UK FHS
Historical sociology
(2014)

Quantitative Data Analysis I.

Contingency tables: bivariate analysis of categorical data – introduction

Jiří Šafr

jiri.safr(AT)seznam.cz

updated **5/6/2014**

Tables as technique of data description

What can a contingency table tell us?

- **Comparison between groups**
- **Mutual relationship between 2 (or more) variables**
- **Patterns** of variation of one variable (phenomenon) in course of time
- **Patterns** of variation of two (and more) variables in their mutual relationship

Further we will consider only
tables for categorical variables,
i.e. situation when we compute absolute/ relative **frequencies**
(N, percent, probability)

Tables can show also other indicators, such as central tendency measures or variance for ratio (numeric) variables (mean, median, StD).

See **Map of bivariate analyses configuration**
http://metodykv.wz.cz/QDA1_map_bivaranal.ppt

Bivariate analysis of categorical variables

Relationship of two categorical variables → comparison of sub-groups

(effect of independent variable on dependent variable)

We use similar principle, when dependent variable is *ratio (numeric)* and the independent *categorical*
→ comparison of *means in subgroups*.

Cross-tabulation
= joint frequency distribution

2×2 Contingency Table

elementary set-up

(both variables are dichotomic)

2×2 table

Variable A	Variable B		
	yes	no	Total
yes	a	b	$a + b$
no	c	d	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d$

Marginal frequencies

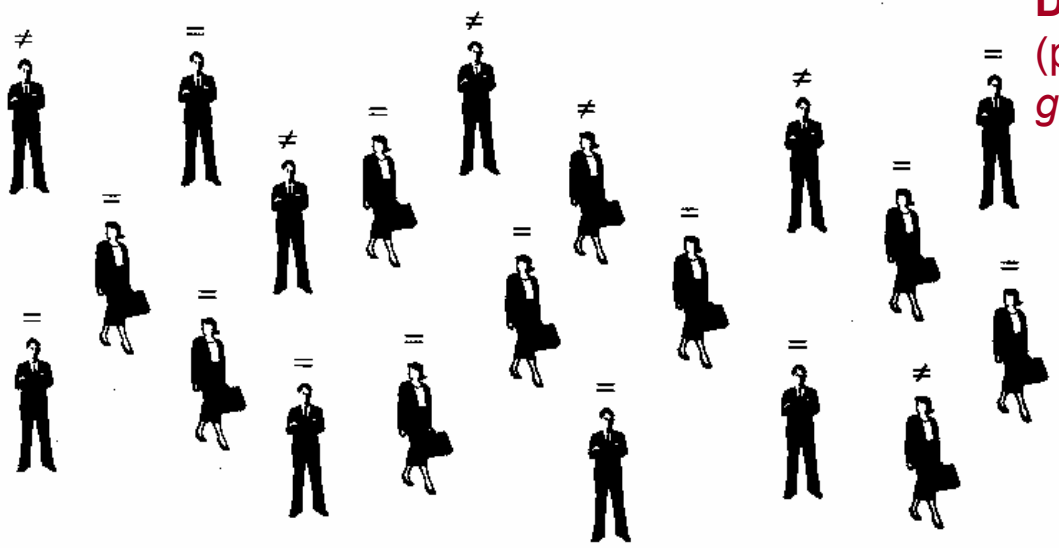
Univariate frequency distribution for each variable

Total number of cases

Source: [Lamser, Růžička 1970: 260]

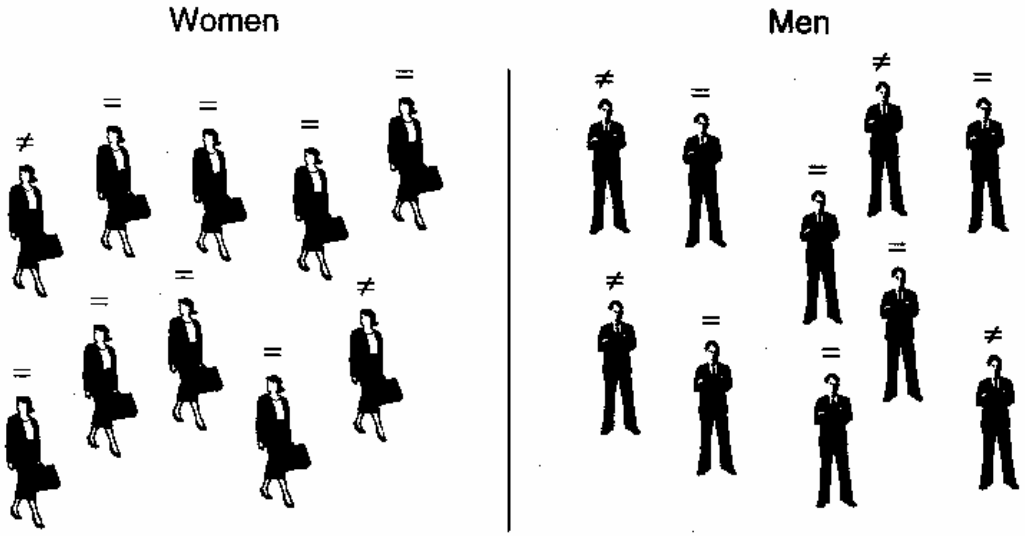
Example: **Percentages in 2x2 table → comparison of sub-populations**

A. Some men and women who either favor (=) sexual equality or don't (≠) favor it.



Dependent variable (preference for *gender equality*)

