

# Biostatistika



**Kontingenční tabulky v Excelu**  
**Základní popisné statistiky**  
**Představení programu Statistica**  
**Import a základní popis dat ve Statistice**

# I. Kontingenční tabulky v Excelu



# Kontingenční tabulka



- Frekvenční sumarizace dvou kategoriálních proměnných (binárních, nominálních nebo ordinálních proměnných).
- Obecně: **R x C kontingenční tabulka** (R – počet kategorií jedné proměnné, C – počet kategorií druhé proměnné).
- Speciální případ: 2 x 2 tabulka = čtyřpolní tabulka.
- Kontingenční tabulky: **absolutních četností, celkových procent, řádkových/sloupcových procent**
- Příklad: Sumarizace vyšetřených osob podle pohlaví a výsledku diagnostického testu.

Pohlaví	Výsledek vyšetření		Celkem
	Nemocný	Zdravý	
Muž	45	11	56
Žena	25	6	31
<b>Celkem</b>	70	17	87



**Jsou více nemocní muži nebo ženy?**

# Ukázka kontingenční tabulky

	Nemocný	Zdravý	Celkem
Muž	45	11	56
Žena	25	6	31
Celkem	70	17	87

Kontingenční tabulka  
absolutních četností

Větší počet nemocných mužů, který je dán pouze vyšším zastoupením mužů v celkovém vzorku (56 z 87)

	Nemocný	Zdravý	Celkem
Muž	80,4 %	19,6 %	100,0 %
Žena	80,6 %	19,4 %	100,0 %

Kontingenční tabulka  
řádkových procent

Po výpočtu relativních četností vidíme, že se muži a ženy neliší ve výskytu onemocnění



**Jsou více nemocní  
muži nebo ženy?**

# Kontingenční tabulky v Excelu: zdroj dat a příprava dat



## Kontingenční tabulka se dá vytvořit:

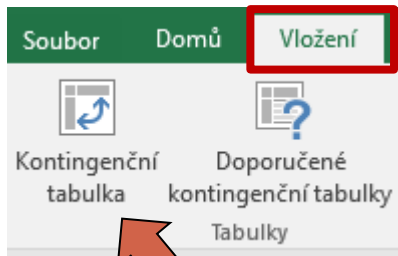
1. z tabulky v daném sešitě
2. z dat z jiného sešitu Excelu
3. z externích dat (např. MS Access)
4. ze sloučených dat z více oblastí - z různých listů nebo různých sešitů
5. z jiné kontingenční tabulky

## Data musí být uspořádána formou standardního databázového seznamu:

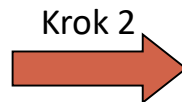
- V prvním řádku: názvy polí
- Další řádky: data

Vzhled tabulky: karta **Domů** → **Formátovat jako tabulku**

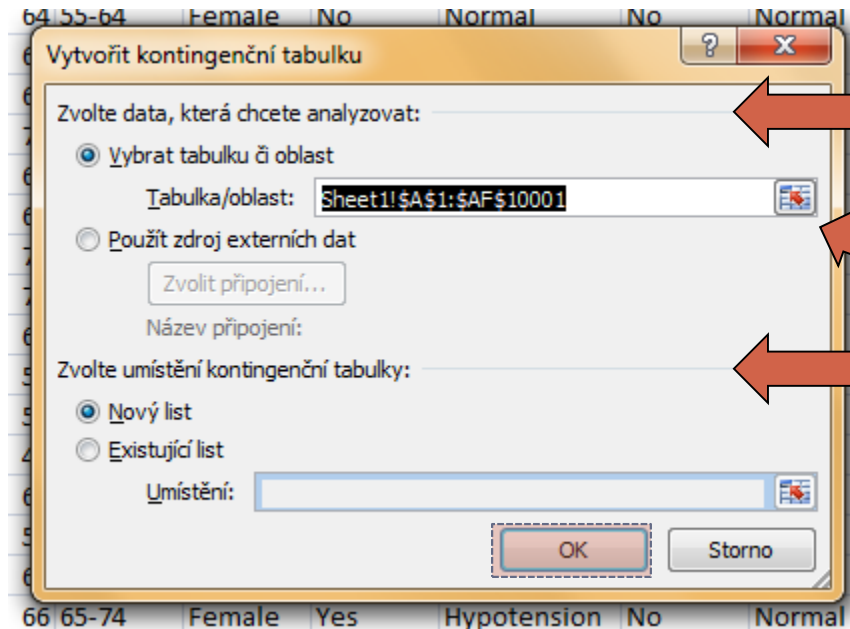
# Vytvoření kontingenční tabulky v Excelu



