

ANALÝZA DAT V R

1. JAK SPRÁVNĚ PŘIPRAVIT A NAČÍST DATA

Mgr. Markéta Pavlíková

Katedra pravděpodobnosti a matematické statistiky MFF UK

www.biostatisticka.cz

JAK VZNIKÁ VĚDECKÁ STUDIE

1. Pozorování
2. Vytvoření hypotéz – úkol a účel studie
3. Vytvoření designu – jak měřit
 - koho porovnávat
 - jakým způsobem
4. Sběr dat
5. Analýza
6. Interpretace a srovnání
7. Publikace výsledků a diskuse
8. Návrh nového experimentu



JAK VZNIKÁ VĚDECKÁ STUDIE

1. Pozorování
2. Vytvoření hypotéz – úkol a účel studie
3. Vytvoření designu – jak měřit
 - koho porovnávat
 - jakým způsobem
4. Sběr dat
5. Analýza
6. Interpretace a srovnání
7. Publikace výsledků a diskuse
8. Návrh nového experimentu



JAK VZNIKÁ VĚDECKÁ STUDIE

1. Pozorování
2. Vytvoření hypotéz – úkol a účel studie
3. Vytvoření designu – jak měřit
 - koho porovnávat
 - jakým způsobem
4. Sběr dat
5. Analýza
6. Interpretace a srovnání
7. Publikace výsledků a diskuse
8. Návrh nového experimentu



KDY PŘEMÝŠLET O DATECH

Už ve chvíli, kdy o studii teprve uvažujeme

- **Co** budeme sbírat?
- Co s tím pak budeme chtít **dělat**? K čemu budou výsledky užitečné?
- **Jak** budeme sbírat?
- Na **kom** to budeme měřit / sledovat / zjišťovat?
- **Kolik** toho bude? Bude to stačit?
- Jak to pak budeme **analyzovat**?

DESIGN STUDIE

Jak a na **kom** budeme měřit?

- popisujeme jen stav nebo porovnáváme intervence?
 - vzorek pacientů z ambulance / hospitalizace
 - volba **zda** a **jaká** kontrolní skupina: zdraví lidé / jiná terapie
 - koho náš vzorek reprezentuje? zobecnitelné?
- jdeme do minulosti?
 - case-control studie, průřezová studie
 - retrospektivně z dokumentace? registr?
- jdeme do budoucnosti? (observace, randomizovaný experiment)

DESIGN STUDIE

- děláme průzkum?
 - je vzorek respondentů reprezentativní?
 - oslovení pacientů v ambulanci
 - výzva na internetu: riziko samovýběru extrémů

Kolik toho budeme měřit?

- cílový počet respondentů / vyšetřených
 - ideálně stanoveny z výpočtů síly studie (power study)
 - uskutečnitelné? → realistický odhad

DESIGN STUDIE

Co budeme měřit?

- výskyt jevu (ano/ne – např. výskyt AI onemocnění)
- odběr a hodnoty z laboratoře (např. koncentrace kappa-lambda FLC)
- popis stavu pomocí škály nebo skóre (MMSE pro Alzheimer. nemoc)
- cílený test (např. TUG test)
- doba od do nějaké události (např. od operace do úmrtí/propuštění)
- dotazníkové šetření (např. kvalita života)

minimalizace zátěže pacienta při maximalizaci informace

Vždy evidujeme i relevantní demografické / anamnestické proměnné

SBĚR DAT

Kvalitně sebraná a připravená data

- umožňují včasnou **detekci problémů**
- významně **urychlují** analýzu
- **omezují riziko** chyby při analýze
- lze lépe interpretovat, činit závěry s větší jistotou
- lze snadno **doplňovat** pro budoucí použití, lze se k nim vracet s minimem nároků na čas
- porozumí jim i někdo jiný a může vaši práci převzít

SBĚR DAT

Jak budeme data **uchovávat**?

- příprava datového formuláře
 - Access, Excel, Calc
 - jasný identifikátor jedince, posléze anonymizovaný, ale přiřazení zaheslované uložit pro budoucí referenci
 - proměnná s kategoriemi: definice, zařazení do skupin, názvy skupin
 - proměnná s hodnotami: možné limity hodnot, jednotky
- přehlednost, popis na zvláštním listě
 - databázi pak může použít i někdo jiný než já sám
- při opisu z karet zapsat vše, co mohu
- nezahazovat primární informaci (BMI: nechat i výšku i hmotnost)

TYPY PROMĚNNÝCH

Základní dva typy proměnných (veličin)

- **kategoriální** proměnná
 - omezené množství hodnot
 - pohlaví, NYHA klasifikace, typ DM, genotyp, ...
 - podtyp binární proměnná (ano/ne)
 - přítomnost onemocnění, pozitivita testu, ...
- **spojitá** proměnná
 - spojité množství hodnot
 - BMI, glykemie, pH, věk, ...

KATEGORIÁLNÍ PROMĚNNÁ

Omezené množství hodnot

- znak přítomen / nepřítomen
- výběr z rovnocenných (neseřaditelných) možností: umístění nádoru, druh terapie, ...
- hodnoty lze seřadit: úroveň vzdělání, stupeň závažnosti onemocnění, ...

Povaha kategorií

- “přirozené” (žena/muž, barva, diagnóza, ...)
- umělé: vznikají pojmenováním nějakých intervalů spojité proměnné
 - podváha / normální váha / nadváha / obezita / morbidní obezita – ve skutečnosti je to BMI rozsekané na kousky

Umělé kategorie znamenají ztrátu informace obsaženou ve spojité proměnné

Lepší zaznamenat tu, pokud to jde, kategorizace se dá vytvořit vždycky

KATEGORIÁLNÍ PROMĚNNÁ

Na co si dát pozor, když zapisuji kategoriální proměnnou

- jasně definovaná množina hodnot, kterých faktor nabývá
- pokud slovně, dávám pozor na **překlepy, mezery**
- pokud číselně, uchovávám si bokem **správné přiřazení** číselných kódů a názvů
- ideálně si v databázi **nastavím povolené hodnoty**, když chci napsat něco jiného, zařve
- pokud vytvořena uměle, uchovávám i původní proměnnou

SPOJITÁ PROMĚNNÁ

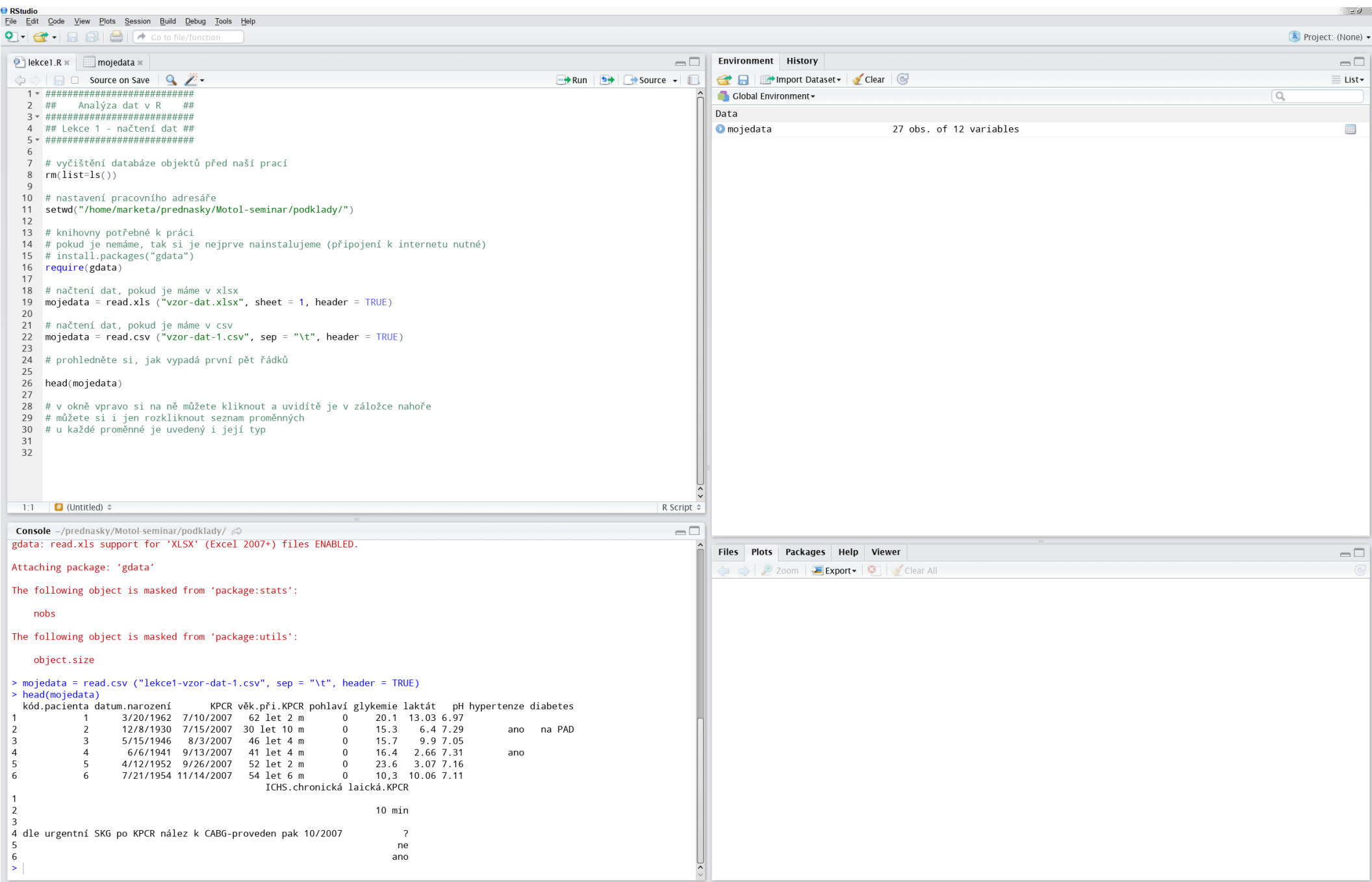
- má zpravidla **jednotky**
- má **logická rozmezí**
 - pH krve = 4 je spíše chyba než reálná hodnota
 - BMI = 10 nebo BMI = 100
- tyto informace v **seznamu proměnných** poslouží skvěle při kontrole dat a analýze
- jednotky nepíšu do dat, ale bokem
- pozor na desetinnou tečku / čárku

UKÁZKA DAT – CO JE ŠPATNĚ?

