



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdelávání
pro konkurenceschopnost



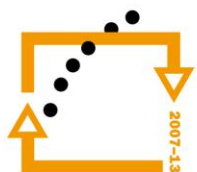
INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Seminář z finanční matematiky

Martin Řezáč
2013



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



**OP Vzdělávání
pro konkurenceschopnost**



INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

Tento učební text vznikl
za přispění Evropského
sociálního fondu a
státního rozpočtu ČR
prostřednictvím
Operačního programu
Vzdělávání pro
konkurenceschopnost v
rámci projektu
Univerzitní výuka
matematiky v měnícím
se světě
(CZ.1.07/2.2.00/15.0203).

Obsah:

1.	Metody průzkumové analýzy	4
2.	Aplikace logistických regresních modelů	85
3.	Neuronové sítě.	120
4.	Bayesovské sítě.	152
5.	Bayesovské sítě – aplikace.	174
6.	Výpočet pojistného v životním pojištění	210
7.	Investování do akcií	238
8.	Fundamentální analýza	278
9.	Technická analýza	313
10.	Oceňování nemovitého majetku	365
11.	Black-Scholesův model	398
12.	Oceňování derivátů, opce	415
13.	Reference.	442

1. Metody průzkumové analýzy



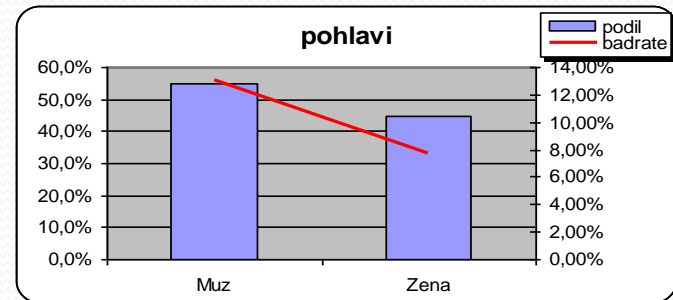
Explorační analýza – PROČ?

- Je třeba pochopit data:
 - najít chyby v datech
 - najít vzory v datech
 - najít porušení statistických předpokladů, testování hypotéz
 - ...a především proto, že pokud to neuděláme, budeme mít velké problémy později.

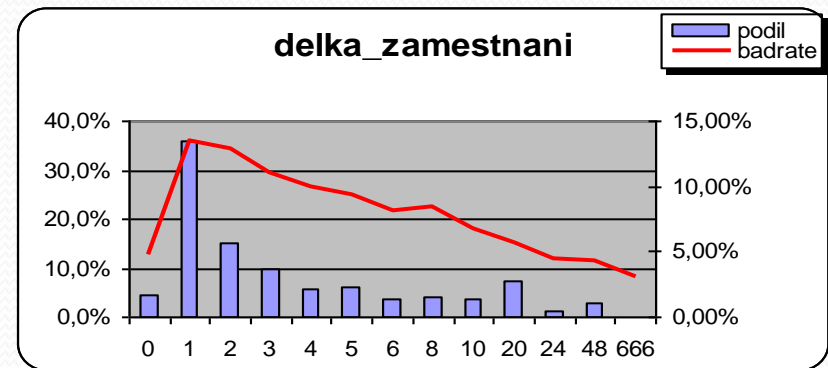
Explorace dat - jednorozměrná

□ Frekvenční tabulky, histogramy:

	pocet	podil	badrate
Muz	248 768	55,0%	13,08%
Zena	203 194	45,0%	7,69%
Total	451 962	100,0%	10,66%



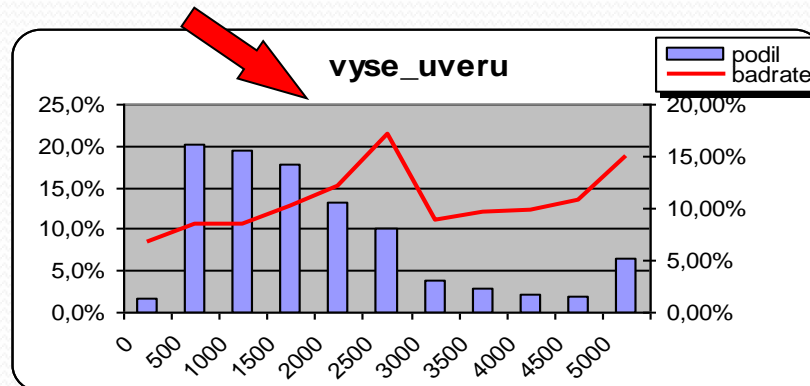
delka_zamestnani	pocet	podil	badrate
0	20 825	4,6%	4,69%
1	163 144	36,1%	13,43%
2	67 462	14,9%	12,80%
3	43 778	9,7%	10,97%
4	26 256	5,8%	10,01%
5	27 526	6,1%	9,32%
6	15 893	3,5%	8,16%
8	18 036	4,0%	8,39%
10	17 195	3,8%	6,72%
20	33 641	7,4%	5,60%
24	5 176	1,1%	4,48%
48	12 934	2,9%	4,28%
666	96	0,0%	3,13%
Total	451 962	100,0%	10,66%



Explorace dat - jednorozměrná

- výše úvěru vs. bad rate

OK? Nebo je to způsobeno jiným faktorem???



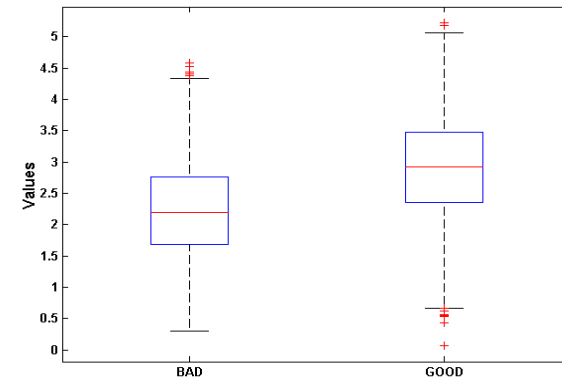
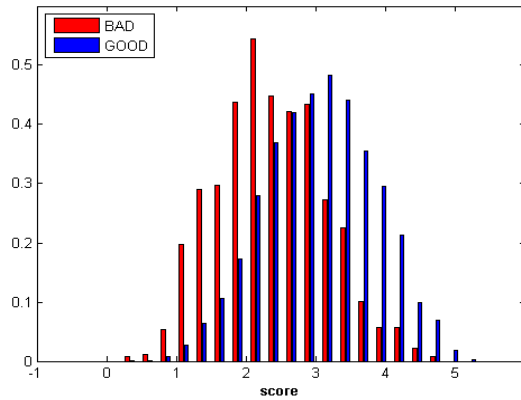
Explorace dat - jednorozměrná

- spojité proměnné:
 - Průměr
 - Modus
 - Kvantily
 - Rozptyl
 - Min./maximální hodnota

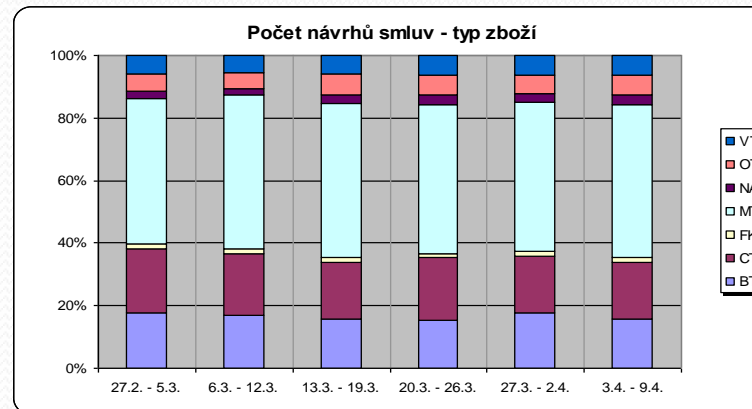
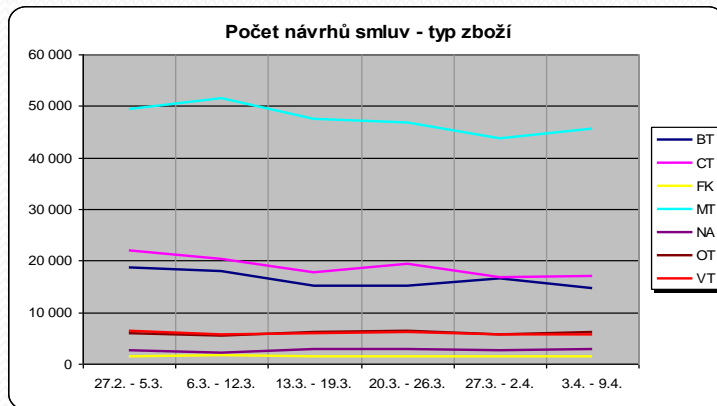
- je vhodná kategorizace

Explorace dat - jednorozměrná

□ Histogramy, box ploty



□ Stabilita v čase



Explorace dat - vícerozměrná

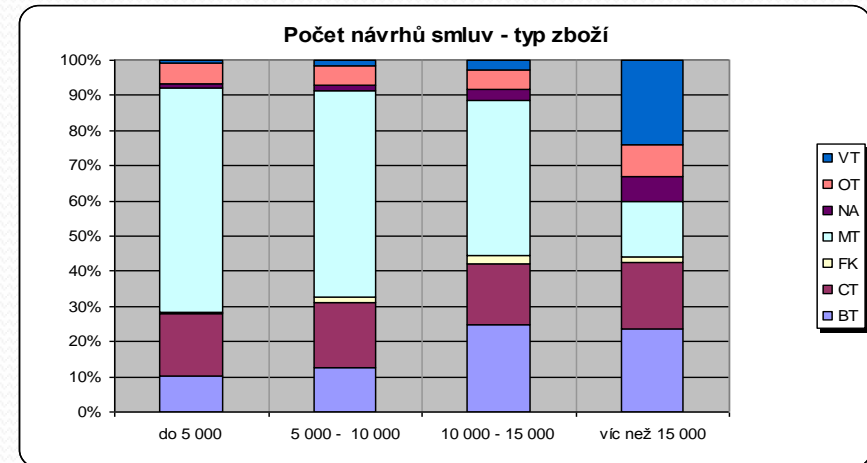
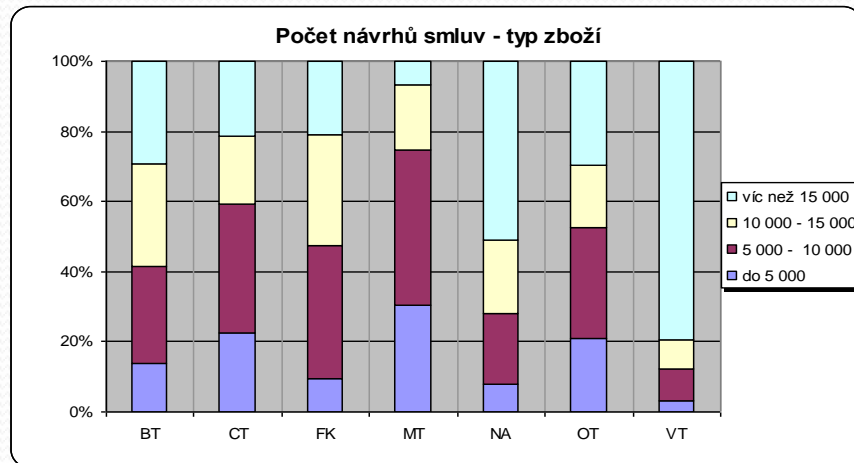
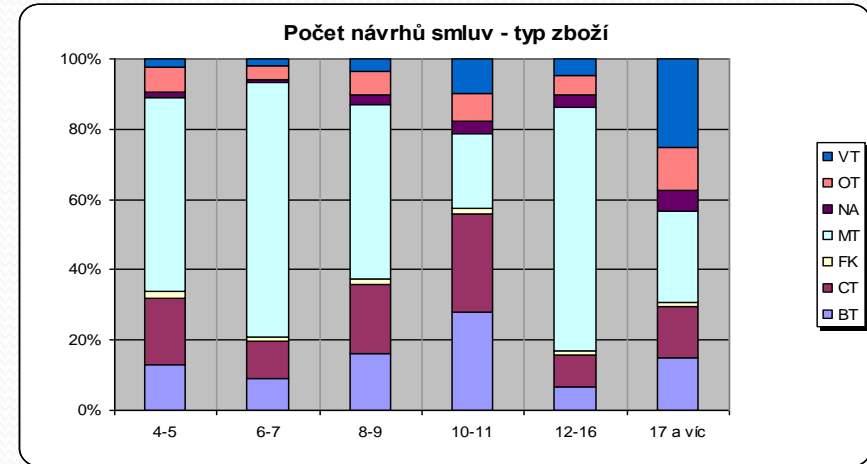
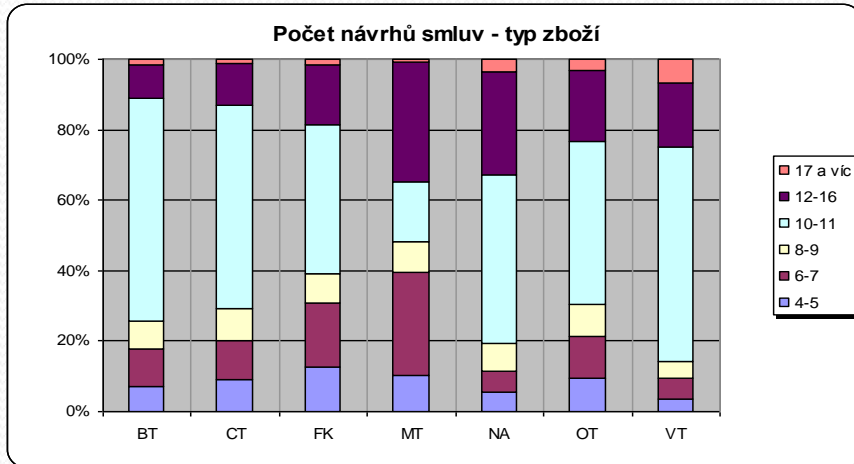
□ Kontingenční tabulky

	do 5 000	5 000 - 10 000	10 000 - 15 000	víc než 15 000
BT	4 291	8 581	9 176	9 044
CT	7 587	12 493	6 500	7 236
FK	258	1 017	851	557
MT	27 191	39 551	16 524	5 992
NA	426	1 088	1 114	2 737
OT	2 478	3 689	2 103	3 475
VT	384	1 001	963	9 086

row%	do 5 000	5 000 - 10 000	10 000 - 15 000	víc než 15 000
BT	13,8%	27,6%	29,5%	29,1%
CT	22,4%	36,9%	19,2%	21,4%
FK	9,6%	37,9%	31,7%	20,8%
MT	30,5%	44,3%	18,5%	6,7%
NA	7,9%	20,3%	20,8%	51,0%
OT	21,1%	31,4%	17,9%	29,6%
VT	3,4%	8,8%	8,4%	79,5%

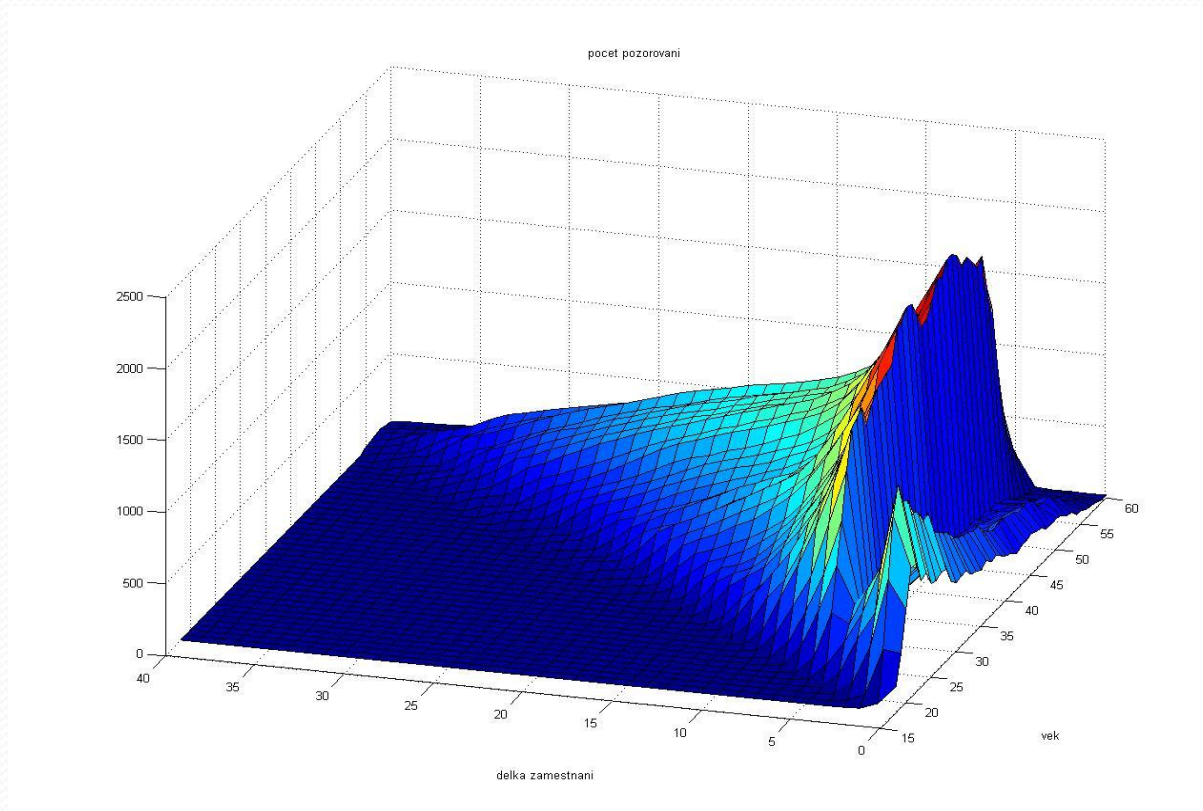
col%	do 5 000	5 000 - 10 000	10 000 - 15 000	víc než 15 000
BT	10,1%	12,7%	24,6%	23,7%
CT	17,8%	18,5%	17,5%	19,0%
FK	0,6%	1,5%	2,3%	1,5%
MT	63,8%	58,7%	44,4%	15,7%
NA	1,0%	1,6%	3,0%	7,2%
OT	5,8%	5,5%	5,6%	9,1%
VT	0,9%	1,5%	2,6%	23,8%

Explorace dat - vícerozměrná



Explorace dat - vícerozměrná

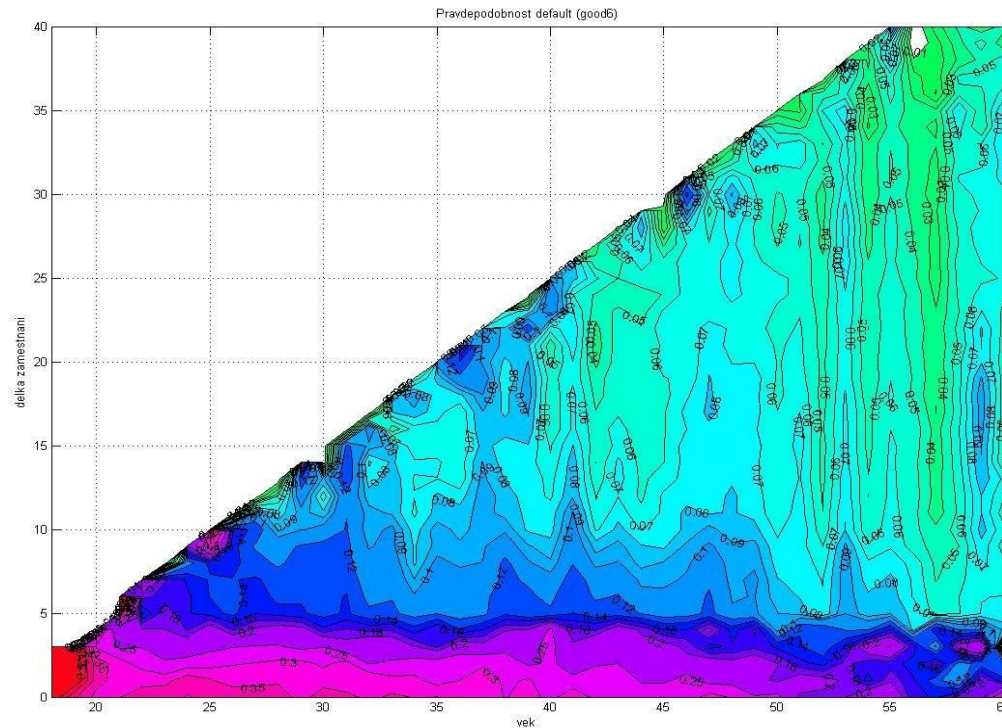
□ Věk vs. délka zaměstnání



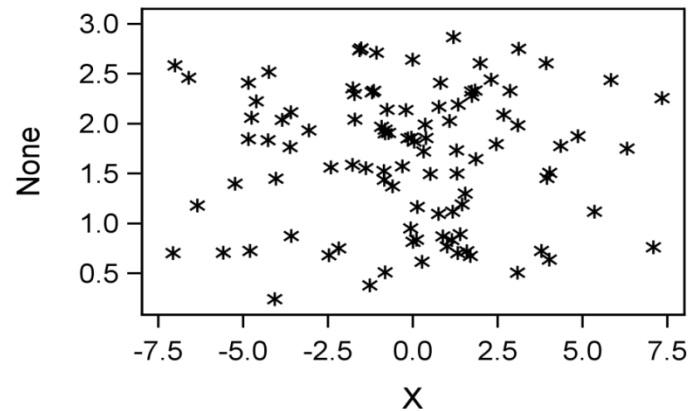
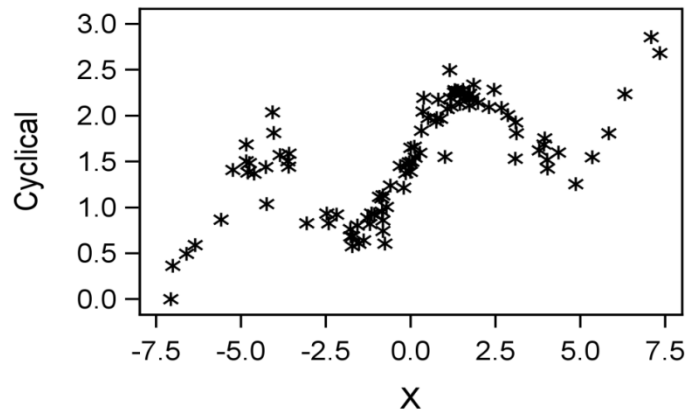
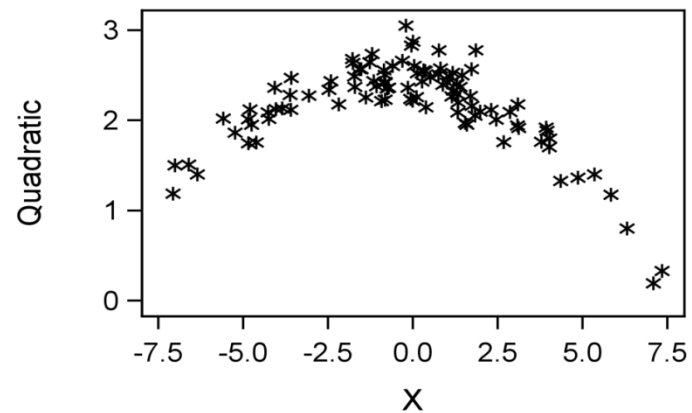
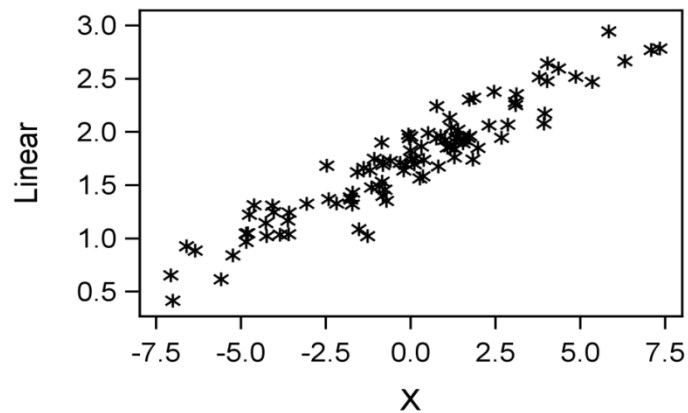
5 let
...defaultní
hodnota???

Explorace dat - vícerozměrná

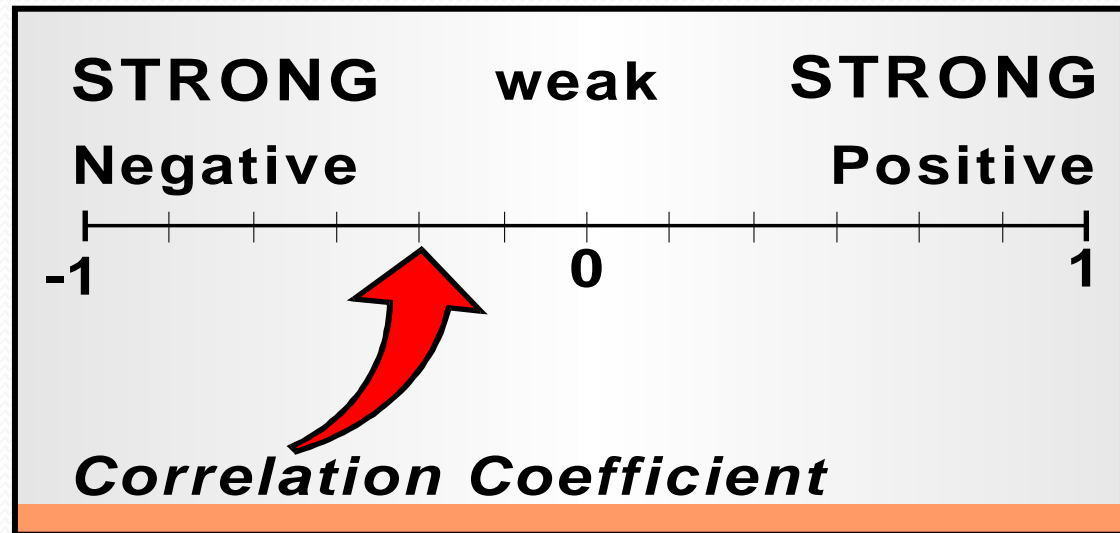
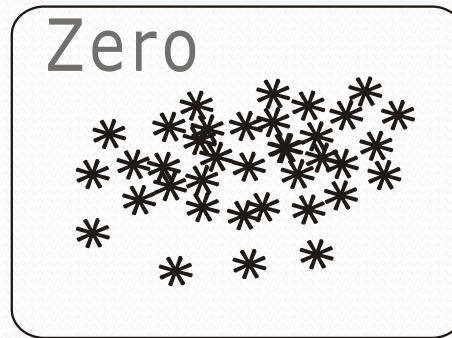
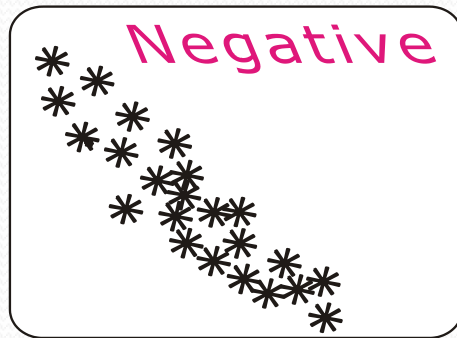
- Věk vs. délka zaměstnání vs. default



Relationships between Continuous Variables – scatter plots

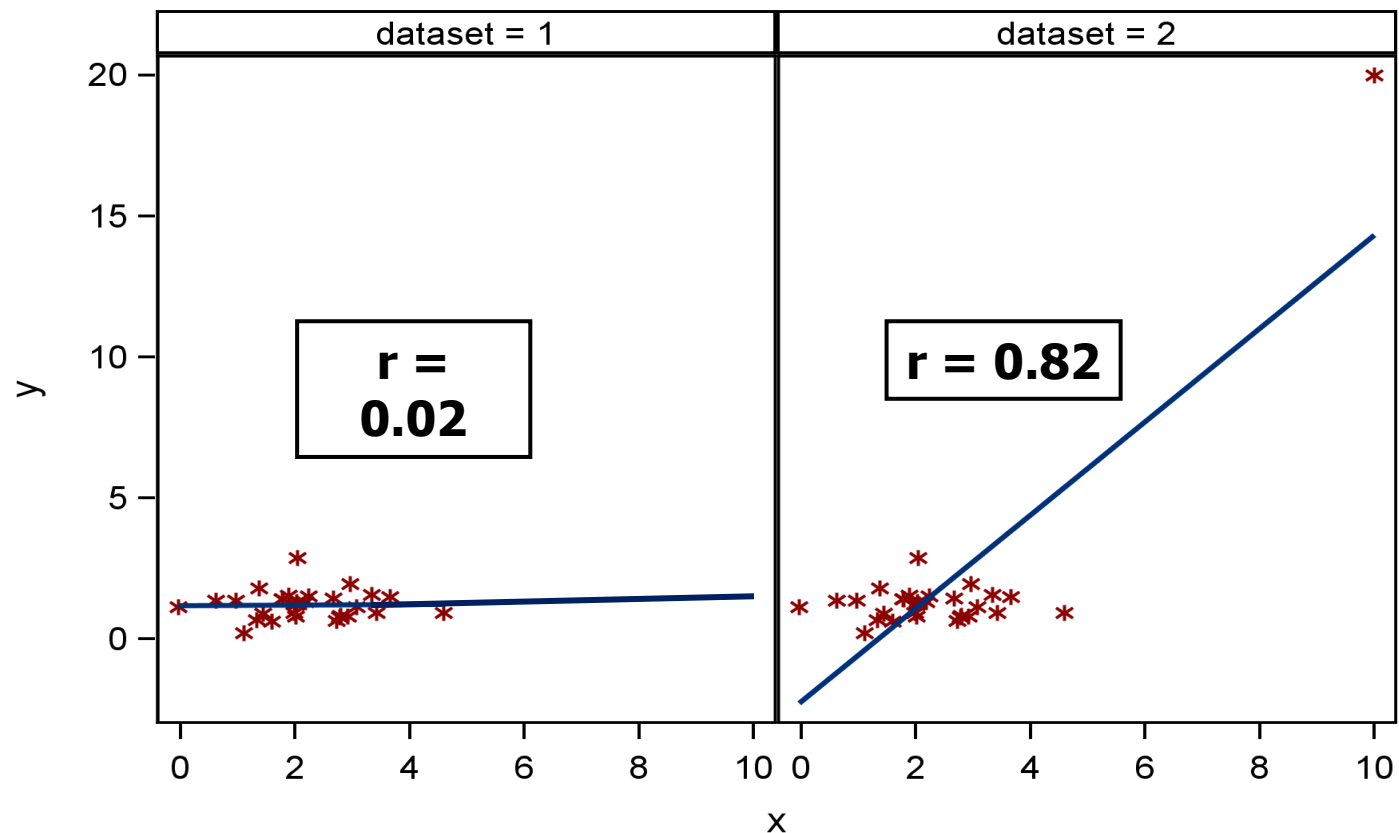


Correlation



Extreme Data Values

Odlehlé (extrémní) hodnoty mohou zcela zkreslit výsledky analýzy.



Diskriminační síla proměnných pro prediktivní modely

Weight of evidence, information value

r ... number of levels (categories) of the categorical variable

g_i ... number of "goods" the in i -th category

b_i ... number of "bads" the in i -th category

$G := \sum g_i$... total number of "goods"

$B := \sum b_i$... total number of "bads"

Weight of evidence for the i -th category:

$$woe_i = \ln (g_i / G) - \ln (b_i / B)$$

Information value for the i -th category:

$$Inf_val_i = [(g_i / G) - (b_i / B)] \cdot$$

woe_i

Total information value for the corresponding variable:

$$Inf_val = \sum inf_val_i$$

Diskriminační síla proměnných

Incorporation Date												
Raw	RegVar	Percant	B	G	TOT	G/B Odds	%Good	%Bad	Bad Rate	WoE	IV	
0 & NOI	inc_1	12%	139	952	1091	7	11%	19%	12,7%	-0,557	0,046116	
1	inc_2	13%	133	1073	1206	8	12%	19%	11,0%	-0,394	0,023731	
2-7	miss	42%	299	3601	3900	12	42%	42%	7,7%	0,007	2,04E-05	
8-15	inc_3	22%	108	1942	2050	18	23%	15%	5,3%	0,408	0,030887	
16+	inc_4	11%	39	1019	1058	26	12%	5%	3,7%	0,781	0,050288	
Total			718	8587	9305	12			7,7%		0,151	

- **<0.02** **unpredictive**
- **0.02 – 0.1** **weak**
- **0.1 – 0.3** **medium**
- **0.3 – 0.5** **strong**
- **> 0.5** **too high ...je třeba prověřit, pravděpodobně je něco špatně**

Diskriminační síla proměnných

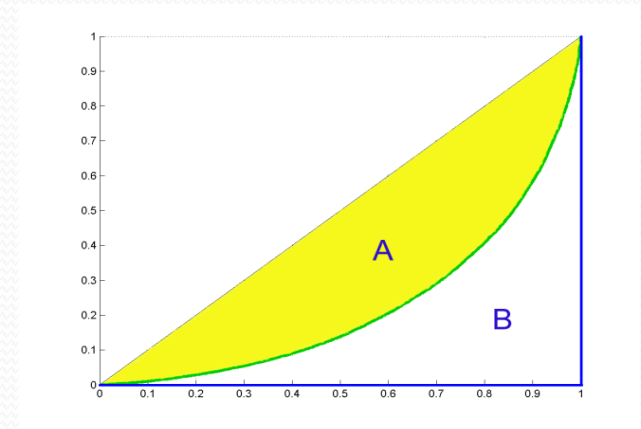
□ Lorenzova křivka, Giniho index

$$x = F_{m.BAD}(a)$$

$$y = F_{n.GOOD}(a), a \in [L, H].$$

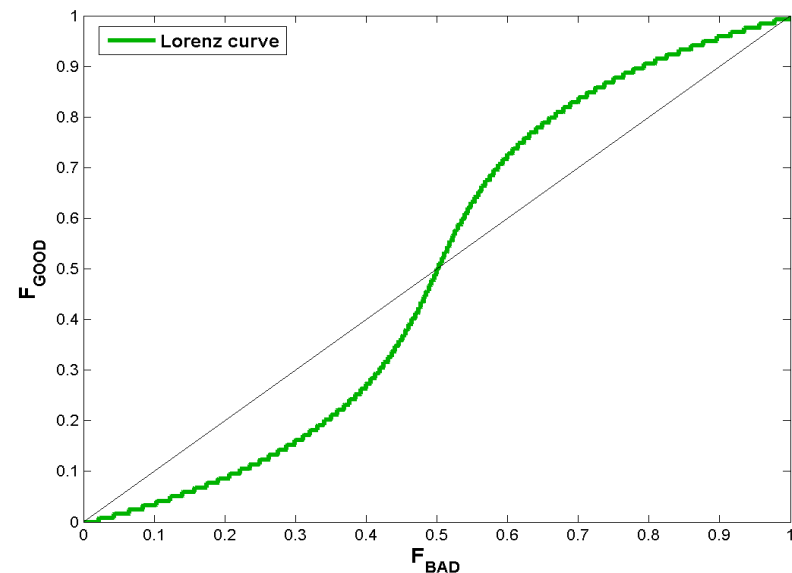
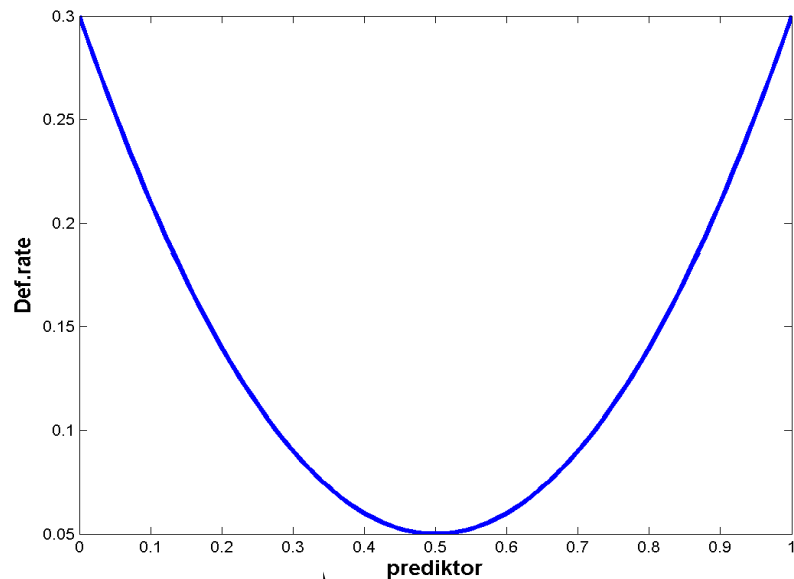
$$Gini = \frac{A}{A+B} = 2A$$

$$Gini = 1 - \sum_{k=2}^{n+m} (F_{m.BAD_k} - F_{m.BAD_{k-1}}) \cdot (F_{n.GOOD_k} + F_{n.GOOD_{k-1}})$$



Diskriminační síla proměnných

- Lorenzova křivka ...kontrola monotónnosti vysvětlované proměnné (def. rate) na dané vysvětlující proměnné



Kategorizace (WOE)

Diskriminační síla proměnných

□ Giniho index

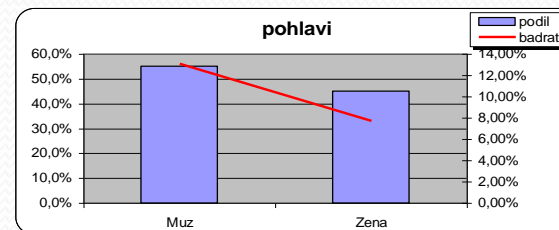
- < 0.05 uninformative
- $0.05 - 0.1$ weak
- $0.1 - 0.2$ medium
- $0.2 - 0.5$ strong
- > 0.5 too high ...je třeba prověřit, pravděpodobně je něco špatně

Diskriminační síla proměnných

pohlavi Gini: **0,1401**

Info.Value: **0,0828**

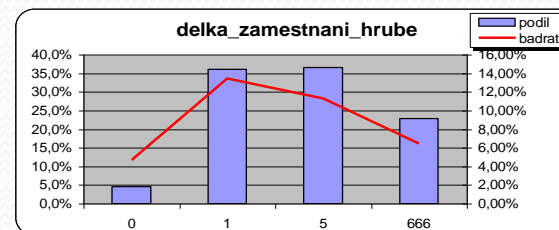
	pocet	podil	badrate
Muz	248 768	55,0%	13,08%
Zena	203 194	45,0%	7,69%
Total	451 962	100,0%	10,66%



delka_zamestnani_hrube Gini: **0,1611**

Info.Value: **0,1100**

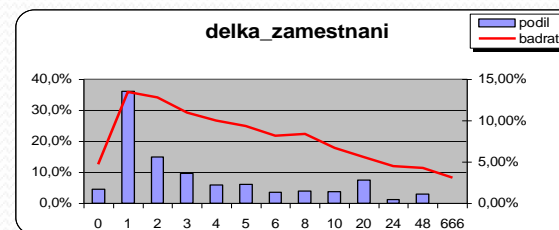
	pocet	podil	badrate
0	20 825	4,6%	4,69%
1	163 144	36,1%	13,43%
5	165 022	36,5%	11,29%
666	102 971	22,8%	6,45%
Total	451 962	100,0%	10,66%



delka_zamestnani_jemne Gini: **0,1762**

Info.Value: **0,1285**

delka_zamestnani	pocet	podil	badrate
0	20 825	4,6%	4,69%
1	163 144	36,1%	13,43%
2	67 462	14,9%	12,80%
3	43 778	9,7%	10,97%
4	26 256	5,8%	10,01%
5	27 526	6,1%	9,32%
6	15 893	3,5%	8,16%
8	18 036	4,0%	8,39%
10	17 195	3,8%	6,72%
20	33 641	7,4%	5,60%
24	5 176	1,1%	4,48%
48	12 934	2,9%	4,28%
666	96	0,0%	3,13%
Total	451 962	100,0%	10,66%



SAS: the FREQ Procedure

- The FREQ procedure can do the following:
 - produce one-way to n -way frequency and crosstabulation (contingency) tables
 - compute chi-square tests for one-way to n -way tables and measures of association and agreement for contingency tables
 - automatically display the output in a report and save the output in a SAS data set
- General form of the FREQ procedure:

```
PROC FREQ DATA=SAS-data-set  
<option(s)>;  
    TABLES variable(s) </option(s)>;  
RUN;
```

The FREQ Procedure

A FREQ procedure with no TABLES statement generates one-way frequency tables for all data set variables.

```
proc freq data=orion.sales;  
run;
```

This PROC FREQ step creates a frequency table for the following nine variables:

- **Employee_ID**
- **First_Name**
- **Last_Name**
- **Gender**
- **Salary**
- **Job_Title**
- **Country**
- **Birth_Date**
- **Hire_Date**

The TABLES Statement

The TABLES statement specifies the frequency and crosstabulation tables to produce.

```
proc freq data=orion.sales;  
  tables Gender Country;  
run;
```

one-way
frequency tables

An asterisk between variables requests a *n*-way crosstabulation table.

```
proc freq data=orion.sales;  
  tables Gender*Country;  
run;
```

two-way
frequency table

The TABLES Statement

A one-way frequency table produces frequencies, cumulative frequencies, percentages, and cumulative percentages.

```
proc freq data=orion.sales;  
  tables Gender Country;  
run;
```

The FREQ Procedure

Gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	68	41.21	68	41.21
M	97	58.79	165	100.00

Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AU	63	38.18	63	38.18
US	102	61.82	165	100.00

The TABLES Statement

An n -way frequency table produces cell frequencies, cell percentages, cell percentages of row frequencies, and cell percentages of column frequencies, plus total frequency and percent.

```
proc freq data=orion.sales;  
  tables Gender*Country;  
run;
```

rows

columns

The TABLES Statement

The FREQ Procedure

Table of Gender by Country

Gender	Country		
	AU	US	Total
F	27	41	68
	16.36	24.85	41.21
	39.71	60.29	
	42.86	40.20	
M	36	61	97
	21.82	36.97	58.79
	37.11	62.89	
	57.14	59.80	
Total	63	102	165
	38.18	61.82	100.00

Additional SAS Statements

- Additional statements can be added to enhance the report.

```
proc format;  
    value $ctryfmt 'AU'='Australia'  
                  'US'='United States';  
run;  
  
options nodate pageno=1;  
  
ods html file='p112d01.html';  
proc freq data=orion.sales;  
    tables Gender*Country;  
    where Job Title contains 'Rep';  
    format Country $ctryfmt.;  
    title 'Sales Rep Frequency Report';  
run;  
ods html close;
```

Additional SAS Statements

- HTML Output

Sales Rep Frequency Report
The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Gender by Country			
	Gender	Country		Total
		Australia	United States	
F	27	40	67	
	16.98	25.16	42.14	
	40.30	59.70		
	44.26	40.82		
M	34	58	92	
	21.38	36.48	57.86	
	36.96	63.04		
	55.74	59.18		
Total	61	98	159	
	38.36	61.64	100.00	

Options to Suppress Display of Statistics

- Options can be placed in the TABLES statement after a forward slash to suppress the display of the default statistics.

Option	Description
NOCUM	suppresses the display of cumulative frequency and cumulative percentage.
NOPERCENT	suppresses the display of percentage, cumulative percentage, and total percentage.
NOFREQ	suppresses the display of the cell frequency and total frequency.
NOROW	suppresses the display of the row percentage.
NOCOL	suppresses the display of the column percentage.

Additional TABLES Statement Options

- Additional options can be placed in the TABLES statement after a forward slash to control the displayed output.

Option	Description
LIST	displays n -way tables in list format.
CROSSLIST	displays n -way tables in column format.
FORMAT=	formats the frequencies in n -way tables.

LIST and CROSSLIST Options

Gender	Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	Australia	27	16.36	27	16.36
F	United States	41	24.85	68	41.21
M	Australia	36	21.82	104	63.03
M	United States	61	36.97	165	100.00

```
tables Gender*Country / list;
```

Table of Gender by Country

Gender	Country	Frequency	Percent	Row Percent	Column Percent
F	Australia	27	16.36	39.71	42.86
	United States	41	24.85	60.29	40.20
	Total	68	41.21	100.00	
M	Australia	36	21.82	37.11	57.14
	United States	61	36.97	62.89	59.80
	Total	97	58.79	100.00	
Total	Australia	63	38.18		100.00
	United States	102	61.82		100.00
	Total	165	100.00		

```
tables Gender*Country / crosslist;
```

PROC FREQ Statement Options

- Options can also be placed in the PROC FREQ statement.

Option	Description
NLEVELS	displays a table that provides the number of levels for each variable named in the TABLES statement.
PAGE	displays only one table per page.
COMPRESS	begins the display of the next one-way frequency table on the same page as the preceding one-way table if there is enough space to begin the table.

NLEVELS Option

```
proc freq data=orion.sales nlevels;  
  tables Gender Country Employee_ID;  
run;
```

Partial PROC FREQ Output

The FREQ Procedure

Number of Variable Levels

Variable	Levels
Gender	2
Country	2
Employee_ID	165

SAS: the MEANS Procedure

- The *MEANS procedure* provides data summarization tools to compute descriptive statistics for variables across all observations and within groups of observations.

General form of the MEANS procedure:

```
PROC MEANS DATA=SAS-data-set <statistic(s)> <option(s)>;  
  VAR analysis-variable(s);  
  CLASS classification-variable(s);  
RUN;
```

The MEANS Procedure

- By default, the MEANS procedure reports the number of nonmissing observations, the mean, the standard deviation, the minimum value, and the maximum value of all numeric variables.

```
proc means data=orion.sales;  
run;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Employee_ID	165	120713.90	450.0866939	120102.00	121145.00
Salary	165	31160.12	20082.67	22710.00	243190.00
Birth_Date	165	3622.58	5456.29	-5842.00	10490.00
Hire_Date	165	12054.28	4619.94	5114.00	17167.00

The VAR Statement

The *VAR statement* identifies the analysis variables and their order in the results.

```
proc means data=orion.sales;  
  var Salary;  
run;
```

The MEANS Procedure

Analysis Variable : Salary

N	Mean	Std Dev	Minimum	Maximum
165	31160.12	20082.67	22710.00	243190.00

The CLASS Statement

- The *CLASS statement* identifies variables whose values define subgroups for the analysis.

```
proc means data=orion.sales;  
  var Salary;  
  class Gender Country;  
run;
```

The MEANS Procedure

Analysis Variable : Salary

Gender	Country	N Obs	N	Mean	Std Dev	Minimum	Maximum
F	AU	27	27	27702.41	1728.23	25185.00	30890.00
	US	41	41	29460.98	8847.03	25390.00	83505.00
M	AU	36	36	32001.39	16592.45	25745.00	108255.00
	US	61	61	33336.15	29592.69	22710.00	243190.00

The CLASS Statement

```
proc means data=orion.sales;  
  var Salary;  
  class Gender Country;  
run;
```

classification variables

The MEANS Procedure

Analysis Variable : Salary

analysis variable

Gender	Country	N Obs	N	Mean	Std Dev	Minimum	Maximum
F	AU	27	27	27702.41	1728.23	25185.00	30890.00
	US	41	41	29460.98	8847.02	25200.00	82505.00
M	AU	36	36	32001.33	10000.00	22710.00	42000.00
	US	61	61	33336.15	29592.69	22710.00	243190.00

statistics for analysis variable

The CLASS statement adds the N Obs column, which is the number of observations for each unique combination of the class variables.

PROC MEANS Statistics

- The statistics to compute and the order to display them can be specified in the PROC MEANS statement.

```
proc means data=orion.sales sum mean range;  
  var Salary;  
  class Country;  
run;
```

The MEANS Procedure

Analysis Variable : Salary

Country	N Obs	Sum	Mean	Range
AU	63	1900015.00	30158.97	83070.00
US	102	3241405.00	31778.48	220480.00

PROC MEANS Statistics

Descriptive Statistic Keywords				
CLM	CSS	CV	LCLM	MAX
MEAN	MIN	MODE	N	NMISS
KURTOSIS	RANGE	SKEWNESS	STDDEV	STDERR
SUM	SUMWGT	UCLM	USS	VAR
Quantile Statistic Keywords				
MEDIAN P50	P1	P5	P10	Q1 P25
Q3 P75	P90	P95	P99	QRANGE
Hypothesis Testing Keywords				
PROBT	T			

SAS: the SUMMARY Procedure

- The SUMMARY procedure provides data summarization tools to compute descriptive statistics for variables across all observations and within groups of observations.

General form of the SUMMARY procedure:

```
PROC SUMMARY DATA=SAS-data-set <statistic(s)>  
                                     <option(s)>;  
  
    VAR analysis-variable(s);  
    CLASS classification-variable(s);  
RUN;
```

The SUMMARY Procedure

- The SUMMARY procedure uses the same syntax as the MEANS procedure.
- The only differences to the two procedures are the following:

PROC MEANS	PROC SUMMARY
The PRINT option is set by default, which displays output.	The NOPRINT option is set by default, which displays no output.
Omitting the VAR statement analyzes all the numeric variables.	Omitting the VAR statement produces a simple count of observations.

SAS: the TABULATE Procedure

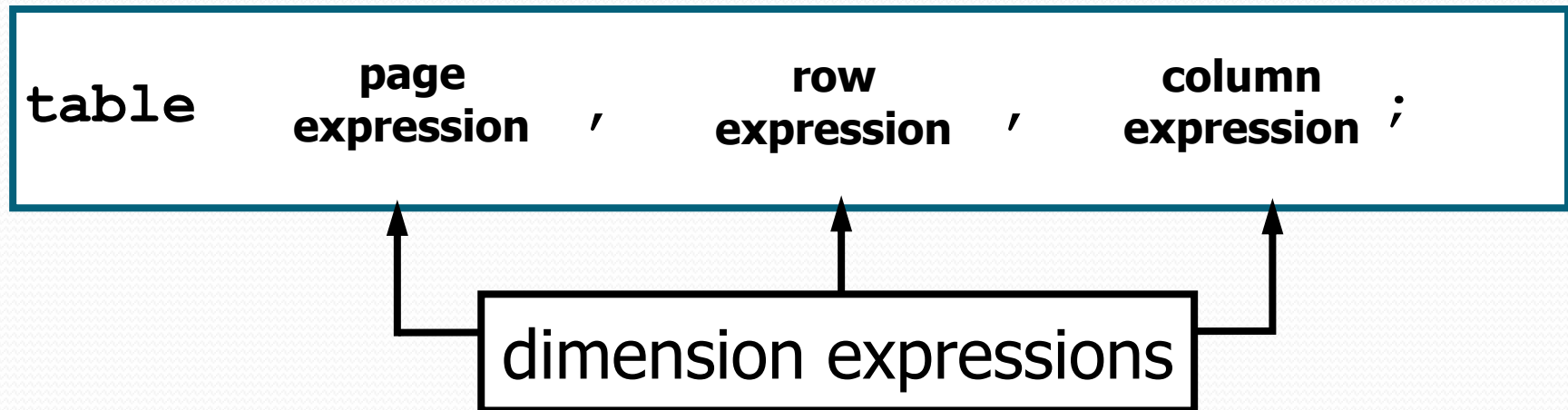
- The TABULATE procedure displays descriptive statistics in tabular format.

General form of the TABULATE procedure:

```
PROC TABULATE DATA=SAS-data-set <options>;  
  CLASS classification-variable(s);  
  VAR analysis-variable(s);  
  TABLE page-expression,  
         row-expression,  
         column-expression </option(s)>;  
RUN;
```

The TABLE Statement

- The TABLE statement describes the structure of the table.



- Commas separate the dimension expressions.
- Every variable that is part of a dimension expression must be specified as a classification variable (CLASS statement) or an analysis variable (VAR statement).

The TABLE Statement

```
table    page  
        expression ,    row  
        expression ,    column  
        expression ;
```

- Examples:

```
table Country ;
```

```
table Gender , Country ;
```

```
table Job_Title , Gender , Country ;
```

The CLASS Statement

- The CLASS statement identifies variables to be used as classification, or grouping, variables.
- General form of the CLASS statement:

```
CLASS classification-variable(s);
```

- N, the number of nonmissing values, is the default statistic for classification variables.
- Examples of classification variables:

Job_Title, **Gender**, and **Country**

The VAR Statement

- The VAR statement identifies the numeric variables for which statistics are calculated.
- General form of the VAR statement:

```
VAR analysis-variable(s);
```

- SUM is the default statistic for analysis variables.
- Examples of analysis variables:

Salary and **Bonus**

One-Dimensional Table

```
proc tabulate data=orion.sales;  
  class Country;  
  table Country;  
run;
```

Country	
AU	US
N	N
63.00	102.00

Two-Dimensional Table

```
proc tabulate data=orion.sales;  
  class Gender Country;  
  table Gender, Country;  
run;
```

	Country	
	AU	US
	N	N
Gender		
F	27.00	41.00
M	36.00	61.00

Three-Dimensional Table

```
proc tabulate data=orion.sales;  
  class Job Title Gender Country;  
  table Job_Title, Gender, Country;  
run;
```

Job_Title Sales Rep. I

	Country	
	AU	US

Gender

F

M

Job_Title Sales Rep. II

	Country	
	AU	US
Gender		
F	10.00	14.00
M	8.00	14.00

p112d08

Dimension Expression

- Elements that can be used in a dimension expression:
 - classification variables
 - analysis variables
 - the universal class variable ALL
 - keywords for statistics
- Operators that can be used in a dimension expression:
 - blank, which concatenates table information
 - asterisk *, which crosses table information
 - parentheses (), which group elements

Dimension Expression

```
proc tabulate data=orion.sales;  
  class Gender Country;  
  var Salary;  
  table Gender all, Country*Salary;  
run;
```

	Country	
	AU	US
	Salary	Salary
	Sum	Sum
Gender		
F	747965.00	1207900.00
M	1152050.00	2033505.00
All	1900015.00	3241405.00

PROC TABULATE Statistics

Descriptive Statistic Keywords

	CSS	CV	LCLM	MAX
MEAN	MIN	MODE	N	NMISS
KURTOSIS	RANGE	SKEWNESS	STDDEV	STDERR
SUM	SUMWGT	UCLM	USS	VAR
PCTN	REPPCTN	PAGEPCTN	ROWPCTN	COLPCTN
PCTSUM	REPPCTSUM	PAGEPCTSUM	ROWPCTSUM	COLPCTSUM

Quantile Statistic Keywords

MEDIAN P50	P1	P5	P10	Q1 P25
Q3 P75	P90	P95	P99	QRANGE

Hypothesis Testing Keywords

PROBT	T			
-------	---	--	--	--

PROC TABULATE Statistics

```
proc tabulate data=orion.sales;  
  class Gender Country;  
  var Salary;  
  table Gender all, Country*Salary* (min max);  
run;
```

	Country			
	AU		US	
	Salary		Salary	
	Min	Max	Min	Max
Gender				
F	25185.00	30890.00	25390.00	83505.00
M	25745.00	108255.00	22710.00	243190.00
All	25185.00	108255.00	22710.00	243190.00

Additional SAS Statements

- Additional statements can be added to enhance the report.

```
proc format;  
  value $ctryfmt 'AU'='Australia'  
                'US'='United States';  
run;  
  
options nodate pageno=1;  
  
ods html file='p112d08.html';  
proc tabulate data=orion.sales;  
  class Gender Country;  
  var Salary;  
  table Gender all, Country*Salary*(min max);  
  where Job_Title contains 'Rep';  
  label Salary='Annual Salary';  
  format Country $ctryfmt.;  
  title 'Sales Rep Tabular Report';  
run;  
ods html close;
```

p112d08

Additional SAS Statements

- HTML Output

Sales Rep Tabular Report

	Country			
	Australia		United States	
	Annual Salary		Annual Salary	
	Min	Max	Min	Max
Gender				
F	25185.00	30890.00	25390.00	32985.00
M	25745.00	36605.00	22710.00	35990.00
All	25185.00	36605.00	22710.00	35990.00

Vice o PROC TABULATE:

- In the SUGI 28 proceedings:
 - “*The Simplicity and Power of the TABULATE Procedure*”,
by Dan Bruns
<http://www2.sas.com/proceedings/sugi28/197-28.pdf>
- Online (from the SUGI 27 proceedings):
 - “*Anyone Can Learn PROC TABULATE*”,
by Lauren Haworth,
<http://www2.sas.com/proceedings/sugi27/po60-27.pdf>

SAS: the UNIVARIATE Procedure

- The UNIVARIATE procedure produces summary reports that display descriptive statistics.
- General form of the UNIVARIATE procedure:

```
PROC UNIVARIATE DATA=SAS-data-set;  
    VAR variable(s);  
RUN;
```

- The VAR statement specifies the analysis variables and their order in the results.

The UNIVARIATE Procedure

The following PROC UNIVARIATE step shows default descriptive statistics for **Salary**.