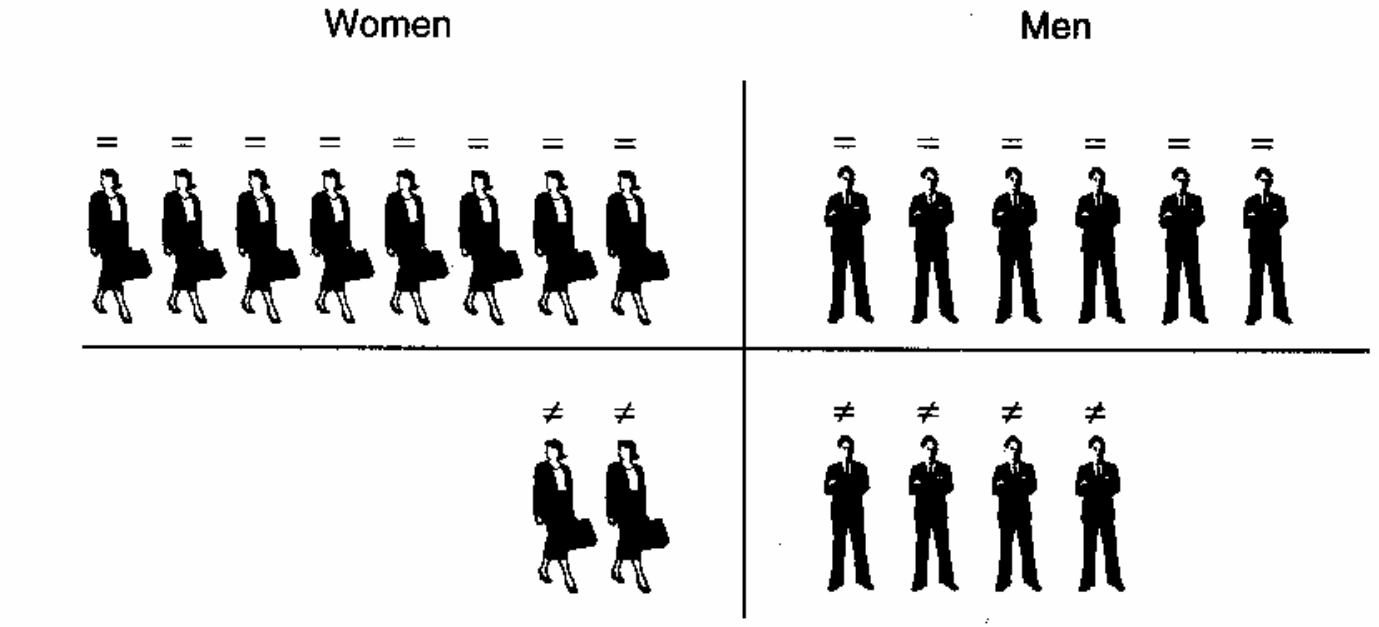
B. Separate the men and the women (the independent variable).



Independent-explanatory variable (*gender-sex of a person*)

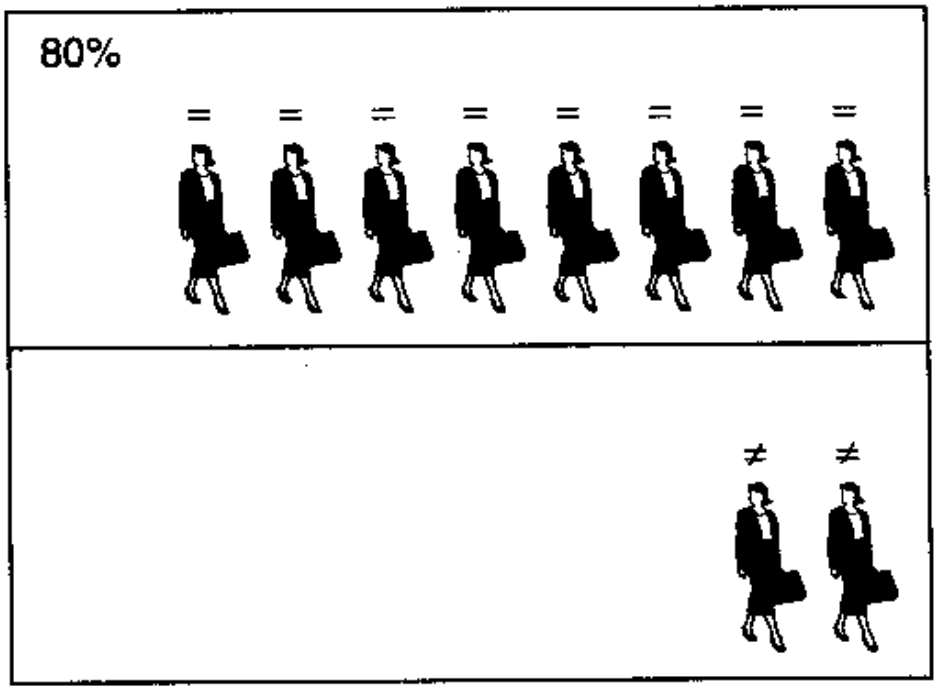
[Babbie 1997: 386]

