MUNI
MED

MIAM021p(s) **Analýza a management dat pro zdravotnické obory – přednáška a cvičení** (jaro 2020)

MICHAL UHER

Institut biostatistiky a analýz LF MU
uher@iba.muni.cz

Ústav zdravotnických informací a statistiky ČR
Michal.Uher@uzis.cz

MUNI
MED

Základy popisné statistiky

Typy proměnných
Popisná statistika

Typy proměnných

- **Kvalitativní (kategoriální) proměnná**
  lze ji řadit do kategorií, ale nelze ji kvantifikovat
  Příklad: ??

- **Kvantitativní (numerická) proměnná**
  můžeme ji přiřadit číselnou hodnotu
  Příklad: ??

Typy proměnných

- **Kvalitativní (kategoriální) proměnná**
  lze ji řadit do kategorií, ale nelze ji kvantifikovat
  Příklad: pohlaví, HIV status, barva vlasů …

- **Kvantitativní (numerická) proměnná**
  můžeme ji přiřadit číselnou hodnotu
  Příklad: výška, hmotnost, teplota, počet hospitalizací …

MUNI MED

Kvalitativní proměnné, znaky

- **Binární znaky:** dvě kategorie, obvykle se kódují pomocí číslic 1 (přítomnost znaku) a 0 (nepřítomnost znaku).
  Příklad: ??

- **Nominální znaky**: několik kategorií (A, B, C), které nelze uspořádat.
  Příklad: ??

- **Ordinální znaky:** několik kategorií, které lze vzájemně seřadit, tedy můžeme se ptát, která je větší/menší (1<2<3).
  Příklad: ??

MUNI MED

Kvalitativní proměnné, znaky

- **Binární znaky:** dvě kategorie, obvykle se kódují pomocí číslic 1 (přítomnost znaku) a 0 (nepřítomnost znaku).
  Příklad: *Diabetes (1 – ano, 0 – ne),*
  Pohlaví (1 – muž, 0 – žena).
- **Nominální znaky**: několik kategorií (A, B, C), které nelze uspořádat.
  Příklad: *krevní skupiny (A – B – AB – 0)*

- **Ordinální znaky:** několik kategorií, které lze vzájemně seřadit, tedy můžeme se ptát, která je větší/menší (1<2<3).
  Příklad: *stupeň bolesti (mírná – střední – velká)*
  stadium maligního onemocnění (I – II – III – IV)

MUNI MED

Kvantitativní proměnné, znaky

- **Intervalové znaky:** interpretace rozdílu dvou hodnot (stejný interval mezi jednou a druhou dvojicí hodnot vyjadřuje i stejný rozdíl v intenzitě zkoumané vlastnosti). Společný znak intervalových znaků: nula byla stanovena uměle, tedy pouhou konvencí.
  Příklad: teplota měřená ve stupních Celsia, letopočet

Den	Teplota	Rozdíl *	Podíl *
1.	2 °C	-	-
2.	4 °C	+2	2x
3.	6 °C	+2	1,5x

* Srovnání s měřením z předchozího dne

← 1,5krát vyšší teplota ve srovnání s 2. dnem, přičemž došlo ke stejnému nárůstu teploty jako při srovnání 2. a 1. dne

MUNI MED

Kvantitativní proměnné, znaky

- **Poměrové znaky:** kromě rozdílu interpretujeme i podíl dvou hodnot.
  Příklad: výška v cm, hmotnost v kg, ...

Pacient	Hmotnost	Rozdíl *	Podíl *
1.	52 kg	-	-
2.	78 kg	+ 26	1,5x
3.	104 kg	+ 52	2x

* Srovnání s prvním pacientem

1,5krát vyšší hmotnost ve srovnání s prvním pacientem

2krát vyšší hmotnost ve srovnání s prvním pacientem

MUNI MED

Popisné statistiky

- **Charakteristiky polohy** (míry střední hodnoty, míry centrální tendence)

  Udávají, kolem jaké hodnoty se data centrují, resp. které hodnoty jsou nejčastější; popis „těžiště“ – míry polohy.

  Aritmetický průměr, medián, modus, geometrický průměr

- **Charakteristiky variability** (proměnlivosti)

  Zachycují rozptýlení hodnot v souboru (proměnlivost dat).

  Variační rozpětí, rozptyl, směrodatná odchylka, variační koeficient, střední chyba průměru

MUNI MED

Charakteristiky polohy

- **Modus:** nejčastěji se vyskytující hodnota proměnné v souboru (u kvalitativních proměnných).
- **α-kvantil:** je-li $\alpha \in (0,1)$, pak α-kvantil x_α je číslo, které rozděluje uspořádaný datový soubor na dolní úsek, obsahující aspoň podíl α všech dat a na horní úsek obsahující aspoň podíl 1-α všech dat.
- $x_{0,50}$ – **medián**, $x_{0,25}$ – **dolní kvartil**, $x_{0,75}$ – **horní kvartil**, $x_{0,1}$.... $x_{0,9}$ – decily
- **Medián:** hodnota, jež dělí řadu podle velikosti seřazených hodnot na dvě stejně početné poloviny.

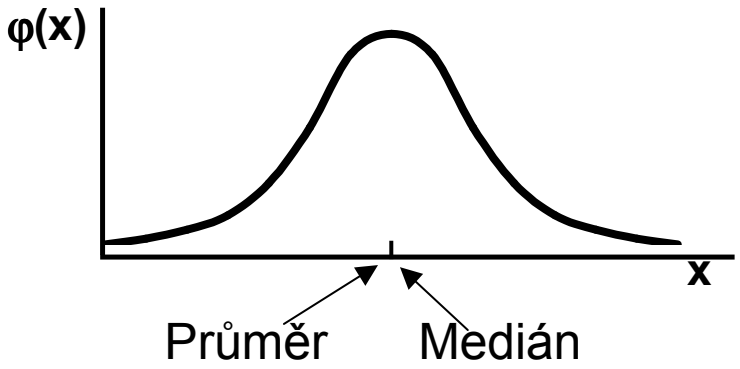
Charakteristiky polohy

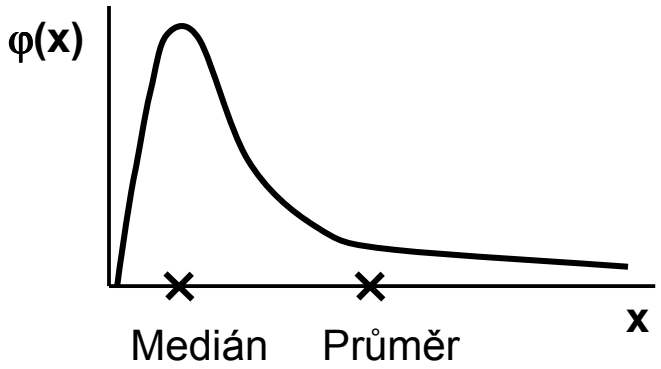
- **Aritmetický průměr**: je definován jako součet všech naměřených údajů vydělený jejich počtem,
  $\bar{x} = \sum_{i=1}^{n} x_i / n$ kde x_i jsou jednotlivé hodnoty a n jejich počet.

