

Typy znaků, které se vyskytují v lékařských datových souborech

Alternativní (binární) znaky: nabývají jen dvou variant, často se kódují 1 (přítomnost sledovaného znaku) a 0 (nepřítomnost sledovaného znaku).

Např. *pohlaví pacienta*: 1 – muž, 0 – žena,

úmrtí pacienta: 1 – zemřel, 0 – žije,

hormonální léčba: 1 – byla, 0 – nebyla,

radioterapie: 1 – byla, 0 – nebyla,

operace: 1 – byla, 0 – nebyla,

relaps (návrat onemocnění, lokální či vzdálený): 1 – nastal, 0 – nenastal.

apod.

Popisují se **tabulkou rozložení absolutních a relativních četností**, dva alternativní znaky pak **čtyřpolní kontingenční tabulkou** simultánních absolutních nebo relativních četností a kontingenční tabulkou řádkově nebo sloupcově podmíněných relativních četností dvojic variant znaků.

Ukázka čtyřpolní kontingenční tabulky řádkově podmíněných relativních četností:

	sex	žije	zemřel/a	Řádk. součty
Četnost	Ž	6	10	16
Řádk. četn.		37,50%	62,50%	
Četnost	M	15	50	65
Řádk. četn.		23,08%	76,92%	
Četnost	Vš.skup.	21	60	81

Nominální znaky: obecně nabývají aspoň dvou variant, které nelze uspořádat.

Např. *histologický typ tumoru* – lymfom, adenom, karcinom, melanom apod.,

grading tumoru - mikroskopické určení stupně diferencovanosti (vyzrálosti)

nádoru. Má varianty:

Gx - nelze stanovit stupeň diferenciacce

G1 - dobře diferencovaný nádor

G2 - středně diferencovaný nádor

G3 - málo diferencovaný nádor

G4 - nediferencovaný nádor

Čím je nádor méně diferencovaný, tím je agresivnější, ale zároveň citlivější k léčbě.

Staging tumoru - určení rozsahu nádoru. Ke stagingu se používá celá řada systémů. Nejpoužívanější je **system TMN**.

T (tumor; označuje velikost nádoru)

T_x - velikost nelze stanovit

T₀ - není

T₁

T₂

T₃

T₄ - prorůstá do okolních tkání

T_{is} - carcinoma in situ (počáteční stadium, nádorové buňky nepřestoupily přes bazální membránu)

N (nodus; říká, zda-li jsou postiženy regionální lymfatické uzliny)

N_x - nelze stanovit

N₀ - regionální lymfatické uzliny nejsou postiženy

N₁

N₂

N₃

M (metastázy; říká, zda-li došlo k založení vzdálených metastáz)

M_x - nelze stanovit

M₀ - metastázy nejsou přítomny

M₁ - metastázy jsou přítomny

Klinické stadium nádoru

Stadium 0: T_{is} N0 M0

Stadium IA: T1 N0 M0

Stadium IB: T1 N1 M0 nebo T2a/b N0 M0

Stadium II: T1 N2 M0 nebo T2a/b N1 M0 nebo T3 N0 M0

Stadium IIIA: T2a/b N2 M0 nebo T3 N1 M0 nebo T4 N0 M0

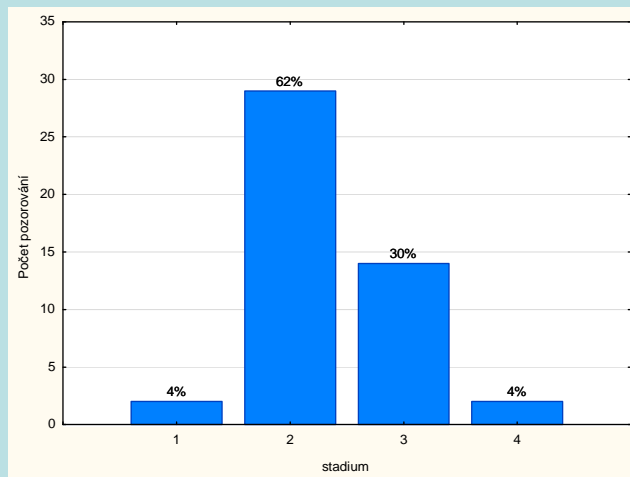
Stadium IIIB: T3 N2 M0

Stadium IV: T4 N1, N2, N3 M0 nebo T1, T2, T3 N3 M0 nebo jakékoliv T
jakékoliv N M1

Nominální znaky se popisují **tabulkou rozložení absolutních a relativních četností variant znaku**, dva nominální znaky pak **kontingenční tabulkou** simultánních absolutních nebo relativních četností a kontingenční tabulkou řádkově nebo sloupcově podmíněných relativních četností dvojic variant znaků.