C. Within each gender group, separate those who favor equality from those who do not (the dependent variable).

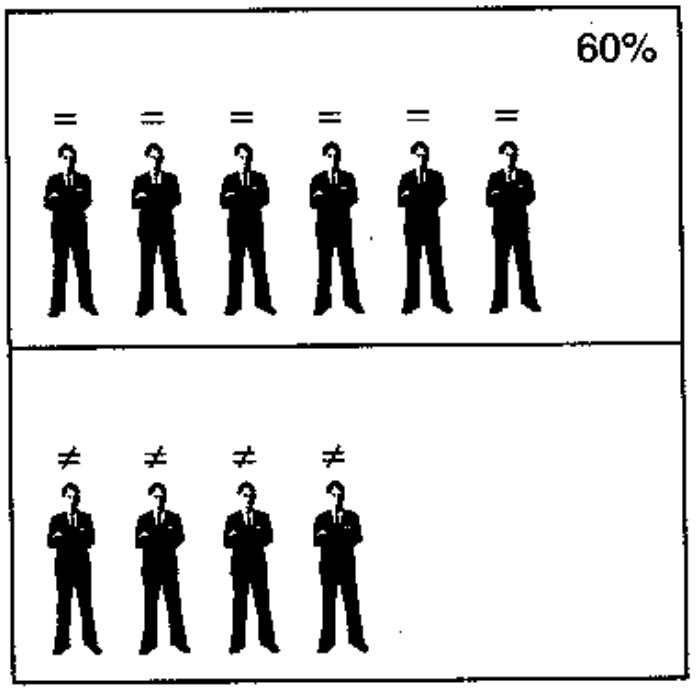


[Babbie 1995: 386-387]

E. What percentage of the women favor equality?



F. What percentage of the men favor equality?



Column percents → for **men** and **women separately**

The difference is 20 percentage points (pp)

G. Conclusions

While a majority of both men and women favored sexual equality, women were more likely than men to do so.

Thus, gender appears to be one of the causes of attitudes toward sexual equality.

	Women	Men
Favor equality	80%	60%
Don't favor equality	20	40
Total	100%	100%

[Babbie 1997: 387]

Relative frequencies – percents in contingency table

- Relative **COLUMN** frequencies = total in each column represents 100%

- Relative **ROW** frequencies = total in each row represents 100%

- There are also **total** percent from the whole table (1 cell from the total) but we don't use them for interpretation of the relationship.
- In the table there are also **marginal** frequencies → univariate distribution for one variable (it depends on whether we use row or column %)

Contingency table

- The situation of four-way (2×2) table can be generalized as *n* × *i*, e.g. 2×3 or 3×3

- When interpreting the table it is important, whether one or both variable is nominal or ordinal.

- **Categorical** variables can be in principle:
  - **dichotomised** → 0/1 (e.g. voted/non-voted)
  - **multinomial** → more than 2 nominal categories (e.g. Studium: HiSo-daily / HiSo-distant / Management&Superv.)
  - **ordinal** → we have ranking of the categories (e.g. Education: 1. Elementary, 2. Vocational training, 3. Secondary w/t diploma, 4. University)
- This distinction results in how we interpret the results (%) and which coefficient of association/correlation we can use.

Before making up a contingency table

always phrase your

research question

(possibly also **hypothesis**).

→ It defines **dependent** and

independent variable

(and possibly also a **control** variable).

Configuration of contingency table

→ **Column** percent:

In the categories of independent variable we show complete (100 %) distribution of dependent variable.

	INDEPENDENT – explanatory variable			
DEPENDENT variable (outcome)		**Gender**		
	Satisfaction	**Men**	**Women**	Total
	1 (not satisfied)	41 % (5)	22 % (2)	7
	2	41 % (5)	11 % (1)	6
	3 (satisfied)	16 % (2)	66 % (6)	8
	Total	100 % (12)	100 % (9)	**21**

Frequently we have **dependent** variable on the left in columns and **independent** (explanatory or predictor) in columns → **column percent**

Illogical configuration of crosstabulation

Zde **řádková** procenta nedávají smysl.
Předchozí tabulku ale lze otočit → spokojenost ve sloupcích, pohlaví v řádcích a pak *řádková* procenta.

		Gender		
	Satisfaction	**Men**	**Women**	Total
	1 (not satisfied)	5 (71 %)	2 (29 %)	7 (100 %)
	2	5 (83 %)	1 (27 %)	6 (100 %)
	3 (satisfied)	2 (25 %)	6 (75 %)	8 (100 %)
	Total	12	9	21 (100 %)

Beliefs can‘t influence gender !

Interpretation of contingency tables (2×2)

Table configuration
„percentage down“

100% is in column → we compare **% → read across row(s)** between categories of subgroups

Table. Church Attendance by gender, USA 1990

	Men	Women
Weekly	28% ⟷	41%
Less often	72	59
100% =	(587)	(746)

Source: General Social Survey, NORC [Table 15-7 in Babbie 1997: 385]

Incorrectly interpreted: “Of the women, only 41 percent attended church weekly, and 59 percent said they attended less often; therefore being a woman makes yon less likely to attend church frequently.”

Correctly interpreted: The conclusion that sex–as a variable–has an effect on church attendance must hinge on a **comparison between men and women**. Specifically, we compare the 41 percent with the 28 percent and note that ***women are more likely than men to attend church weekly***. The **comparison of subgroups, then, is essential in reading an explanatory bivariate table**.

[Babbie 1997: 388]

Interpreting bivariate percentage tables

- "**percentage down**" and "**read across**" in making the subgroup comparisons,
  → COLUMN percentages (mostly preferred)
- or "percentage **across**" and "read **down**" in making subgroup comparisons
  → ROW percentages

[Babbie 1997: 393]

Interpretation of contingency table

dependent variable = it is influenced in the hypothesis, caused (→mostly in rows)

dependent variable(s) = it explains the dependent variable

We show in categories of independent variable complete (100 %) distribution of dependent variable.