```
proc univariate data=orion.nonsales;  
    var Salary;  
run;
```

- Without the VAR statement, SAS will analyze all numeric variables.

The UNIVARIATE Procedure

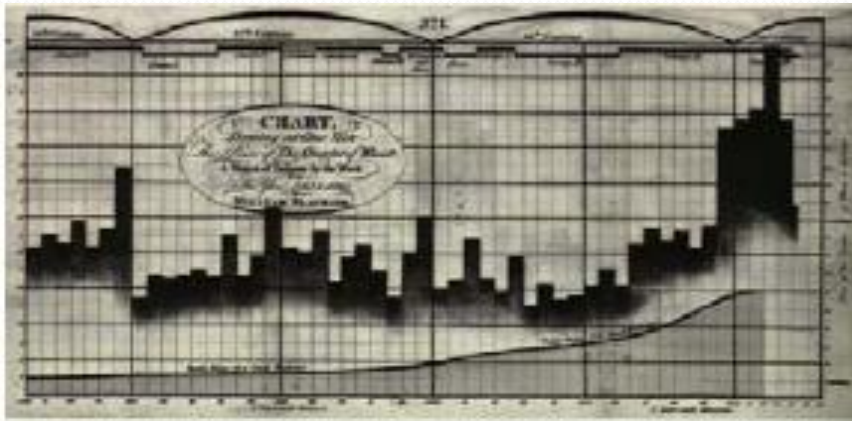
- The UNIVARIATE procedure can produce the following sections of output:
 - Moments
 - Basic Statistical Measures
 - Tests for Locations
 - Quantiles
 - Extreme Observations
 - Missing Values

Explorace - Vizualizace

- Na prvním místě se obvykle citují knihy prof. Tufteho, např. Tufte E.R. (1983) The Visual Display of Quantitative Information, Graphic Press, Chesire, Conn.
- Weby o vizualizaci, např.
 - <http://www.math.yorku.ca/SCS/Gallery/noframes.html> - galerie s poučným výkladem a příklady i nezdařených či lživých grafů
 - <http://www.agocg.ac.uk/> - John Lansdown (1992) Aspects of Design in Computer Graphics: Some Notes –
<http://www.agocg.ac.uk/train/hitch/hitch.htm>
- Jiné weby, např. stránky různých vizualizačních programů a organizací
 - <http://www.cybergeography.org/atlas/atlas.html> nebo <http://miner3d.com/products/gallery.html>

Vizualizace – historie

- William Playfair, 1786: první publikovaná prezentační grafika



- Dr. John Snow, 1845: epidemie cholery v Londýně



Vizualizace – historie

- Florence Nightingale, 1858: důvody úmrtí v průběhu Krymské války (1853-1856)



- Harry Beck, 1931: schéma Londýnského metra



Vizualizace – investigativní analýza



Law Enforcement

- » Counterterrorism
- » Narcotics investigations
- » Organized crime
- » Intelligence analysis
- » Fraud
- » Missing persons
- » Major investigations
- » Counterfeiting
- » Immigration control
- » Major event security
- » Money laundering
- » Gang investigations

Government

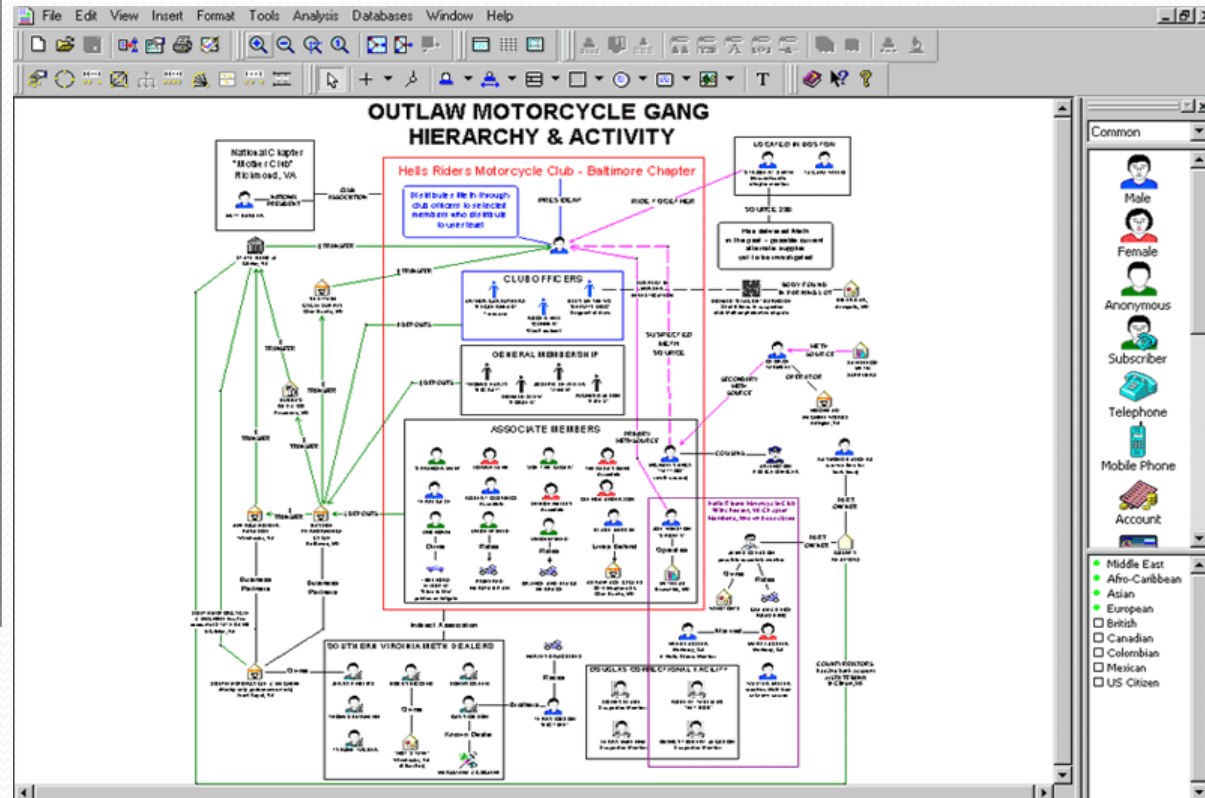
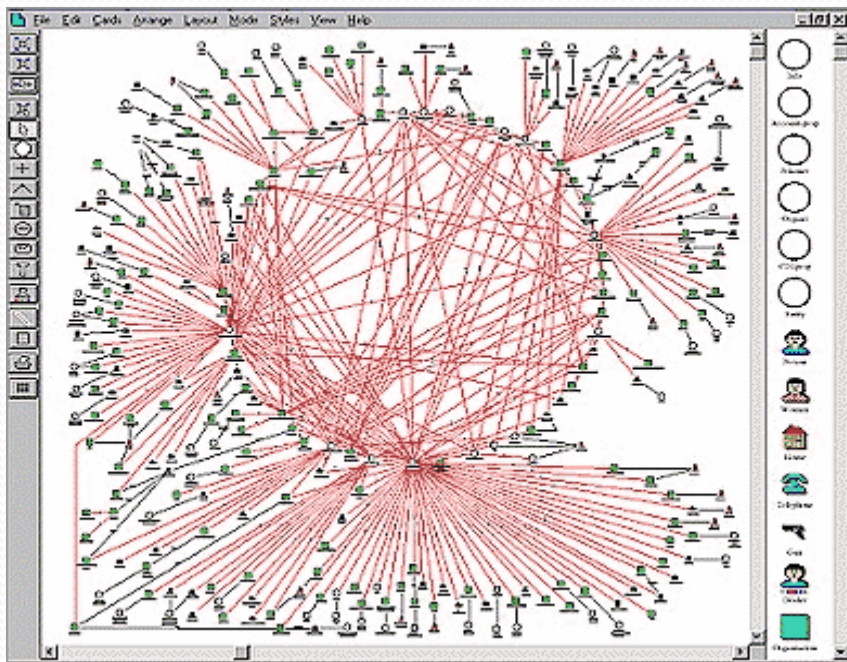
- » Criminal prosecutions
- » National security
- » Military intelligence
- » Embassy security
- » Postal inspection and fraud
- » Prison investigations
- » Park and wildlife services
- » Antitrust investigations
- » Tax fraud investigations
- » Customs investigations

Commercial

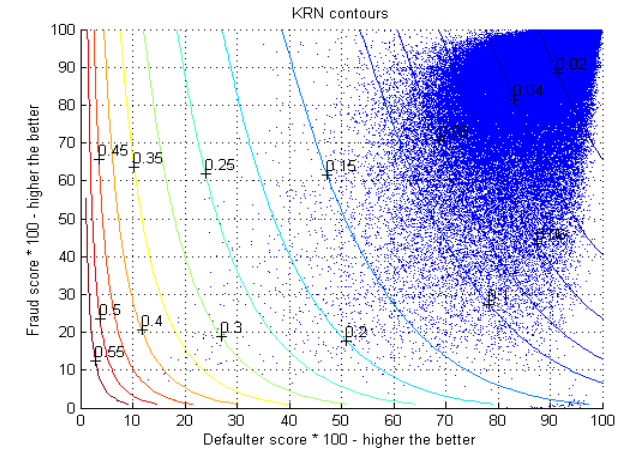
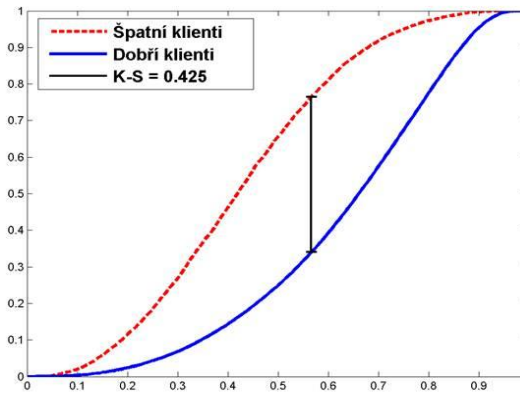
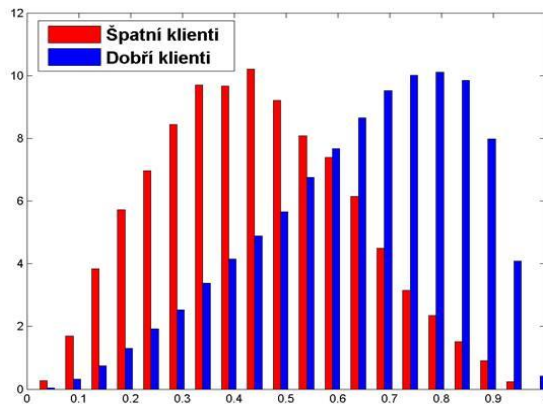
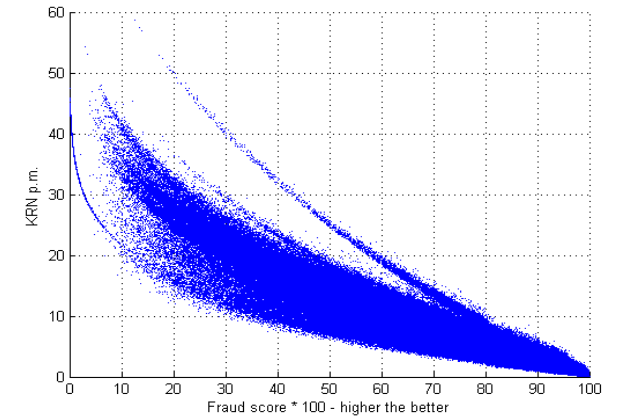
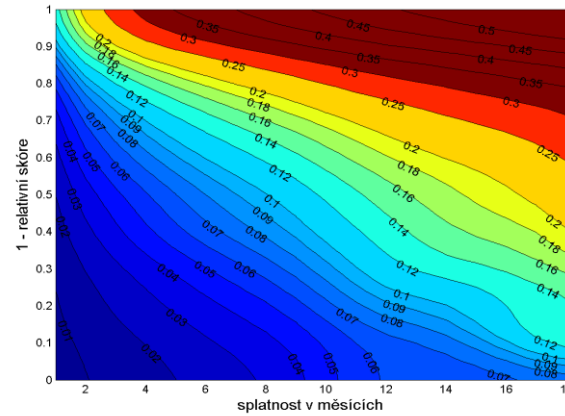
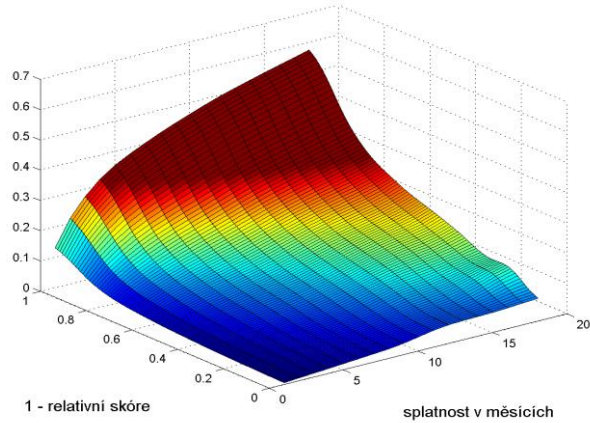
- » Forensic accounting
- » Money laundering
- » Insider trading violations
- » Corporate security
- » Anti-pirating investigations
- » Entertainment copyright violations
- » Competitive intelligence
- » Civil lawsuits
- » Fraud:
 - » Credit card
 - » Insurance
 - » Retail
 - » Health care
 - » Commercial
 - » Telephone

Vizualizace – investigativní analýza

- ❑ Praní špinavých peněz, kriminální gangy



Vizualizace – risk management



Vizualizace - dendrogram

Credit ranking (1=default)

Node 0		
Category	%	n
Bad	52,01	168
Good	47,99	155
Total	(100,00)	323

Paid Weekly/Monthly
Adj. P-value=0,0000, Chi-square=179,6665, df=1

Weekly pay

Node 1		
Category	%	n
Bad	86,67	143
Good	13,33	22
Total	(51,08)	165

Monthly salary

Node 2		
Category	%	n
Bad	15,82	25
Good	84,18	133
Total	(48,92)	158

Social Class
Adj. P-value=0,0004, Chi-square=20,3674, df=2

Age Categorical
Adj. P-value=0,0000, Chi-square=58,7255, df=1

Management;Professional

Clerical;Skilled Manual

Unskilled

Young (< 25)

Middle (25-35);Old (> 35)

Node 3		
Category	%	n
Bad	71,11	32
Good	28,89	13
Total	(13,93)	45

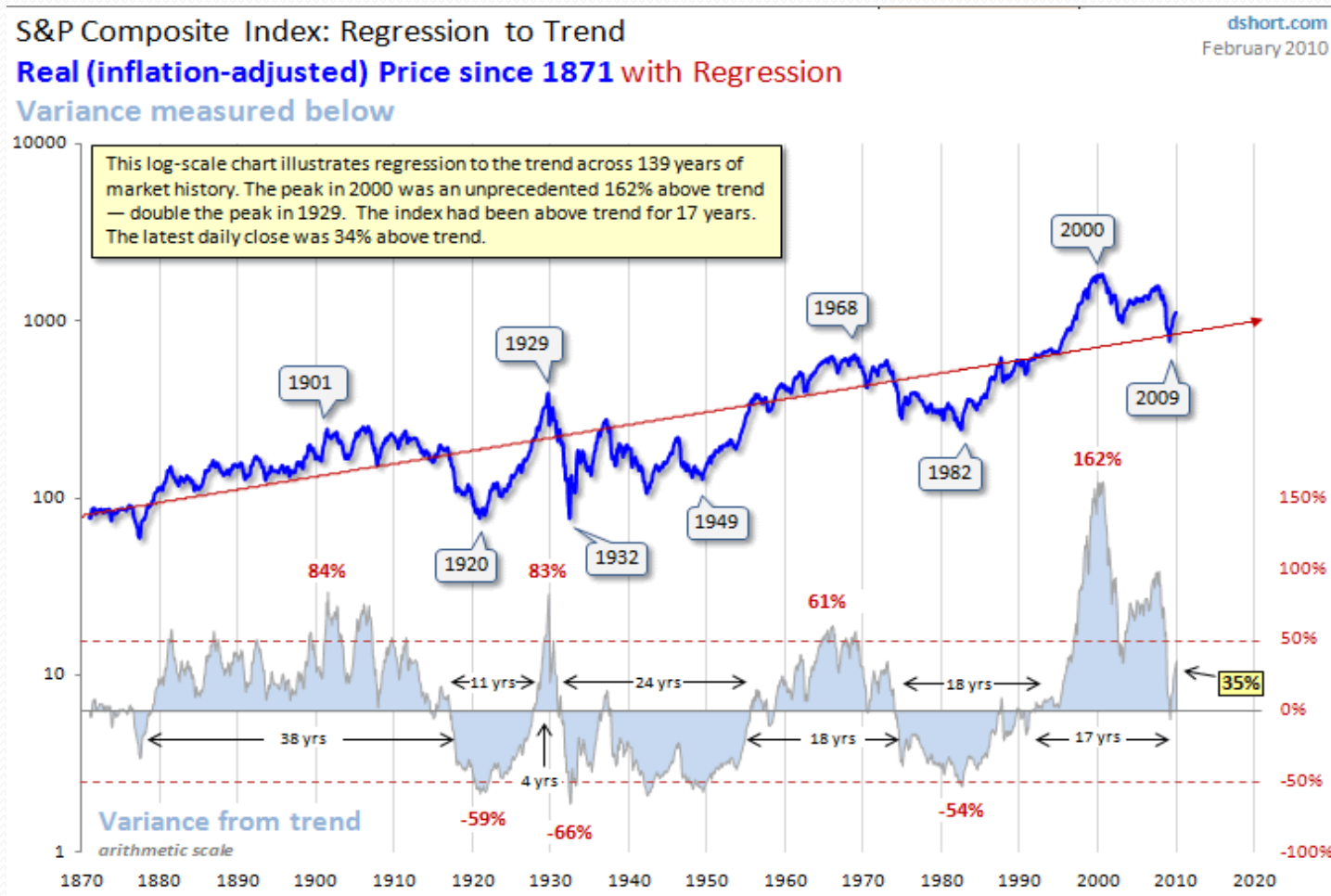
Node 4		
Category	%	n
Bad	97,56	80
Good	2,44	2
Total	(25,39)	82

Node 5		
Category	%	n
Bad	81,58	31
Good	18,42	7
Total	(11,76)	38

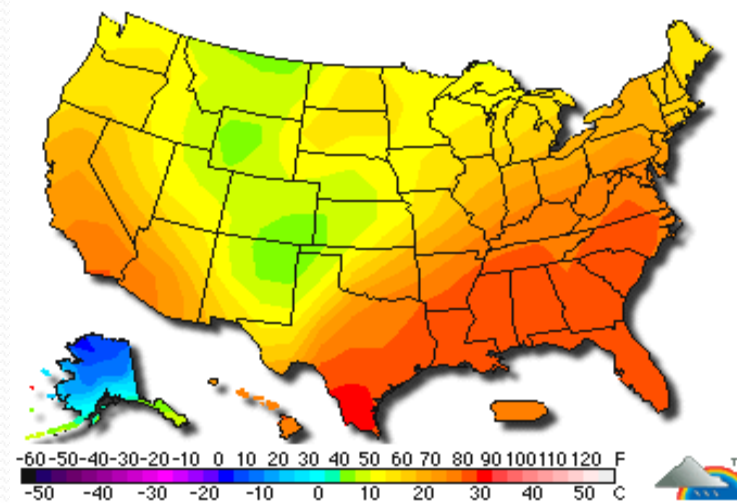
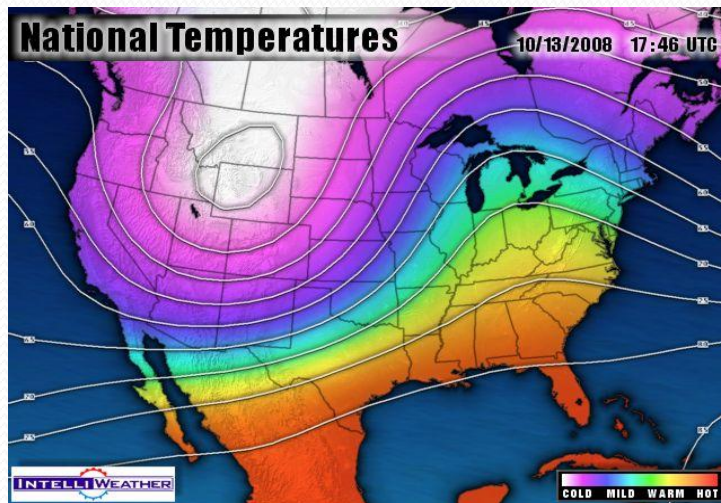
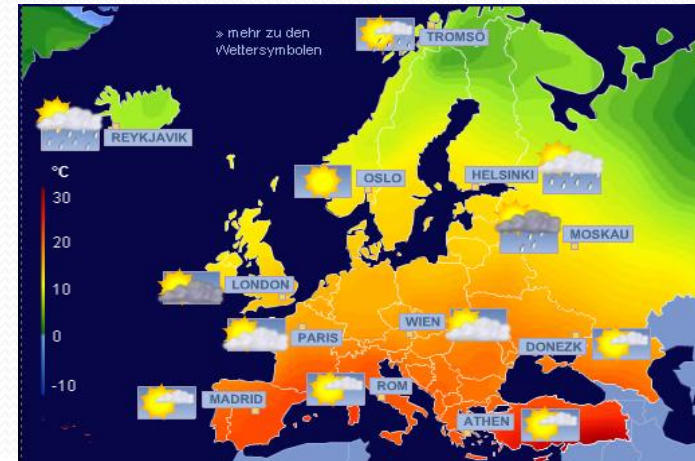
Node 6		
Category	%	n
Bad	48,98	24
Good	51,02	25
Total	(15,17)	49

Node 7		
Category	%	n
Bad	0,92	1
Good	99,08	108
Total	(33,75)	109

Vizualizace – ekonomie

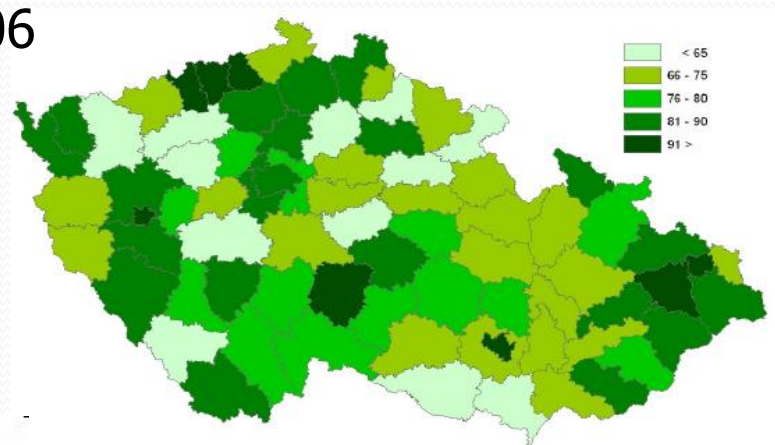


Meteo-vizualizace



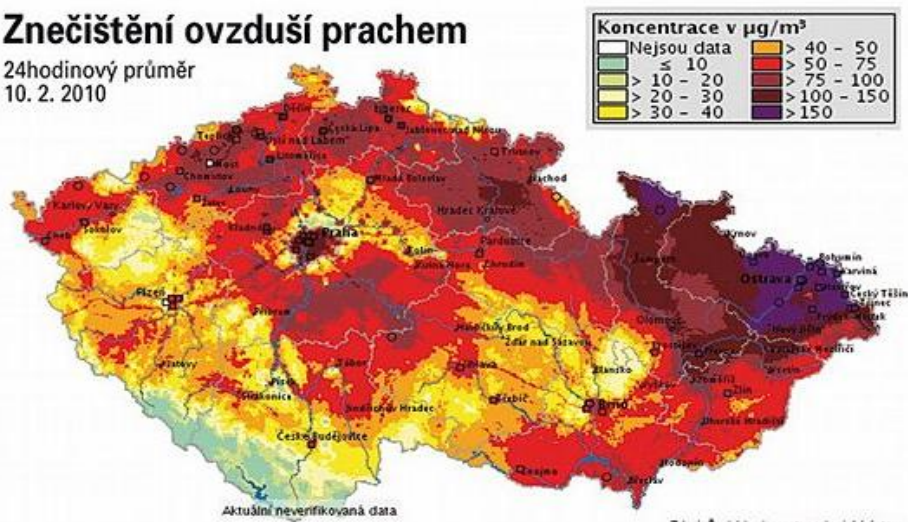
Kartogram

☐ Obce s počtem 500 a více obyvatel s vysokorychlostním připojením k internetu, podle okresů (%), k 31.12.2006



Znečištění ovzduší prachem

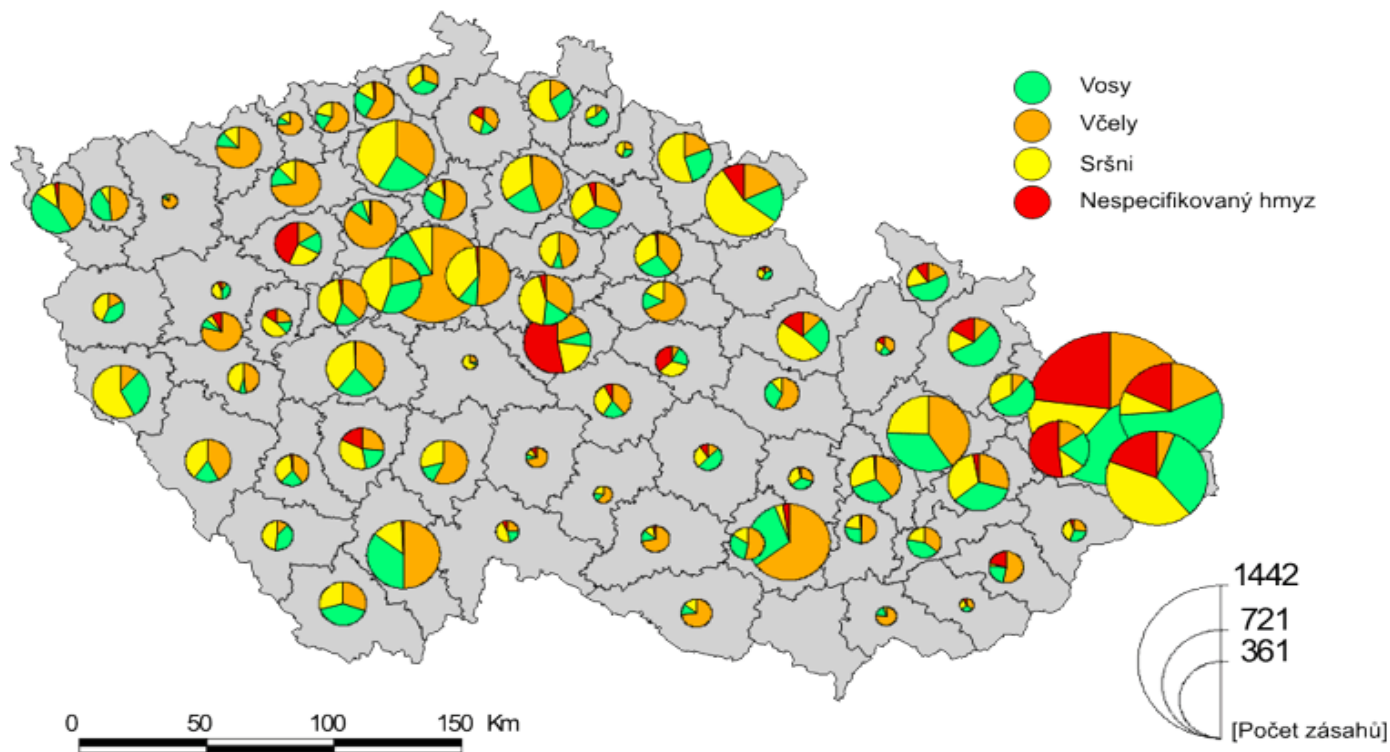
24hodinový průměr
10. 2. 2010



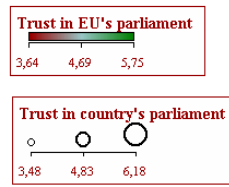
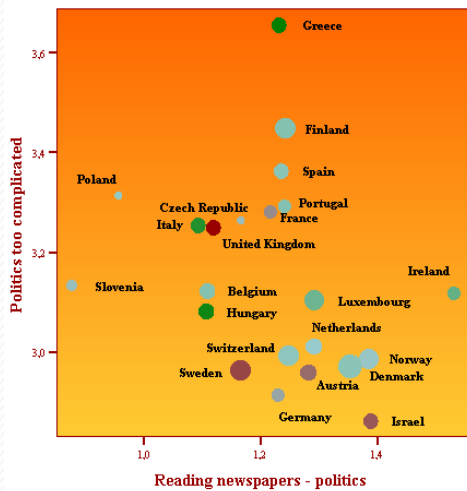
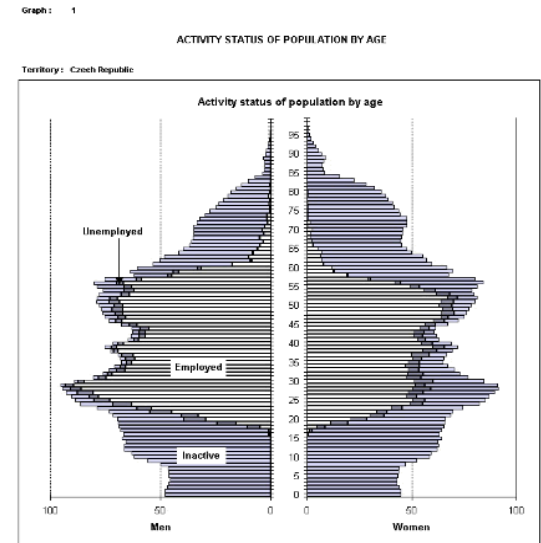
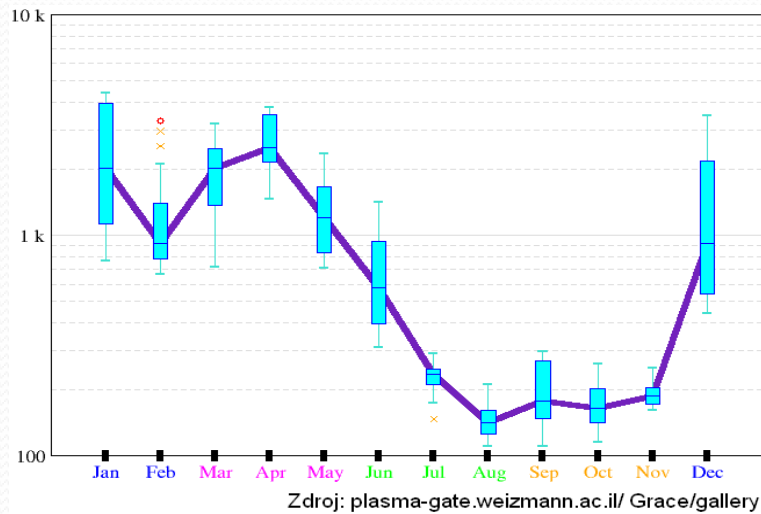
Zdroj: Český hydrometeorologický ústav

Kartodiagram

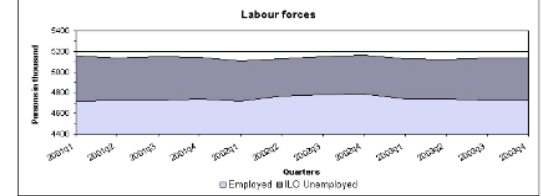
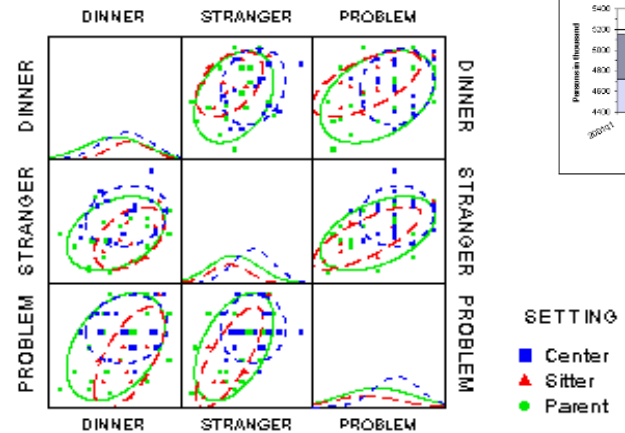
ZÁSAHY JEDNOTEK PO PROTI HMYZU v okresech České republiky v letech 1997-2000



Grafy –další typy

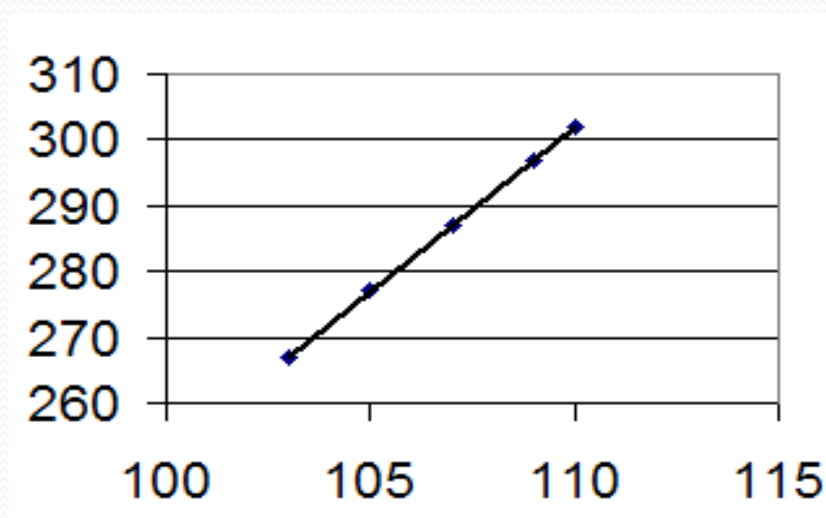
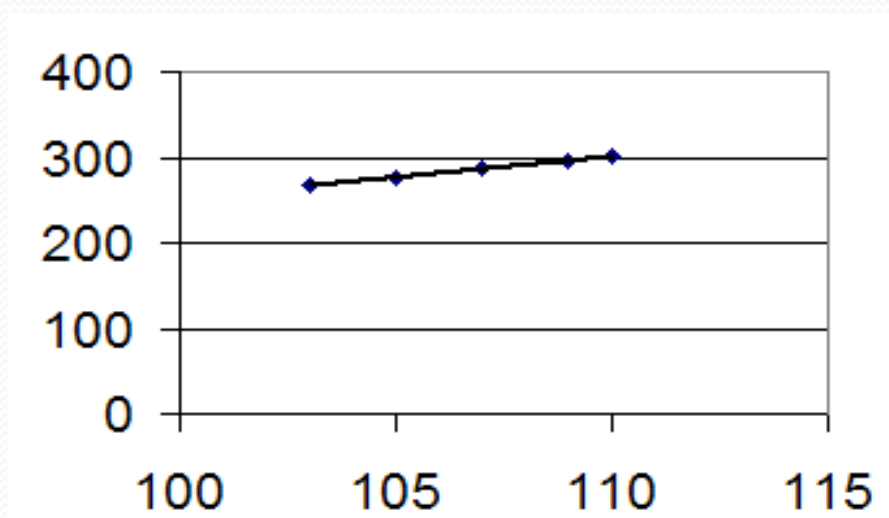


Social Competence Measures Across Setting



Měřítko grafu

□ Která přímka roste strměji?



x	y
103	567
105	577
107	587
109	597
110	602

Měřítko grafu

- ❑ Pohled tvůrce grafu:
 - Zvýraznění trendu – pozitivní výsledky.
 - Potlačení trendu – negativní výsledky.

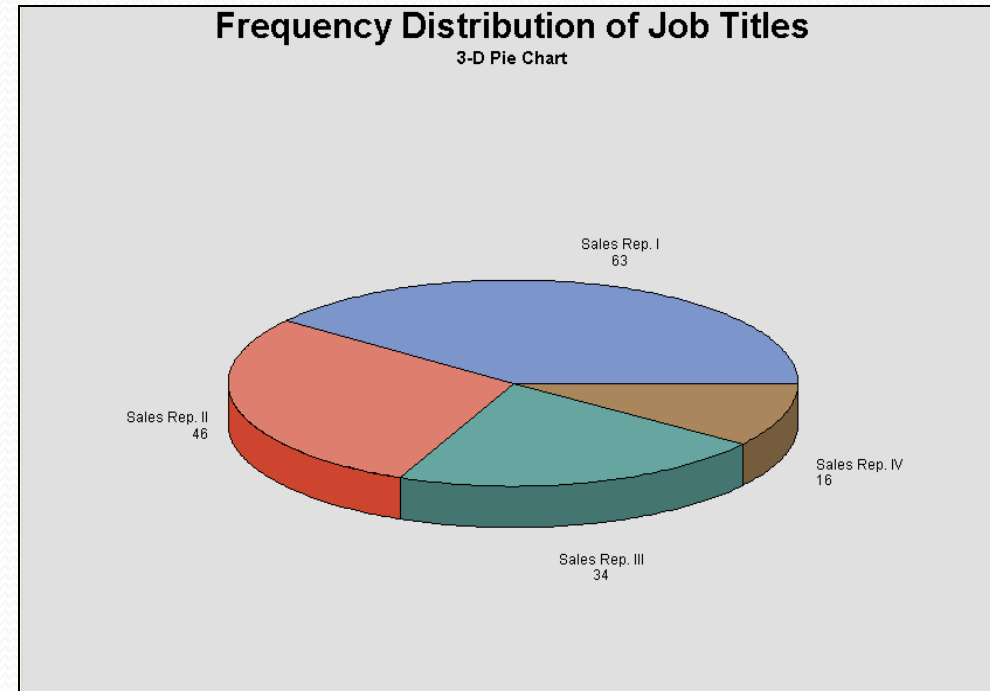
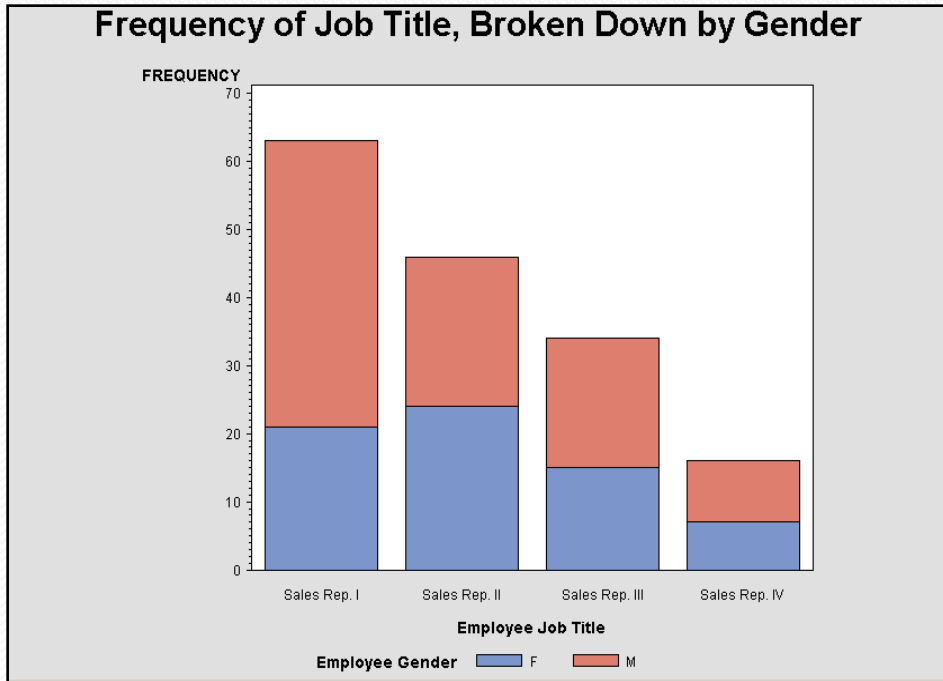
- ❑ Pohled uživatele grafu:
 - Grafy bez uvedeného měřítka jsou silně podezřelé.
 - Nepodléhat podsouvané informaci o růstu/poklesu.

What Is SAS/GRAPH Software?

- *SAS/GRAPH software* is a component of SAS software that enables you to create the following types of graphs:
 - bar, block, and pie charts
 - two-dimensional scatter plots and line plots
 - three-dimensional scatter and surface plots
 - contour plots
 - maps
 - text slides
 - custom graphs

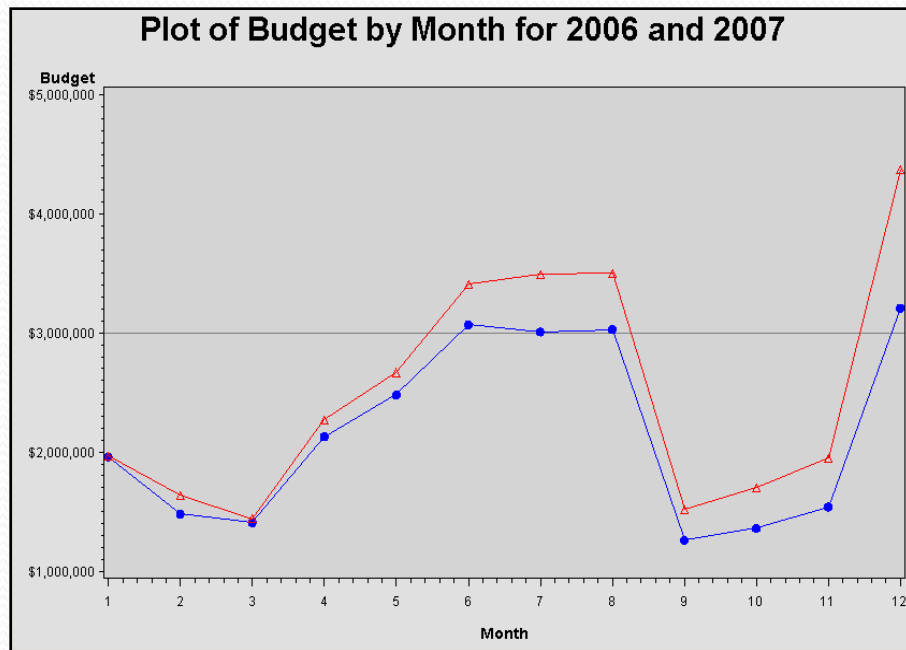
Základní typy grafů

- Bar Charts (GCHART Procedure)
- Pie Charts (GCHART Procedure)

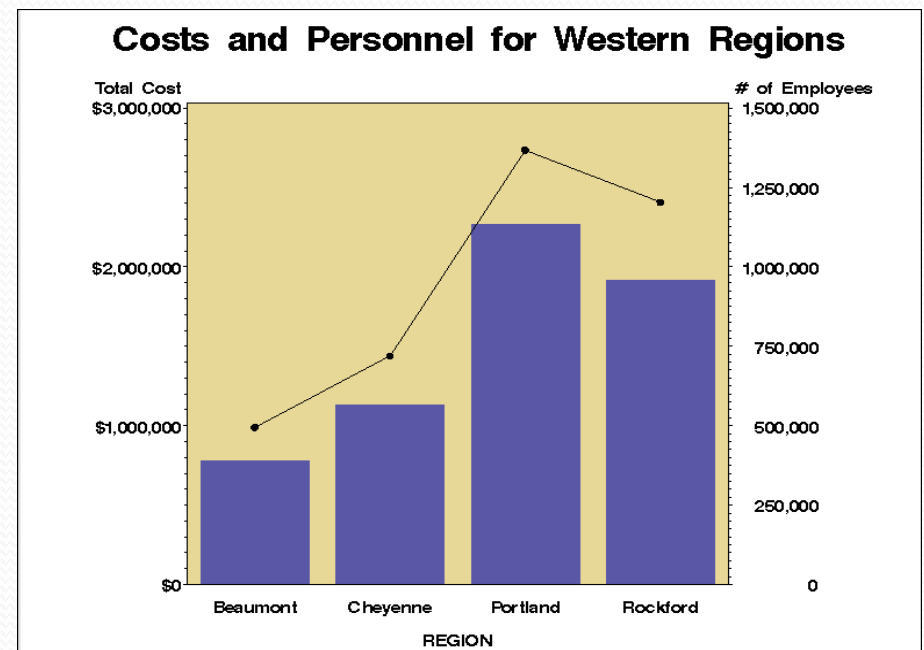


Základní typy grafů

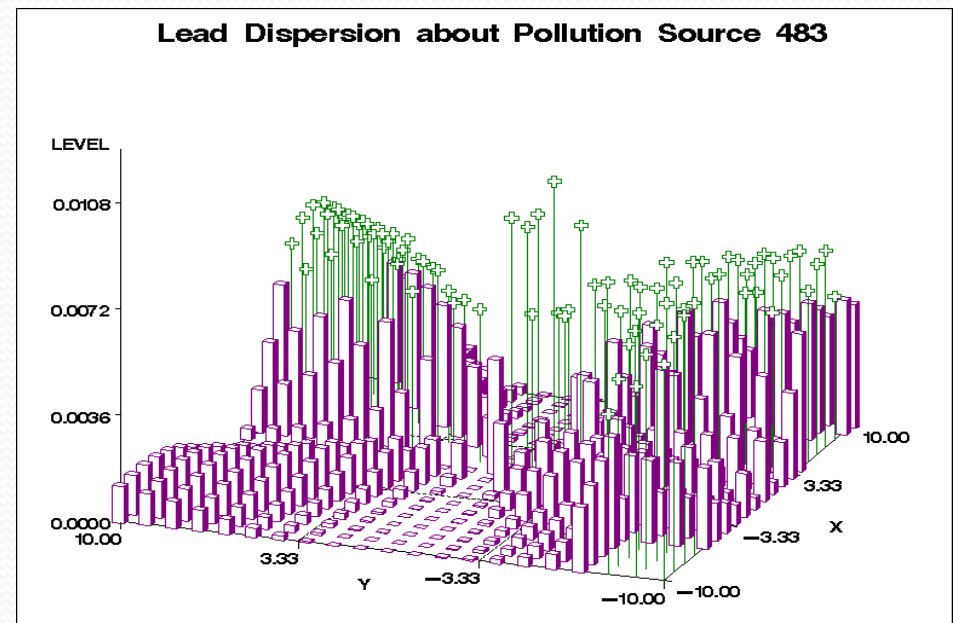
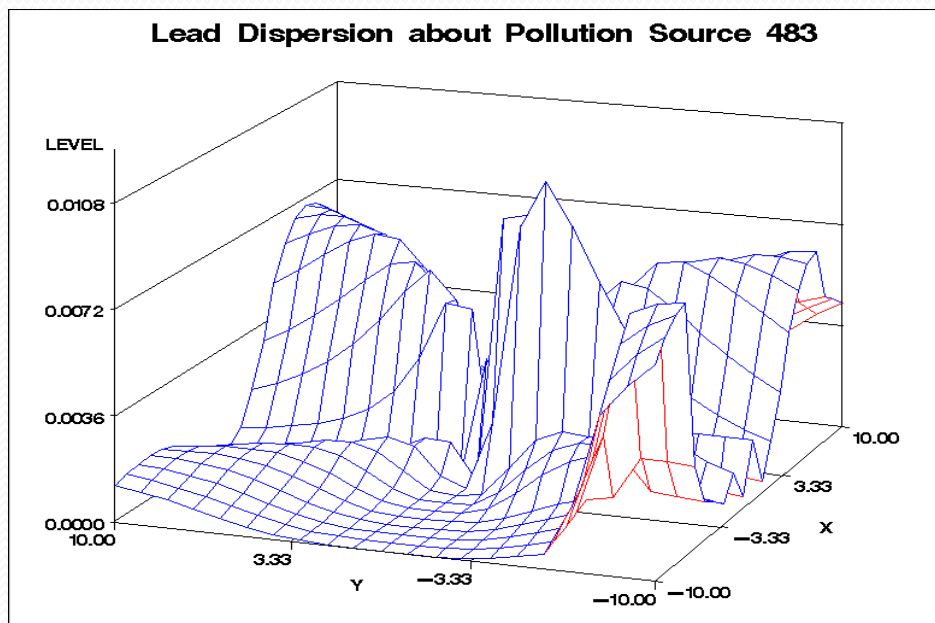
- Scatter and Line Plots (GPLOT Procedure)



- Bar Charts with Line Plot Overlay (GBARLINE Procedure)

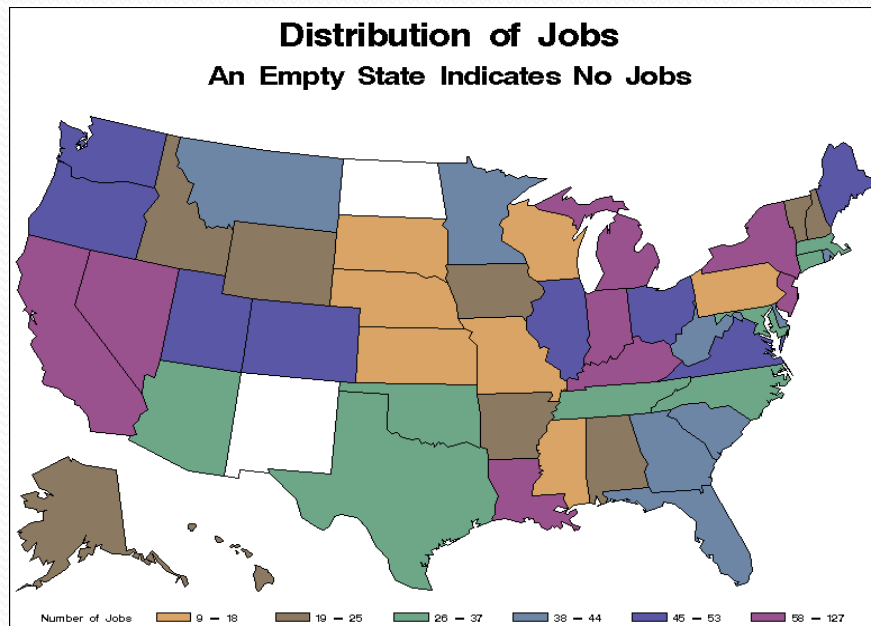


Three-Dimensional Surface and Scatter Plots (G3D Procedure)

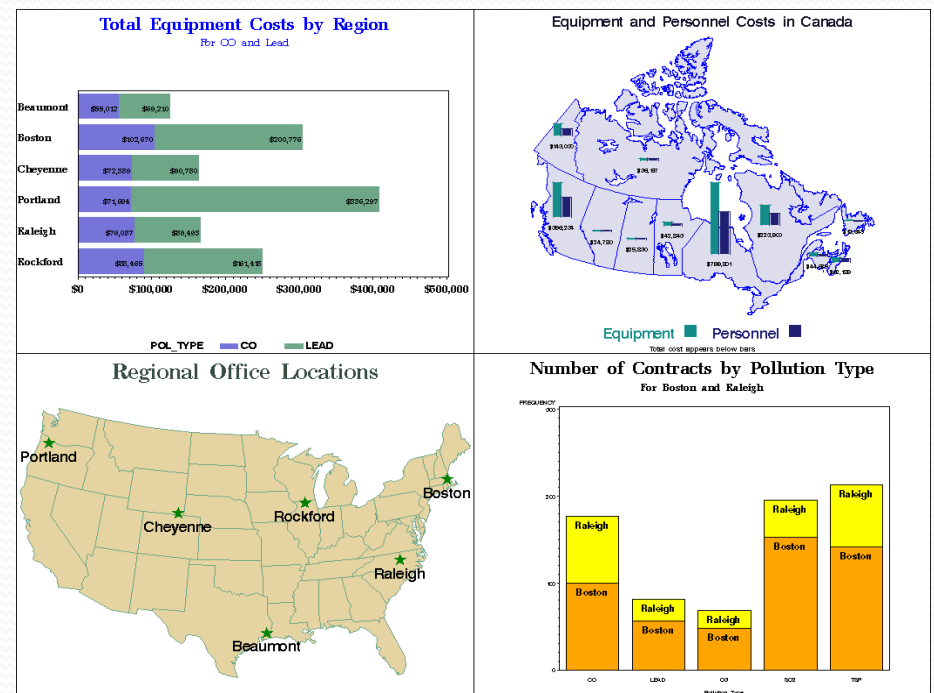


Maps (GMAP Procedure)

- Maps (GMAP Procedure)



- Multiple graphs on a page (GREPLAY Procedure)



Producing Bar and Pie Charts with the GCHART Procedure

- General form of the PROC GCHART statement:

```
PROC GCHART DATA=SAS-data-set;
```

- Use one of these statements to specify the chart type:

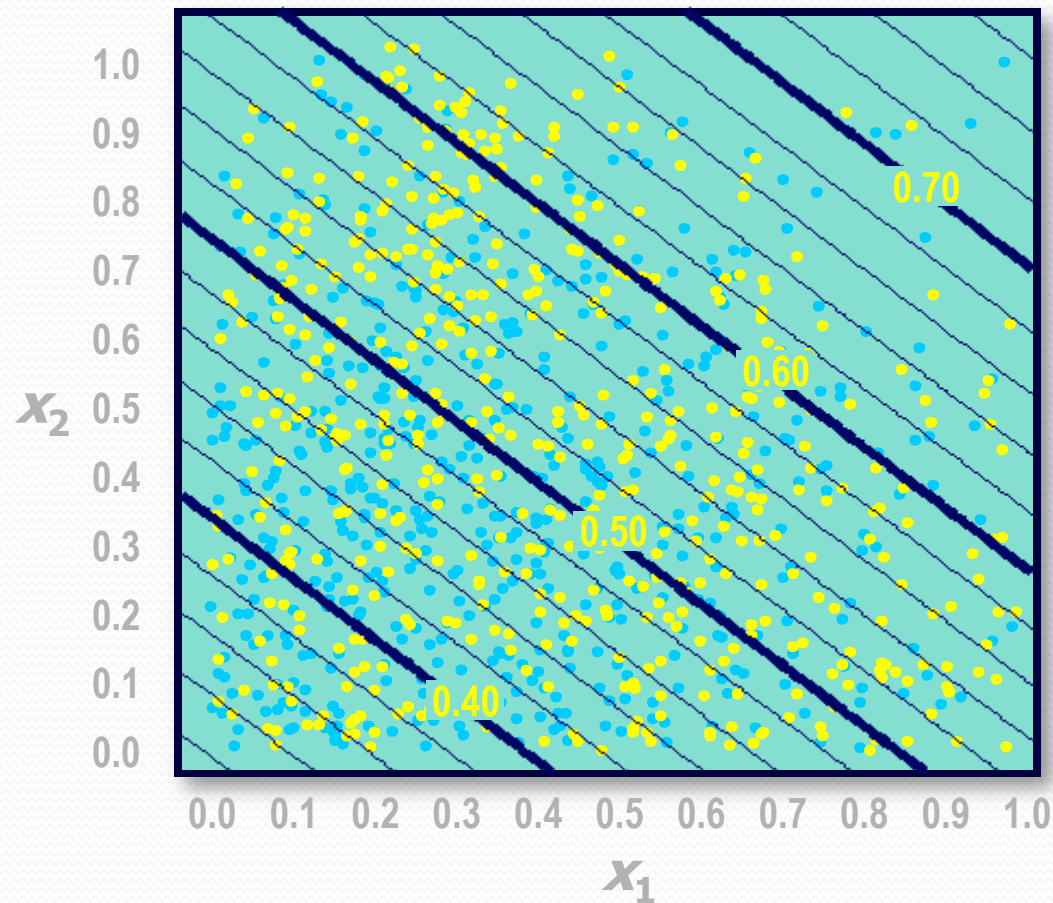
```
HBAR chart-variable . . . </ options>;  
HBAR3D chart-variable . . . </ options>;  
  
VBAR chart-variable . . . </ options>;  
VBAR3D chart-variable . . . </ options>;  
  
PIE chart-variable . . . </ options>;  
PIE3D chart-variable . . . </ options>;
```

Producing Plots with the GPLOT Procedure

- You can use the GPLOT procedure to plot one variable against another within a set of coordinate axes.
- General form of a PROC GPLOT step:

```
PROC GPLOT DATA=SAS-data-set;  
    PLOT vertical-variable * horizontal-variable </ options>;  
RUN;  
QUIT;
```

2. Aplikace logistických regresních modelů



Logistická regrese

Simultánní efekt nezávislých (explanačních) proměnných na šanci

$$\text{Odds} = P/1-P = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}$$

Jestliže logaritmujeme obě strany

$$\text{Log}\{P/1-P\} = \log e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}$$

$$\text{Logit } P = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Logit Transformation

- Logistic regression models transform probabilities called logits*.

$$\text{logit}(p_i) = \ln \left(\frac{p_i}{(1 - p_i)} \right)$$

where

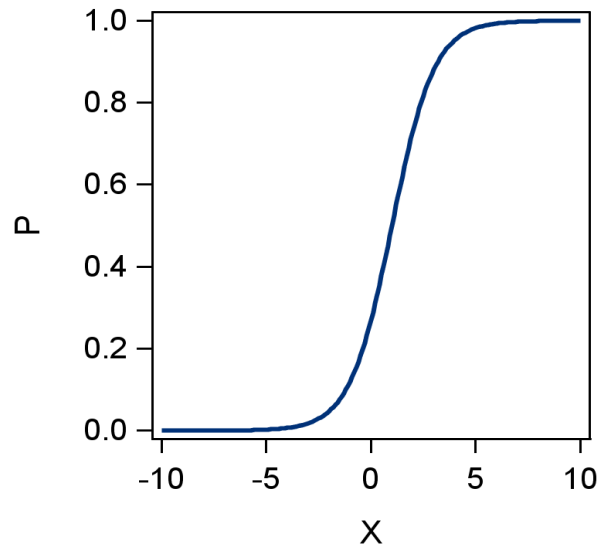
i indexes all cases (observations)

p_i is the probability the event (a default, for example) occurs in the i^{th} case

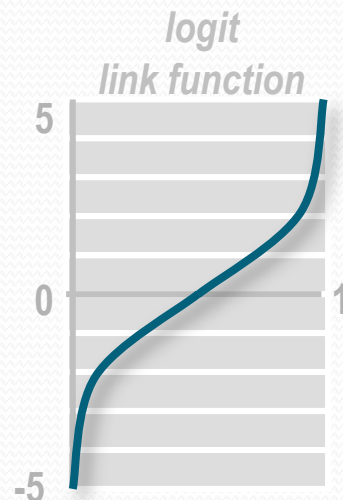
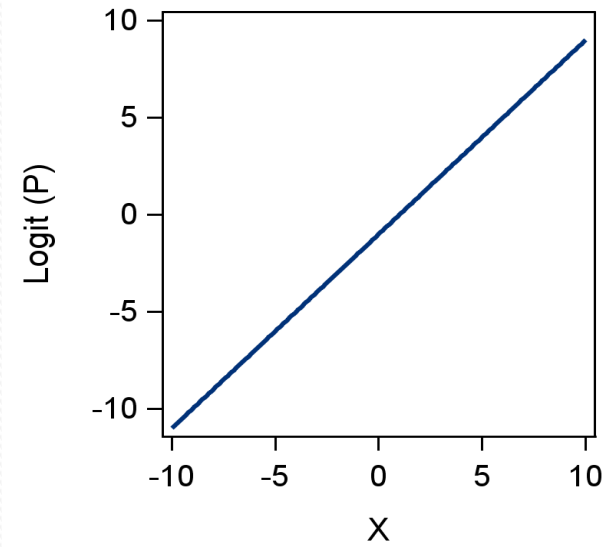
\ln is the natural log (to the base e).

* The logit is the natural log of the odds.

Logit link function



Logit Transform



The logit link function transforms probabilities (between 0 and 1) to logit scores (between $-\infty$ and $+\infty$).

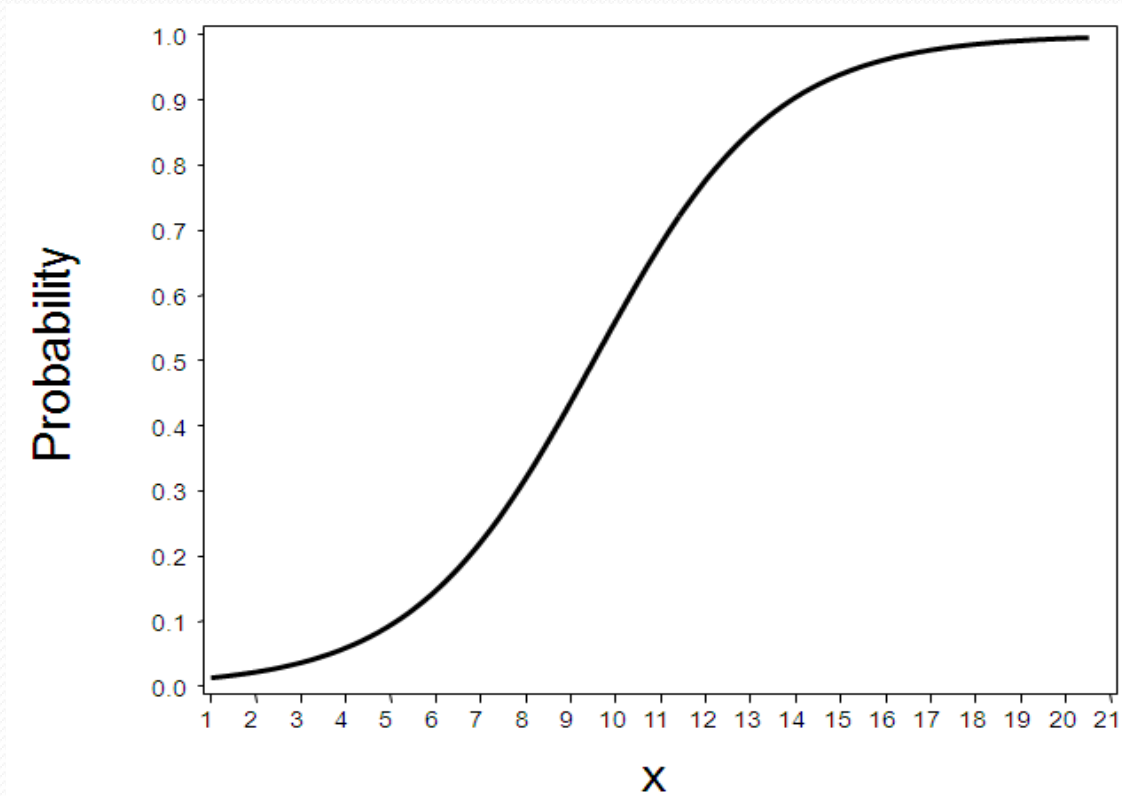
Logistic Regression Model

$$\text{logit}(p_i) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

where

- $\text{logit}(p_i)$ = logit of the probability of the event
- β_0 = intercept of the regression equation
- β_k = parameter estimate of the k^{th} predictor variable

Logistic Regression Curve



$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}}$$

Logistic Regressions -example

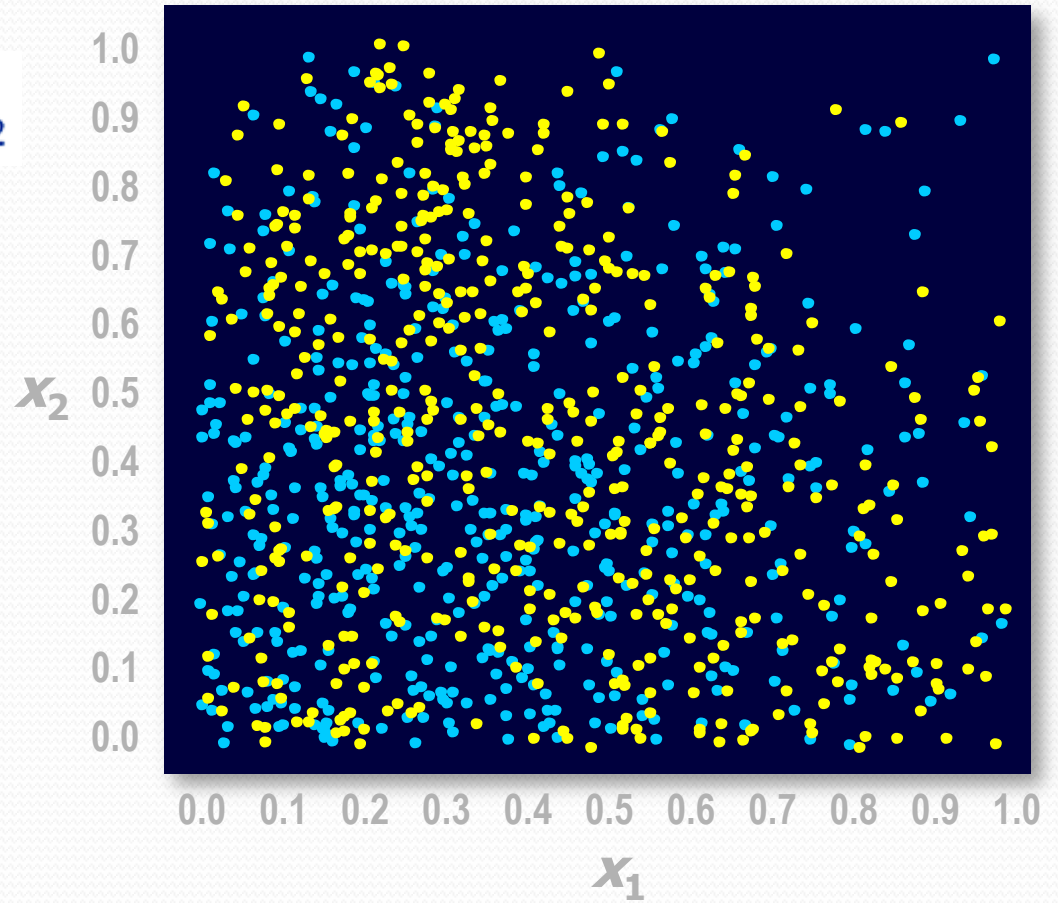
$$\text{logit}(\hat{p}) = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2$$

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

Find parameter estimates by
maximizing

$$\sum_{\substack{\text{primary} \\ \text{outcome} \\ \text{training cases}}} \log(\hat{p}_i) + \sum_{\substack{\text{secondary} \\ \text{outcome} \\ \text{training cases}}} \log(1 - \hat{p}_i)$$

log-likelihood function



Logistic Regressions -example

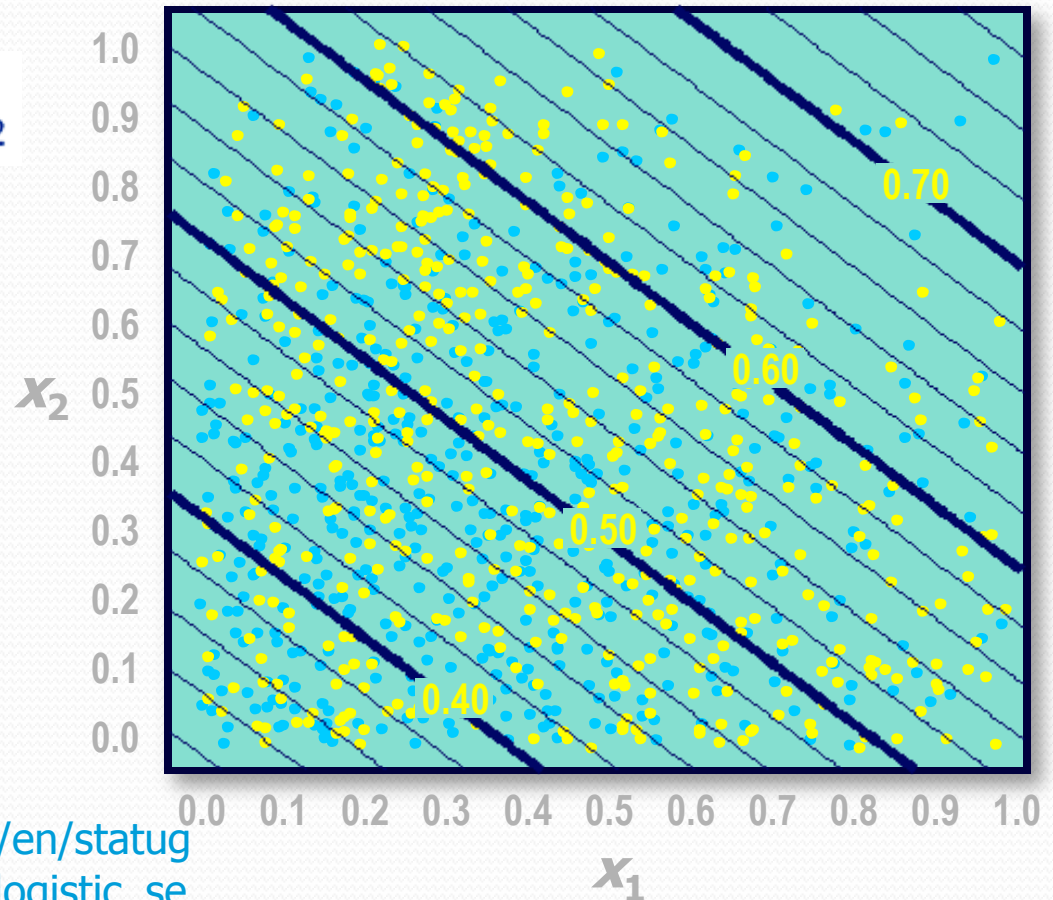
$$\text{logit}(\hat{p}) = -0.81 + 0.92x_1 + 1.11x_2$$

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

Using the maximum likelihood estimates, the prediction formula assigns a logit score to each x_1 and x_2 .

Další příklad na:

http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_logistic_section002.htm



Odhad parametrů

- Metoda maximální věrohodnosti vede na soustavu nelineárních rovnic.
- Tuto soustavu řešíme Newton-Raphsonovou iterační metodou.
- Více na:
 - <http://www.stat.cmu.edu/~cshalizi/402/lectures/14-logistic-regression/lecture-14.pdf>
 - <http://czep.net/stat/mlelr.pdf>
 - <http://www.stat.psu.edu/~jiali/course/stat597e/notes2/logit.pdf>

Maximálně věrohodný odhad (MLE)

MLE is a general purpose method for parametric model estimation. We will make use of it to estimate the logistic regression.

If we have a model with parametric structure θ , we can compute the **likelihood** that the model will generate a sequence of n observations $\mathbf{D} = (d_1, \dots, d_n)$.

$$L(\theta|\mathbf{D}) = P(\mathbf{D}|\theta)$$

The model which best fits the data is selected as the one which maximizes this likelihood.

$$\hat{\theta} = \arg \max_{\theta} L(\theta|\mathbf{D})$$

If we *assume independence between the observations*, this then gives

$$\hat{\theta} = \arg \max_{\theta} \prod_{i=1}^m P(d_i|\theta)$$

Maximálně věrohodný odhad

This MLE can be expressed more conveniently in terms of log-likelihoods (since log is monotonic on its argument):

$$\hat{\theta} = \arg \max_{\theta} \sum_{i=1}^m \log P(d_i | \theta)$$

Remember:

- We do not know the true value of the parameter θ , but we want to estimate it.
- To distinguish the estimate from the true value, in our notation, we put a "hat" on the estimate: $\hat{\theta}$.

MLE has several nice asymptotic properties:

- Consistency
- Asymptotic normality
- Efficiency.

Maximálně věrohodný odhad

Consider the training data set D_{train} with n observations (borrowers).

Remember

- \mathbf{x}_i denotes values for predictor variables for observation i .
- y_i denotes the outcome for observation i , either 0 or 1.

Then the likelihood of the outcome for each observation i is given by

$$\begin{array}{ll} P(y_i = 0 | \mathbf{x}_i, \boldsymbol{\beta}) & \text{if } y_i = 0, \\ 1 - P(y_i = 0 | \mathbf{x}_i, \boldsymbol{\beta}) & \text{if } y_i = 1 \end{array}$$

which is

$$P(y_i = 0 | \mathbf{x}_i, \boldsymbol{\beta})^{1-y_i} (1 - P(y_i = 0 | \mathbf{x}_i, \boldsymbol{\beta}))^{y_i}$$

giving log-likelihood for each observation:

$$(1 - y_i) \log P(y_i = 0 | \mathbf{x}_i, \boldsymbol{\beta}) + y_i \log(1 - P(y_i = 0 | \mathbf{x}_i, \boldsymbol{\beta}))$$

Maximálně věrohodný odhad

Assuming independence between observations, this gives the log-likelihood function for β :

$$\log L(\beta|D_{\text{train}}) = \sum_{i=1}^n (1 - y_i) \log\left(\frac{1}{1 + e^{-(\beta_0 + \beta \cdot \mathbf{x}_i)}}\right) + y_i \log\left(\frac{1}{1 + e^{\beta_0 + \beta \cdot \mathbf{x}_i}}\right)$$

Differentiating by each coefficient in β and setting the derivative equal to zero to find the maxima gives

$$\sum_{i=1}^n \left(1 - y_i - \left(\frac{1}{1 + e^{-(\beta_0 + \beta \cdot \mathbf{x}_i)}}\right)\right) = 0$$

and

$$\sum_{i=1}^n x_{ij} \left(1 - y_i - \left(\frac{1}{1 + e^{-(\beta_0 + \beta \cdot \mathbf{x}_i)}}\right)\right) = 0$$

for each attribute $j=1$ to m .

These are non-linear equations that can be solved by computer intensive processes such as Newton-Raphson methods.

Standard errors on the MLE

Since $\hat{\theta}$ is only an estimate of the best model to explain the data, it is possible to derive standard errors \hat{s} on the estimates.

Asymptotic normality for MLE is such that

$$\frac{(\hat{\theta}_j - \theta_j)}{\hat{s}_j} \rightarrow N(0,1) \text{ as } n \rightarrow \infty$$

where $\hat{\theta}_j$, θ_j and \hat{s}_j are the j th components of $\hat{\theta}$, θ and \hat{s} respectively and $N(0,1)$ is the standard normal distribution.

This property then allows us to generate:-

- Generate a hypothesis tests using the Wald chi-square statistic;
- Generate confidence intervals around the estimate.

MLE- testování hypotéz

We test the hypothesis that an estimated coefficient is not zero against the null hypothesis that it is zero. That is, we testing if a parameter has a genuine effect in the model.

- Null hypothesis: $H_0: \theta_j = 0$
- Alternative hypothesis: $H_1: \theta_j \neq 0$

The Wald test says reject H_0 if $\frac{|\hat{\theta}_j|}{\hat{s}_j} > Z_{\alpha/2}$ for some significance level α , where $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ and Φ is the CDF for the standard normal distribution.

MLE – konfidenční intervaly

The asymptotic normality property also allows us to compute confidence intervals (CIs):

$$P(\hat{\theta}_j - z_{\alpha/2}\hat{S}_j < \theta_j < \hat{\theta}_j + z_{\alpha/2}\hat{S}_j) \rightarrow 1 - \alpha$$

as $n \rightarrow \infty$.

This is a range of possible values of the parameter within a given confidence level $1 - \alpha$.

Note: the larger the confidence level, the broader the confidence interval.

Likelihood Ratio Test

The maximized likelihood gives a measure of how well the model fits the data (1=perfect fit, 0=no fit). The ratio of likelihoods between two models, A "nested" in B, can be used to test whether the fit of A improves on B.

Definitions

Suppose we have two models A and B with the same structure except A has more parameters than B:

$$\theta_A = (\theta_1, \dots, \theta_{m+r}) \text{ and } \theta_B = (\theta_1, \dots, \theta_m)$$

Then *A is nested in B*.

The *likelihood ratio statistic* is $\lambda = 2 \log \left(\frac{L(\hat{\theta}_A)}{L(\hat{\theta}_B)} \right)$.

Newton-Raphsonova metoda

- Základní princip metody:

$$p(x, \beta) = \frac{1}{1 + e^{-\beta^T x}} \quad L(\beta) = \sum_{i=1}^n y_i \beta^T x_i - \log(1 + e^{\beta^T x_i}) \quad \beta^{new} = \beta^{old} - \frac{\partial^2 L(\beta)^{-1}}{\partial \beta \partial \beta^T} \frac{\partial L(\beta)}{\partial \beta}$$

- Maticový zápis:

$$\beta^{new} = (X^T W X)^{-1} X^T W (X \beta^{old} + W^{-1} (y - p))$$

y ... vektor pozorování vysvětlované proměnné

X ... matice plánu, typu $n \times (p + 1)$

p ... vektor pravděpodobností $p(x_i, \beta^{old})$

W ... $n \times n$ diagonální matice vah, s diag. prvky $p(x_i, \beta^{old}) \cdot (1 - p(x_i, \beta^{old}))$

- Jde o numerickou iterační metodu -> je třeba zkontrolovat, zda byla splněna podmínka konvergence (metoda „dokonvergovala“ k optimálnímu řešení)

Výhody logistické regrese

- Málo parametrů
- Snadné použití i interpretace
- Lze snadno začlenit i diskrétní prediktory
- Funguje dobře i na datech, která se poměrně značně liší od gaussovských směsí
- A především většinou dobře funguje, pokud věnujeme odpovídající pozornost přípravě dat
 - praktická zkušenost: ve čtyřech případech z pěti je logistická regrese na datech, která analyzuji, buď nejlepší nebo zhruba stejně dobrá jako jiné metody.

Interpretace, rozdíly proti OLS

- Regresní koeficienty b : kladné znamenají, že proměnná svým růstem zvyšuje šanci zařazení do skupiny kódované číslem 1, a naopak záporné indikují pokles této šance
- Často se používá $\exp(b_i)$: je to faktor, kterým se násobí šance $p/(1-p)$ při jednotkovém nárůstu x_i a neměnných ostatních x_k
 - Pozor na různá měřítka, v nichž x_i mohou být měřena;
- Místo F-testu celkové validity nyní máme chí-kvadrátový test pro totéž
- Místo t-testu signifikance proměnných v modelu jsou Waldovy statistiky; je to v podstatě totéž a čteme to stejně
- Místo R^2 jsou jen pseudo- R^2

Příklad

The following logistic regression output was produced on a data set of 40,000 credit cards.

Likelihood Ratio = 1819 (p-value < 0.001)

Variable	Coefficient	Estimate	Standard error	Wald chi-square	P > chi-square
Intercept	β_0	-0.181	0.084	4.6	0.032
Age	β_1	+0.0353	0.0013	757.6	<0.001
Income (log)	β_2	-0.0164	0.0100	2.67	0.10
Residential phone	β_3	+0.622	0.030	430.8	<0.001
Home owner *		0			
Renter	β_4	-0.155	0.039	15.6	<0.001
Lives with parents	β_5	+0.256	0.045	32.1	<0.001
Months in residence	β_6	-0.00025	0.00011	5.4	0.020
Months in current job	β_7	+0.00210	0.00025	72.9	<0.001

* Notice that the Home owner category is set as base residency category and so has no coefficient estimate. We will discuss this in a later lecture.

Příklad

We have used logistic regression to model the negative outcome (ie $y = 0$).

- This may seem odd given that the outcome of interest is the positive one (eg default).
- However, this model ensures the log-odds scores are the right way round: ie increasing scores imply increasing creditworthiness.
- There is no material difference. If we had modelled $y = 1$, the signs on the coefficient estimates would be reversed but everything else would be the same.

Interpretations:

- The estimates (highlighted) form the scorecard.
- Estimates greater than 0 indicate relative decrease in risk.
- Estimates less than 0 indicate relative increase in risk.
- Small p-values indicate coefficients that are statistically significantly different to zero (how small?).
- Large p-values indicate coefficients that have a good chance of actually being zero.

Příklad

Remember in the exercise in Chapter 1 we gave details of six borrowers. You were asked to select three to accept and three to reject.

Here the scores assigned by the model above are shown. The observations with the three lowest scores are rejected by the model. The actual outcome in each case is also shown. *How does your performance compare with the model?*

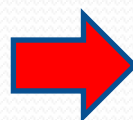
Age	Monthly Income (£)	Residential phone?	Residence type?	Months in residence	Months in current job	Score	Model accept or reject?	Actual outcome
22	1,145	Yes	Home owner	48	12	1.11	Reject	Good
46	15,500	Yes	Renter	48	192	2.14	Accept	Good
71	900	Yes	Renter	96	12	2.68	Accept	Good
32	5,000	Yes	Renter	48	168	1.61	Accept	Bad
25	1,385	Yes	Renter	12	0	1.05	Reject	Bad
43	3,145	No	Home owner	96	36	1.25	Reject	Bad

Příklad

Variable	Value	Coefficient	Estimate	Value × Estimate
Intercept	n/a	β_0	-0.181	-0.181
Age	22	β_1	+0.0353	+0.777
Income (log)	log(1145) =7.04	β_2	-0.0164	-0.116
Residential phone	1	β_3	+0.622	+0.622
Home owner *	1		0	0
Renter	0	β_4	-0.155	0
Lives with parents	0	β_5	+0.256	0
Months in residence	48	β_6	-0.00025	-0.012
Months in current job	12	β_7	+0.00210	0.025
Score (sum)				+1.115

Compute the PD of the borrower.

Score = 1.115



$$P(y = 1|s) = \frac{1}{1+e^s} \approx 0.25.$$

Multinomiální logistická regrese

- Taktéž polytomická regrese
- Závisle proměnná má M kategorií, více než dvě. Např.: kterou stranu respondent volí?
- Základní idea:
 - Prohlásit jednu kategorii za referenční
 - Spočítat $M-1$ obyčejných logistických modelů pro každou ze zbylých kategorií oproti referenční
 - A predikovat tu kategorii, kde vyšla největší pravděpodobnost přes všechny modely



Budování modelu

- ❑ Forward
 - začíná se s prázdným modelem
 - postupné přidávání proměnných
- ❑ Backward
 - začíná se s plným modelem (všechny proměnné)
 - postupné odebrání proměnných
- ❑ Stepwise
 - začíná se s prázdným modelem
 - postupně se přidávají a odebírají proměnné
- ❑ Enter
 - je předepsán seznam proměnných v modelu

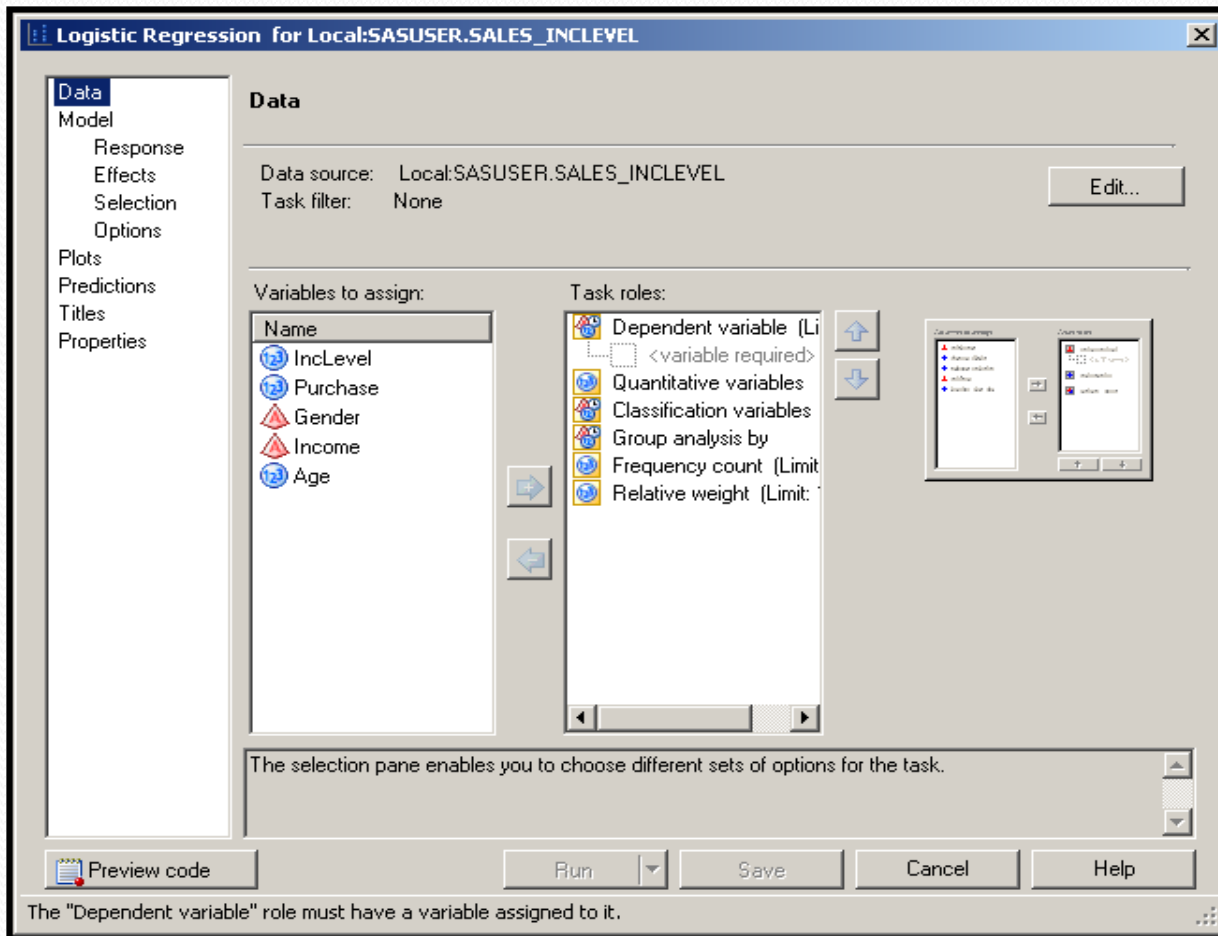
Logistic Regression with Sequential Steps

- Forward regression
 - starts with a baseline model (intercept-only)
 - searches all variables and finds the strongest one
 - keeps adding variables in order of strength until no significant improvement is achieved in the model.
- Backwards regression
 - starts with a full model using all variables
 - removes the weakest input variable provided that taking it out does not cause a significant reduction in the fit of the model
 - continues removing the weakest input variables in order unless there is a significant reduction in the fit of the model; at which point the algorithm stops.

Logistic Regression with Sequential Steps

- Stepwise regression
 - is a combination of forward and backward regression
 - begins the same way as forward
 - re-evaluates the statistical significance of all included variables after each new variable is added.
-  If a previously included variable becomes statistically insignificant when a new variable is added, that variable is then removed.
-  The algorithm stops when no more variables can be found that add significantly to the fit of the model **and** all variables remaining in the model are statistically significant.

The Logistic Regression Task

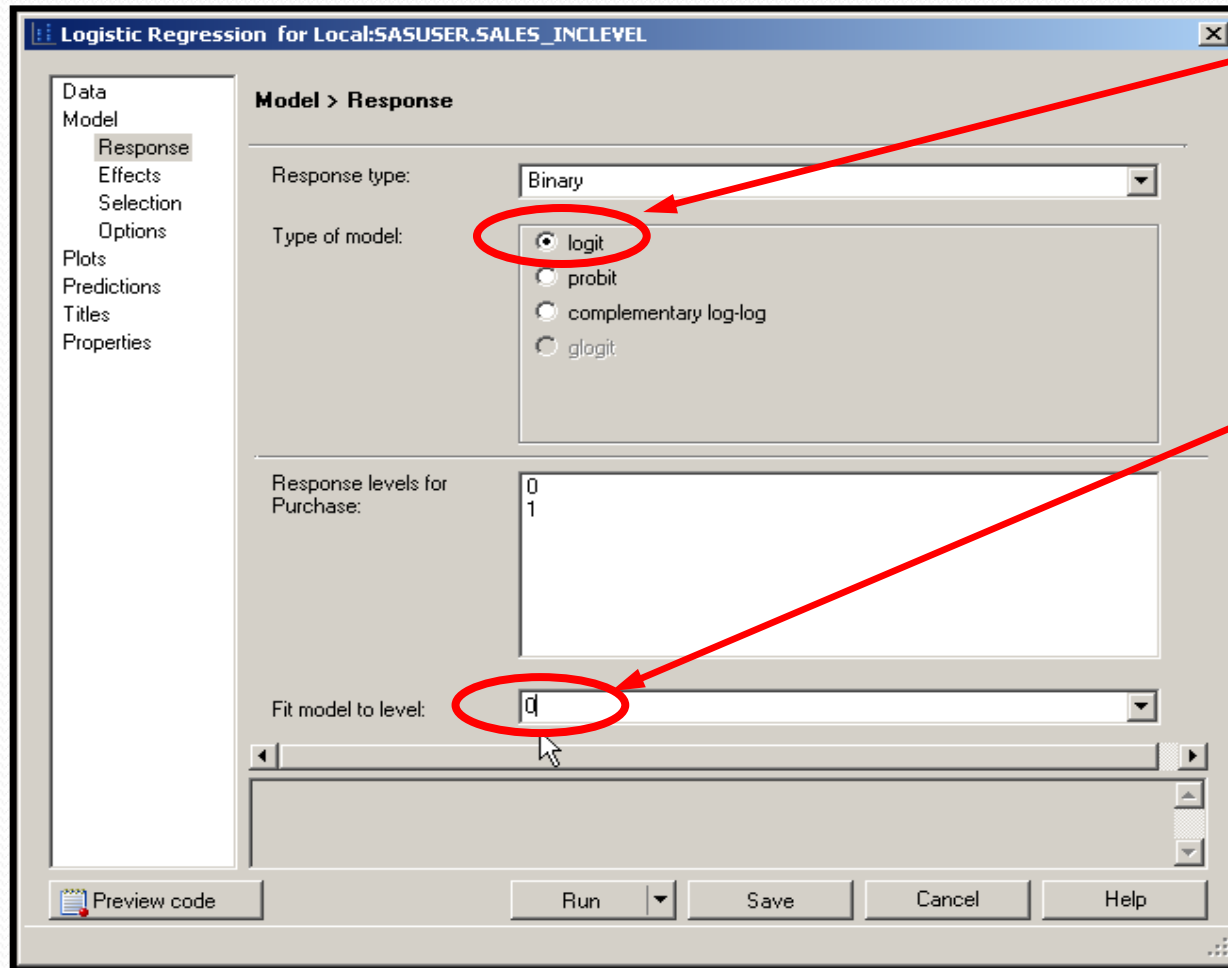


Which link function, which response Level to Model?

Volba linkovací funkce.

Specify the level of the response variable that you want to model.

For example, do you want to model the probability of a 0 or a 1?



LOGISTIC Procedure

- General form of the LOGISTIC procedure:

```
PROC LOGISTIC DATA=SAS-data-set
  <options>;
CLASS variables </ options>;
MODEL response=predictors </ options>;
UNITS independent1=list ... </ options>;
ODDSRATIO <'label'> variable </ options>;
OUTPUT OUT=SAS-data-set
  keyword=name </ options>;
RUN;
```

Více např. na: <http://www.okstate.edu/sas/v8/sashtml/onldoc.htm>

<http://www.okstate.edu/sas/v8/saspdf/stat/chap39.pdf>

LOGISTIC Procedure - příklad

```
ods html file="logistic_vyvoj.html" style=sasweb;  
proc logistic data=dm1.data_vyvoj descending;  
model good4=goods_type_w phone_w a_uver_w  
      fam_state_w income_w credit_w vek_w  
      ;  
run;  
ods html close;
```

LOGISTIC Procedure - příklad

```
proc logistic data=dm1.score_base outest=work.model_def;  
  
CLASS AGE_d EDUCATION_d CAR_AGE_d / param=glm;  
MODEL def_bad = AGE_d EDUCATION_d CAR_AGE_d  
total_income_d(init_pay_by_INCOME_d)  
  
/ SELECTION=FORWARD HIERARCHY=MULTIPLECLASS;  
score out=work.tab_scored_def;  
run;
```

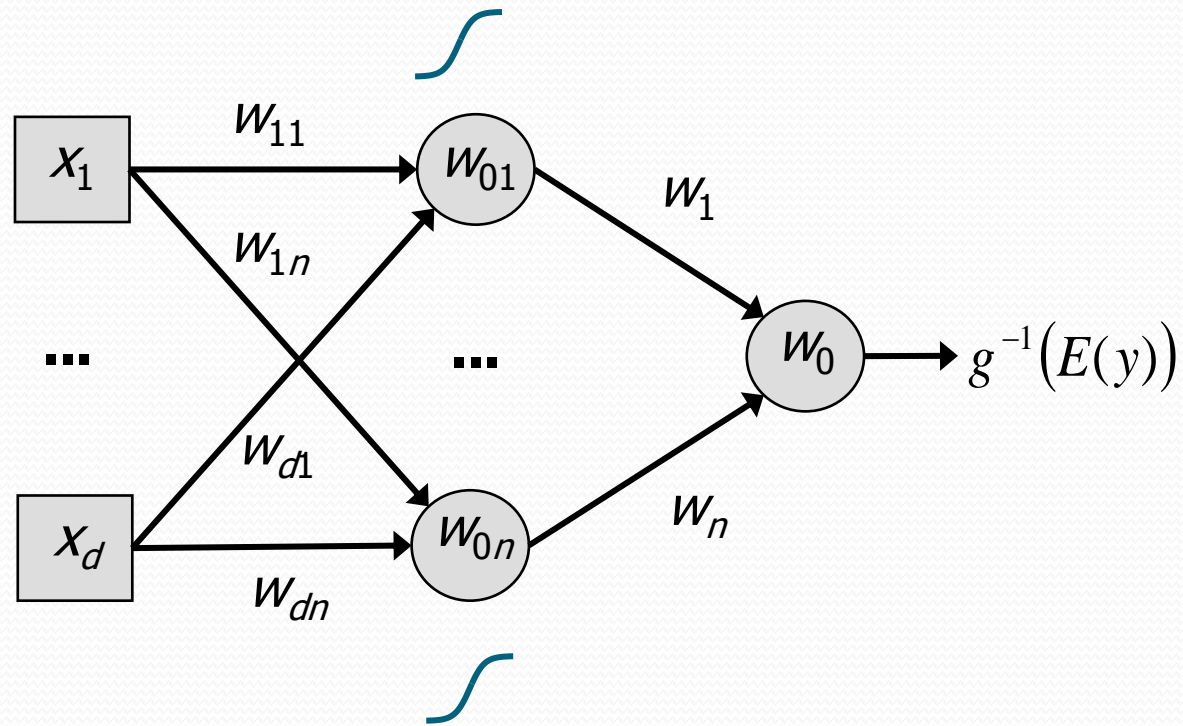
LOGISTIC Procedure - příklad

```
proc logistic  
data=dm1.score_base outest=work.model_def namelen=200;  
where client_type="1-Novy";  
CLASS sex_k child_num_k fam_state_k age_k;  
MODEL def_bad = AGE_w EDUCATION_w  
    AGE_w*EDUCATION_w  
    sex_k|child_num_k|fam_state_k|age_k@4  
  
/selection=stepwise slentry=0.6 slstay=0.1 details corrb  
;  
run;
```

LOGISTIC Procedure - příklad

```
proc logistic  
data=dm1.score_base inest=hc.modelSU namelen=200;  
CLASS sex_k child_num_k fam_state_k age_k;  
MODEL def_bad = AGE_w EDUCATION_w  
    AGE_w*EDUCATION_w  
    sex_k|child_num_k|fam_state_k|age_k@4  
  