- ukážeme si data
- společně budeme identifikovat, co je špatně
- ukážeme si, co to udělá při natažení do R
- ukážeme si, jak to napravit

ZAČÁTEK PRÁCE S R STUDIEM

- vytvořte si někde pracovní adresář
- <http://biostatisticka.cz/vyuka>
- do adresáře si stáhněte **lekce1.R** a data
- kliknutím nebo otevřením v Rstudiu spustíte
- kurzor na řádku + **Ctrl+Enter** spustí řádek
- blok označený myší + **Ctrl+Enter** spustí blok



```
1 #####
2 ##    Analýza dat v R    ##
3 #####
4 ## Lekce 1 - načtení dat ##
5 #####
6
7 # vyčištění databáze objektů před naší prací
8 rm(list=ls())
9
10 # nastavení pracovního adresáře
11 # toto si musíte nastavit sami podle sebe, podle toho, kam jste si soubory uložili
12 setwd("/home/marketa/prednasky/Motol-seminar/podklady/")
13
14 # knihovny potřebné k práci
15 # pokud je nemáme, tak si je nejprve nainstalujeme (připojení k internetu nutné)
16 # install.packages("gdata")
17 require(gdata)
18
19 # pokud to nejde, pracujte se souborem .csv
20
21 # načtení dat, pokud je máme v xls
22 mojedata = read.xls ("lekce1-vzor-dat.xlsx", sheet = 1, header = TRUE)
23
24 # načtení dat, pokud je máme v csv
25 mojedata = read.csv ("lekce1-vzor-dat-1.csv", sep = "\t", header = TRUE)
26
27 # prohledněte si, jak vypadá první pět řádků
28
29 head(mojedata)
30
31 # v okně vpravo si na ně můžete kliknout a uvidíte je v záložce nahoře
32 # můžete si i jen rozkliknout seznam proměnných
33 # u každé proměnné je uvedený i její typ
```

RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function

lekce1.R xmojedata x

27 observations of 12 variables

	kód.pacienta	datum.narozeni	KPCR	věk.při.KPCR.	pohlavi	glykemie	X.laktát	pH	hypertenze	diabetes
1	1	3/20/1962	7/10/2007	62 let 2 m	0	20.1	13.03	6.97		
2	2	12/8/1930	7/15/2007	30 let 10 m	0	15.3	6.4	7.29	ano	na PAD
3	3	5/15/1946	8/3/2007	46 let 4 m	0	15.7	9.9	7.05		
4	4	6/6/1941	9/13/2007	41 let 4 m	0	16.4	2.66	7.31	ano	
5	5	4/12/1952	9/26/2007	52 let 2 m	0	23.6	3.07	7.16		
6	6	7/21/1954	11/14/2007	54 let 6 m	0	10,3	10.06	7.11		
7	7	10/18/1955	2/3/2008	55 let 8 m	0	12,8	2.15	7.33		
8	8	11/24/1939	3/29/2008	39 let 10 m	0	12.9	nelze	7.36	ano	
9	9	10/30/1935	6/7/2008	35 let 9 m	1	9.2	7.54	6.96		DM II na P/
10	10	1/30/1926	6/24/2008	26 let 0 m	0			NA		DM II na d:
11	11	10/29/1950	9/2/2008	50 let 9 m	0	12.5		NA		
12	12	7/25/1952	2/9/2009	52 let 6 m	1			NA	ano	
13	13	6/7/1966	5/29/2009	66 let 4 m	0			NA		
14	14	10/22/1932	9/4/2009	32 let 8 m	0	8.3		NA	ano	DM 2 na PAD
15	15	3/16/1955	9/19/2009	55 let 1 m	0	15.1	5.78	7.23	ano	
16	16	12/18/1953	10/13/2009	53 let 10 m	0	10.7	0.83	7.32	ano	
17	17	12/4/1964	10/30/2009	64 let 10 m	0	23.9		NA		
18	18	7/23/1943	12/29/2009	43 let 5 m	0	30.6	13.9	4.86		DM II na P/
19	19	12/18/1947	2/4/2010	47 let 10 m	0	17.8		NA	ano	
20	20	9/19/1943	2/15/2010	43 let 7 m	0		3.09	7.13	ano	DM II na P/
21	21	11/23/1975	5/16/2011	75 let 9 m	1	nelze zjistit čas zástavy		NA	gestační	DM 1. typu
22	22	4/18/1943	6/1/2010	43 let 2 m	0	25.3	14.8	7.05	ano	DM II na P/
23	23	12/16/1931	7/30/2010	31 let 10 m	1			NA	ano	
24	24	5/18/1943	8/7/2010	43 let 3 m	0			NA	ano	
25	25	10/15/1938	11/7/2010	38 let 8 m	0			NA	ano	DM II na P/
26	26	11/30/1937	12/21/2010	37 let 9 m	0	22.4	10.5	7.09		DM II na ir
27	27	7/20/1955	12/23/2010	55 let 5 m	1	6.7	1.1	7.51		

EnvironmentHistory

Global Environment

Data

mojedata27 obs. of 12 variables

Console~/prednasky/Motol-seminar/podklady/

gdata: read.xls support for 'XLSX' (Excel 2007+) files ENABLED.

Attaching package: 'gdata'

The following object is masked from 'package:stats':

nobs

The following object is masked from 'package:utils':

object.size

> mojedata = read.csv ("lekce1-vzor-dat-1.csv", sep = "\t", header = TRUE)

> head(mojedata)

kód.pacienta datum.narozeni KPCR věk.při.KPCR pohlaví glykemie laktát pH hypertenze diabetes

1 1 3/20/1962 7/10/2007 62 let 2 m 0 20.1 13.03 6.97

2 2 12/8/1930 7/15/2007 30 let 10 m 0 15.3 6.4 7.29

3 3 5/15/1946 8/3/2007 46 let 4 m 0 15.7 9.9 7.05

4 4 6/6/1941 9/13/2007 41 let 4 m 0 16.4 2.66 7.31

5 5 4/12/1952 9/26/2007 52 let 2 m 0 23.6 3.07 7.16

6 6 7/21/1954 11/14/2007 54 let 6 m 0 10,3 10.06 7.11

ICHS.chronická laická.KPCR

1

2 10 min

3

4 dle urgentní SKG po KPCR nález k CABG-proveden pak 10/2007 ?

5 ne

6 ano

> |

UKÁZKA DAT – CO JE ŠPATNĚ?

Environment

History

Import Dataset

Clear

List

Global Environment

Data

mojedata

27 obs. of 12 variables

kód.pacienta : int 1 2 3 4 5 6 7 8 9 10 ...

datum.narození: Factor w/ 27 levels "1/30/1926","10/15/1938",...: 16 14 19 21 17 24 3 8 6 1 ...

KPCR : Factor w/ 27 levels "10/13/2009","10/30/2009",...: 18 19 21 23 26 3 9 12 17 16 ...

věk.při.KPCR. : Factor w/ 27 levels "26 let 0 m","30 let 10 m",...: 24 2 14 9 17 20 23 8 5 1 ...

pohlaví : int 0 0 0 0 0 0 0 0 1 0 ...

glykemie : Factor w/ 21 levels "", "10,3", "10.7",...: 12 8 9 10 14 2 4 6 20 1 ...

X.laktát : Factor w/ 17 levels "", "0.83", "1.1",...: 6 14 16 10 11 4 9 17 15 1 ...

pH : num 6.97 7.29 7.05 7.31 7.16 7.11 7.33 7.36 6.96 NA ...

hypertenze : Factor w/ 3 levels "", "ano", "gestační": 1 2 1 2 1 1 1 2 1 1 ...

diabetes : Factor w/ 7 levels "", "DM 1. typu na insulinu",...: 1 7 1 1 1 1 1 1 6 4 ...

ICHS.chronická: Factor w/ 10 levels "", "3 -VD, ad CABG elektivně",...: 1 1 1 6 1 1 1 1 1 10 ...

laická.KPCR : Factor w/ 12 levels "", "?", "10 min",...: 1 3 1 2 11 5 6 1 1 1 ...