Caution! **The direction of causality is always matter of the theory, we can not determine it from the data itself.**

[Treiman 2009]

Interpretation of table for **Ordinal variables**

		Education				
		1 Elementary	2 Vocational training	3 Secondary w/t diploma	4 University	Total
Income (quartiles)	1 I. (< 7 t.)	63	26	29	20	32
	2 II. (7–9 t.)	26	32	16	18	24
	3 III. (9–15 t.)	7	30	29	22	26
	4 IV. (> 15 t.)	4	12	26	41	18
Total		100%	100%	100%	100%	100%

Comparisons are made by across the categories of the independent variable.

Comparing the extreme categories (ignoring the middles) is usually sufficient for assessing ordinal correlation (when both variables are ordinal).

		Education				
		1 Elementary	2 Vocational training	3 Secondary w/t diploma	4 University	Total
Income (quartiles)	1 I. (< 7 t.)	63	26	29	20	32
	2 II. (7–9 t.)	26	32	16	18	24
	3 III. (9–15 t.)	7	30	29	22	26
	4 IV. (> 15 t.)	4	12	26	41	18
Total		100%	100%	100%	100%	100%

The relationship of ordinal variables is often indicated by **cumulation of high % on the diagonal** (but not necessarily!)

		Income (quartiles)				
		1 I. (< 7 t.)	2 II. (7–9 t.)	3 III. (9–15 t.)	4 IV. (> 15 t.)	Total
Education	1 Elementary	63	26	7	4	100%
	2 Vocational training	26	32	30	12	100%
	3 Secondary w/t diploma	29	16	29	26	100%
	4 University	20	18	22	41	100%
Total		32	24	26	18	100%

We can pivot the table through ninety degrees: changing **rows with columns** and **column % with row %.**

Bivariate analysis: how to read the table and what collapsing categories can bring about

Table 15-4 Attitudes Toward the United Nations
"How is the U.N. doing in solving the problems it has had to face?"

	West Germany	Britain	France	Japan	United States
Very good job	2%	7%	2%	1%	5%
Good job	46	39	45	11	46
Poor job	21	28	22	43	27
Very poor job	6	9	3	5	13
Don't know	26	17	28	41	10
	100 %				

Source: "5-Nation Survey Finds Hope for U.N.," *The New York Times*, June 26, 1985, p. 6.

Collapsing categories and omitting „Don‘t know“

Table 15-5 Collapsing Extreme Categories

	West Germany	Britain	France	Japan	United States
Good job or better	48%	46%	47%	12%	51%
Poor job or worse	27	37	25	48	40
Don't know	26	17	28	41	10

Table 15-6 Omitting the "Don't Knows"

	West Germany	Britain	France	Japan	United States
Good job or better	65%	55%	65%	20%	57%
Poor job or worse	35%	45%	35%	81%	44%

[Babbie 1997: 383-84]

Organisation of crosstabulation: conditional probability

Organise the contingency tables (almost) always in the way they express

relative probability, that respondents (cases) will fall into separate categories of dependent variable, **provided that it falls to given category of independent** variable(s).

Probabilities can be expressed as percent (% = probability multiplied by 100).

[Treiman 2009: ch. 1]

Bivariate analysis →

Groups comparison (general principle)

1. Divide cases into adequate groups in terms of their attributes on some independent variable (according your hypothesis, e.g. by education)
2. Describe each subgroup (of independent variable) in terms of some dependent variable using adequate statistics (e.g. percentage /probability, or *for ratio-numerical variables* median, mean)
3. Compare these measures – the dependent variable descriptions among the subgroups.
4. Interpret any observed differences as a association between the independent and dependent variables.

[Babbie 1997: 393]

How to interpret crosstabulation

1. Divide cases into adequate groups according the independent variable (e.g. men/women)
2. Each subgroup is described according attributes of dependent variable (e.g. satisfaction)
3. We read the table in a way, that **we compare subgroups of independent variable** (e.g. men/women) from point of view of **characteristics (statistics such as %) of dependent variable** (e.g. satisfaction).

[Babbie 1997]

Relationship of two variables in crosstabulation

- If both variables are **ordinal**:
  Cumulation of **high values (%) on a diagonal** of the table indicates, that there is (linear) **association** (rank-correlation) between **ordinal** variables.

- **However association can have different form**, e.g. in each column cases can be cumulated into only one cell, which position would be in each column different (i.e. not on diagonal).

[Kreidl 2000]

Interpretation of cross-tabulation

- For ordinal variables: When interpreting percents, it is usually sufficient to compare only extreme values-categories and ignore middle categories.
- If we have ordinal variables it is not reasonable to draw a conclusion from percents within each category of independent variable.
- It is meaningful to **compare of distributions across categories of independent variable**.
- Be careful and don‘t take labels of categories literally (→ operationalisation of variables).

[Treiman 2009]

CROSSTABS basic entry in SPSS

- **Categorical** X **Categorical** variables:
  CROSSTABS *var1-DEPENDENT* **BY** *var2-INDEPENDENT*.
- → counts (absolute frequency), but we need PERCENT which we can have **COLUMN %** or **ROWS** %.

CROSSTABS *var1-dependent* BY *var2-independent* /CELL COL.

or reversed

CROSSTABS *var2-independent* **BY** *var1-dependent* /CELL **ROW**.

- Notice in CROSSTABS it is similar principle as in MEANS:

MEANS *var1-dependent-**numeric*** **BY** *var2-independent-categorical*.