- **Geometrický průměr:** n kladných hodnot x_i, $\sqrt[n]{x_1 * ..... * x_n}$, má smysl všude, kde má nějaký informační smysl součin hodnot proměnné. Z praktického hlediska platí, že logaritmus geometrického průměru je roven aritmetickému průměru logaritmovaných hodnot souboru.

Průměr vs. medián

- **POZOR:** Průměr je silně ovlivněn extrémními hodnotami (tzv. odlehlá pozorování), medián jimi ovlivněn není.
- **Průměr** je vhodný ukazatel středu souboru u normálního, resp. symetrického rozložení, **medián** i v případě proměnných s neznámým rozdělením.
- V případě symetrického rozložení jsou průměr a medián v podstatě shodné, v případě asymetrického rozložení nikoliv!





MUNI MED

Charakteristiky variability

- **Kvartilové rozpětí:** $q = x_{0,75} - x_{0,25}$
- **Rozptyl (variance):** ukazatel šířky rozložení získaný na základě odchylky jednotlivých hodnot od průměru (jeho vypovídací schopnost je nejvyšší v případě symetrického/normálního rozložení).

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

- **Směrodatná odchylka (SD):** druhá odmocnina z rozptylu
- **Koeficient variance:** podíl SD ku průměru u poměrových znaků. Vyjadřuje se v procentech. Umožňuje porovnat variabilitu několika znaků.

Další popisné statistiky

- **Počet hodnot:** důležitý ukazatel, znamená jak moc lze na data spoléhat
- **Suma hodnot**
- **Minimum**
- **Maximum**
- **Variační rozpětí (rozsah):** rozdíl mezi největší a nejmenší hodnotou
- **Střední chyba průměru (SE):** měří rozptýlenost vypočítaného aritmetického průměru v různých výběrových souborech vybraných z jednoho základního souboru

Popis a vizualizace kvalitativních proměnných

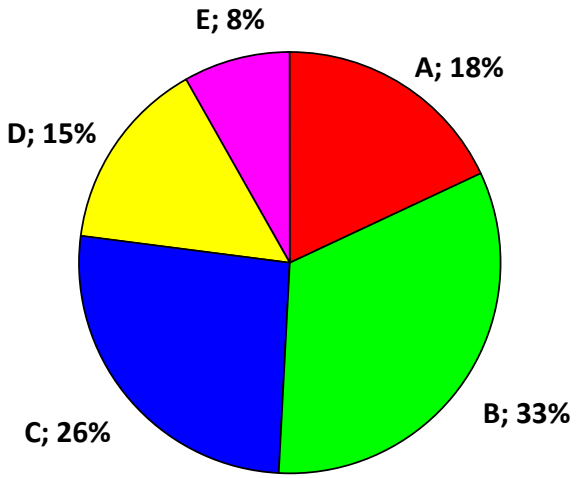
- **Popis kvalitativních dat:** četnost jednotlivých kategorií
- **Vizualizace kvalitativních dat:** koláčový nebo sloupcový graf

Příklad: Známka z biostatistiky (podzim 2014)

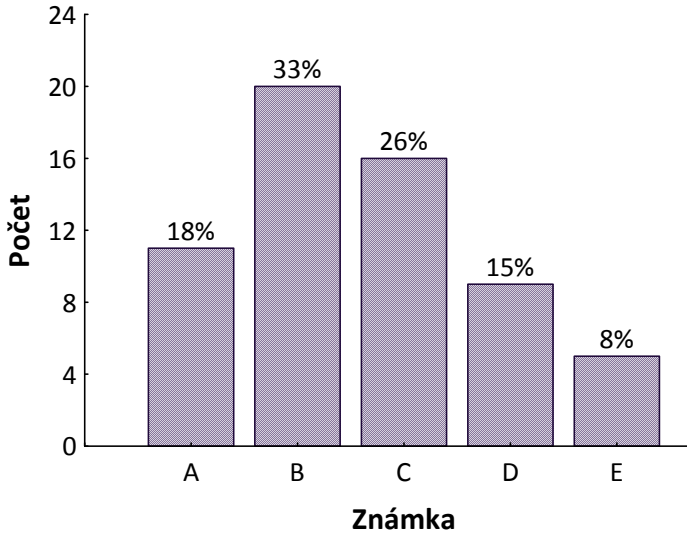
Frekvenční tabulka

Známka	n	%
A	11	18,0
B	20	32,8
C	16	26,2
D	9	14,8
E	5	8,2
F	0	0,0
Celkem	61	100,0

Koláčový graf



Sloupcový graf



MUNI MED

Popis kvantitativních dat

- **Popis kvantitativních dat:** charakteristika středu (průměr, medián aj.), charakteristika variability (rozptyl, rozsah hodnot, interkvartilové rozpětí aj.)

Příklad: Popis výšky pacientů (cm)

Popisné statistiky

Charakteristika	
N	61
Průměr (cm)	161,5
Medián (cm)	161,0
Sm. odchylka (cm)	4,7
Rozptyl (cm^2)	22,2
min-max (cm)	144 – 169
dolní-horní kvartil (cm)	158 - 164

Průměr a medián se téměř shodují. Co nám to říká?