UKÁZKA DAT – CO JE ŠPATNĚ?

Environment

History

Import Dataset

Clear

List

Global Environment

Data

mojedata

27 obs. of 12 variables

kód.pacienta : int 1 2 3 4 5 6 7 8 9 10 ...

datum.narození: Factor w/ 27 levels "1/30/1926","10/15/1938",...: 16 14 19 21 17 24 3 8 6 1 ...

KPCR : Factor w/ 27 levels "10/13/2009","10/30/2009",...: 18 19 21 23 26 3 9 12 17 16 ...

věk.při.KPCR. : Factor w/ 27 levels "26 let 0 m","30 let 10 m",...: 24 2 14 9 17 20 23 8 5 1 ...

pohlaví : int 0 0 0 0 0 0 0 0 1 0 ...

glykemie : Factor w/ 21 levels "", "10.3", "10.7",...: 12 8 9 10 14 2 4 6 20 1 ...

X.laktát : Factor w/ 17 levels "", "0.83", "1.1",...: 6 14 16 10 11 4 9 17 15 1 ...

pH : num 6.97 7.29 7.05 7.31 7.16 7.11 7.33 7.36 6.96 NA ...

hypertenze : Factor w/ 3 levels "", "ano", "gestační": 1 2 1 2 1 1 1 2 1 1 ...

diabetes : Factor w/ 7 levels "", "DM 1. typu na insulinu",...: 1 7 1 1 1 1 1 1 6 4 ...

ICHS.chronická: Factor w/ 10 levels "", "3 -VD, ad CABG elektivně",...: 1 1 1 6 1 1 1 1 1 10 ...

laická.KPCR : Factor w/ 12 levels "", "?", "10 min",...: 1 3 1 2 11 5 6 1 1 1 ...

- že je diabetes, hypertenze apod faktor (tj. kategoriální) je OK
- že je věk, glykemie atd. faktor je špatně
- co se stalo?

UKÁZKA DAT – CO JE ŠPATNĚ?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	kód pacienta	datum narození	KPCR	věk při KPCR	pohlaví	glykemie	laktát	pH	hypertenze	diabetes	ICHS chronická	laická KPCR		
2	1	3/20/1962	7/10/2007	62 let 2 m	0	20.1	13.03	6.97						
3	2	12/8/1930	7/15/2007	30 let 10 m	0	15.3	6.4	7.29	ano	na PAD		10 min		
4	3	5/15/1946	8/3/2007	46 let 4 m	0	15.7	9.9	7.05						
5	4	6/6/1941	9/13/2007	41 let 4 m	0	16.4	2.66	7.31	ano		dle urgentní SKG po KPCR nález k CABG- t	?		
6	5	4/12/1952	9/26/2007	52 let 2 m	0	23.6	3.07	7.16				ne		
7	6	7/21/1954	11/14/2007	54 let 6 m	0	10,3	10.06	7.11				ano		
8	7	10/18/1955	2/3/2008	55 let 8 m	0	12,8	2.15	7.33				ano, 15 min		
9	8	11/24/1939	3/29/2008	39 let 10 m	0	12.9	nelze	7.36	ano					
10	9	10/30/1935	6/7/2008	35 let 9 m	1	9.2	7.54	6.96		DM II na PAD				
11	10	1/30/1926	6/24/2008	26 let 0 m	0					DM II na dietě	sy AP			
12	11	10/29/1950	9/2/2008	50 let 9 m	0	12.5					3 -VD, ad CABG elektivně	laická 6 min.		
13	12	7/25/1952	2/9/2009	52 let 6 m	1				ano					
14	13	6/7/1966	5/29/2009	66 let 4 m	0									
15	14	10/22/1932	9/4/2009	32 let 8 m	0	8.3			ano	DM 2 na PAD	3- VD ad CABG časné překladem			
16	15	3/16/1955	9/19/2009	55 let 1 m	0	15.1	5.78	7.23	ano			ano, čas neuveden		
17	16	12/18/1953	10/13/2009	53 let 10 m	0	10.7	0.83	7.32	ano		3-VD ad CABG časné překladem	ano, telefonicky as.		
18	17	12/4/1964	10/30/2009	64 let 10 m	0	23.9						ano		
19	18	7/23/1943	12/29/2009	43 let 5 m	0	30.6	13.9	4.86		DM II na PAD	st.p. IM bez podrobností	ne		
20	19	12/18/1947	2/4/2010	47 let 10 m	0	17.8			ano			15-20 min. pouze masáž, poté ALS 20-25 min.		
21	20	9/19/1943	2/15/2010	43 let 7 m	0		3.09	7.13	ano	DM II na PAD	st.p. IM 1997, difuzní postižení tepen			
22	21	11/23/1975	5/16/2011	75 let 9 m	1	nelze zjistit čas zástavy			gestační	DM 1. typu na inzulinu		sporně		
23	22	4/18/1943	6/1/2010	43 let 2 m	0	25.3	14.8	7.05	ano	DM II na PAD				
24	23	12/16/1931	7/30/2010	31 let 10 m	1				ano		difuzní postižení tepen ke konzerv. Postupu	laická 8 min.		
25	24	5/18/1943	8/7/2010	43 let 3 m	0				ano					
26	25	10/15/1938	11/7/2010	38 let 8 m	0				ano	DM II na PAD				
27	26	11/30/1937	12/21/2010	37 let 9 m	0	22.4	10.5	7.09		DM II na inzulinu	st.p. QIM spodní stěny 5/06, 3-VD, st.p opakovaných PCI			
28	27	7/20/1955	12/23/2010	55 let 5 m	1	6.7	1.1	7.51						
29														

pozn.: data v Excelu

UKÁZKA DAT – CO JE ŠPATNĚ?

[illegible]

UKÁZKA DAT – CO JE ŠPATNĚ?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	kód pacienta	datum narození	KPCR	věk při KPCR	pohlaví	glykemie	laktát	pH	hypertenze	diabetes	ICHS chronická	laická KPCR		
2	1	3/20/1962	7/10/2007	62 let 2 m	0	20.1	13.03	6.97						
3	2	12/8/1930	7/15/2007	30 let 10 m	0	15.3	6.4	7.29	ano	na PAD		10 min		
4	3	5/15/1946	8/3/2007	46 let 4 m	0	15.7	9.9	7.05						
5	4	6/6/1941	9/13/2007	41 let 4 m	0	16.4	2.66	7.31	ano		dle urgentní SKG po KPCR nález k CABG- t	?		
6	5	4/12/1952	9/26/2007	52 let 2 m	0	23.6	3.07	7.16				ne		
7	6	7/21/1954	11/14/2007	54 let 6 m	0	10,3	10.06	7.11				ano		
8	7	10/18/1955	2/3/2008	55 let 8 m	0	12,8	2.15	7.33				ano, 15 min		
9	8	11/24/1939	3/29/2008	39 let 10 m	0	12.9	nelze	7.36	ano					
10	9	10/30/1935	6/7/2008	35 let 9 m	1	9.2	7.54	6.96		DM II na PAD				
11	10	1/30/1926	6/24/2008	26 let 0 m	0					DM II na dietě	sy AP			
12	11	10/29/1950	9/2/2008	50 let 9 m	0	12.5					3 -VD, ad CABG elektivně	laická 6 min.		
13	12	7/25/1952	2/9/2009	52 let 6 m	1				ano					
14	13	6/7/1966	5/29/2009	66 let 4 m	0									
15	14	10/22/1932	9/4/2009	32 let 8 m	0	8.3			ano	DM 2 na PAD	3- VD ad CABG časné překladem			
16	15	3/16/1955	9/19/2009	55 let 1 m	0	15.1	5.78	7.23	ano			ano, čas neuveden		
17	16	12/18/1953	10/13/2009	53 let 10 m	0	10.7	0.83	7.32	ano		3-VD ad CABG časné překladem	ano, telefonicky as.		
18	17	12/4/1964	10/30/2009	64 let 10 m	0	23.9						ano		
19	18	7/23/1943	12/29/2009	43 let 5 m	0	30.6	13.9	4.86		DM II na PAD	st.p. IM bez podrobností	ne		
20	19	12/18/1947	2/4/2010	47 let 10 m	0	17.8			ano			15-20 min. pouze masáž, poté ALS 20-25 min.		
21	20	9/19/1943	2/15/2010	43 let 7 m	0		3.09	7.13	ano	DM II na PAD	st.p. IM 1997, difuzní postižení tepen			
22	21	11/23/1975	5/16/2011	75 let 9 m	1	nelze zjistit čas zástavy			gestační	DM 1. typu na inzulínu		sporně		
23	22	4/18/1943	6/1/2010	43 let 2 m	0	25.3	14.8	7.05	ano	DM II na PAD				
24	23	12/16/1931	7/30/2010	31 let 10 m	1				ano		difuzní postižení tepen ke konzerv. Postupu	laická 8 min.		
25	24	5/18/1943	8/7/2010	43 let 3 m	0				ano					
26	25	10/15/1938	11/7/2010	38 let 8 m	0				ano	DM II na PAD				
27	26	11/30/1937	12/21/2010	37 let 9 m	0	22.4	10.5	7.09		DM II na inzulínu	st.p. QIM spodní stěny 5/06, 3-VD, st.p opakovaných PCI			
28	27	7/20/1955	12/23/2010	55 let 5 m	1	6.7	1.1	7.51						
29														

věk není číslo ale text! Hezké, ale nepoužitelné.