CROSSTABS in SPSS examples: 2×3nominal and 3n×3n table

2×3nominal **CROSSTABS** Church BY Region3 /CELLS **COLUMN** /STATIST **PHI.**

Visits church services (>monthly) * Region3 Crosstabulation

% within Region3

		Region3			
		Prague	Bohemia	Moravia	Total
Visits church services (>monthly)	less often-never	95,8%	92,5%	80,0%	88,0%
	>monthly	4,2%	7,5%	20,0%	12,0%
Total		100,0%	100,0%	100,0%	100,0%

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,200	,000
	Cramer's V	,200	,000
N of Valid Cases		1067	

In 2×3n table we can compare only **one row of „positive" category of dependent variable** (>monthly visits) but each with each category (if independent var. is ordinal we can look at trend only). Suitable **coefficient of association** is **Cramer's V** (or Contingency coefficient, Lambda). Don't use correlation here.

3nominal×3nominal **CROSSTABS** Relig3 BY Region3 /CELLS **COLUMN** /STATIST **PHI.**

Religious denomination (3 ctg.) * Region3 Crosstabulation

% within Region3

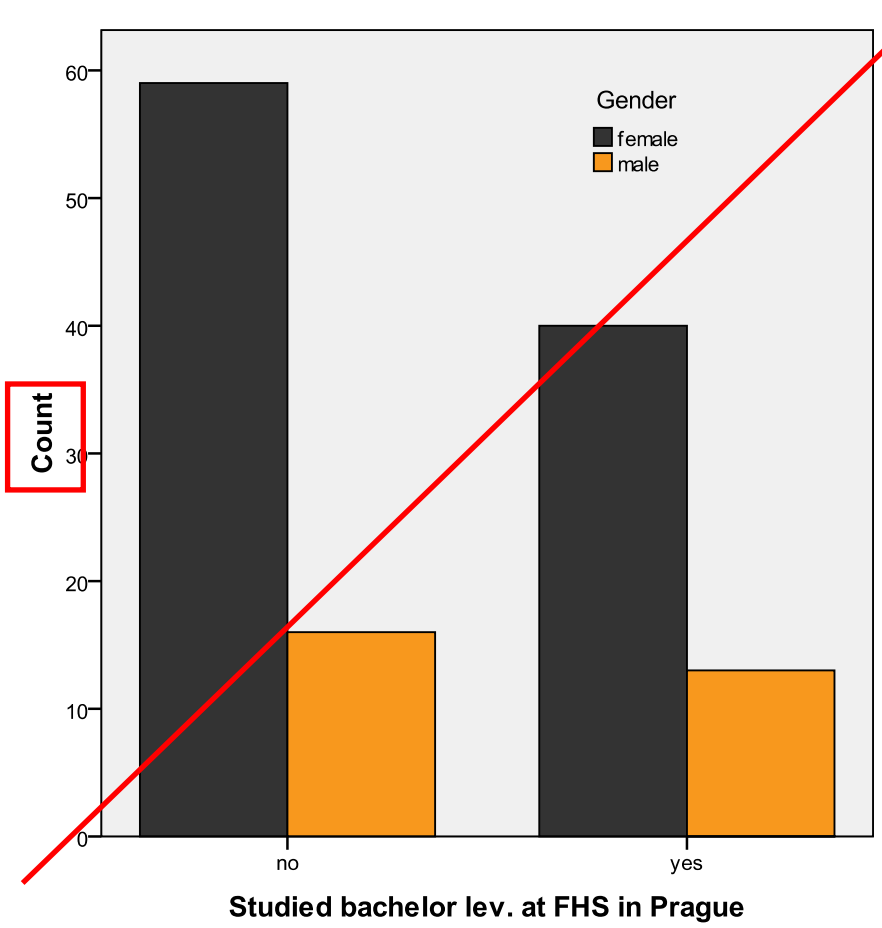
		Region3			
		Prague	Bohemia	Moravia	Total
Religious denomination (3 ctg.)	Roman Catholic	15,5%	20,7%	44,1%	29,3%
	Other Catholic	2,7%	6,4%	4,4%	5,2%
	None/atheist	81,8%	73,0%	51,5%	65,4%
Total		100,0%	100,0%	100,0%	100,0%