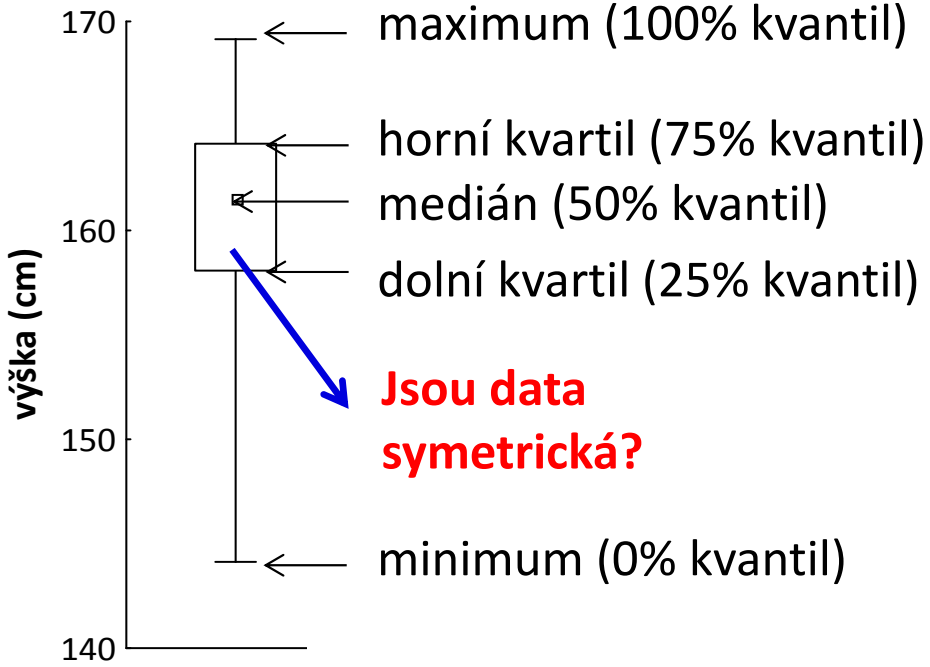
MUNI MED

Vizualizace kvantitativních dat

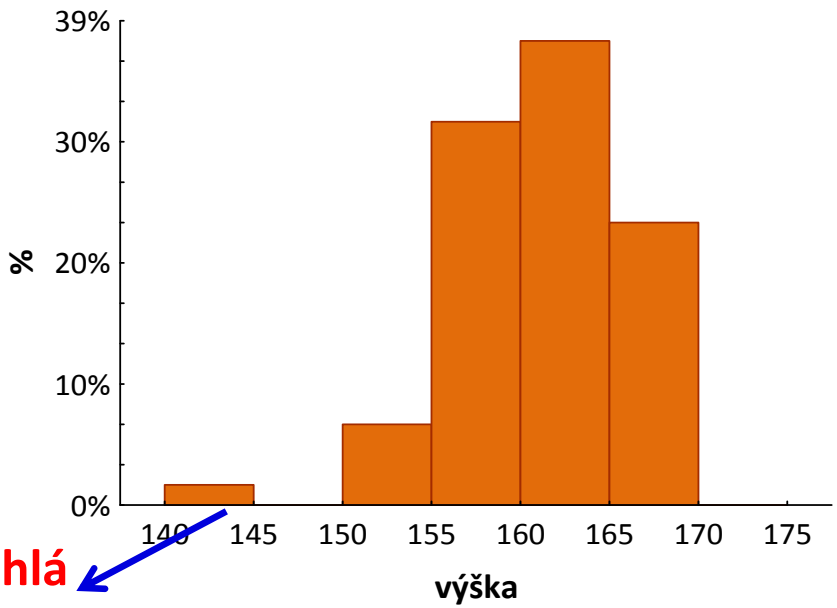
- **Vizualizace kvantitativních dat:** nejčastěji pomocí krabicového grafu nebo histogramu

Příklad: Popis výšky pacientů (cm)

Krabicový graf



Histogram



MUNI MED

MUNI
MED

Program Statistica

Představení programu Statistica
Praktické cvičení v programu Statistica

Program Statistica

Jak získat program Statistica:

https://inet.muni.cz

Login a heslo: UČO a primární heslo jako do IS-u.

V nabídce zvolit: **Provozní služby – Software – Nabídka softwaru**

Nalézt: **Statistica 13** – kliknout **Získat** a postupovat dle návodu

MUNI MED

MUNI
MED

Praktické cvičení v programu Statistica

Datový soubor

Rehabilitace po mozkovém infarktu

Data: 02_Biostatistika_Data02.sta* (24v by 407c)

Rehabilitace po mozkovem infarktu: data

	1 ID	2 Pohlavi	3 Vek	4 Etiologie	5 Lokalizace	6 Terapie	7 Komorbid	8 Barthel_inc	9 Kategorie_zavislosti_p	10 Ukoncen
1	1	muž	82	okluze neb	mozkové tepny	jiná farmakolog	0	25	vysoce závislý	propuště
2	2	žena	81	embolie	mozkové tepny	jiná farmakolog	2	20	vysoce závislý	přeložen
3	3	muž	55	okluze neb	mozkové tepny	jiná farmakolog	0	35	vysoce závislý	propuště
4	4	žena	46	embolie	mozkové tepny	intravenózní tro	0	20	vysoce závislý	propuště
5	5	muž	76	okluze neb	mozkové tepny	jiná farmakolog	0	45	částečně soběstačný	propuště
6	6	muž	72	okluze neb	mozkové tepny	jiná farmakolog	0	25	vysoce závislý	přeložen
7	7	muž	62	trombóza	mozkové tepny	jiná farmakolog	0	40	vysoce závislý	propuště
8	8	muž	64	trombóza	přívodní tepny	jiná farmakolog	0	15	vysoce závislý	propuště
9	9	žena	82	okluze neb	mozkové tepny	jiná farmakolog	0	10	vysoce závislý	přeložen
10	10	muž	58	trombóza	mozkové tepny	jiná farmakolog	0	25	vysoce závislý	propuště
11	11	muž	84	okluze neb	mozkové tepny	jiná farmakolog	0	40	vysoce závislý	propuště
12	12	žena	92	okluze neb	mozkové tepny	jiná farmakolog	0	30	vysoce závislý	propuště
13	13	žena	79	embolie	mozkové tepny	jiná farmakolog	1	40	vysoce závislý	propuště
14	14	muž	69	trombóza	mozkové tepny	jiná farmakolog	3	45	částečně soběstačný	propuště

MUNI MED

Rehabilitace po mozkovém infarktu

- Cvičný datový soubor obsahuje záznamy o **celkem 407 pacientech hospitalizovaných pro mozkový infarkt** na neurologickém oddělení akutní péče, kde jim byla poskytnuta terapie pro obnovu krevního oběhu v postižené části mozku.
- Po zvládnutí akutní fáze byl u pacientů vyhodnocen stupeň soběstačnosti v základních denních aktivitách (ADL) pomocí tzv. **indexu Barthelové** (BI) a byli přeloženi na **rehabilitační oddělení**.
- Po dvou týdnech byl opět dle BI vyhodnocen stupeň soběstačnosti a pacienti byli buď propuštěni do ambulantní péče, nebo přeloženi na oddělení následné péče.

Rehabilitace po mozkovém infarktu

Sbírané informace:

- základní demografické údaje (**pohlaví** a **věk**),
- informace o samotné diagnóze mozkové příhody (**etiologie** a **lokalizace uzávěru cévy**),
- informace o léčbě (typ indikované **terapie** a **výskyt komplikací**)
- informace o **způsobu ukončení rehabilitace**.
- Stupeň soběstačnosti před rehabilitací byl dodatečně zjištěn z neurologie a na konci rehabilitace byl vyplněn nový dotazník pro určení výsledného **indexu Barthelové**.