/selection=none maxiter=0;  
output out=dm1.data_all_scr (keep=id_credit score def_bad  
    compress=yes)  
prob=score;  
run;
```

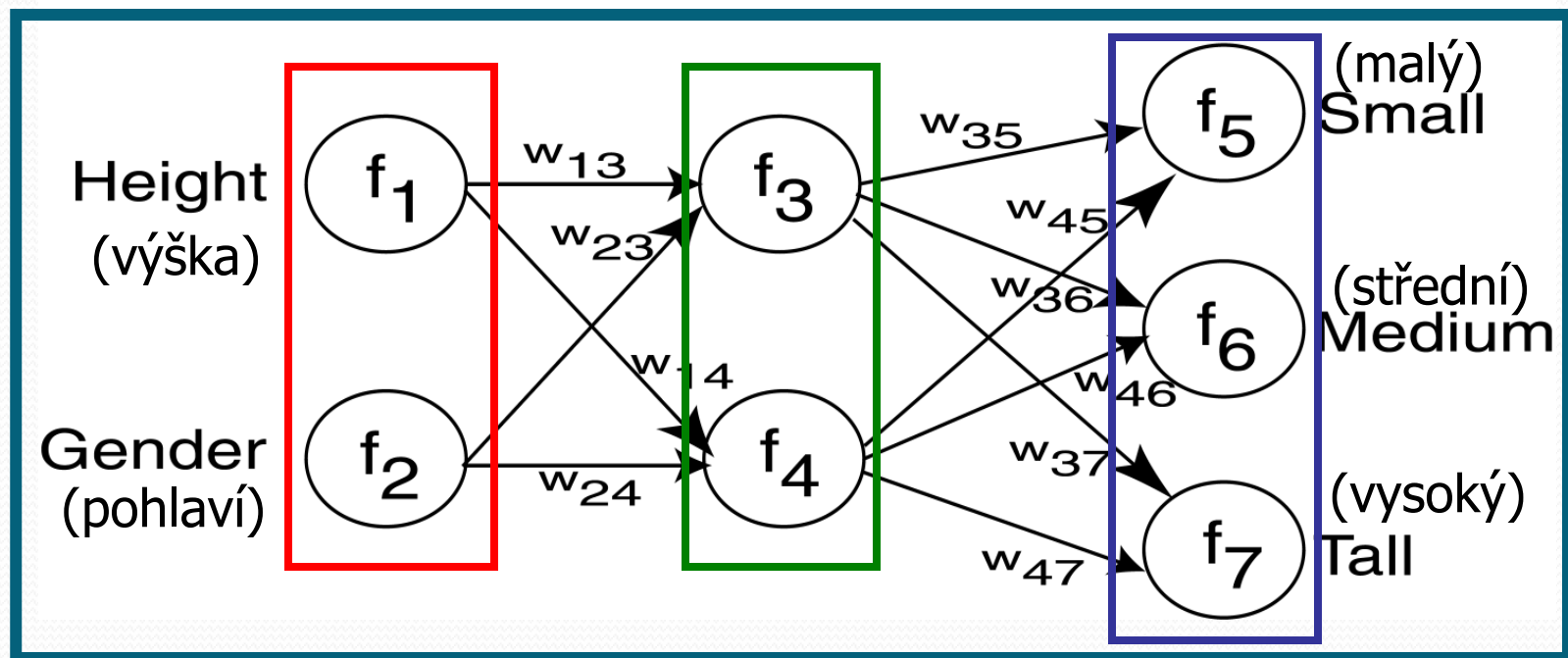
3. Neuronové sítě



Neuronové sítě (Neural Networks)

- Někdy se také uvádí název *Artificial Neural Networks (ANN)*, tj. *umělé neuronové sítě*.
- Založené na pozorované funkcionalitě lidského mozku.
- Ovšem v porovnání s mozkiem jde o velmi zjednodušený matematický model.
- Často jde u NN o adaptivní systém, který mění svou strukturu na základě vnějších či vnitřních informací získaných v průběhu učící fáze.
- Využívají se např. při vyhledávání vzorů v datech, rozpoznávání řeči nebo klasifikačních problémech.
- http://en.wikipedia.org/wiki/Artificial_neural_network

Příklad neuronové sítě

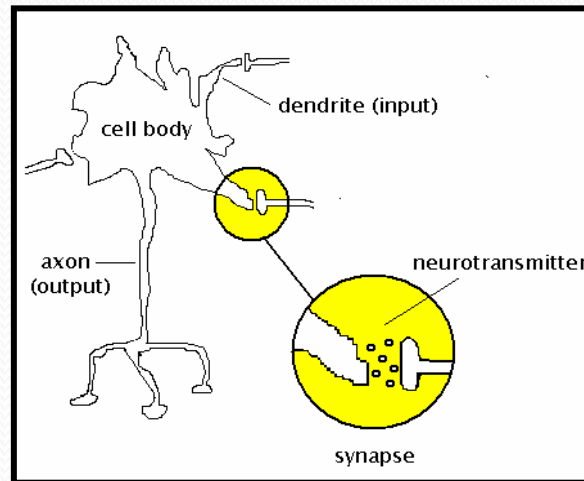


vstup
(input)

skrytá vrstva
(hidden layer)

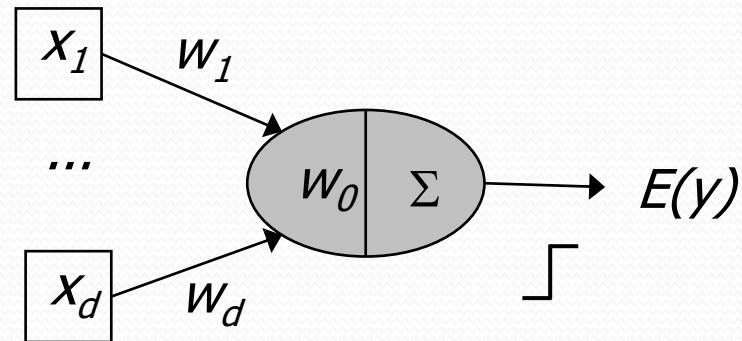
výstup
(output)

The Neuron



- Excitatory (+) and inhibitory (-) inputs, arriving at the dendrites, are weighted by adaptable synapses.
- The weighted inputs are added together.
- If the sum is greater than an adaptable threshold (bias) value, the neuron sends activation down its axon.

The McCulloch-Pitts Neuron



- A McCulloch-Pitts neuron with d inputs is formally defined by the following equation:

$$E(y) = f\left(w_0 + \sum_{i=1}^d w_i x_i\right)$$

- The step function, $f(\cdot)$, turns each McCulloch-Pitts neuron into a linear classifier/discriminator.

The Hebb Rule

- The strength of the connection between neurons i and j should be adjusted in accordance with the equation:

$$\Delta w_{ij} = \eta \hat{y}_i x_j$$

- *The eta (η) term is the neuron's learning rate, which scales the amount of weight adjustment.*
- Permitted learning rate values range from 0 to 1.
- Large learning rate values risk divergence.

The Widrow-Hoff Delta Rule

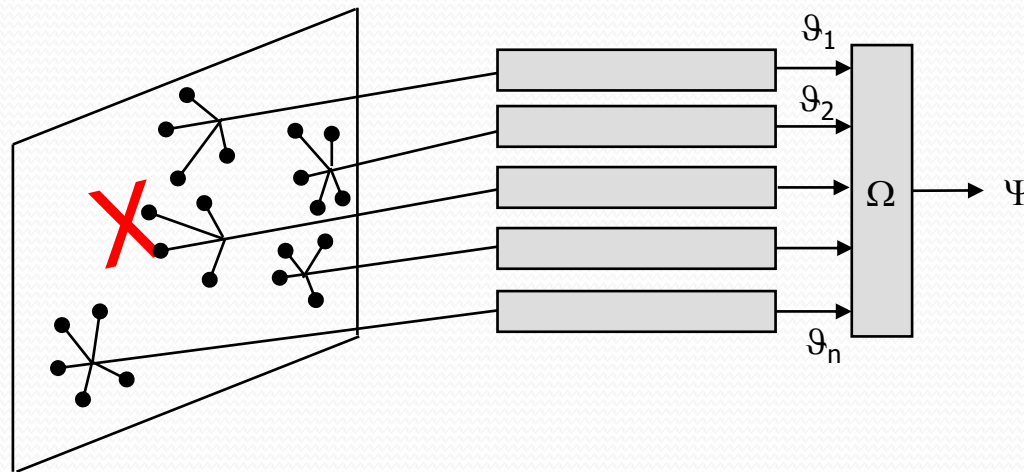
- Hebb's learning rule is unstable.
- Widrow and Hoff proposed a variant of Hebb's rule, one that is stable under a range of learning rates:

$$\Delta w_{ij} = \eta(y_i - \hat{y}_i)x_j$$

- They called their learning model the *delta rule*.
- Because the delta rule reduces the sum of squared error, it is also known as the *least mean squares rule*.

The Perceptron

- The perceptron is a pattern-recognition machine invented in the 1950s for optical character recognition.

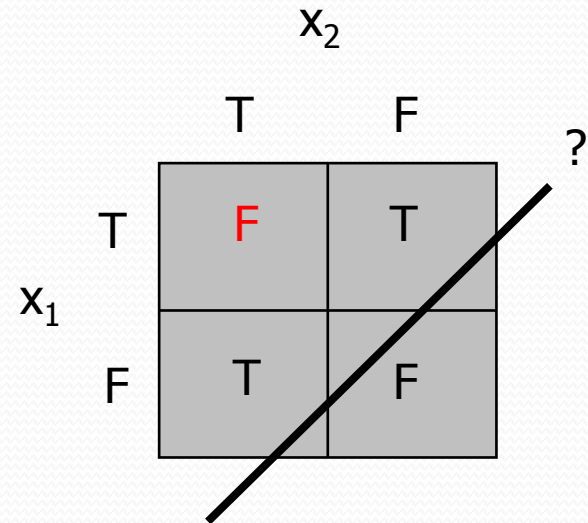


- Each processing unit is a *McCulloch-Pitts neuron*.
- A perceptron with n outputs is a discriminator function that divides the input space into n distinct regions.

The Limitations of a Simple Perceptron

- The simple (linear) perceptron can only solve linearly separable problems.
- The EXCLUSIVE OR truth table (below) is an example of a problem that is **not** linearly separable.

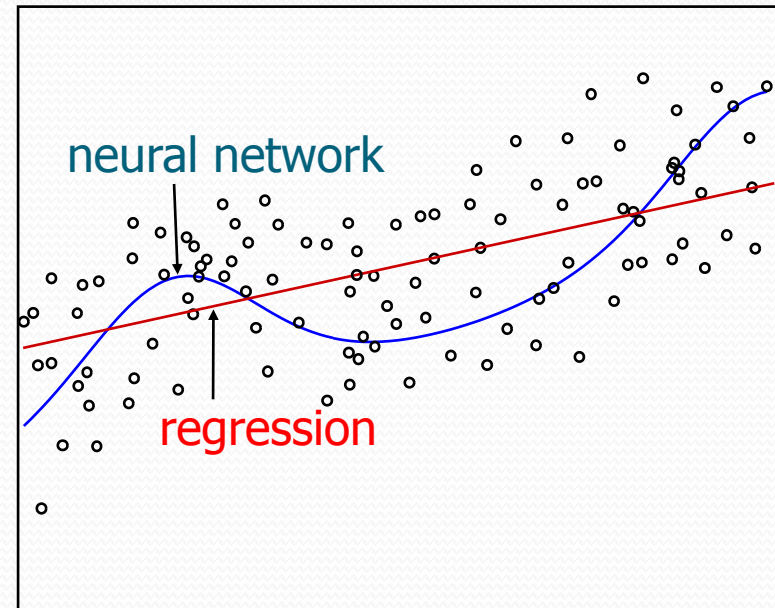
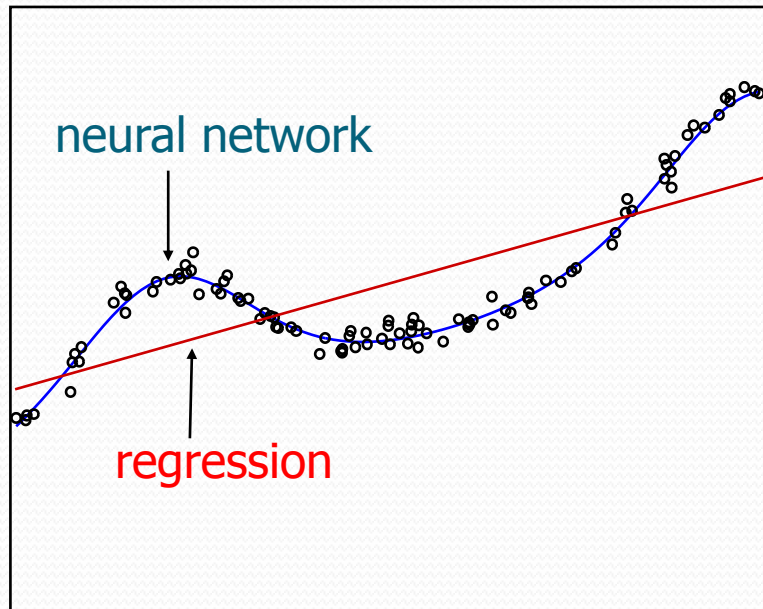
Inputs		Output
x_1	x_2	
F	F	F
T	F	T
F	T	T
T	T	F



Výhody NN

- Schopnost učení.
- Snadná parametrizace.
- Robustnost.
- Řeší mnoho problémů.

The Impact of Noisy Data



Nevýhody NN

- Nesnadné porozumění/interpretace.
- Můžou trpět přeučením (overfitting).
- Vstupy musí být numerické.
- Obtížná verifikace.

Typy neuronových sítí

- Existuje celá řada typů neuronových sítí, přičemž každý z nich se hodí na jinou třídu úlohy.
- Podle přítomnosti „učitele“ dělíme neuronové sítě na
 - sítě s učitelem (srovnávání výstupu s požadovaným)
 - sítě bez učitele (bez vnějšího arbitru).

Typy neuronových sítí podle zpracování signálu

Symbol	Způsob zpracování signálu
–	chybí vrstva
*	žádné
L	lineární kombinace
V	vzdálenost
Z	znaménko
S	sigmoida
G	Gaussova funkce
E	exponenciála
MIN	nejmenší vyhrává
MAX	největší vyhrává

Typy neuronových sítí podle zpracování signálu

Typ sítě	Vrstvy			Autoři
	Vstupní	Skrytá	Výstupní	
OLAM	*	–	L(+Z)	Haykin
HEBB	*	–	L+Z	Hopfield
HAMM	*	L+MAX	L+Z	Lipmann
MLP ₁	*	L+Z	L+Z	Widrow, Hoff
MLP ₂	*	L+S	L+S	Rummelhart
SOM	*	–	V+MIN	Kohonen
RBF	*	V+G	L	Poggio, Girosi
MOD	*	L+E	L	Jacobs, Jordan
COUNT	*	V+MIN	L	Nielsen

Typy neuronových sítí podle zpracování signálu

- V předchozí tabulce je základních devět typů sítí:
 - optimální lineární asociativní paměť (Optimum Linear Associative Memory – OLAM),
 - Hebbova síť (HEBB),
 - Hammingova síť (HAMM),
 - vícevrstvá síť s bipolárními neurony (Multi Layer Perceptron 1 – MLP₁),
 - vícevrstvá síť se spojitým chováním (MLP₂), Kohonenovy mapy (SOM),
 - síť s radiální bází (RBF),
 - modulární síť (MOD) a
 - síť se zpětným šířením (counterpropagation – COUNT).
- Další sítě lze vytvářet jejími kombinacemi.

Asociativní neuronové sítě

- U asociativní paměti probíhá vybavení příslušné informace na základě její částečné znalosti (asociace).
- Rozlišujeme sítě s pamětí
 - autoasociativní (upřesnění či zúplnění vstupní informace na základě již naučeného)
 - heteroasociativní (vybavení si sdružené informace na základě vstupní asociace)

Učení neuronových sítí

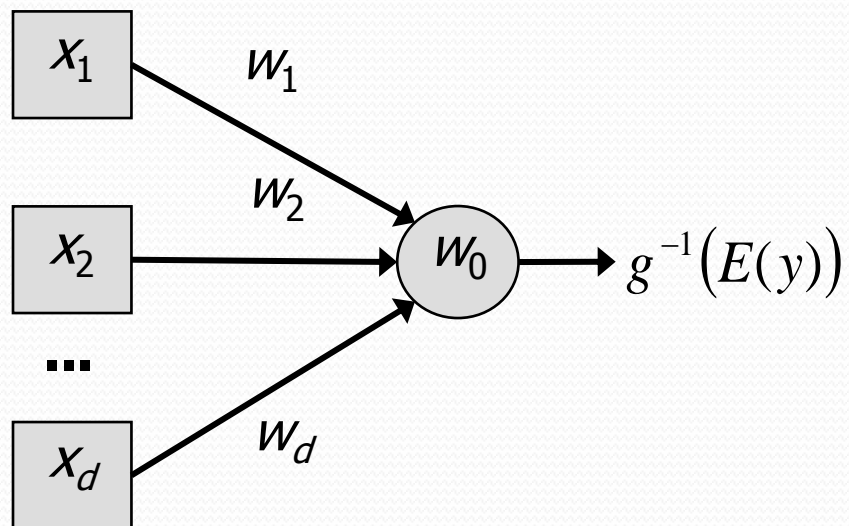
- Algoritmus učení je různý, nicméně obecně má tyto kroky:
 - inicializace vah (malé náhodné hodnoty)
 - předložení nového vzoru (vektor reálných hodnot X)
 - výpočet aktuálního vstupu (podle f aktivační funkce)
 - přizpůsobení vah (přepočtení vah podle zjištěné odchylky)
 - opakování procesu učení (až do stabilizace vah w_i)
- Fáze učení sítě se nazývá adaptivní a po naučení je síť ve fázi vybavování (aktivní fázi).

Využití neuronových sítí

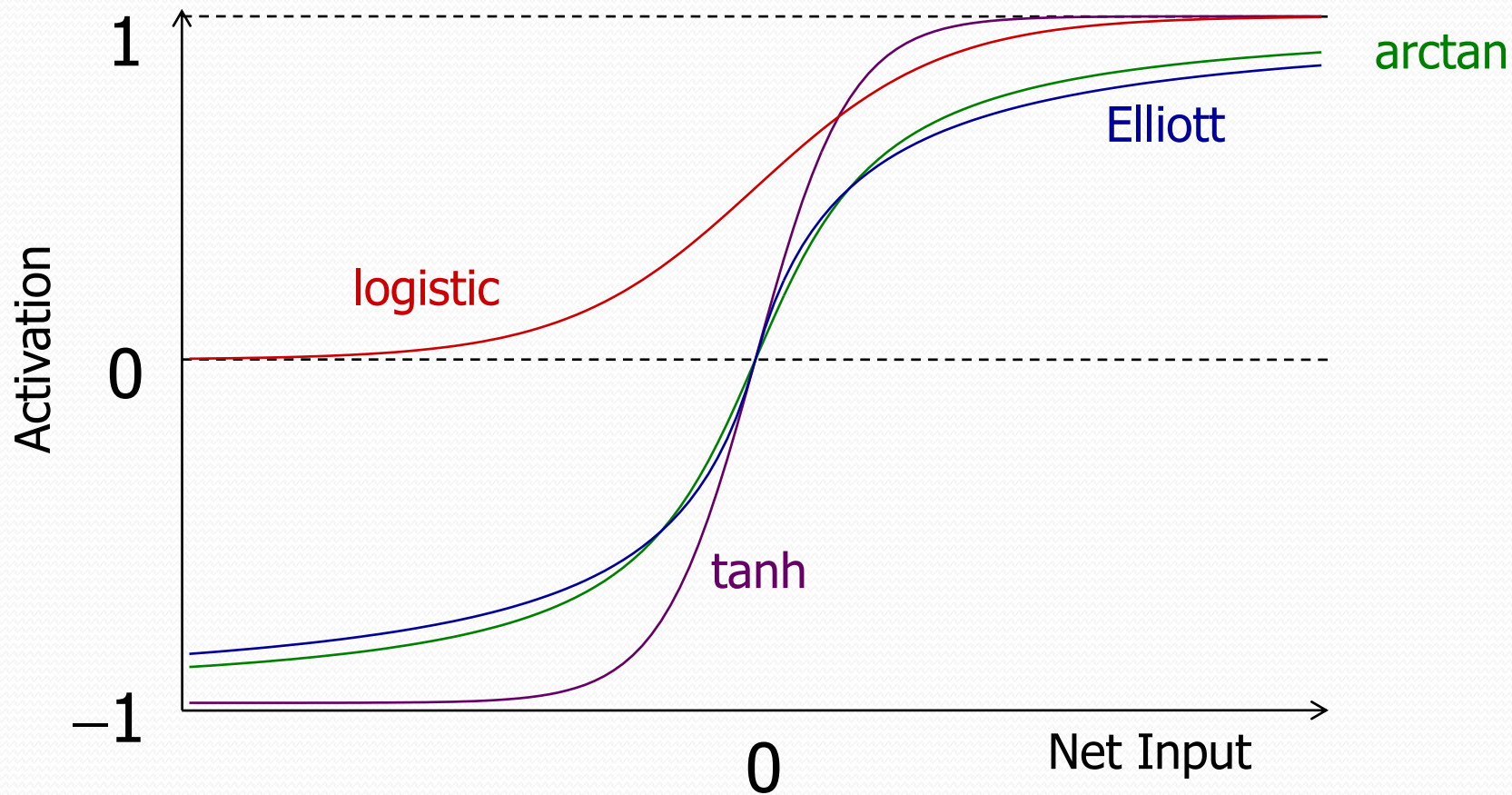
Úloha	Vhodné neuronové sítě
logické obvody	HEBB, HAMM, MLP ₁
odstranění šumu	MLP ₁ , MLP ₂ , RBF, MOD
řeč a výslovnost	MLP ₂ , SOM
kompresce	COUNT
data mining	OLAM, HEBB, SOM
optické rozpoznávání znaků	HEBB, OLAM, HAMM, MLP ₁ , MLP ₂ , RBF, SOM

Linear Perceptron

$$g^{-1}(E(y)) = w_0 + \sum_{i=1}^d w_i x_i$$

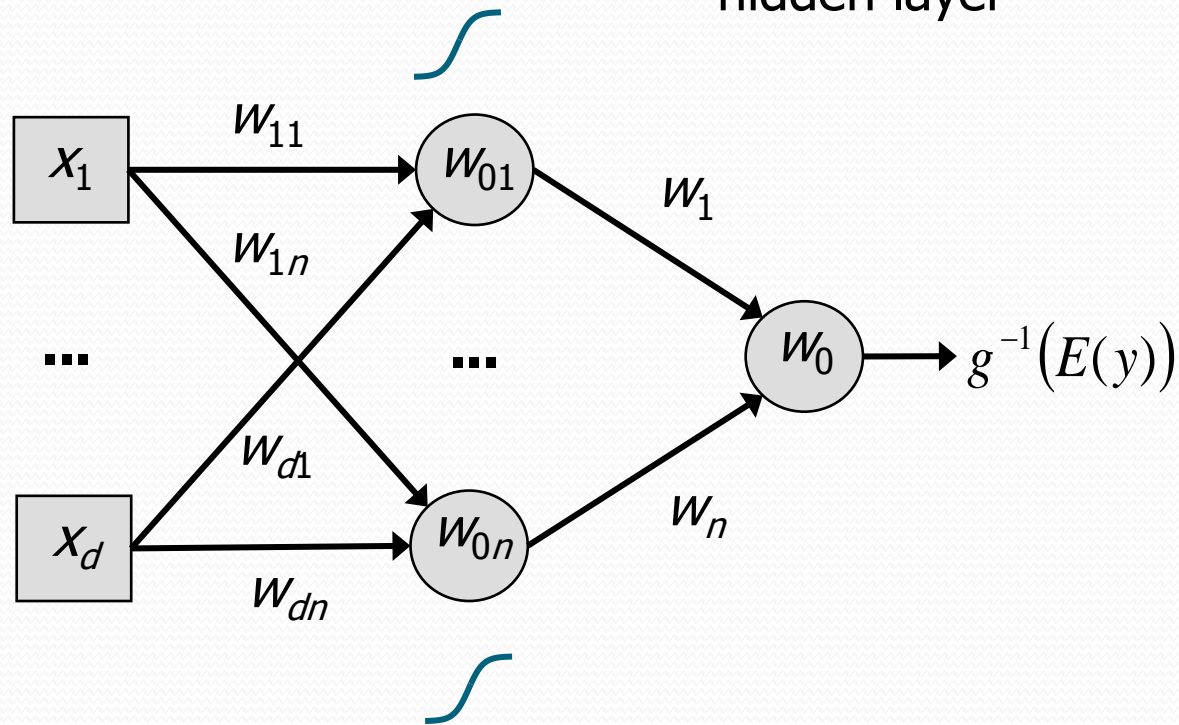


Activation Functions



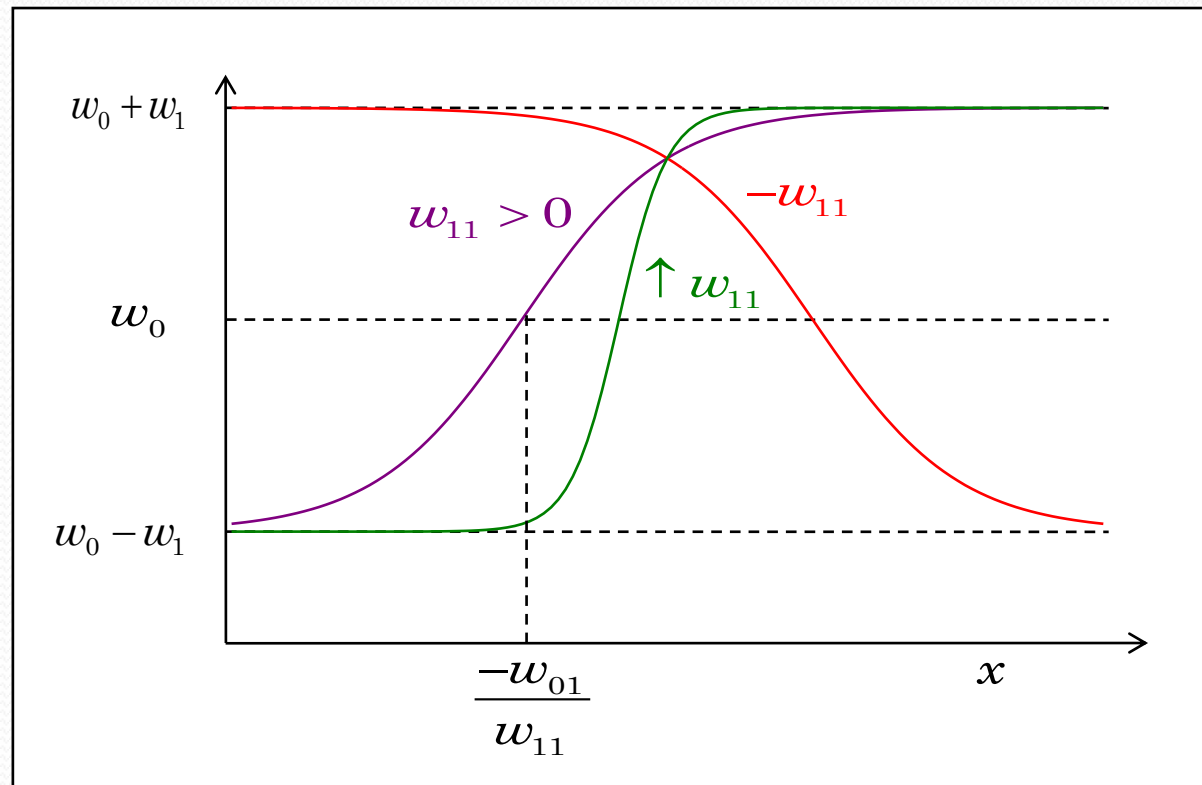
Multilayer Perceptron

$$g^{-1}(E(y)) = w_0 + \underbrace{\sum_{i=1}^h w_i g_i \left(w_{0i} + \sum_{j=1}^d w_{ij} x_j \right)}_{\text{hidden layer}}$$

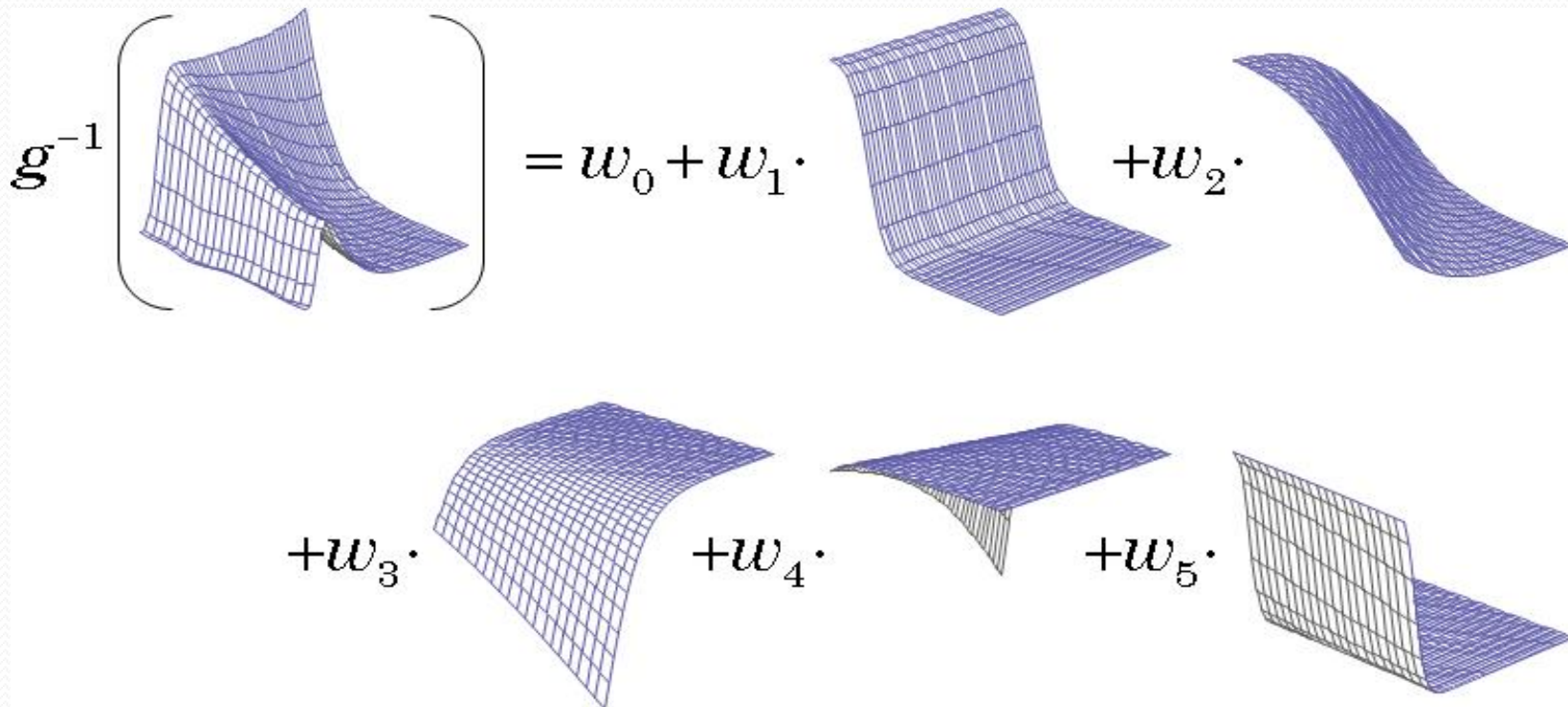


Shaping the Sigmoid

$$w_0 + w_1 \tanh(w_{01} + w_{11}x)$$

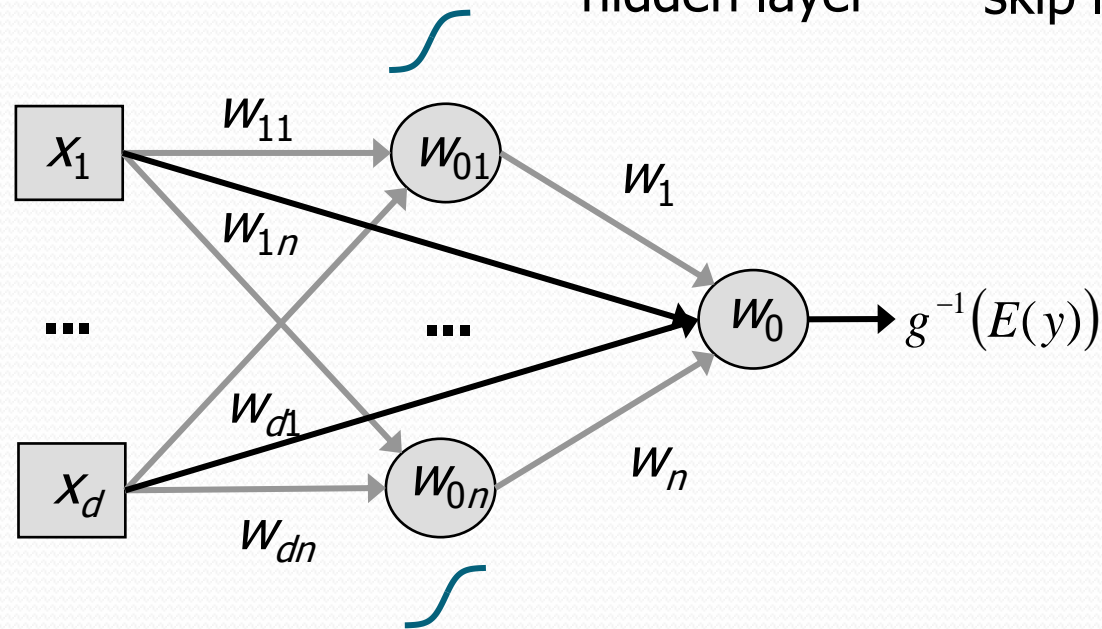


Sigmoidal Basis Functions



Skip-Layer Perceptron

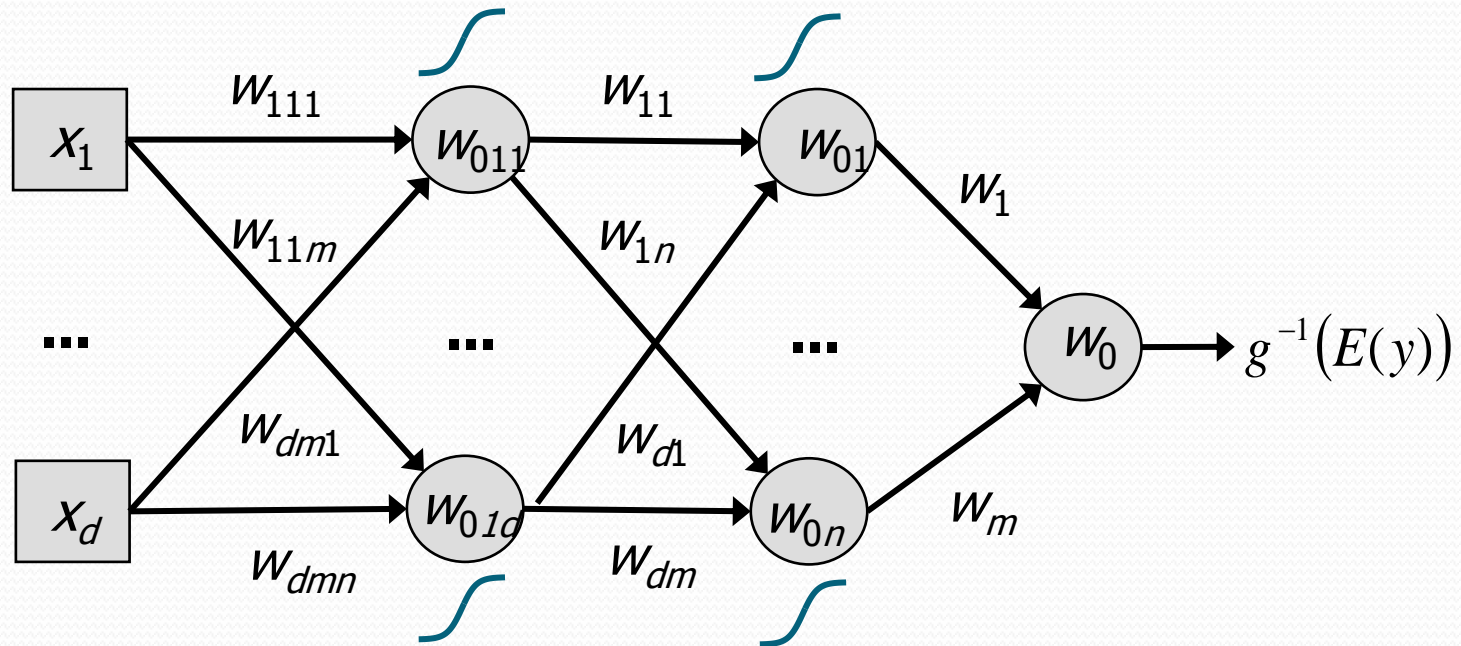
$$g^{-1}(E(y)) = w_0 + \underbrace{\sum_{i=1}^h w_i g_i \left(w_{0i} + \sum_{j=1}^d w_{ij} x_j \right)}_{\text{hidden layer}} + \underbrace{\sum_{k=1}^d w_k x_k}_{\text{skip layer}}$$



MLP with Two Hidden Layers

$$g^{-1}(E(y)) = w_0 + \sum_{k=1}^m w_k g_k \left(w_{0k} + \sum_{j=1}^n w_{jk} g_j \left(w_{0jk} + \sum_{i=1}^d w_{ijk} x_i \right) \right)$$

nested hidden layers

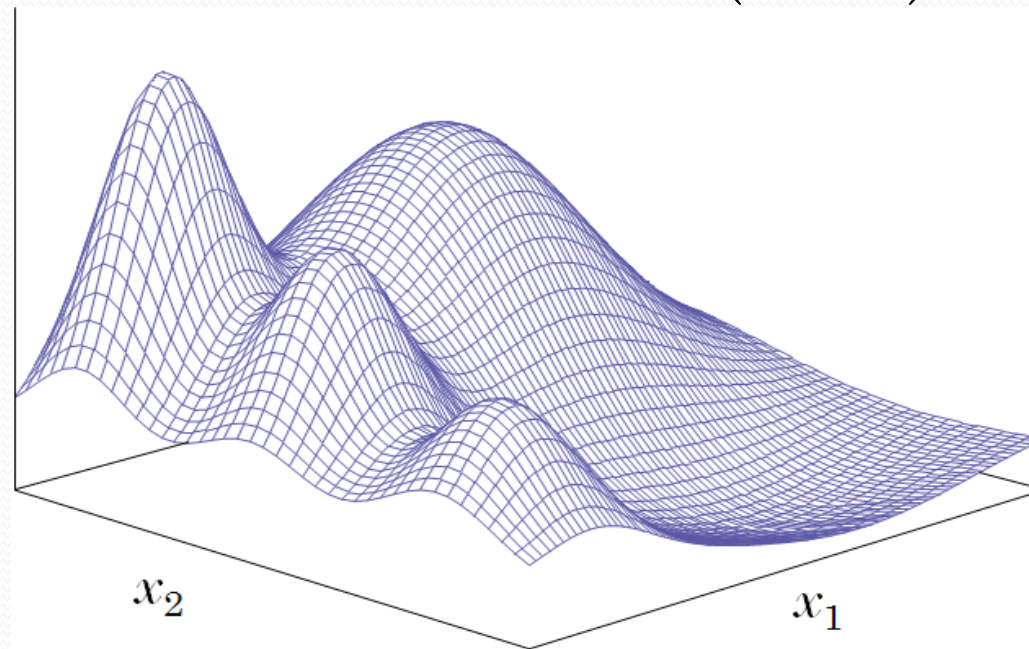


How Many?

- A single hidden layer network models any **continuous** relationship between the inputs and outputs.
- Two hidden layers model **discontinuous** relationships.
- The number of hidden units that will be required in each defined hidden layer is problem specific.

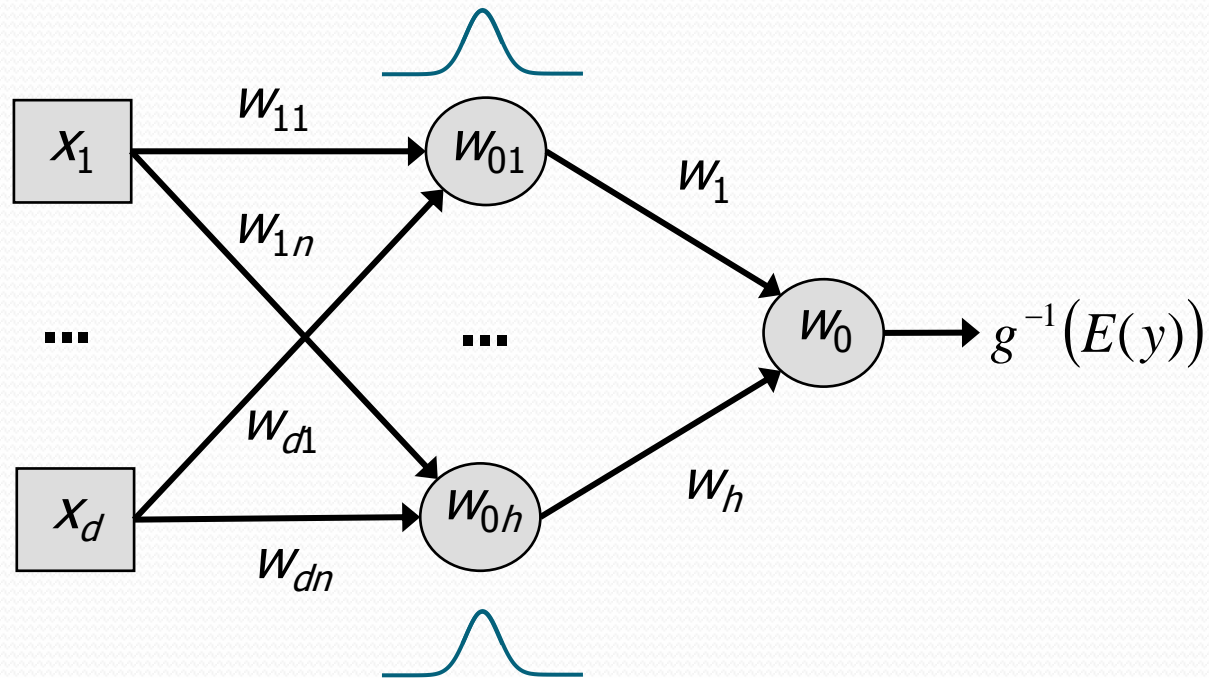
Overview of Radial Basis Functions

- Ordinary Radial Basis Functions (ORBF).
- Normalized Radial Basis Functions (NRBF).



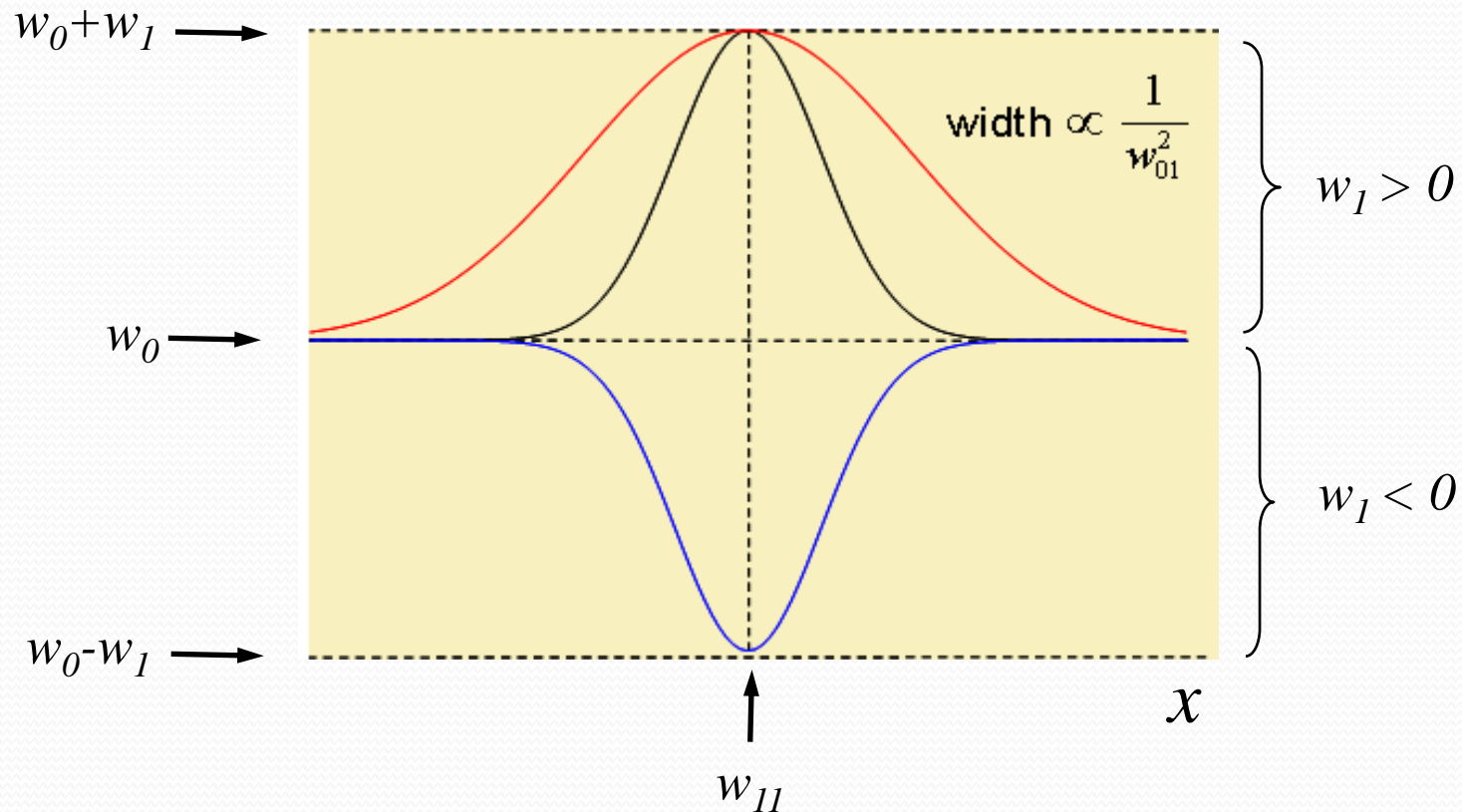
Ordinary Radial Basis Functions

$$g^{-1}(E(y)) = w_0 + \sum_{i=1}^h w_i \underbrace{\exp \left[-w_{0i} \left(\sum_{j=1}^d (x_j - w_{ij})^2 \right) \right]}_{\text{hidden unit}}$$



Shaping the Gaussian

$$w_0 + w_1 \exp\left(-w_{01}^2(x - w_{11})^2\right)$$



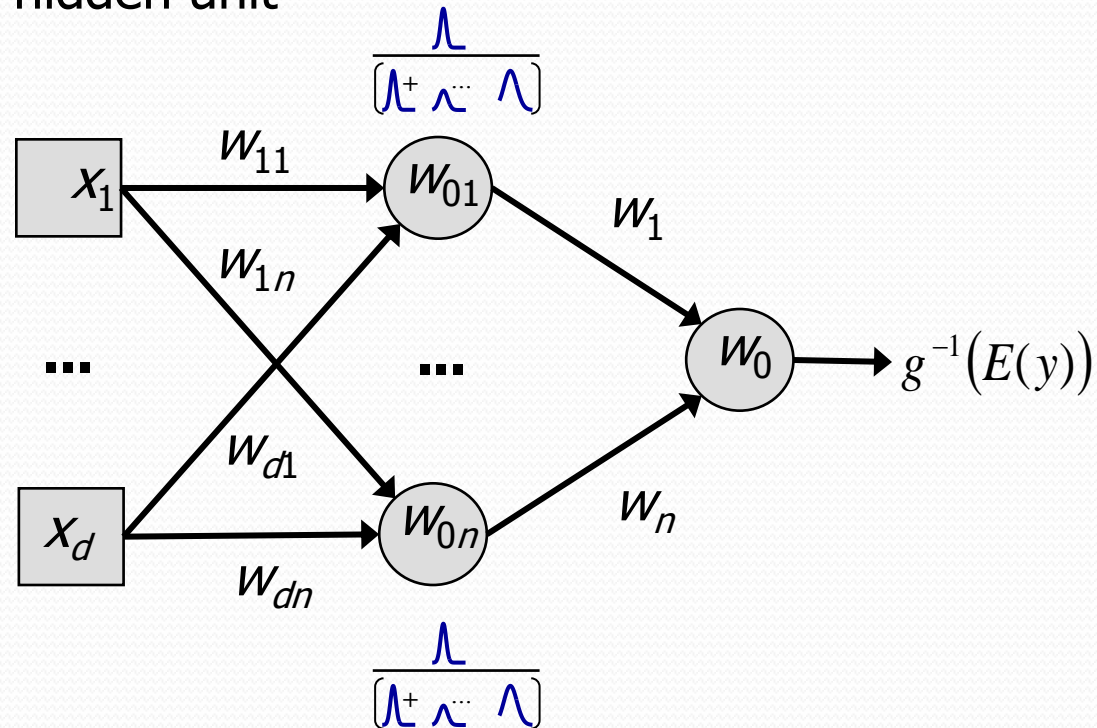
RBF Combination Functions

- **XRADIAL** Unequal Heights and Widths.
- **EQRADIAL** Equal Heights and Widths.
- **EWRADIAL** Equal Widths.
- **EHRADIAL** Equal Heights.
- **EVRADIAL** Equal Volumes.

Normalized Radial Basis Functions

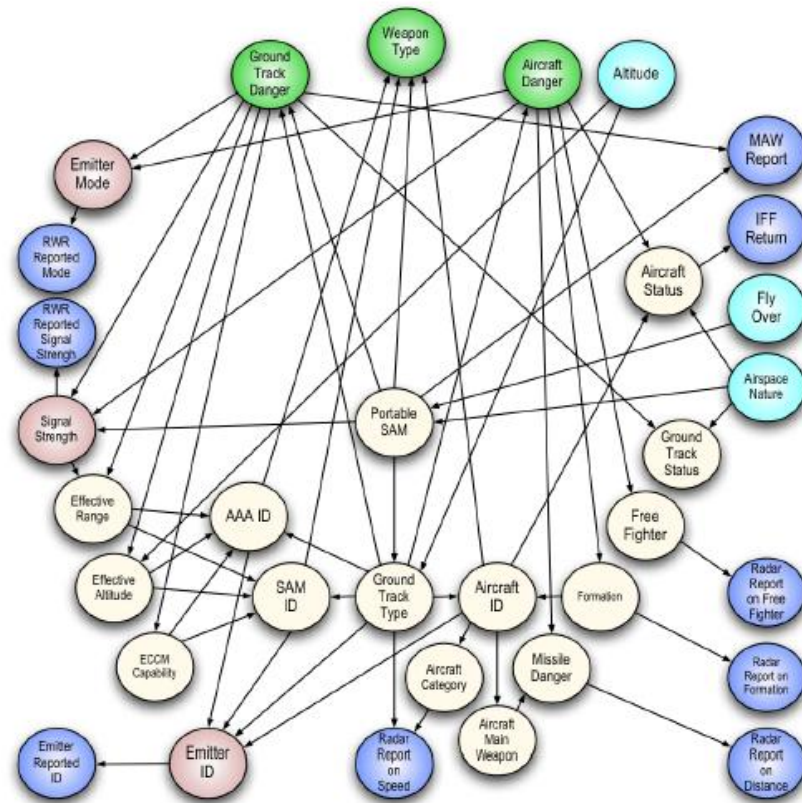
$$g^{-1}(E(y)) = w_0 + \sum_{i=1}^h w_i \underbrace{\left(\frac{e_i}{\sum_{j=1}^k e_j} \right)}_{\text{hidden unit}} \text{ where } e_i = \exp \left[f \cdot \ln(a_i) - w_{0i}^2 \left(\sum_{j=1}^d (x_j - w_{ji})^2 \right) \right]$$

hidden unit



4. Bayesovské sítě

Bayesian Networks (Belief Networks)

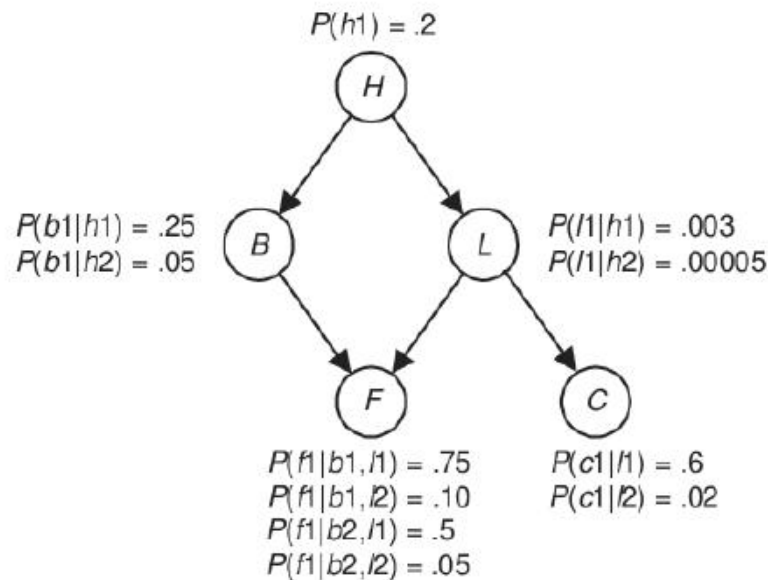


- Bayesův vzorec:
$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

- Typický příklad využití Bayesova vzorce (test na TBC):

Test má senzitivitu 90%, specifitu 1% a TBC trpí 5 lidí z 10 000. Jaká je pravděpodobnost, že osoba, které test určil přítomnost TBC, touto chorobou skutečně trpí? [0,043] --> Vidíme, že (ne)přítomnost TBC má vliv na výsledek testu.

- V reálném světě jsou komplikovanější závislosti. Např. fakt zda osoba kouří má vliv na to, zda trpí bronchitidou nebo rakovinou. Každá z těchto chorob má vliv na kondici a navíc (ne)přítomnost rakoviny plic má vliv na výsledky RTG. Otázkou tedy je, jak tuto situaci řešit.

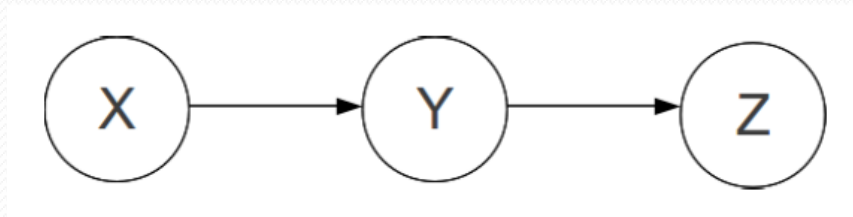


Feature	Value	When the Feature Takes this Value
H	$h1$	There is a history of smoking
	$h2$	There is no history of smoking
B	$b1$	Bronchitis is present
	$b2$	Bronchitis is absent
L	$l1$	Lung cancer is present
	$l2$	Lung cancer is absent
F	$f1$	Fatigue is present
	$f2$	Fatigue is absent
C	$c1$	Chest X-ray is positive
	$c2$	Chest X-ray is negative

Zdroj: Neapolitan, R.E. (2004)

- Pojem Bayesovské sítě jako první použil Judea Pearl v roce 1985.
- Za základní práce na toto téma lze považovat:
 - Pearl, J.: Probabilistic Reasoning in Intelligent Systems (1988)
 - Neapolitan, R.E.: Probabilistic Reasoning in Expert Systems (1989)

- Formalizmus na podporu rozhodování pod vlivem nejistoty.
 - Reprezentuje sdružené rozdělení pravděpodobnosti vektoru náhodných veličin.
 - Při volbě pravděpodobnosti kombinuje historická data (např. zpoždění vlaku) a názor odborníka (např. kvalita testu).
 - Uplatňují se v Risk managementu v oblastech s nedostatečnými nebo žádnými daty, např. při určování rizika teroristického útoku nebo selhání nového systému.
 - Využití v medicínské diagnostice, vyhodnocování rizika, chyb materiálu, kvality softwaru,...
-
- Komerční software: Agenarisk, BayesLab, BNet,...
 - Freeware/ open source: MSBNx, OpenBayes, Powersoft,...



- Orientovaný acyklický graf (DAG).
- Uzel vyjadřuje náhodnou veličinu (diskrétní nebo spojitou).
- Orientovaná hrana vyjadřuje závislost mezi danými uzly.
- Každý uzel je závislý na svých „rodičích“, jinak jsou uzly navzájem podmíněně nezávislé (D-separované).
- Zjednodušení:
 - z $P(X, Y, Z) = P(X) * P(Y|X) * P(Z|Y, X)$
 - na $P(X, Y, Z) = P(X) * P(Y|X) * P(Z|Y)$
- Pro Y platí: X je jeho „rodič“ (přímý předchůdce)
Z je jeho „dítě“ (přímý následovník)

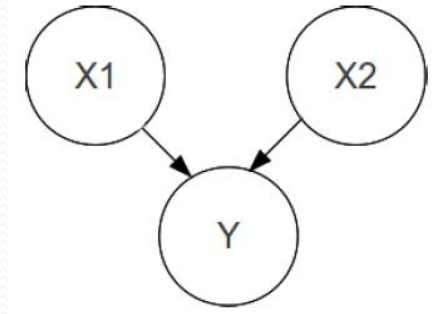
- Sdružené rozdělení pravděpodobnosti vektoru náhodných veličin

$$P_B(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P_B(X_i | \pi_i) = \prod_{i=1}^n \theta_{X_i | \pi_i}$$

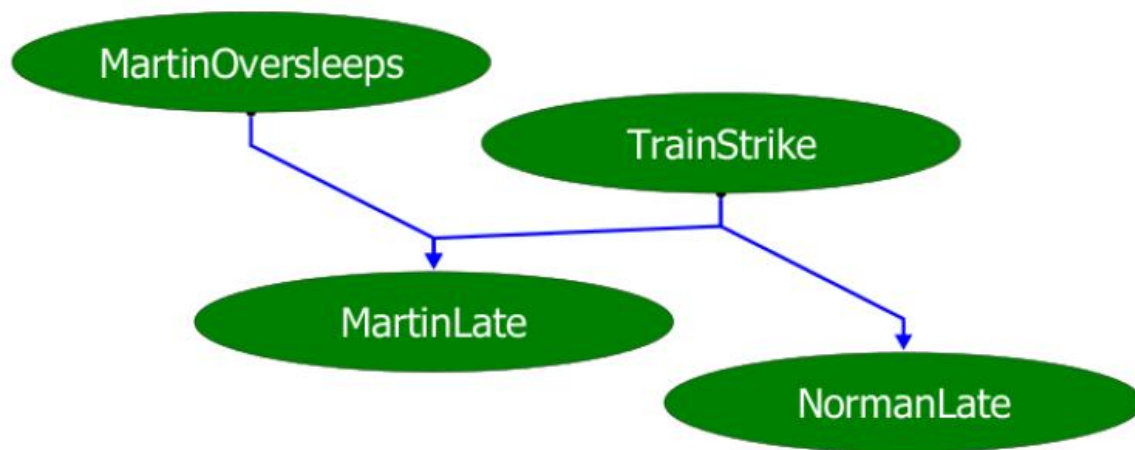
- Síť $B = \langle G, \Theta \rangle$, kde G je DAG s uzly X_1, X_2, \dots, X_n a Θ je množina parametrů $\theta_{X_i | \pi_i}$ pro všechny náhodné veličiny X_i .
- Podmínné rozdělení pravděpodobnosti každého uzlu X_i je závislé na množině jeho rodičů π_i .
- Diskrétní případ: podmíněné rozdělení pravděpodobnosti (CPD) náhodné veličiny Z obsahuje parametry $\theta_{Z|Y}$ uspořádané v pravděpodobnostní tabulce (NPT, CPT):

Y	$P(Z = z_1 Y)$	$P(Z = z_2 Y)$	$P(Z = z_3 Y)$
y_1	$P(Z = z_1 Y = y_1)$	$P(Z = z_2 Y = y_1)$	$P(Z = z_3 Y = y_1)$
y_2	$P(Z = z_1 Y = y_2)$	$P(Z = z_2 Y = y_2)$	$P(Z = z_3 Y = y_2)$

- Zpětné vyhodnocování podmíněných pravděpodobností při pozorování skutečné hodnoty některé náhodné veličiny.
- Bayesovská síť obsahuje skryté i pozorované uzly



- Máme dané $P(X_1)$, $P(X_2)$ a $P(Y|X_1, X_2)$.
- Při pozorování skutečné hodnoty $Y=y$ se změní $P(X_1)$ na $P(X_1|y)$ a $P(X_2)$ na $P(X_2|y)$.
- Pomocí Bayesova vzorce
$$P(X_1|y) = \frac{P(y|X_1)P(X_1)}{P(y)}$$



MartinOverleeps		
Yes	No	bar charts
0,4	0,6	

TrainStrike		
Yes	No	bar charts
0,1	0,9	

Parent Node(s)		MartinLate		
MartinOverleeps	TrainStrike	Yes	No	bar charts
Yes	Yes	0,8	0,2	
	No	0,6	0,4	
No	Yes	0,6	0,4	
	No	0,3	0,7	

Parent Node(s)		NormanLate		
TrainStrike		Yes	No	bar charts
Yes		0,8	0,2	
No		0,1	0,9	

O = Martin oversleeps	M = Martin late
T = Train strike	N = Norman late

Node Name	State 0	State 1
Martin late	Yes	No
	0,4460	0,5540
Martin oversleeps	Yes	No
	0,4000	0,6000
Norman late	True	False
	0,1700	0,8300
Train strike	Yes	No
	0,1000	0,9000

$$P(N) = P(N|T)P(T) + P(N|\neg T)P(\neg T) = 0.8 * 0.1 + 0.1 * 0.9 = \mathbf{0.17}$$

$$\begin{aligned}
P(M) &= P(M|T, O)P(T)P(O) + P(M|T, \neg O)P(T)P(\neg O) + \\
&\quad + P(M|\neg T, O)P(\neg T)P(O) + P(M|\neg T, \neg O)P(\neg T)P(\neg O) \\
&= 0.8 * 0.1 * 0.4 + 0.6 * 0.1 * 0.6 + 0.6 * 0.9 * 0.4 + 0.3 * 0.9 * 0.6 \\
&= 0.032 + 0.036 + 0.216 + 0.162 = \mathbf{0.446}
\end{aligned}$$

Pravděpodobnost, že bude mít Martin zpoždění je 44,6%.

Pravděpodobnost, že bude mít Norman zpoždění je 17%.

Node Name	State 0	State 1
Martin late	Yes	No
	0,5424	0,4576
Martin oversleeps	Yes	No
	0,4000	0,6000
Norman late	Yes	No
	1,0000	0,0000
Train strike	Yes	No
	0,4706	0,5294

Pozorujeme, že Norman měl zpoždění, tj. $P(N)=1$.

$$P(T|N) = \frac{P(N|T)P(T)}{P(N)} = \frac{0.8 * 0.1}{0.17} = 0.471 \text{ (výrazně více než 0.1)}$$

$$\begin{aligned} \mathbf{P(M)} &= P(M|(T|N), O)P(T|N)P(O) + P(M|(T|N), \neg O)P(T|N)P(\neg O) + \\ &+ P(M|\neg(T|N), O)P(\neg(T|N))P(O) + \\ &+ P(M|\neg(T|N), \neg O)P(\neg(T|N))P(\neg O) \\ &= 0.8 * 0.471 * 0.4 + 0.6 * 0.471 * 0.6 + 0.6 * 0.529 * 0.4 \\ &+ 0.3 * 0.529 * 0.6 = 0.151 + 0.169 + 0.127 + 0.095 = \mathbf{0.542} \end{aligned}$$

Pokud má zpoždění Norman, roste pravděpodobnost, že bude mít zpoždění i Martin.

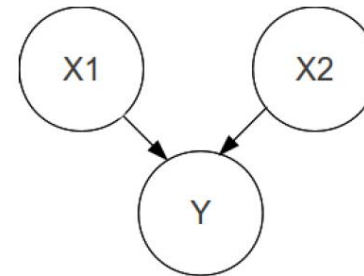
- Objektově orientované BNs.
- Určování NPTs pro rozsáhlé BNs.
- Bayesovské sítě učící se z dat.
- Dynamické bayesovské sítě.
- Hybridní bayesovské sítě.

- V klasické BN je množina uzlů a vazba mezi nimi fixní a použitelná jen pro daný případ.
- Programátorské řešení přes abstraktní datové typy a objektově orientované programování.
- ADT = implementačně nezávislá specifikace struktury dat s operacemi na této struktuře, např. zásobník.
- Objektově orientované programování = organizování ADT.
- Základním prvkem OOBNs je objekt.
- Nejzákladnější objekt je náhodná veličina (jako u klasické BN).
- Např. objekt auto má atributy barva, majitel, motor,.... Barva je tedy základní objekt (nabývá konečného množství hodnot), ale majitel má své další atributy.

- Komplexní objekt je definovaný přiřazením stochastických funkcí každému jeho atributu a propojením atributů pomocí BN, tj. vytvoření pravděpodobnostního modelu.
- „Bayesovská síť bayesovských sítí“.
- Stochastická funkce pro všechny hodnoty vstupů přiřadí rozdělení pravděpodobnosti hodnot vstupů.
- Třída objektů = množina objektů popsaných stejným pravděpodobnostním modelem.
- Třídy umožňují sestavit obecnou strukturu, kterou je možné použít při řešení různých problémů.
- Používají se při modelování dynamických BNs.

- U rozsáhlých sítí je problém s ručním plněním NPTs, především pro uzly s mnoha možnostmi (časová náročnost, riziko překlepu).
- Uspořádané uzly reprezentují kvalitativní proměnné, chápeme je jako diskretizaci intervalu $[0, 1]$.
- Např. $X_1, X_2, Y \in \{\text{velmi nízká, nízká, vysoká, velmi vysoká}\}$

X_1 : kvalita testovacích nástrojů
 X_2 : kvalita testovacích postupů
 Y : efektivnost testu



=> NPT pro Y bude mít přes 125 položek.

- Možnost zjednodušení pomocí Noisy-OR nebo Noisy-MAX modelů.

- Jen pro diskrétní náhodné veličiny.
- Noisy-OR pro binární proměnné, Noisy-MAX je rozšířený model pro vícehodnotové proměnné.
- Předpokládá se, že „rodiče“ jsou navzájem nezávislí při ovlivňování „dítěte“.
- Logaritmická redukce rozsahu NPTs.
- Deterministický OR model předpokládá, že $X_i = \text{True} \Rightarrow Y = \text{True}$ nezávisle na hodnotách ostatních rodičů.
- U Noisy-OR modelu tato implikace nemusí nastat:

$$P(Y_i = \text{True} | X_i = \text{True}) = p_i$$

$$P(Y_i = \text{True} | X_i = \text{False}) = 0$$

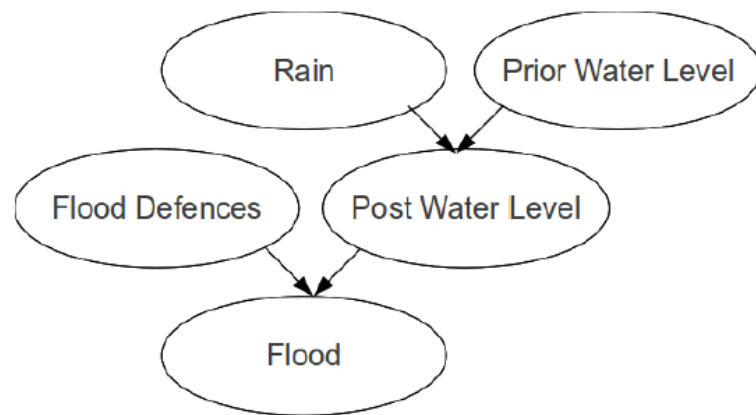
- $\forall i : X_i = \text{False} \Rightarrow Y = \text{False}$ reálně nemusí platit...Leaky Noisy-OR.

- Možnost naučit se kvantitativní, ale i kvalitativní část.
- Naučená pravděpodobnost vyjadřuje relativní četnost (ne subjektivní pravděpodobnost).
- Učení využívá bayesovský princip: $\pi(\theta|y) \propto f(y|\theta)\pi(\theta)$
- Neinformativně apriorní rozdělení – jakákoli pravděpodobnost je stejně možná.
- Informativně apriorní rozdělení – např. Beta rozdělení.
- Vytvoření rozšířené BN s uzly reprezentující naše přesvědčení o relativní četnosti (rodič daného uzlu).

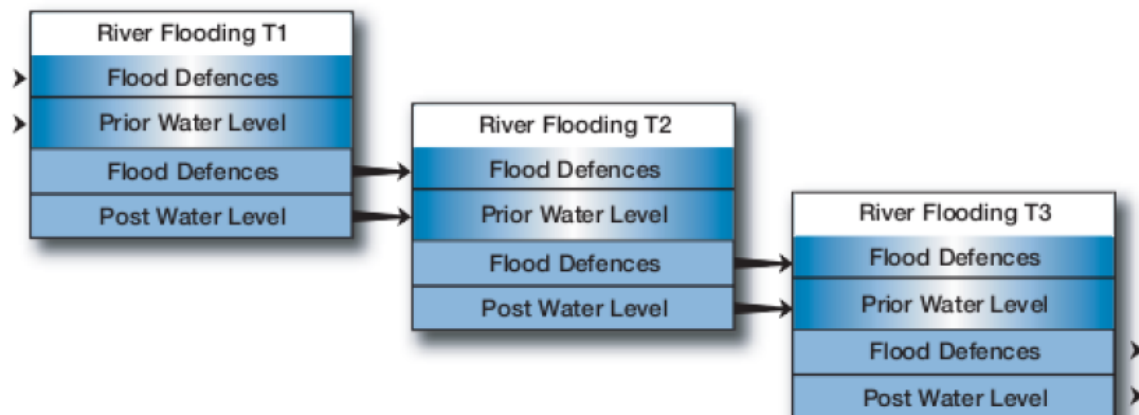
- Rozšíření BNs pro modelování rozdělení pravděpodobnosti nekonečné posloupnosti náhodných veličin Z_1, Z_2, \dots
- Nejčastěji jde o časovou řadu (např. při rozpoznávání hlasu) nebo posloupnost znaků (např. proteiny).
- Stochastické procesy s diskrétním časem.
- Modelace dynamických systémů, samotná BN se v čase nemění.
- DBN je definovaná jako dvojice (B_1, B_{\rightarrow}) , kde B_1 je BN definující apriorní $P(Z_1)$ a B_{\rightarrow} je 2TBN, která definuje $P(Z_t|Z_{t-1})$ jako

$$P(Z_t|Z_{t-1}) = \prod_{i=1}^N P(Z_t^i | \text{Pa}(Z_t^i)),$$

- kde Z_t^i je i -tý bod v čase t a $\text{Pa}(Z_t^i)$ jsou jeho rodiče.



Statická BN, nutnost replikovat celou síť při každé iteraci



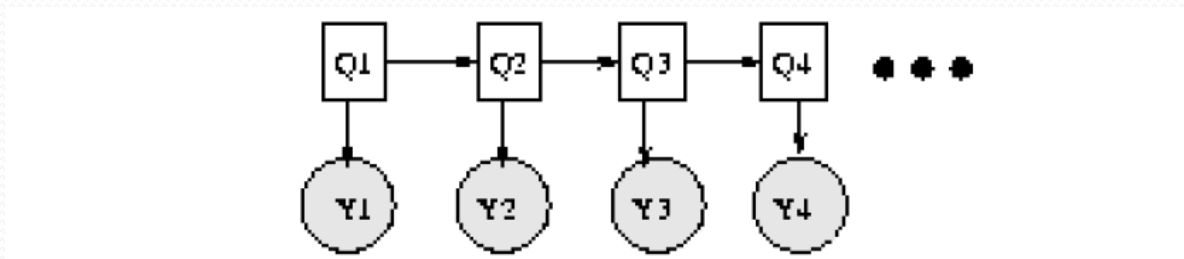
Obr.: Dynamická BN, zdroj: Fenton a Neil

Hybridní Bayesovské sítě:

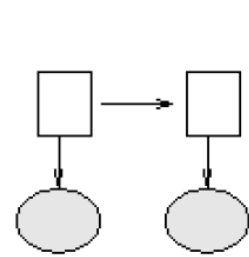
- Obsahují diskrétní i spojité náhodné veličiny.
- Nejčastěji je používán podmíněný lineární Gaussův model: dítě má normální rozdělení se střední hodnotou závislou na spojitéch i diskrétních rodičích a s rozptylem nezávislým na spojitéch rodičích.
- Není možné, aby diskrétní dítě mělo spojité rodiče.
- Je ale možná diskretizace.
- Příklad: cash-flow společnosti (spojitý bod) ovlivní pravděpodobnost kapitálové investice (diskrétní bod).

Skryté Markovské modely (HMMs):

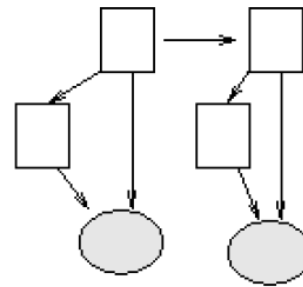
- Nejjednodušší typ dynamických a zároveň hybridních BNs.
- Pro každý časový okamžik má jeden diskrétní skrytý bod a jeden diskrétní nebo spojitý pozorovaný bod.



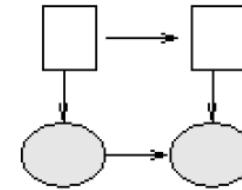
- Kruh je spojitý bod, čtverec bod diskrétní. Bílý bod je skrytý, šedý bod pozorovaný.



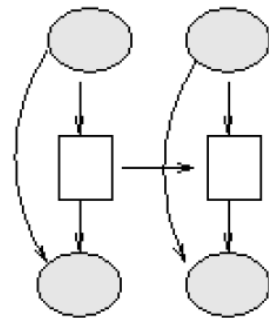
HMM with Gaussian output



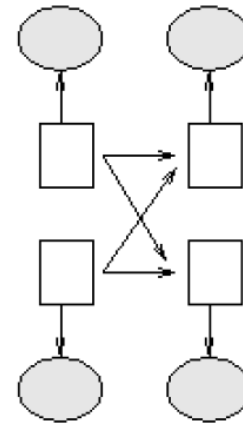
HMM with mixture of Gaussians output



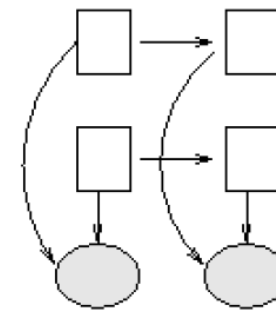
Auto Regressive HMM



Input-output HMM



Coupled HMM

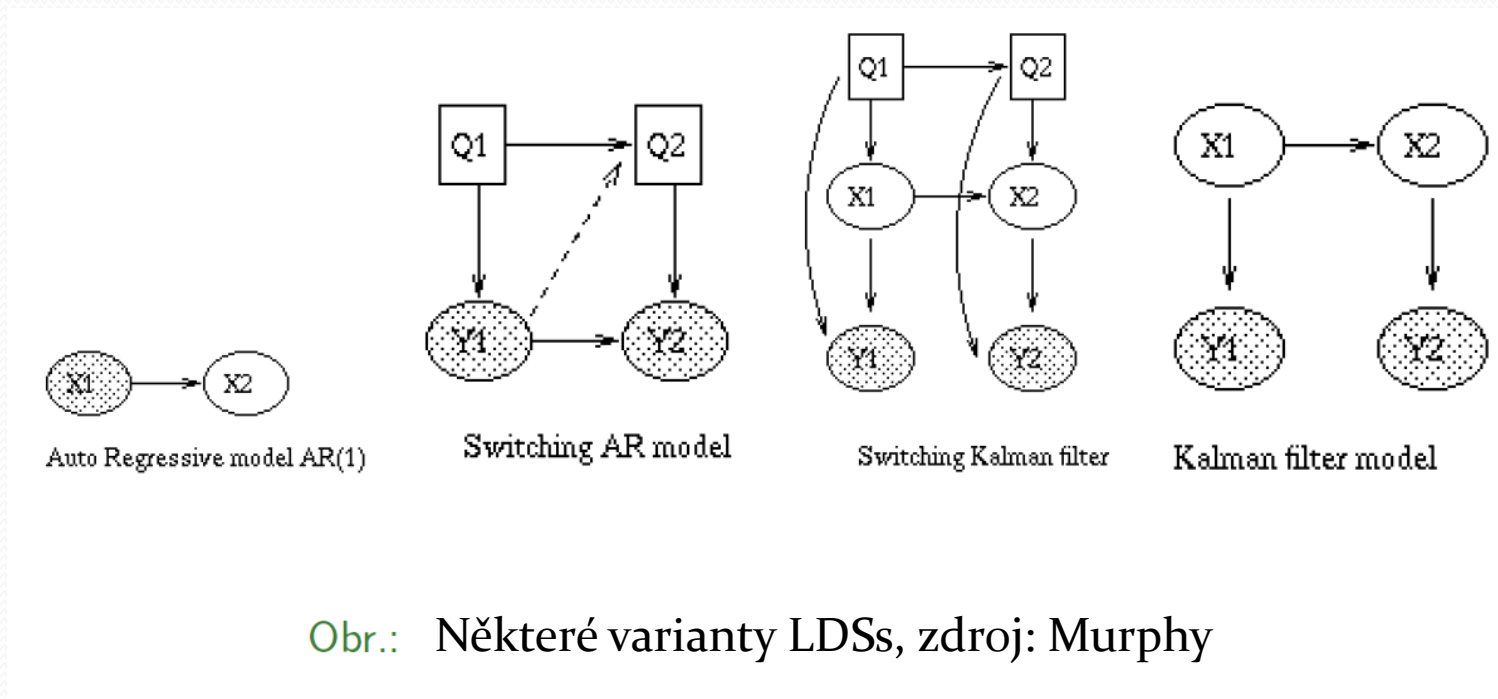


Factorial HMM

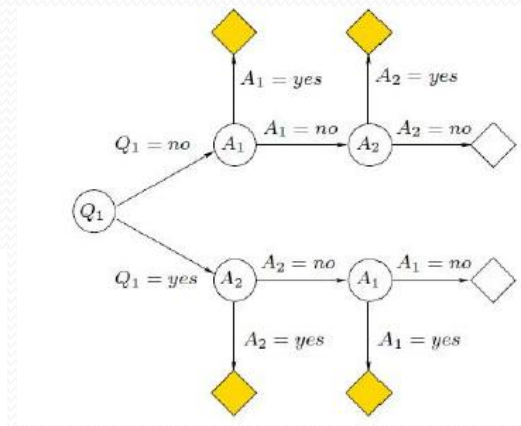
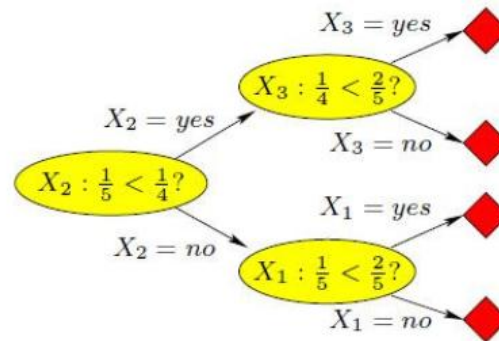
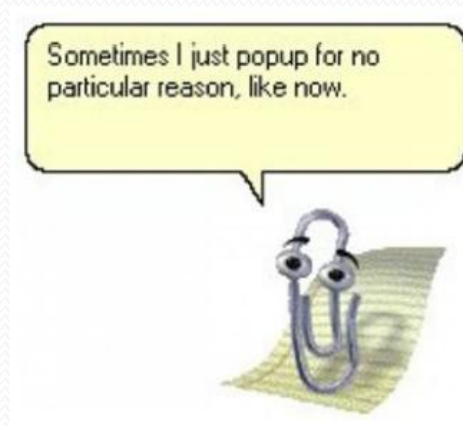
Obr.: Některé varianty HMMs, zdroj: Murphy

Lineární dynamické systémy (LDSs):

- Další příklad DBNs; stejná topologie jako HMMs.
- Navíc předpoklad lineárního Gaussova rozdělení všech bodů.



5. Bayesovské sítě - aplikace



- První pracovní aplikace BNs
 - Období 1988 – 1995
 - Diagnostika
 - Lékařství – MUNIN
HUGIN
 - Diagnostika poruch (závad)

MUNIN

- Dokončen jako systém pro podporu rozhodování lékařů.
- Je těžké odhadnout, do jaké míry byl v té době použit. Lékaři byli vůči němu skeptičtí – nevěřili těmto UI aplikacím.

BNs a Microsoft

- Rok 1996 - "Konkurenční výhodou Microsoftu jsou jeho odborné znalosti v Bayesovských sítích."
- Vysoké investice do výzkumu
 - Eric Horvitz
 - David Heckerman
- Oblasti využití
 - uživatelská podpora
 - automatizovaná diagnostika chyb

Uživatelská podpora

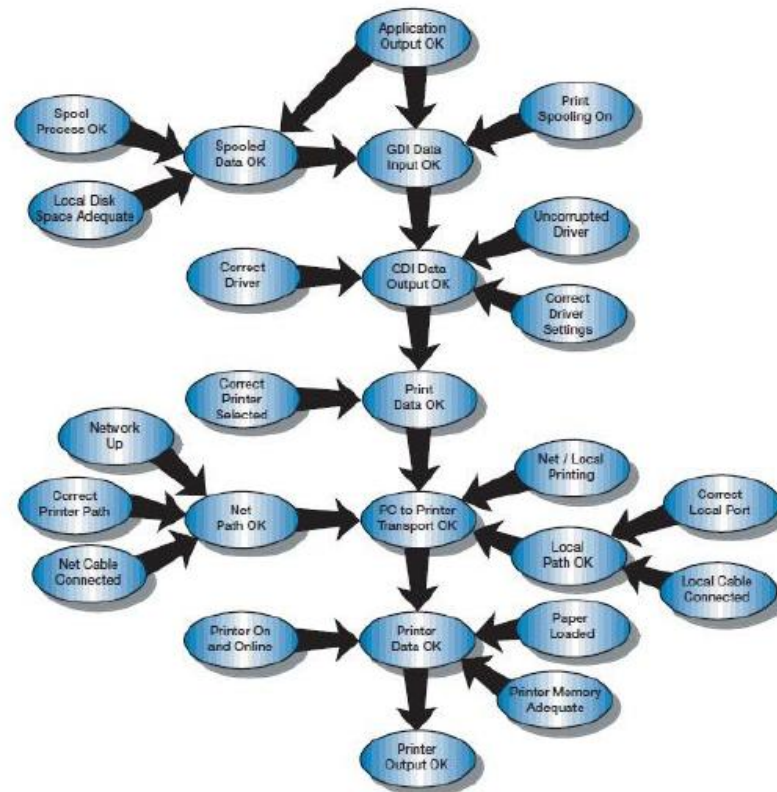
- MS Office – Paperclip
- Jedna z nejslavnějších, ale také nejvíce nepochopených aplikací BNs v Microsoftu.

Sometimes I just popup for no particular reason, like now.



Automatická diagnostika chyb

- Především diagnostický systém selhání tiskáren



Farmaceutické společnosti

- Vysoké využití – možnost vyšší účinnosti a správnosti jejich testů. Snížení nákladů.
- Rok 2006 - FDA (*Food and Drug Administration*) oznámila svou snahu využívat Bayesiánské techniky.
 - Cíl – urychlení procesu schvalování a zvýšení celkové bezpečnosti.

Vývoj softwaru

- Pro nový software je důležité, aby obsahoval co nejméně chyb. Ale i přes mnoho testů, tam stále nějaké chyby budou.
- Cíl – aby uvolněný software obsahoval co nejméně chyb a aby případné chyby měly na koncové uživatele co nejmenší vliv.
- BNs se využívají pro tvorbu předpovědí o zbytkových chybách.
- Důležité především pro bezpečnostní software.

System TRACS

- System pro předpovídání spolehlivosti pozemních vozidel.
- QinetiQ (*Společnost, která poskytuje technické poradenství zákazníkům v celosvětovém letectví, obraně a bezpečnosti trhu*).
- Výběr vozidla pro Ministerstvo obrany (UK)
 - Obtížné, časově náročné a náchylné k chybám
 - Původně: kombinace testování prototypů a modelování analýz – velmi drahé a neuspokojivé
 - Následně vyvinut BNs model.

System TRACS

- Vytvořen na základě šablon bayesovských modelů založených na konkrétních subsystémech architektury dané specifikace vozidla.
- Možnost komunikovat s modelem BNs prostřednictvím jednoduchého dotazníku GUI (*Graphical User Interface*)
- System TRACS šetří čas i peníze – předpovědi jsou přesnější, protože jsou schopné s pevnými daty kombinovat i subjektivní údaje.

Meteorologie

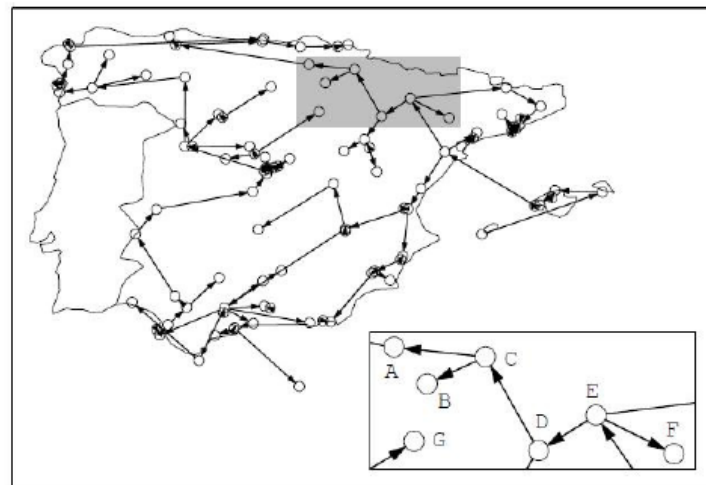
- Místo studie – Pyrenejský poloostrov
 - 100 pozorovacích stanic
 - Vyšší počet dostupných dat – vývoj statistických a data miningových technik



- Zdroj: Cano, R., Sordo, C., Gutiérrez, J.M.: Applications of Bayesian Networks in Meteorology, 2004

Meteorologie

- Proč používat BNs?
 - Jsou schopny podat kompaktní a srozumitelné znázornění pravděpodobnostní informace.
 - Jsou schopny pracovat s velkým počtem proměnných.
 - Uzel = proměnná hrana = závislost



Meteorologie

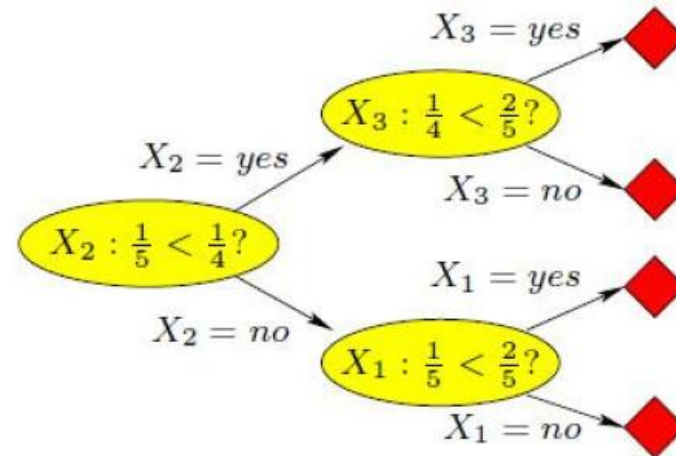
- Využití BNs
 - Doplnování chybějících dat
 - Nutnost úplných údajů
 - Problém s daty s vysokou odchylkou
 - Chybějící údaje musí být nahrazeny při zachování hlavních vlastností původních dat
 - Regresní modely X Gaussovské BNs
 - Možnost využití všech informací

Adaptivní testování

- Adaptivní testy - automaticky se přizpůsobují zjištěné úrovni znalostí zkoušeného.
- Po získání odpovědi na položenou otázku systém automaticky vybere následující otázku s využitím znalosti získané na základě předchozích odpovědí zkoušeného.

Adaptivní testování

- Jednoduchý příklad



Zdroj: Vomlel, J.: Dvě aplikace bayesovských sítí, VŠE Praha

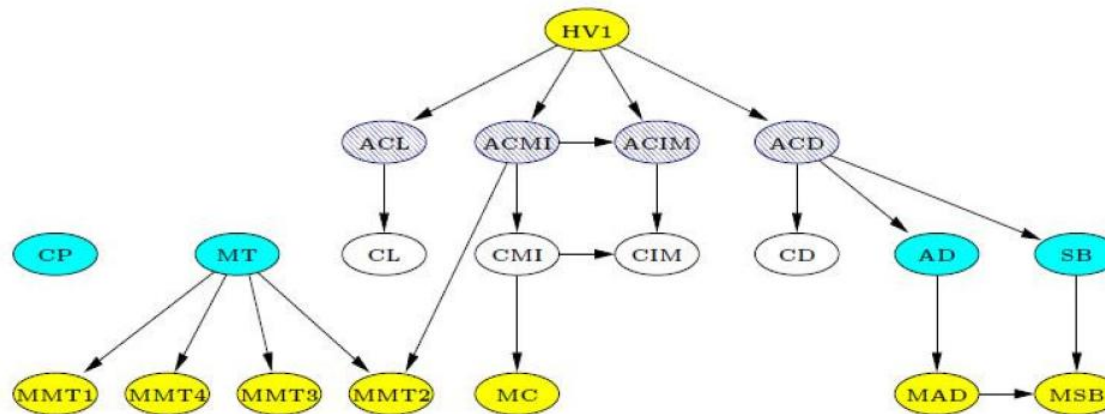
- Nutnost využití počítačů – počítačové adaptivní testování (*computerized adaptive testing*)

Adaptivní testování

- Model složený ze dvou modelů:
 - Model studenta - popisuje vztahy mezi znalostmi a dovednostmi studenta
 - Pozorovací modely - pozorovací model odpovídá jedné otázce
- Testové znalosti a dovednosti: $\mathcal{Y} = \{Y_1, \dots, Y_k\}$
- Soubor otázek: $\mathcal{X} = \{X_1, \dots, X_m\}$
- Model studenta – připravené papírové testy, které řešili studenti prvního ročníku střední školy
 - Testovány čtyři základní dovednosti a čtyři operační dovednosti na zlomcích.
 - Shrnutí výsledků do datových záznamů o jednotlivých studentech.
 - Objeveno několik typických špatných přístupů k řešení některých operací.
 - Pomocí PC-algoritmu vytvořena struktura modelu studenta.
 - Některé vztahy mezi veličinami vysvětleny pomocí latentních veličin.

Adaptivní testování

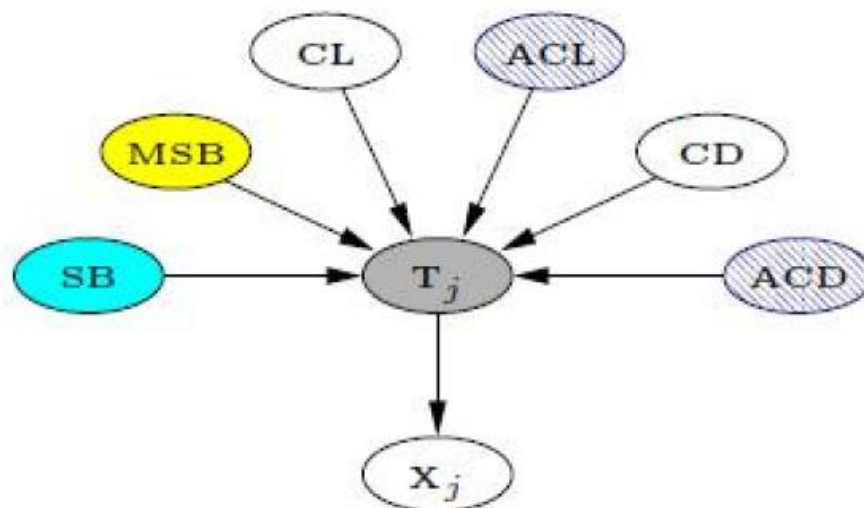
- Výsledný model studenta



Zdroj: Vomel, J.: Dvě aplikace bayesovských sítí, VŠE Praha

Adaptivní testování

- Příklad úlohy: $\frac{1}{3} - \frac{1}{12} = \frac{4}{12} - \frac{1}{12} = \frac{3}{12} = \frac{1}{4}$



Adaptivní testování

- Výsledek a porovnání
 - Klasický přístup pro testování znalostí – teorie odezvy (vícerozměrná teorie odezvy) – od 60.let 20.století
 - Není schopen poskytnout podrobnější diagnostiku.
 - Výhody Bayesovských sítí:
 - BN lépe modeluje přemýšlení studenta.
 - Možnost kratší doby testování.
 - Výsledky ukázaly, že testy využívající BNs mají vysokou úspěšnost.
 - Více než 90% dovedností bylo odhadnuto již po 7 otázkách.
 - V papírových testech bylo potřeba 20 otázek.

Technická diagnostika

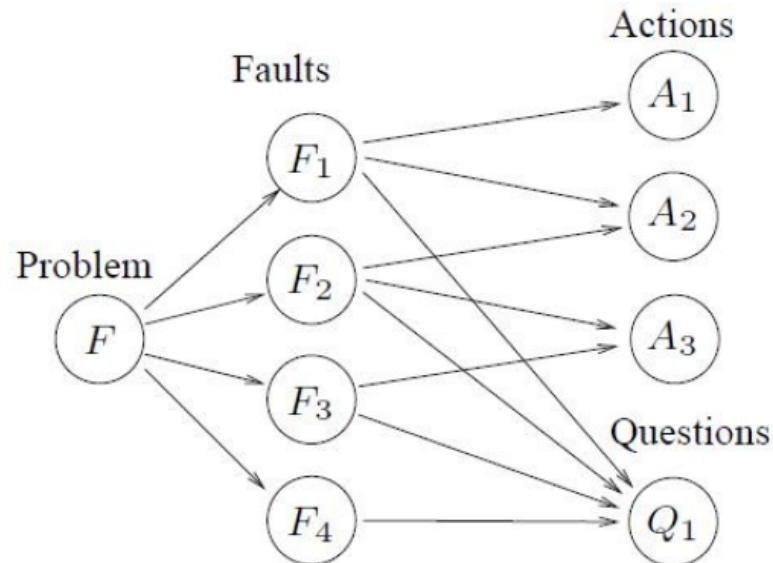
- Diagnostika odhalující technické problémy zařízení.
- Pomocí počítačových systémů může být podstatně urychlena.
- Uvedený příklad má tři typy veličin:
 - Poruchy zařízení: $F \in \mathcal{F}$
 - Akce (opravné kroky): $A \in \mathcal{A}$
 - Pozorování: $Q \in \mathcal{Q}$

Technická diagnostika

- Příklad: Předpokládejme, že tiskárna vytiskne stránku, která je příliš světlá. Tento problém může mít mnoho příčin. Uvažujme zjednodušený model obsahující pouze 4 možné příčiny světlého tisku: F_1 distribuční problém toneru, F_2 vadný toner, F_3 narušený tok dat a F_4 špatné nastavení ovladače. Necht' akce, které mohou vyřešit tento problém jsou: A_1 "vyjměte toner, zatřeste s ním a uložte ho zpět" s cenou $c_1 = 5$, A_2 "zkuste jiný toner" s cenou $c_2 = 15$ a A_3 "vypněte a zase zapněte tiskárnu" s cenou $c_3 = 1$. Pro každou akci expert zadal podmíněnou pravděpodobnost $P(A_i = \text{yes} \mid F_j)$. Například, akce A_2 "zkuste jiný toner" vyřeší distribuční problém toneru a vadný toner s pravděpodobností 0.9, t.j. $P(A_2 = \text{yes} \mid F_i) = 0.9$, $i = 1, 2$, ale nevyřeší špatné nastavení ovladače, t.j. $P(A_2 = \text{yes} \mid F_4) = 0$.

Technická diagnostika

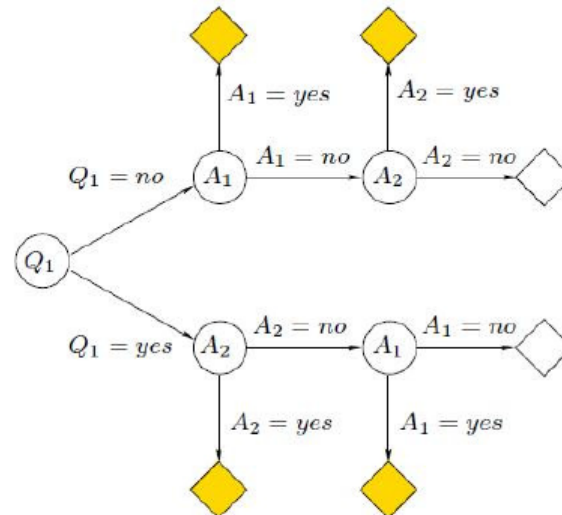
- Můžeme předpokládat dvě věci:
 - Akce a otázky jsou navzájem podmíněně nezávislé je-li známa příčina.
 - Špatnou funkci zařízení způsobuje pouze jedna příčina.



Zdroj: Vomlel, J.: Dvě aplikace bayesovských sítí, VŠE Praha

Technická diagnostika

- Strategie - máme dvě možná řešení problému
 - Problém je vyřešen X Problém zůstane nevyřešen.
 - Definujeme dva typy terminálních uzlů:
 - Úspěšné terminální uzly
 - Neúspěšné terminální uzly



Technická diagnostika

- Mějme hodnotící funkci $CR(e_l)$ a její dvě složky $t(e_l)$ a $c(e_l)$.
- Dostaneme kritérium, které nazýváme očekávaná cena opravy.

$$E_{CR}(s) = \sum_{l \in L(s)} P(e_l) \cdot f(e_l) = \sum_{l \in L(s)} P(e_l) \cdot (t(e_l) + c(e_l))$$

- Základní úlohou pak je najít takovou strategii s , která minimalizuje $E_{CR}(s)$.

