184,000 Kč	Silver
Peugeot	307	94,500	184,000 Kč	Yellow
Citroen	Pluriel	131,000	185,000 Kč	Silver
Peugeot	307	124,000	185,000 Kč	Silver
Renault	Laguna	125,500	185,000 Kč	Silver
Peugeot	307	90,000	186,000 Kč	Silver
Ford	Escort	82,500	186,000 Kč	Silver
Peugeot	307	125,500	186,000 Kč	Silver
Peugeot	307	120,000	186,000 Kč	Silver
Skoda	Felicia	79,500	187,000 Kč	Blue

Fiat	Coupe	97,000	187,000 Kč	Blue
Ford	Escort	59,000	188,000 Kč	Red
Skoda	Fabia	118,000	188,000 Kč	Silver
Skoda	Octavia	50,000	189,000 Kč	Blue
Citroen	Pluriel	90,500	189,000 Kč	Blue
Opel	Zafira	75,000	189,000 Kč	Silver
Opel	Corsa	96,000	189,000 Kč	Silver
Renault	Laguna	86,500	190,000 Kč	Silver
Citroen	Saxo	86,000	190,000 Kč	Silver
Skoda	Felicia	70,500	190,000 Kč	Yellow
Opel	Zafira	70,500	190,000 Kč	Silver
Opel	Corsa	70,500	190,000 Kč	Red
Skoda	Fabia	114,500	190,000 Kč	Silver
Fiat	Coupe	114,500	190,000 Kč	Blue
Alfa Romeo	146	114,500	190,000 Kč	Red
Renault	Laguna	86,500	190,000 Kč	Silver
Citroen	Saxo	86,000	190,000 Kč	Red
Citroen	Pluriel	85,500	191,000 Kč	Blue
Renault	Kangoo	85,500	191,000 Kč	Silver
Citroen	Saxo	118,000	191,000 Kč	Silver
Citroen	Pluriel	85,500	191,000 Kč	Yellow
Renault	Kangoo	85,500	191,000 Kč	Silver
Fiat	Coupe	48,500	193,000 Kč	Silver
Peugeot	307	92,500	193,000 Kč	Silver
Opel	Corsa	70,000	194,000 Kč	Red
Opel	Corsa	90,000	194,000 Kč	Red
Citroen	Pluriel	80,000	195,000 Kč	Silver
Skoda	Octavia	111,000	195,000 Kč	Red
Citroen	Pluriel	80,000	195,000 Kč	Silver
Citroen	Pluriel	126,000	196,000 Kč	Blue
Skoda	Octavia	62,500	196,000 Kč	Yellow
Alfa Romeo	Spider	106,500	196,000 Kč	Yellow
Fiat	Coupe	61,500	197,000 Kč	Silver
Fiat	Bravo	104,000	197,000 Kč	Silver
Citroen	Pluriel	106,000	197,000 Kč	Silver
Skoda	Fabia	105,500	197,000 Kč	Silver
Renault	Kangoo	70,000	198,000 Kč	Blue
Renault	Laguna	105,500	198,000 Kč	Blue
Ford	Mondeo	74,000	199,000 Kč	Silver
Skoda	Fabia	110,000	199,000 Kč	Yellow
Ford	Mondeo	74,000	199,000 Kč	Blue
Skoda	Octavia	39,500	200,000 Kč	Silver
Opel	Corsa	74,000	200,000 Kč	Silver
Skoda	Fabia	62,500	200,000 Kč	Silver
Fiat	Coupe	100,000	200,000 Kč	Silver
Skoda	Octavia	39,500	200,000 Kč	Silver
Opel	Zafira	70,500	201,000 Kč	Blue
Opel	Zafira	70,500	201,000 Kč	Silver
Opel	Zafira	70,500	201,000 Kč	Red
Skoda	Fabia	104,000	201,000 Kč	Silver