Úkol č. 1 – Popis kategoriálních dat

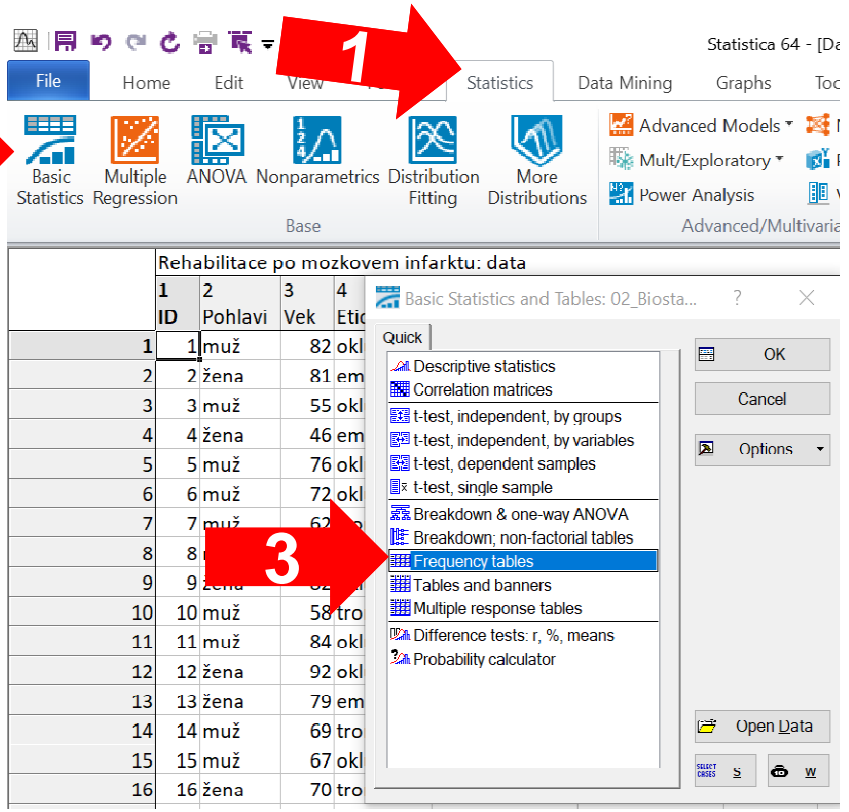
Zadání: „Proveďte základní popis zastoupení pohlaví u pacientů s mozkovým infarktem. Následně také srovnejte zastoupení pohlaví mezi třemi skupinami pacientů dle etiologie mozkové příhody.“

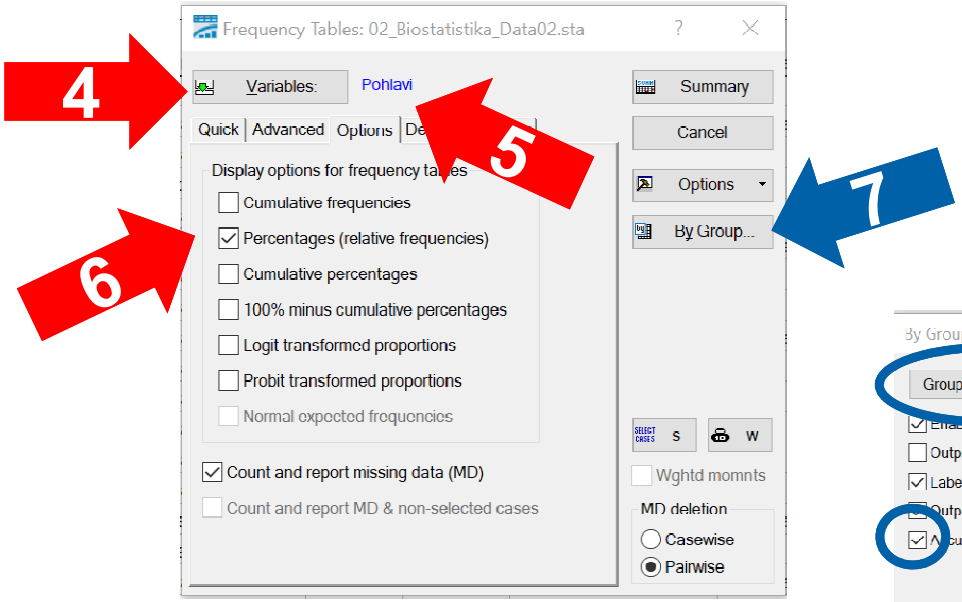
Postup:

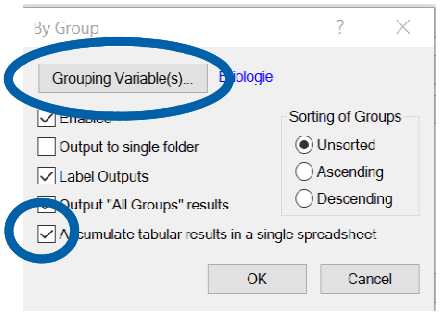
1. Pro popis dat je vhodné určit absolutní a relativní četnosti.
2. Grafické znázornění je ideální pomocí koláčového grafu.
3. V programu Statistica lze získat výsledky pro jakoukoli podskupinu souboru pomocí obecné funkce „By Group“ nebo „Select Cases“.

Úkol č. 1 – Řešení v programu Statistica

- V menu ***Statistics*** zvolíme ***Basic statistics*** a vybereme ***Frequency tables***.
- Vybereme proměnnou (***Variables***), kterou chceme analyzovat a na záložce ***Options*** zaškrtneme možnost ***Percentages (relative frequencies)***.







Chceme-li získat výsledky zvlášť pro podskupiny jiné proměnné, aktivujeme funkci ***By Group*** (vybrat třídící proměnnou a zaškrtnout ***Accumulate tabular results in a single spreadsheet***).