Symmetric Measures

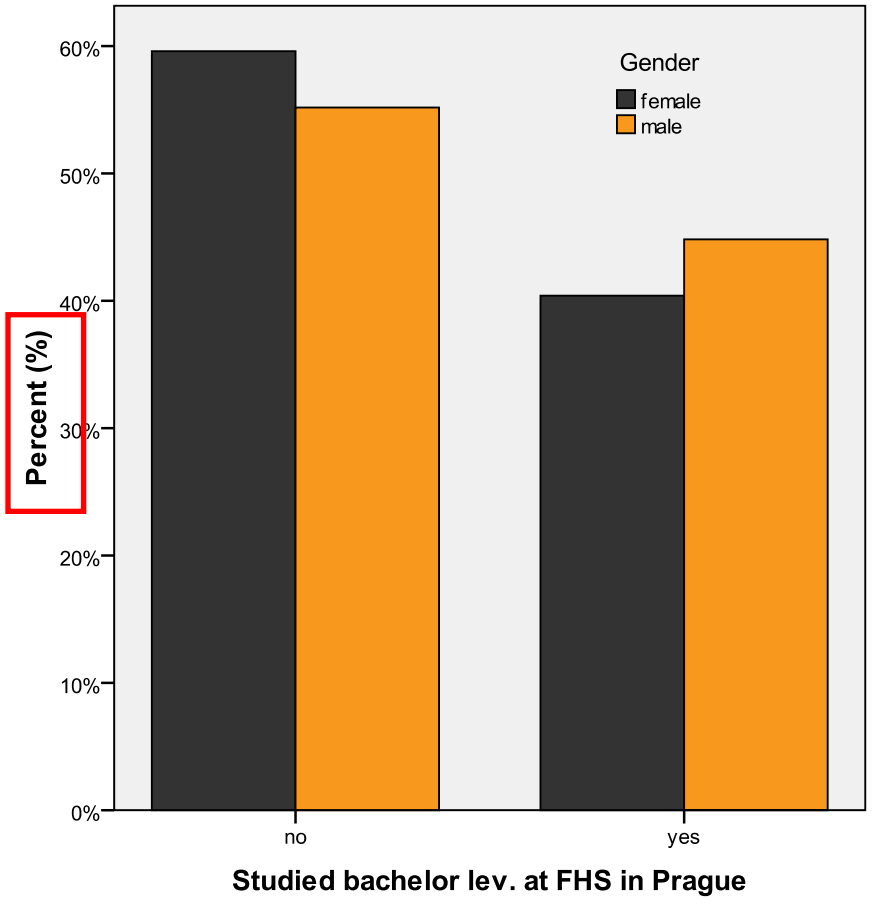
		Value	Approx. Sig.
Nominal by Nominal	Phi	,269	,000
	Cramer's V	,191	,000
N of Valid Cases		1036	

In 3n×3n table, in addition we need to compare **each row category of dependent variable** (but for example here we can focus only on kinds of Catholics leaving Atheists aside). Suitable **coefficient of association** is **Cramer's V** (or Contingency coefficient, Lambda). Don't use correlation here.

Attention – we conduct comparison of sub-groups using **relative** (%) not absolute (count) frequencies



GRAPH /BAR(GROUPED)=**COUNT** BY BC_FHS BY gender.



GRAPH /BAR(GROUPED)=**PCT** BY BC_FHS BY gender.

Source: dataset [TV&Books FHS 2014]

Note:
We can (and in fact we **should) extent bivariate contingency table to multivariate analysis introducing 3rd test variable which effect we control (i.e. 3-rd level data sorting).**

See next presentation

Contingency tables:
third level of data sorting – multivariate analysis and elaboration – introduction

http://metodykv.wz.cz/QDA1_crosstab2multivar.ppt

Measures of association (ordinal correlation)

in contingency table

→ „one number“ measuring **strength of association** between two **categorical** variables

Measures of association in contingency table

- When interpreting as well as measuring strength of relationship of categorical variables, it is crucial whether one or both variables are nominal or ordinal.
- The very basic tool is always comparison of **percent point differences**.
- In addition we can measure strength of mutual relationship using:
- for **nominal** variables coefficients of **association** (Contingency coefficient, Cramer‘s V, Lambda etc.). → it measures
- for **ordinal** variables further (besides coefficients of association) coefficients of **ordinal correlation** (Sperman‘s Rho, Gamma, Kendall‘ Tau B etc.).

How to compute these coefficients in SPSS see later; for more in details *2. Korelace a asociace: vztahy mezi kardinálními/ ordinálními znaky* at http://metodykv.wz.cz/AKD2_korelace.ppt

When our data are from random sample (from a population) then we first need to test for statistical significance of the coefficients of association/correlation (i.e. it is not zero in the whole population) More on this in QDA II.

- We can analyse contingency table also using:
  odds ratio = ratio of mutually conditioned probabilities of different cells

  More on this in QDA II., see *5. Poměry šancí (Odds Ratio)* http://metodykv.wz.cz/AKD2_odds_ratio.ppt

  measures of variation/dispersion for example **Dissimilarity index (Δ)**

  More on this in QDA II., see *9. Míry variability: variační koeficient a další indexy* http://metodykv.wz.cz/AKD2_variacni_koef.ppt

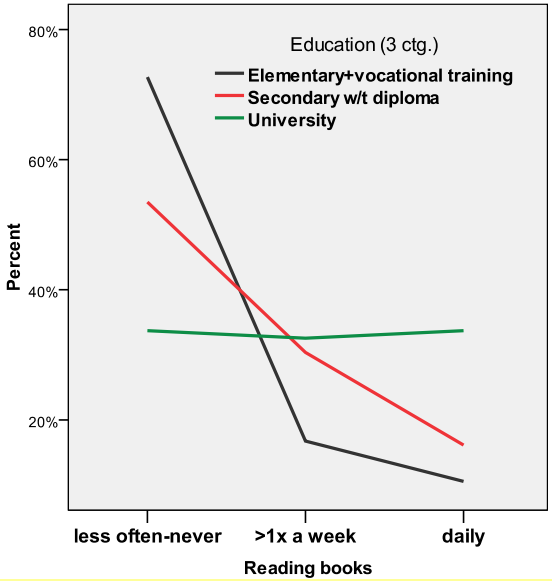
Measures of association for nominal variables

- **Generally coefficients of association**:
  - range from **0** = **no association** to **1** = **complete association** between the variables
  - in principle they say how much variability of one variable can be explained via the other. Note that „explanation“ should be understood as reduction of statistical dispersion of data not as causal interpretation.
  - **There is no direction** (as in case of correlation however some coefficients of association are directional, i.e. you have to assign which variable is dependent)
- **Contingency coefficient C**

  The simplest formula. Don‘t use it to compare associations among tables with different numbers of categories.
- **Cramer's V** (**CV** or **Cr**) generally recommended
- When **both variables are dichotomic** (2×2 table) we use **Phi coefficient** (for 2×2 table it is equivalent to CV)
- **Lambda Λ** (symmetric/ asymmetric) measures the percentage improvement when prediction of one variable is done on the basis of values of the other (in both directions – symmetric or just for predicting dependent variable – asymmetric)
- All these coefficients are available in SPSS command CROSSTABS (see later)
- You can use them also for ordinal variables but in that case you can also use correlation coefficient.

