Analýza dat pro Neurovědy

RNDr. Eva Janoušová
doc. RNDr. Ladislav Dušek, Dr.

Jaro 2012

4. Hledání diagnostického cut-off pomocí ROC křivek.

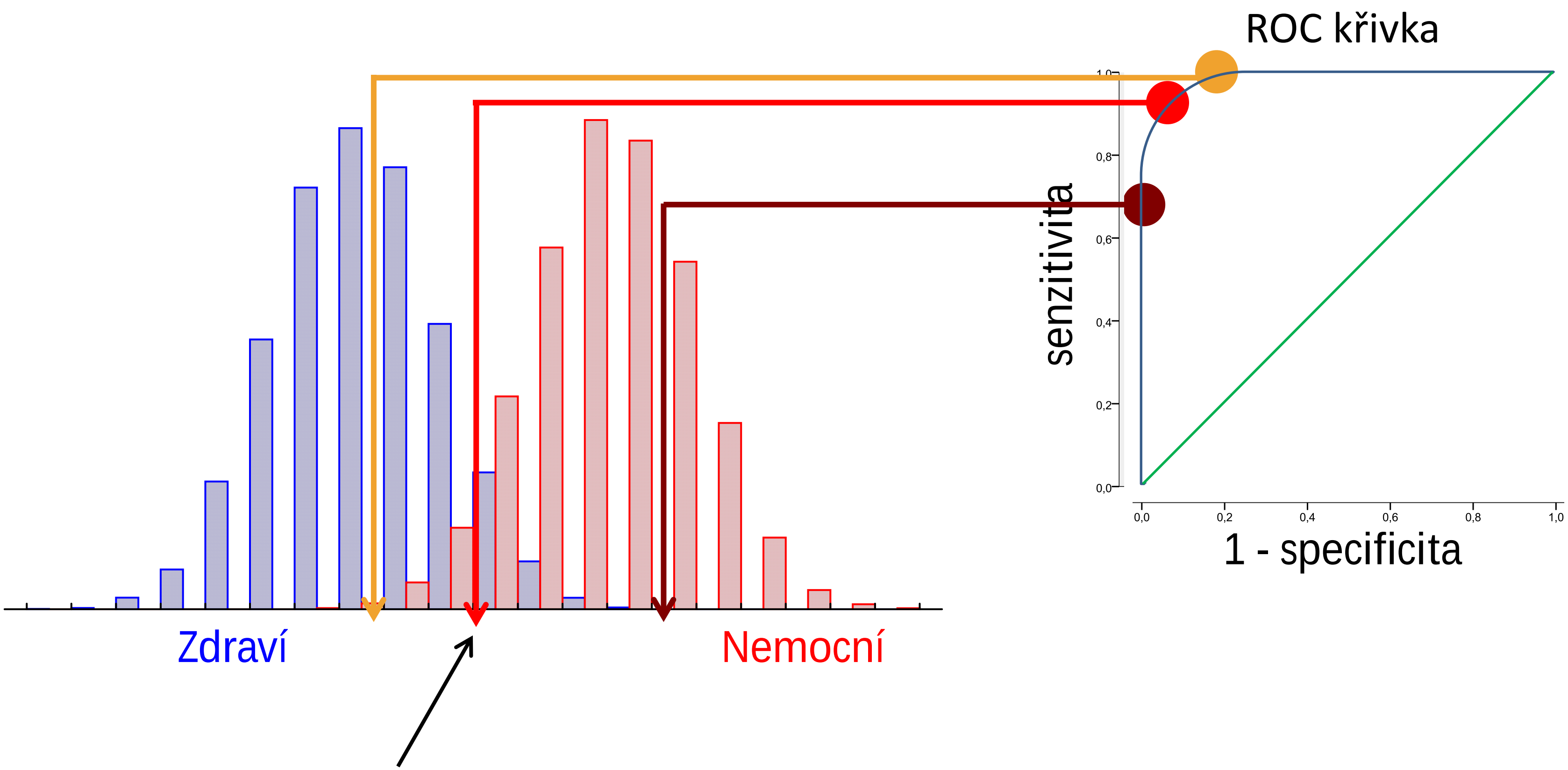
ROC analýza – motivace

- Dříve probrané ukazatele diagnostické síly testů (senzitivita, specificita apod.) **nelze použít u diagnostických testů, jejichž výstupem je spojitá (kvantitativní) proměnná** (např. koncentrace analytu v krevním séru, systolický krevní tlak).
- Na základě předchozích výzkumů známe dělicí body, které odlišují normální a patologické hodnoty spojité proměnné, pomocí nichž můžeme spojitou proměnnou binarizovat – tzn. vytvoření dvou kategorií „pozitivní“ / „negativní“ (např. „pod normou“ / „v normě“).
- Pokud dělicí body nejsou známy předem, můžeme se je snažit nalézt pomocí **ROC („Receiver Operating Characteristic“) křivky**.

- **Cíle ROC analýzy:**
  1. Určit, zda je spojitá proměnná vhodná pro diagnostické odlišování zdravých a nemocných jedinců.
  2. Nalezení dělicího bodu („cut-off point“) na škále hodnot spojité proměnné, který nejlépe odlišuje zdravé a nemocné jedince.

ROC analýza

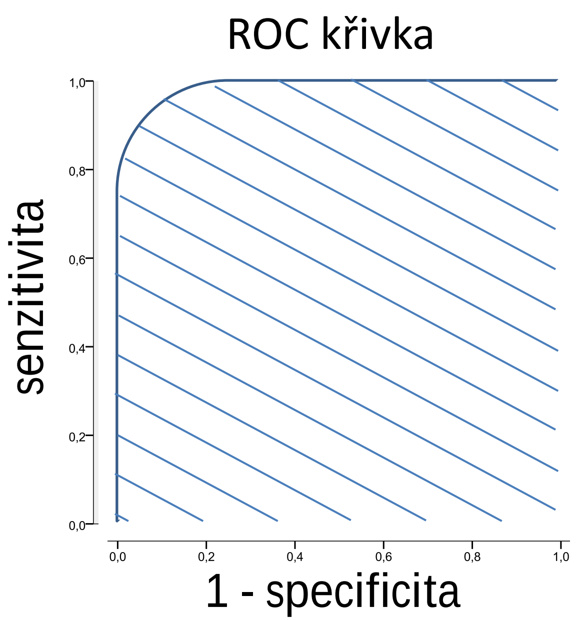
- Princip: Jakákoli hodnota spojité proměnné nějak rozlišuje zdravé a nemocné jedince, tzn. je spojena s nějakou senzitivitou a specificitou.



Nejlepší dělící bod („cut-off“) – nejvyšší sensitivita a specificita pro odlišení skupin – tzn. maximální součet hodnot senzitivity a specificity.

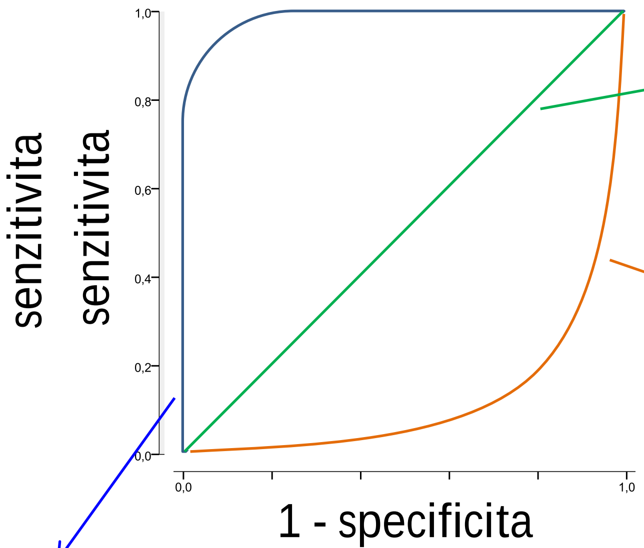
ROC analýza – plocha pod ROC křivkou

- Plocha pod ROC křivkou = „Area Under the Curve“ (AUC).
- Nabývá hodnot od 0 do 1.
- Slouží k vyjádření diagnostické síly (efektivity) testu.
- Čím větší hodnota AUC, tím lepší diagnostický test je (hodnota AUC nad 0,75 většinou poukazuje na uspokojivou diskriminační schopnost testu).



ROC analýza – srovnání diagnostické síly různých testů

ROC křivka



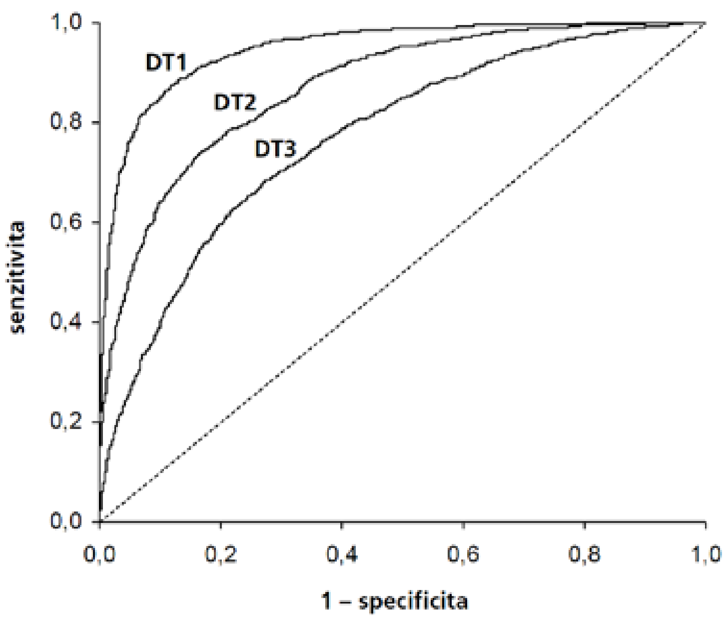
test nediskriminuje vůbec

test diskriminuje „obráceně“

dobře diskriminující test

ROC analýza – srovnání diagnostické síly různých testů

- Lze srovnat i velmi rozdílné testy (např. testy založené na různých proměnných).



Diagnostický test	AUC
DT1	0,949
DT2	0,872
DT3	0,770

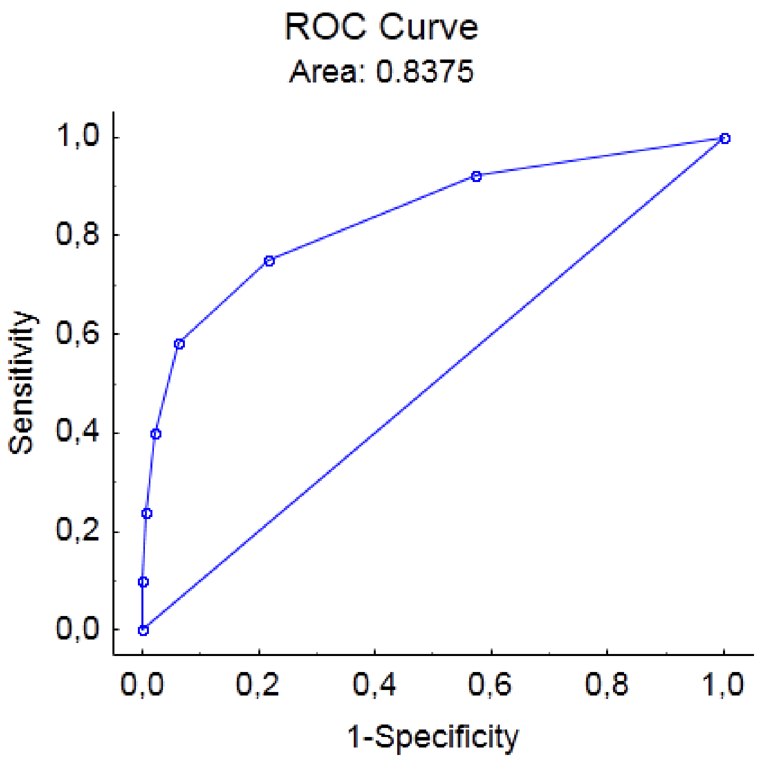
→ nejlepší

→ nejhorší

Zdroj: Dušek, Pavlík, Jarkovský, Koptíková, Analýza dat v Neurologii, Cesk Slov Neurol N 2011; 74/ 107(4)

ROC analýza

Příklad: Zjistěte, zda je MMSE skóre vhodné na diagnostiku mírné kognitivní poruchy (MCI). Najděte dělící bod (cut-off), který nejlépe odlišuje pacienty s MCI od kontrolních subjektů.



MMSE skóre	Sensitivity	1-Specificity	Specificity	Sensitivity + Specificity
-23	0,002	0,000	1,000	1,002
-24	0,101	0,000	1,000	1,101
-25	0,239	0,004	0,996	1,235
-26	0,399	0,022	0,978	1,377
-27	0,581	0,061	0,939	1,520
-28	0,749	0,217	0,783	1,531
-29	0,924	0,574	0,426	1,350
-30	1,000	1,000	0,000	1,000

Blok 5

Jak hodnotit vztah spojitých proměnných a základy regresního modelování.