Úkol č. 1 – Výsledky v Statistica

Frekvenční tabulka proměnné pohlaví u pacientů s CMP

Category	Frequency table: Pohla	
	Count	Percent
muž	248	60,93366
žena	159	39,06634

Frekvenční tabulka proměnné pohlaví u pacientů s CMP dle etiologie centrální mozkové příhody

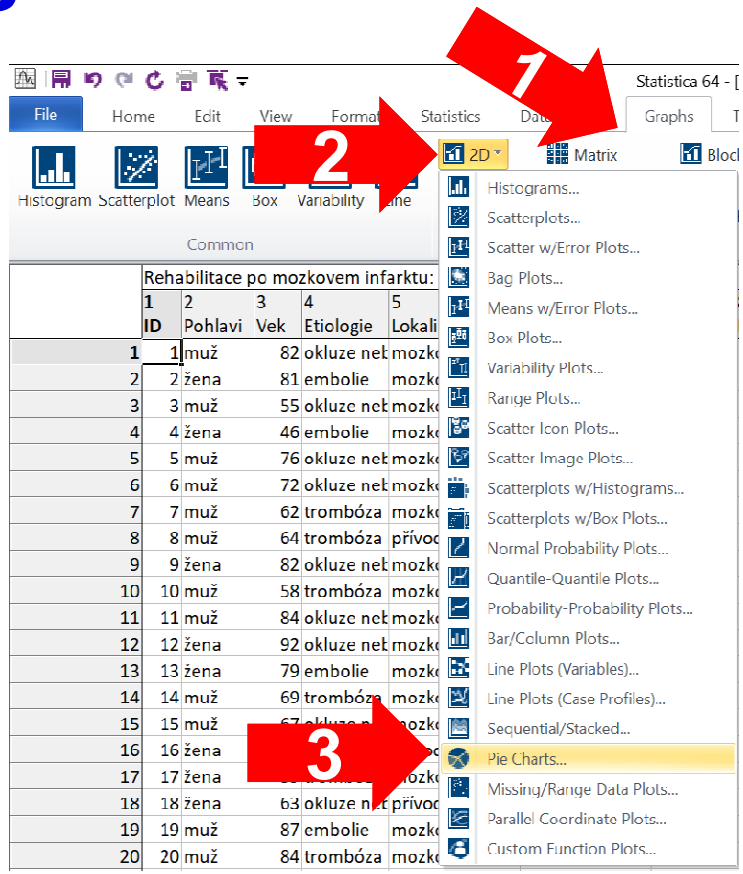
Category	Aggregate Results Frequency table: Pohlavi (02_Biostatistika_Dat		
	Etiologie	Count	Percent
muž	okluze nebo stenóza	125	62,18905
žena	okluze nebo stenóza	76	37,81095
žena	embolie	36	46,15385
muž	embolie	42	53,84615
muž	trombóza	81	63,28125
žena	trombóza	47	36,71875

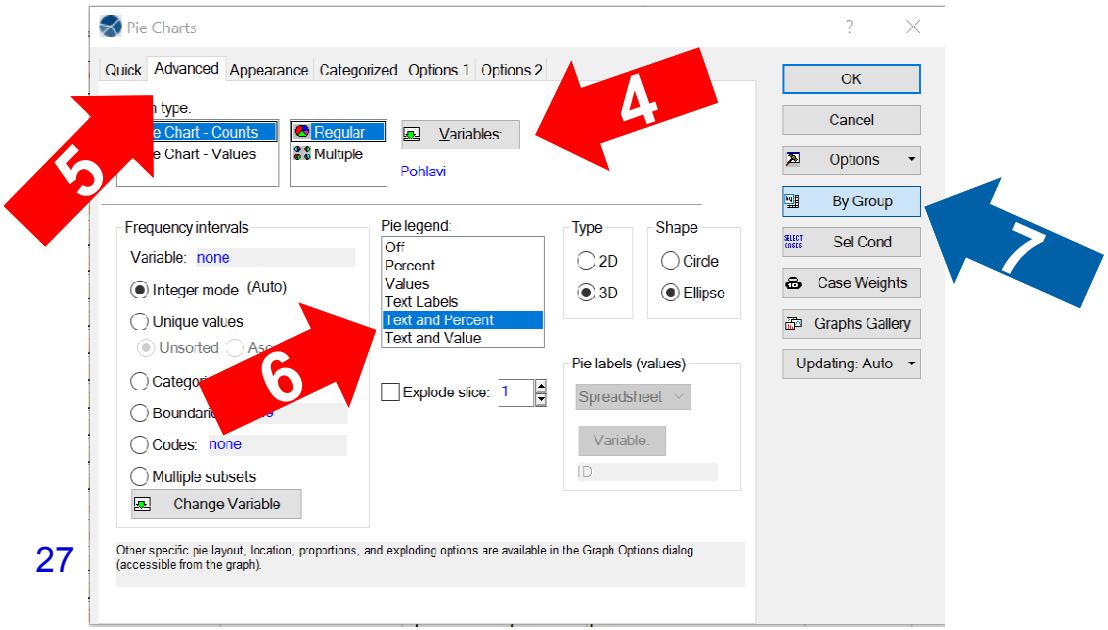
Zastoupení mužů a žen v celém souboru je 61 % oproti 39 %. Při srovnání pacientů dle etiologie mozkového infarktu se nejvíce liší pacienti s embolií, u kterých je podíl mužů jen 54 % oproti 46 % žen.

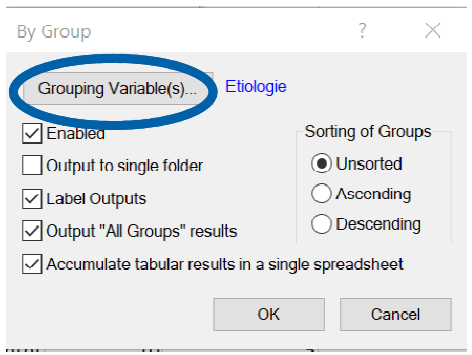
MUNI MED

Úkol č. 1 – Řešení v programu Statistica

- V menu ***Graphs*** zvolíme nabídku ***2D grafů*** a vybereme ***Pie Charts.***
- Vybereme proměnnou (***Variables***), kterou chceme analyzovat a na záložce ***Advanced*** vybereme v nastavení legendy možnost ***Text and Percent.***
- Chceme-li získat výsledky zvlášť pro podskupiny jiné proměnné, aktivujeme ***By Group*** (opět vybrat třídící proměnnou).