Krok 1



Krok 2



Zdroj dat (kromě Excelu i např. externí databáze)

Zdrojová oblast dat

Umístění tabulky

# Kontingenční tabulky – rozvržení

A3

Seznam polí kontingenční tabulky

Zvolte pole, které chcete přidat do sestavy:

- age
- agecat
- gender
- diabetes
- bp
- smoker
- choles
- active
- obesity
- angina
- mi
- nitro
- antictot

Přetáhnout pole mezi následujícími oblastmi:

- Filtr sestavy
- Popisky sloupců
- Popisky řádků
- Σ Hodnoty

Odložit aktualizaci rozlo... Aktualizovat

parametry, které je možné zobrazit v kontingenční tabulce

Seznam polí kontingenční tabulky

Zvolte pole, které chcete přidat do sestavy:

- age
- agecat
- gender
- diabetes
- bp
- smoker
- choles

Přetáhnout pole mezi následujícími oblastmi:

- Filtr sestavy
- Popisky sloupců
- Popisky řádků
- Σ Hodnoty

smoker

agecat

Počet z agecat

filtr

parametry ve sloupcích

parametry na řádcích

parametry dat

# Kontingenční tabulky – nastavení II.

Kontingenční tabulka

Počet z agecat	Popisky sloupců		
Popisky řádků	No	Yes	Celkový součet
45-54	1694	501	2195
55-64	3015	863	3878
65-74	2200	661	2861
75+	816	250	1066
<b>Celkový součet</b>	<b>7725</b>	<b>2275</b>	<b>10000</b>

Seznam polí kontingenční tabulky

Zvolte pole, které chcete přidat do sestavy:

- age
- agecat
- gender
- diabetes
- bp
- smoker
- choles

Přetáhnout pole mezi nás

Filtr sestavy

Popisky řádků

agecat

- Přesunout nahoru
- Přesunout dolů
- Přesunout na začátek
- Přesunout na konec
- Přejít k filtru sestavy
- Přejít k popiskům řádků
- Přejít k popiskům sloupců
- Přejít k hodnotám
- Odstranit pole
- Nastavení polí hodnot...**

Počet z agecat

Nastavení polí hodnot

Název zdroje: agecat

Vlastní název: Počet z agecat

Souhrn Zobrazit hodnoty jako

### Kritéria shrnutí pole hodnoty

Zvolte typ kalkulačky, který chcete použít pro shrnutí dat z vybraného pole:

- Součet
- Počet**
- Průměr
- Maximum
- Minimum
- Součin

Způsob  
sumarizace  
položky

Formát čísla

OK

Storno



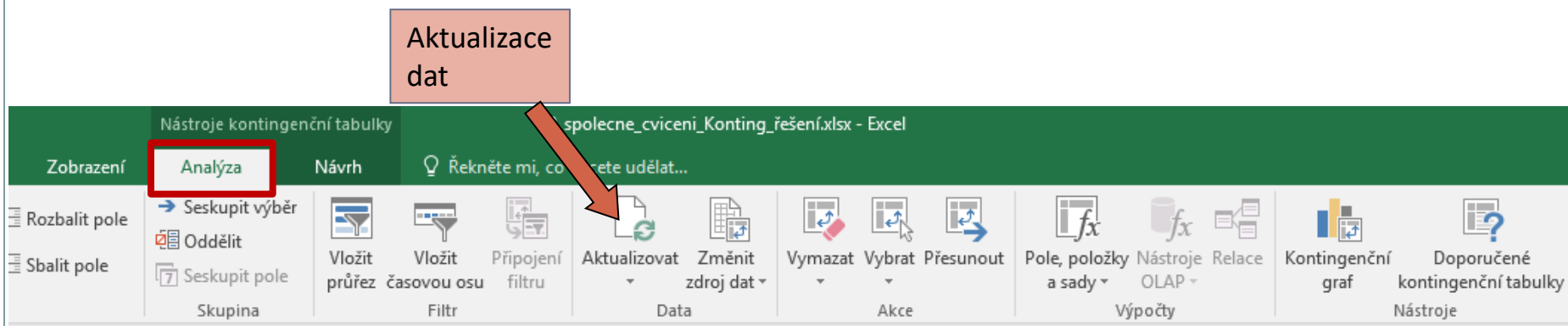
# Aktualizace dat v kontingenční tabulce



Při změně dat v tabulce se zdrojovými daty **nedojde** automaticky k aktualizaci dat v kontingenční tabulce.