UKÁZKA DAT – CO JE ŠPATNĚ?

[illegible]

UKÁZKA DAT – CO JE ŠPATNĚ?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	kód pacienta	datum narození	KPCR	věk při KPCR	pohlaví	glykemie	laktát	pH	hypertenze	diabetes	ICHS chronická	laická KPCR		
2	1	3/20/1962	7/10/2007	62 let 2 m	0	20.1	13.03	6.97						
3	2	12/8/1930	7/15/2007	30 let 10 m	0	15.3	6.4	7.29	ano	na PAD		10 min		
4	3	5/15/1946	8/3/2007	46 let 4 m	0	15.7	9.9	7.05						
5	4	6/6/1941	9/13/2007	41 let 4 m	0	16.4	2.66	7.31	ano		dle urgentní SKG po KPCR nález k CABG- t	?		
6	5	4/12/1952	9/26/2007	52 let 2 m	0	23.6	3.07	7.16				ne		
7	6	7/21/1954	11/14/2007	54 let 6 m	0	10,3	10.06	7.11				ano		
8	7	10/18/1955	2/3/2008	55 let 8 m	0	12,8	2.15	7.33				ano, 15 min		
9	8	11/24/1939	3/29/2008	39 let 10 m	0	12.9	nelze	7.36	ano					
10	9	10/30/1935	6/7/2008	35 let 9 m	1	9.2	7.54	6.96		DM II na PAD				
11	10	1/30/1926	6/24/2008	26 let 0 m	0					DM II na dietě	sy AP			
12	11	10/29/1950	9/2/2008	50 let 9 m	0	12.5					3 -VD, ad CABG elektivně	laická 6 min.		
13	12	7/25/1952	2/9/2009	52 let 6 m	1				ano					
14	13	6/7/1966	5/29/2009	66 let 4 m	0									
15	14	10/22/1932	9/4/2009	32 let 8 m	0	8.3			ano	DM 2 na PAD	3- VD ad CABG časné překladem			
16	15	3/16/1955	9/19/2009	55 let 1 m	0	15.1	5.78	7.23	ano			ano, čas neuveden		
17	16	12/18/1953	10/13/2009	53 let 10 m	0	10.7	0.83	7.32	ano		3-VD ad CABG časné překladem	ano, telefonicky as.		
18	17	12/4/1964	10/30/2009	64 let 10 m	0	23.9						ano		
19	18	7/23/1943	12/29/2009	43 let 5 m	0	30.6	13.9	4.86		DM II na PAD	st.p. IM bez podrobností	ne		
20	19	12/18/1947	2/4/2010	47 let 10 m	0	17.8			ano			15-20 min. pouze masáž, poté ALS 20-25 min.		
21	20	9/19/1943	2/15/2010	43 let 7 m	0		3.09	7.13	ano	DM II na PAD	st.p. IM 1997, difuzní postižení tepen			
22	21	11/23/1975	5/16/2011	75 let 9 m	1	nelze zjistit čas zástavy			gestační	DM 1. typu na inzulinu		sporně		
23	22	4/18/1943	6/1/2010	43 let 2 m	0	25.3	14.8	7.05	ano	DM II na PAD				
24	23	12/16/1931	7/30/2010	31 let 10 m	1				ano		difuzní postižení tepen ke konzerv. Postupu	laická 8 min.		
25	24	5/18/1943	8/7/2010	43 let 3 m	0				ano					
26	25	10/15/1938	11/7/2010	38 let 8 m	0				ano	DM II na PAD				
27	26	11/30/1937	12/21/2010	37 let 9 m	0	22.4	10.5	7.09		DM II na inzulinu	st.p. QIM spodní stěny 5/06, 3-VD, st.p opakovaných PCI			
28	27	7/20/1955	12/23/2010	55 let 5 m	1	6.7	1.1	7.51						
29														

text uprostřed čísel, navíc nenesoucí příliš informace

UKÁZKA DAT – CO JE ŠPATNĚ?

[illegible]

UKÁZKA DAT – CO JE ŠPATNĚ?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	kód pacienta	datum narození	KPCR	věk při KPCR	pohlaví	glykemie	laktát	pH	hypertenze	diabetes	ICHS chronická	laická KPCR		
2	1	3/20/1962	7/10/2007	62 let 2 m	0	20.1	13.03	6.97						
3	2	12/8/1930	7/15/2007	30 let 10 m	0	15.3	6.4	7.29	ano	na PAD		10 min		
4	3	5/15/1946	8/3/2007	46 let 4 m	0	15.7	9.9	7.05						
5	4	6/6/1941	9/13/2007	41 let 4 m	0	16.4	2.66	7.31	ano		dle urgentní SKG po KPCR nález k CABG- t	?		
6	5	4/12/1952	9/26/2007	52 let 2 m	0	23.6	3.07	7.16				ne		
7	6	7/21/1954	11/14/2007	54 let 6 m	0	10,3	10.06	7.11				ano		
8	7	10/18/1955	2/3/2008	55 let 8 m	0	12,8	2.15	7.33				ano, 15 min		
9	8	11/24/1939	3/29/2008	39 let 10 m	0	12.9	nelze	7.36	ano					
10	9	10/30/1935	6/7/2008	35 let 9 m	1	9.2	7.54	6.96		DM II na PAD				
11	10	1/30/1926	6/24/2008	26 let 0 m	0					DM II na dietě	sy AP			
12	11	10/29/1950	9/2/2008	50 let 9 m	0	12.5					3 -VD, ad CABG elektivně	laická 6 min.		
13	12	7/25/1952	2/9/2009	52 let 6 m	1				ano					
14	13	6/7/1966	5/29/2009	66 let 4 m	0									
15	14	10/22/1932	9/4/2009	32 let 8 m	0	8.3			ano	DM 2 na PAD	3- VD ad CABG časné překladem			
16	15	3/16/1955	9/19/2009	55 let 1 m	0	15.1	5.78	7.23	ano			ano, čas neuveden		
17	16	12/18/1953	10/13/2009	53 let 10 m	0	10.7	0.83	7.32	ano		3-VD ad CABG časné překladem	ano, telefonicky as.		
18	17	12/4/1964	10/30/2009	64 let 10 m	0	23.9						ano		
19	18	7/23/1943	12/29/2009	43 let 5 m	0	30.6	13.9	4.86	ne? NA?	DM II na PAD	st.p. IM bez podrobností	ne		
20	19	12/18/1947	2/4/2010	47 let 10 m	0	17.8			ano			15-20 min. pouze masáž, poté ALS 20-25 min.		
21	20	9/19/1943	2/15/2010	43 let 7 m	0		3.09	7.13	ano	DM II na PAD	st.p. IM 1997, difúzní postižení tepen			
22	21	11/23/1975	5/16/2011	75 let 9 m	1	nelze zjistit čas zástavy			gestační	DM 1. typu na inzulínu		sporně		
23	22	4/18/1943	6/1/2010	43 let 2 m	0	25.3	14.8	7.05	ano	DM II na PAD				
24	23	12/16/1931	7/30/2010	31 let 10 m	1				ano		difúzní postižení tepen ke konzerv. Postupu	laická 8 min.		
25	24	5/18/1943	8/7/2010	43 let 3 m	0				ano					
26	25	10/15/1938	11/7/2010	38 let 8 m	0				ano	DM II na PAD				
27	26	11/30/1937	12/21/2010	37 let 9 m	0	22.4	10.5	7.09		DM II na inzulínu	st.p. QIM spodní stěny 5/06, 3-VD, st.p opakovaných PCI			
28	27	7/20/1955	12/23/2010	55 let 5 m	1	6.7	1.1	7.51						
29														

prázdné pole: je to NE nebo chybějící pozorování?

UKÁZKA DAT – CO JE ŠPATNĚ?