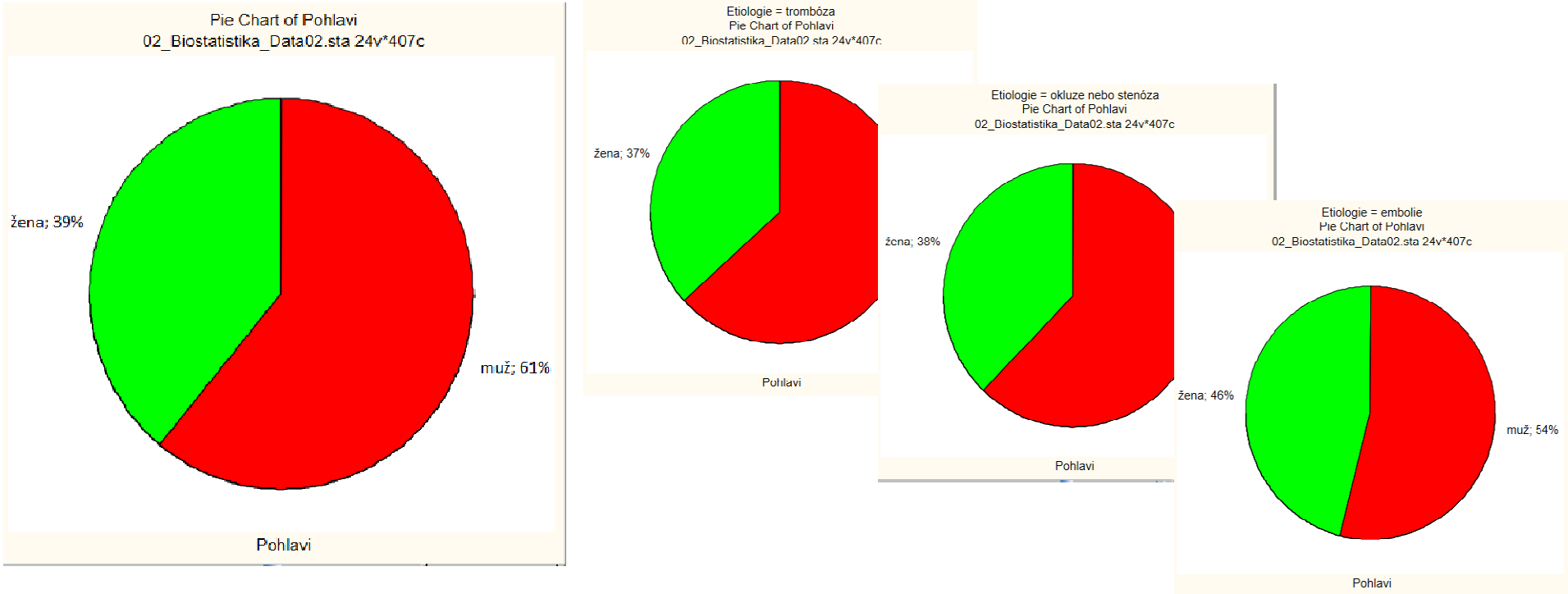




MUNI MED

Úkol č. 1 – Výsledky v Statistica

Koláčový graf proměnné pohlaví u pacientů s CMP



Zastoupení mužů v celém souboru je 61 % oproti 39 % žen.
Při srovnání pacientů dle etiologie mozkového infarktu se nejvíce liší pacienti s embolií, u kterých je podíl mužů jen 54 % oproti 46 % žen.

MUNI MED

Úkol č. 2 – Popis kvantitativních dat

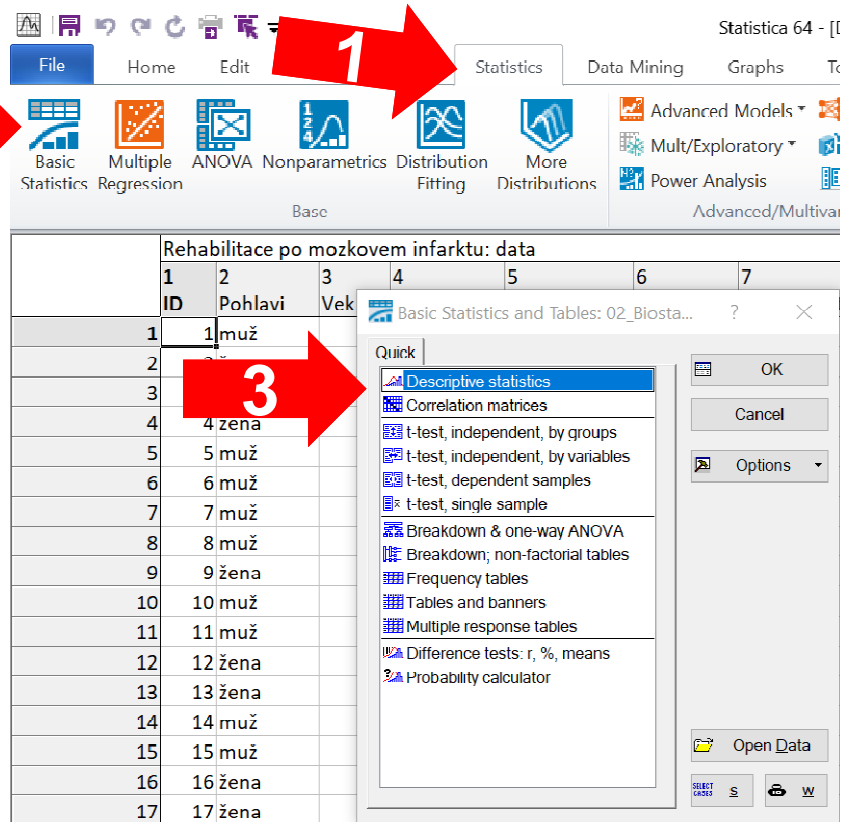
Zadání: „Proveďte základní popis soběstačnosti dle indexu Barthelové na konci rehabilitace po mozkovém infarktu. Následně také tento ukazatel srovnejte podle míry komplikací během léčby.“