Fiat	Coupe	98,000	201,000 Kč	Silver
Alfa Romeo	146	94,500	201,000 Kč	Blue
Opel	Zafira	70,500	201,000 Kč	Silver
Opel	Zafira	70,500	201,000 Kč	Silver
Opel	Zafira	70,500	201,000 Kč	Silver
Fiat	Bravo	166,000	204,000 Kč	Silver
Citroen	Saxo	80,500	204,000 Kč	Red
Skoda	Octavia	59,500	205,000 Kč	Silver
Alfa Romeo	146	94,000	205,000 Kč	Silver
Ford	Fiesta	62,500	207,000 Kč	Yellow
Alfa Romeo	Spider	94,500	207,000 Kč	Silver
Ford	Fiesta	62,500	207,000 Kč	Blue
Opel	Zafira	61,500	208,000 Kč	Silver
Ford	Fiesta	61,500	208,000 Kč	Silver
Citroen	Pluriel	94,500	208,000 Kč	Silver
Skoda	Fabia	86,500	208,000 Kč	Silver
Opel	Zafira	61,500	208,000 Kč	Silver
Ford	Fiesta	61,500	208,000 Kč	Red
Renault	Kangoo	59,000	209,000 Kč	Blue
Renault	Laguna	86,000	209,000 Kč	Blue
Renault	Kangoo	59,000	209,000 Kč	Yellow
Citroen	Pluriel	56,000	210,000 Kč	Blue
Opel	Zafira	91,500	210,000 Kč	Blue
Skoda	Fabia	56,000	211,000 Kč	Red
Peugeot	307	85,500	211,000 Kč	Red
Skoda	Fabia	56,000	211,000 Kč	Silver
Ford	Mondeo	42,500	215,000 Kč	Silver
Skoda	Fabia	90,500	215,000 Kč	Silver
Skoda	Octavia	50,000	216,000 Kč	Yellow
Peugeot	307	85,500	216,000 Kč	Yellow
Skoda	Octavia	50,000	216,000 Kč	Silver
Citroen	Pluriel	150,000	221,000 Kč	Silver
Ford	Escort	72,000	221,000 Kč	Silver
Ford	Mondeo	35,000	226,000 Kč	Blue
Opel	Corsa	70,500	226,000 Kč	Blue
Ford	Mondeo	35,000	226,000 Kč	Silver
Skoda	Octavia	138,000	232,000 Kč	Silver
Peugeot	307	81,500	232,000 Kč	Silver
Skoda	Octavia	11,000	243,000 Kč	Silver
Citroen	Saxo	62,500	243,000 Kč	Silver
Skoda	Octavia	11,000	243,000 Kč	Silver
Alfa Romeo	Spider	90,000	247,000 Kč	Silver
Skoda	Felicia	91,500	247,000 Kč	Silver
Alfa Romeo	Spider	151,000	258,000 Kč	Silver
Peugeot	307	72,500	258,000 Kč	Silver
Alfa Romeo	146	47,500	267,000 Kč	Silver
Ford	Escort	92,000	267,000 Kč	Silver
Alfa Romeo	156	146,000	269,000 Kč	Yellow
Peugeot	307	81,500	269,000 Kč	Yellow
Renault	Kangoo	158,000	278,000 Kč	Blue

Peugeot	307	80,500	278,000 Kč	Yellow
Alfa Romeo	156	22,500	280,000 Kč	Silver
Porsche	911	70,500	280,000 Kč	Silver
Alfa Romeo	156	22,500	280,000 Kč	Silver
Alfa Romeo	156	31,000	289,000 Kč	Red
Renault	Laguna	82,500	289,000 Kč	Red
Alfa Romeo	156	30,000	300,000 Kč	Silver
Opel	Corsa	70,500	300,000 Kč	Silver
Alfa Romeo	156	30,000	300,000 Kč	Blue

Find out if the car price is dependent on the color!
before you start, categorize the prize into the following 10 price bins:
up to 30k 56000 90000 120000 147000 180000 210000 232000 269000

$$\chi^2 = \sum\sum \frac{(o-e)^2}{e}$$

Find out if the car price is dependent on the mileage!
before you start, categorize the prize into the following 10 price bins:
up to 30k 56000 90000 120000 147000 180000 210000 232000 269000

300000

300000

four-fold (2x2) tables degrees of freedom=1 compared t

Find out if laterality is dependent on gender.

H0: There is no gender determinati
H1: Laterality is dependend on the

	right-handed	left-handed	total
men	38	8	46
women	49	5	54
total	87	13	**100**

alpha = 0.05

for large samples

"double-sum" chi2-test

			Check-total
	40	6	46
	47	7	54
chech total	87	13	**100**

if there is no laterality dependence
if there is no laterality dependence
>> in this table are expected value

0.101959 0.682341
0.086854 0.581254 1.452408 = X2
3.84145882 = X2critical H0 is accepted