Technická diagnostika

- Dosažené výsledky
 - Z důvodu jednoduchosti modelu je možné provádět větší počet výpočtů podmíněných pravděpodobností.
 - Diagnostické strategie byly porovnány s optimálními strategiemi a hodnoty E_{CR} jim byly velmi blízké. Průměrná odchylka byla méně než 2%.

Právní argumentace

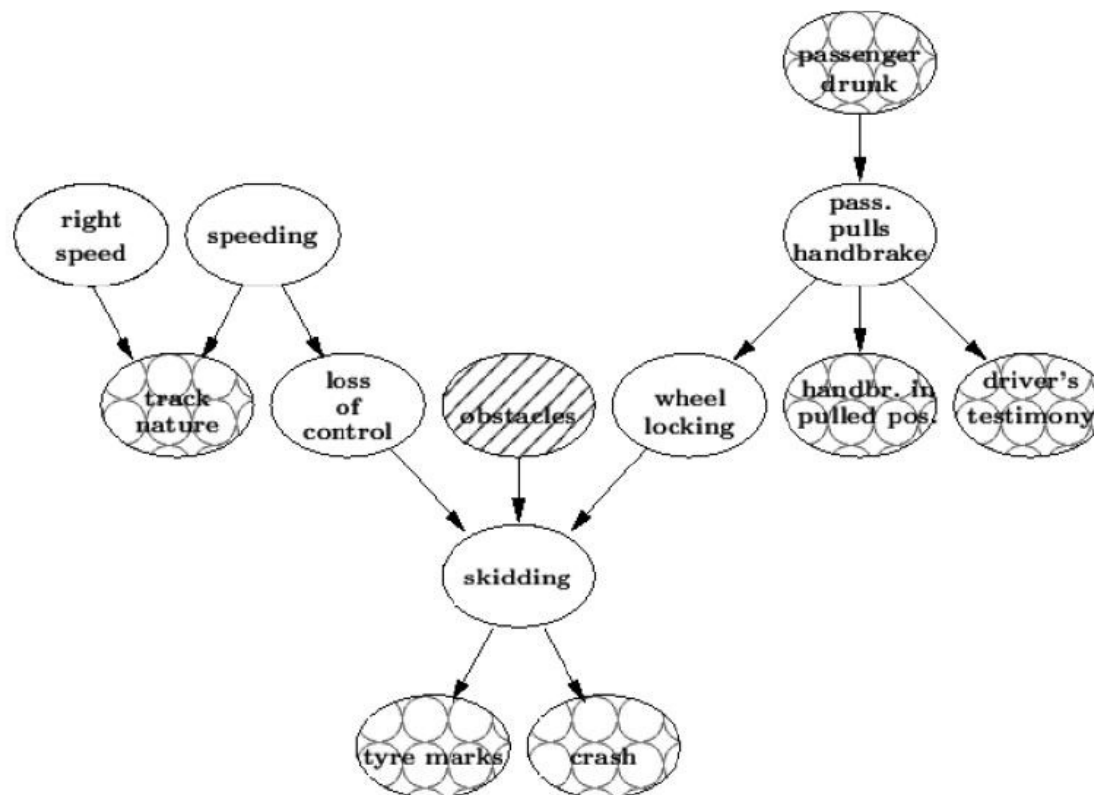
- Uvažujme tento případ – u soudu byly předloženy dva možné scénáře autonehody.
- Fakta: Havárie jednoho automobilu se stala kolem 3:00, kdy auto narazilo do stromu. Policisté, kteří dorazili na místo nehody zjistili následující:
 - Řidička seděla na zemi vedle auta, a uvedla, že „on“ zatáhl ruční brzdu.
 - Těžce zraněný cestující seděl na předním sedadle spolujezdce.
 - Ruční brzda byla v zatažené pozici.
 - Auto projelo správně zatáčku, do chvíle než nabouralo.
 - Na silnici byly stopy brzdění v zatáčce.
 - Byly tam stopy smyku vedoucí k místu nehody.

Právní argumentace

- Co následovalo:
 - Ani jeden si nemohl na nehodu pořádně vzpomenout.
 - Vraceli se z párty a spolujezdec byl opilý.
 - Spolujezdec podal na řidičku žalobu o náhradu škody.
 - Řidička se hájila tím, že ruční brzdu zatáhl spolujezdec.
 - Nizozemské právo: povinností řidiče je, aby měl auto pod kontrolou.
 - Závěr vrchního soudu: vzhledem k pravděpodobnosti s jakou pasažér zatáhl ruční brzdu, nemůže být fakt, že se jedná o nehodu jednoho auta, použit jako důkaz ve prospěch navrhovatele.

Právní argumentace

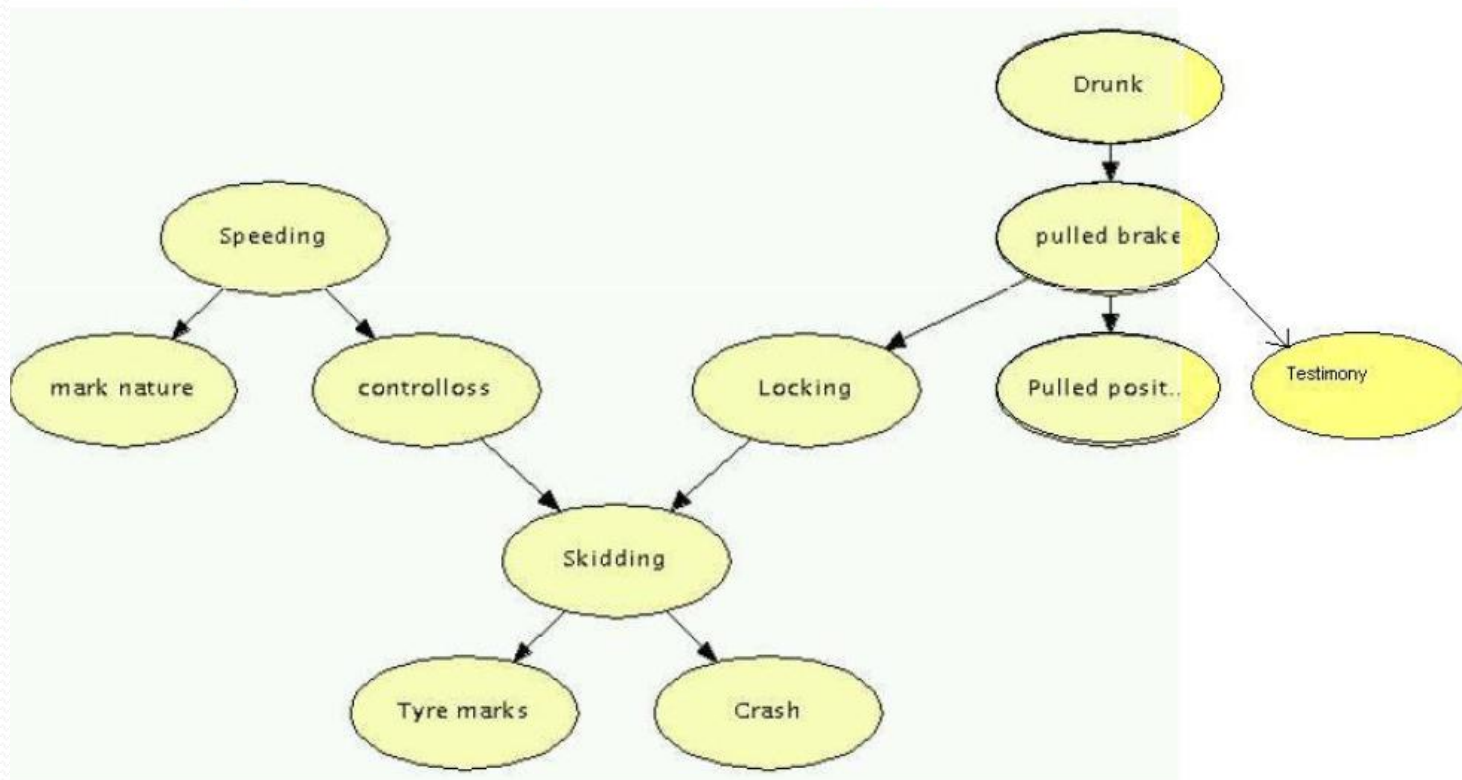
- Kauzální struktura tohoto případu:



Zdroj: P.E.M. Huygen: Use of Bayesian Belief Networks in legal reasoning, 2002, Free University, Amsterdam.

Právní argumentace

- Stejná struktura zakreslena jako BN



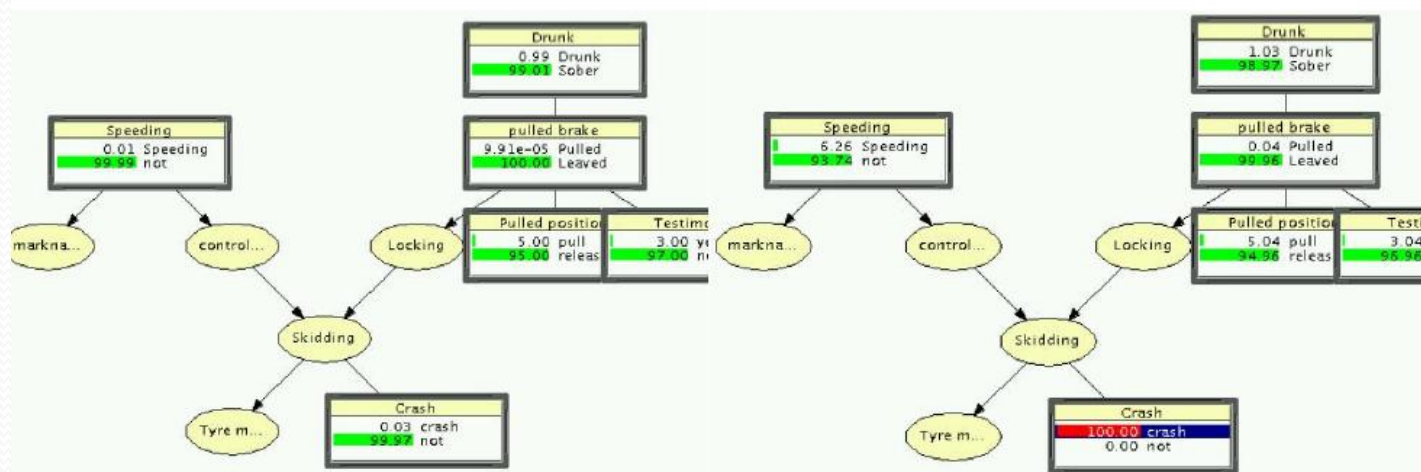
Zdroj:P.E.M. Huygen: Use of Bayesian Belief Networks in legal reasoning, 2002,Free University, Amsterdam.

Právní argumentace

- Počáteční pravděpodobnosti

- Objektivní hodnoty

Zadána skutečnost „crash“



- Zdroj: P.E.M. Huygen: Use of Bayesian Belief Networks in legal reasoning, 2002, Free University, Amsterdam.

Právní argumentace

- Z modelu vyplývá, že je více pravděpodobné, že srážka nastala zatáhnutím ruční brzdy než překročením rychlosti.
- Pokud se zadaly hodnoty výhodné pro žalobce, byly výsledné pravděpodobnosti zavinění srovnatelné.

PATERNITIES SEARCH WITH OBJECT-ORIENTED BAYESIAN NETWORKS

ANDRADE Marina, (PT), FERREIRA, Manuel Alberto M., (PT)

Abstract. Paternity dispute problems are examples of situations in which forensic approach DNA profiles study is a common procedure. To implement this approach an efficient tool the object-oriented Bayesian networks (OOBN). Along this paper are presented the various OOBN adequate to solve the simple paternity dispute and more complex paternity dispute problems with incomplete DNA profiles data about the putative father such as: only putative grandfather information, only putative uncle information, only putative father's uncle information and only simultaneously putative uncle and putative father's uncle information. Here it is exhibited an algebraic treatment, for the simple problem and with those the use of object-oriented Bayesian networks is shown. Then the most complex kind of problems that occur is presented. Although these are not the most common cases there is notice of occurrence at least in Portuguese courts.

2 Simple Paternity Dispute

In a disputed paternity decision problem there are formally two challenging hypotheses (prosecution and defense):

H_p : The true father is the putative father.

vs

H_D : The true father is another individual randomly drawn from the population, and not genetically related with the mother or the putative father.

The court has to decide about the paternity of the child, and so, after Bayes' Law

$$\frac{P(H_p | E)}{P(H_D | E)} = \frac{P(E | H_p)}{P(E | H_D)} \times \frac{P(H_p)}{P(H_D)} \quad (1)$$

with E the vector containing the available evidence, genetic information of the mother (mgf), of the child (cgt) and of the putative father (pgf), being the algebraic approach simple.

It is needed to assess the likelihood function over the hypotheses as to the true father, i.e., to evaluate the likelihood ratio:

$$LR = \frac{P(E | H_p)}{P(E | H_D)} \quad (2)$$

Naturally the court has to answer to the truly paternity of the child. So it has to evaluate the ratio of the hypotheses in dispute. Admitting that $P(H_p) = P(H_D)$ then (1) becomes

$$\frac{P(H_p | E)}{P(H_D | E)} = \frac{P(E | H_p)}{P(E | H_D)} \quad (3)$$

In fact, knowing that the markers are in different chromosomes (*linkage equilibrium*) and assuming random mating (*Hardy-Weinberg equilibrium*) there is independence between and within markers. Thus, it is possible to obtain the *LR* for each marker separately and finally multiply the values to determine the overall likelihood ratio based on the data available for all markers.

To determine algebraically the probability of the triplet E , under the two hypotheses, it is reasonable to consider that before knowing any data on the child it is reasonable to assume that the identity of the true father is independent of the mother's and the putative father's. And supported on that, it is easily seen that it is possible to determine the conditional probability of the child's genotype, given the other two available genotypes. Thus, to determine $P(E | H_p)$ one has only to apply Mendel's laws. But the calculus of $P(E | H_D)$ necessarily demands the knowledge of the population allele frequencies for the considered markers.

If for a certain marker the triplet $E = (mgt, cgt, pfgt)$ is $E = ((A, B); (B, B); (A, B))$, and p_A and p_B are the population allele frequencies then

$$\begin{aligned} P(E | H_p) &= P[(mgt, cgt, pfgt) | (mgt, pfgt)] \\ &= P[cgt | (mgt, pfgt)] \\ &= 0.5 \times 0.5 \end{aligned}$$

and

$$\begin{aligned} P(E | H_D) &= P[(mgt, cgt, pfgt) | (mgt, rgt)] \\ &= P[cgt | (mgt, rgt)] \\ &= 0.5 \times p_B \end{aligned}$$

where rgt assigns the genotype of a random individual of the population, not related to the mother or the putative father.

Therefore,

$$LR = \frac{0.5}{p_B}$$

The considered problem is, as shown, easily algebraically solved. It is used to illustrate the simplicity and the advantages of this tool in more complex situations. Given the freedom of choice for the variables to include in the graphical approach, different representations can be obtained. Some of them simpler than others. To get a 'good' representation is very important to the efficiency and the viability of the computational routines. These are extremely sensible to the organization of the graphical structure. The first step consists on the identification and definition of the nodes for all the variables of interest to the problem.

Then the graphical representation can be obtained. According to Dawid *et al.* (2002), *in order to maximize the efficiency of the calculations as well as the logical clarity of the representation we chose to disaggregate each individual's genotype into its constituent, unobserved, paternally and maternally inherited genes.*

Figure 1 exhibits the OOBN for a paternity case as the discussed above considering a single marker. Each node (instance) in the network represents itself a Bayesian network.

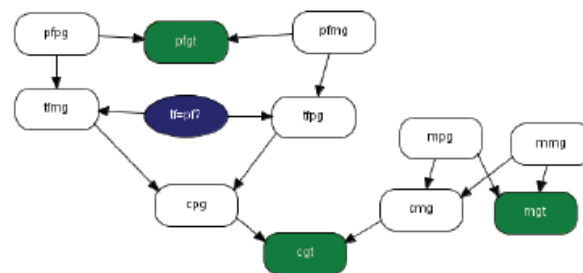


Figure 1: Simple paternity network.

In this simple paternity case instances pfp , pfg , mp and mm are all of class *founder*, a single node *gene*, having for its space of states all the possible alleles that can be presented for the specific case, and the correspondent population gene frequencies. Instances mgt , cgt and $pfgt$ are of class *genotype*, an unordered pair of alleles inherited from paternal, pg , and maternal, mg , genes, here represented by $gmin := \min\{pg, mg\}$ and $gmax := \max\{pg, mg\}$, where pg and mg are input nodes identical to the *gene* node of founder. Instances tfm and $tfpg$ are of class *whom*, describing the true father's allele origin. If $tf=pf?$ has true for value then the true father's allele, tfp , will be identical with the putative father's, pfp , otherwise the true father's allele is randomly chosen from another man in the population. And cpg and cmg instances are of class *inherit*, modelling the Mendel's inheritance in which the child's allele is chosen at random from the two parents, pg and mg , here as the sequence of the observed outcome of a fair coin toss.

For illustration according to Dawid *et al.* (2002), the data for marker FES are child genotype $cgt = \{B, B\}$, mother's genotype $mgt = \{A, B\}$ and putative father's genotype = $\{A, B\}$. The population allele frequencies are $p_A = 0.28425$ and $p_B = 0.25942$.

After specifying the network, put it to run and then insert the evidence. Considering equal prior probabilities for the query node representing the hypotheses, the likelihood is got after inserting the evidence. The likelihood ratio, based on the data for this marker, is obtained from the marginal posterior distribution of the query node. Thus, $P(tf = pf? := true | E) = 0.6584$ and $P(tf = pf? := false | E) = 0.3416$, and $LR = 1.9274$, being these results in agreement with the algebraic approach (note that $0.5/0.25942 = 1.9274$).

3 Paternities search in more uncommon situations

When the data E are not in the form $(mgt, cgt, pfgt)$ it is not possible to determine in algebraic form the likelihood function for the various hypotheses, i.e. to determine the weight of the genetic

connection of the child with the putative father ancestor(s). The use of Bayesian networks allow to overcome these problems. These networks are a good tool to compute the likelihood functions. Forwarding and backwarding the information a measure of the “strength” of the information available in each case is obtained.

In the sequence the networks for the uncommon cases described in the introduction are presented each one together with a numerical example.

The data considered are the same for the whole cases and are in Table 1 where five different markers are considered and the respective genotypes for the mother, the child, the grandfather, the uncle and the grandfather brother, where * indicates rare alleles, and (a) signs alleles considered as good discriminate markers, with more than 10 alleles in each marker.

Table 1:

Marker	<i>mgt</i>	<i>cgt</i>	<i>gfgt</i>	<i>ungf</i>	<i>gfbgt</i>
D3S1358	16, 18	13*, 16	13*, 17	13*, 16	13*, 15
VWA	16, 17	13*, 16	13*, 16	16, 18	13*, 15
D16S539	11, 12	12, 12	9, 12	10, 12	12, 13
D8S1179	12, 13	13, 17*	14, 17*	14, 15	12, 17*
D21S11(a)	29, 31.2	29, 31.2	29, 31.2	28, 31.2	29, 30

Genetic profiles

In Table 2 the respective allelic frequencies are presented: p_i is the i allele frequency in the population.

Marker	Frequencies				
D3S1358	p_{13}	p_{15}	p_{16}	p_{17}	p_{18}
	0.0032	0.2611	0.2477	0.2065	0.1606
VWA	p_{13}	p_{15}	p_{16}	p_{17}	p_{18}
	0.0023	0.1216	0.2300	0.2649	0.1859
D16S539	p_9	p_{10}	p_{11}	p_{12}	p_{13}
	0.1431	0.0545	0.3009	0.2876	0.1654
D8S1179	p_{12}	p_{13}	p_{14}	p_{15}	p_{17}
	0.1351	0.3028	0.2178	0.1223	0.0031
D21S11(a)	p_{28}	p_{29}	p_{30}	$p_{31.2}$	
	0.1674	0.2136	0.2437	0.1138	

Table 2: Allele frequencies

The allelic frequencies used were collected in www.uni-duesseldorf.de/WWW/MedFak/Serology/dna.htm for Portugal (Azores and Madeira archipelagos not included).

4 Only putative grandfather information

Bayesian networks for more complex problems can be built out of the same fundamental local modules that we have already described for the simple paternity dispute problem, Dawid et al. (2002).

The object-oriented Bayesian network for the “only putative grandfather information” case is shown in Figure 2. Note, for example, the node *gfgt* (grandfather genotype) and the respective connections with the other nodes.

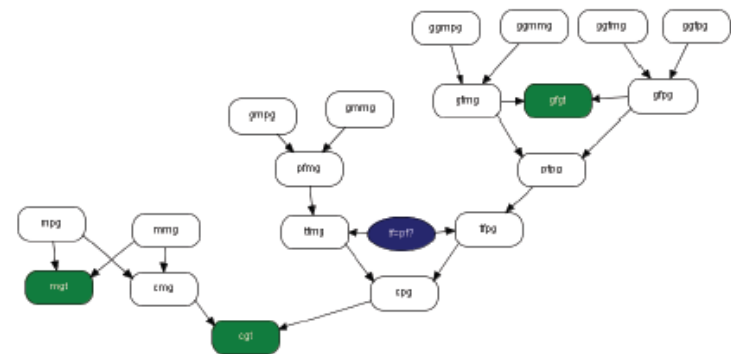


Figure 2: Only putative grandfather network

The results obtained are in Table 3. In the last column Rescaled – corrected so that the sum of the entries is equal to 1 – is presented the result for the 5 markers. Since the markers are independent the final result is obtained by multiplying the result obtained for each marker.

	D3S1358	VWA	D16S539	D8S1179	D21S11	Rescaled
$P(H_p E)$	0.9874	0.9909	0.5779	0.9878	0.6255	0.999999
$P(H_d E)$	0.0126	0.0091	0.4221	0.0122	0.3745	6.33E-07

Table 3: Analysis results with only putative grandfather information

5 Only putative uncle information

The object-oriented Bayesian network for the “only putative uncle information” case is shown in Figure 3.

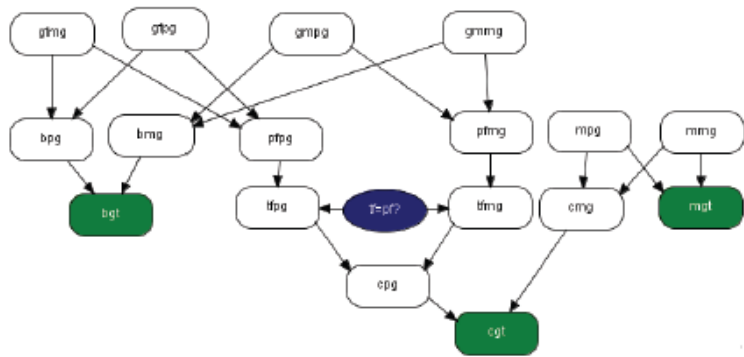


Figure 3: Only putative uncle network

The results obtained are in Table 4 following the same methodology as in section 5.

	D3S1358	VWA	D16S539	D8S1179	D21S11	Rescaled
$P(H_1 E)$	0.9874	0.3333	0.5779	0.3333	0.5582	0.97133
$P(H_2 E)$	0.0126	0.6667	0.4221	0.6667	0.4418	0.02867

Table 4: Analysis results with only putative uncle information

6 Only putative father 's uncle information

In Figure 4 the object-oriented Bayesian network for the "only putative father 's uncle information" case is shown.

It is a network more complex than the former ones owing to the further parentage relationship considered, that implies more complex genetic connections.

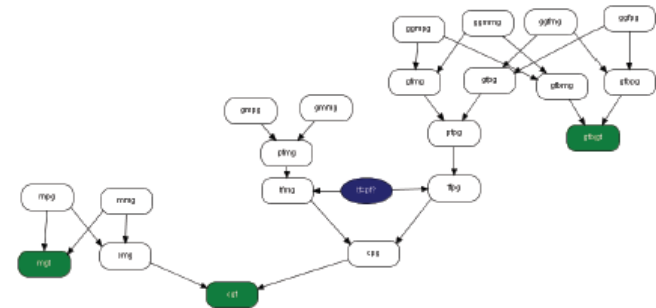


Figure 4: Only putative father 's uncle network

The results obtained are in Table 5.

	D3S1358	VWA	D16S539	D8S1179	D21S11	Rescaled
$P(H_1 E)$	0.9755	0.9822	0.5423	0.9762	0.5309	0.999992
$P(H_2 E)$	0.0245	0.0178	0.4577	0.0238	0.4691	8.28E-06

Table 5: Analysis results with only putative father 's uncle information

7 Only simultaneously putative uncle and putative father 's uncle information

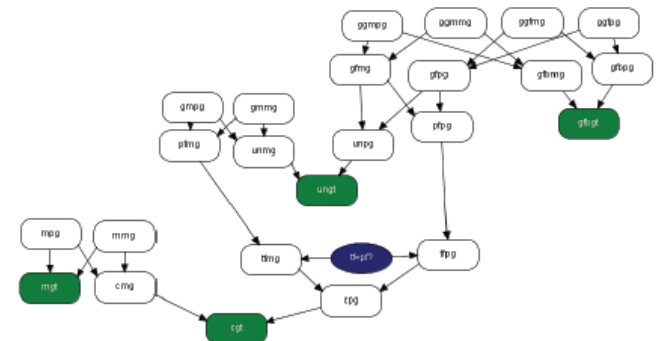


Figure 5: Only simultaneously putative uncle and putative father's uncle

The "only simultaneously putative uncle and putative father's uncle information" case network is the last one presented (in Figure 5) and the results are presented in Table 6.

	D3S1358	VWA	D16S539	D8S1179	D21S11	Rescaled
$P(H_r E)$	0.9875	0.9650	0.5764	0.9536	0.5707	0.999988
$P(H_u E)$	0.0125	0.0350	0.4236	0.0464	0.4293	1.23E-05

Table 6: Analysis results with only simultaneously putative uncle and putative father's uncle information

8 Conclusions

The paternities search in more uncommon cases demands the calculation of probabilities in the context of numerous and complex successive uses of Bayes Law. This situation is impossible to be treated algebraically. It was shown that the object-oriented Bayesian networks are a very powerful tool, very simple to use, that allows the referred calculations in an efficient way.

The major problem is to build the network taking in account the various and complex connections that may occur in parentage relationships. Then the use of an adequate software as Hugin or SPSS makes easy to apply it in practical cases. In this work Hugin was the chosen.

Inspecting the tables of results one can note that, as expected, rare alleles shared lead to greater probabilities of true paternity. On the contrary, more frequent alleles shared lead to lesser probabilities.

With the particular data used the final probabilities for true paternity were in general great.

References

- ABRANTES, D., PONTES, M. L., PINHEIRO, M. F., ANDRADE, M. and FERREIRA, M. A. M.: *Towards a systematic probabilistic evaluation of parentage casework in forensic genetics: A modest attempt to define a general standardized approach to simple and complex cases*. Forensic Science International: Genetics Supplement Series 1, pp. 635-637, 2008.
- ANDRADE, M.: *A Estatística Bayesiana na Identificação Forense – análise e avaliação de vestígios de DNA com redes Bayesianas*. PhD Thesis, ISCTE, Lisboa, 2007.
- ANDRADE, M.: *A Note on Foundations of Probability*. Journal of Mathematics and Technology, vol. 1 (1), pp 96-98, 2010.
- ANDRADE, M., FERREIRA, M. A. M. and FILIPE, J. A.: *Evidence evaluation in DNA mixture traces*. Journal of Mathematics, Statistics and Allied Fields (Scientific Journals International-Published online), vol. 2 (2), 2008.
- ANDRADE, M., FERREIRA, M. A. M., FILIPE, J. A. and COELHO, M.: *Paternity dispute: is it important to be conservative?*. Aplimat – Journal of Applied Mathematics, vol. 1 (2), 2008.

- ANDRADE, M. and FERREIRA, M. A. M.: *Bayesian networks in forensic identification problems*. Aplimat - Journal of Applied Mathematics, vol. 2 (3), pp. 13-30, 2009.
- ANDRADE, M. and FERREIRA, M. A. M.: *Civil Identification Problems with Bayesian Networks Using Official DNA Databases*. Aplimat-Journal of Applied Mathematics, vol. 3 (3), pp. 155-162, 2010.
- ANDRADE, M. e FERREIRA, M. A. M.: *Solving civil identification cases with DNA profiles databases using Bayesian networks*. Journal of Mathematics and Technology, 1(2), pp. 37-40, 2010.
- ANDRADE, M. e FERREIRA, M. A. M.: *Evaluation of Paternities with less usual Data using Bayesian Networks*. IEEE Xplore (BMEI 2010 IEEE Catalog Number CFP1093D-PRT, ISBN: 978-1-4244-6496-8), 2010.
- ANDRADE, M., FERREIRA, M. A. M., ABRANTES, D., PONTES, M. L. e PINHEIRO, M. F.: *Object-oriented Bayesian Networks in the evaluation of paternities in less usual environments*. Journal of Mathematics and Technology, 1(1), pp. 161-164, 2010.
- DAWID, A. P., MORTERA, J., PASCALI, V. L. and van BOXEL, D. W.: *Probabilistic expert systems for forensic inference from genetic markers*. Scandinavian Journal of Statistics vol. 29, pp. 577-595, 2002.
- FERREIRA, M. A. M. and ANDRADE, M.: *A note on Dawnie Wolfe Steadman, Bradley J. Adams, and Lyle W. Konigsberg, Statistical Basis for Positive Identification in Forensic Anthropology. American Journal of Physical Anthropology 131: 15-26 (2006)*. International Journal of Academic Research, vol. 1 (2), pp. 23-26, 2009.
- LAURITZEN, S. L.: *Bayesian networks for forensic identification Problems*. Tutorial 19th Conference on Uncertainty in Artificial Intelligence, Mexico, 2003.

6. Výpočet pojistného v životním pojištění

Dělení životního pojištění

- pojištění riziková - jedná se o pojištění, u kterých se předem neví, zda dojde k pojistné události a následně výplatě pojistného plnění.
- pojištění rezervotvorná – v případě tohoto druhu pojištění musí pojišťovna počítat s výplatou pojistného plnění vždy. Ať dříve v případě smrti nebo později při dožití konce pojištění.

Pojištění pro případ úmrtí

- rizikové pojištění
- výplata pouze v případě úmrtí

Smíšené životní pojištění

- pojištění pro případ smrti nebo dožití
- pojistná částka je vyplacena vždy
- může být kladen větší důraz na jedno z rizik a podle toho mohou být nastaveny pojistné částky pro každé riziko zvlášť
- může být sjednáno na zvyšující se pojistnou částku při dožití se v průběhu pojistné doby, nebo může být sjednáno pro dvojici osob apod.
- varianty smíšeného pojištění, kdy oprávněným osobám je vyplaceno plnění v případě úmrtí pojištěného a poté ještě jedno plnění v době smlouveného konce pojištění
- bývá rozšiřováno o krytí dalších pojistných rizik neživotního charakteru (úraz, invalidita, vážná nemoc apod.)

Důchodové pojištění

- kryje pouze riziko dožití
- výplata jednorázově nebo pravidelné důchodové splátky
- varianty
 - **Základní doživotní důchod** – vyplácen od data nároku na starobní důchod
 - **Dočasný důchod** – výplata v případě trvalé invalidity

Finanční matematika v pojištění

a) Spojité úročení

b) Hodnota důchodů

- systém opakujících se plateb, jejichž výše zůstává v čase stejná nebo se mění dle určitého schématu

- ocenění důchodu vztažením všech jeho plateb ke stejnému časovému okamžiku s použitím úrokové míry

Pojistné

= úplata za poskytnutou pojistnou ochranu

- a) Netto pojistné
- b) Brutto pojistné

Netto pojistné

- označované taky jako ryzí, představuje hodnotu veškerých závazků pojišťovny, které připadají na všechny klienty vzhledem k jejich předpokládaným pojistným událostem.

Při stanovování výše netto pojistného se přihlíží zejména k následujícím faktorům:

- výši sjednané pojistné částky
- ohodnocení rizika
- výši technické úrokové míry

Faktory ovlivňují výši netto pojistného

- výši sjednané pojistné částky

Platí zde vztah přímé úměry mezi pojistnou částkou a pojistným, tzn. čím vyšší pojistná částka, tím vyšší pojistné. Předcházení jejího znehodnocování by mělo zabránit sjednání dynamizace pojistné smlouvy.

- ohodnocení rizika – riziko úmrtí nebo dožití

Výše pojistného se odvíjí především od pohlaví a věku pojištěného. Do ceny pojištění vstupuje taky zohlednění zdravotního stavu, výše pojistné částky a délka trvání pojištění. Ohodnocování rizika slouží k určení velikosti netto pojistného v životním pojištění, v některých případech může vést k odmítnutí sjednání životního pojištění ze strany pojišťovny. K určení pravděpodobnosti dožití určitého věku a pravděpodobnosti úmrtí před dosažením určitého věku slouží pojišťovnám úmrtnostní tabulky.

Faktory ovlivňují výši netto pojistného

- výše technické úrokové míry

V případě technické úrokové míry se jedná o garantovanou výnosnost pro klienta, se kterou musí pojistní matematici počítat, aby se nestala situace, že v případě pojišťovny, která bude počítat s vyšším zhodnocením rezerv se dostane do nerovnovážné situace v důsledku toho, že od klientů vybírá nižší pojistné. V současné době je horní limita výše technické úrokové míry regulovaná státem vyhláškou č. 303/2004 Sb., kterou se provádí některá ustanovení zákona o pojišťovnictví, ve znění pozdějších předpisů.

- Aktuálně 2,4 % p.a.

Úmrtnostní tabulka

Skládá se ze sloupců a řádků. Ve sloupcích jsou uvedeny jednotlivé veličiny (počet osob, počet žijících osob v daném věku,...). Řádky představují hodnoty veličin uvedených ve sloupcích pro konkrétní věk.

Úmrtnostní tabulka

x	l_x	d_x	p_x	q_x	D_x	C_x	N_x	M_x	S_x	R_x	L_x	T_x	e_x
3	99554	25	0,9997488800	0,0002511200	93 811,96	23,10	3 609 095,24	23 045,38	103 437 600,62	1 580 906,99	99 541	7 208 709	72,41
4	99529	14	0,9998593375	0,0001406625	91 949,41	12,68	3 515 283,28	23 022,29	99 828 505,38	1 557 861,61	99 522	7 109 168	71,43
5	99515	13	0,9998693664	0,0001306336	90 133,80	11,54	3 423 333,87	23 009,61	96 313 222,10	1 534 839,32	99 509	7 009 646	70,44
6	99502	11	0,9998894495	0,0001105505	88 354,93	9,58	3 333 200,07	22 998,06	92 889 888,23	1 511 829,71	99 497	6 910 137	69,45
7	99491	13	0,9998693349	0,0001306651	86 612,90	11,10	3 244 845,14	22 988,49	89 556 688,16	1 488 831,65	99 484	6 810 640	68,45
8	99478	16	0,9998391604	0,0001608396	84 903,52	13,39	3 158 232,24	22 977,39	86 311 843,02	1 465 843,16	99 470	6 711 156	67,46
9	99462	15	0,9998491886	0,0001508114	83 225,35	12,31	3 073 328,72	22 964,00	83 153 610,78	1 442 865,77	99 454	6 611 686	66,47
10	99447	14	0,9998592215	0,0001407785	81 581,18	11,26	2 990 103,37	22 951,70	80 080 282,06	1 419 901,76	99 440	6 512 232	65,48
11	99433	14	0,9998592017	0,0001407983	79 970,29	11,04	2 908 522,19	22 940,44	77 090 178,69	1 396 950,06	99 426	6 412 792	64,49
12	99419	16	0,9998390650	0,0001609350	78 391,20	12,37	2 828 551,91	22 929,40	74 181 656,50	1 374 009,62	99 411	6 313 365	63,50
13	99403	19	0,9998088589	0,0001911411	76 841,75	14,40	2 750 160,70	22 917,03	71 353 104,59	1 351 080,22	99 394	6 213 954	62,51
14	99384	20	0,9997987604	0,0002012396	75 320,65	14,86	2 673 318,95	22 902,63	68 602 943,89	1 328 163,19	99 374	6 114 560	61,52
15	99364	25	0,9997483998	0,0002516002	73 828,92	18,21	2 597 998,30	22 887,77	65 929 624,94	1 305 260,56	99 352	6 015 186	60,54
16	99339	29	0,9997080703	0,0002919297	72 363,08	20,71	2 524 169,38	22 869,56	63 331 626,64	1 282 372,78	99 325	5 915 835	59,55
17	99310	43	0,9995670124	0,0004329876	70 923,48	30,11	2 451 806,30	22 848,85	60 807 457,26	1 259 503,22	99 288	5 816 510	58,57
18	99267	52	0,9994761603	0,0005238397	69 502,72	35,69	2 380 882,82	22 818,74	58 355 650,95	1 236 654,37	99 241	5 717 222	57,59
19	99215	58	0,9994154110	0,0005845890	68 104,23	39,03	2 311 380,10	22 783,05	55 974 768,13	1 213 835,63	99 186	5 617 981	56,62
20	99157	62	0,9993747290	0,0006252710	66 729,82	40,91	2 243 275,87	22 744,02	53 663 388,03	1 191 052,58	99 126	5 518 796	55,66
21	99095	63	0,9993642464	0,0006357536	65 380,48	40,75	2 176 546,05	22 703,11	51 420 112,16	1 168 308,56	99 064	5 419 669	54,69
22	99032	67	0,9993234510	0,0006765490	64 057,76	42,49	2 111 165,57	22 662,36	49 243 566,10	1 145 605,45	98 999	5 320 606	53,73
23	98965	71	0,9992825746	0,0007174254	62 759,24	44,14	2 047 107,81	22 619,87	47 132 400,54	1 122 943,09	98 930	5 221 607	52,76
24	98894	69	0,9993022833	0,0006977167	61 484,52	42,06	1 984 348,57	22 575,73	45 085 292,73	1 100 323,22	98 859	5 122 677	51,80
25	98825	67	0,9993220339	0,0006779661	60 236,89	40,04	1 922 864,04	22 533,67	43 100 944,16	1 077 747,49	98 791	5 023 818	50,84
26	98758	65	0,9993418255	0,0006581745	59 015,74	38,08	1 862 627,15	22 493,63	41 178 080,12	1 055 213,82	98 725	4 925 027	49,87
27	98693	63	0,9993616569	0,0006383431	57 820,48	36,19	1 803 611,42	22 455,55	39 315 452,97	1 032 720,18	98 662	4 826 301	48,90
28	98630	67	0,9993206935	0,0006793065	56 650,56	37,73	1 745 790,93	22 419,37	37 511 841,55	1 010 264,63	98 597	4 727 640	47,93
29	98563	69	0,9992999401	0,0007000599	55 502,04	38,09	1 689 140,37	22 381,64	35 766 050,62	987 845,26	98 528	4 629 043	46,97

Popis úmrtnostní tabulky

q_x je pravděpodobnost úmrtí x -letých (před $x + 1$ narozeninami)

p_x je pravděpodobnost dožití se $x + 1$ narozenin (přežití věku x)

l_x je počet osob dožívajících se věku x (pojištění na doživotní důchod)

d_x je počet zemřelých ve věku x » $l_x - l_{x+1}$ (pojištění na dožití a pro případ úmrtí)

Komutační čísla

Slouží pro zrychlení výpočtů z důvodu často se opakujících součtů a součinů

D_x – diskontovaný počet osob dožívajících se věku x

C_x – diskontovaný počet zemřelých ve věku x

N_x - součet D_x až do konce tabulky

M_x - součet C_x až do konce tabulky

S_x – součet N_x až do konce tabulky

R_x - součet M_x až do konce tabulky

$$N_x = \sum_{K=0}^{\omega-x} D_{x+K} = \sum_{K=0}^{\omega-x} l_{x+K} \cdot v^{x+K}$$

Brutto pojistné

Brutto (celkové) Pojistné		
Netto (ryzí) pojistné		správní náklady
rezervotvorná (spořicí) složka	riziková složka	správní náklady

Brutto pojistné