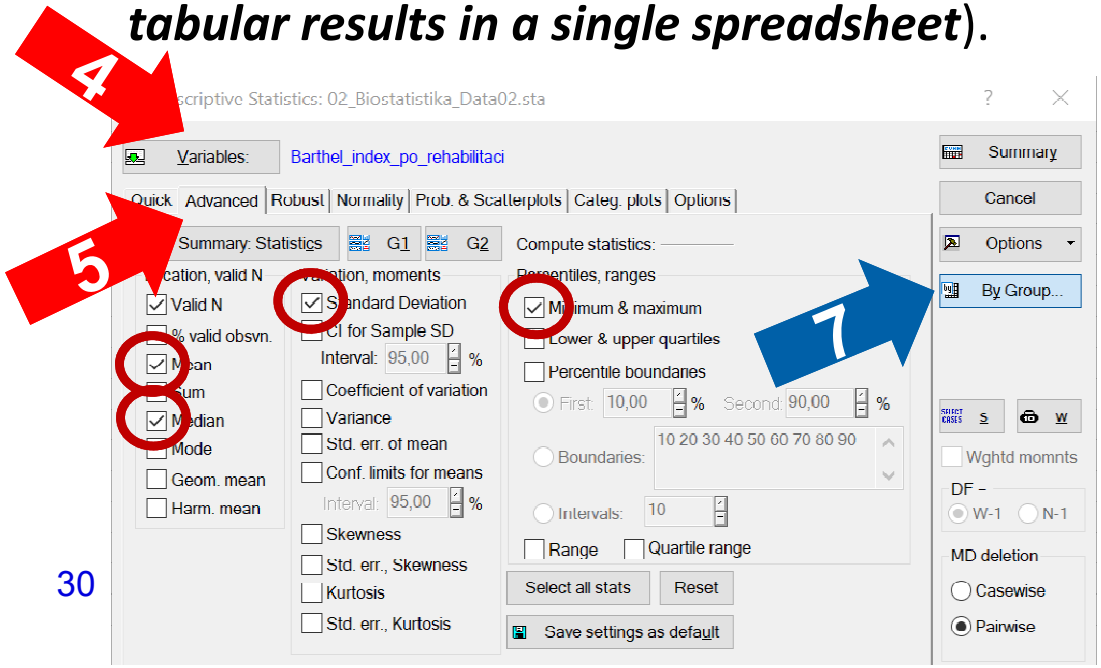
Postup:

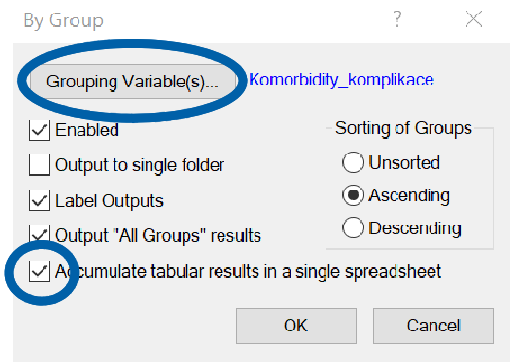
1. Pro popis dat je vhodné určit průměr, medián, směrodatnou odchylku, případně minimum a maximum.
2. Grafické znázornění je ideální pomocí histogramu. V případě srovnávání různých skupin je vhodný krabicový graf.
3. V programu Statistica lze získat výsledky pro jakoukoli podskupinu pomocí „By Group“ nebo „Select Cases“.

Úkol č. 2 – Řešení v programu Statistica

- V menu ***Statistics*** zvolíme ***Basic statistics*** a vybereme ***Descriptive statistics.***
- Vybereme proměnnou (***Variables***), kterou chceme analyzovat a na záložce ***Advanced*** zaškrtneme možnosti výpočtu ***Mean, Median, Std. Dev, Min. & Max.***
- Chceme-li získat výsledky zvlášť pro podskupiny jiné proměnné, použijeme ***By Group*** (vybrat třídící proměnnou a zaškrtnout ***Accumulate tabular results in a single spreadsheet***).







MUNI MED

Úkol č. 2 – Výsledky v Statistica

Popisná statistika indexu Barthelové na konci rehabilitace u pacientů s CMP

All Groups
Descriptive Statistics (02_Biostatistika_Data02.sta)

Variable	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.
Barthel_index_po_rehabilitaci	407	62,01474	70,00000	10,00000	95,00000	19,44095

Popisná statistika indexu Barthelové na konci rehabilitace u pacientů s CMP dle stupně komplikací

Aggregate Results
Descriptive Statistics (02_Biostatistika_Data02.sta)

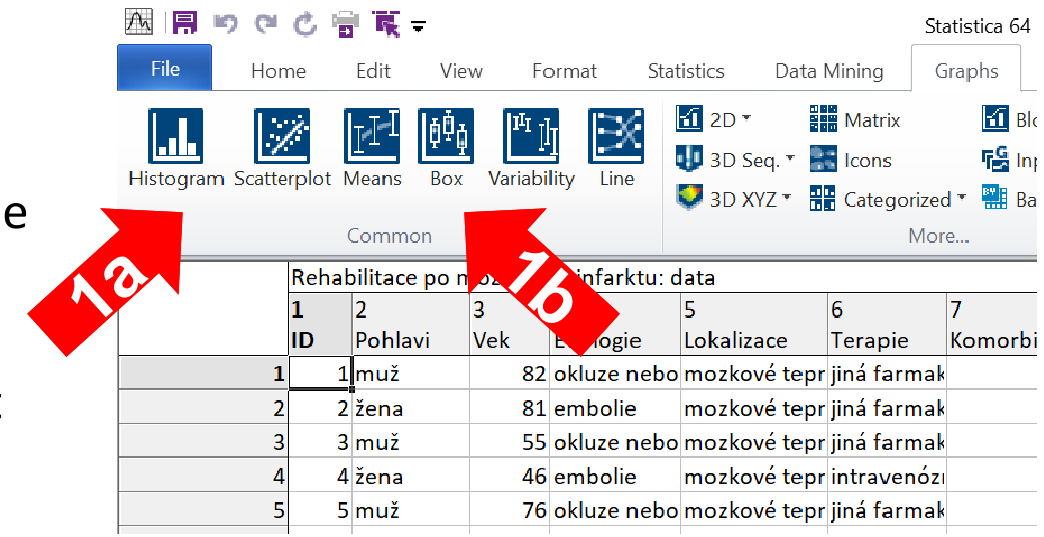
Variable	Komorbidity_komplikace	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.
Barthel_index_po_rehabilitaci	0	239	66,84100	70,00000	15,00000	95,00000	17,45221
Barthel_index_po_rehabilitaci	1	79	57,15190	60,00000	10,00000	90,00000	19,15998
Barthel_index_po_rehabilitaci	2	50	56,70000	60,00000	10,00000	90,00000	20,01556
Barthel_index_po_rehabilitaci	3	39	49,10256	50,00000	10,00000	85,00000	21,36376