$$\chi^2 =$$

simplified formula

$$X^2 = \frac{N(ad - bc)^2}{(a + b)(a + c)(b + d)(c + d)}.$$

a **b**
c **d**

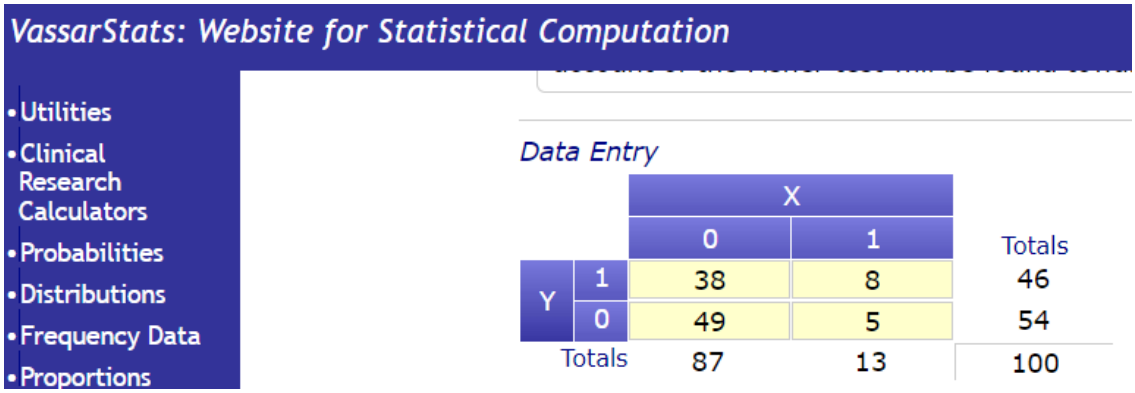
chi2N **1.452408** H0 is accepted
3.841459 = X2critical

Fisher exact test
(eg. VassarStats)

Fisher's test calculates an exact p value,
while chi-square test only calculates an approximation.
With large samples, the difference is trivial.
With small samples, the difference can be important.

http://vassarstats.net/tab2x2.html (two-tailed test) alpha=0.05 >> It is not

p (two-tailed) = 0.2496648878063239
p > alpha >> H0 is accepted



to X2 critical (alpha;1): χ2crit.=CHIsq.iNV.rt(alfa;dof)

ion of laterality.
gender.

on gender then we should observe 40 right-handed man and 47 right-handed women
on gender then we should observe 6 left-handed man and 7 left-handed women
s if the H0 is true

$$\sum\sum \frac{(o-e)^2}{e}$$

>> We can not compare with ""double-sum" chi2-test" and "simplified formula" because it is exact test
>> Exact test based on combinatorial calculation

X

>> Fischer test calculate exact probability value while chi-square is only approximation that is valid for larg
>> It works for relative large sample too but

straightforward algorithm; use on-calculator

ge sample the difference is trivial but for small sample could be important

Cross-clasification - risk ratio

Cross classification usually means that the data are classified according to multiple criteria (usually coded as factors, that is, categorical variables) at the same time, giving rise to a contingency
prospective studies - based on characteristics of a group obese/not obese patients - Cardiovascular d

The relative risk is the ratio of the proportions of cases having a positive outcome in the two group

	obese	not obese = control
CVD	46	60
no CVD	254	640

$$RR = \frac{P_1}{P_0} = \frac{\frac{a}{a+c}}{\frac{b}{b+d}}.$$

P1 P0

RR= x higher

conclusion

Obese persons have ..... times the risk of developing CVD as compared to non-obese persons in t

=in the future

table.
'isease

s.

his study.

Teeth erosion

a case-control study carried out among swimmers to investigate the possible association between exposure to chlorinated swimming pool water and erosion of dental enamel. Among 49 swimmers with enamel erosion (the cases) 32 reported swimming six or more hours per week, compared with 118 of 245 swimmers without enamel erosion (the controls).

2x2

H0: Erosion of dental enamel is indep

time	erosion	controls	total
> 6h	32	118	150
< 6h	17	127	144
total	49	245	**294**

$$X^2 = \frac{N(ad - bc)^2}{(a + b)(a + c)(b + d}$$

dof: 1

a **b**

c **d**

chi2N **4.802**

3.841459 = X2critical

H0 is rejected.

Conclusion: Erosion of dental enamel is dependent

)endent on exposure to chlorinated swimming pool water.

compared to X2 critical (alpha;1): χ2crit.=CHIsq.iNV.rt(alfa;dof)

$$\overline{\quad\quad\quad\quad}$$
$$!)(c + d)$$

on exposure to chlorinated swimming pool water.

	Math. Mag.	Science	
math	5	0	$R_1 = 5$
biology	1	4	$R_2 = 5$
	$C_1 = 6$	$C_2 = 4$	$N = 10.$

Computing P_{cutoff} gives

$$P_{\text{cutoff}} = \frac{5!^2\ 6!\ \ 4!}{10!\ (5!\ \ 0!\ \ 1!\ \ 4!)} = 0.0238,$$

and the other possible matrices and their Ps are

$$\begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} P = 0.2381$$

$$\begin{bmatrix} 3 & 2 \\ 3 & 2 \end{bmatrix} P = 0.4762$$

$$\begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix} P = 0.2381$$

$$\begin{bmatrix} 1 & 4 \\ 5 & 0 \end{bmatrix} P = 0.0238,$$

which indeed sum to 1, as required. The sum of P-values less than or equal to $P_{\text{cutoff}} = 0.0238$ is then 0.(case, there would be a statistically significant association between the journal and type of article appearing.

0476 which, because it is less than 0.05, is significant. Therefore, in this