$$PB = PN + \alpha + \beta + \gamma + \delta + \varepsilon$$

PB – brutto pojistné

PN – netto pojistné

α – jednorázové počáteční náklady

β – běžné správní náklady po celou dobu pojištění

γ – běžné inkasní náklady

δ – běžné správní náklady spojené s výplatou důchodu

Náklady pojištění

- **jednorázové počáteční náklady (α)** – bývají vynakládány pojišťovnou hned na počátku pojistné doby, při sjednání pojistné smlouvy. Těmito náklady pojišťovna pokrývá provize prodejců životního pojištění, výdaje na vystavení pojistné smlouvy, lékařskou vstupní prohlídku apod. a zpravidla bývají úměrné sjednané pojistné částce nebo důchodu ve formě nějaké výše procenta z pojistné částky,
- **běžné správní náklady (β)** – vynakládány během celého trvání pojištění nezahrnuté v ostatních nákladových položkách a jsou spojeny s udržováním daného pojištění, korespondenci s pojištěným, administrativou apod. a udávají se opět jako procenta z pojistné částky nebo důchodu, ale bývají o řád nižší než náklady počáteční
- **inkasní náklady (γ)** – jsou spojené s inkasem běžného pojistného ale tentokrát jsou stanoveny jako procenta z ročního brutto pojistného,
- **náklady při výplatě důchodu (δ)** – týkají se pouze pojištění, kde dochází k výplatě důchodu a tedy souvisí pouze s výplatami důchodu. V současné době dochází ke zmenšování γ a δ nákladů díky bezhotovostním platbám.
- **Bezpečnostní přírážku** si pojišťovna většinou připočítává pro případ nepříznivých výkyvů náhodné povahy v souboru pojištěných, kterými mohou být např. náhlé zvýšení úmrtnosti v některých věkových kategoriích, hromadné rušení pojistných smluv klienty, epidemie atd. Pojišťovna by měla dbát na to, aby se tato bezpečnostní přírážka nestávala dodatečným zdrojem nadměrných zisků pojišťovny.

Principy při výpočtu netto pojistného

Fiktivní soubor

pojišťovna předpokládá, že všechny osoby se narodily 1.1. a zemřeli 31.12.

předpoklad, že počet osob, které ve věku x uzavřou stejný typ pojištění, je lx z použité úmrtnostní tabulky. Tedy, že daný typ pojištění uzavřou všechny osoby, které jsou ve věku x naživu. Ačkoliv je skutečnost zcela jiná, jde o značné zjednodušení, které vede k dostatečně přesným výsledkům a k jeho praktickému využití.

Princip ekvivalence

příjmy a výdaje pojišťovny jsou v rovnováze
zohlednění:

a) časové rozložení příjmů a výdajů – finanční matematika

finanční toky rozložené v čase se vztáhnou diskontováním do jejich počáteční hodnoty nebo naopak

b) náhodný charakter fin. toků – očekávání (stř. hodnota)

Princip ekvivalence

očekávaná počáteční hodnota pojistného
=
očekávaná počáteční hodnota pojistného plnění

Předpoklady:

P – jednorázové pojistné

O – běžné pojistné

$P\check{C} = 1$

v ... Diskontní faktor = $1/(1+i)$

Pojištění pro případ smrti

Pravděpodobnostní vzorec – vznik vydělením l_x

Vzorec $P = \sum_{t=0}^{\omega-x} {}_tP_x \cdot q_{x+t} \cdot v^{t+1}$ ísel - vznik vynásobením v^x

$$P = \frac{M_x}{D_x}$$

Pojištění pro případ dožití

- X-letá osoba dostane PČ v případě, že bude naživu
- Rovnice ekvivalence
- a) Vzorec pomocí komutačních čísel

$$P = \frac{l_{x+n}}{l_x} \cdot v^n = \frac{l_{x+n} \cdot v^n}{l_x} = \frac{l_{x+n} \cdot v^{x+n}}{l_x \cdot v^x} = \frac{D_{x+n}}{D_x}$$

- b) Pravděpodobnostní vzorec

$$P = {}_n p_x \cdot v^n$$

Doživotní důchod předlhůtní

X-letá osoba dostane vyplacenu PČ v případě, že je vždy na začátku období naživu

- Vzorec pomocí komutačních čísel

- Pravo $P = \frac{N_x}{D_x}$ vzorec

$$P = \sum_{t=0}^{\infty} {}_tP_x \cdot v^t$$

Doživotní důchod polhůtní

Výpočet pomocí komutačních čísel

$$P = \frac{N_{x+1}}{D_x}$$

Doživotní důchod s garancí vyplácení n let

x-letá osoba se pojistí tak, že po dobu n let mu bude vyplácena PČ, ať už je na živu nebo ne

$$l_x \cdot v^x \cdot P = 1 \cdot l_x \cdot v^x + 1 \cdot l_x \cdot v^1 \cdot v^x + \dots + 1 \cdot l_x \cdot v^{n-1} \cdot v^x + 1 \cdot l_{x+n} \cdot v^n \cdot v^x + \dots + 1 \cdot l_{\omega} \cdot v^{\omega-x} \cdot v^x$$

$$l_x \cdot v^x \cdot P = 1 \cdot l_x \cdot v^x + 1 \cdot l_x \cdot v^1 \cdot v^x + \dots + 1 \cdot l_x \cdot v^{n-1} \cdot v^x + 1 \cdot l_{x+n} \cdot v^{x+n} + \dots + 1 \cdot l_{\omega} \cdot v^{\omega}$$

$$D_x \cdot P = D_x \cdot (1 + v + \dots + v^{n-1}) + D_{x+n} + D_{x+n+1} + D_{x+n+2} + \dots + D_{\omega}$$

$$D_x \cdot P = D_x \cdot \frac{v^n - 1}{v - 1} + N_{x+n}$$

$$P = \frac{v^n - 1}{v - 1} \cdot \frac{1}{D_x} + \frac{N_{x+n}}{D_x} = \ddot{a}_{\overline{n}|} + {}_n|\ddot{a}_x$$

Doživotní důchod rostoucí lineárně

x- leté osobě je 1. rok vyplacena 1 p.j., 2.rok 2 p.j.,...

$$l_x \cdot v^x \cdot \mathbf{P} = 1 \cdot l_x \cdot v^x + 2 \cdot l_{x+1} \cdot v^x \cdot v^1 + \dots + n \cdot l_{x+n-1} \cdot v^x \cdot v^{n-1} + \dots + (\omega - x + 1) \cdot l_{\omega} \cdot v^x \cdot v^{\omega-x}$$

$$D_x \cdot \mathbf{P} = 1 \cdot D_x + 2 \cdot D_{x+1} + \dots + n \cdot D_{x+n-1} + \dots + (\omega - x + 1) \cdot D_{\omega}$$

$$D_x \cdot \mathbf{P} = S_x$$

$$\mathbf{P} = \frac{S_x}{D_x}$$

Smíšené pojištění

PČ je vyplacena v případě že kdykoliv do doby $x+n$ zemře nebo se dožije věku $x+n$

$$l_x \cdot v^x \cdot \mathbf{P} = 1 \cdot d_x \cdot v^x \cdot v + 1 \cdot d_{x+1} \cdot v^x \cdot v^2 + \dots + 1 \cdot d_{x+n-1} \cdot v^x \cdot v^n + 1 \cdot l_{x+n} \cdot v^x \cdot v^n$$

$$l_x \cdot v^x \cdot \mathbf{P} = 1 \cdot d_x \cdot v^{x+1} + 1 \cdot d_{x+1} \cdot v^{x+2} + \dots + 1 \cdot d_{x+n-1} \cdot v^{x+n} + l_{x+n} \cdot v^{x+n}$$

$$D_x \cdot \mathbf{P} = 1 \cdot C_x + 1 \cdot C_{x+1} + \dots + 1 \cdot C_{x+n-1} + 1 \cdot D_{x+n}$$

$$D_x \cdot \mathbf{P} = M_x - M_{x+n} + D_{x+n}$$

$$\mathbf{P} = \frac{M_x - M_{x+n} + D_{x+n}}{D_x}$$

Běžné netto pojistné

Předpokládáme, že doba placení je kratší než doba trvání pojištění

Pojištění na dožití

$$O = \frac{D_{x+n}}{N_x - N_{x+m}}$$

Motivační příklad

Jaké bude běžné netto pojistné pokud budete 15 let platit pojistné a chcete od 40. roku prvních 10 let garanci důchodu 12 000 Kč ročně, poté chcete, aby 10 let důchod rostl o 500 Kč, poté chcete opět garanci 10 let ve výši narostlého ročního důchodu a poté chcete, aby od 70. roku života důchod 5 roků klesal o 5 % z výše důchodu, kterou obdržíte v 69. roku života? (výpočet viz tabule)

7. Investování do akcií

Akcie

- Akcie je majetkový cenný papír
- S držením akcie vzniká majiteli
 - Vlastnické právo na určitý podíl firmy,
 - Právo podílet se na chodu společnosti, tj. účastnit se valných hromad,
 - Právo na likvidační zůstatek
- Je to dlouhodobý cenný papír, bez pevné doby splatnosti
- Převoditelný na jinou osobu
- Musí obsahovat:
 - Obchodní jméno a sídlo společnosti
 - Jmenovitou hodnotu
 - Označení formy akcie
 - U akcie na jméno i jméno akcionáře
 - Datum emise

Výnos z akcií

- Dividenda
 - Peněžitý podíl z čistého zisku akciových společností vyplácený akcionářům
 - Není pevně daná
- Prodání akcie na trhu cenných papírů
 - Zisk nebo ztráta je rozdíl mezi cenou, za kterou jsme akcii koupili a cenou, za kterou ji prodáváme

- Cena jednotlivých akcií je dána střetem poptávky a nabídky, nedá se určit její budoucí hodnota
- Jako každá komodita na trhu je i cena akcie tím vyšší, čím větší je poptávka
- Efektivní mechanismus převodu volných finančních prostředků investorů ke společnosti

Druhy akcií

- Dle formy:
 - Akcie na majitele (doručitele)
 - Na jméno
 - Na řád
- Dle účasti na řízení společnosti
 - Kmenové akcie
 - Prioritní akcie
- Dále specifické akcie
 - Zaměstnanecké akcie
 - Zlaté akcie (zakladatelské)
 - Požitkové akcie
- Další investiční instrumenty, zajišťující nárok na budoucí cash flow jsou finanční investiční instrumenty (např. cenné papíry, finanční deriváty), reálné investiční instrumenty (např. nemovitosti, umělecké sbírky, drahé kovy)

Výhody spojené s akciemi z hlediska emitenta

- Získaný kapitál má společnost na neomezeně dlouhou dobu, akcie jsou nesplacitelné
- Akciová společnost může shromáždit velké množství kapitálu
- Možnost získávat další kapitál emisí nových akcií
- Nemusí vyplácet dividendy, pokud tak rozhodne valná hromada
- Obchodováním na sekundárním trhu je zabezpečena likvidita akcií, snížení nákladů emitenta
- Rostoucí prestiž eminenta, pokud jsou akcie obchodovány na sekundárním trhu
- Obchodováním na sekundárním trhu dochází k dennímu oceňování hodnoty společnosti připadající na jednu akcii
- Emisí akcií dochází k diverzifikaci rizika mezi větší počet akcionářů^[1]

[1] Veselá, Jitka. Investování na kapitálových trzích.

Nevýhody spojené s akciemi z hlediska emitenta

- Akcionáři mají právo zasahovat a podílet se na řízení společnosti
- Informační povinnosti spojené s vysokými informačními náklady
- Vysoké emisní náklady s novými akciemi
- Za porušení závazků ručí celým svým majetkem
- Dividendy vypláceny až ze zisku po zdanění
- Přísná regulace právní formy v podobě akciové společnosti ve snaze zabránit vzniku monopolní či oligopolní tržní struktury
- Společnosti, u kterých jsou akcie veřejně obchodovatelné, musí provádět transparentní a kontinuální dividendovou, investiční a finanční politiku
- Ve společnosti může docházet ke konfliktu zájmů mezi managementem a akcionáři společnosti^[1]

Výhody spojené s akcemi z hlediska akcionáře (investora)

- Investicí do akcie může investor dosáhnout kapitálového zisku
- Inkasuje důchod ve formě dividendy, pokud je vyplácena
- Existuje omezené ručení, akcionář neručí za závazky akciové společnosti
- Má možnost podílet se na řízení akciové společnosti, účastnit se valné hromady a hlasovat
- Má právo na likvidační zůstatek
- Má předkupní právo na nákup nových akcií
- Investování do akcií představuje anonymní investováníNemá právo podílet se na řízení společnostiU obchodovatelných akcií je zajištěna likvidita^[1]

Nevýhody spojené s akciami z hlediska akcionáře (investora)

- Investicí do akcie může investor dosáhnout kapitálové ztráty
- Získává nulový důchod, pokud dividenda vyplácena není
- Ve společnosti může docházet ke konfliktu zájmů mezi managementem a akcionáři společnosti
- Minoritní vlastník má z praktického hlediska omezenou možnost zasahovat do řízení firmy
- Nemá nárok na vrácení svého vkladu
- Může docházet k vysokému zdanění kapitálových zisků či dividend
- S neobchodovatelnými akciami je spojena nízká likvidita
- Obchodování s malým počtem kusů akcií je spojeno s vysokými transakčními náklady^[1]

Oceňování akcií

- Spolehlivá metoda stanovení správné hodnoty akcie neexistuje
- Cena akcie je následkem mnoha událostí a nelze ji předvídat
- Nelze pokrýt všechny faktory, které cenu akcie ovlivňují
- Existuje mnoho teorií, které vysvětlují chování cen akcií (např. náhodná procházka)
- Pro oceňování akcií je důležitým parametrem především zisk společnosti

Dlouhodobý vývoj cen akcií

- Odráží trendy fundamentálních veličin a to
 - Makroekonomických (HDP, inflace, zaměstnanost, ...)
 - Odvětvových (dostupnost surovin, vývoj technologií, ...)
 - Individuálních (zadlužení, úroveň vedení, ...)

Metody oceňování akcií

- Fundamentální analýza
- Technická analýza
- Psychologická analýza

Kritéria investičního rozhodování

- Investice je záměrné obětování dnešní hodnoty za účelem získání vyšší hodnoty v budoucnu
 - Výnos plyne z nároku na plynoucí cash flow, nebo cena podkladového aktiva stoupne a prodáme ho za vyšší cenu, než za kterou jsme ho nakoupili
- Existují 3 základní investiční faktory tvořící tzv. magický trojúhelník
 - Výnosnost
 - Rizikovitost
 - Likvidita



[2]

Výnosnost, riziko, likvidita

- Výnosnost udává míru zhodnocení peněžních prostředků
 - ex post (již zrealizovaná)
 - ex ante (očekávaná výnosnost)
- Riziko je nebezpečí, že nedosáhneme očekávaného výnosu. Za vyšší riziko investor zpravidla požaduje vyšší výnos
 - Riziko změn tržní úrokové míry, inflační, měnové, právní, události, ztráty likvidity předmětného finančního instrumentu, ...
 - Riziku lze předcházet diverzifikací portfolia, tj. „nevsázet vše na jednu kartu“
- Likvidita je rychlost, s jakou je možnost přeměnit finanční instrument bez ztráty zpět v hotové peníze

Akciové indexy

- Akciový index je skupina konkrétně vybraných jednotlivých akciových titulů. Jeho hlavní úlohou je okamžitě prozradit investorovi, jak si vedou akciové tituly, jež jsou v indexu zahrnuty, jako celek.
- Při změně vah se vychází z aktuálních cen za tu kterou akcii k danému dni, maximální váha může být vždy maximálně jen 25%, u akcií, které jsou mimo pražskou burzu obchodované i na jiné burze v zahraničí, se počítá dále poměr zobchodovaných akcií Česká republika: zahraničnímu trhu.^[3]

$$PX(t) = K(t) \frac{M(t)}{M(0)} \times 1000$$

- M(t) tržní kapitalizace báze v čase t
- M(0) = 379 786 853 620,0 Kč představuje tržní kapitalizaci báze indexu PX ve výchozím dni 5. 4. 1994
- K(t) faktor zohledňující změny uskutečněné v bázi indexu, s účinností od 19. 3. 2012 hodnota faktoru zřetězení K(t) = 0,4680501949
- Dow Jones Industrial Average (DJIA)
- National Association of Securities Dealers Automated Quotations 100 (NASDAQ)
- S&P 500
- Nikkei 225
- Index PX, což je oficiální index Burzy cenných papírů Praha.

^[3]IHNED.CZ [online]. 1998-2011. 2012 [cit. 2012-3-11]. Váha indexu pro výpočet PX. Dostupné z WWW: <<http://byznys.ihned.cz/trhy-a-investice/c1-35147550-akcie-aaa-prestanou-mit-vahu-pro-vypocet-indexu-px>>

Akciové indexy

- Akciové indexy mají za úkol stručně informovat o vývoji celku určitého akciového trhu či jeho části. Indexy se používají jako standard (benchmark), což je měřítko průměrné výnosnosti daného trhu, se kterým se měří úspěšnost či neúspěšnost investování portfoliových manažerů. Je-li výkonnost portfoliového manažera vyšší než růst indexu, říká se, že manažer překonal trh. Indexy představují indikátory akciového trhu, koncentrující pohyby cen akcií do jednoho čísla a vypovídají o vývojových tendencích trhu.

Index PX

- SPAD (Systém pro podporu trhu akcií a dluhopisů) je nejvýznamnější segment BCPP. V rámci tohoto segmentu se obchoduje s cennými papíry vybraných významných podniků – jsou to ty nejlikvidnější a nejobchodovatelnější.
 - Úkolem SPADU je zajistit fungování burzovního trhu na pražské burze.
- Blue chip, blue chip stocks (česky „modrý žeton“) je termín označující akcie největších a nejziskovějších společností, které jsou obchodovány na burze, mají stabilní růst a pravidelně vyplácejí dividendy. V ČR to jsou zejména tituly společností obchodované na hlavním trhu SPAD BCPP tvořící Index PX.

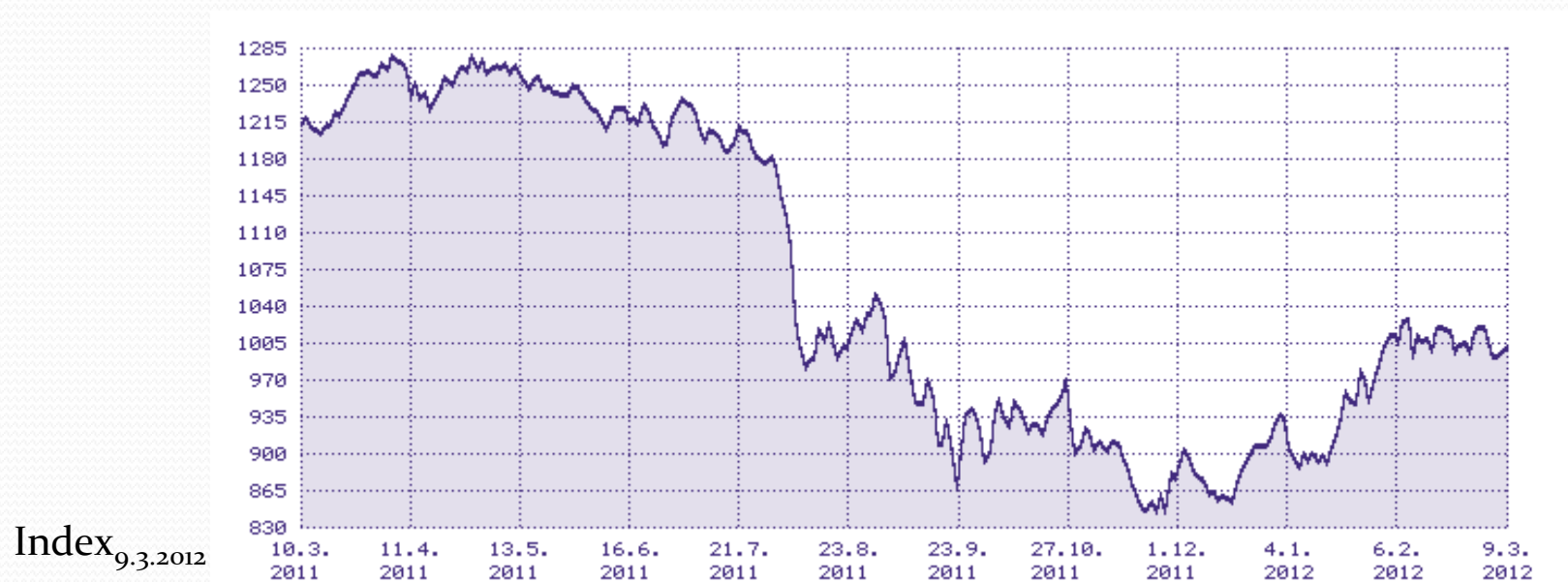
Index PX

- Index sdružuje následující společnosti, v závorkách uvedeny váhy: ^[3]
 - Erste group bank (21,72 %)
 - ČEZ (19,02 %)
 - KB (17,38 %)
 - Telefónica C.R. (15,53 %)
 - VIG (12,55 %)
 - NWR (4,71 %)
 - Unipetrol (3,83 %)
 - Philip Morris ČR (2,68 %)
 - CETV (0,95 %)
 - Fortuna (0,63 %)
 - Pegas Nonwovens (0,51 %)
 - Orco (0,21 %)
 - AAA (0,16 %)
 - Kitd (0,13 %)

^[3]Burza cenných papírů Praha [online]. 1998-2011. 2012 [cit. 2012-3-11]. Báze indexu PX ke dni 11.3.2012. Dostupné z WWW: <<http://www.bcpcz.cz/Statistika/Burzovni-Indexy/>>.

Index PX

- Dlouhodobý vývoj PX indexu [4]



[4] *Burza cenných papírů Praha* [online]. 1998-2011. 2012 [cit. 2012-3-11]. Index PX. Dostupné z WWW: <<http://www.bcpcz.cz/Statistika/Burzovni-Indexy/Detail.aspx?bi=1>>.

Obchodování na sekundárním trhu

- Akcie nejsou prodávány pouze jejich konečným majitelům (primární trh), ale běžně se s nimi obchoduje a to prostřednictvím sekundárního trhu

Burzy

- New York Stock Exchange Euronext
 - V roce 2007 fúze NYSE a Euronext (Francie, Belgie, Holansko, Portugalsko, Luxembursko, Velká Británie)
- Tokyo Stock Exchange (Japonsko)
- National Association of Securities Dealers Automated Quotations (NASDAQ, USA)
- London Stock Exchange (Velká Británie)
- Shanghai Stock Exchange (Čína)
- Frankfurt Stock Exchange (Německo)
- Burza cenných papírů Praha (BCPP, ČR)

^[5]Wikipedie [online]. 2012 [cit. 2012-3-11]. Pohled na obchodování v roce 2008. Dostupné <http://upload.wikimedia.org/wikipedia/commons/thumb/3/3e/NYSE127.jpg/800px-NYS>



[5]

Burza cenných papírů Praha, a. s.

- Založena na členském principu, tj. přístup do burzovního systému a právo obchodovat mají pouze licencovaní obchodníci s cennými papíry, kteří jsou zároveň členy burzy
- 24.11.1992 Vznik BCCP, a.s.
- 5.4.1994 Zahájení výpočtu oficiálního burzovního indexu PX 50, dnes již nahrazeno PX
- 15.3.1996 Zahájení obchodování v systému KOBOS (průběžné obchodování při proměnlivé ceně) s 5 emisemi akcií a 2 emisemi obligací
- 25.5.1998 Zahájení obchodování v systému SPAD (Systém pro podporu trhu akcií a dluhopisů)
- 8.12.2005 Objemy obchodů v průběhu roku poprvé překročili bilionovou hranici, celkový objem obchodů za rok 2005 dosáhl 1041,2 mld. korun^[6]



[6] *Miras.cz* [online]. 2000-2011. 2012 [cit. 2012-3-12]. Burza cenných papírů Praha - historie. Dostupné z WWW: <<http://www.miras.cz/akcie/burza-praha-historie.php>>

[7] *Wikipedie* [online]. 2000-2011. 2012 [cit. 2012-3-12]. Burza cenných papírů Praha v datech. Dostupné z WWW: <<http://cs.wikipedia.org/>>

Investiční doporučení

Titul	Sektor	Doporučení	Cílová cena	Měna
ČEZ	Energetika	akumulovat	970	CZK
NWR	Těžba uhlí	akumulovat	310	CZK
KB	Bankovníctví	koupit	4 150	CZK
TELEFONICA	Telekomunikace	držet	387	CZK
FORTUNA	Sázkové služby	koupit	145	CZK

[8] *Cyrrus* [online]. 1998-2011. 2012 [cit. 2012-3-11]. Aktuální doporučení. Dostupné z WWW:
<<http://www.cyrrus.cz/aktualni-doporuceni>>.

Portfolio

Titul	Kupní hodnota ^[9]	Množství	Akcie dohromady
ČEZ	829,00	40	33 160,00
KB	3 782,00	10	37 820,00
NWR	145,60	210	30 576,00
Telefónica O2 C.R.	395,80	51	20 185,80
Fortuna	98,70	300	29 610,00
Hodnota portfolia			151 351,80

Vývoj akcií a příslušného portfolia byl sledován od

24. 2. 2012 do 9. 3. 2012

[9] *Burza cenných papírů Praha* [online]. 1998-2012. 2012 [cit. 2012-3-11]. Index PX. Dostupné z WWW: <<http://www.bcpcp.cz/Statistika/Burzovni-Indexy/Detail.aspx?bi=1>>.

České energetické závody (ČEZ), a. s.

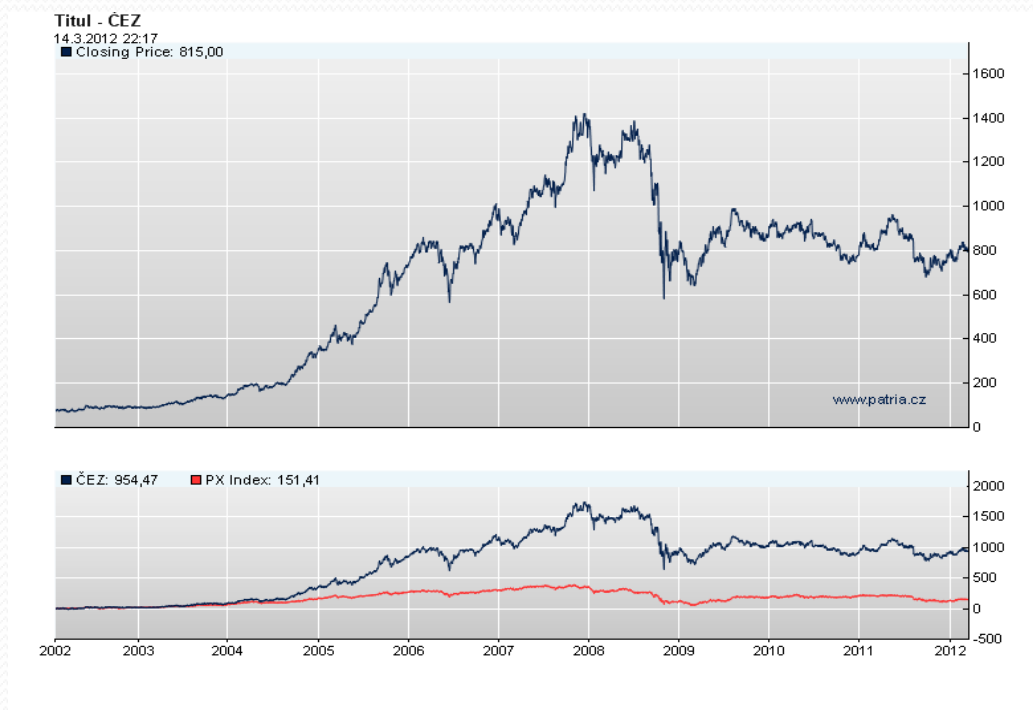
- Založeny roku 1992
- Hlavním předmětem činnosti je výroba a prodej elektřiny
- Dále se zabývá výrobou, rozvodem a prodejem tepla
- V roce 2003 se ČEZ spojila s distribučními společnostmi SeČ energetika, SeM en., StČ en., VČ en., a ZČ en. a vznikla Skupina ČEZ
- Postupně akvizice v mnoha státy Evropy
- Skupina ČEZ patří do evropské desítky největších energetických koncernů
- V České republice je Skupina ČEZ největším výrobcem elektřiny a tepla

Hlavní akcionáři ČEZ

	stav k 31.12.2007	stav k 31.12.2008	stav k 31.12.2009	stav k 31.12.2010
Česká republika	65,99 %	63,39 %	69,78 %	69,78 %
Ostatní právnické osoby	19,04 %	12,97 %	4,02 %	2,34 %
domácí	10,24 %	11,59 %	2,73 %	1,22 %
zahraniční	8,80 %	1,38 %	1,29 %	1,12 %
Fyzické osoby	4,14 %	4,55 %	5,41 %	5,13 %
domácí	4,04 %	4,43 %	5,26 %	4,43 %
zahraniční	0,10 %	0,12 %	0,15 %	0,70 %

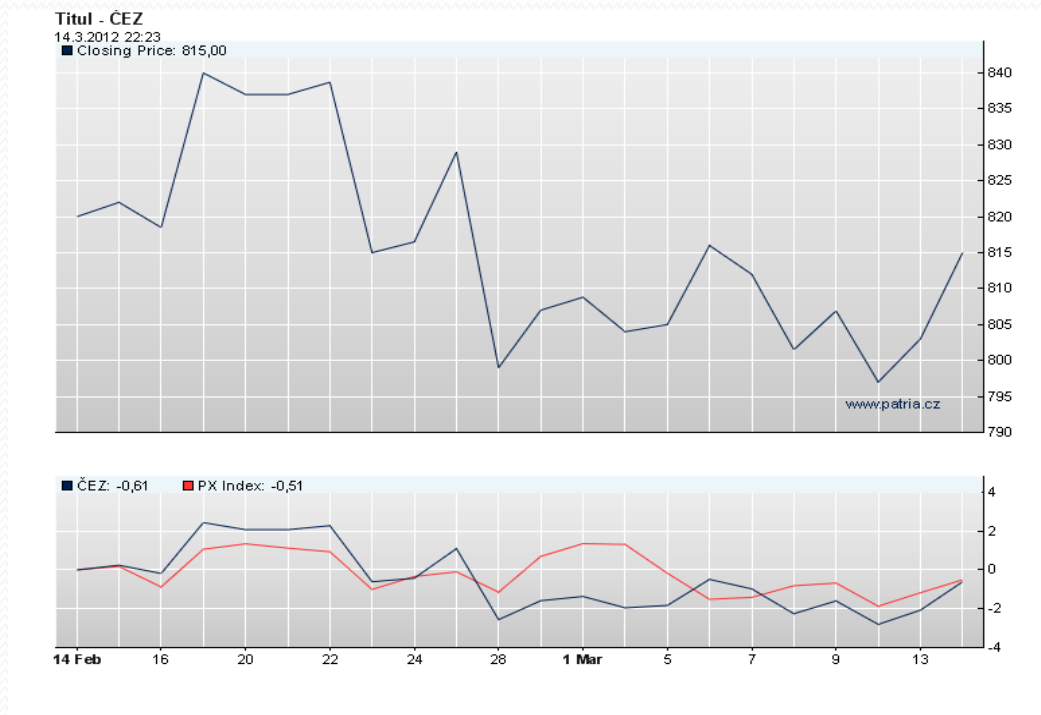
[10] ČEZ, a. s. [online]. 1998-2012. 2012 [cit. 2012-3-11]. Struktura akcionářů. Dostupné z WWW: <<http://www.cez.cz/cs/o-spolecnosti/cez/struktura-akcionaru.html>>

Dlouhodobý vývoj akcií ČEZ



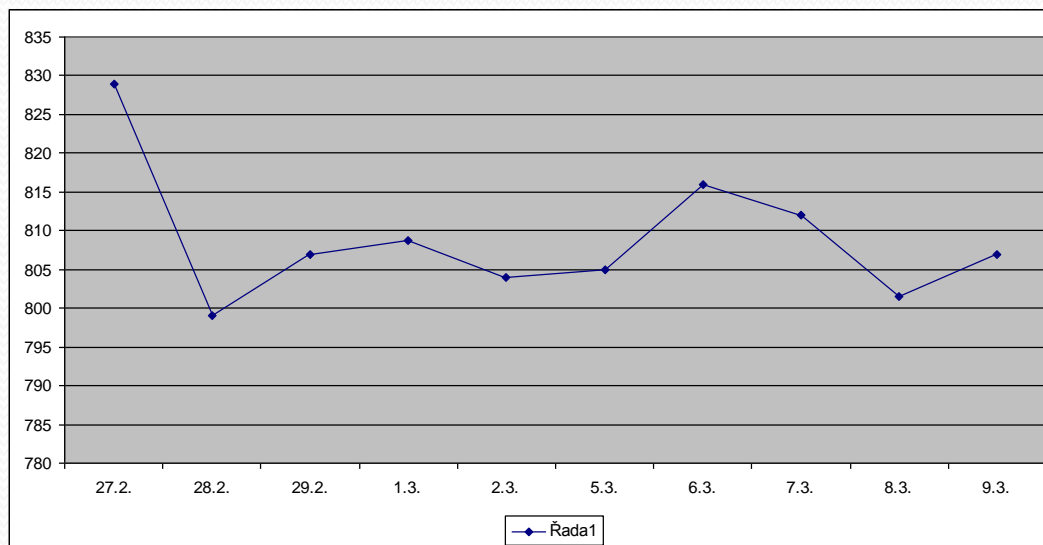
[1] Patria [online]. 1990-2012, 2012 [cit. 2012-3-14]. Detail graf. Dostupné z www.<<http://www.patria.cz/akcie/CEZPsp.PR/cez/graf.html>>

Vývoj akcií ČEZ v průběhu měsíce



[12] Patria [online]. 1998-2012. 2012 [cit. 2012-3-14]. Detail graf. Dostupné z www. <<http://www.patria.cz/akcie/detail/graf.html>>

Vývoj akcií ČEZ v průběhu 14 dní



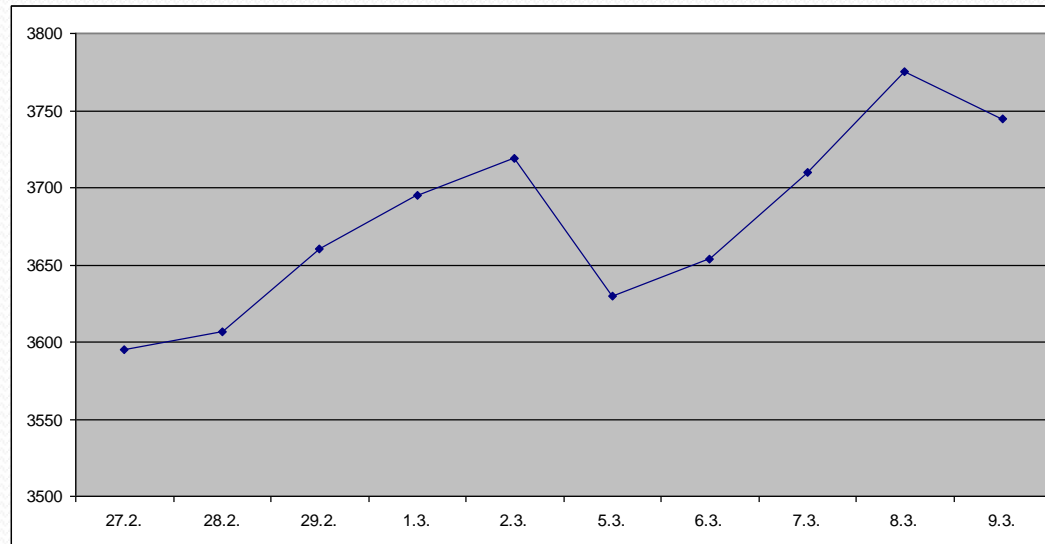
27.2.	829
28.2.	799
29.2.	807
1.3.	808,8
2.3.	804
5.3.	805
6.3.	816
7.3.	812
8.3.	801,5
9.3.	806,9

[13] *Investiční web* [online]. 2012 [cit. 2012-3-14]. Kurzy. Dostupné z WWW:
<<http://www.investicniweb.cz/kurzy/bcpp/charts/detail/cz0005u2300.html>>

KB

- Založena v roce 1990 jako státní instituce
- V roce 1992 byla transformována na akciovou společnost
- KB je univerzální bankou se širokou nabídkou služeb v oblasti podnikového a investičního bankovníctví
- Nabízejí další specializované služby, mezi něž patří penzijní připojištění, stavební spoření, spotřebitelské úvěry a pojištění, dostupné prostřednictvím sítě poboček KB, přímého bankovníctví a vlastní distribuční sítě
- Služby samotné Komerční banky využívá 1,59 milionu zákazníků prostřednictvím 395 poboček a 677 bankomatů po celé České republice

Vývoj akcií KB v průběhu 14 dní

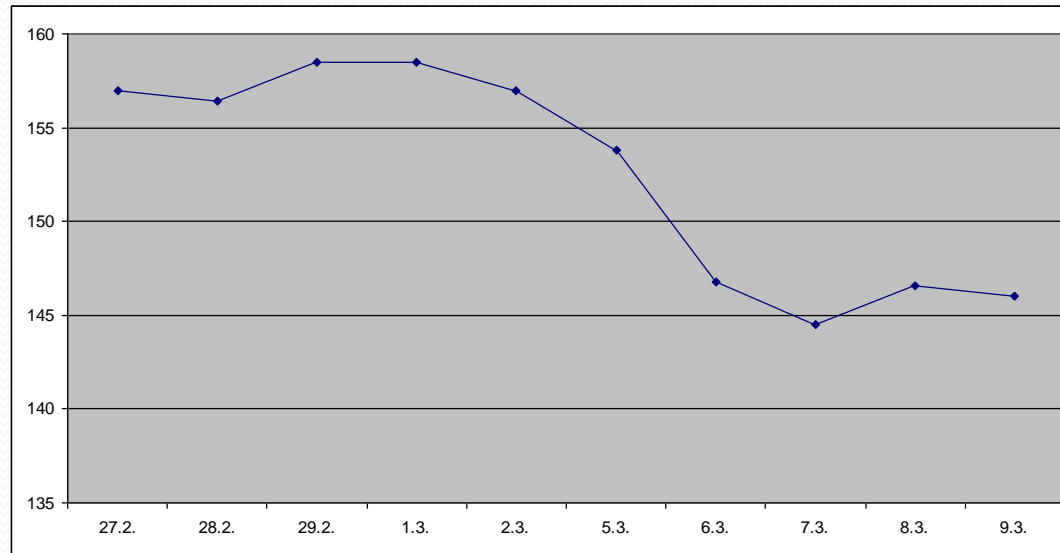


27.2.	3595
28.2.	3607
29.2.	3660
1.3.	3695
2.3.	3719
5.3.	3630
6.3.	3654
7.3.	3710
8.3.	3775
9.3.	3745

NWR

- Společnost NWR je předním středoevropským producentem černého uhlí a koksu
- V současné době společnost těží na území České republiky
- NWR zaměstnává 18 553 lidí
- Společnost NWR sídlí v Nizozemsku a jejím majitelem je jeden z nejbohatších Čechů Zdeněk Bakal

Vývoj akcií NWR v průběhu 14 dní

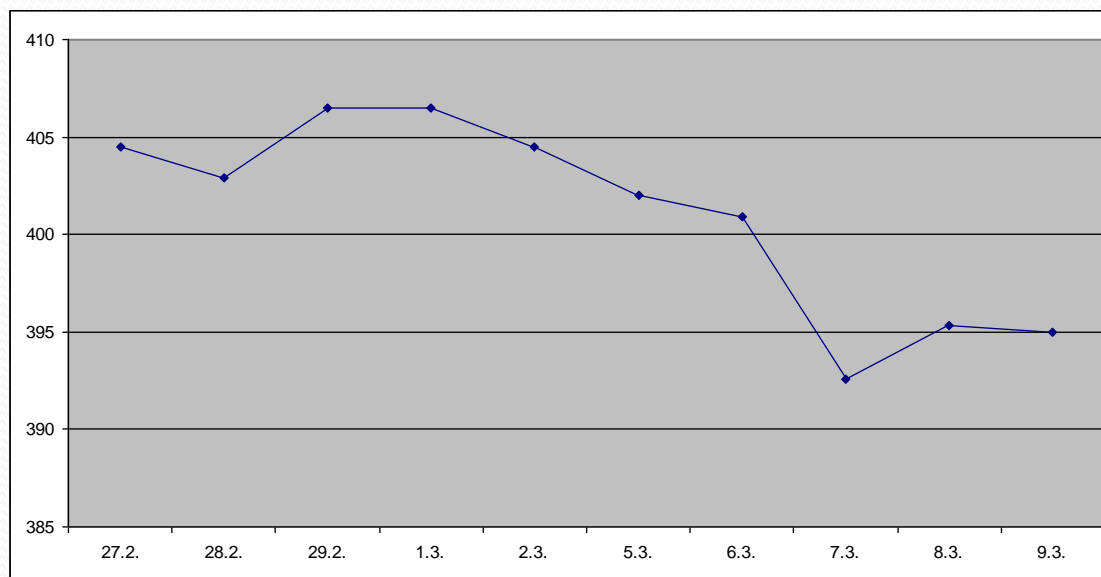


27.2.	157
28.2.	156,4
29.2.	158,5
1.3.	158,5
2.3.	157
5.3.	153,8
6.3.	146,8
7.3.	144,5
8.3.	146,6
9.3.	146

Telefónica C.R.

- Vznikla 1. července 2006 spojením společností ČESKÝ TELECOM, a.s. a Eurotel Praha, spol. s r.o.
- Je předním telekomunikačním operátorem na českém trhu
- V současnosti provozuje téměř 7mil. mobilních a pevných linek

Vývoj akcií Telefónica C.R. v průběhu 14 dní

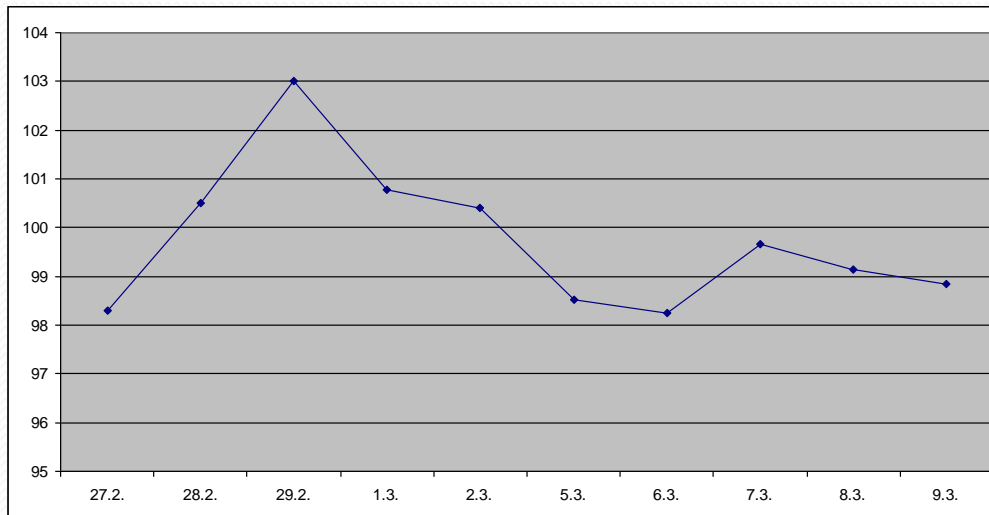


27.2.	404,5
28.2.	402,9
29.2.	406,5
1.3.	406,5
2.3.	404,5
5.3.	402
6.3.	400,9
7.3.	392,6
8.3.	395,3
9.3.	395

Fortuna

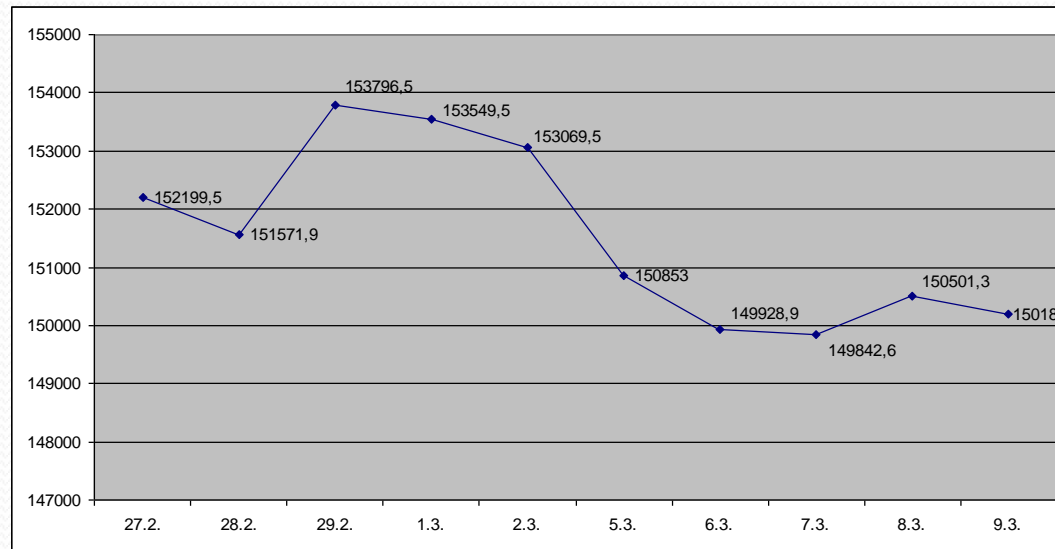
- Zakládající společnost tohoto seskupení Fortuna sázková kancelář, a. s. vznikla v roce 1990 v Praze
- V roce 2009 vznikla Fortuna Entertainment Group a stala se největším středoevropským provozovatelem kurzových sázek
- Vzestup online sázení jak v Česku, tak i na Slovensku

Vývoj akcií Fortuna v průběhu 14 dní



27.2.	98,3
28.2.	100,5
29.2.	103
1.3.	100,77
2.3.	100,4
5.3.	98,51
6.3.	98,25
7.3.	99,65
8.3.	99,15
9.3.	98,85

Vývoj portfolia v průběhu 14 dní

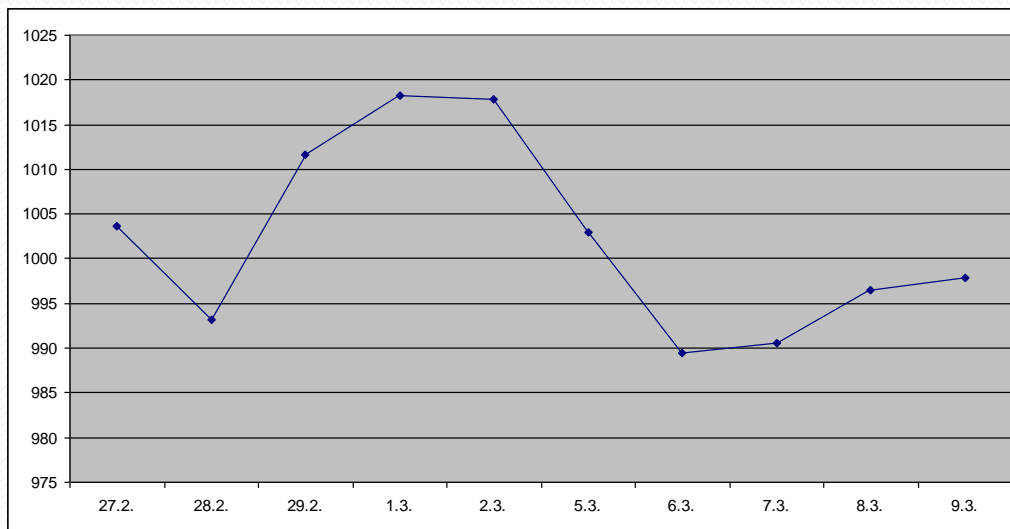


27.2.	152199,5
28.2.	151571,9
29.2.	153796,5
1.3.	153549,5
2.3.	153069,5
5.3.	150853
6.3.	149928,9
7.3.	149842,6
8.3.	150501,3
9.3.	150186

Vyhodnocení portfolia

Titul	Množství	Prodejní hodnota	Změna(%)	Akcie dohromady	Změna v korunách
ČEZ	40	806,9	-2,74	32 276	-884
KB	10	3745	-0,99	37 450	-370
NWR	210	146	-0,41	30 660	-84
Telefónica	51	395	-0,20	20 145	-40,8
Fortuna	300	98,85	-0,15	29 655	-55
Hodnota portfolia				150 186	-1 433,8

Vývoj indexu PX v průběhu 14 dní



27.2.	1003,7
28.2.	993,2
29.2.	1011,7
1.3.	1018,3
2.3.	1017,9
5.3.	1003
6.3.	989,5
7.3.	990,5
8.3.	996,5
9.3.	997,9

8. Fundamentální analýza

Typy analýz:

- Technická analýza
- Psychologická analýza

je založena na předpokladu, že investování je ve značné míře ovlivněno emocemi

- Fundamentální analýza:
 - Globální FA
 - Odvětvová FA
 - Firemní FA

Globální FA

- jde o celkové zhodnocení hospodářské situace na daném trhu a její vliv na vývoj akciového trhu jako celku
- působí zde celá řada faktorů, z nichž za nejdůležitější můžeme považovat zejména:
 - Vývoj HDP
 - Vývoj úrokových měr
 - Změna Inflace
 - Vývoj peněžní zásoby

Vývoj HDP

- vztah HDP a akciových kurzů je kladný
- jedná se o předbíhající faktor, a to o 3 až 9 měsíců
- pokud předpokládáme růst HDP, investoři nakupují akcie, což ve výsledku pozitivně ovlivní reálný výstup ekonomiky

Vývoj úrokových měr

- negativní vztah
- při růstu úrokových měr dochází k poklesu kurzů akcií a naopak, což může být vysvětlováno například jako:
 - změna budoucí vnitřní hodnoty akcií
 - odliv peněžních prostředků z akciových trhů

Změna inflace

- její vliv na vývoj akciových kurzů není zcela zřejmý, avšak spíše budeme mluvit o slabém negativním vztahu
- růst inflace je velice často doprovázen růstem úrokových měr a taktéž v inflačním prostředí roste nejistota v ekonomice, tudíž pro investory stoupá riziko investic do cenných papírů

Vývoj peněžní zásoby

- kladný vztah, hlavně v krátkém období
- předbíhající faktor
- při růstu peněžní zásoby je více peněz investováno do jednotlivých akciových titulů, což způsobuje růst jejich ceny

Odvětвовá FA

- odvětvová analýza zkoumá specifika a vztahy v odvětví, ve kterém daná společnost působí a jejich vliv na kurzy
- důležitou roli zde hrají faktory jako:
 - citlivost odvětví na hospodářský cyklus
 - životní cyklus daného odvětví
 - struktura trhu
 - regulace v odvětví

Citlivost odvětví na hospodářský cyklus

- **Cyklická odvětví**

- firmy produkující statky zbytné spotřeby, tedy výrobky a služby, jejichž spotřeba není nutná a lze ji odložit do budoucna. Cena akcie se poté vyvíjí podobně jako hospodářský cyklus.
- např.: stavebnictví, automobilový průmysl, cestovní ruch, elektrotechnika

- **Neutrální odvětví**

- společnosti produkující statky nezbytné spotřeby (nulová cenová elasticita), či návykové produkty
- např.: potravinářské produkty, farmaceutický průmysl, tabákový průmysl a výroba alkoholických nápojů.

- **Anticyklická odvětví**

- odvětví, která profitují během recese
- firmy z anticyklických odvětví produkují **levnější substituty drahých produktů**,
- např.: levné oděvy, obuv a potraviny nahrazující drahé výrobky

Životní cyklus v odvětví

- **Pionýrská fáze**
 - společnost produkuje nové, či silně inovované produkty
 - po těchto stadiích prudce narůstá poptávka, firmy mohou dosahovat nadprůměrných zisků
 - to způsobuje rostoucí konkurenci a může docházet ke krachu některých podniků
 - investor může dosahovat mimořádných výnosů, avšak za vyššího rizika
 - např.: boom ve výpočetní technice v průběhu 90. let
- **Fáze rozvoje**
 - stabilizace odvětví, firmy, které přečkaly pionýrskou fázi
 - upevňování pozice na trhu
- **Fáze stability**
 - na trhu se nacházejí silné firmy, které mají zavedené jméno
 - stabilní vývoj tržeb, marží a zisku
 - tato fáze se vyznačuje nízkými výnosy a postupně rostoucími náklady, zejména na marketing a propagaci produktů
 - některé společnosti zde již nevidí budoucnost a odvětví postupně opouštějí.
- **Období útlumu**
 - odvětví zastarává, objem produkce pozvolna klesá a firmy zde ukončují činnost

Firemní FA

- věnuje se samotné akciové společnosti
- zkoumá, jak firma hospodaří
- jaký je její očekávaný vývoj v budoucnosti

- zda je příslušný cenný papír správně ohodnocen – pomocí vnitřní hodnoty akcie
- cílem je najít na trhu nadhodnocené a podhodnocené akcie a následně provádět případný nákup či prodej

Vnitřní hodnota