Síla závislosti mezi dvěma nominálními znaky se měří pomocí **Cramérova koeficientu**. Má hodnoty mezi 0 a 1, čím silnější závislost, tím vyšší hodnoty.

Ukázka sloupkového diagramu stadia rakoviny žaludku:



Ordinální znaky: obecně nabývají aspoň dvou variant, které lze uspořádat podle velikosti.

Např. *výsledky léčby* – má varianty:

kompletní remise (CR)

stabilizované onemocnění (SD)

částečná remise (PR)

progredující onemocnění (PD)

Toxicita léčby – má varianty $G_0 < G_1 < G_2 < \dots$

Ordinální znaky se popisují **tabulkou rozložení absolutních** a relativních četností a číselnými charakteristikami, k nimž patří **kvantily** či **kvantilové rozpětí**. Síla pořadové závislosti mezi dvěma ordinálními znaky se měří pomocí **Spearmanova koeficientu pořadové korelace**. Nabývá hodnot mezi -1 a 1. Hodnoty blízké -1 svědčí o silné nepřímé pořadové závislosti, hodnoty blízké 1 pak o silné přímé pořadové závislosti.

Intervalové znaky: obecně nabývají počtu variant, který je blízký rozsahu datového souboru. Varianty se dají uspořádat podle velikosti a dají se odečítat. Nula na měřicí stupnici je stanovena konvencí.

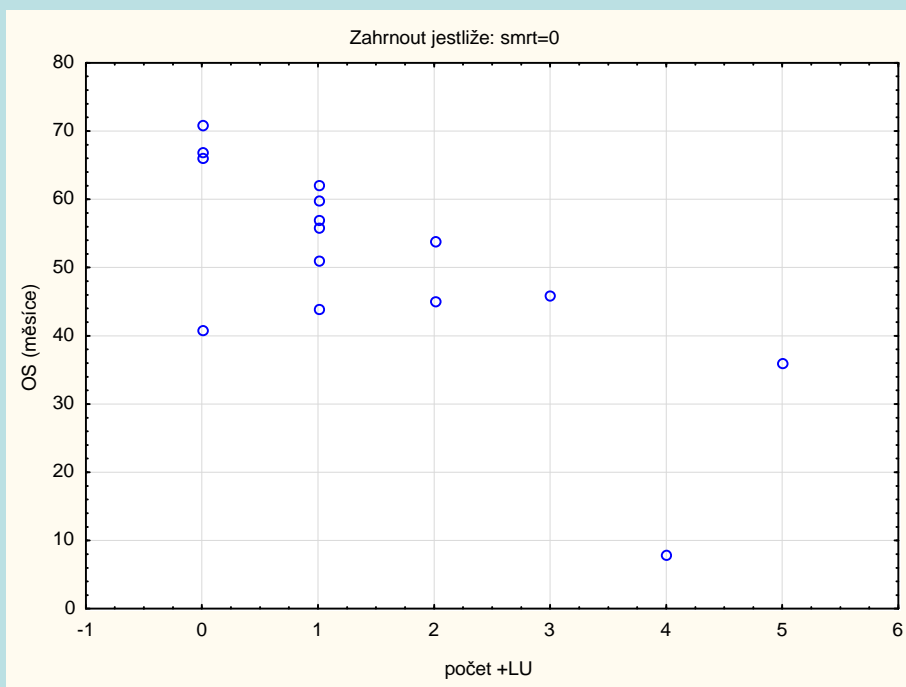
Např. *datum operace*, *datum návratu onemocnění*, *datum úmrtí* apod.

Poměrové znaky: obecně nabývají počtu variant, který je blízký rozsahu datového souboru. Varianty se dají uspořádat podle velikosti a dají se nejen odečítat, ale i dělit.

Např. *věk pacienta v době stanovení diagnózy*, *doba sledování* (tj. doba od vstupu do studie do jejího ukončení resp. do smrti pacienta či jeho ztrátu ze sledování), *dávka radioterapie* apod.

Intervalové a poměrové znaky se popisují **tabulkou intervalového rozložení absolutních a relativních četností intervalů hodnot znaku**. K číselným charakteristikám patří **aritmetický průměr, směrodatná odchylka, minimum, maximum, rozpětí, šikmost, špičatost**. U poměrových znaků lze použít i **koeficient variace** (tj. poměr směrodatné odchylky a průměru). Síla lineární závislosti dvou intervalových či poměrových znaků se měří pomocí **Pearsonova koeficientu korelace**. Nabývá hodnot mezi -1 a 1. Hodnoty blízké -1 svědčí o silné nepřímé lineární závislosti, hodnoty blízké 1 pak o silné přímé lineární závislosti.

Ukázka dvourozměrného tečkového diagramu zachycujícího závislost doby sledování žijících pacientů trpících rakovinou žaludku na počtu odstraněných pozitivních uzlin:



Koeficient korelace = -0,73