[illegible]

UKÁZKA DAT – CO JE ŠPATNĚ?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	kód pacienta	datum narození	KPCR	věk při KPCR	pohlaví	glykemie	laktát	pH	hypertenze	diabetes	ICHS chronická	laická KPCR		
2	1	3/20/1962	7/10/2007	62 let 2 m	0	20.1	13.03	6.97						
3	2	12/8/1930	7/15/2007	30 let 10 m	0	15.3	6.4	7.29	ano	na PAD		10 min		
4	3	5/15/1946	8/3/2007	46 let 4 m	0	15.7	9.9	7.05						
5	4	6/6/1941	9/13/2007	41 let 4 m	0	16.4	2.66	7.31	ano		dle urgentní SKG po KPCR nález k CABG- t	?		
6	5	4/12/1952	9/26/2007	52 let 2 m	0	23.6	3.07	7.16				ne		
7	6	7/21/1954	11/14/2007	54 let 6 m	0	10,3	10.06	7.11				ano		
8	7	10/18/1955	2/3/2008	55 let 8 m	0	12,8	2.15	7.33				ano, 15 min		
9	8	11/24/1939	3/29/2008	39 let 10 m	0	12.9	nelze	7.36	ano					
10	9	10/30/1935	6/7/2008	35 let 9 m	1	9.2	7.54	6.96		DM II na PAD				
11	10	1/30/1926	6/24/2008	26 let 0 m	0					DM II na dietě	sy AP			
12	11	10/29/1950	9/2/2008	50 let 9 m	0	12.5					3 -VD, ad CABG elektivně	laická 6 min.		
13	12	7/25/1952	2/9/2009	52 let 6 m	1				ano					
14	13	6/7/1966	5/29/2009	66 let 4 m	0									
15	14	10/22/1932	9/4/2009	32 let 8 m	0	8.3			ano	DM 2 na PAD	3- VD ad CABG časné překladem			
16	15	3/16/1955	9/19/2009	55 let 1 m	0	15.1	5.78	7.23	ano			ano, čas neuveden		
17	16	12/18/1953	10/13/2009	53 let 10 m	0	10.7	0.83	7.32	ano		3-VD ad CABG časné překladem	ano, telefonicky as.		
18	17	12/4/1964	10/30/2009	64 let 10 m	0	23.9						ano		
19	18	7/23/1943	12/29/2009	43 let 5 m	0	30.6	13.9	4.86		DM II na PAD	st.p. IM bez podrobností	ne		
20	19	12/18/1947	2/4/2010	47 let 10 m	0	17.8			ano			15-20 min. pouze masáž, poté ALS 20-25 min.		
21	20	9/19/1943	2/15/2010	43 let 7 m	0		3.09	7.13	ano	DM II na PAD	st.p. IM 1997, difuzní postižení tepen			
22	21	11/23/1975	5/16/2011	75 let 9 m	1	nelze zjistit čas zástavy			gestační	DM 1. typu na inzulinu		sporně		
23	22	4/18/1943	6/1/2010	43 let 2 m	0	25.3	14.8	7.05	ano	DM II na PAD				
24	23	12/16/1931	7/30/2010	31 let 10 m	1				ano		difuzní postižení tepen ke konzerv. Postupu	laická 8 min.		
25	24	5/18/1943	8/7/2010	43 let 3 m	0				ano					
26	25	10/15/1938	11/7/2010	38 let 8 m	0				ano	DM II na PAD				
27	26	11/30/1937	12/21/2010	37 let 9 m	0	22.4	10.5	7.09		DM II na inzulinu	st.p. QIM spodní stěny 5/06, 3-VD, st.p opakovaných PCI			
28	27	7/20/1955	12/23/2010	55 let 5 m	1	6.7	1.1	7.51						
29														

desetinná čárka a desetinná tečka zároveň

UKÁZKA DAT – CO JE ŠPATNĚ?

[illegible]

UKÁZKA DAT – CO JE ŠPATNĚ?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	kód pacienta	datum narození	KPCR	věk při KPCR	pohlaví	glykemie	laktát	pH	hypertenze	diabetes	ICHS chronická	laická KPCR		
2	1	3/20/1962	7/10/2007	62 let 2 m	0	20.1	13.03	6.97						
3	2	12/8/1930	7/15/2007	30 let 10 m	0	15.3	6.4	7.29	ano	na PAD		10 min		
4	3	5/15/1946	8/3/2007	46 let 4 m	0	15.7	9.9	7.05						
5	4	6/6/1941	9/13/2007	41 let 4 m	0	16.4	2.66	7.31	ano		dle urgentní SKG po KPCR nález k CABG- t	?		
6	5	4/12/1952	9/26/2007	52 let 2 m	0	23.6	3.07	7.16				ne		
7	6	7/21/1954	11/14/2007	54 let 6 m	0	10,3	10.06	7.11				ano		
8	7	10/18/1955	2/3/2008	55 let 8 m	0	12,8	2.15	7.33				ano, 15 min		
9	8	11/24/1939	3/29/2008	39 let 10 m	0	12.9	nelze	7.36	ano					
10	9	10/30/1935	6/7/2008	35 let 9 m	1	9.2	7.54	6.96		DM II na PAD				
11	10	1/30/1926	6/24/2008	26 let 0 m	0					DM II na dietě	sy AP			
12	11	10/29/1950	9/2/2008	50 let 9 m	0	12.5					3 -VD, ad CABG elektivně	laická 6 min.		
13	12	7/25/1952	2/9/2009	52 let 6 m	1				ano					
14	13	6/7/1966	5/29/2009	66 let 4 m	0									
15	14	10/22/1932	9/4/2009	32 let 8 m	0	8.3			ano	DM 2 na PAD	3- VD ad CABG časné překladem			
16	15	3/16/1955	9/19/2009	55 let 1 m	0	15.1	5.78	7.23	ano			ano, čas neuveden		
17	16	12/18/1953	10/13/2009	53 let 10 m	0	10.7	0.83	7.32	ano		3-VD ad CABG časné překladem	ano, telefonicky as.		
18	17	12/4/1964	10/30/2009	64 let 10 m	0	23.9						ano		
19	18	7/23/1943	12/29/2009	43 let 5 m	0	30.6	13.9	4.86		DM II na PAD	st.p. IM bez podrobností	ne		
20	19	12/18/1947	2/4/2010	47 let 10 m	0	17.8			ano			15-20 min. pouze masáž, poté ALS 20-25 min.		
21	20	9/19/1943	2/15/2010	43 let 7 m	0		3.09	7.13	ano	DM II na PAD	st.p. IM 1997, difuzní postižení tepen			
22	21	11/23/1975	5/16/2011	75 let 9 m	1	nelze zjistit čas zástavy			gestační	DM 1. typu na inzulinu		sporně		
23	22	4/18/1943	6/1/2010	43 let 2 m	0	25.3	14.8	7.05	ano	DM II na PAD				
24	23	12/16/1931	7/30/2010	31 let 10 m	1				ano		difuzní postižení tepen ke konzerv. Postupu	laická 8 min.		
25	24	5/18/1943	8/7/2010	43 let 3 m	0				ano					
26	25	10/15/1938	11/7/2010	38 let 8 m	0				ano	DM II na PAD				
27	26	11/30/1937	12/21/2010	37 let 9 m	0	22.4	10.5	7.09		DM II na inzulinu	st.p. QIM spodní stěny 5/06, 3-VD, st.p opakovaných PCI			
28	27	7/20/1955	12/23/2010	55 let 5 m	1	6.7	1.1	7.51						
29														

nesmyslná hodnota – obvykle nenajdeme na první pohled

UKÁZKA DAT – CO JE ŠPATNĚ?

[illegible]

UKÁZKA DAT – CO JE ŠPATNĚ?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	kód pacienta	datum narození	KPCR	věk při KPCR	pohlaví	glykemie	laktát	pH	hypertenze	diabetes	ICHS chronická	laická KPCR		
2	1	3/20/1962	7/10/2007	62 let 2 m	0	20.1	13.03	6.97						
3	2	12/8/1930	7/15/2007	30 let 10 m	0	15.3	6.4	7.29	ano	na PAD		10 min		
4	3	5/15/1946	8/3/2007	46 let 4 m	0	15.7	9.9	7.05						
5	4	6/6/1941	9/13/2007	41 let 4 m	0	16.4	2.66	7.31	ano		dle urgentní SKG po KPCR nález k CABG- t	?		
6	5	4/12/1952	9/26/2007	52 let 2 m	0	23.6	3.07	7.16				ne		
7	6	7/21/1954	11/14/2007	54 let 6 m	0	10,3	10.06	7.11				ano		
8	7	10/18/1955	2/3/2008	55 let 8 m	0	12,8	2.15	7.33				ano, 15 min		
9	8	11/24/1939	3/29/2008	39 let 10 m	0	12.9	nelze	7.36	ano					
10	9	10/30/1935	6/7/2008	35 let 9 m	1	9.2	7.54	6.96		DM II na PAD				
11	10	1/30/1926	6/24/2008	26 let 0 m	0					DM II na dietě	sy AP			
12	11	10/29/1950	9/2/2008	50 let 9 m	0	12.5					3 -VD, ad CABG elektivně	laická 6 min.		
13	12	7/25/1952	2/9/2009	52 let 6 m	1				ano					
14	13	6/7/1966	5/29/2009	66 let 4 m	0									
15	14	10/22/1932	9/4/2009	32 let 8 m	0	8.3			ano	DM 2 na PAD	3- VD ad CABG časné překladem			
16	15	3/16/1955	9/19/2009	55 let 1 m	0	15.1	5.78	7.23	ano			ano, čas neuveden		
17	16	12/18/1953	10/13/2009	53 let 10 m	0	10.7	0.83	7.32	ano		3-VD ad CABG časné překladem	ano, telefonicky as.		
18	17	12/4/1964	10/30/2009	64 let 10 m	0	23.9						ano		
19	18	7/23/1943	12/29/2009	43 let 5 m	0	30.6	13.9	4.86		DM II na PAD	st.p. IM bez podrobností	ne		
20	19	12/18/1947	2/4/2010	47 let 10 m	0	17.8			ano			15-20 min. pouze masáž, poté ALS 20-25 min.		
21	20	9/19/1943	2/15/2010	43 let 7 m	0		3.09	7.13	ano	DM II na PAD	st.p. IM 1997, difuzní postižení tepen			
22	21	11/23/1975	5/16/2011	75 let 9 m	1	nelze zjistit čas zástavy			gestační	DM 1. typu na insulínu		sporně		
23	22	4/18/1943	6/1/2010	43 let 2 m	0	25.3	14.8	7.05	ano	DM II na PAD				
24	23	12/16/1931	7/30/2010	31 let 10 m	1				ano		difusní postižení tepen ke konzerv. Postupu	laická 8 min.		
25	24	5/18/1943	8/7/2010	43 let 3 m	0				ano					
26	25	10/15/1938	11/7/2010	38 let 8 m	0				ano	DM II na PAD				
27	26	11/30/1937	12/21/2010	37 let 9 m	0	22.4	10.5	7.09		DM II na inzulinu	st.p. QIM spodní stěny 5/06, 3-VD, st.p opakovaných PCI			
28	27	7/20/1955	12/23/2010	55 let 5 m	1	6.7	1.1	7.51						
29														

tři informace (přítomnost, typ, terapie) naráz, takto neanalyzovatelné

UKÁZKA DAT – NÁPRAVA

[illegible]