- Vnitřní hodnotu počítáme v různých časových periodách:
 - v delším období - abychom zjistili, jak se mění a jakým směrem můžeme očekávat další pohyb kurzu
 - v krátkém období - je možné považovat vnitřní hodnotu za konstantní a tudíž schopnou porovnání se skutečným kurzem akcie
- Pro budoucí použití označíme aktuální vnitřní hodnotu cenného papíru jako V_o a skutečný tržní kurz P_o .
- $P_o > V_o$ – nadhodnocený cenný papír
- $P_o \approx V_o$ – relativně správně ohodnocený cenný papír
- $P_o < V_o$ – cenný papír je podhodnocený

Metody stanovení vnitřní hodnoty akcie

- Dividendové diskontní modely – DDM
- Ziskové modely
- Cash Flow modely
- Další ohodnocovací modely

Vstupní hodnoty pro modely

- Míra růstu dividend
- Požadovaná výnosová míra

Míra růstu dividend

Historická data

- Průměr hodnot za delší období
 - aritmetický
 - geometrický
- Nevýhoda - nezahrnuje žádné předpoklady do budoucna

$$g = \sqrt[n]{\frac{D_{n+m}}{D_n}} - 1$$

$$g = \sqrt[m]{g_1 \times g_2 \times \dots \times g_m}$$

Míra růstu dividend

Z firemních ukazatelů

- Míra růstu dividend se rovná míře růstu zisku na akcii

$$g = \frac{D_{n+1}}{D_n} - 1 = \frac{e_{n+1}}{e_n} - 1$$

- Budeme vycházet z následujících vztahů:

$$e_{n+1} = ROE_n \times VK_n$$
$$VK_n - VK_{n-1} = b \times e_n$$

- Dosazením získáme:

$$g = \frac{ROE_n \times VK_n - ROE_{n-1} \times VK_{n-1}}{ROE_{n-1} \times VK_{n-1}} = \frac{VK_n - VK_{n-1}}{VK_{n-1}} = \frac{b \times e_n}{VK_{n-1}} \approx ROE \times b$$

Požadovaná výnosová míra

- Nejznámější CAPM
 - (Capital Asset Pricing Model)
 - patří mezi nejpoužívanější modely
 - zahrnuje celou řadu faktorů

$$r = r_f + P_{ri}$$

$$r = r_f + \beta (r_m - r_f)$$

Metody stanovení vnitřní hodnoty akcie

- Dividendové diskontní modely – DDM
- Ziskové modely
- Cash Flow modely
- Další ohodnocovací modely

Dividendové diskontní modely

- nejčastěji používané modely
- diskontováním budoucí očekávané hodnoty akcie i jednotlivých dividendových výnosů v jednotlivých letech

$$VH = \frac{D_1}{(1+r)} + \frac{D_2}{(1+r)^2} + \dots + \frac{D_n + C_n}{(1+r)^n} = \sum_{i=1}^n \frac{D_i}{(1+r)^i} + \frac{C_n}{(1+r)^n}$$

Dividendové diskontní modely

- Modely s nulovým růstem
- Jednostupňové DDM
 - Gordonův model
- Vícestupňové modely

Dividendové diskontní modely

Jednostupňové DDM a Gordonův model

- Jednostupňové

$$VH = \sum_{i=1}^n \frac{D_0(1+g)^i}{(1+r)^i} + \frac{C_n}{(1+r)^n}$$

- Gordonův model

$$VH = \frac{D_0(1+g)}{r-g} = \frac{D_1}{r-g}$$

- Musí být splněn
 - $(r - g) > 0$, tudíž $g < r$

Dividendové diskontní modely

Vícestupňové DDM

$$VH = \sum_{i=1}^n \frac{D_0(1+g_1)^i}{(1+r)^i} + \sum_{j=1}^m \frac{D_0(1+g_2)^n(1+g_2)^j}{(1+r)^{n+m}} + \frac{C_{n+m}}{(1+r)^{n+m}}$$

- **Růstová** nadprůměrná míra růstu dividendy je způsobena růstovými faktory firmy
- **Přechodná** během této fáze je nadprůměrná míra růstu postupně snižována až na normální míru růstu pro dané odvětví
- **Finální** zde již počítáme s průměrnou mírou růstu dividendy v našem odvětví po celou dobu, tedy v našem případě nekonečnou dobu držby akcie

Ziskové modely

- počítají se samotným ziskem – ziskové m.
- bývají považovány za přesnější než dividendové
- zaměřují se na kratší investiční horizont - přibližně tři roky, maximálně pět let

- 3 základní typy:
 - P/E ratio
 - P/BV ratio
 - P/S ratio

Ziskové modely – P/E ratio

- nejčastější ukazatel
- kolika násobek zisku si člověk cení
- porovnává se v rámci odvětví ne napříč trhem

$$\frac{VH}{E_0} = \frac{D_1/E_1}{(1+r)} + \frac{D_2/E_2}{(1+r)^2} + \dots + \frac{D_n/E_n}{(1+r)^n}$$

$$P/E = \frac{VP_1}{(1+r)} + \frac{VP_2}{(1+r)^2} + \dots + \frac{VP_n}{(1+r)^n}$$

$$VH = E_0 \times \left(\frac{VP_1}{(1+r)} + \frac{VP_2}{(1+r)^2} + \dots + \frac{VP_n}{(1+r)^n} \right)$$

$$VH = E_0 \times \frac{VP(1+g)}{r-g} = E_1 \times \frac{VP}{r-g}$$

Ziskové modely – P/BV ratio

- využívá účetní hodnotu vlastního kapitálu
- P/BV (price-to-book-value ratio)
- očekávaný zisk v příštím roce nahradíme součinem rentability vlastního kapitálu a očekávané účetní hodnoty vlastního kapitálu na akcii
- hodnota vyšší než 1 - tak si investoři cenní akcií více než podílu hodnoty majetku společnosti, který připadá na jednu akcii
- vypovídající schopnost je nízká a často opožděná
- účetnictví jednotlivých společností bývají odlišná

$$VH = P = \frac{E_1 \times VP}{r - g} = \frac{BV_1 \times ROE \times VP}{r - g}$$

Ziskové modely – P/S ratio

- vyjadřuje, jak moc si investor cenní jedné koruny z tržby podniku
- můžeme použít i v případě, kdy podnik dosahuje nízkého zisku či ztráty
- lze lépe porovnávat podniky navzájem

- S_1 - očekávané tržby na akcii v prvním roce

- M_1 - očekávaná zisková marže na akcii v prvním roce

$$VH = \frac{E_1 \times VP}{r - g} = \frac{M_1 \times S_1 \times VP}{r - g}$$

$$\frac{VH}{S} = P/S = \frac{M_1 \times VP}{r - g}$$

Příklad – společnost ČEZ

- České Energetické Závody.
- v roce 2009 byla největší českou firmou podle tržeb
- 32 tisíce zaměstnanců - třetí největší

- Struktura vlastníků:
 - Ministerstvo financí ČR - 69,369 %
 - Ostatní právnické osoby - 4,427 %
 - Fyzické osoby - 5,416 %
 - Správci celkem - 20,788 %

Požadovaná výnosová míra

ČEZ

Bezriziková výnosová míra	r_f	6,55 %
Tržní výnosová míra	r_m	14,25 %
Beta faktor	β	0,73

$$\beta(r_m - r_f) = P_{ri}$$

$$0,73 * (0,1425 - 0,0655) = 0,05621$$

$$r_f + P_{ri} = r$$

$$0,0655 + 0,05621 = 0,12171$$

Míra růstu dividend

Rok	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Dividenda	2,00	2,50	4,50	8,00	9,00	15,0	20,00	40,00	50,00	53,00	50,00
Míra růstu dividendy v %	x	25 %	80 %	78 %	13 %	67%	33%	100%	25 %	6 %	-6%

Aritmetický průměr – 42,06 %

Geometrický průměr – 37,97 %

	Jednot.	Zkratka	Vztahy	2 000	2 001	2 002	2 003	2 004	2 005	2 006	2 007	2 008	2 009	2 010
Dividend a	Kč	D	Bloomberg	2	3	5	8	9	15	20	40	50	53	50
Čistý zisk	mil. Kč	e	Bloomberg	7 237	9 123	8 421	8 869	13 213	21 438	27 697	41 555	46 510	51 547	47 232
Počet akcií	mil	a	Bloomberg	592,0	590,1	590,3	591,5	592,2	589,8	588,8	541,8	533,0	533,4	533,9
Zisk na akcii		E	E=e/a	12,23	15,46	14,27	14,99	22,31	36,35	47,04	76,69	87,25	96,63	88,47
Vlastní kapitál	mil. Kč	VK	Bloomberg	129 442	136 726	143 675	171 075	178 447	191 289	207 653	184 226	185 410	206 675	227 051
Rentabilit a vl. kapitálu	%	ROE	ROE=(e/VK)*100	5,59	6,67	5,86	5,18	7,40	11,21	13,34	22,56	25,08	24,94	20,80
Div. výplatní poměr	%	VP	VP=(D/E)*100	16,36	16,17	31,54	53,35	40,34	41,27	42,51	52,16	57,30	54,85	56,52
Podíl zadržené ho zisku	%	b	b=100-VP	83,64	83,83	68,46	46,65	59,66	58,73	57,49	47,84	42,70	45,15	43,48
Míra růstu dividend	%	g	g = ROE *b	4,68	5,59	4,01	2,42	4,42	6,58	7,67	10,79	10,71	11,26	9,05

Míra růstu dividend

Způsob výpočtu	Míra růstu dividend
Aritmetický průměr	42,06 %
Geometrický průměr	37,97 %
Udržovací růstový model	7,25 %

DDM

ČEZ		
Požadovaná výnosová míra	r	12,171 %
Míra růstu dividend	g	7,25 %
Poslední vyplacená dividend	D	50

Použijeme Gordonův model

$$VH = \frac{D_0(1 + g)}{r - g} = \frac{50(1 + 0,0725)}{0,12171 + 0,0725} \doteq 1089,72$$

Ziskové modely – P/E ratio

$$VP = \frac{D}{E} = \frac{50}{88,47} = 0,57$$

ČEZ		
Požadovaná výnosová míra	r	12,171 %
Míra růstu dividend	g	7,25 %
Dividendový výplatní poměr	VP	0,57
Očekávaný zisk v dalším roce	E ₁	81,02

$$VH = E_1 \times \frac{VP}{r - g} = E_1 \times P/E$$

$$VH = 81,02 \times \frac{0,57}{0,12171 - 0,0725} = 81,02 \times 11,485 = 930,53$$

Ziskové modely P/BV ratio

ČEZ

Požadovaná výnosová míra	r	12,171 %
Míra růstu dividend	g	7,25 %
Dividendový výplatní poměr	VP	0,57
Rentabilita vlastního kapitálu	ROE	20,8 %
Očekávaná účetní hodnota vlastního kapitálu na akcii	BV	431,43

$$VH = \frac{BV_1 \times ROE \times VP}{r - g} = BV_1 \times P/BV$$

$$VH = \frac{431,43 \times 0,208 \times 0,57}{0,12171 - 0,0725} = 431,43 \times 2,389 = 1030,77$$

Shrnutí

ČEZ

Současný kurz	787,5 Kč	
DDM	1089,72 Kč	KOUPIT
Ziskový model – P/E ratio	930,53 Kč	KOUPIT
Ziskový model – P/BV ratio	1030,77 Kč	KOUPIT
Doporučení	KOUPIT	

9. Metody technické analýzy

Technická analýza

- analytický přístup zabývající se vývojem kurzů cenných papírů či cen komodit. Je to způsob rozhodování o koupích a prodeích finančních instrumentů na základě minulého vývoje jejich tržních cen a objemu obchodů.
- postavena na analýze publikovaných tržních dat, kterými jsou kurzy nebo objemy obchodů.
- Hlavním cílem je prognózování krátkodobých pohybů akcií nebo akciových indexů. Zaměřuje se na změny tržních cen jako na indikátory nabídky a poptávky.
- je založena na předpokladu, že všechno, co potřebuje investor vědět, je obsaženo v tržních cenách. Dává nám doporučení, kdy provést obchod.

Základní předpoklady technické analýzy

- Tržní ceny odrážejí a zahrnují veškeré informace.
- Ceny se pohybují v trendech.
- Historie se opakuje.

DOW THEORY

- odvozena z článků, které publikoval Charles H. Dow ve Wall Street Journal v letech 1900 až 1902 .
- Podle Ch. Dowa se většina akcií chová podobným způsobem. Nemusíme tedy zkoumat každou akcii zvlášť, ale situaci na trhu by měla dostatečně popisovat průměrná tržní cena akcií.
- Vytvořil indexy: Dow-Jones-Rail-Average (DJRA) a Dow-Jones-Industrial-Average (DJIA)

Principy Dowovy teorie I.

- Akciové indexy v sobě zahrnují všechny relevantní informace.
- Pohyby akciových kurzů lze rozložit na tři základní trendové pohyby, kterými jsou primární (1 rok a déle), sekundární (3 týdny až 3 měsíce) a terciární trend (méně než 3 týdny).
- Primární trendy obsahují tři fáze – akumulární, rostoucí a fáze distribuce.
- Akciové indexy se musí navzájem potvrzovat. To, co je vyrobeno, musí být též dopraveno, proto by měl být vývoj obou indexů (DJIA a DJTA) stejný.
- Objem obchodů musí potvrzovat trend. Nastoupený trend na trhu je potvrzen, pokud ho doprovází rostoucí objem obchodů. Oproti tomu klesající objem obchodů naznačuje pravděpodobnou změnu trendu.
- Nastoupený trend trvá až k jasné změně trendu.

Principy Dowovy teorie II.

- Budoucí vývoj kurzů (pokračování trendu nebo jeho změnu) lze odvodit z minulé tržní situace. Býčím, neboli rostoucím trendem (Bull Trend) nazýváme trend, jehož každý vrchol je vyšší než předchozí, a taktéž každé dno je vyšší než dno předchozí. Naopak, pokud je každý vrchol nižší než vrchol předchozí a každé dno je nižší než dno předchozí, mluvíme o medvědím neboli klesajícím trendu (Bear Trend). Změna trendu z býčího na medvědí nastane, pokud nový vrchol nedosáhne úrovně předchozího vrcholu a nové dno leží níž než předchozí dno. Jestliže byl doposud na trhu medvědí trend a nový vrchol leží výše než předchozí vrchol, zatímco nové dno je výš než předchozí dno, značí to změnu trendu na býčí trend.

Kritika Dowovy teorie

- Signály k nákupu a prodeji přicházejí příliš pozdě a můžou být falešné nebo nejednoznačně interpretovatelné.
- Pomáhá pouze při analýze primárního trendu, přičemž na sekundárním a terciárním trendu lze též docílit vysokých zisků.
- Soustředí se pouze na trh jako celek, a proto není schopna umožnit výběr jednotlivých akcií.

Metody technické analýzy

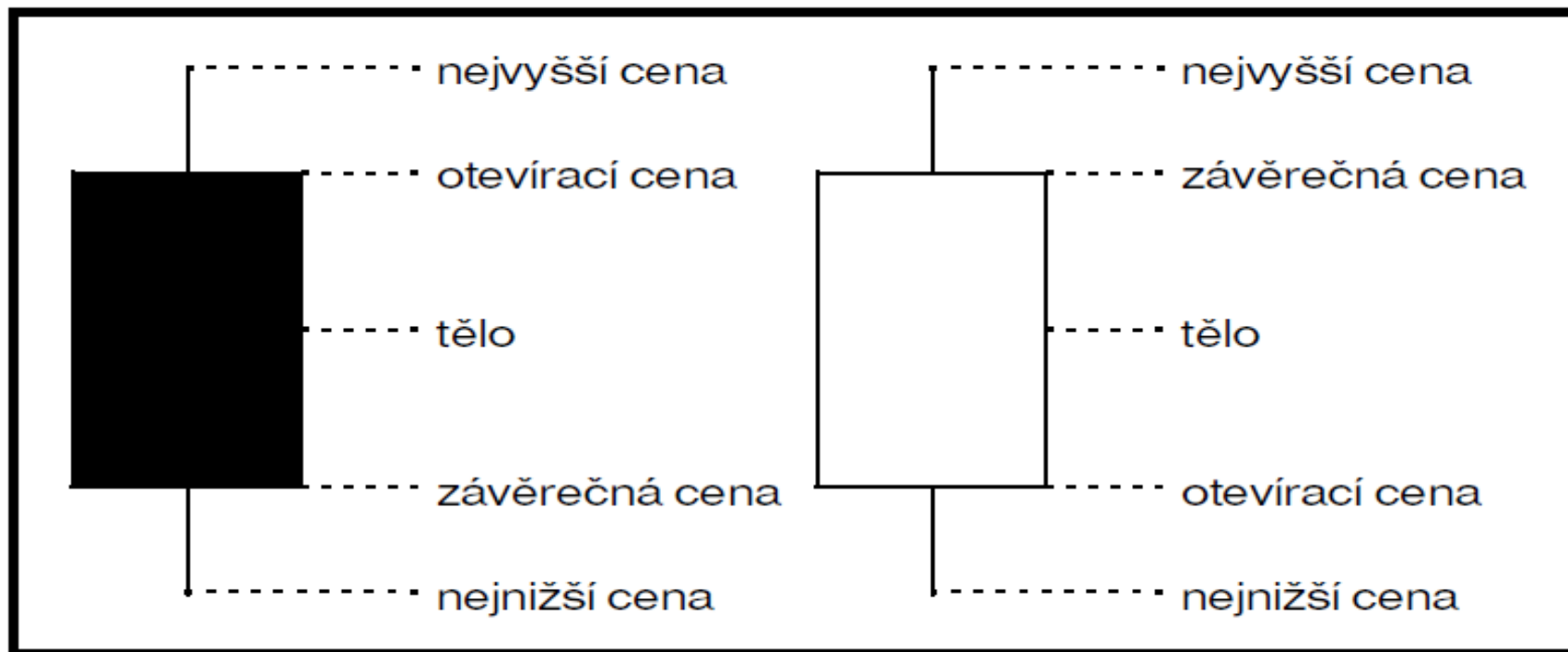
- Grafické metody – hledají opakující se formace, které vznikají na grafech ceny nebo objemu akcie.
- Technické indikátory – Indikátor je funkce času, vektoru parametrů, historických cen a objemů, která konkrétním hodnotám přiřazuje vektor reálných čísel.

Grafické metody

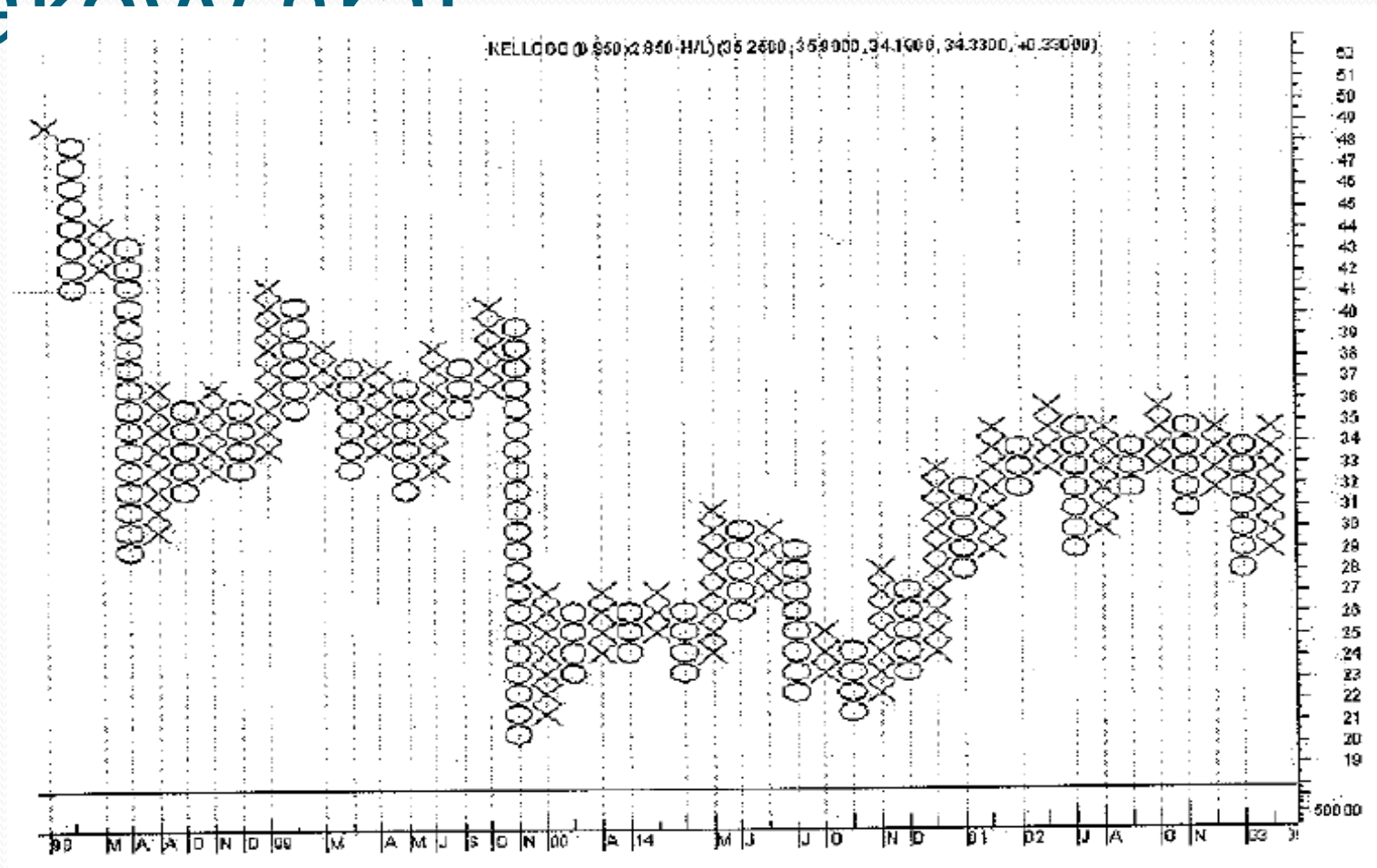
- Grafy – pomocné nástroje technické analýzy
 - Čárový graf
 - Úsečkový graf
 - Graf typu svíce
 - Graf objemu obchodů
 - Znakový graf (Point & Figure)

Grafický znak typu svíce

(Japonské svíčky)



Znakový graf



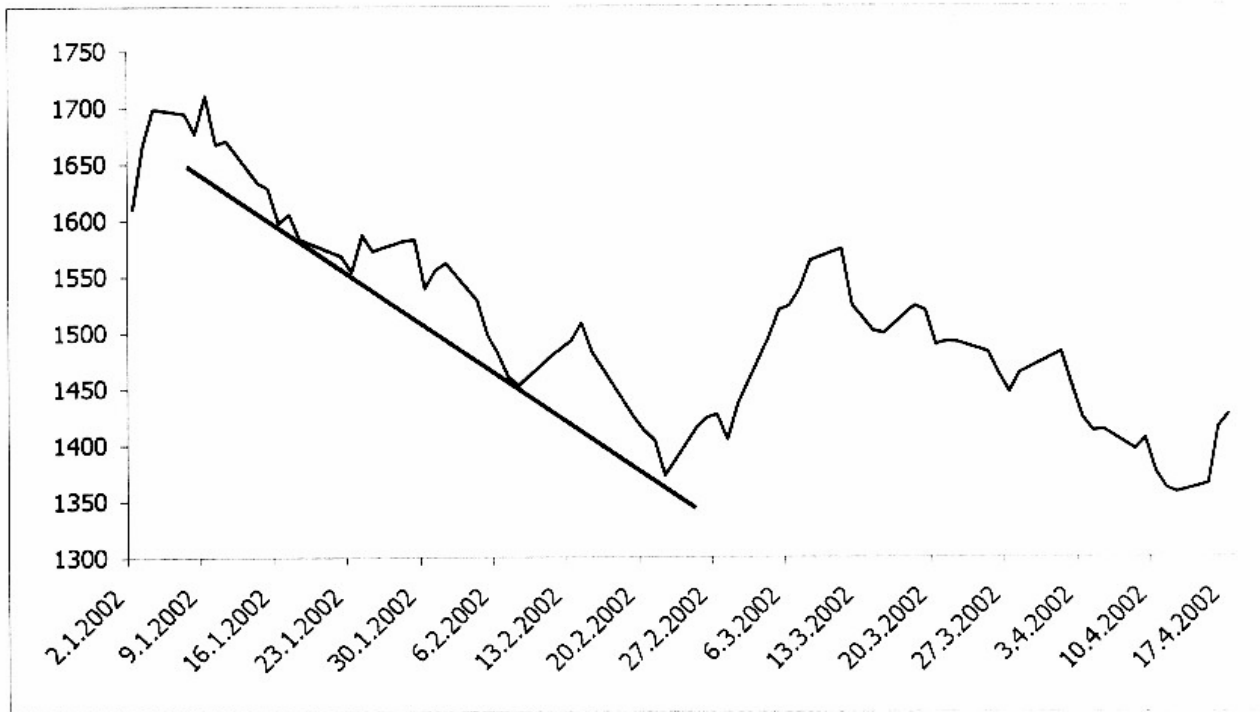
Grafické metody a formace

- Hranice podpory (Support Level) a hranice odporu (Resistance Level)
- Konsolidační formace (signalizují pokračování původního trendu po jeho dočasném přerušení) – trojúhelníky, vlajky, praporky a klíny, atd.
- Reverzní formace (signalizují změnu trendu) – hlava a ramena, vrcholy, dna, atd.
- Trendové kanály
- Mezery

Hranice podpory

- Podpora je taková úroveň ceny akcie, při které je poptávka tak silná, že zastaví pokles ceny.
- Linie podpory znázorňuje hladinu, od které by se měl kurz cenného papíru odrazit směrem vzhůru.
- Pokud kurz linií podpory propadne a vzroste přitom objem, je to známka změny trendu. Prolomená linie podpory se pak stává linií odporu.

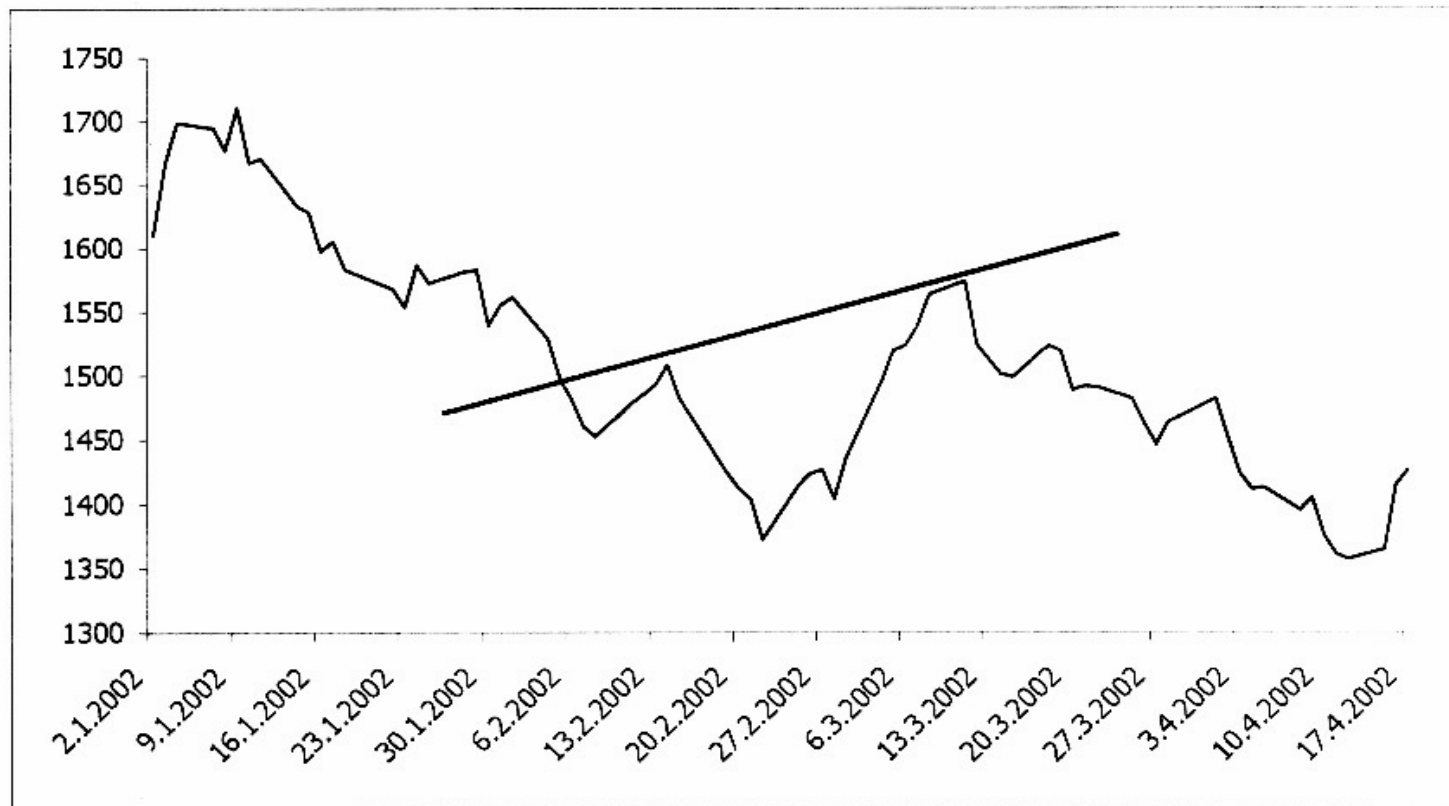
Hranice podpory



Hranice odporu

- Odpor je úroveň ceny akcie, při které je nabídka taková, že zastaví růst ceny.
- Linie odporu znázorňuje hladinu, od které by se měl kurz cenného papíru odrazit směrem dolů.
- Pokud je linie odporu prolomena za rostoucího objemu, je to známka změny trendu a linie odporu se obvykle mění v linii podpory.

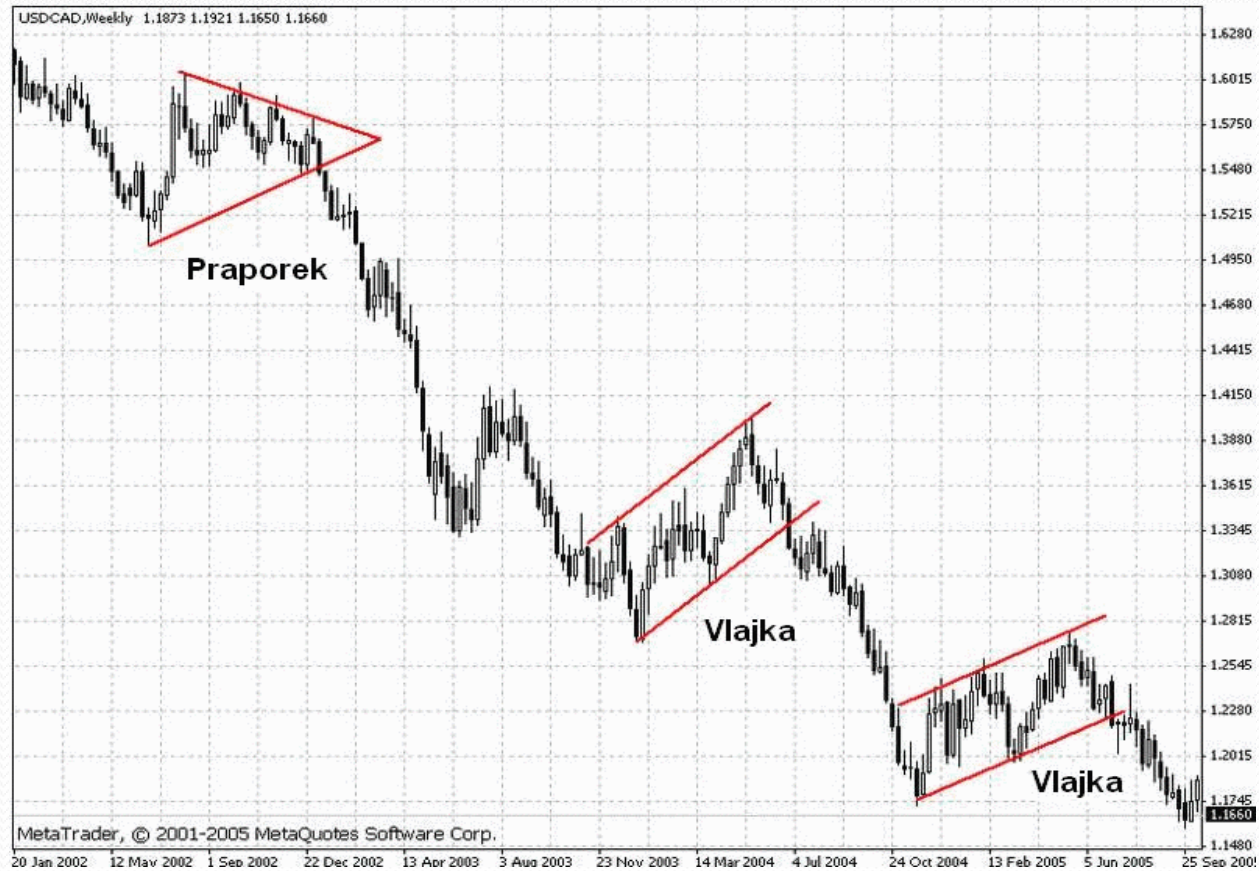
Linie odporu



Trojúhelník

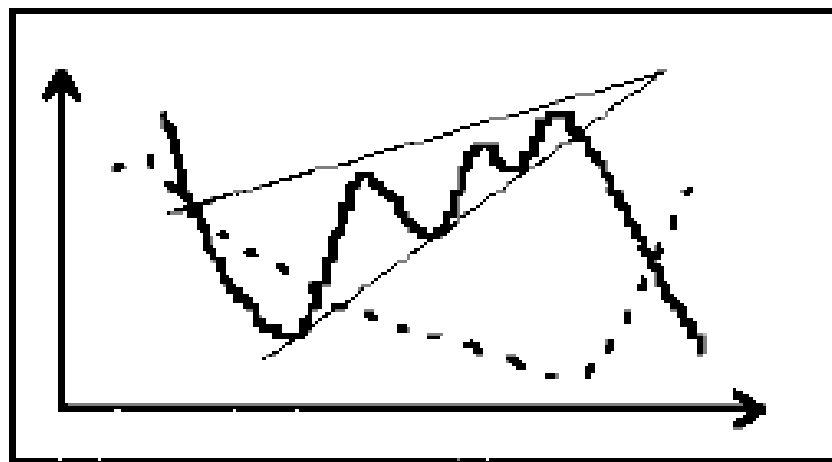


Vlajka a praporek

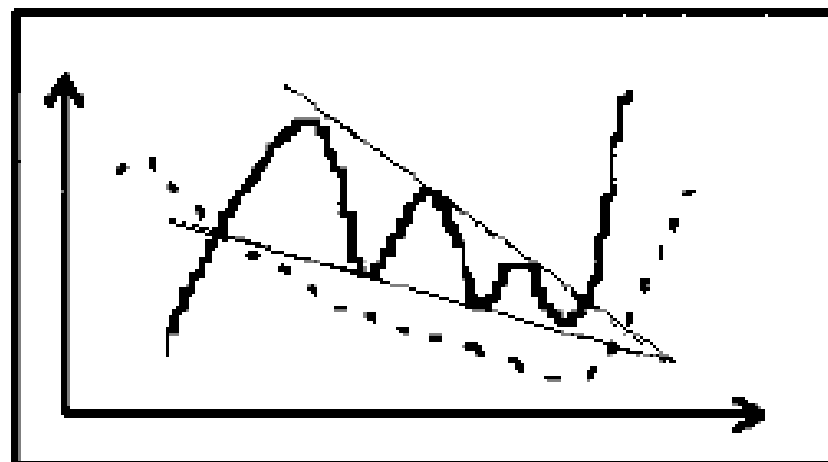


Klíny

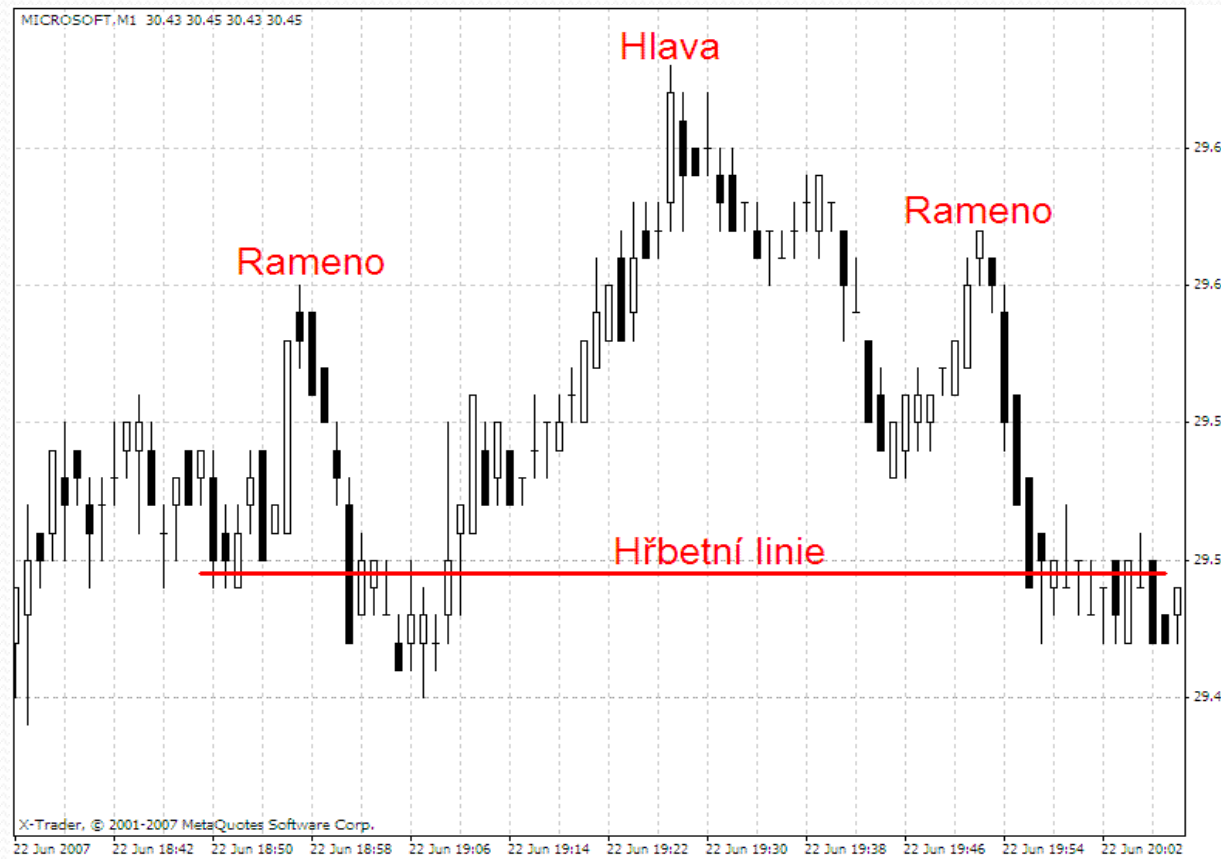
vzestupný klín



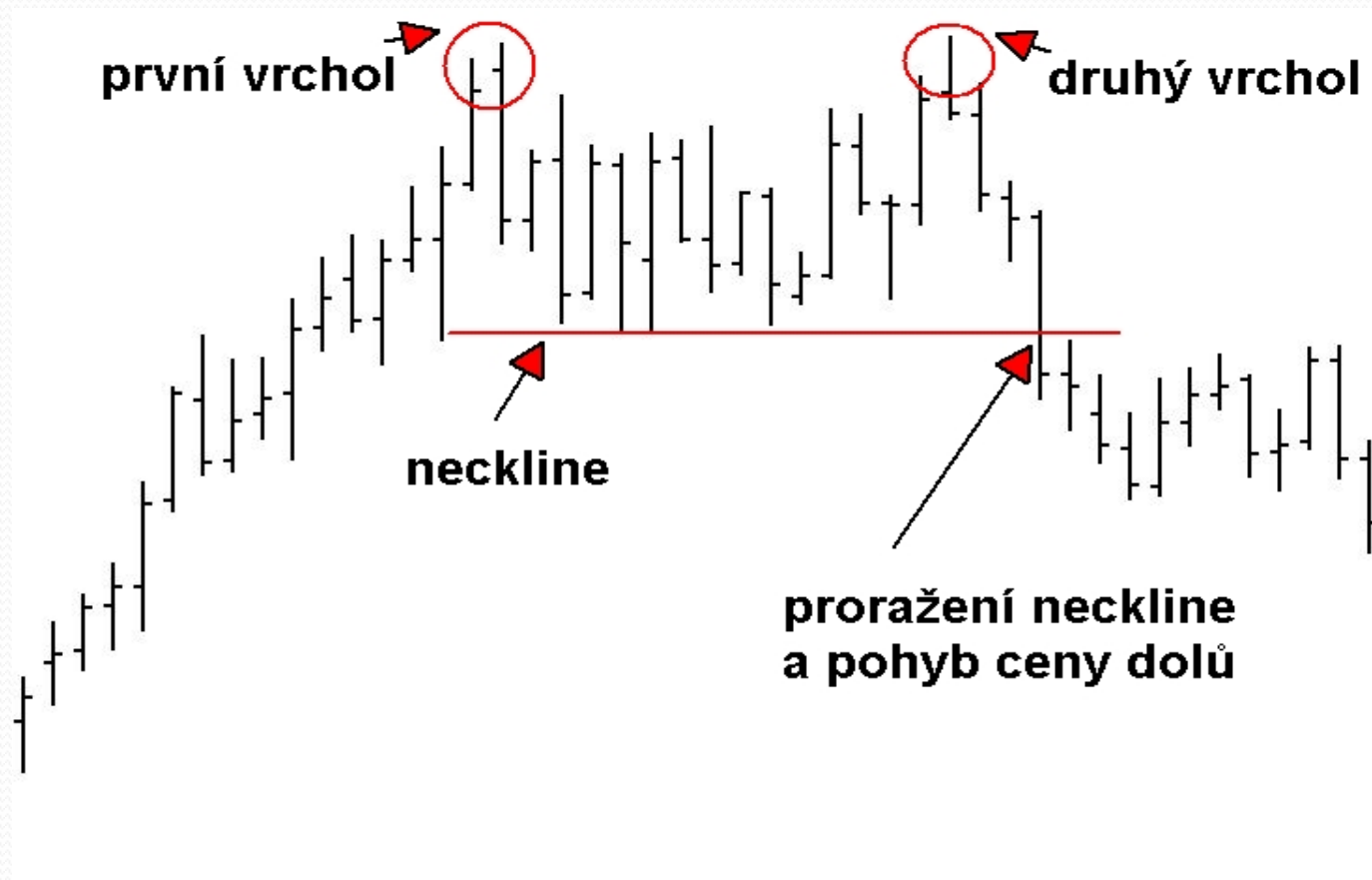
sestupný klín



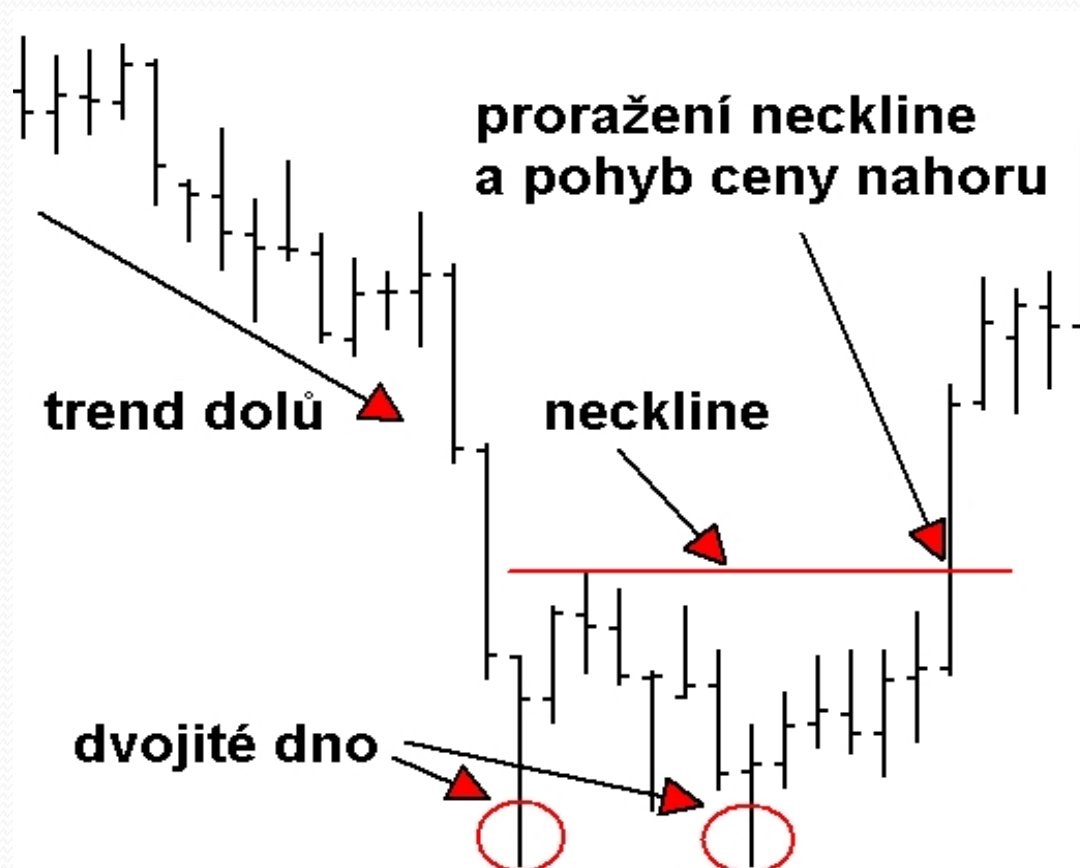
Hlava a ramena



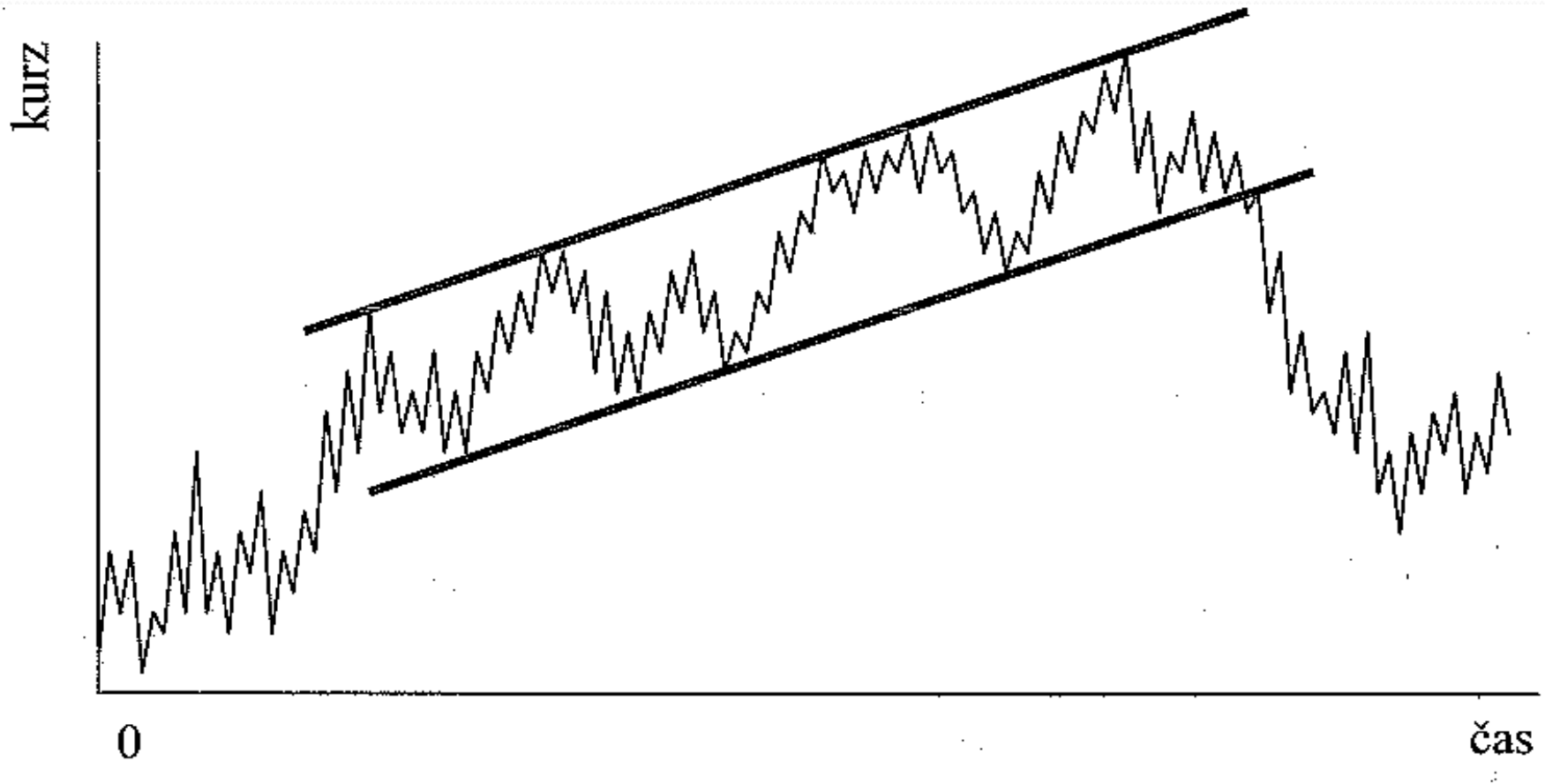
Dvojitý vrchol



Dvojité dno



Trendový kanál



Mezery I.



Mezery II.



Mezery III.



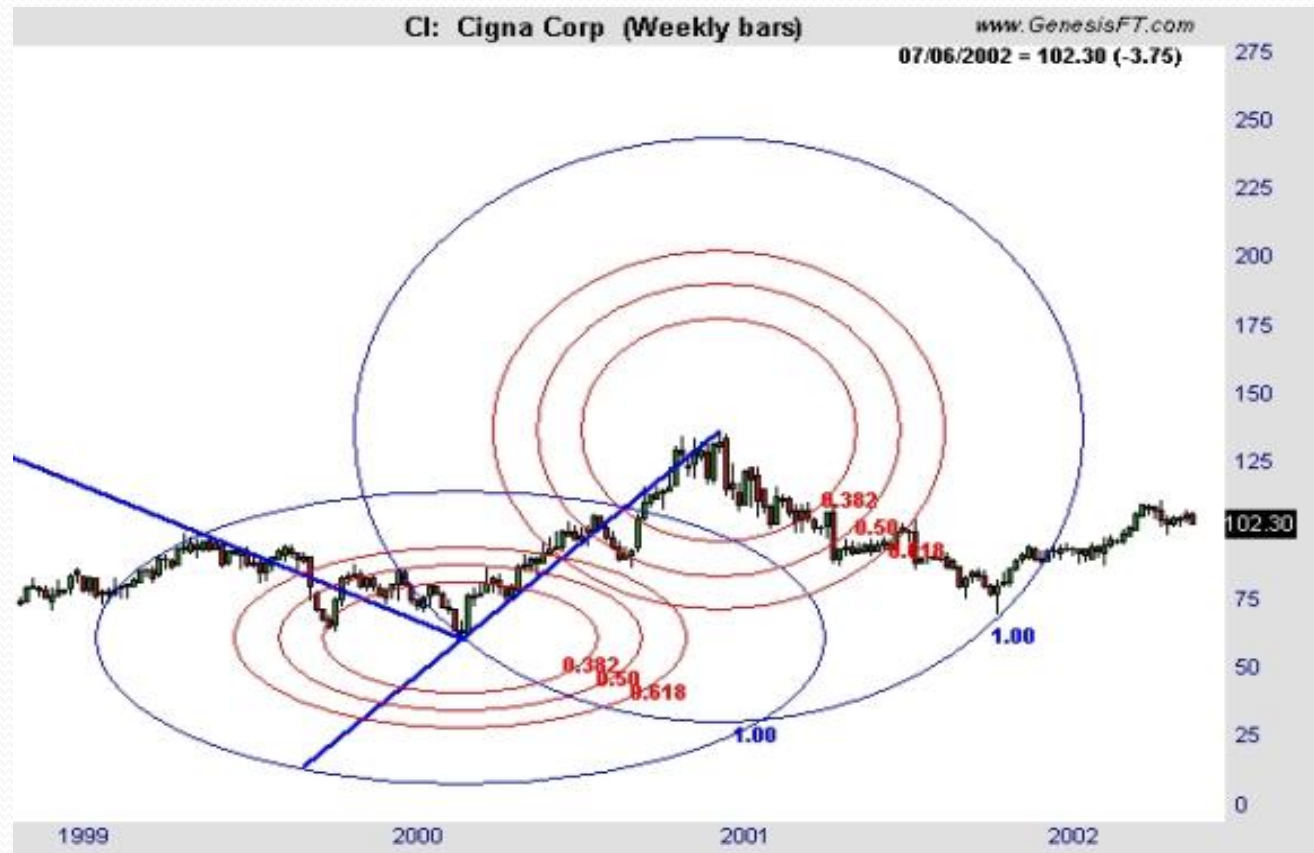
Mezery IV.



Fibonacciho studie

- Fibonacciho řada se skládá z členů, kde každý následující člen je roven součtu dvou předchozích: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55,...
- Následující člen je vždy přibližně 1,618 násobek předcházejícího a zároveň 0,618 násobek členu následujícího.
- Fibonacciho oblouky – zakreslení trendové linie mezi dva extrémní body a dále zobrazení tří oblouků se středem v druhém extrému o poloměrech 38,2%, 50,0% a 61,8% délky trendové linie.
- Cílem je najít úrovně podpory a odporu v blízkosti Fibonacciho oblouků.
- Používají se i Fibonacciho vějíře, hladiny návratu a časové zóny jako svislé linie zobrazené v intervalech rovnajících se jednotlivým hodnotám Fibonacciho čísel.

Fibonacciho oblouky



Technické indikátory

- Klouzavé průměry a metody na nich založené
- Oscilátory
- Objemové indikátory
- Sentiment indikátory
- Indikátory šíře trhu

Klouzavé průměry a metody na nich založené

- „trend-following metody“ – svými signály a doporučeními zpravidla následují trend.
- Nejlépe fungují v dobře trendujících trzích.
- Nejčastějším vyhodnocováním signálů je sledování situace, kdy se indikátor protne s cenou.
- Protne-li indikátor cenu zezdola nahoru – signál k nákupu.
- Protne-li indikátor cenu shora dolů – prodejní signál.

Klouzavé průměry a metody na nich založené

- Druhy klouzavých průměrů se liší podle váhy přiřazované údajům různého stáří:
 - Jednoduchý
 - Vážený
 - Exponenciální
 - Trojúhelníkový
 - Proměnlivý
- Metody založené na klouzavých průměrech:
 - MACD
 - Obálky
 - Bollingerovy pásy
 - Procentní pásy
 - Klouzavá regrese

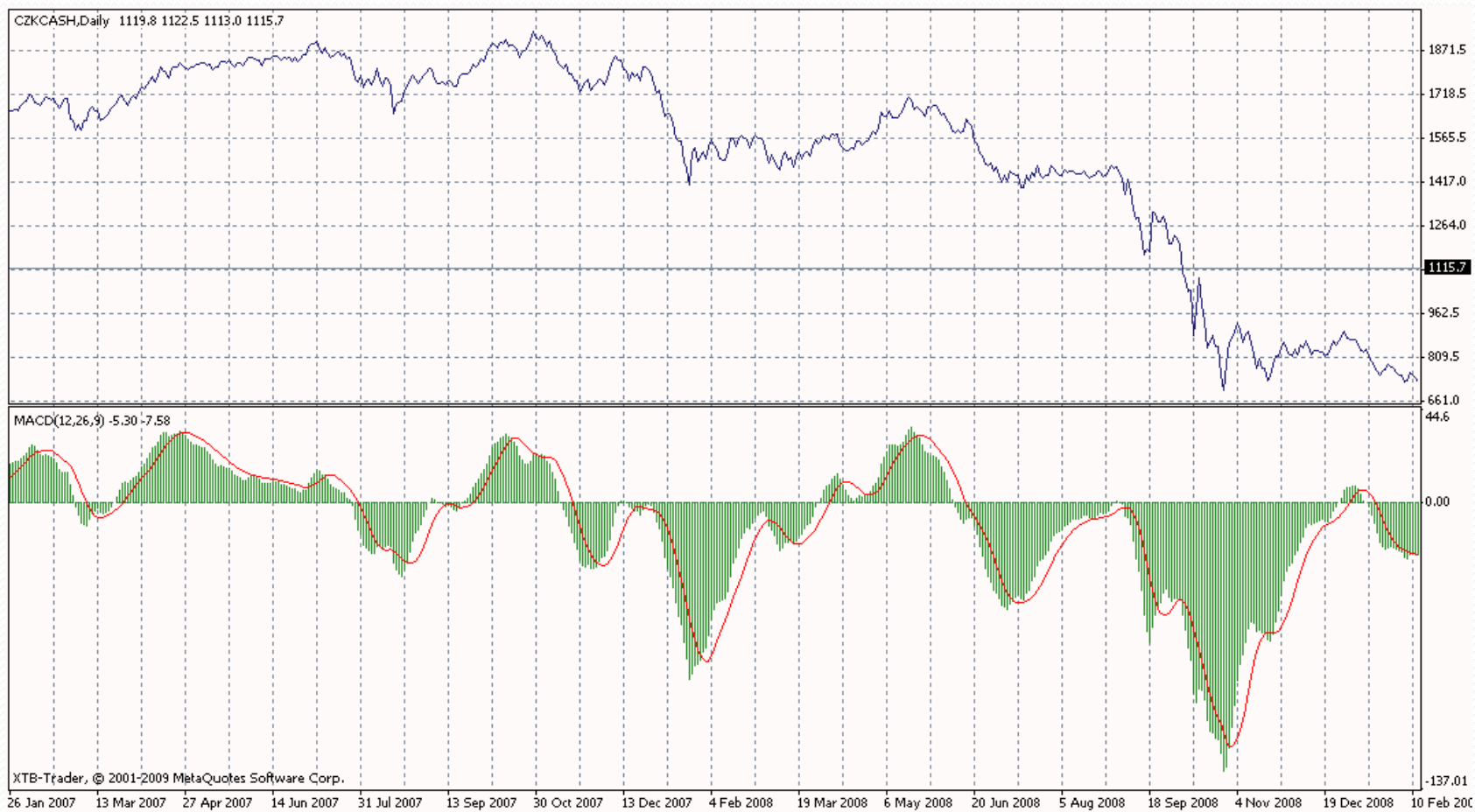
Jednoduchý klouzavý průměr



MACD

- Moving Average Convergence Divergence
- je tvořen dvěma křivkami. První křivku dostaneme po odečtení dlouhodobého exponenciálního klouzavého průměru od krátkodobého exponenciálního klouzavého průměru a označujeme ji jako MACD. Druhou, nebo-li signální křivku získáme vyhlazením MACD pomocí dalšího exponenciálního klouzavého průměru.

MACD



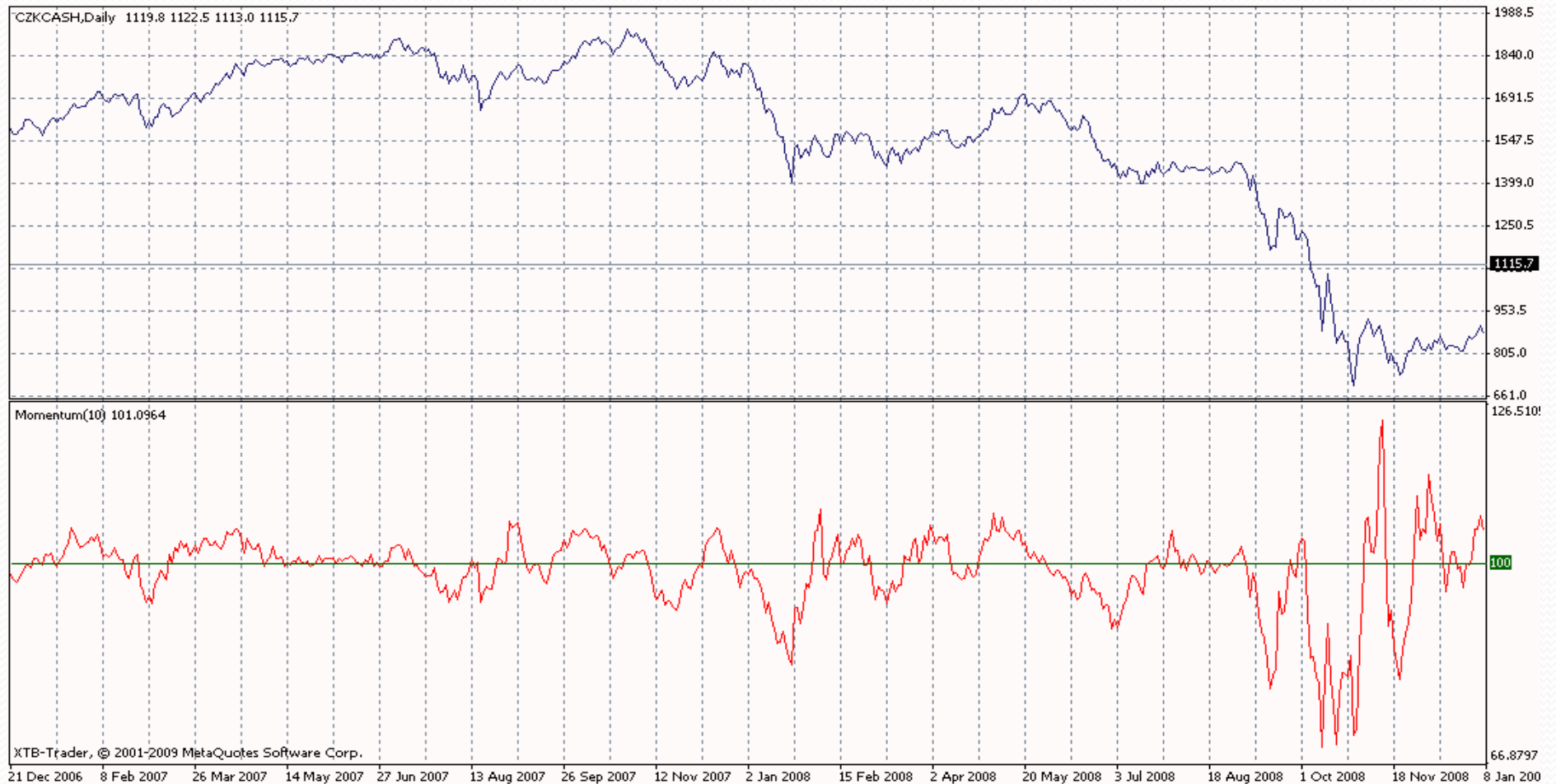
Oscilátory

- Hodnota kolísá zpravidla buď kolem nějaké úrovně nebo v rámci pásma.
 - Momentum
 - Price Rate of Change
 - Moving Average Spread (MAS)
 - Cenový oscilátor (Price Oscillator)
 - Trix
 - Index relativní síly (RSI)
 - Stochastik

Momentum

- Zjišťuje velikost změny kurzu za určité období, čímž měří zrychlení nebo zpomalení trendu.
- Hodnoty indikátoru se pohybují kolem oscilační linie 0 v případě absolutního Momenta nebo kolem oscilační linie 1 (popř. 100) v případě relativního Momenta.
- Protne-li oscilační linii zezdola nahoru – signál k nákupu.
- Protne-li oscilační linii seshora dolů – signál k prodeji.
- Na silně trendujícím trhu může vysílat ukvapené signály.

Momentum

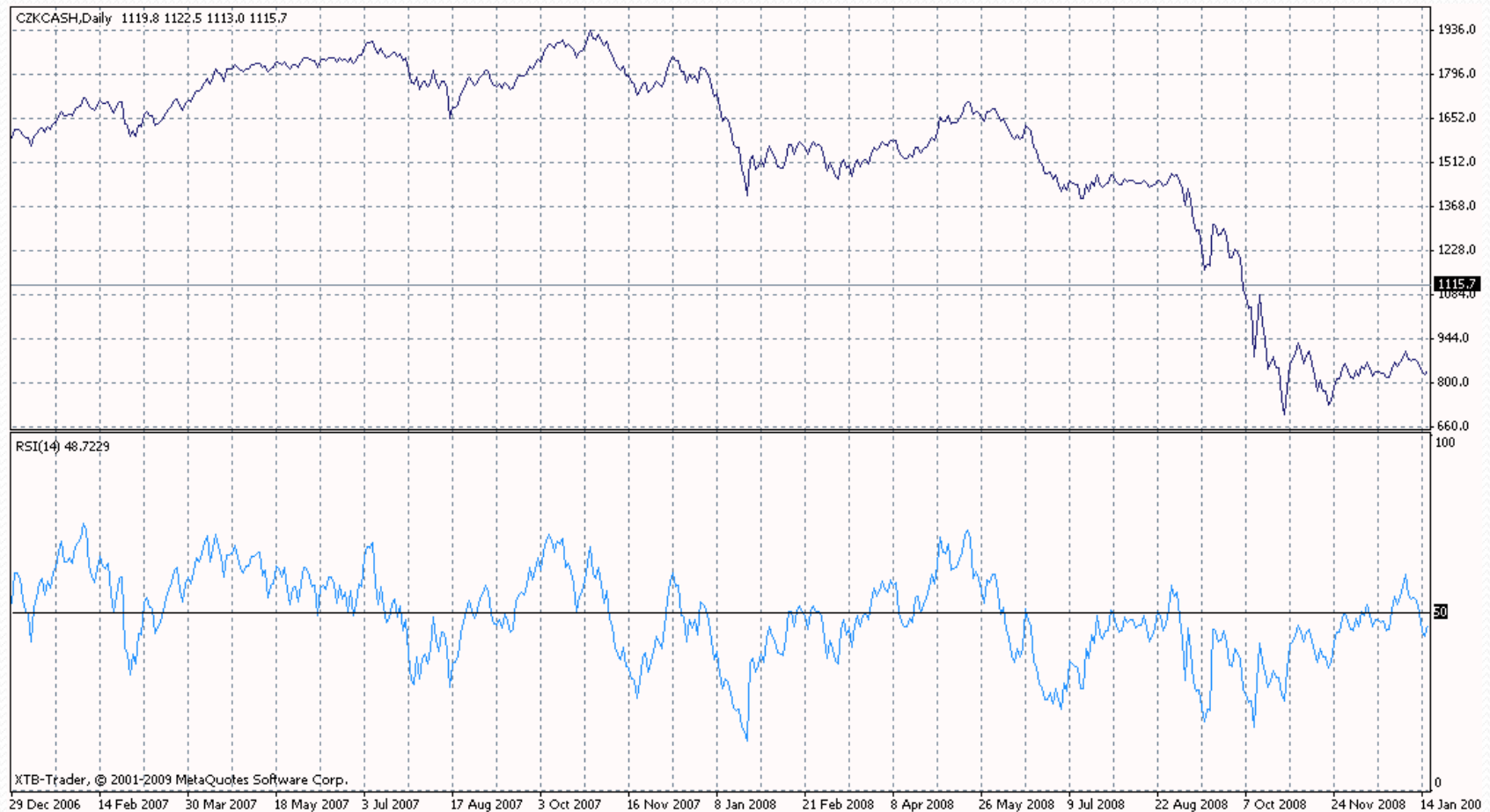


Index relativní síly (RSI)

$$RSI = 100 - \left(\frac{100}{1 + RS} \right)$$

- kde RS je podíl průměrných kladných změn v kurzu a průměrných záporných změn v kurzu během stanovené časové periody.
- RSI vyjadřuje vnitřní sílu jednotlivého cenného papíru.
- Hodnoty RSI se pohybují v intervalu 0 až 100.
- Dá se dobře využít k analýze překoupeného a přeprodaného trhu.

Index relativní síly (RSI)



Objemové indikátory

- Pracují s údaji o objemech obchodů, které jsou doplňovány údaji o vývoji kurzů.
- Pokud objem obchodů roste, značí to větší aktivitu investorů. Při upadající aktivitě na trhu je málo obchodníků ochotných za tuto cenu obchodovat, tudíž můžeme očekávat změnu trendu.
- Využití především na trzích s malou likviditu.

Objemové indikátory

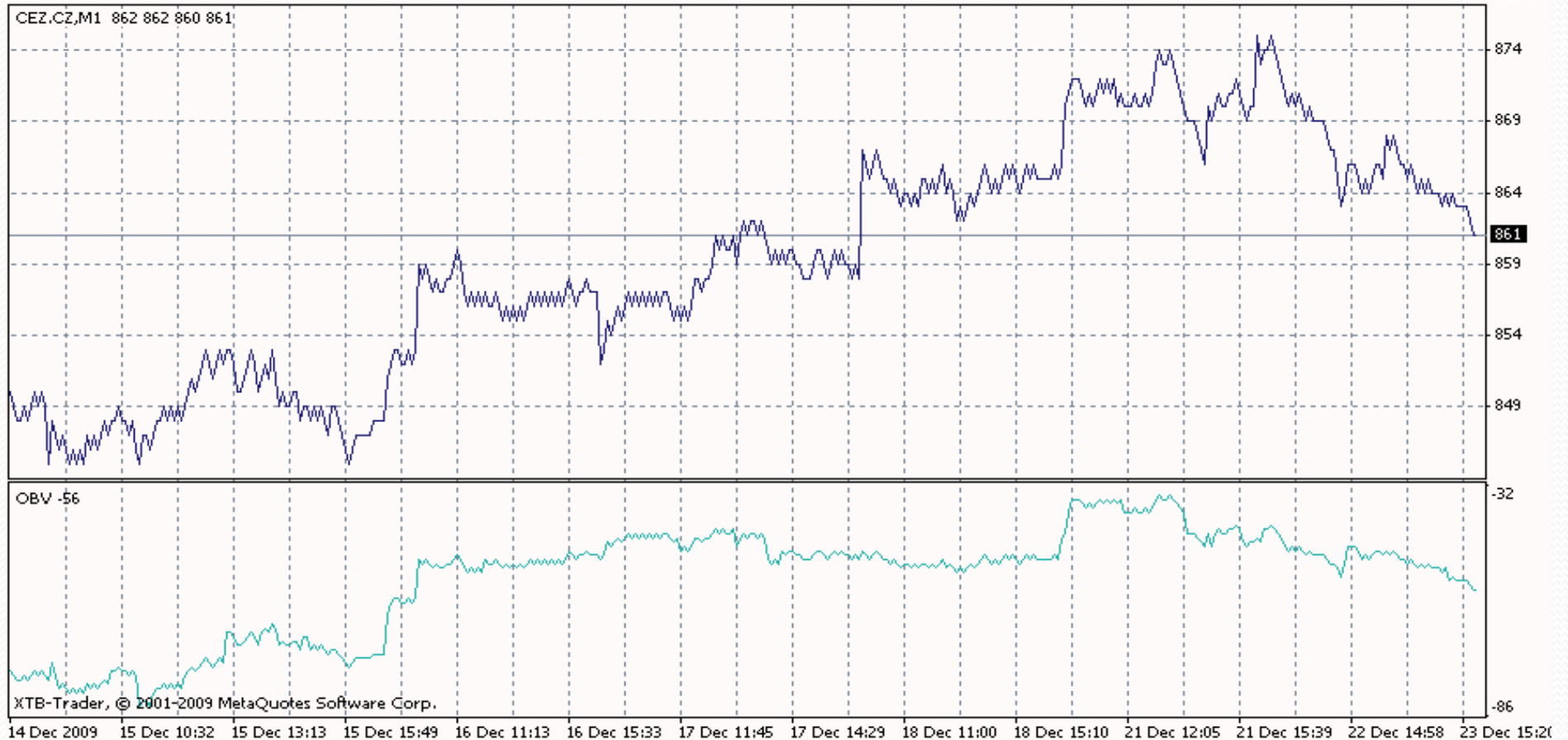
- On Balance Volume (Bilance objemu)
- Price and Volume Trend (PVT)
- Volume Rate-Of-Change (ROC)
- Volume Oscillator (Oscilátor objemu)
- Indexy PVI a NVI (Volume indexes)
- Volume Accumulation/ Distribution Indicator (AD)
- Chaikinův oscilátor
- Money Flow Index (MFI)

On Balance Volume

$$OBV_t = \sum_{p_t > p_{t-1}} v_t - \sum_{p_t < p_{t-1}} v_t$$

- kde p je kurz akcie a v je objem obchodů.
- Vychází z předpokladu, že na trhu obchodují dva typy investorů – smart money (profesionální investoři) a general public.
- Nákupní signál – indikátor začne stoupat, zatímco cena klesá.
- Prodejní signál – indikátor začne klesat, zatímco cena roste.
- Indikátor, který předbíhá trend.

On Balance Volume



Sentiment indikátory

- Mají blízko k psychologické analýze.
- Předmětem jejich zkoumání jsou nálady, očekávání a mínění investorů.
- Investoři jsou ovládáni optimismem či pesimismem a podle toho, která nálada na trhu převládá, kurz roste nebo klesá.
- Podle výkladu se dělí na:
 - Anticyklické
 - Cyklické

Anticyklické sentiment indikátory

- Snaží se zachytit chování široké investorské veřejnosti, která své obchody většinou uzavírá v nevhodné tržní situaci a se zpožděním.
- Proto by investor měl jednat opačně, než signalizuje anticyklický indikátor.
 - Odd-lot Theory
 - Short Sales Ratio
 - Doporučení investičních poradců
 - Put/Call Ratio

Cyklické sentiment indikátory

- Snaží se zmapovat chování profesionálních investorů, kteří představují obzor úspěšného investorského chování.
- Doporučení – chovat se v souladu s indikátory.
 - Barron's Confidence index (BCI) neboli Barronův index důvěry
 - Struktura portfolia fondů

Indikátory šíře trhu

- Sledují kvantitativní pohyb celého trhu, a to na základě údajů o počtu akcií, které klesly a počtu akcií, které stouply.
- Snaha zmapovat výkonnost vybrané akcie, odvětví či jistého tržního segmentu – vždy však k relaci k jinému odvětví či segmentu.
- Dobré pro předpověď pravděpodobné extrémní změny trendu, kterou odhalí s dostatečným předstihem.
- Kvůli svému pohledu na celkový trh nejsou určeny pro analyzování jednotlivých akcií.

Indikátory šíře trhu

- Advance/ Decline Line (A/D)
- Advance-Decline-Ratio
- Advance-All-Ratio
- McClellanův Sumation index (MSI)
- McClellanův oscilátor
- Relativní síla

Advance/ Decline Line (A/D)

$$ADL_t = ADL_{t-1} + (AI - DI)$$

- vyjadřuje číselně kumulativní rozdíl mezi počtem emisí akcií, jejichž kurz stoupá a klesá na dané množině akcií (báze indexu) v čase.
- reaguje na změnu podmínek dříve než akciový index.
- Tržní index roste, zatímco A/D Line již klesá – signál k prodeji.
- Tržní index klesá, zatímco A/D Line již roste – signál k nákupu.

Advance/ Decline Line (A/D)



Advance-Decline-Ratio (ADR)

$$ADR_t = ADR_{t-1} + \left(\frac{AI}{DI} \right)$$

- vyjadřuje číselně kumulativní podíl mezi počtem emisí akcií, jejichž kurz stoupá a klesá na dané množině akcií (báze indexu) v čase.
- reaguje na změnu podmínek dříve než akciový index.
- Tržní index roste, zatímco A/D Line již klesá – signál k prodeji.
- Tržní index klesá, zatímco A/D Line již roste – signál k nákupu.

10. Oceňování nemovitého majetku



Oceňování majetku představuje soubor činností, kdy je určitému předmětu nebo souboru předmětů přiřazována určitá peněžní hodnota

Účel ocenění

- Převod, přechod, dělení nebo navyšování vlastnictví
- Financování a úvěrování
- Škody na majetku
- Účetnictví
- Daně a poplatky
- Investiční, arbitrážní a tržní poradenství
- Pojišťovnictví

Oceňování majetku v ČR

- Oceňování administrativní ze zákona č. 151/1997 Sb., o oceňování majetku, prováděcí vyhlášky Ministerstva financí
- Oceňování tržní

Administrativní ocenění

- Založeno na přesně definovaných postupech a krocích
- Slouží účelům a potřebám státní správy: daňové účely, ocenění majetku investičních a penzijních fondů, vyvlastnění apod.
- Zajištění spravedlivého ocenění
- Striktně dáno zákonem, není zde žádný prostor pro individuální přístup
- Výsledná cena při řádném dodržení metodiky by měla být jednoznačná a pokud možno jediná

Negativum úředního oceňování

Vzniká cena uměle vytvořená, která se skutečnou objektivní hodnotou majetku přijímanou trhem má jen málo společného!

Tržní oceňování

- Systematický, ale zároveň individuální tvůrčí proces, spočívající v hledání cenotvorných faktorů, v jejich analýze a následném vážení všech vlivů, které na hodnotu věci působí
- Neurčujeme cenu, ale pouze její ODHAD!
- Výběr metodiky ocenění je čistě na odborném a zodpovědném uvážení odhadce nebo soudního znalce

Postup při oceňování

- Zadání – přesná charakteristika předmětu a účelu ocenění, datum, místo, kde se nemovitost nachází, identifikace zadavatele, cena posudku a datum jeho plnění
- Ověření existence nemovitosti
- Provedení místního šetření
- Určení a použití oceňovacích metod
- POSUDEK

Příklad posudku

http://www.mmdrazby.cz/data/auctions/8/znalecky-posudek-varvazov_132670241932.040.pdf

Základní dokumenty potřebné k ocenění nemovitosti

- Výpis z katastru nemovitostí
- Kopie příslušné části katastrální mapy
- Výpisy z pozemkové knihy
- Cenová mapa pozemků - pokud je v dané obci vypracovaná
- Výkresová dokumentace
- Stavební povolení, územní rozhodnutí a projektová dokumentace k němu
- Kupní, nájemní a další převodní smlouvy
- Smlouvy či doklady o správě, službách, pojištění, odpisech, nákladech na opravy a údržbu, daních
- Fyzická i právní břemena vztahující se k nemovitosti, přehled a současný stav úvěrů
- Fotodokumentace
- Výsledky místního ohledání nemovitosti provedeného osobně odhadcem

Základní metody oceňování nemovitého majetku

- Nákladová
- Výnosová
- Porovnávací

- Jejich kombinace

Nákladová metoda

- Vychází z vynaložených nákladů na výrobu či sestavení dané věci
- Založena na fyzických a technických vlastnostech oceňovaného předmětu
- Historicky nejstarší a nejpracnější způsob oceňování
- Kolik by stálo postavení této nemovitosti dnes? (započítává se materiál i práce)
- Opotřebení

Metody založené na nákladovém principu

- Individuální cenová kalkulace
- Podrobný položkový rozpočet
- Použitím technicko-hospodářských ukazatelů
- Podle vyhlášky č. 504/2002

Zjištění věcné hodnoty pomocí THU

- Zjistíme základní tabulkovou cenu ZC z katalogu THU
- Vypočítáme základní upravenou cenu ZCU a obestavený prostor

$$ZCU = ZC \times K_v \times K_{pod} \times K_z \times K_m \times K_d$$

$$RC = ZCU \times OP$$

$$VH = RC - OPOTŘ$$

Výnosová metoda

- Ekonomický pohled na nemovitost
- Vlastníkovi nemovitosti patří i veškeré požitky z vlastnictví nemovitosti
- vychází z hrubého nájemného, které je potřeba dále snížit o náklady na jeho dosažení
- Čisté příjmy se diskontují na současnou hodnotu kapitalizační mírou

Věčná renta

- Konstantní výnosy po dlouhou dobu

$$C_v = \frac{Z}{i}$$

Dočasná renta

- V určitém časovém období lze předpokládat totální opotřebení nemovitosti

$$C_v = z \times \left(\frac{q^n - 1}{q^n \times i} \right)$$

Konstantní výnos po určitou dobu, s prodejem na konci

- Je potřeba odhadnout cenu nemovitosti za n let

$$C_v = z \times \left(\frac{q^n - 1}{q^n \times i} \right) + \frac{R}{q^n}$$

Pomocí diskontovaných peněžních toků

- Kalkulováno s příjmy a výdaji

$$C_v = \left(\sum_{t=1}^{n-1} \frac{ZD_t}{q^{t-1}} \right) + \frac{\text{zbytek v roce } n}{q^n}$$

Míra kapitalizace

- Významně ovlivňuje výslednou hodnotu nemovitosti
- Výše úroku je přímo úměrná riziku, které sebou přináší daná investice
- Její určení závisí pouze na úvaze odhadce či soudního znalce

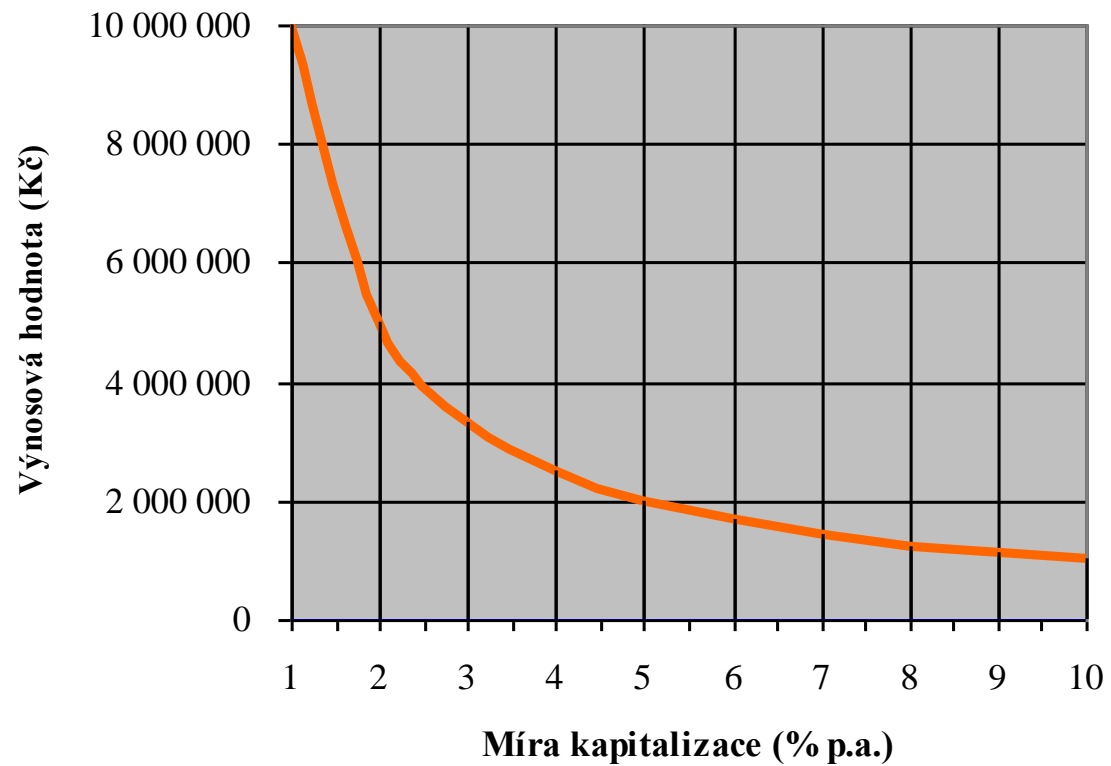
Jak určit míru kapitalizace

- Odvozením od úrokové sazby v bankovních institucích
- Zjištěním z již realizovaných prodejů staveb, které jsou následně pronajímány
- Použitím míry kapitalizace podle cenového předpisu

Míra kapitalizace podle cenového předpisu

Číslo položky	Název položky	Míra kapitalizace %
1	Nemovitosti pro výrobu a garážování	10
2	Nemovitosti pro obchod a administrativu	7
3	Nemovitosti pro hromadné ubytování a stravování	8
4	Nemovitosti pro dopravu, spoje a školství	9
5	Nemovitosti pro kulturu	8
6	Nemovitosti pro zdravotnictví	8
7	Nemovitosti pro zemědělství	7
8	Nemovitosti pro skladování	6
9	Bytové domy	5
10	Ostatní nemovitosti	8
11	Majetková práva	12

Závislost výnosové hodnoty na použité míře kapitalizace



Příklad

Zjistěte výnosovou hodnotu obytného domu

- 6 bytů: 5x byt 1+1 a 1x byt 2+1
- Dobrý technický stav, postaven v roce 1940
- Pozemek pod domem není ve vlastnictví majitele
- Byt 1+1 obývá chronický neplatič, jeden byt 1+1 je volný, ostatní byty jsou pronajímány na základě nájemní smlouvy na dobu neurčitou za regulované nájemné
- Rohový dům, který se nachází v Ostravě Přívozu na ulici Jílová (jedná se o méně atraktivní lokalitu)
- Dům má 3NP a 1PP, sedlovou střechu
- Obestavěný prostor činí 1 991m³, základní cena je 3870Kč/m³.
- Míra kapitalizace činí 5%.

Identifikace výnosů

Jednotka č.	Počet pokojů	Kategorie	Cena za m2	Výměra	Nájem za vyb. bytu (Kč)	Celkem nájem v Kč	Poznámka
1	1+1	II	18,55	49,50	25,-	918,-	
2	1+1	-	50,00*	45,50	-	2275	Volný byt
3	1+1	II	18,55	48,00	50,-	890,-	
4	1+1	II	18,55	44,00	-	-	Neplatič
5	1+1	II	18,55	42,34	45,-	785,-	
6	2+1	II	18,55	63,44	150,-	1177,-	
Celkem měsíční nájem					270,-	6045,-	
Celkem roční nájem z provozu domu					75780,-		

Identifikace nákladů

Položka	Náklad roční	Poznámka
Daň z nemovitostí	1 764,-	
Pojištění stavby	4 506,-	
Provoz a údržba	38 525,-	