Osnova

1. Základy korelační analýzy
2. Základy regresní analýzy
3. Analýza přežití a Coxův model proporcionálních rizik

1. Základy korelační analýzy

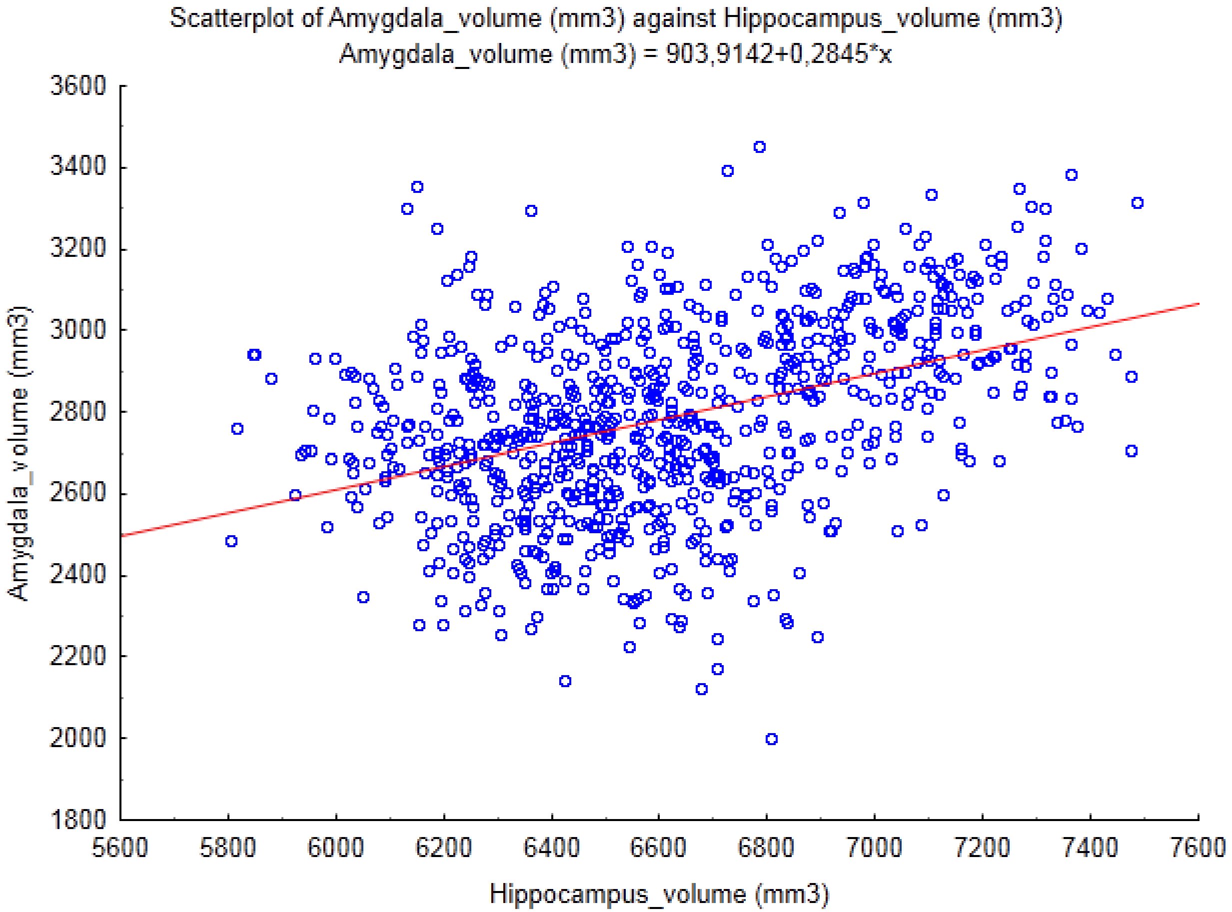
Motivace

- Zatím jsme se zabývali spojitou proměnnou v jedné skupině, spojitou proměnnou ve více skupinách, diskrétní proměnnou v jedné skupině, diskrétní proměnnou ve více skupinách, vztahem dvou diskrétních proměnných.

- Teď se chceme zabývat dvěma spojitými proměnnými:
  1. **Chceme zjistit, jestli mezi nimi existuje vztah** – např. jestli vyšší hodnoty jedné proměnné znamenají nižší hodnoty jiné proměnné.
  2. **Chceme kvantifikovat vztah mezi dvěma spojitými proměnnými** – např. pro použití jedné proměnné na místo druhé proměnné.
  3. **Chceme predikovat hodnoty jedné proměnné na základě znalosti hodnot jiných proměnných.**

Jak hodnotit vztah dvou spojitých proměnných?

- Nejjednodušší formou je **bodový graf (x-y graf)**.
- Např. vztah objemu hipokampu a amygdaly:



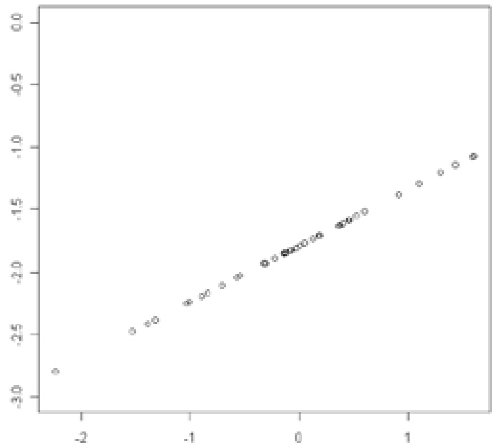
Korelace

- **Korelační koeficient** – kvantifikuje míru vztahu mezi dvěma spojitými proměnnými (X a Y).

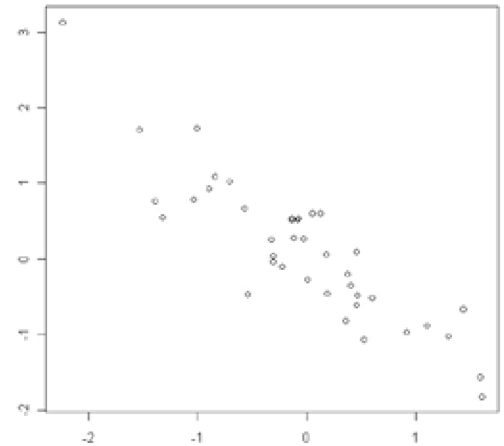
- Standardní metodou je výpočet **Pearsonova korelačního koeficientu (r)**:
  - Charakterizuje **linearitu** vztahu mezi X a Y – jinak řečeno variabilitu kolem lineárního trendu.
  - Nabývá hodnot od -1 do 1.
  - Hodnota r je kladná (kladná korelace), když vyšší hodnoty X souvisí s vyššími hodnotami Y, a naopak je záporná (záporná korelace), když nižší hodnoty X souvisí s vyššími hodnotami Y.
  - Proměnné jsou nekorelované, pokud *r* = 0.
  - Hodnoty 1 nebo -1 získáme, když body x-y grafu leží na přímce.

- Lze statistickým testem **otestovat, zda jsou dvě spojité proměnné nezávislé** – hypotézy mají tvar: $H_0: r = 0$ (tzn. korelační koeficient je roven nule) a $H_1: r \neq 0$.

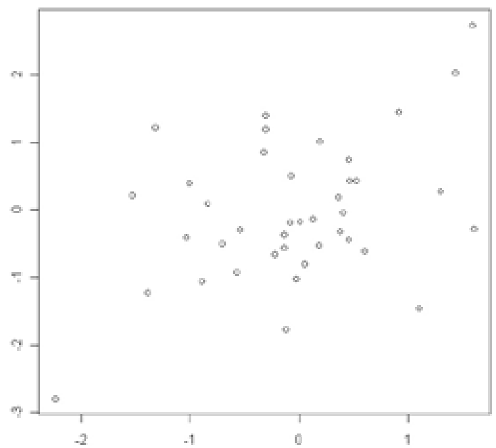
Pearsonův korelační koeficient (*r*)



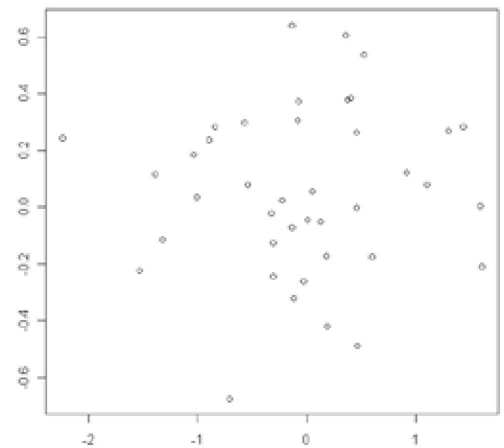
r = 1,0



r = -0,9



r = 0,4

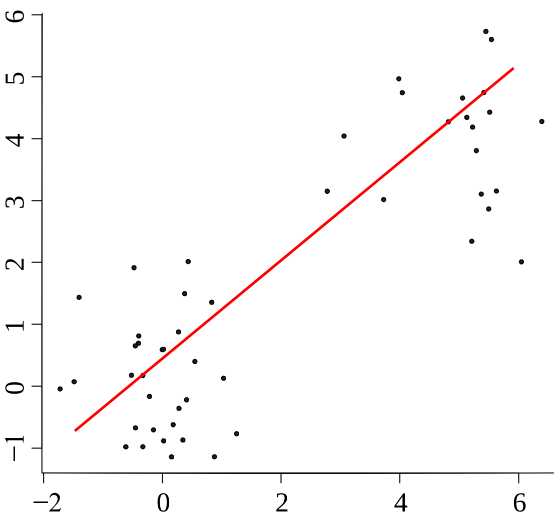


r = 0,05

Pearsonův korelační koef. – problematické situace I.

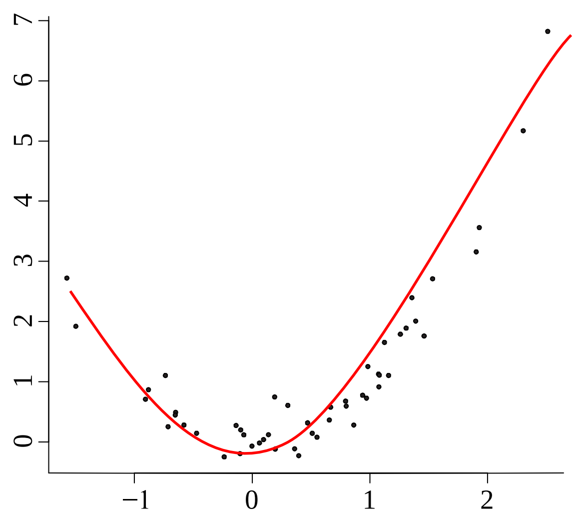
- Pearsonův korelační koeficient není vhodné počítat v situaci, kdy:
  - se v datech vyskytuje více skupin
  - proměnné mají nelineární vztah
  - se v datech vyskytují odlehlé hodnoty

Více skupin



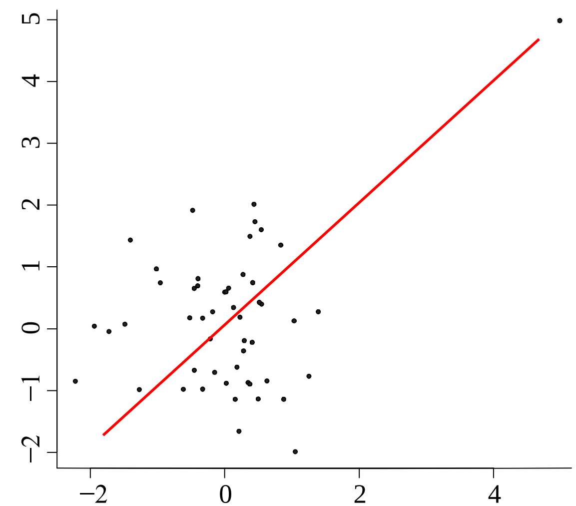
r = 0,84
(p < 0,001)

Nelineární vztah



r = 0,58
(p < 0,001)

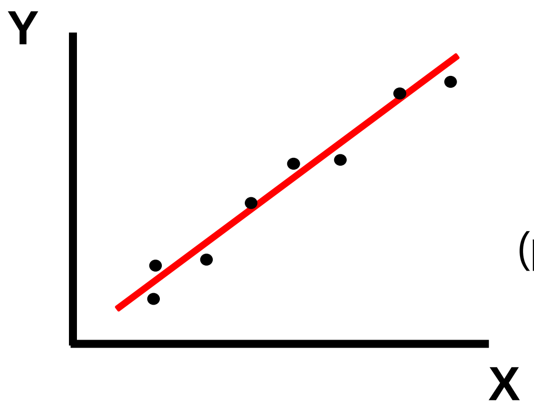
Odlehlá hodnota



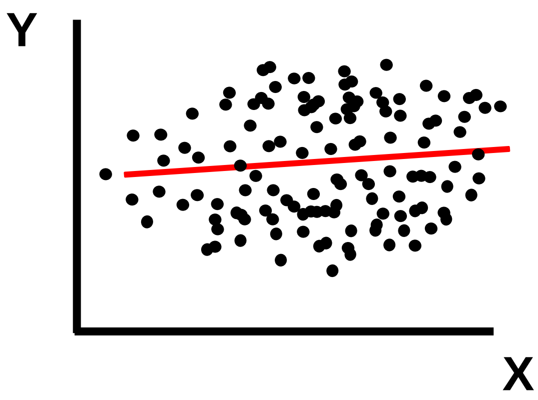
r = 0,36
(p = 0,009)

Pearsonův korelační koef. – problematické situace II.