UKÁZKA DAT – NÁPRAVA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	ID	datum.narozeni	KPCR	vek.KPCR	pohlavi	glykemie	laktát	pH	hypertenze	hypertenze.pozn	DM	DM.typ	DM.terapie	ICHS	ICHS.pozn	laická KPCR	
2	1	3/20/1962	7/10/2007	45	0	20.1	13.03	6.97	ne		0			0			
3	2	12/8/1930	7/15/2007	77	0	15.3	6.4	7.29	ano		1	NA	PAD	0		10 min	
4	3	5/15/1946	8/3/2007	61	0	15.7	9.9	7.05	ne		0			0			
5	4	6/6/1941	9/13/2007	66	0	16.4	2.66	7.31	ano		0			1	dle urgentní SKG po KPCR nález k CA	?	
6	5	4/12/1952	9/26/2007	55	0	23.6	3.07	7.16	ne		0			0		ne	
7	6	7/21/1954	11/14/2007	53	0	10.3	10.06	7.11	ne		0			0		ano	
8	7	10/18/1955	2/3/2008	52	0	12.8	2.15	7.33	ne		0			0		ano, 15 min	
9	8	11/24/1939	3/29/2008	68	0	12.9	NA	7.36	ano		0			0			
10	9	10/30/1935	6/7/2008	73	1	9.2	7.54	6.96	ne		1	2	PAD	0			
11	10	1/30/1926	6/24/2008	82	0		NA				1	2	dieta	1	sy AP		
12	11	10/29/1950	9/2/2008	58	0	12.5	NA				0			1	3 -VD, ad CABG elektivně	laická 6 min.	
13	12	7/25/1952	2/9/2009	57	1		NA		ano		0			0			
14	13	6/7/1966	5/29/2009	43	0		NA				NA			NA			
15	14	10/22/1932	9/4/2009	77	0	8.3	NA		ano		1	2	PAD	1	3- VD ad CABG časné překladem		
16	15	3/16/1955	9/19/2009	55	0	15.1	5.78	7.23	ano		0			0		ano, čas neuveden	
17	16	12/18/1953	10/13/2009	56	0	10.7	0.83	7.32	ano		0			1	3-VD ad CABG časné překladem	ano, telefonicky as.	
18	17	12/4/1964	10/30/2009	45	0	23.9	NA		ne		0			0		ano	
19	18	7/23/1943	12/29/2009	66	0	30.6	13.9	4.86	ne		1	2	PAD	1	st.p. IM bez podrobností	ne	
20	19	12/18/1947	2/4/2010	62	0	17.8	NA		ano		0			0		15-20 min. pouze masáž, p	
21	20	9/19/1943	2/15/2010	66	0		3.09	7.13	ano		1	2	PAD	1	st.p. IM 1997, difuzní postižení tepen		
22	21	11/23/1975	5/16/2011	35	1		NA		ano	gestační	1	1	insulin	0		sporně	
23	22	4/18/1943	6/1/2010	67	0	25.3	14.8	7.05	ano		1		PAD	0			
24	23	12/16/1931	7/30/2010	79	1		NA		ano		0			1	difuzní postižení tepen ke konzerv. Pos	laická 8 min.	
25	24	5/18/1943	8/7/2010	67	0		NA		ano		0			0			
26	25	10/15/1938	11/7/2010	72	0		NA		ano		1	2	PAD	0			
27	26	11/30/1937	12/21/2010	73	0	22.4	10.5	7.09	ne		1	2	insulin	1	st.p. QIM spodní stěny 5/06, 3-VD, st.p opakovaných PCI		
28	27	7/20/1955	12/23/2010	55	1	6.7	1.1	7.51	ne		0	1		0			
29																	

věk přepočítán (zaokrouhlení vizuální, jinak des. číslo)

UKÁZKA DAT – NÁPRAVA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	ID	datum.narozeni	KPCR	vek.KPCR	pohlavi	glykemie	laktát	pH	hypertenze	hypertenze.pozn	DM	DM.typ	DM.terapie	ICHS	ICHS.pozn	laická KPCR	
2	1	3/20/1962	7/10/2007	45	0	20.1	13.03	6.97	ne		0			0			
3	2	12/8/1930	7/15/2007	77	0	15.3	6.4	7.29	ano		1	NA	PAD	0		10 min	
4	3	5/15/1946	8/3/2007	61	0	15.7	9.9	7.05	ne		0			0			
5	4	6/6/1941	9/13/2007	66	0	16.4	2.66	7.31	ano		0			1	dle urgentní SKG po KPCR nález k CA	?	
6	5	4/12/1952	9/26/2007	55	0	23.6	3.07	7.16	ne		0			0		ne	
7	6	7/21/1954	11/14/2007	53	0	10.3	10.06	7.11	ne		0			0		ano	
8	7	10/18/1955	2/3/2008	52	0	12.8	2.15	7.33	ne		0			0		ano, 15 min	
9	8	11/24/1939	3/29/2008	68	0	12.9	NA	7.36	ano		0			0			
10	9	10/30/1935	6/7/2008	73	1	9.2	7.54	6.96	ne		1	2	PAD	0			
11	10	1/30/1926	6/24/2008	82	0		NA				1	2	dieta	1	sy AP		
12	11	10/29/1950	9/2/2008	58	0	12.5	NA				0			1	3 -VD, ad CABG elektivně	laická 6 min.	
13	12	7/25/1952	2/9/2009	57	1		NA		ano		0			0			
14	13	6/7/1966	5/29/2009	43	0		NA				NA			NA			
15	14	10/22/1932	9/4/2009	77	0	8.3	NA		ano		1	2	PAD	1	3- VD ad CABG časné překladem		
16	15	3/16/1955	9/19/2009	55	0	15.1	5.78	7.23	ano		0			0		ano, čas neuveden	
17	16	12/18/1953	10/13/2009	56	0	10.7	0.83	7.32	ano		0			1	3-VD ad CABG časné překladem	ano, telefonicky as.	
18	17	12/4/1964	10/30/2009	45	0	23.9	NA		ne		0			0		ano	
19	18	7/23/1943	12/29/2009	66	0	30.6	13.9	4.86	ne		1	2	PAD	1	st.p. IM bez podrobností	ne	
20	19	12/18/1947	2/4/2010	62	0	17.8	NA		ano		0			0		15-20 min. pouze masáž, p	
21	20	9/19/1943	2/15/2010	66	0		3.09	7.13	ano		1	2	PAD	1	st.p. IM 1997, difuzní postižení tepen		
22	21	11/23/1975	5/16/2011	35	1		NA		ano	gestační	1	1	insulin	0		sporně	
23	22	4/18/1943	6/1/2010	67	0	25.3	14.8	7.05	ano		1		PAD	0			
24	23	12/16/1931	7/30/2010	79	1		NA		ano		0			1	difuzní postižení tepen ke konzerv. Pos	laická 8 min.	
25	24	5/18/1943	8/7/2010	67	0		NA		ano		0			0			
26	25	10/15/1938	11/7/2010	72	0		NA		ano		1	2	PAD	0			
27	26	11/30/1937	12/21/2010	73	0	22.4	10.5	7.09	ne		1	2	insulin	1	st.p. QIM spodní stěny 5/06, 3-VD, st.p	opakovaných PCI	
28	27	7/20/1955	12/23/2010	55	1	6.7	1.1	7.51	ne		0	1		0			
29																	