**Musíte provést aktualizaci dat.**

1. Stůjíte kdekoliv v kontingenční tabulce
2. Na kartě **Analýza** ve skupině **Data** klikněte na **Aktualizovat** (Alt+F5), nebo na **Aktualizovat vše** (Ctrl+Alt+F5)



# Rozložení kontingenční tabulky



Po vytvoření se kontingenční tabulka zobrazí v tzv. **kompaktním formátu**. Lze ji zobrazit ale i ve formě **tabulky**, nebo ve formě **osnovy**.

1. Stůjte kdekoliv v kontingenční tabulce
2. Na kartě **Návrh** vyberte tlačítko **Rozložení sestavy** a volbu **Zobrazit ve formě osnovy nebo zobrazit ve formě tabulky**

**Kompaktní formát** - uspořádání tabulky aby zabírala co nejméně místa

**Forma osnovy** - řádková pole nižší úrovně je od vyšších úrovní odsazena, řádky nejsou odděleny čarami

**Forma tabulky** - klasická forma tabulky, pole nižší úrovně jsou v dalším sloupci

**Vyzkoušej!**

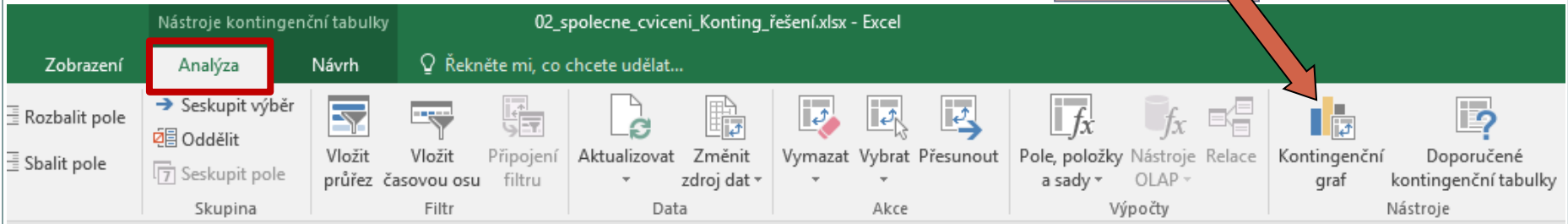
# Kontingenční graf



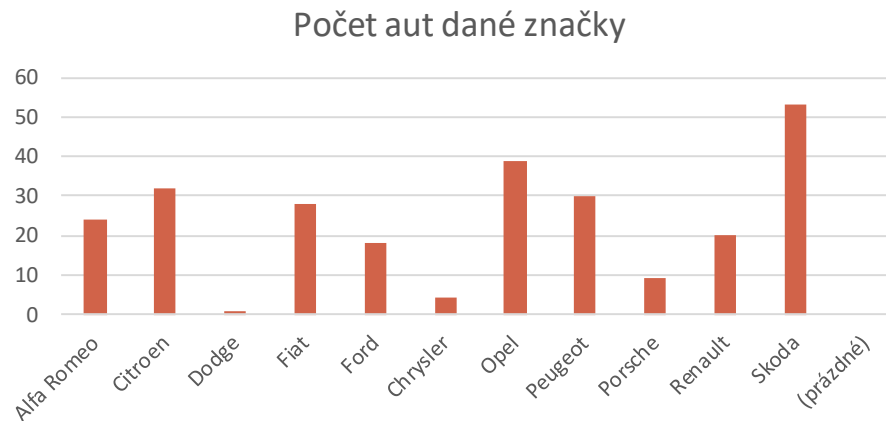
Data z kontingenční tabulky lze vizualizovat pomocí kontingenčního grafu.