- Problém velikosti vzorku:



r = 0,891
(p < 0,214)

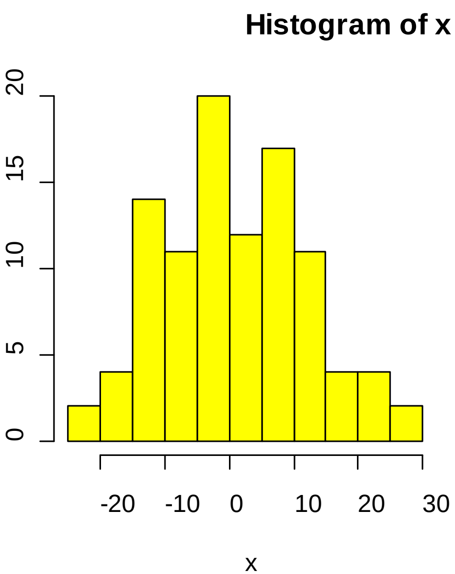


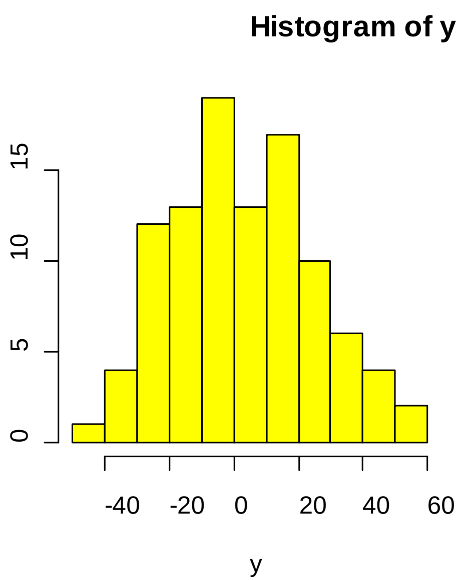
r = 0,012
(p < 0,008)

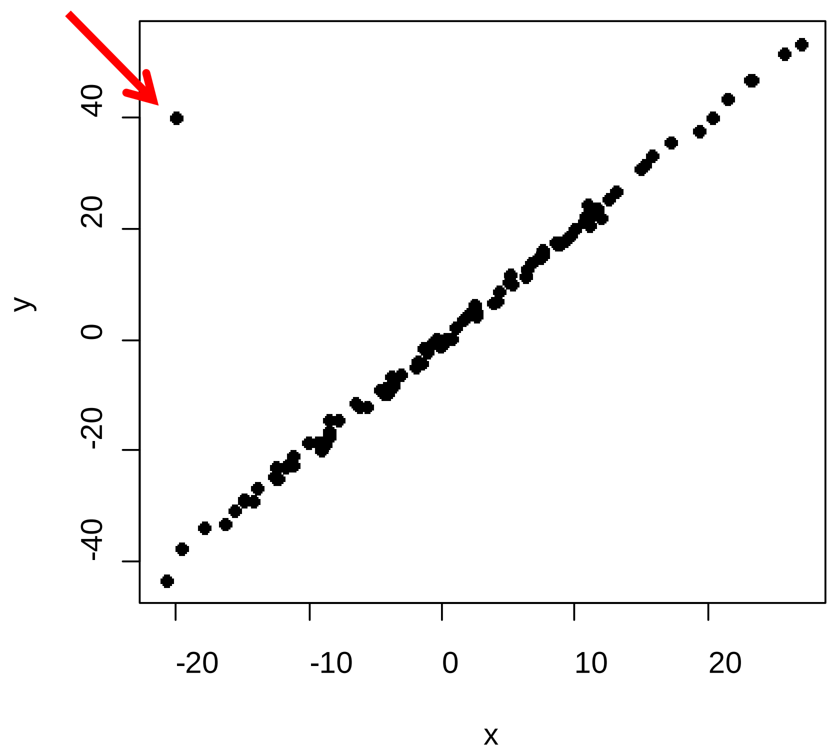
- Test na ověření, zda je Pearsonův korelační koeficient různý od nuly, je parametrický test – předpoklad normality srovnávaných spojitých proměnných!

Pearsonův korelační koef. – problematické situace III.

- Při srovnání dvou spojitých proměnných je nutné vykreslovat bodový graf, protože histogramy pro jednotlivé proměnné zvlášť nám nemusejí odhalit odlehlé hodnoty!



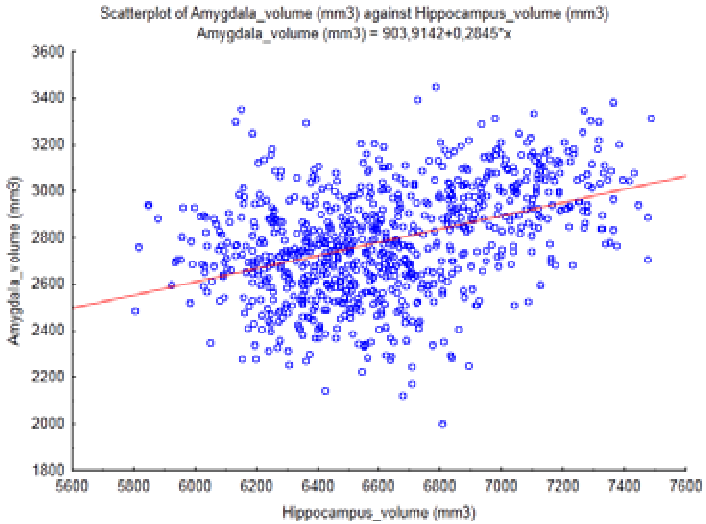




Pearsonův korelační koeficient

- **Příklad:** Ověřte, zda existuje vztah objemu hipokampu a amygdaly v souboru 833 subjektů.

- **Řešení:**



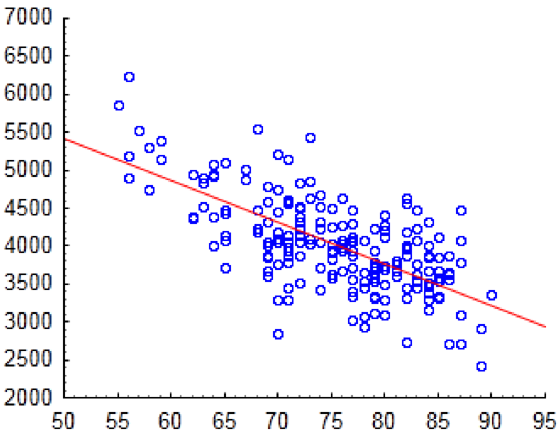
Variable	Correlations (Data_neuro_vycistena4) Marked correlations are significant at p < ,05000 N=833 (Casewise deletion of missing data) **Hippocampus_ volume (mm3)**	Amygdala_volu me (mm3)
Hippocampus_volume (mm3)	1,0000	,4173
	p= ---	p=0,00
Amygdala_volume (mm3)	,4173	1,0000
	p=0,00	p= ---

Úkol 1.

- **Zadání:** Ověřte, zda existuje vztah objemu nucleus caudatus a věku u pacientů s AD, pacientů s MCI a u kontrol. Nezapomeňte ověřit normalitu srovnávaných proměnných.

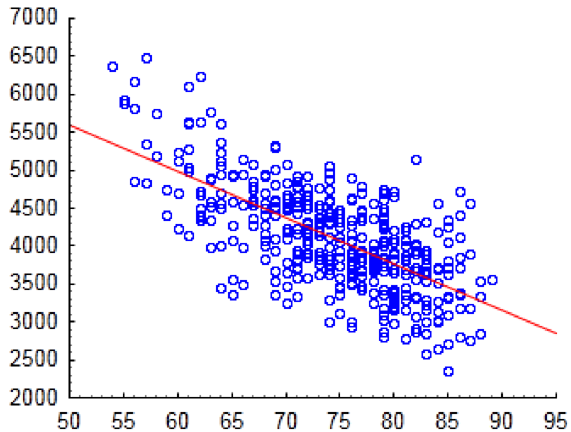
- **Řešení:**

AD



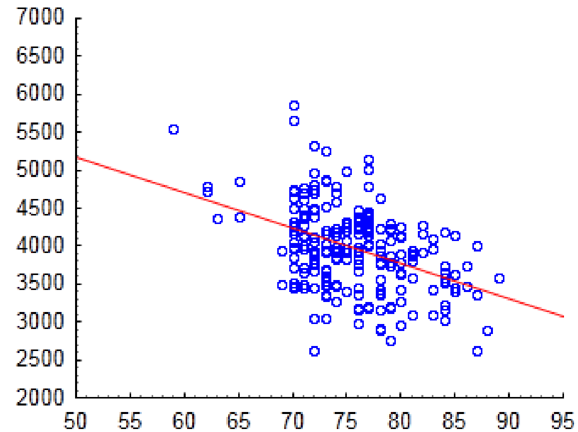
r = 0,68
(*p* < 0,001)

MCI



r = 0,67
(*p* < 0,001)

CN



r = 0,43
(*p* < 0,001)

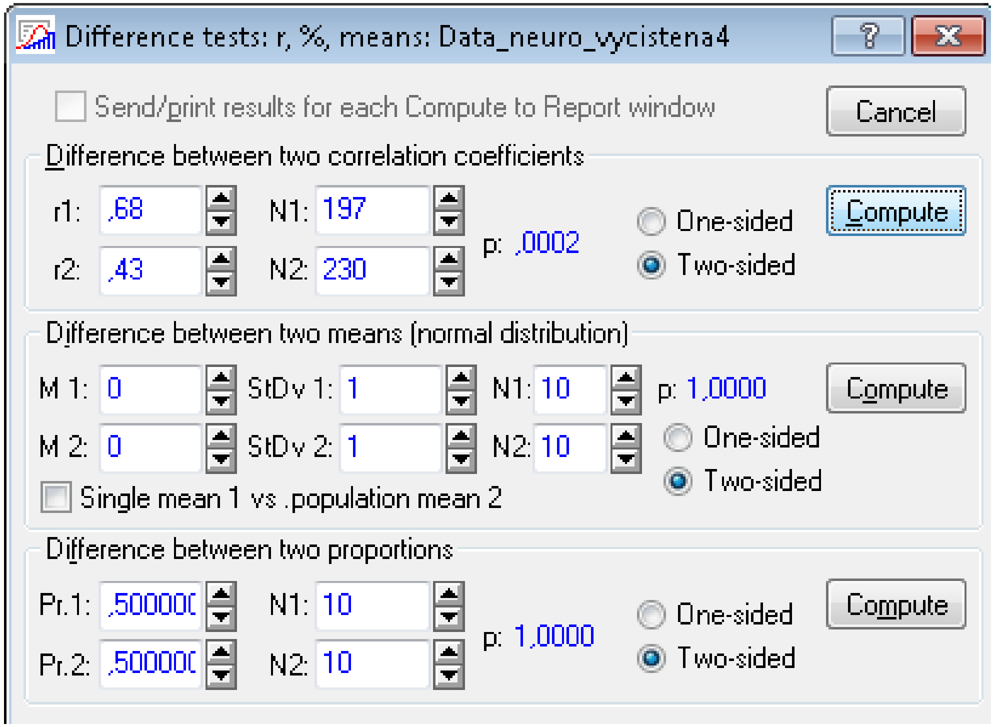
Srovnání dvou korelačních koeficientů

- **Příklad:** Srovnejte korelační koeficienty objemu nucleus caudatus a věku u pacientů s AD a kontrolních subjektů.

- **Postup:**

Z předchozího úkolu víme, že:

$r_1 = 0{,}68$
$N_1 = 197$
$r_2 = 0{,}43$
$N_2 = 230$



Poznámka

- Korelace dvou náhodných veličin se často interpretuje pomocí druhé mocniny Pearsonova korelačního koeficientu: r^2.

- Hodnota r^2 vyjadřuje, kolik % své variability sdílí jedna veličina s druhou, jinak řečeno, kolik % variability jedné veličiny může být predikováno pomocí té druhé.

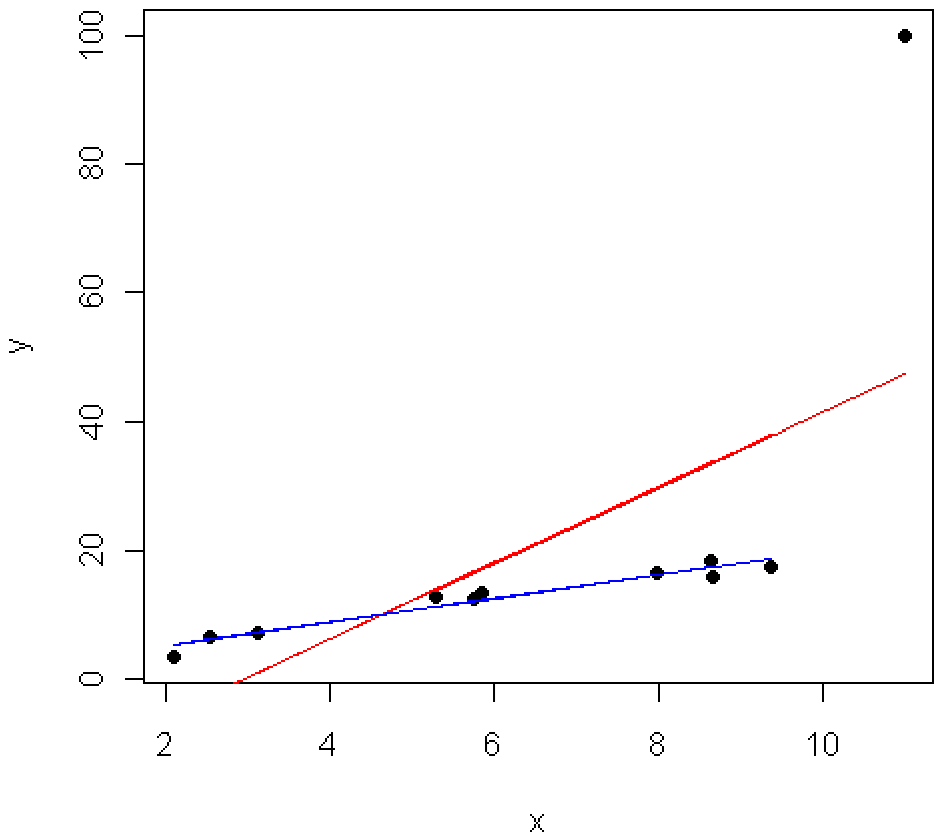
- S hodnotou r^2 se setkáte v lineárních modelech.