čárka opravena na tečku, nejsou texty; NA je ekvivalentní prázdnému poli

UKÁZKA DAT – NÁPRAVA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	ID	datum.narozeni	KPCR	vek.KPCR	pohlavi	glykemie	laktát	pH	hypertenze	hypertenze.pozn	DM	DM.typ	DM.terapie	ICHS	ICHS.pozn	laická KPCR	
2	1	3/20/1962	7/10/2007	45	0	20.1	13.03	6.97	ne		0			0			
3	2	12/8/1930	7/15/2007	77	0	15.3	6.4	7.29	ano		1	NA	PAD	0		10 min	
4	3	5/15/1946	8/3/2007	61	0	15.7	9.9	7.05	ne		0			0			
5	4	6/6/1941	9/13/2007	66	0	16.4	2.66	7.31	ano		0			1	dle urgentní SKG po KPCR nález k CA	?	
6	5	4/12/1952	9/26/2007	55	0	23.6	3.07	7.16	ne		0			0		ne	
7	6	7/21/1954	11/14/2007	53	0	10.3	10.06	7.11	ne		0			0		ano	
8	7	10/18/1955	2/3/2008	52	0	12.8	2.15	7.33	ne		0			0		ano, 15 min	
9	8	11/24/1939	3/29/2008	68	0	12.9	NA	7.36	ano		0			0			
10	9	10/30/1935	6/7/2008	73	1	9.2	7.54	6.96	ne		1	2	PAD	0			
11	10	1/30/1926	6/24/2008	82	0		NA				1	2	dieta	1	sy AP		
12	11	10/29/1950	9/2/2008	58	0	12.5	NA				0			1	3-VD, ad CABG elektivně	laická 6 min.	
13	12	7/25/1952	2/9/2009	57	1		NA		ano		0			0			
14	13	6/7/1966	5/29/2009	43	0		NA				NA			NA			
15	14	10/22/1932	9/4/2009	77	0	8.3	NA		ano		1	2	PAD	1	3- VD ad CABG časné překladem		
16	15	3/16/1955	9/19/2009	55	0	15.1	5.78	7.23	ano		0			0		ano, čas neuveden	
17	16	12/18/1953	10/13/2009	56	0	10.7	0.83	7.32	ano		0			1	3-VD ad CABG časné překladem	ano, telefonicky as.	
18	17	12/4/1964	10/30/2009	45	0	23.9	NA		ne		0			0		ano	
19	18	7/23/1943	12/29/2009	66	0	30.6	13.9	4.86	ne		1	2	PAD	1	st.p. IM bez podrobností	ne	
20	19	12/18/1947	2/4/2010	62	0	17.8	NA		ano		0			0		15-20 min. pouze masáž, p	
21	20	9/19/1943	2/15/2010	66	0		3.09	7.13	ano		1	2	PAD	1	st.p. IM 1997, difuzní postižení tepen		
22	21	11/23/1975	5/16/2011	35	1		NA		ano	gestační	1	1	insulin	0		sporně	
23	22	4/18/1943	6/1/2010	67	0	25.3	14.8	7.05	ano		1		PAD	0			
24	23	12/16/1931	7/30/2010	79	1		NA		ano		0			1	difuzní postižení tepen ke konzerv. Pos	laická 8 min.	
25	24	5/18/1943	8/7/2010	67	0		NA		ano		0			0			
26	25	10/15/1938	11/7/2010	72	0		NA		ano		1	2	PAD	0			
27	26	11/30/1937	12/21/2010	73	0	22.4	10.5	7.09	ne		1	2	insulin	1	st.p. QIM spodní stěny 5/06, 3-VD, st.p	opakovaných PCI	
28	27	7/20/1955	12/23/2010	55	1	6.7	1.1	7.51	ne		0	1		0			
29																	

rozlišeno ne a chybějící, poznámka je vedle

UKÁZKA DAT – NÁPRAVA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	ID	datum.narozeni	KPCR	vek.KPCR	pohlavi	glykemie	laktát	pH	hypertenze	hypertenze.pozn	DM	DM.typ	DM.terapie	ICHS	ICHS.pozn	laická KPCR	
2	1	3/20/1962	7/10/2007	45	0	20.1	13.03	6.97	ne		0			0			
3	2	12/8/1930	7/15/2007	77	0	15.3	6.4	7.29	ano		1	NA	PAD	0		10 min	
4	3	5/15/1946	8/3/2007	61	0	15.7	9.9	7.05	ne		0			0			
5	4	6/6/1941	9/13/2007	66	0	16.4	2.66	7.31	ano		0			1	dle urgentní SKG po KPCR nález k CA	?	
6	5	4/12/1952	9/26/2007	55	0	23.6	3.07	7.16	ne		0			0		ne	
7	6	7/21/1954	11/14/2007	53	0	10.3	10.06	7.11	ne		0			0		ano	
8	7	10/18/1955	2/3/2008	52	0	12.8	2.15	7.33	ne		0			0		ano, 15 min	
9	8	11/24/1939	3/29/2008	68	0	12.9	NA	7.36	ano		0			0			
10	9	10/30/1935	6/7/2008	73	1	9.2	7.54	6.96	ne		1	2	PAD	0			
11	10	1/30/1926	6/24/2008	82	0		NA				1	2	dieta	1	sy AP		
12	11	10/29/1950	9/2/2008	58	0	12.5	NA				0			1	3-VD, ad CABG elektivně	laická 6 min.	
13	12	7/25/1952	2/9/2009	57	1		NA		ano		0			0			
14	13	6/7/1966	5/29/2009	43	0		NA				NA			NA			
15	14	10/22/1932	9/4/2009	77	0	8.3	NA		ano		1	2	PAD	1	3- VD ad CABG časné překladem		
16	15	3/16/1955	9/19/2009	55	0	15.1	5.78	7.23	ano		0			0		ano, čas neuveden	
17	16	12/18/1953	10/13/2009	56	0	10.7	0.83	7.32	ano		0			1	3-VD ad CABG časné překladem	ano, telefonicky as.	
18	17	12/4/1964	10/30/2009	45	0	23.9	NA		ne		0			0		ano	
19	18	7/23/1943	12/29/2009	66	0	30.6	13.9	4.86	ne		1	2	PAD	1	st.p. IM bez podrobností	ne	
20	19	12/18/1947	2/4/2010	62	0	17.8	NA		ano		0			0		15-20 min. pouze masáž, p	
21	20	9/19/1943	2/15/2010	66	0		3.09	7.13	ano		1	2	PAD	1	st.p. IM 1997, difuzní postižení tepen		
22	21	11/23/1975	5/16/2011	35	1		NA		ano	gestační	1	1	insulin	0		sporně	
23	22	4/18/1943	6/1/2010	67	0	25.3	14.8	7.05	ano		1		PAD	0			
24	23	12/16/1931	7/30/2010	79	1		NA		ano		0			1	difuzní postižení tepen ke konzerv. Pos	laická 8 min.	
25	24	5/18/1943	8/7/2010	67	0		NA		ano		0			0			
26	25	10/15/1938	11/7/2010	72	0		NA		ano		1	2	PAD	0			
27	26	11/30/1937	12/21/2010	73	0	22.4	10.5	7.09	ne		1	2	insulin	1	st.p. QIM spodní stěny 5/06, 3-VD, st.p	opakovaných PCI	
28	27	7/20/1955	12/23/2010	55	1	6.7	1.1	7.51	ne		0	1		0			
29																	

rozlišeno ne a chybějící, poznámka je vedle

UKÁZKA DAT – NÁPRAVA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ID	datum.narozeni	KPCR	vek.KPCR	pohlavi	glykemie	laktát	pH	hypertenze	hypertenze.pozn	DM	DM.typ	DM.terapie	ICHS	ICHS.pozn	laická KPCR
2	1	3/20/1962	7/10/2007	45	0	20.1	13.03	6.97	ne		0			0		
3	2	12/8/1930	7/15/2007	77	0	15.3	6.4	7.29	ano		1	NA	PAD	0		10 min
4	3	5/15/1946	8/3/2007	61	0	15.7	9.9	7.05	ne		0			0		
5	4	6/6/1941	9/13/2007	66	0	16.4	2.66	7.31	ano		0			1	dle urgentní SKG po KPCR nález k CA	?
6	5	4/12/1952	9/26/2007	55	0	23.6	3.07	7.16	ne		0			0		ne
7	6	7/21/1954	11/14/2007	53	0	10.3	10.06	7.11	ne		0			0		ano
8	7	10/18/1955	2/3/2008	52	0	12.8	2.15	7.33	ne		0			0		ano, 15 min
9	8	11/24/1939	3/29/2008	68	0	12.9	NA	7.36	ano		0			0		
10	9	10/30/1935	6/7/2008	73	1	9.2	7.54	6.96	ne		1	2	PAD	0		
11	10	1/30/1926	6/24/2008	82	0		NA				1	2	dieta	1	sy AP	
12	11	10/29/1950	9/2/2008	58	0	12.5	NA				0			1	3-VD, ad CABG elektivně	laická 6 min.
13	12	7/25/1952	2/9/2009	57	1		NA		ano		0			0		
14	13	6/7/1966	5/29/2009	43	0		NA				NA			NA		
15	14	10/22/1932	9/4/2009	77	0	8.3	NA		ano		1	2	PAD	1	3- VD ad CABG časné překladem	
16	15	3/16/1955	9/19/2009	55	0	15.1	5.78	7.23	ano		0			0		ano, čas neuveden
17	16	12/18/1953	10/13/2009	56	0	10.7	0.83	7.32	ano		0			1	3-VD ad CABG časné překladem	ano, telefonicky as.
18	17	12/4/1964	10/30/2009	45	0	23.9	NA		ne		0			0		ano
19	18	7/23/1943	12/29/2009	66	0	30.6	13.9	4.86	ne		1	2	PAD	1	st.p. IM bez podrobností	ne
20	19	12/18/1947	2/4/2010	62	0	17.8	NA		ano		0			0		15-20 min. pouze masáž, p
21	20	9/19/1943	2/15/2010	66	0		3.09	7.13	ano		1	2	PAD	1	st.p. IM 1997, difúzní postižení tepen	
22	21	11/23/1975	5/16/2011	35	1		NA		ano	gestační	1	1	insulin	0		sporně
23	22	4/18/1943	6/1/2010	67	0	25.3	14.8	7.05	ano		1		PAD	0		
24	23	12/16/1931	7/30/2010	79	1		NA		ano		0			1	difúzní postižení tepen ke konzerv. Pos	laická 8 min.
25	24	5/18/1943	8/7/2010	67	0		NA		ano		0			0		
26	25	10/15/1938	11/7/2010	72	0		NA		ano		1	2	PAD	0		
27	26	11/30/1937	12/21/2010	73	0	22.4	10.5	7.09	ne		1	2	insulin	1	st.p. QIM spodní stěny 5/06, 3-VD, st.p opakovaných PCI	
28	27	7/20/1955	12/23/2010	55	1	6.7	1.1	7.51	ne		0	1		0		
29																