3o×3o table (both variables **ordinal**)

		edu3 Education (3 ctg.)			
		1 Elementary+ vocational training	2 Secondary w/t diploma	3 University	Total
Read3 Reading books	1 less often-never	72,7%	53,5%	33,7%	62,4%
	2 >1x a week	16,7%	30,4%	32,6%	23,2%
	3 daily	10,5%	16,1%	33,7%	14,4%
Total		100,0%	100,0%	100,0%	100,0%



The highest proportion (in the rows) is mostly **on the diagonal** indicating ordinal correlation (linear trend in mutual ranking). However, **this trend is not absolute**: there is 40 % points difference between the most distant categories (Element./vocc. and Univ.) within „Less often-never“ but only 22 % points between „Daily“ readers. See the graph.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Nominal by Nominal	Phi	,264			,000
	Cramer's V	,186			,000
	Contingency Coefficient	,255			,000
Ordinal by Ordinal	Kendall's tau-b	,228	,025	8,985	,000
	Gamma	,392	,042	8,985	,000
	Spearman Correlation	,243	,028	8,985	,000[c]
Interval by Interval	Pearson's R	,240	,028	8,465	,000[c]
N of Valid Cases		1202			

Coefficients of **association**

Coefficients of **ordinal correlation**

When our **data is from random sample** (i.e. not whole population) we have to in addition first test statistical **hypothesis, that the coefficient is not zero** (i.e. it is not zero in the whole population and not only in our sample). **Approx. Significance** (also ***p***) is here < 5% → we reject the null hypothesis that Gamma/TauB/Spearman is zero in whole population). More on this in QDA II.

Both variables are ordinal so correlation can be measured (and compared with nominal association e.g. Cramer‘s V).

Gamma can be recommended but it usually gives higher number so compare it with other coefficients.
(Spearman‘s Rho is rank-order version of Pearson‘s R which is only for ratio-numerical variables.)

CROSSTABS Read3 BY edu3 /**STATISTICS CC Phi GAMMA CORR BTAU**.

We will further elaborate these bivariate **„zero order“ contingency tables/ associations** into **„first order conditional“ tables/ associations**

Contingency tables: multivariate analysis and elaboration – introduction to third level of data sorting

http://metodykv.wz.cz/QDA1_crosstab2multivar.ppt

n×n table: when at least one variable is **multi-nominal**

- The principle is the same as with ordinal variables but **we can NOT** compute correlation, **only coefficients of association** (Contingency coefficient, Cramer‘s V, Lambda etc.).
- If only 3rd – controlling variable Z is nominal (and the others are ordinal), then we can compute correlation in these groupings defined by Z and mutually compare them (Is there trend in correlation along Z categories?).
- **When interpreting proportional differences (%) in nominal variables we have to care about ALL categories** of dependent variable as well as independent variable.
- The situation is easier when at least one variable is ordinal because then we can look (only) for trend between categories. However, the differences can be present in other (nonlinear) form.
- It is optimal when dependent variable is dichotomic or ordinal.
- When dependent variable is dichotomic (perhaps we can collapse some categories), then it is equivalent of means comparison in between subgroups (if dependent variable coded as 0/1 then means represent probabilities).

Examples of **bivariate** association/correlation in contingency table for different types of categorical variables

2×2

		vek50plus		
		<49	>50	Total
Čte knihy >týdně	<měsíc	65,5%	59,8%	62,5%
	>týdně	34,5%	40,2%	37,5%
Total		100,0%	100,0%	100,0%

		Value
Nominal by Nominal	Phi	,059
	Cramer's V	,059
	Contingency Coefficient	,059

2×3nominal

		Náboženská orientace			
		Římskokatolické	Katolické_ostat	žádné	Total
Čte knihy >týdně	<měsíc	58,1%	51,7%	65,0%	62,3%
	>týdně	41,9%	48,3%	35,0%	37,7%
Total		100,0%	100,0%	100,0%	100,0%

		Value
Nominal by Nominal	Phi	,081
	Cramer's V	,081
	Contingency Coefficient	,081
N of Valid Cases		1155

2×3ordinal

		vek3			
		19-29	30-49	50+	Total
Čte knihy >týdně	<měsíc	64,7%	66,0%	59,8%	62,5%
	>týdně	35,3%	34,0%	40,2%	37,5%
Total		100,0%	100,0%	100,0%	100,0%

		Value
Nominal by Nominal	Phi	,060
	Cramer's V	,060
	Contingency Coefficient	,060
Ordinal by Ordinal	Gamma	,096
	Spearman Correlation	,054

3o×3o

		vek3			
		19-29	30-49	50+	Total
Čte knihy, 3k	>rok/nikdy	45,9%	43,6%	39,6%	41,9%
	<měsíc	18,8%	22,5%	20,2%	20,6%
	>týdně	35,3%	34,0%	40,2%	37,5%
Total		100,0%	100,0%	100,0%	100,0%

		Value
Nominal by Nominal	Phi	,068
	Cramer's V	,048
	Contingency Coefficient	,068
Ordinal by Ordinal	Gamma	,084
	Spearman Correlation	,058

For tables larger than 2×2 **you can always use Cramer‘s V** and Contingency coefficient.