Spearmanův korelační koeficient (r_s)

- Pearsonův korelační koeficient je náchylný k odlehlým hodnotám a obecně odchylkám od normality. **Spearmanův korelační koeficient** stejně jako řada dalších neparametrických metod **pracuje pouze s pořadími** pozorovaných hodnot.

- Hodnoty Spearmanova korelačního koeficientu r_s se pohybují stejně jako u Pearsonova korelačního koeficientu r od -1 do 1.

Srovnání Pearsonova a Spearmanova korelačního koeficientu



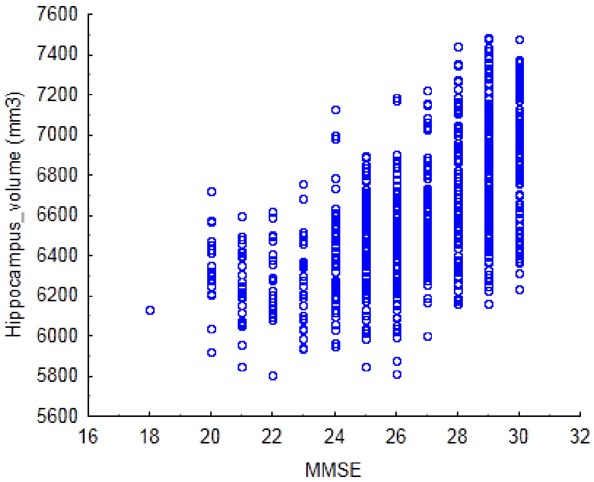
Pearsonův korelační koeficient:
$r = 0{,}65$
($p = 0{,}029$)

Spearmanův korelační koeficient:
$r_S = 0{,}95$
($p < 0{,}001$)

Spearmanův korelační koeficient není náchylný k odlehlým hodnotám.

Spearmanův korelační koeficient

- **Příklad:** Zjistěte, zda existuje vztah objemu hipokampu a MMSE skóre.
- **Řešení:**



Variable	Spearman Rank Order Correlations (Data_neuro_vycistena4) MD pairwise deleted Marked correlations are significant at p <,05000 **MMSE**	Hippocampus_volume (mm3)
MMSE	1,000000	0,626892
Hippocampus_volume (mm3)	0,626892	1,000000

Pair of Variables	Spearman Rank Order Correlations (Data_neuro_vycistena4) MD pairwise deleted Marked correlations are significant at p <,05000 **Valid N**	Spearman R	t(N-2)	p-value
MMSE & Hippocampus_volume (mm3)	833	0,626892	23,19513	0,00

Úkol 2.

- **Zadání:** Zjistěte, zda existuje vztah objemu všech dalších pěti mozkových sktruktur s MMSE skóre (nezapomeňte vykreslit bodové grafy).

- **Řešení:**

Pair of Variables		Spearman Rank Order Correlations (Data_neuro_vycistena4) MD pairwise deleted Marked correlations are significant at p <,05000 Valid N	Spearman R	t(N-2)	p-value
MMSE	**& Amygdala_volume (mm3)**	833	0,338742	10,37852	0,000000
MMSE	& Thalamus_volume (mm3)	833	-0,000759	-0,02187	0,982557
MMSE	& Pallidum_volume (mm3)	833	0,039167	1,12992	0,258834
MMSE	& Putamen_volume (mm3)	833	0,324925	9,90402	0,000000
MMSE	& Nucl_caud_volume (mm3)	833	0,011837	0,34124	0,733012

2. Základy regresní analýzy

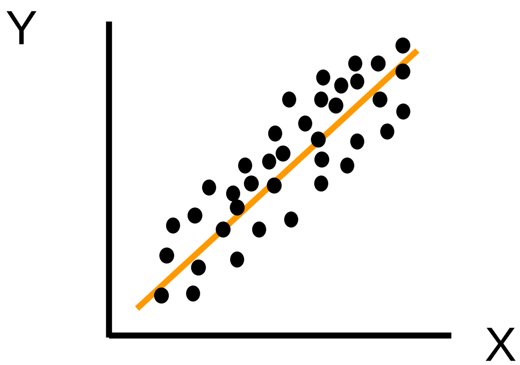
Motivace

- Cílem regresní analýzy je popsat závislost hodnot jedné proměnné na hodnotách druhé proměnné.
- Např. závislost objemu hipokampu na věku.

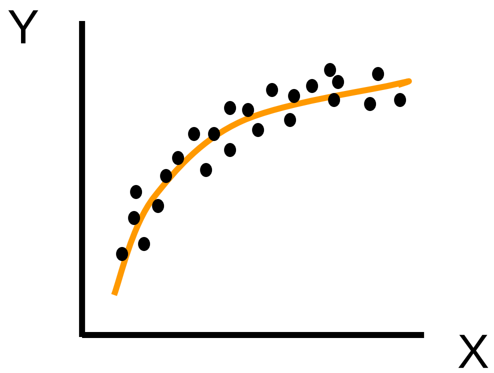
- Dva problémy:
  - Vybrat správnou funkci k popisu dané závislosti.
  - Stanovit konkrétní parametry daného typu funkce.

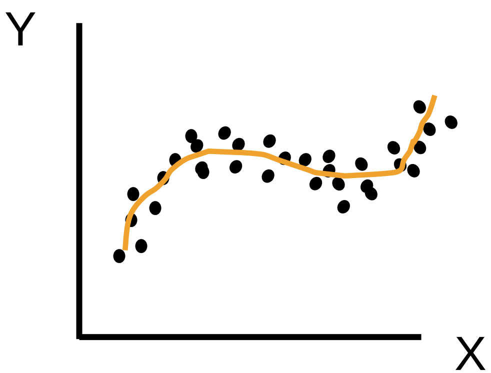
Příklady závislostí

Lineární



Nelineární

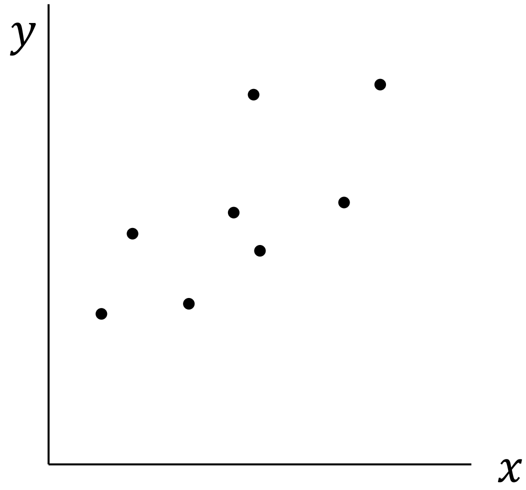




Lineární regrese

Obecný zápis: $\boldsymbol{y} = \boldsymbol{X} * \boldsymbol{\beta} + \boldsymbol{\varepsilon}$

Zápis, pokud máme pouze jednu nezávisle proměnnou: $\boldsymbol{y} = \beta_0 + \beta_1 * \boldsymbol{x} + \boldsymbol{\varepsilon}$



y – závisle proměnná (vysvětlovaná proměnná)

x – nezávisle proměnná (vysvětlující proměnná, regresor)

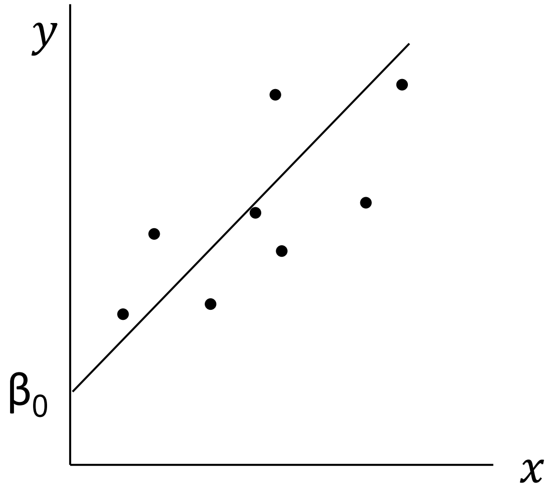
ε – náhodná složka modelu přímky (rezidua přímky)

β_0 – intercept
β_1 – regresní koeficient – „sklon regresní přímky“

Lineární regrese

Obecný zápis: $\boldsymbol{y} = \boldsymbol{X} * \boldsymbol{\beta} + \boldsymbol{\varepsilon}$

Zápis, pokud máme pouze jednu nezávisle proměnnou: $\boldsymbol{y} = \beta_0 + \beta_1 * \boldsymbol{x} + \boldsymbol{\varepsilon}$



y – závisle proměnná (vysvětlovaná proměnná)

x – nezávisle proměnná (vysvětlující proměnná, regresor)

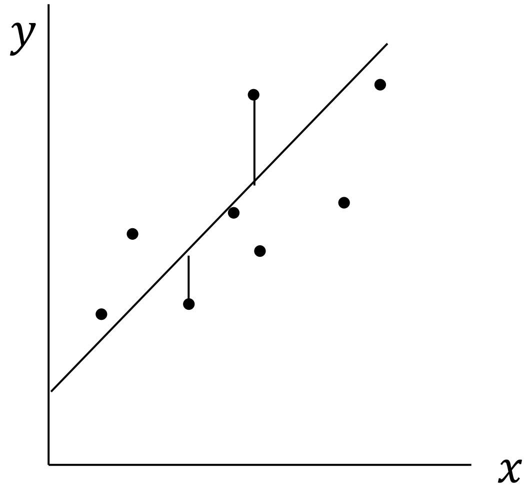
$\boldsymbol{\varepsilon}$ – náhodná složka modelu přímky (rezidua přímky)

β_0 – intercept
β_1 – regresní koeficient – „sklon regresní přímky“

Lineární regrese

Obecný zápis: $\boldsymbol{y} = \boldsymbol{X} * \boldsymbol{\beta} + \boldsymbol{\varepsilon}$

Zápis, pokud máme pouze jednu nezávisle proměnnou: $\boldsymbol{y} = \beta_0 + \beta_1 * \boldsymbol{x} + \boldsymbol{\varepsilon}$



y – závisle proměnná (vysvětlovaná proměnná)

x – nezávisle proměnná (vysvětlující proměnná, regresor)

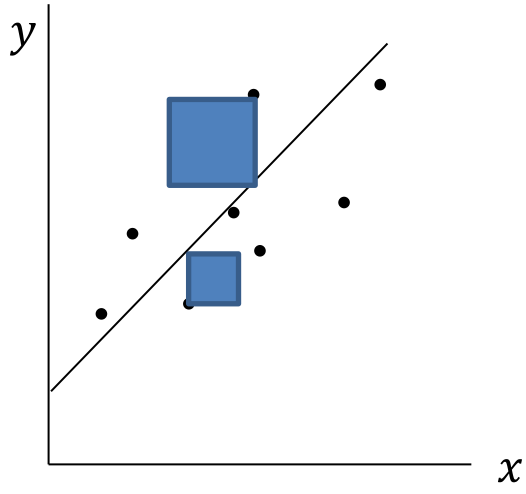
ε – náhodná složka modelu přímky (rezidua přímky)

β_0 – intercept
β_1 – regresní koeficient – „sklon regresní přímky“

Lineární regrese

Obecný zápis: $\boldsymbol{y} = \boldsymbol{X} * \boldsymbol{\beta} + \boldsymbol{\varepsilon}$

Zápis, pokud máme pouze jednu nezávisle proměnnou: $\boldsymbol{y} = \beta_0 + \beta_1 * \boldsymbol{x} + \boldsymbol{\varepsilon}$



y – závisle proměnná (vysvětlovaná proměnná)

x – nezávisle proměnná (vysvětlující proměnná, regresor)