rozděleno na výskyt a poznámku

UKÁZKA DAT – NÁPRAVA

J	K	L	M
diabetes	DM	DM.typ	DM.terapie
na PAD	0		
	1	NA	PAD
	0		
	0		
	0		
	0		
	0		
DM II na PAD	1	2	PAD
DM II na dietě	1	2	diet
	0		
	0		
DM 2 na PAD	NA		
	1	2	PAD
	0		
	0		
DM II na PAD	1	2	PAD
	0		
DM II na PAD	1	2	PAD
DM 1. typu na insulinu	1	1	insulin
DM II na PAD	1		PAD
	0		
	0		
DM II na PAD	1	2	PAD
DM II na insulinu	1	2	insulin
	0	1	



rozlišen výskyt, typ, terapie do samostatných proměnných

UKÁZKA DAT – OPRAVA

```
34
35 # načtení opravených dat, pokud je máme v xlsx - jsou na druhém listu
36 mojedata.oprava = read.xls ("lekce1-vzor-dat.xlsx", sheet = 2, header = TRUE, na.strings = c("", "NA"))
37
38 # načtení opravených dat, pokud je máme v csv - jsou v jiném souboru
39 mojedata.oprava = read.csv ("lekce1-vzor-dat-2.csv", sep = "\t", header = TRUE, na.strings = c("", "NA"))
40
```

Environment

History

Import Dataset

Clear

List

Global Environment

Data

mojedata

27 obs. of 12 variables

kód.pacienta : int 1 2 3 4 5 6 7 8 9 10 ...

datum.narození: Factor w/ 27 levels "1/30/1926","10/15/1938",...: 16 14 19 21 17 24 3 8 6 1 ...

KPCR : Factor w/ 27 levels "10/13/2009","10/30/2009",...: 18 19 21 23 26 3 9 12 17 16 ...

věk.při.KPCR. : Factor w/ 27 levels "26 let 0 m","30 let 10 m",...: 24 2 14 9 17 20 23 8 5 1 ...

pohlaví : int 0 0 0 0 0 0 0 0 1 0 ...

glykemie : Factor w/ 21 levels "", "10,3", "10.7",...: 12 8 9 10 14 2 4 6 20 1 ...

X.laktát : Factor w/ 17 levels "", "0.83", "1.1",...: 6 14 16 10 11 4 9 17 15 1 ...

pH : num 6.97 7.29 7.05 7.31 7.16 7.11 7.33 7.36 6.96 NA ...

hypertenze : Factor w/ 3 levels "", "ano", "gestační": 1 2 1 2 1 1 1 2 1 1 ...

diabetes : Factor w/ 7 levels "", "DM 1. typu na insulinu",...: 1 7 1 1 1 1 1 6 4 ...

ICHS.chronická: Factor w/ 10 levels "", "3 -VD, ad CABG elektivně",...: 1 1 1 6 1 1 1 1 1 10 ...

laická.KPCR : Factor w/ 12 levels "", "?", "10 min",...: 1 3 1 2 11 5 6 1 1 1 ...

mojedata.oprava

27 obs. of 16 variables

ID : int 1 2 3 4 5 6 7 8 9 10 ...

datum.narození : Factor w/ 27 levels "1/30/1926","10/15/1938",...: 16 14 19 21 17 24 3 8 6 1 ...

KPCR : Factor w/ 27 levels "10/13/2009","10/30/2009",...: 18 19 21 23 26 3 9 12 17 16 ...

vek.KPCR : num 45.2 76.6 61.2 66.2 55.4 ...

pohlaví : int 0 0 0 0 0 0 0 0 1 0 ...

glykemie : num 20.1 15.3 15.7 16.4 23.6 10.3 12.8 12.9 9.2 NA ...

laktat : num 13.03 6.4 9.9 2.66 3.07 ...

pH : num 6.97 7.29 7.05 7.31 7.16 7.11 7.33 7.36 6.96 NA ...

hypertenze : Factor w/ 2 levels "ano", "ne": 2 1 2 1 2 2 2 1 2 NA ...

hypertenze.pozn: Factor w/ 1 level "gestační": NA NA NA NA NA NA NA NA NA NA ...

DM : int 0 1 0 0 0 0 0 0 1 1 ...

DM.typ : int NA NA NA NA NA NA NA NA 2 2 ...

DM.terapie : Factor w/ 3 levels "dieta", "insulin",...: NA 3 NA NA NA NA NA NA 3 1 ...

ICHS : int 0 0 0 1 0 0 0 0 0 1 ...

ICHS.pozn : Factor w/ 9 levels "3 -VD, ad CABG elektivně",...: NA NA NA 5 NA NA NA NA 9 ...

laicka.KPCR : Factor w/ 11 levels "?", "10 min", "15-20 min. pouze masáž, poté ALS 20-25 min.",...: NA 2 NA 1 10 4...

	A	B	C	D	E
1	<u>proměnná</u>	<u>jednotka / hodnota</u>	<u>popis</u>		
2	ID		<u>identifikační kód pacienta</u>		
3					
4	<u>datum.narozeni</u>	<u>datum</u>	<u>datum narození</u>		
5					
6	KPCR	<u>datum</u>	<u>datum kardiopulmonární resuscitace</u>		
7					
8	<u>vek.KPCR</u>	<u>rok</u>	<u>věk v okamžiku KPCR</u>		
9					
10	<u>pohlavi</u>	<u>0 = muž</u>	<u>pohlaví pacienta</u>		
11		<u>1 = žena</u>			
12					
13	<u>glykemie</u>	<u>mmol/l</u>	<u>hodnota glykemie v séru</u>		
14			<u>norma 3.30 – 5.80</u>		
15					
16	<u>laktát</u>	<u>mmol/l</u>	<u>hodnota laktátu v séru</u>		
17			<u>norma 0.50 – 2.00</u>		
18					
19	pH	<u>bezrozměrné</u>	<u>pH krve</u>		
20			<u>norma 7.36-7.44</u>		
21					
22	<u>hypertenze</u>	<u>ano/ne</u>	<u>diagnóza hypertenze v anamnéze</u>		
23	<u>hypertenze.pozn</u>	<u>text</u>	<u>poznámka k hypertenzi</u>		
24					
25	DM	<u>0 = ne</u>	<u>diagnóza diabetu v anamnéze</u>		
26		<u>1 = ano</u>			
27					
28	<u>DM.typ</u>	<u>1 = 1. typu</u>	<u>typ diabetu</u>		
29		<u>2 = 2. typu</u>			
30		<u>3 = gestační</u>			
31					
32	<u>DM.terapie</u>	<u>PAD</u>	<u>aktuální terapie diabetu</u>		
33		<u>insulin</u>			
34		<u>dieta</u>			
35					
36	ICHS	<u>ano/ne</u>	<u>diagnóza ICHS v anamnéze</u>		
37	<u>ICHS.pozn</u>	<u>text</u>	<u>poznámka k ICHS</u>		
38					

- přehledný seznam slouží jako reference
 - pro autora databáze
 - pro analytika
 - pro návrat v budoucnosti
- snižuje riziko chyby, záměny, nepochopení

VÝSLEDEK OPRAV

- data jsou v R přímo čitelná se správným typem
- dají se na ně rovnou uplatnit analytické nástroje (tabulky, výpočet průměrů atd)
- radikálně rychlejší, efektivnější práce s menším rizikem chyb
- dobrá příprava předem se rozhodně vyplatí

DOMÁCÍ ÚKOL

- navrhnout, jak přepracovat poslední sloupec (laická KPCR)
- zkusit si natáhnout do Rstudia nějaká svoje data a podívat se na ně

KDE HLEDAT POMOC S R

- R pro biology (Karel Zvára, KPMS)
 - <http://www.karlin.mff.cuni.cz/~zvara/biostat/2009/eRkoBio.pdf>
- Materiály k Rku (Arnošt Komárek, KPMS MFF UK)
 - http://www.karlin.mff.cuni.cz/~komarek/vyuka/2011_12/Rko-2011.html
 - Hrátky s R (správný odkaz):
https://www.karlin.mff.cuni.cz/~kulich/vyuka/Rdoc/arnost_introR.pdf
- Cookbook for R (www.cookbook-r.com)
- Stackoverflow (stackoverflow.com)
 - dotazy s řešeními od uživatelů
 - typicky se sem dostanu po dotazu v googlu

DĚKUJI ZA POZORNOST

www.biostatisticka.cz