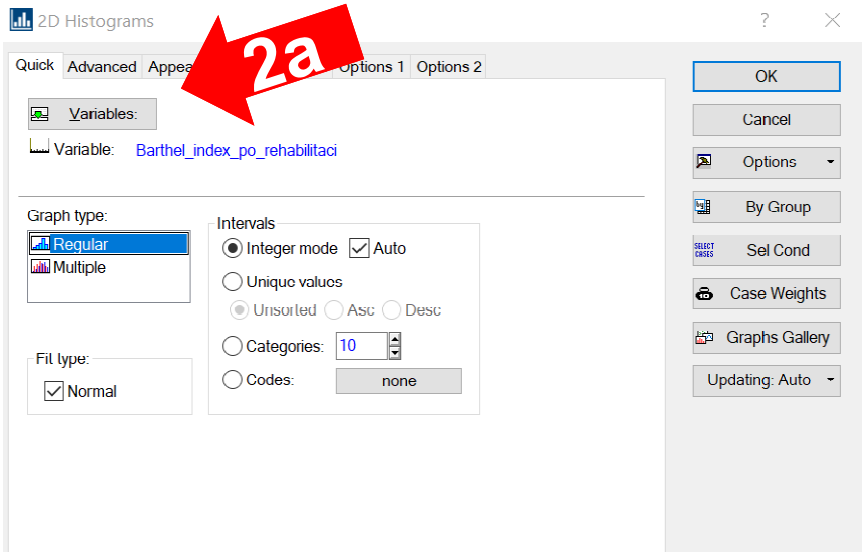
Celková průměrná hodnota indexu Barthelové je 62 bodů s mediánem 70 bodů. V závislosti na skóre komorbidit a komplikací během léčby je zřetelný pokles výsledné soběstačnosti s průměrem 67 bodů u nekomplikovaných případů až k 49 bodům u pacientů se stupněm komplikací 3.

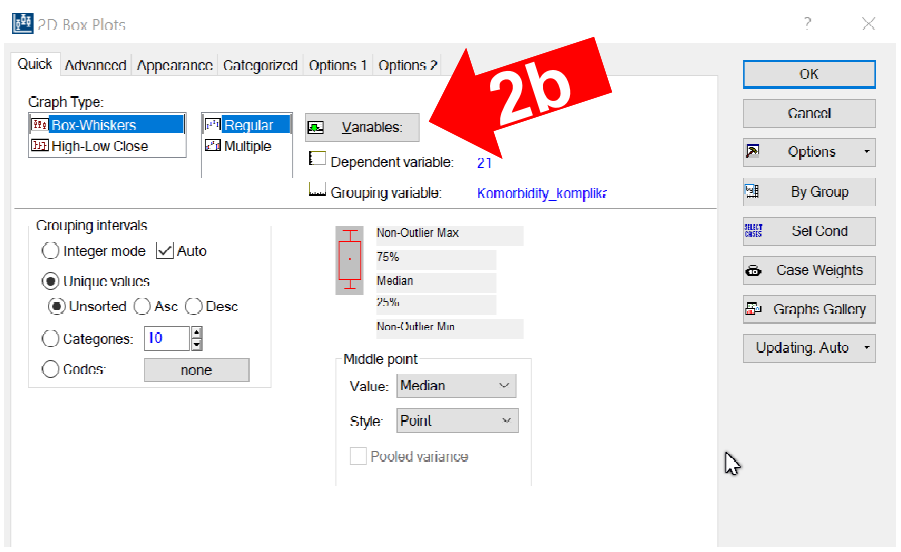
MUNI MED

Úkol č. 2 – Řešení v programu Statistica

- V menu ***Graphs*** zvolíme rovnou ***Histogram (a)*** nebo ***Box (b).***
- U <u>histogramu</u> pouze vybereme proměnnou (***Variables***), kterou chceme analyzovat.
- U <u>box-plotu</u> vybereme proměnnou (***Variables***), kterou chceme analyzovat (***dependent***), a proměnnou obsahující skupiny, které srovnáváme (***grouping***).

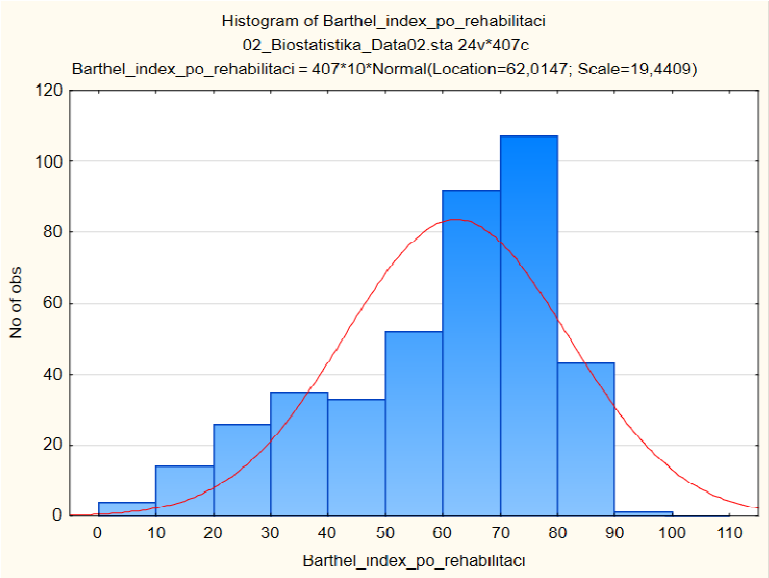




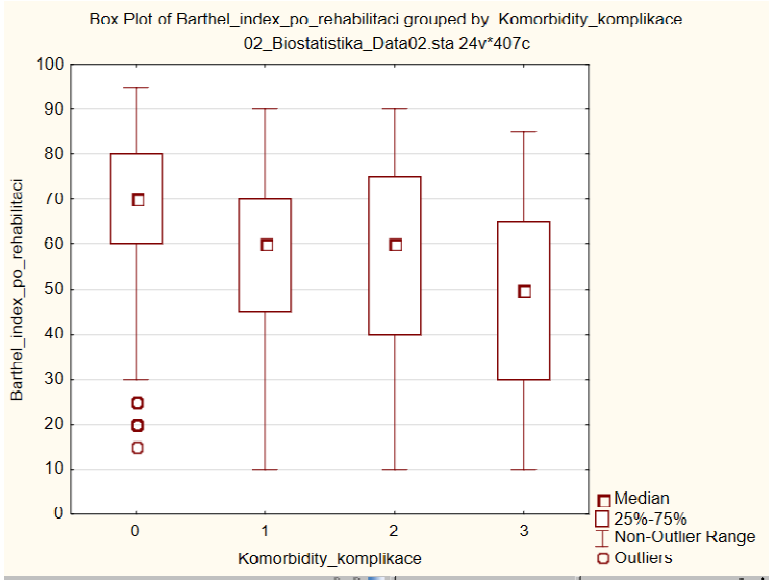


Úkol č. 2 – Výsledky v Statistica

Histogram indexu Barthelové na konci rehabilitace u pacientů s CMP



Krabicový graf indexu Barthelové na konci rehabilitace dle stupně komplikací



Celková průměrná hodnota indexu Barthelové je 62 bodů a tvar distribuce je asymetrický s hodnotami vyskytujícími se hlavně v rozmezí cca 50 až 90 bodů. V závislosti na skóre komorbidit a komplikací během léčby je zřetelný pokles výsledné soběstačnosti.

MUNI MED