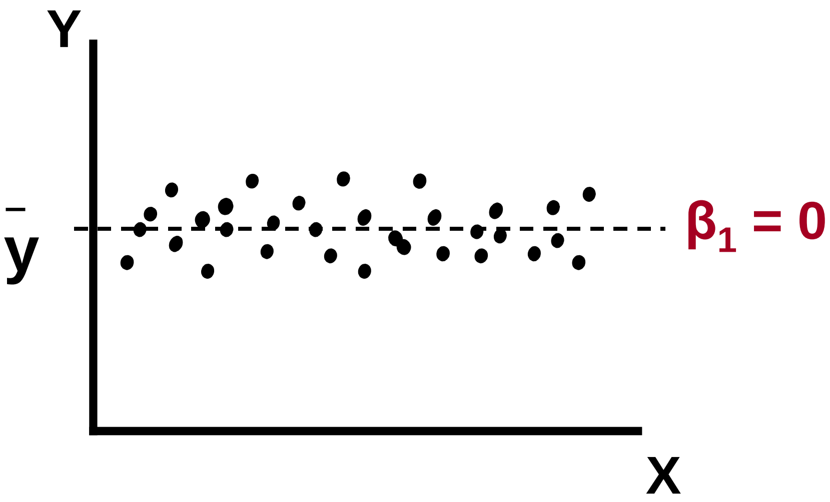
ε – náhodná složka modelu přímky (rezidua přímky)

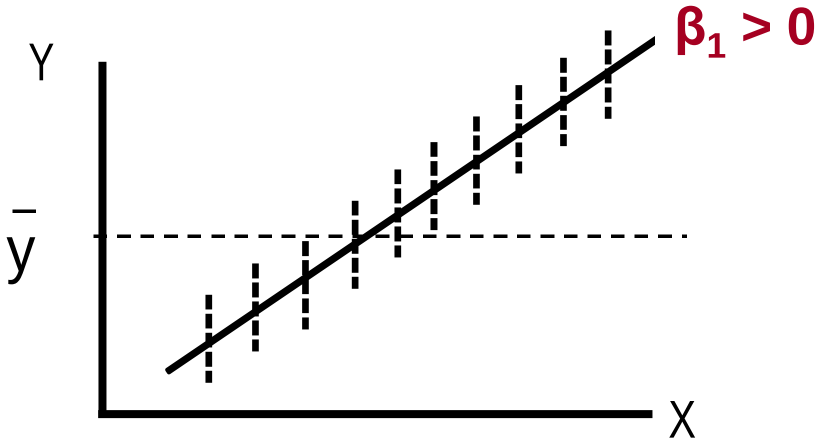
β_0 – intercept
β_1 – regresní koeficient – „sklon regresní přímky“

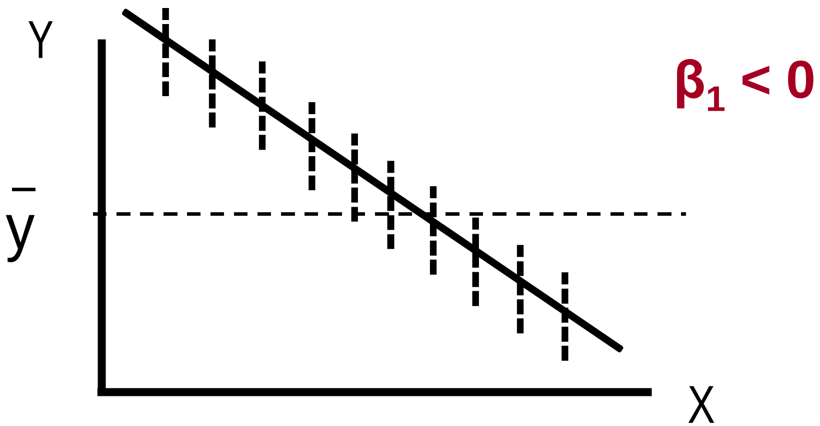
Odhad koeficientů β metodou nejmenších čtverců:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

Lineární regrese - příklady

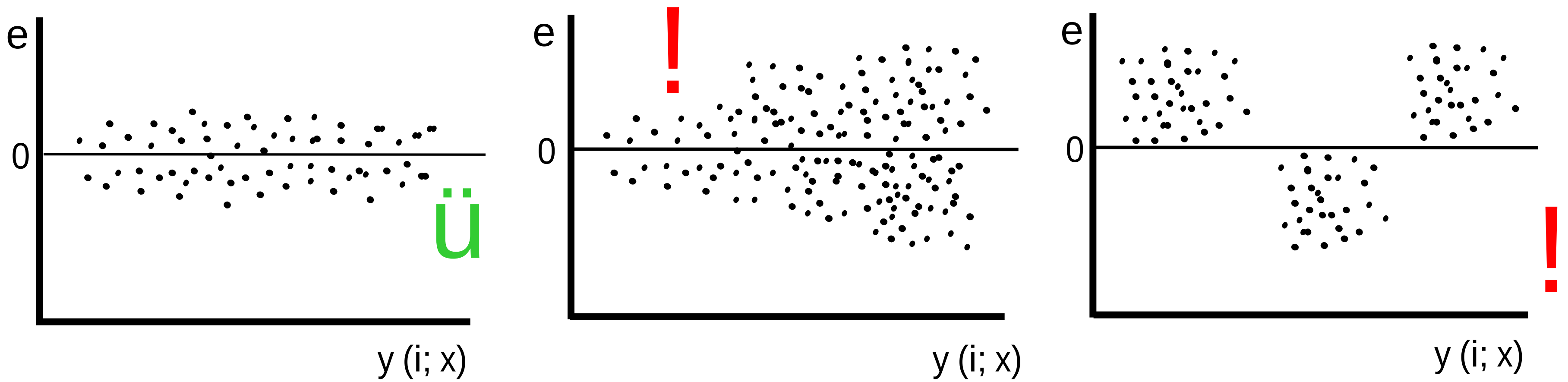




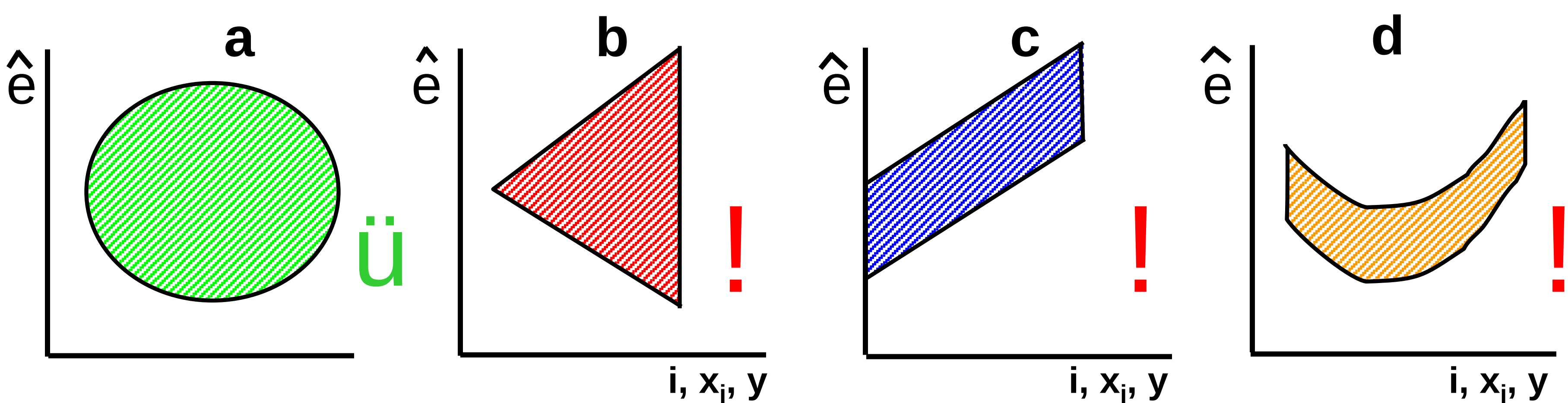


Regresní analýza v grafech

Grafy residuí modelů (příklady)



Obecné tvary residuí modelů (schéma)



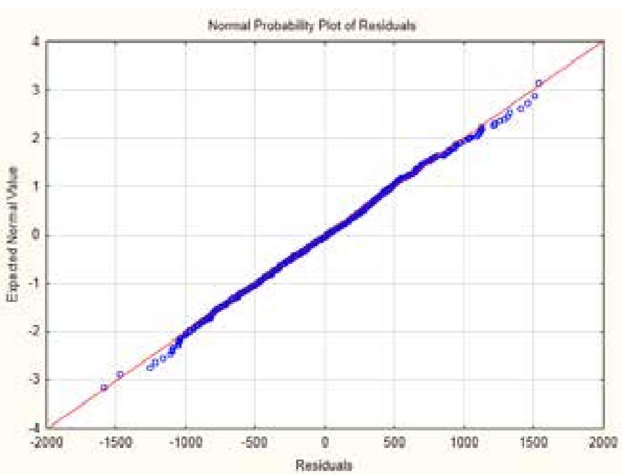
Lineární regrese – příklad I

- **Příklad:** Proveďte regresní analýzu, v níž budete modelovat závislost objemu nucleus caudatus na věku.

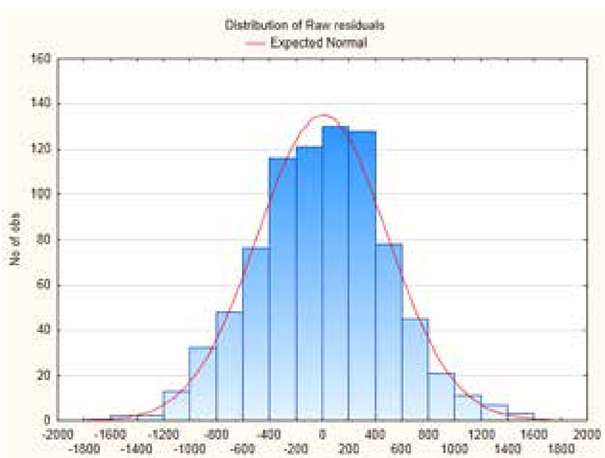
Regression Summary for Dependent Variable: Nucl_caud_volume (mm3) (Data_neuro_vycistena4)
R= ,62657661 R2= ,39259825 Adjusted R2= ,39186732
F(1,831)=537,12 p<0,0000 Std.Error of estimate: 494,97

N=833	b*	Std.Err. of b*	b	Std.Err. of b	t(831)	p-value
Intercept			8348,848	186,0558	44,8728	0,00
Age	-0,626577	0,027036	-57,369	2,4754	-23,1759	0,00

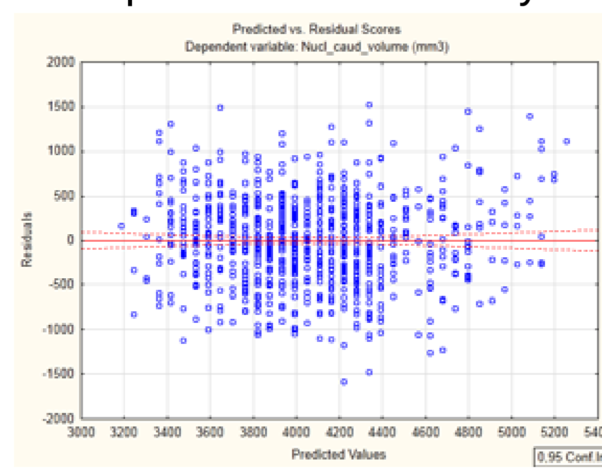
Q-Q graf reziduí



Histogram reziduí



Bodový graf reziduí vs. predikované hodnoty



Lineární regrese – příklad II

- **Příklad:** Chceme zjistit, zda se liší objem nucleus caudatus podle typu onemocnění (pacienti s AD, pacienti s MCI, kontroly). Srovnávané skupiny subjektů však obsahují jiný poměr mužů a žen a liší se i věkovým složením. Odstraňte vliv věku a pohlaví, aby výsledek srovnání objemu nucleus caudatus podle typu onemocnění nebyl ovlivněn tím, že skupiny nejsou srovnatelné.