Náklady na pronájem pozemku	4 500,-	
Náklady na správu nemovitostí	0,-	neuvažujeme
Odpisy	0,-	neuvažujeme
Celkem	49 295,-	

Výpočet výnosové hodnoty

Položka	Míra kapitalizace v %	Výnosová hodnota
Diskontní sazba	1,25	1 958 800,-Kč
Běžný úrok v bance	2,00	1 324 250,-Kč
Porovnávací způsob č.m.	5,00	529 700,-Kč
Porovnávací způsob h.m.	8,00	947 250,-Kč
Cenový předpis	4,00	757 800,-Kč

Porovnávací metoda

- nejpoužívanější oceňovací princip
- Setkáváme se s ní v každodenním životě
- Nemovitost je výrazně heterogenní věcí
- Předpoklad: shodné fyzické vlastnosti a stejná kombinace vlastnických práv
- Založena na porovnání s již prodanými nemovitostmi - prodány za poslední rok
- nejspolehlivější a nejobektivnější nástroj sloužící k určení tržní hodnoty
- nutné mít k dispozici databázi s dostatečným počtem nemovitostí, u nichž jsou známy základní technické parametry a také cena, za kterou se obchod realizoval

Metody založené na principu porovnání

- Odbornou rozvahou
- Pomocí koeficientu prodejnosti
- Metoda přímého porovnání
- Metoda nepřímého porovnání

Příklad

Zjistěte porovnávací hodnotu bytu

- lokalita: předměstí
- 3+1, výměra je 60m², balkon, cihla
- průměrný technický stav
- osobní vlastnictví
- Byt není zatížen žádnými zástavními ani jinými právy, které by mohly omezovat nové vlastníky v nakládání s bytem

Výpočet porovnávací hodnoty odbornou rozvahou

Číslo bytu	Typ vl.	Pramen	Lokalita bytu	Technický popis	Cena nabídková	Koeficient	Upravená Cena (Kč)
1	<u>osv</u>	inzerce	Centrum	2+1, 65m ² , balkon, panel	650 000	0,90	585 000
2	<u>dr</u>	inzerce	Centrum	1+1, 35m ² , panel	550 000	0,90 * 1,15	569 000
3	id	kupní smlouva	Okrajová oblast	2+1, 55m ² , balkon, cihla	640 000	1,00	640 000
4	<u>osv</u>	inzerce	Předměstí (navazující centrum)	3+1, 70m ² , balkon, panel	900 000	0,90	810 000
5	<u>osv</u>	inzerce	Úpadková oblast	3+1, 100m ² , balkon, cihla	450 000	0,90	405 000
Průměrná hodnota							601 800

Metoda přímého porovnání

Stanovení koeficientů:

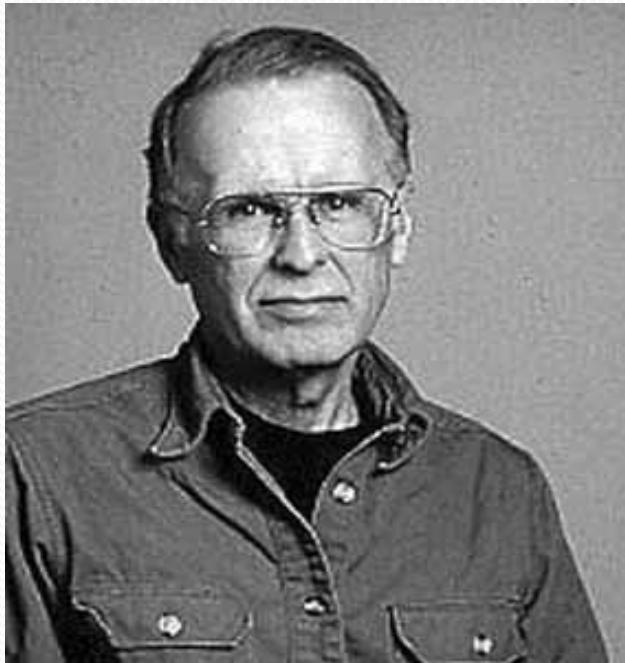
- koeficient polohy K_p
- Koeficient velikosti bytu K_v
- Koeficient konstrukce K_k
- Koeficient balkonu K_b

Výpočet porovnávací hodnoty přímým porovnáním

Byt	Cena	<u>Koef.</u> na pramen ceny	<u>Kor.</u> Vlastnictví	<u>Kor.</u> Cena	<u>Kp</u>	<u>Kv</u>	<u>Kk</u>	Kb	I	Cena (tis. Kč)
1	650	0,90	1,00	585	1,1	1,035	0,9	1,0	1,02465	570
2	550	0,90	1,15	570	1,1	0,825	0,9	0,95	0,7759	734
3	640	1,00	1,05	672	0,9	0,965	1,0	1,0	0,8685	773
4	900	0,90	1,00	810	1,0	1,07	0,9	1,0	0,963	841
5	450	0,90	1,00	405	0,5	1,28	1,0	1,0	0,64	632
Průměr										710

11. Black-Scholesův model, jištění, citlivosti

BS model – otcové myšlenky



Fischer Black (1938-1995)



Myron Scholes (1941-)

BS model - pojmy

= model pro oceňování opcí

- S_T ... cena podkladového aktiva v čase T
- S_0 ... cena podkladového aktiva v čase 0 (spotová)
- K ... realizační cena
- σ ... volatilita
- T ... doba expirace
- r ... bezriziková úroková míra
- c ... cena evropské call opce
- p ... cena evropské put opce

BS model - předpoklady

- neexistence transakčních nákladů
- neměnné zdanění opcí
- úroková míra stejná pro půjčku a výpůjčku
- trhy fungují nepřetržitě (odstraňují arbitráž okamžitě)
- ceny akcií jsou spojitou náhodnou veličinou
- akcie nenesou dividendy
- opce je evropského typu
- akcie lze prodávat i na krátko
- můžeme obchodovat i s částí akcie

BS model - tvar

$$c = S_0 N(d_1) - Ke^{-rT} N(d_2)$$

$$p = Ke^{-rT} N(-d_2) - S_0 N(-d_1)$$

- kde

$$d_1 = \frac{\ln \frac{S_0}{K} + \left(r + \frac{\sigma^2}{2}\right) T}{\sigma \sqrt{T}}$$

$$d_2 = \frac{\ln \frac{S_0}{K} + \left(r - \frac{\sigma^2}{2}\right) T}{\sigma \sqrt{T}}$$

- N je distribuční funkce standardizovaného normálního rozdělení

Citlivost - Greeks

- Δ ... delta
- θ ... theta
- Γ ... gamma
- V ... vega
- ρ ... rho

Delta a delta-neutrální portfolio

- Delta vyjadřuje rychlost změny ceny opce vzhledem ke změně ceny akcie:

$$\Delta = \frac{\partial c}{\partial S}$$

- Delta-neutrální portfolio: $\Delta = 0$
- Výhoda: Při malých změnách ceny akcie se hodnota celého portfolia nemění.
- Platí: $\Delta c = N(d_1)$, $\Delta p = N(d_1) - 1$, $\Delta_{\text{port.}} = \text{suma}(n_i \Delta_i)$,
 $\Delta(\text{akcie})=1$
- Pozor na vysoké transakční náklady.

Theta

- Theta vyjadřuje rychlost změny ceny opce vzhledem ke změně času:

$$\theta = \frac{\partial c}{\partial T}$$

- Proti času nemá smysl se jistit – víme, jak se mění.
- θ v praxi náhražka za Γ
- θc získáme derivací podle T

Gamma a gamma-neutrální portfolio

- Gamma vyjadřuje rychlost změny ceny opce vzhledem ke změně delta:

$$\Gamma = \frac{\partial \Delta}{\partial S} = \frac{\partial^2 c}{\partial S^2}$$

- Čím je Γ větší, tím častěji musíme rebalancovat portfolio.
- Γ -neutrální portfolio: $\Gamma = 0$
- Výhoda: I při velkých změnách ceny akcie se hodnota celého portfolia nemění.
- Platí: $\Gamma(\text{akcie})=0$, Γc dostaneme druhou derivací podle S

Vega

- Vega vyjadřuje rychlost změny ceny opce vzhledem ke změně volatility:

$$v = \frac{\partial c}{\partial \sigma}$$

- Platí: V (akcie) = 0
- Obvykle pokud $\Gamma = 0 \leftrightarrow V \neq 0$
- Pro $\Gamma = 0 \wedge V = 0$ potřebuje dva různé deriváty na stejné podkladové aktivum

Rho

- Rho vyjadřuje rychlost změny ceny opce vzhledem ke změně úrokové míry:

$$\rho = \frac{\partial c}{\partial r}$$

- ρc získáme derivací podle r

Jištění – možnosti investora

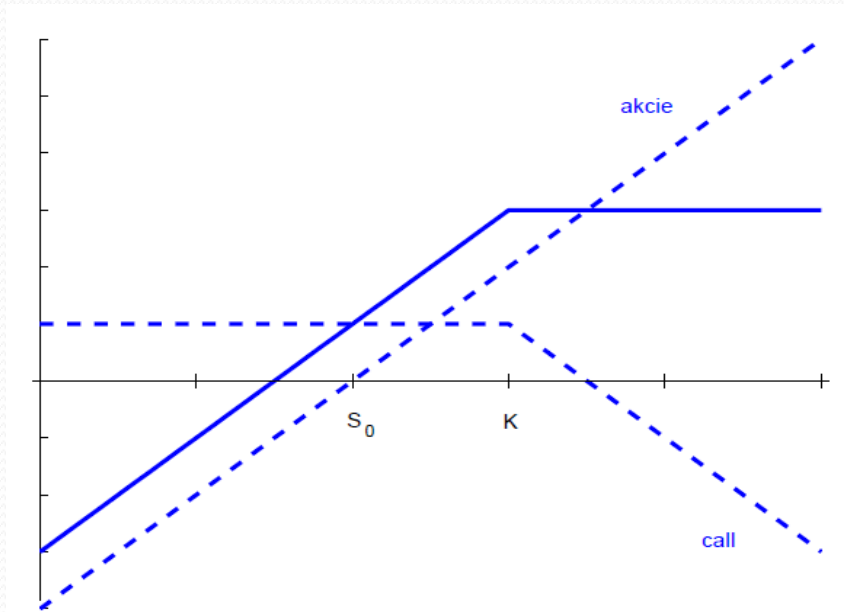
- nicnedělání (naked position)
- prodej akcií

Jištění pomocí opcí:

- krytý call
- ochranný put
- dynamické jištění (delta hedging)

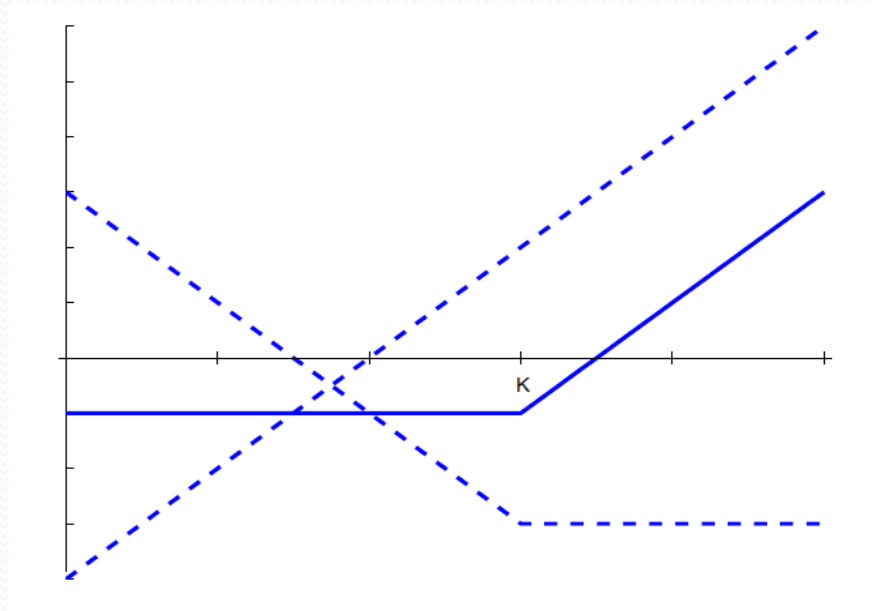
- Jištění = chránění investice v akciích.

Krytý call



- skládá se z short call opce a long akcie
- případná ztráta z akcií je kryta výnosem z opcí

Ochranný put



- skládá se z long put opce a long akcie
- garance minimální hodnoty investice

Dynamické jištění

- utvoříme delta-neutrální portfolio => je imunní vůči malým výkyvům ceny akcie
- za daný časový úsek měníme svou pozici v opcích tak, aby portfolio zůstávalo delta-neutrální
- existuje delta-hedging s call a s put

Příklad

- Vypočítejte cenu evropské call opce pomocí BS vzorce, když znáte:
 - doba expirace 1 rok
 - spotová cena je 2Kč
 - úroková míra 82%,
 - realizační cena je 2Kč
 - volatilita 0,6
- nápověda: $N(1,666)=0,9521$, $N(1,366)=0,9141$, $N(1,066)=0,8569$, $\exp(-0,82)=0,44$
- výsledek $c = 1,15\text{Kč}$

Zdroje

- Hull, J.C. *Options, futures and other derivatives*. 8. vyd. Boston: Pearson, 2012, 847 s. ISBN 9780273759072.
- Ambrož, L. *Oceňování opcí*. Vyd. 1. Praha: C.H. Beck, 2002, xvi, 313 s. ISBN 8071795313.
- přednášky do MFoo3

12. Oceňování derivátů, hraniční opce, opce závislé na cestě

Spôsoby oceňování derivátů

Black-Scholesov vzorec a jeho modifikácie

$$c = kS_0N(kd_1) - kXe^{-rT}N(kd_2)$$

$$d_1 = \frac{\ln(S_0 / X) + (r + \sigma^2 / 2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(S_0 / X) + (r - \sigma^2 / 2)T}{\sigma\sqrt{T}}$$

c – cena kúpnej opcie

S_0 – cena podkladového aktíva v čas 0

X – realizačná cena

$N()$ – distribučná funkcia normálneho štandardizovaného rozdelenia

k – je 1 pre call opciu a -1 pre put opciu

r – tržná úroková miera

T – doba do expirácie opcie

σ – volatilita podkladového aktíva

Predpoklady:

1. Log-normálne rozdelenie podkladového aktíva
2. Úroková miera sa nemení
3. Nevyplácajú sa dividendy
4. Opcia je európska
5. Žiadne transakčné náklady
6. Nekonečná deliteľnosť podkladového aktíva
7. Obchodovanie je spojité
8. Short-sale je povolený

Spůsoby oceňování derivátů

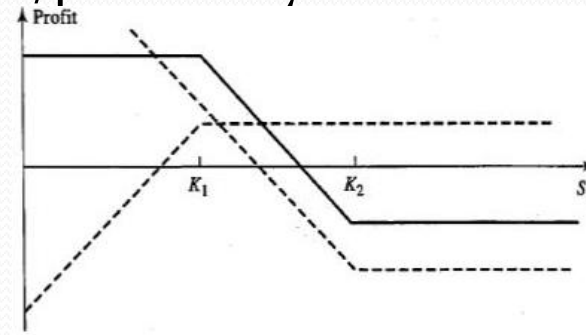
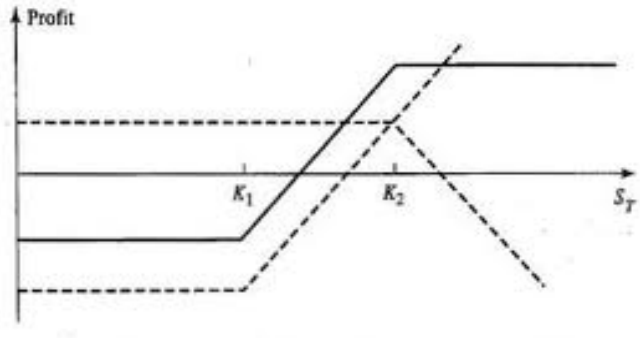
- Risk - neutrálné oceňovanie
- Binomické a trinomické stromy
- 3D binomické a trinomické stromy
 - Spread opcie resp. tam kde sa využívajú 2 procesy súčasne

Spůsoby oceňování derivátů

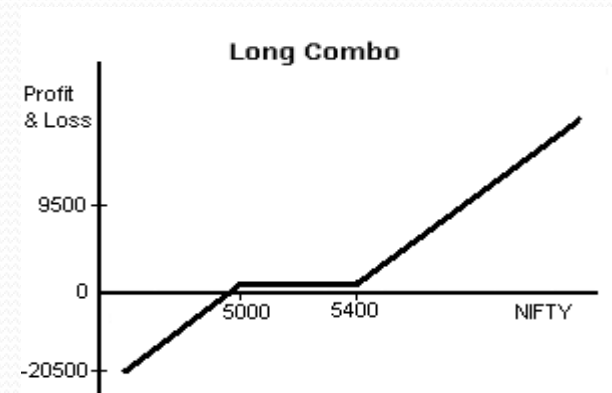
- Simulácia Monte Carlo
 - Určenie strednej hodnoty z určitého počtu náhodných pokusov
 - Vhodná pre path-dependend opcie
 - Opcie s viac podkladovými aktívami
 - Opcie ktoré neboli analyticky ocenené
 - Náchylná na vysoké chyby pri nízkom počte opakovaní

Opční formace

- **Bull a Bear spread** (short a long call/put s rôznými realizačnými cenami)

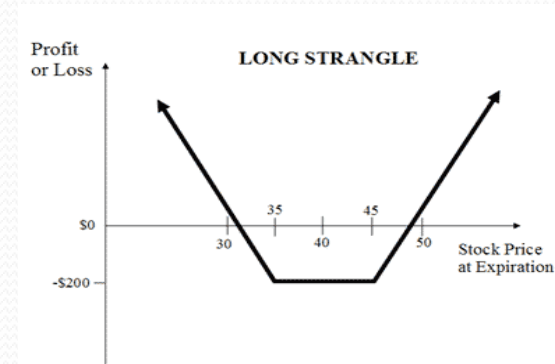
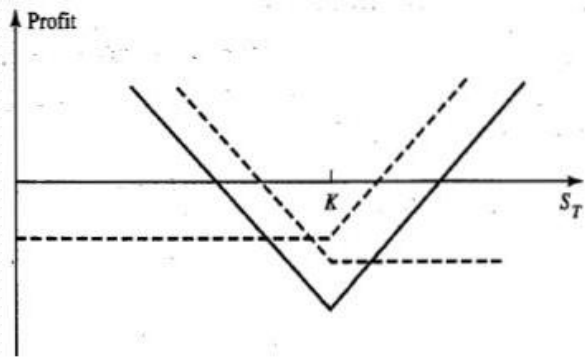


- **Combo** (long OTM call a short OTM put)

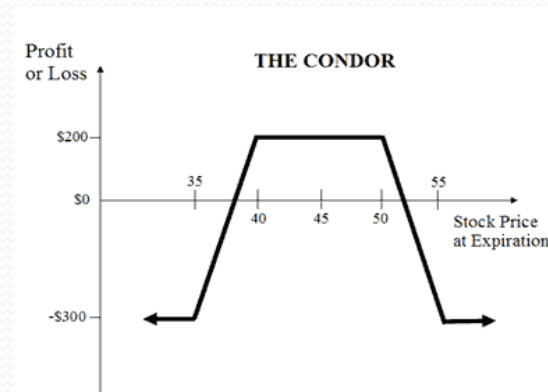
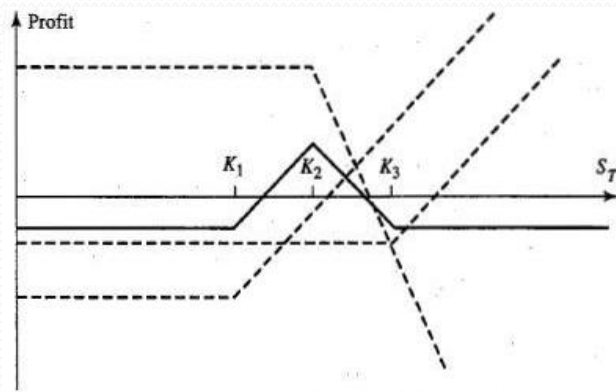


Opční formace

- **Straddle, Strangle** (1 call +1 put)

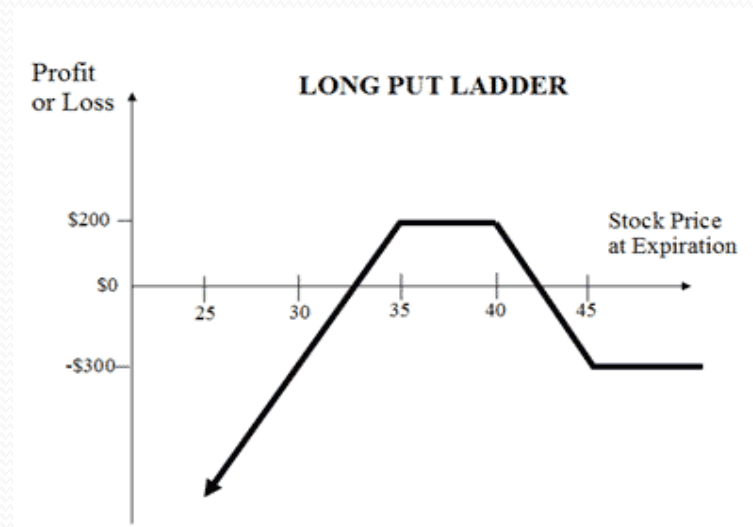
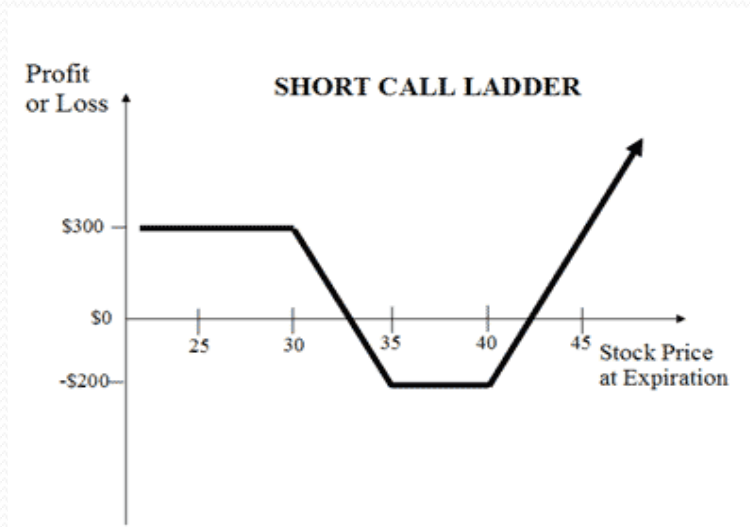


- **Butterfly, Condor** (long call+2 short call+long call) call/put



Opční formace

- Ladder (short call+2 long call) alebo (2 long put+short put)



Exotické opcie

- Path-dependent opcie (závislé na ceste)
 - Ázijské opcie
 - Forward-start opcie
 - One-clique option
 - Barirové opcie
 - Roll-up/Roll-down opcie
 - Lookback opcie
 - Shout opcie
 - Lock-in opcie
- Korelačné opcie
 - Výmenné opcie
 - Best-of/Worst-of and cash opcie
 - Korelačné binárne opcie
 - Kvocientné opcie
 - Produkčné opcie
 - Opcie zahraničného kapiálu
- Quanto opcie
- Rainbow opcie
- Spread opcie
- Dual-strike opcie
- Out-performance opcie
- Basket opcie
- Ostatné
 - Opcie s nelineárnou výplatom
 - Zložené opcie
 - Chooser opcie
 - Opcie s podmienenou prémieou
 - Explodujúce opcie
 - Reset opcie
 - Konvexné opcie

Kombinace

- Kanárské opcie
- Bermudské opcie
- Havajské opcie
- Ruské opcie
- Parížske opcie
- Izraelské opcie
- Napoleonské opcie

Barrierové opcie

- Opcie závislé na ceste, ktoré sa spustia resp. ukončia svoju platnosť pri prerazení vopred stanovenej hranice, ktorá je buď pod alebo nad spotovou cenou v dobe vypísania opcie
- Opcia môže vyplácať rabat, ak nedôjde k spusteniu, alebo zanikne
- Delenie
 - » Put/Call
 - » Knock-In/Knock-Out
 - » Up/Down
- Ocenenie

$$A = \phi S e^{-gT} N(\phi x_1) - \phi X e^{-rT} N(\phi x_1 - \phi \sigma \sqrt{T})$$

$$B = \phi S e^{-gT} N(\phi x_2) - \phi X e^{-rT} N(\phi x_2 - \phi \sigma \sqrt{T})$$

$$C = \phi S e^{-gT} \left(\frac{H}{S}\right)^{2(\mu+1)} N(\eta y_1) - \phi X e^{-rT} \left(\frac{H}{S}\right)^{2\mu} N(\eta y_1 - \eta \sigma \sqrt{T})$$

$$D = \phi S e^{-gT} \left(\frac{H}{S}\right)^{2(\mu+1)} N(\eta y_2) - \phi X e^{-rT} \left(\frac{H}{S}\right)^{2\mu} N(\eta y_2 - \eta \sigma \sqrt{T})$$

Barierové opce

$$E = Ke^{-rT} \left[N(\eta x_2 - \eta\sigma\sqrt{T}) - \left(\frac{H}{S}\right)^{2\mu} N(\eta y_2 - \eta\sigma\sqrt{T}) \right]$$
$$F = K \left[\left(\frac{H}{S}\right)^{\mu+\lambda} N(\eta z) + \left(\frac{H}{S}\right)^{\mu-\lambda} N(\eta z - 2\eta\lambda\sigma\sqrt{T}) \right]$$

Kde

$$x_1 = \frac{\ln\left(\frac{S}{X}\right)}{\sigma\sqrt{T}} + (1 + \mu)\sigma\sqrt{T}$$

$$y_1 = \frac{\ln\left(\frac{H^2}{SX}\right)}{\sigma\sqrt{T}} + (1 + \mu)\sigma\sqrt{T}$$

$$z = \frac{\ln\left(\frac{H}{S}\right)}{\sigma\sqrt{T}} + \lambda\sigma\sqrt{T}$$

$$\lambda = \sqrt{\mu^2 + \frac{2(r-g)}{\sigma^2}}$$

$$x_2 = \frac{\ln\left(\frac{S}{H}\right)}{\sigma\sqrt{T}} + (1 + \mu)\sigma\sqrt{T}$$

$$y_2 = \frac{\ln\left(\frac{H}{S}\right)}{\sigma\sqrt{T}} + (1 + \mu)\sigma\sqrt{T}$$

$$\mu = \frac{r-g-\frac{\sigma^2}{2}}{\sigma^2}$$

S- spotová cena podkladového aktíva pri ocenení

H- úroveň bariery

X- strike

K- rabat

r- tržná úroková miera

g- miera rastu dividend

σ - volatilita

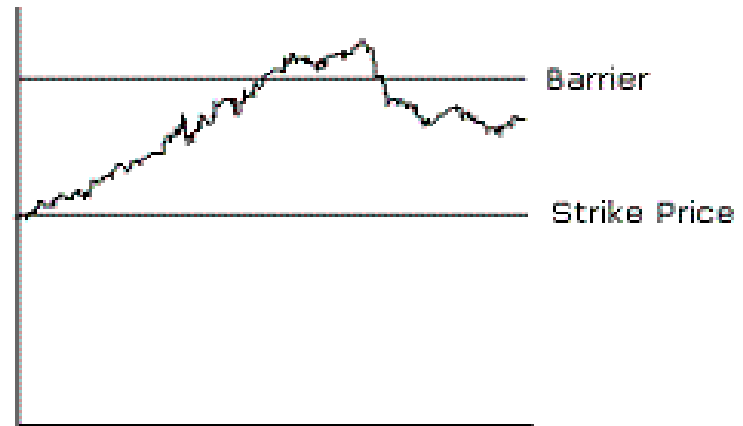
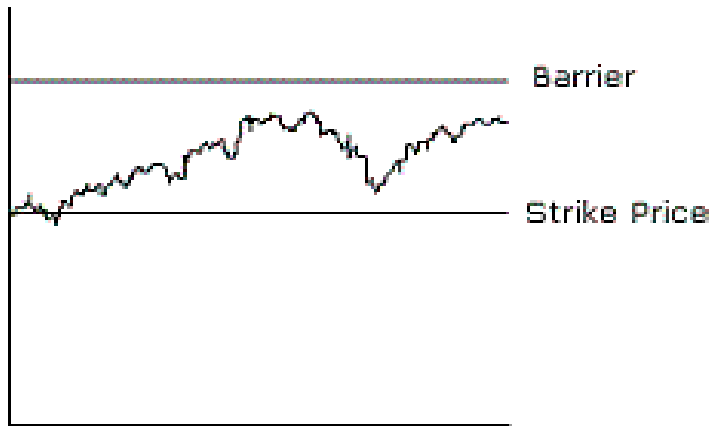
Bariérové opce

$A_k X > H$	Hodnota	η	ϕ
Down-In Call	$C+E$	1	1
Down-Out Call	$A-C+F$	1	1
Up-In Call	$A+E$	-1	1
Up-Out Call	F	-1	1
Down-In Put	$B-C+D+E$	1	-1
Down-Out Put	$A-B+C-D+F$	1	-1
Up-In Put	$A-B+D+E$	-1	-1
Up-Out Put	$B-D+F$	-1	-1

$A_k X < H$	Hodnota	η	ϕ
Down-In Call	$A-B+D+E$	1	1
Down-Out Call	$B-D+F$	1	1
Up-In Call	$B-C+D+E$	-1	1
Up-Out Call	$A-B+C-D+F$	-1	1
Down-In Put	$A+E$	1	-1
Down-Out Put	F	1	-1
Up-In Put	$C+E$	-1	-1
Up-Out Put	$A-C+F$	-1	-1

Bariérové opcie

- Bariérové up opcie



Azijské opcie

- Opcie kde je výplata závislá na rozdiel medzi priemerom cien podkladového aktíva za určité vopred stanovené obdobie a vopred stanovenou realizačnou cenou.
- Môže byť aj rozdiel medzi spotovou cenou a priemerom (fluktuujúca realizačná cena)
- Aritmetický a geometrický priemer
- Výhody: lacnejšie než vanilla opcie, nižšia volatilita
- Ocenenie Geometrickej call opcie:

Geometrický priemer:

$$B^{sa}(0) = 1, \quad B^{sa}(j) = \left(\prod_{i=1}^j \frac{S_i}{S_0} \right)^{\frac{1}{j}}$$

Časové funkcie:

$$T_{r,n-j}^{sa} = \frac{n-j}{n} \quad T_{n-j}^{sa} = \tau \left(\frac{n-j}{n} \right)^2 - \frac{(n-j)(n-j-1)(4n-4j+1)}{6n^3}$$

Pomocné funkcie:

$$A^{sa}(j) = e^{\frac{-r(\tau - T_{r,n-j}^{sa}) - \sigma^2(T_{r,n-j}^{sa} - T_{n-j}^{sa})}{2}} B^{sa}(j) \quad d_{n-j}^{sa} = \frac{\left\{ \ln\left(\frac{S}{X}\right) + \left(r - g - \frac{1}{2}\sigma^2\right)T_{r,n-j}^{sa} + \ln[B^{sa}(j)] \right\}}{\sigma\sqrt{T_{n-j}^{sa}}}$$

Cena opcie:

$$C^{sa} = kS_0 A^{sa}(j) e^{-gT_{r,n-j}^{sa}} N\left(kd_{n-j}^{sa} + k\sigma\sqrt{T_{n-j}^{sa}}\right) - kXe^{-r\tau} N\left(kd_{n-j}^{sa}\right)$$

n – celkový počet prvkov z ktorých sa počíta priemer
j – priebežný počet prvkov z ktorých sa počíta priemer

k – 1 pre call a -1 pre put
τ - doba do expirácie opcie

Lookback opcie

- Opcie závislé na ceste, ktorých payoff je závislí na maxime resp. minime, ktorý vývoj ceny podkladového aktíva dosiahol za určité obdobie.
- **Delenie**
 - Fixný strike – payoff je určený ako rozdiel medzi maximálnou cenou a fixnou realizačnou cenou pre call opciu a rozdielom fixnej realizačnej ceny a minimom podkladového aktíva pre put opciu
 - Pohyblivý strike - payoff je daný rozdielom maxima a spotovej ceny v rozhodný deň pre call a spotovej ceny a minima pre put opciu. Vždy dôjde k realizácii.
- Relatívne drahé, vhodné na komodity
- Ocenenie:

Fixed strike Lookback call ak $X > \text{Max}(S)$

$$C = Se^{-gT} N(a_1) - Xe^{-rT} N(a_2) + Se^{-rT} \frac{\sigma^2}{2(r-g)} \left[e^{(r-g)T} N(a_1) - \left(\frac{S}{X} \right)^{\frac{2(r-g)}{\sigma^2}} N\left(a_1 - \frac{2(r-g)\sqrt{T}}{\sigma} \right) \right]$$

Fixed strike Lookback call ak $X < \text{Max}(S)$

$$C = Se^{-gT} N(b_1) + e^{-rT} [(S_{MAX} - X) - S_{MAX} N(b_2)] + Se^{-rT} \frac{\sigma^2}{2(r-g)} \left[e^{(r-g)T} N(b_1) - \left(\frac{S}{S_{MAX}} \right)^{\frac{2(r-g)}{\sigma^2}} N\left(a_1 - \frac{2(r-g)\sqrt{T}}{\sigma} \right) \right]$$

Lookback opcie

Fixed strike Lookback Put ak $X < \text{Min}(S)$

$$P = Xe^{-rT} N(-a_2) - Se^{-gT} N(-a_1) + Se^{-rT} \frac{\sigma^2}{2(r-g)} \left[\left(\frac{S}{X} \right)^{\frac{2(r-g)}{\sigma^2}} N\left(-a_1 - \frac{2(r-g)\sqrt{T}}{\sigma}\right) - e^{(r-g)T} N(-a_1) \right]$$

Fixed strike Lookback Put ak $X > \text{Min}(S)$

$$P = e^{-rT} [(X - S_{MIN}) + S_{MIN} N(-c_2)] - Se^{-gT} N(-c_1) + Se^{-rT} \frac{\sigma^2}{2(r-g)} \left[\left(\frac{S}{S_{MIN}} \right)^{\frac{2(r-g)}{\sigma^2}} N\left(-c_1 + \frac{2(r-g)\sqrt{T}}{\sigma}\right) - e^{(r-g)T} N(-c_1) \right]$$

Kde

$$a_1 = \frac{\ln\left(\frac{S}{X}\right) + \left(r - g + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}$$

$$b_1 = \frac{\ln\left(\frac{S}{S_{MAX}}\right) + \left(r - g + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}$$

$$c_1 = \frac{\ln\left(\frac{S}{S_{MIN}}\right) + \left(r - g + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}$$

S- spotová cena pri uzatváraní kontraktu

X- realizačná cena

S_{\max} - historicky maximálna cena

S_{\min} - historicky minimálna cena

r - tržná úroková miera

T - doba do maturity opcie

g - miera rastu dividend

σ - volatilita

Lookback opcie

$$C = Se^{-gT}N(c_1) - S_{MIN}e^{-rT}N(c_2) + Se^{-rT} \frac{\sigma^2}{2(r-g)} \left[\left(\frac{S}{S_{MIN}} \right)^{\frac{2(r-g)}{\sigma^2}} N\left(-c_1 + \frac{2(r-g)\sqrt{T}}{\sigma}\right) - e^{-gT}N(-c_1) \right]$$

- Lookback s pohyblivou realizačnou cenou

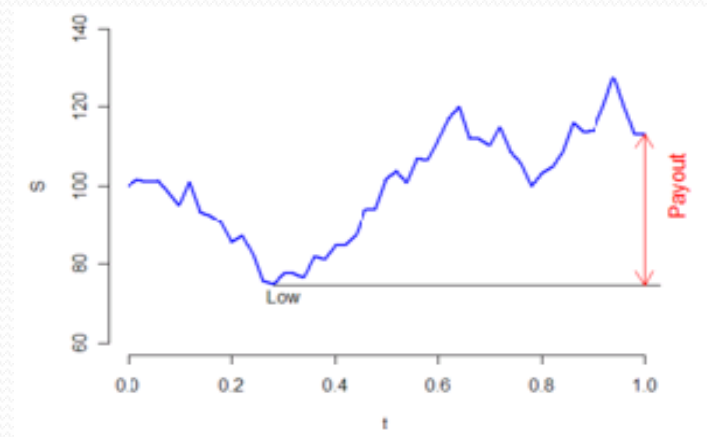
$$P = S_{MAX}e^{-rT}N(d_2) - Se^{-gT}N(-d_1) + Se^{-rT} \frac{\sigma^2}{2(r-g)} \left[-\left(\frac{S}{S_{MAX}} \right)^{\frac{2(r-g)}{\sigma^2}} N\left(d_1 + \frac{2(r-g)\sqrt{T}}{\sigma}\right) + e^{-gT}N(d_1) \right]$$

$$c_1 = \frac{\ln\left(\frac{S}{S_{MIN}}\right) + \left(r - g + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}$$

$$c_2 = c_1 - \sigma\sqrt{T}$$

$$d_1 = \frac{\ln\left(\frac{S_{MAX}}{S}\right) + \left(r - g + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$



Shout opcie

- Sú to modifikácie európskej opcie. Ako bonus má majiteľ opcie právo “zakričať” na vypisovateľa jedenkrát (prípadne viac krát) na vypisovateľa v priebehu životnosti opcie v čase τ , keď predpokladá, že je cena dostatočne vysoko (nízko).
- Payoff je väčšia hodnota z výplaty z európskej vanilla opcie alebo z rozdielu spotovej ceny pri shout-e a realizačnej ceny.
- Podobná lookback opcií, ale lacnejšia
- Kombinácia vlastností lookback a americkej opcie
- Využívajú sa v ako súčasť úrokových produktov a v segregovaných poisťných fondoch
- Payoff: call

	$\max(0, S_T - S_\tau) + (S_\tau - X)$	S_T – cena pri expirácii
put		S_τ – cena pri shout-e
	$\max(0, S_\tau - S_T) + (X - S_\tau)$	X – realizačná cena

- Oceňujú sa simuláciami, napríklad Monte Carlo. Prípadne binomickými a trinomickými stromami. Presné analytické riešenie neexistuje.

Zložené opcie

- Call na call opciu

$$C_C = Se^{-gT_1} M\left(a_1, b_1; \sqrt{\frac{T_2}{T_1}}\right) - X_1 e^{-rT_1} M\left(a_2, b_2; \sqrt{\frac{T_2}{T_1}}\right) - e^{-rT_2} X_2 N(a_2)$$

- Call na put opciu

$$C_P = Se^{-gT_1} M\left(a_1, -b_1; -\sqrt{\frac{T_2}{T_1}}\right) - X_1 e^{-rT_1} M\left(a_2, -b_2; -\sqrt{\frac{T_2}{T_1}}\right) + e^{-rT_2} X_2 N(a_2)$$

- Put na call opciu

$$P_C = X_1 e^{-rT_1} M\left(-a_2, b_2; -\sqrt{\frac{T_2}{T_1}}\right) - Se^{-gT_1} M\left(-a_1, b_1; -\sqrt{\frac{T_2}{T_1}}\right) + e^{-rT_2} X_2 N(-a_2)$$

- Put na put opciu

$$P_P = X_1 e^{-rT_1} M\left(-a_2, -b_2; \sqrt{\frac{T_2}{T_1}}\right) - Se^{-gT_1} M\left(-a_1, -b_1; \sqrt{\frac{T_2}{T_1}}\right) - e^{-rT_2} X_2 N(-a_2)$$

$$a_1 = \frac{\ln\left(\frac{S}{S^*}\right) + \left(r - g + \frac{1}{2}\sigma^2\right)T_2}{\sigma\sqrt{T_2}}$$

$$b_1 = \frac{\ln\left(\frac{S}{X_2}\right) + \left(r - g + \frac{1}{2}\sigma^2\right)T_1}{\sigma\sqrt{T_1}}$$

$$a_2 = a_1 - \sigma\sqrt{T_2}$$

$$b_2 = b_1 - \sigma\sqrt{T_1}$$

- Pre kritickú hodnotu S^* musí platiť že hodnota európskej call/put opcie v čase T_2 je X_2
 $C(S^*, X_1, g, r, \sigma, T_1 = T_2) = X_2$

Chooser opcie

- Tieto opcie majú vlastnosť, že v určitom vopred stanovenom čase v budúcnosti T_1 sa držiteľ opcie rozhodne či je daná opcia call alebo put.
- Drahšie ako vanilla opcie
- Lacnejšia alternatíva obchodovania straddlov
- Hodnota v čase T_1 je:

$$\text{Max}(c,p)$$

kde

c – je hodnota kúpnej opcie v čase T_1 s realizáciou v čase T_2

p – je hodnota predajnej opcie v čase T_1 s realizáciou v čase T_2

- Ocenenie:

Na základe call-put parity dostaneme

$$\max(c, p) = \max\left(c, c + Xe^{-r(T_2-T_1)} - S_1 e^{-g(T_2-T_1)}\right) = c + e^{-g(T_2-T_1)} \max(0, Xe^{-(r-g)(T_2-T_1)} - S_1)$$

S_1 je cena podkladového aktíva v čase voľby typu opcie

T_1 doba do voľby typu opcie

T_2 doba do maturity opcie

Chooser opcie

- Cena chooser opcie je určená ako:

$$Ch = Se^{-gT_2} N(d_1) - Xe^{-rT_2} N(d_2) - Se^{-gT_2} N(-e_1) + Xe^{-rT_2} N(-d_2)$$

$$d_1 = \frac{\ln\left(\frac{S}{X}\right) + \left(r - g + \frac{1}{2}\sigma^2\right)T_2}{\sigma\sqrt{T_2}} \quad d_2 = d_1 - \sigma\sqrt{T_2}$$

$$e_1 = \frac{\ln\left(\frac{S}{X}\right) + (r - g)T_2 + \frac{1}{2}\sigma^2T_1}{\sigma\sqrt{T_1}} \quad e_2 = e_1 - \sigma\sqrt{T_1}$$

kde

S – cena podkladového aktíva pri oceňovaní opcie

X – realizačná cena opcie

S₁ – je cena podkladového aktíva v čase voľby typu opcie

T₁ – doba do voľby typu opcie

T₂ – doba do maturity opcie

r – tržná úroková miera

σ – volatilita

g – miera rastu dividend

Binárne/Digitálne opcie

- Opcie, ktoré majú binárnu výplatnú funkciu. Ak podkladové aktívum prekročí realizačnú cenu, držiteľ opcie získa podkladové aktívum alebo vopred stanovenú sumu peňazí.
- Relatívne drahé opcie
- **Cash-or-nothing**
 - Call – držiteľ získa sumu K ak cena podkladového aktíva je vyššia ako realizačná cena ($S > X$), inak nezíska nič
 - Put – držiteľ získa sumu K ak cena podkladového aktíva je nižšia ako realizačná cena ($S < X$), inak nič

Ocenenie:

$$C_{con} = Ke^{-rT} N(d)$$

$$P_{con} = Ke^{-rT} N(-d)$$

- **Asset-or-nothing**
 - Call – držiteľ získa podkladové aktívum ak cena podkladového aktíva je vyššia ako realizačná cena ($S > X$), inak nezíska nič
 - Put – držiteľ získa podkladové aktívum ak cena podkladového aktíva je nižšia ako realizačná cena ($S < X$), inak nič

Ocenenie:

$$C_{aon} = Se^{-gT} N(d)$$

$$P_{aon} = Se^{-gT} N(-d)$$

$$d = \frac{\ln\left(\frac{S}{X}\right) + \left(r - g + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}$$

c – cena kúpnej opcie

S_0 – cena podkladového aktíva v čas 0

X – realizačná cena

N() – distribučná funkcia normálneho štandardizovaného rozdelenia

r – tržná úroková miera

T – doba do expirácie opcie

σ – volatilita podkladového aktíva

Ďalšie typy opcií

- **Amortizované/Veľerybie opcie**- rozdiel medzi priemernou cenou podkladového aktíva a realizačnou cenou (určenej percentuálne zo spotovej ceny)
- **Cliquet/Ratchet Options** – opcie s periodickým vysporiadaním, pri prenastavení realizačnej ceny podľa posledného vysporiadania. Zisk sa môže priebežne vyplácať alebo akumulovať.
- **Alfa-kvantilové opcie** – lacnejšia verzia Lookback opcie. Ako rozhodujúca sa berie najnižšia, resp. najvyššia hodnota, kde cena zotrvá aspoň určitú dobu. Neobchoduje sa ešte.
- **Bermudské opcie** – kombinácia amerických a európskych opcií. Majú niekoľko realizačných dní. Využívajú sa na úrokových trhoch a Forexe.
- **Kanárske opcie** – opäť kombinácia amerických a európskych. Opcia sa správa ako európska až do istého vopred stanoveného dátumu a následne funguje ako americká, alebo bermudská opcia.
- **Košíkové opcie** – podkladové aktívum je index, alebo skupina podkladových aktív
- **Instalment opcie** – prémie za opciu sa spláca po častiach. Séria zložených opcií.
- **Coupe option** – podobné Cliquet opciám, ale po výplate sa nová realizačná cena nastaví na horšiu zo spotovej ceny a pôvodnej realizačnej ceny.

Ďalšie typy opcií

- **Parížske opcie** – kombinácia bariérových a ázijských opcií. Na rozdiel od bariérových knock-in/out je spustený ak podkladové aktívum zotrvá určitú dobu na druhej strane bariéry resp. ak ju pretne priemer cien.
- **Edokko (Tokíjske) opcie** – rozšírenie Parížskych opcií. Redukuje potrebu predaja podkladového aktíva vypisovateľom opcie na krátko ak je opcia ITM tesne pred expiráciou
- **Výmenné opcie** – umožňujú vo vopred stanovený termín výmenu podkladových aktív. Používajú sa najmä na menové páry.
- **Napoleónske opcie** – vyplácajú fixný kupón a najhorší výsledok indexu za určité obdobie.
- **Dúhové opcie** – majú 2 a viac podkladových aktív z ktorých sa vyberá výsledok najlepšieho/najhoršieho/priemer
- **Opcie Spreadu** – Payoff je funkciou 2 alebo viacerých podkladových aktív
- **Extreme spread opcie** – modifikácia lookback opcie. Payof je rozdiel maxima v druhej časti životnosti opcie a maxima v prvej časti životnosti opcie.
- **Predĺžiteľné opcie** – vypisovateľ alebo držiteľ môžu predĺžiť životnosť opcie

Ďalšie typy opcií

- **Havajské opcie** – kombinácia amerických a ázijských opcií. Payoff je rozdiel medzi strikom a priemernou cenou a je možná predčasná realizácia.
- **Pasové (Passport) opcie** – podkladové aktívum je zostatok obchodného účtu bánk. Európsky typ opcie.
- **Dovolenkové (Vacation) opcie** – to isté ako pasové, len majú charakteristiku americkej opcie.
- **Izraelské opcie** – americká opcia s možnosťou vypisovateľa opcie ukončiť jej platnosť kedykoľvek počas jej životnosti. Držiteľ opcie dostane nejaký rabat.
- **Ruské opcie** – kombinácia americkej a lookback opcie. Držiteľ opcie má právo na určité percento z maximálnej hodnoty, ktorú opcia dosiahla. Neobchoduje sa este.

Zdroje

- Higham Desmond J. *An introduction to Financial Option Valuation* 4.th edition. Cambridge University Press: 2004. 273 s. ISBN 978-0-52183884-9
- HULL, J. C. *Options, futures, and other derivatives*. 8.th edition. Harlow: Pearson Edusation, 2012. 845 s. ISBN 0-13-009056-5
- ZHANG, P. G. *Exotic options*. 2nd edition. vyd. Singapore: World Scientific Publishing Co.Pte.Ltd., 1998. 692 s. ISBN 9810234821
- www.global-derivatives.com

13. Reference



- Neapolitan R.E.: *Learning Bayesian Networks*. Prentice Hall. 2004
- Ben-Gal I.: Bayesian Networks. *Encyclopedia of Statistics in Quality and Reliability*. Wiley. 2007
- Fenton N.E., Neil M.: *Managing Risk in Modern World*. London Mathematical Society. 2007
- Koller D., Pfeffer A.: *Object-oriented Bayesian networks*. Proceedings of 13th Annual Conference on Uncertainty in Artificial Intelligence, Providence, Rhode Island, pp. 302-313. 1997
- Fenton N.E., Neil M., Caballero J.G.: *Using Ranked Nodes to Model Quantitative Judgments in Bayesian Networks*. IEEE Transactions on Knowledge and Data Engineering, Vol. 19, No. 10, pp. 1420-1432. 2007
- Charniak E.: *Bayesian Networks without Tears*. In: AI Magazine, vol. 12, pp. 50-63. 1991
- Zagorecki A., Druzdel M.: *An Empirical Study of Probability Elicitation under Noisy-OR Assumption*. Proceedings on the 17th Int'l Florida Artificial Intelligence Research Soc. Conference, 2004, pp. 880-885.
- Murphy K.P.: *Dynamic Bayesian Networks: Representations, Inference and Learning*. University of California, Berkeley, 2002.
- Cobb B.R., Shenoy P.P.: *Inference in Hybrid Bayesian Networks with Mixtures of Truncated Exponentials*. School of Business Working Paper no. 294, 2005
- Murphy K.: *A Brief Introduction to Graphical Models and Bayesian Networks*, 1998.
<http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html>
- WIKIPEDIA: Bayesian Network. http://en.wikipedia.org/wiki/Bayesian_network
- <http://www2.cs.cas.cz/~sima/kniha.pdf>
- <http://www.root.cz/clanky/neuronove-siete-su-ciernou-skrinkou/>

Užitečné zdroje dat

- <http://archive.ics.uci.edu/ml/>
- <http://kdd.ics.uci.edu/>



- <http://sede.neurotech.com.br:443/PAKDD2009/>
- <http://www.dataminingbook.com/>
- http://www.stat.uni-muenchen.de/service/datenarchiv/welcome_e.html