Note: **if correlation absent, there still can be (nominal) association**

- If ordinal dependency – correlation is absent, it doesn't imply statistical independency. It only means that there is no ordinal relationship (~ linearity). There still can be strong **association**, i.e. joint frequency is e.g. cumulated in one cell (or several cells out of diagonal or without any other „trend“).
- This will be indicated by **significant coefficient of association** (e.g. Cramer‘s V) whereas **ordinal correlation is around zero** (e.g. Gamma)**.**
- **Only absence of nominal dependency – association represents (total) statistical independency**. (e.g. CV = 0)
- → **compute both** coefficients of **association** (Cramer‘s V etc.) and ordinal **correlation** (Gamma etc.) and **compare them**.

Coeff. of association/correlation in bivariate analysis
in SPSS within CROSSTABS

- Within CROSSTABS we can compute several measures of bivariate association and correlation (as well as separately in categories of controlling factor – see presentation *4. Contingency tables: multivariate analysis and elaboration*).
- **For nominal variables coefficients of association** (they range 0-1 and have no direction):

```
CROSSTABS var1 BY var2 /CELLS COL /STATISTICS CC PHI.
```

Coefficients of association: **CC** = Contingency coefficient, **PHI** = *Cramer V* (+ equivalent for dichotomised variables is *Phi*); there are also other coefficients of association and correlation (e.g. *Lambda*).

- **for ordinal variables** (in addition to association coeff.) **ordinal correlation** (they range -1–0–1 and **direction**):

```
CROSSTABS var1 BY var2 /CELLS COL
/STATISTICS CC PHI GAMMA CORR BTAU.
```

Correlation coefficients: **GAMMA** = Goodman&Kruskal *Gamma*, **BTAU** = Kendaull *Tau B*, **CORR** = Spearman *Rho* (+ Pearson correl. coef. R for ratio variables)

- **Notice, if we don‘t find correlation, it doesn't mean that, there is no (strong) relationship–association.**
- Moreover with ordinal variables comparison of correlations and coefficients of association can help us indicate what is the relationship (nonlinearity).

How to preset tables (some rules)

For more details see

[Treiman 2009: Chapter 1]

Rules for presenting tables

- Only percents say not enough. Always include **number of cases on which percentages are based**. → **Don't hold back counts (absolute frequency)**
  Optimally we show counts for all cells (in brackets) but it is space consuming so **marginal counts are mostly enough** (row or column) from which a reader can reconstruct a table of frequencies and possibly reorganize data. But uncompromisingly you have to minimally quote **the whole number of valid cases + how many missing values are there.**

Table 1. Percent Militant by Religiosity Among Urban Negros in the USA, 1964

	Very religious	Somewhat religious	Not very religious/Not at all	Total
Militant	27%	30%	48%	33%
Nonmilitant	73	70	52	67
Total	100%	100%	100%	100%
N	(230)	(532)	(231)	(993)

Source: adapted from table 1.2 in [Treiman 2009: 10]

- Always **include percentages totals** (the row or column of 100%). Together with **% signs on the top row** (column) clearly indicates that it is percentage table and how it is organised.

Source: adapted from [Treiman 2009: 9-10]

Rules for presenting tables

- When constructing a table check the accuracy of your entries: count up the entries in each row confirming that they correspond to the column marginal (the same for rows, and for total marginals and grand total).

- **Round decimal numbers of %.** Whole percentages are precise enough.
  23,48 % → **23 %**

[Treiman 2009: 9-10]

Rules for presenting tables

[Kreidl 2000; Babbie 1997; Treiman 2009]

- Table must have a **heading** and **variables labeled** (rows and columns).
- Quote **original content of the variable**, notably when it is an attitude → **quote wording of question** as well as possible answers from questionnaire (perhaps in a note).
- Quote the **source of the data**.
- Quote the **grand total of valid cases** (marginal frequencies - counts).
- Quote, how **percentages were computed** (percentage base), in table using % state at least **grand total count** (N)
- Don't use % and counts concurrently in each cell.
- Remark if some **categories were omitted** (e.g. „Don‘t know“).
- **Missing values** → always quote how many people didn‘t answer (or generally how many observations we are missing). But it is not necessary to keep it in percentage base, i.e. we use only valid cases (see how to cope with missing values)

Don‘t forget to quote in the heading:

- type of table e.g. Percent distribution ... or ... (%)
- variables included in the table, e.g. Religiosity and education level
- From what sample is the data → to what population it can generalised
- year of data collection

Example: Percent users of marihuana by education attainment, secondary students in CR, 1997.

References

- Babbie, E. 1995. **Elementary Analyses**. (chapter 15) Pp. 375-394 in *The Practice of social Research. 7th Edition*. Belmont: Wadsworth.
- Treiman, D. J. 2009. *Quantitative data analysis: doing social research to test ideas*. San Francisco: Jossey-Bass. (chapters 1. **Cross-tabulations** and 2. **More on tables**)
- de Vaus, D., A. (1985) 2002. *Surveys in Social Research, Fifth Edition*. St Leonards NSW: Allen & Unwin / London: Routledge.
(chapter 11. **Bivariate analysis: crosstabulation**).