Data: Spreadsheet1 (8v by 833c)

Data_neuro_vycistena4

	1 Predicted	2 Residuals	3 StandardPredicted	4 StandardResidual	5 StdErrorPredicted	6 MahalanobisDistanc	7 DeletedResidual	8 CookDistance
1	3543.61	-15.89	-1.29	-0.03	31.38	2.35	-15.95	0.00
2	3967.80	-194.34	-0.22	-0.39	26.73	1.43	-194.91	0.00
3	3831.22	463.23	-0.56	0.94	24.44	1.03	464.36	0.00
4	3220.00	365.00	-2.10	0.74	44.91	5.85	368.03	0.00
5	4255.41	-532.15	0.50	-1.08	27.70	1.61	-533.82	0.00
6	4312.93	-343.56	0.65	-0.69	28.54	1.77	-344.71	0.00
7	3277.52	-391.29	-1.95	-0.79	42.93	5.26	-394.25	0.00
8	3392.57	348.66	-1.66	0.70	39.14	4.20	350.85	0.00
9	3507.61	229.79	-1.38	0.46	35.63	3.31	230.99	0.00
10	3335.05	-704.48	-1.81	-1.42	41.00	4.71	-709.34	0.00
11	3852.75	-960.02	-0.51	-1.94	27.92	1.65	-963.08	0.00
12	3543.61	7.62	-1.29	0.02	31.38	2.35	7.65	0.00
13	4348.93	285.38	0.74	0.58	25.97	1.29	286.17	0.00
14	3831.22	50.65	-0.56	0.10	24.44	1.03	50.77	0.00
15	4291.41	-837.95	0.59	-1.69	24.83	1.10	-840.07	0.00

Vícenásobná lineární regrese

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

$$\begin{array}{c}\boldsymbol{y}\\ \begin{bmatrix}25\\36\\58\\\cdots\\ \\ \\ \end{bmatrix}\end{array} = \text{individuals}\begin{array}{c}\boldsymbol{X}\\ \begin{array}{|c|llll|}\hline & \mathbf{x}_1 & \mathbf{x}_2 \; \ldots & & \mathbf{x}_p \\ \hline \mathbf{I}_1 & \mathbf{1} & & & \\ \mathbf{I}_2 & \mathbf{1} & & & \\ \mathbf{I}_3 & \mathbf{1} & & & \\ \mathbf{I}_4 & \mathbf{1} & & & \\ \ldots & & & & \\ \mathbf{I}_n & \mathbf{1} & & & \\ \hline \end{array}\\ \text{voxels}\end{array} * \begin{array}{c}\boldsymbol{\beta}\\ \begin{bmatrix}\beta_1\\\beta_2\\\beta_3\\\beta_4\\\cdots\\ \\ \beta_p\end{bmatrix}\end{array} + \begin{array}{c}\boldsymbol{\varepsilon}\\ \begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\\\varepsilon_4\\\cdots\\\varepsilon_n\end{bmatrix}\end{array}$$

***X* – matice plánu (design matice)**

3. Analýza přežití a Coxova regrese

Analýza přežití

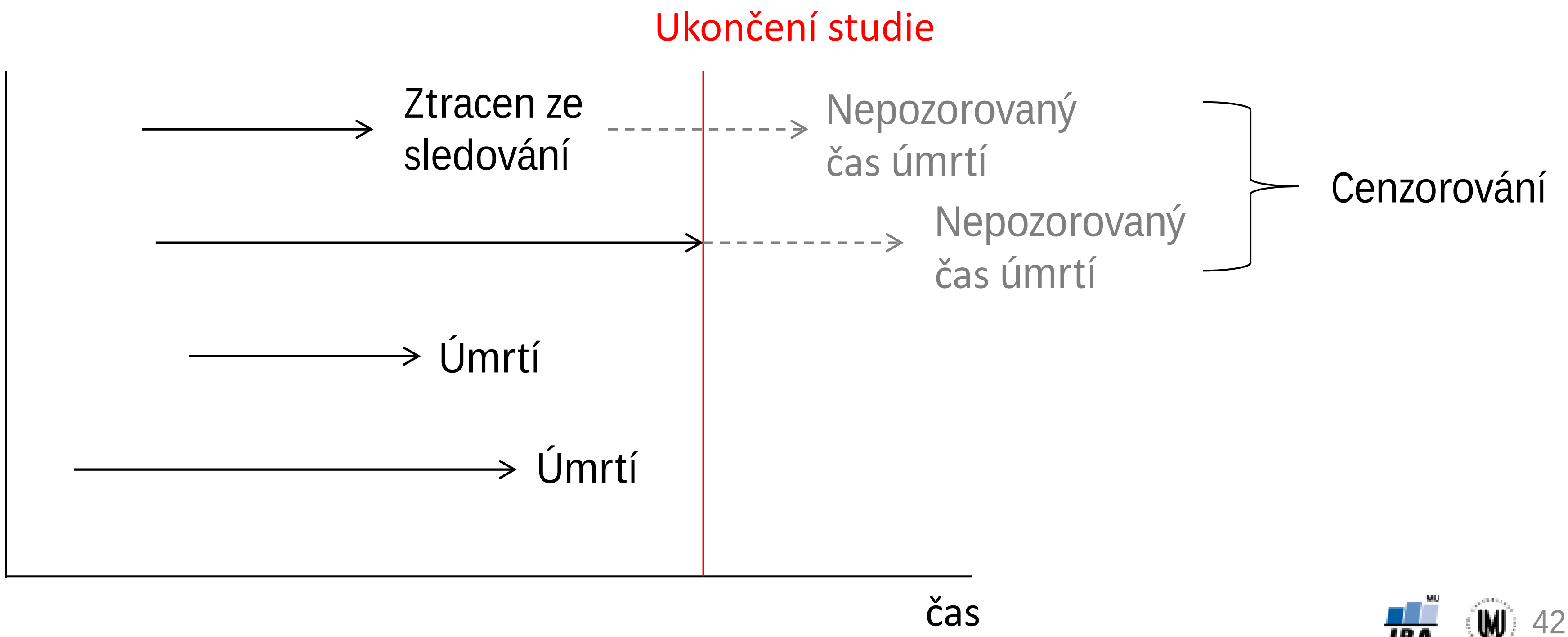
- Analýza doby od vstupní události do výskytu sledované (koncové) události.
- **Vstupní událost** – např.:
  - narození
  - začátek léčby
  - počátek nemoci
  - vstup do studie
  - dosažení remise (uzdravení pacienta)
  - operace
  - začátek používání přístroje
- **Koncová (sledovaná) událost** – např.:
  - úmrtí
  - výskyt progrese onemocnění
  - výskyt relapsu onemocnění (tzn. návrat onemocnění)
  - dosažení remise (uzdravení pacienta)
  - porucha přístroje
- Dobu mezi vstupní a koncovou událostí nazýváme jako **doba přežití**.

Analýza přežití

- Uplatnění analýzy přežití:
  - v lékařském výzkumu:
    - hodnocení celkového přežití („Overall Survival“ – OS)
    - hodnocení času do progrese („Progression-free Survival“ – PFS)
    - a další
  - v průmyslu
  - v zemědělství

Cenzorování

- Doba přežití dané osoby je cenzorována, pokud během pozorování nenastane sledovaná událost – důvody:
  - ztráta kontaktu s danou osobou (osoba se přestěhuje nebo přestane chodit na prohlídky)
  - v době uzavření studie se sledovaná událost ještě nevyskytla
  - pozorovaná osoba zemřela na jinou nemoc



Analýza přežití – vstupní data

- Pro každého člověka dvojice hodnot:
  - T – čas přežití:
    - u osob, u nichž sledovaná událost nastala, je to čas od vstupní události do sledované události
    - u osob, u nichž sledovaná událost nenastala, je to čas od vstupní události do ukončení studie
  - C – identifikátor cenzorování – pouze 2 hodnoty:
    - 1 ... sledovaná událost nastala
    - 0 ... sledovaná událost nenastala (osoba je cenzorována)

- Cenzorovaná pozorování (tzn. osoby, u nichž sledovaná událost nenastala) nesmíme z analýzy vyhodit – obsahují informaci!!!

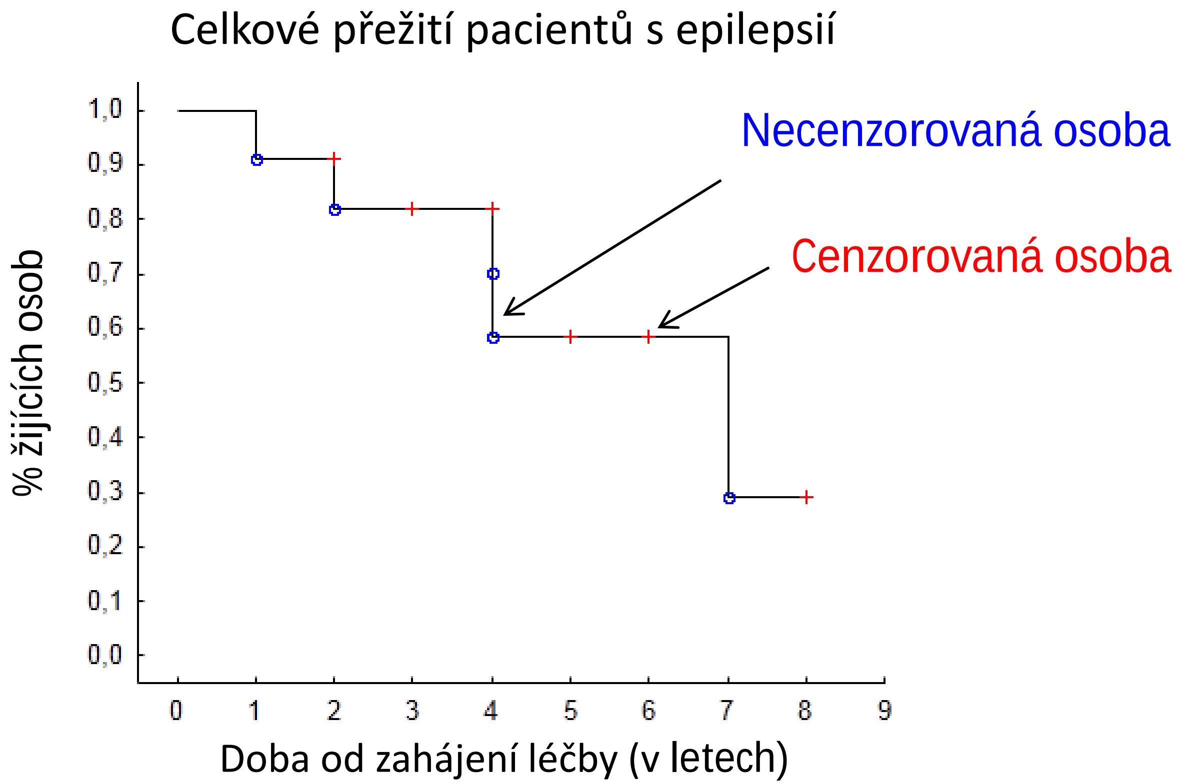
Neparametrické odhady křivky přežití

- V klinickém výzkumu i populačním modelování jsou pro popis přežití standardem neparametrické metody – Kaplan-Meierova metoda a metoda Life-tables.

1. **Kaplan-Meierova metoda** – založena na jednotlivých pozorovaných časech přežití, vhodná zejména pro hodnocení dat klinických studií – vyžaduje přesný záznam doby sledování.

2. **Life-tables** – založena na agregaci pozorování do časových intervalů, vhodná zejména pro popis přežití na populační úrovni, kde není k dispozici tak kvalitní záznam doby sledování.

Kaplan-Meierova křivka přežití



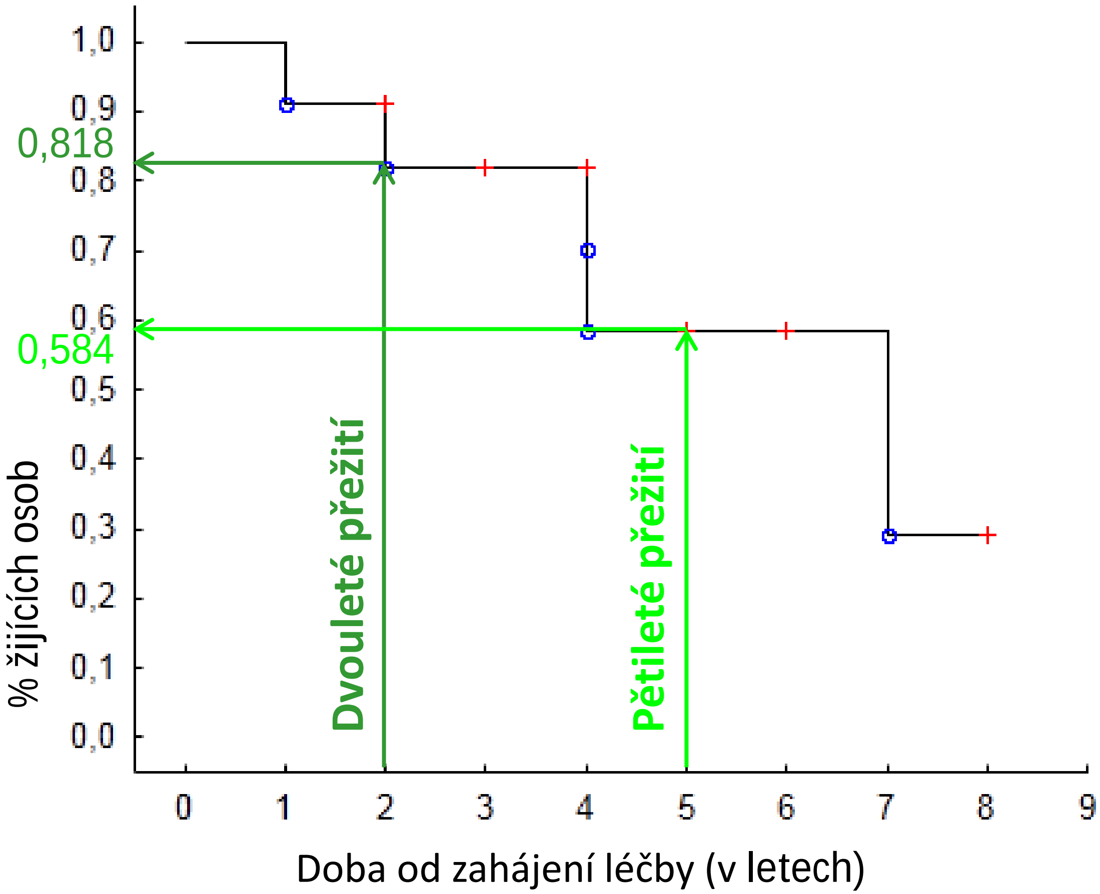
Vstupní data

Pacient	Čas	Úmrtí
1	1	1
2	2	1
3	2	0
4	3	0
5	4	1
6	4	1
7	4	0
8	5	0
9	6	0
10	7	1
11	8	0

- Křivka přežití odráží **procento žijících osob v daném čase**.
- Křivka přežití je **klesající**:
  - pokles křivky přežití nastane v čase koncové události u necenzorované osoby
  - velikost „schodu“ je dána počtem osob, kteří v daném čase zůstávají „v riziku“

Křivka přežití – hodnocení x-letého přežití

Celkové přežití pacientů s epilepsií



Kaplan-Meier (Product-limit) analysis (Data_preziti) Note: Censored cases are marked with +

	Time	Cumulative Survival	Standard Error
1	1	0,909	0,087
2	2	0,818	0,116
3+	2		
4+	3		
5	4	0,701	0,147
6	4	0,584	0,163
7+	4		
8+	5		
9+	6		
10	7	0,292	0,222
11+	8		

2-leté přežití: 81,8% (tzn. po dvou letech od zahájení léčby žije 81,8% pacientů)
5-leté přežití: 58,4% (tzn. po pěti letech od zahájení léčby žije 58,4% pacientů)

Křivka přežití – hodnocení x-letého přežití – výpočet intervalu spolehlivosti

Výpočet intervalu spolehlivosti pro x-leté přežití:

Odhad pravděpodobnosti přežití v čase t ± 1,96 × SE (odhadu)

Kaplan-Meier (Product-limit) analysis (Data_preziti) Note: Censored cases are marked with +			
	Time	Cumulative Survival	Standard Error
1	1	0,909	0,087
2	2	0,818	0,116
3+	2		
4+	3		
5	4	0,701	0,147
6	4	0,584	0,163
7+	4		
8+	5		
9+	6		
10	7	0,292	0,222
11+	8		

Dolní mez IS:

0,818 - 1,96 * 0,116 = 0,591

0,584 - 1,96 * 0,163 = 0,265

Horní mez IS:

0,818 + 1,96 * 0,116 = 1,045

0,818 + 1,96 * 0,116 = 0,903

Závěr: **2-leté přežití: 81,8%** (59,1%; 100,0%)
5-leté přežití: 58,4% (26,5%; 90,3%)

Upozornění

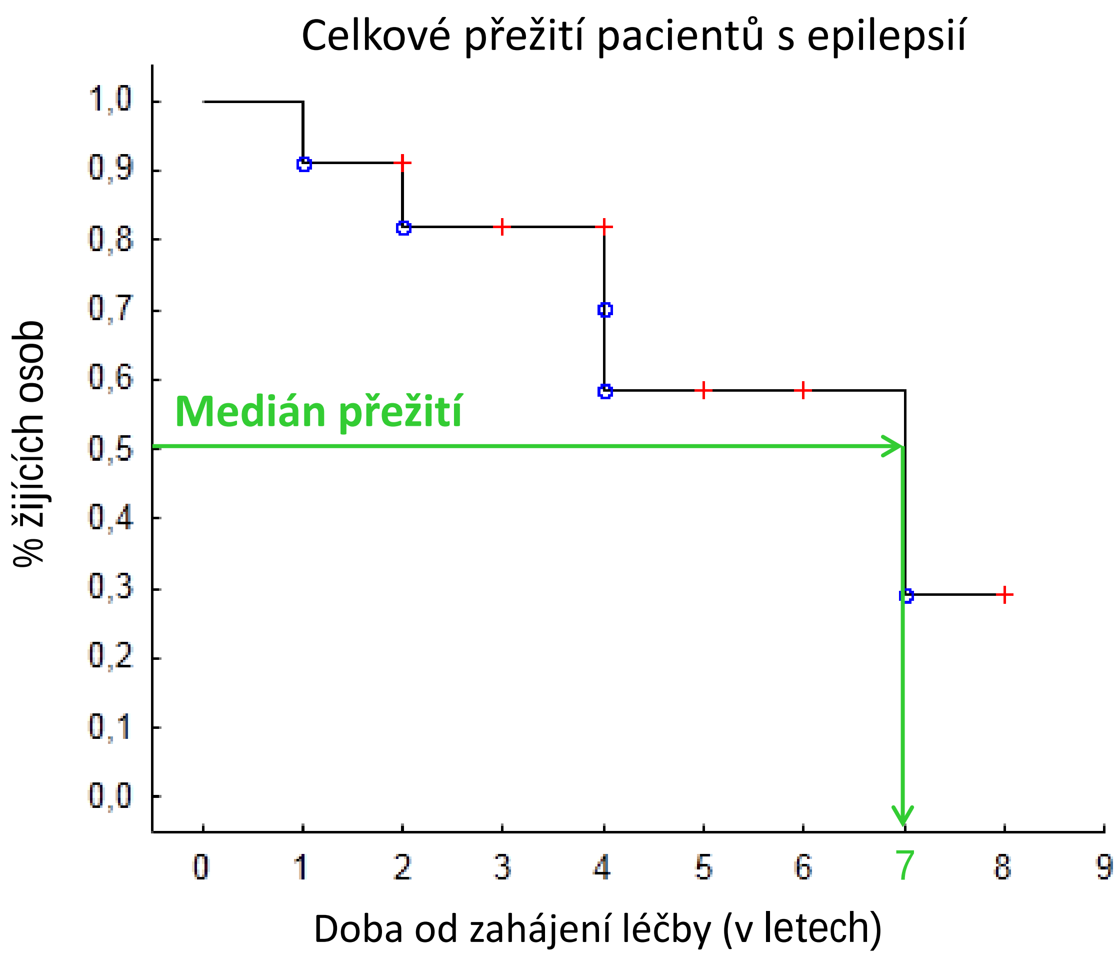
- Pokud horní mez intervalu spolehlivosti vyjde větší než 1 (tzn. 100%), je třeba toto číslo nahradit hodnotou 1 (resp. 100%).

- Pokud dolní mez intervalu spolehlivosti vyjde menší než 0 (tzn. 0%), je třeba toto číslo nahradit hodnotou 0 (resp. 0%).

- **Procento žijících lidí totiž nemůže být větší než 100% a menší než 0% !**

Křivka přežití – medián přežití

Medián přežití je čas od vstupní události, v němž je pravděpodobnost přežití 50%, tedy čas, kterého se podle očekávání dožije polovina sledovaných pacientů.

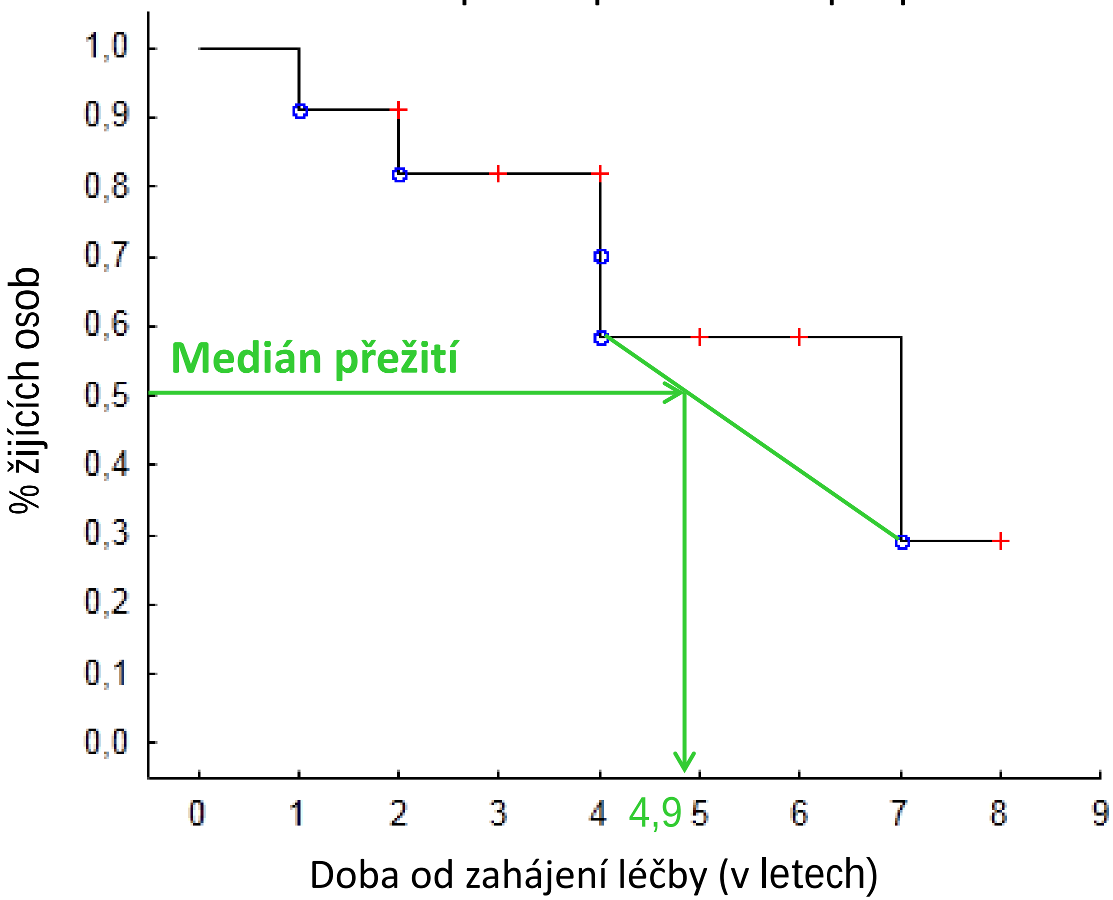


Závěr:
Medián přežití je 7 let.
(Tzn. čas, kterého se dožila polovina sledovaných pacientů, je 7 let.)

Křivka přežití – medián přežití v softwaru STATISTICA

Software STATISTICA provádí interpolaci – počítá aproximativní medián. Nevadí to, pokud je velký počet sledovaných událostí.

Celkové přežití pacientů s epilepsií



	Survival Time
25'th percentile (lower quartile)	3,2
50'th percentile (median)	4,9
75'th percentile (upper quartile)	

Závěr:
Medián přežití je 4,9 let.

Poznámky

- **Sestrojovat křivky přežití za každou cenu je mnohdy zavádějící** – zvláště v případě použití stratifikačních kritérií vedoucích k nízkým N ve skupinách.
 è riziko zkreslení a dezinterpretace výsledků!!

- Podíl cenzorovaných pozorování je důležitou charakteristikou – je vhodné uvádět:
  – Podíl pacientů ztracených ze sledování (lost to follow-up).
  – Podíl pacientů „bez události“ na konci studie.

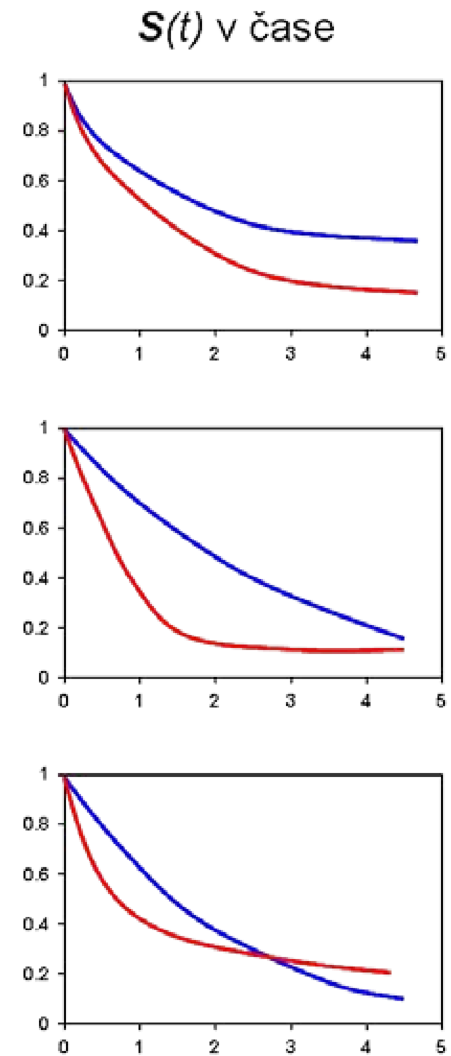
- S křivkami přežití by měla být vždy reportována **maximální a minimální doba sledování** dosažitelná ve studii (dáno začátkem náboru a datem ukončení studie) – samozřejmě ve vztahu k události, která nás zajímá.

- Je nutné mít na paměti **nízkou věrohodnost „konce“ křivky přežití** – zůstává-li ve studii 10 pacientů nebo méně, „skoky“ v přežití s každou další událostí jsou dramatické.

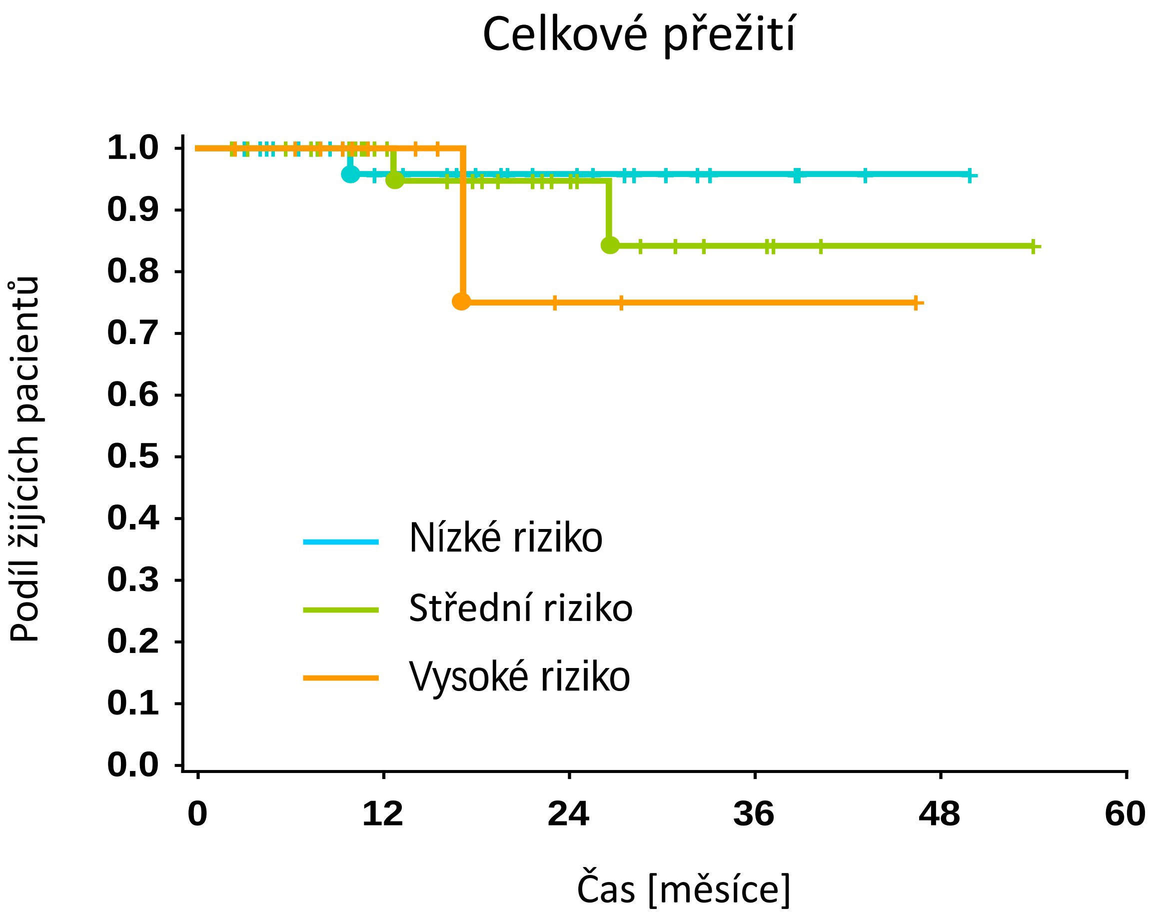
Srovnání křivek přežití

- Častým cílem klinického výzkumu je srovnání přežití dvou a více skupin pacientů
- Standardem v analýze klinických dat jsou opět neparametrické testy:
  - Log-rank test
  - Gehanův-Wilcoxonův test

- **Log-rank test** je zaměřen na srovnání očekávaných a pozorovaných počtů událostí v jednotlivých skupinách
- **Gehanův-Wilcoxonův test** umožňuje klást větší důraz na rozdíly v raných fázích sledování pacientů.

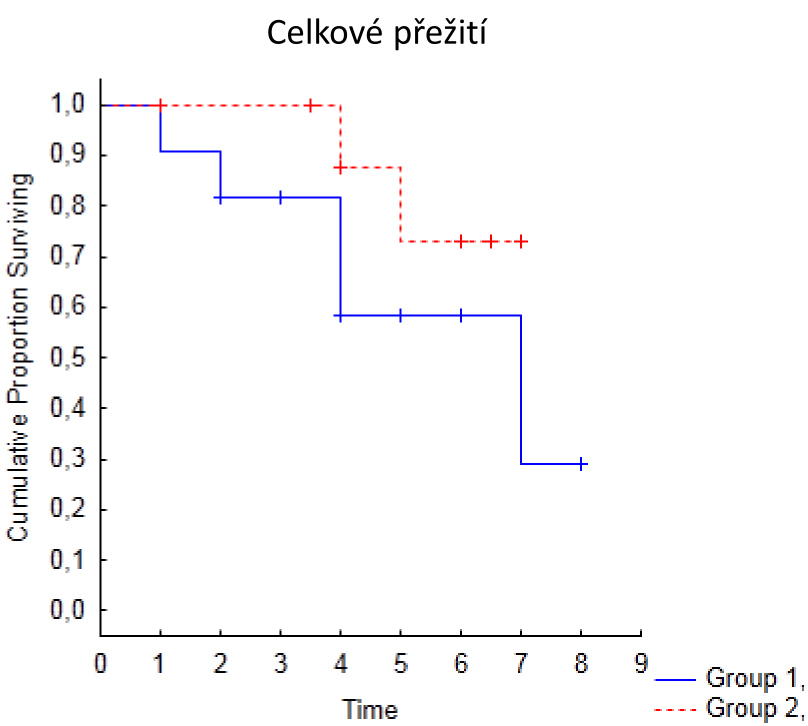


Srovnání křivek přežití – ukázka



Srovnání dvou křivek přežití – příklad

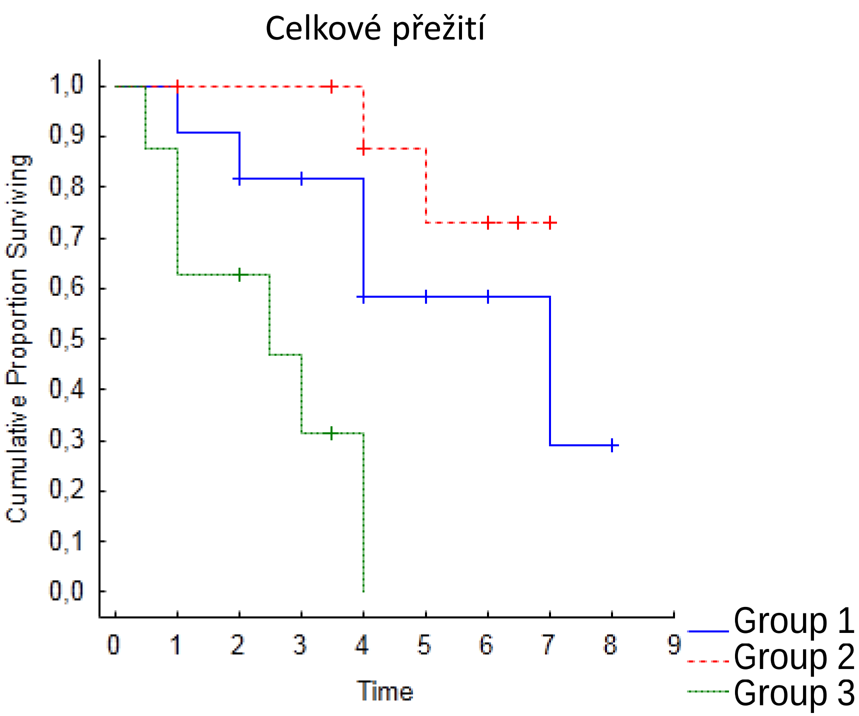
- **Příklad:** Srovnejte přežití pacientů, kteří byli léčeni dvěma různými preparáty.



Life Table for Group 1 and Group 2 (Data_preziti_2kat) Group 1: Code 1, Group 2: Code 2		
	Group 1: Cum.%	Group 2: Cum.%
1,00	100,0	100,0
1,78	90,9	100,0
2,56	81,3	100,0
3,33	81,3	100,0
4,11	56,3	87,5
4,89	56,3	87,5
5,67	56,3	72,9
6,45	56,3	72,9
7,22	28,2	72,9
8,00	28,2	

Srovnání třech křivek přežití – příklad

- **Příklad:** Srovnejte přežití pacientů se třemi diagnózami (CML, CLL a AML).



	Group 1	Group 2	Group 3
,50	100,0	100,0	100,0
1,33	90,9	100,0	62,5
2,17	81,3	100,0	62,5
3,00	81,3	100,0	46,9
3,83	81,3	100,0	28,1
4,67	56,3	86,7	0,0
5,50	56,3	72,2	0,0
6,33	56,3	72,2	0,0
7,17	28,2	72,2	0,0
8,00	28,2		0,0

Coxův model proporcionálních rizik

- Analýza vlivu prognostických faktorů onemocnění na přežití pacientů, na dosažení remise apod.

- **Příklad: Testování vlivu binární proměnné** (např. užívání léčby B) **na celkové přežití.**

Výsledek analýzy:

Variable	Hazard ratio (HR) (poměr rizik)	95% conf. Int.	P-value
DRUG B	2.18	1.4 – 3.5	0.001

Interpretace:
U pacientů užívajících v období před vstupem do studie přípravek B, je více jak dvojnásobně vyšší riziko úmrtí ve sledovaném období než u pacientů neužívajících přípravek B.

Coxův model proporcionálních rizik

- **Příklad: Testování vlivu kategoriální proměnné** (např. věk při diagnóze) **na celkové přežití.**

Výsledek analýzy:

Age group	HR	95% conf. Int.	P-value
1: [20-29]	1.0		
2: [30-34]	3.31	1.37-8.01	<0.001
3: [35-39]	3.72	1.51-9.14	<0.001
4: [40-54]	6.43	2.56-16.12	<0.001

Interpretace:
S rostoucím věkem při diagnóze roste riziko úmrtí ve sledovaném období.
Nárůst rizika je vztažen k věkové skupině 20 - 29 let.

Coxův model proporcionálních rizik

- **Příklad: Testování vlivu spojité proměnné** (např. věk při diagnóze) **na celkové přežití.**

Výsledek analýzy:

Variable	HR	95% conf. Int.	P-value
$AGE_{[5\ years\ interval]}$	1.50	1.3 – 1.8	<0.001

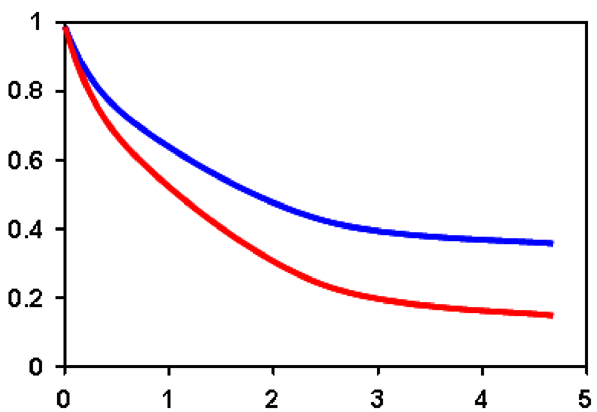
<u>Interpretace:</u>

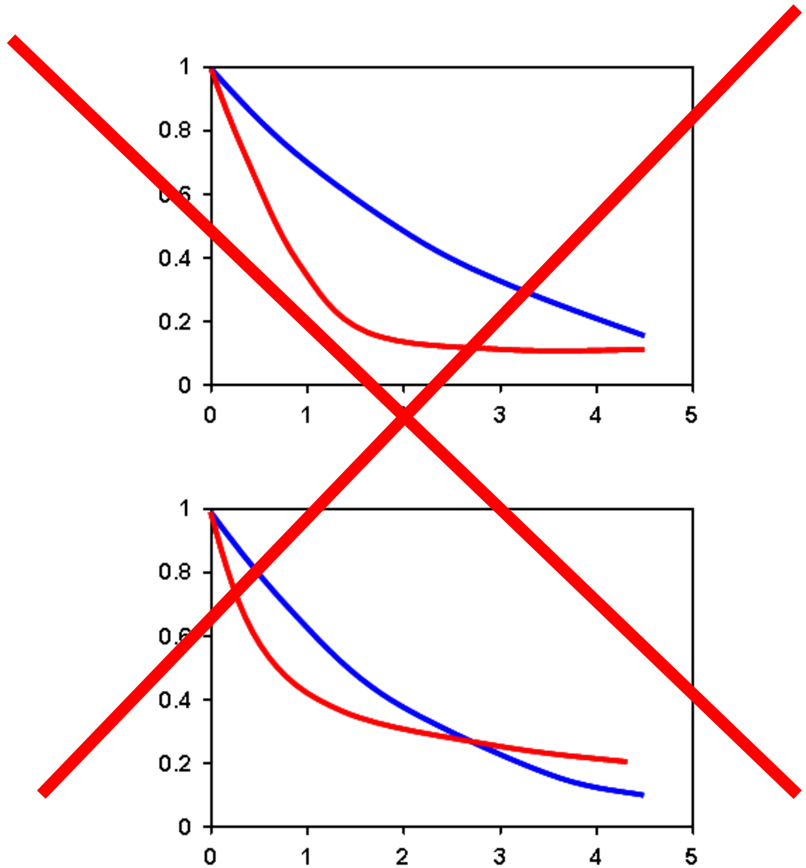
Nárůst věku při diagnóze o 5 let zvyšuje riziko úmrtí 1,5-krát.

Coxův model proporcionálních rizik – důležité!

- Coxův model má smysl počítat, pokud se křivky přežití od sebe postupně oddalují. Pokud se křivky oddálí a pak se zase přiblíží nebo pokud se dokonce křivky překříží, tak nelze Coxův model pro výpočet poměru rizik (hazard ratio) použít!

OK





Coxův model proporcionálních rizik – příklad

- **Příklad:** Vypočtěte poměr rizika úmrtí podle ECOG skóre.

	Parameter Estimates (Data_Cox)							
	Standard Error	Chi-square	P value	95% Lower CL	95% Upper CL	Hazard Ratio	95% Hazard Ratio Lower CL	95% Hazard Ratio Upper CL
ECOG1	0,192764	20,60900	0,000006	0,497283	1,252903	2,399098	1,644247	3,500491