**Návod:** Na kartě **Analýza** ve skupině **Data** klikněte na **Kontingenční graf**.

Kontingenční graf



Příklad kontingenčního grafu:



# II. Základy popisné statistiky



# Typy proměnných



## Kvalitativní (kategoriální) proměnná

- lze ji řadit do kategorií, ale nelze ji kvantifikovat

*Příklad: ??*

## Kvantitativní (numerická) proměnná

- můžeme jí přiřadit číselnou hodnotu

*Příklad: ??*

# Typy proměnných



## Kvalitativní (kategoriální) proměnná

- Ize ji řadit do kategorií, ale nelze ji kvantifikovat
- Příklady: *pohlaví, HIV status, barva vlasů ...*

## Kvantitativní (numerická) proměnná

- můžeme ji přiřadit číselnou hodnotu
- Příklady: *výška, váha, teplota, počet hospitalizací ...*

# Kvalitativní znaky



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

# Kvalitativní znaky



- **Binární znaky**: dvě kategorie, obvykle se kódují pomocí číslic 1 (přítomnost sledovaného znaku) a 0 (nepřítomnost sledovaného znaku).  
*Příklady: 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/O).*
- **Ordinální znaky**: několik kategorií, které lze vzájemně seřadit, tedy můžeme se ptát, která je větší/menší ( $1 < 2 < 3$ ).  
*Příklady: stupeň bolesti (mírná/střední/velká),  
stádium maligního onemocnění (I/II/III/IV).*



# Kvantitativní 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 <sup>1</sup>	Podíl <sup>1</sup>
1.	2 °C	-	-
2.	4 °C	+2	2x
3.	6 °C	+2	1.5x

<sup>1</sup>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

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

# 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**

# Charakteristiky polohy



## Charakteristiky polohy u nominálních znaků

- **Modus**: nejčastěji se vyskytující hodnota proměnné v souboru.

## Charakteristiky polohy u ordinálních znaků

- **$\alpha$ -kvantil**: je-li  $\alpha \in (0,1)$ , pak  $\alpha$ -kvantil  $x_\alpha$  je číslo, které rozděluje uspořádaný datový soubor na dolní úsek, obsahující aspoň podíl  $\alpha$  všech dat a na horní úsek obsahující aspoň podíl  $1-\alpha$  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 výsledků na dvě stejně početné poloviny.

# Charakteristiky polohy



## Charakteristiky polohy u intervalových a poměrových znaků

- **Aritmetický průměr:** je definován jako součet všech naměřených údajů vydělený jejich počtem,

$$E(x) = \bar{x} = \sum_{i=1}^n x_i / n \quad \text{kde } x_i \text{ jsou jednotlivé hodnoty a } n \text{ jejich počet}$$

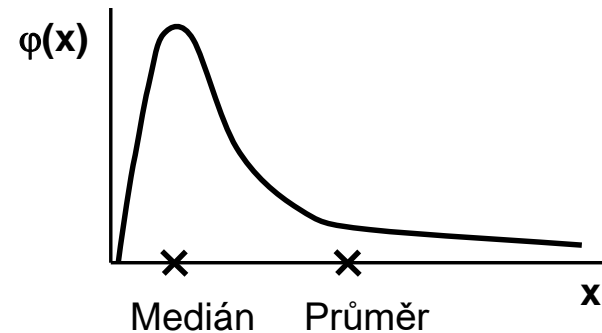
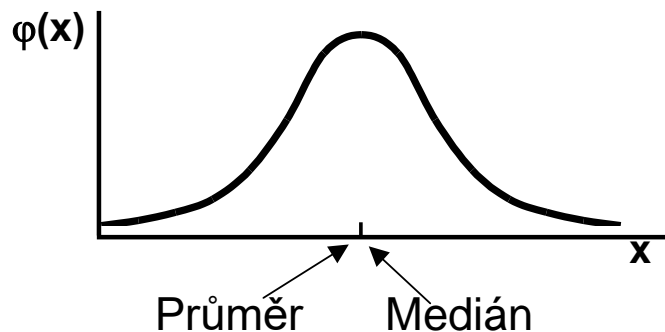
- **Geometrický průměr:**  $n$  kladných hodnot  $x_i$ ,  $\sqrt[n]{x_1 * \dots * x_n}$ , má smysl všude, kde má nějaký informační smysl součin hodnot proměnné. Z praktického hlediska platí, že logaritmus geometrického průměru je roven aritmetickému průměru logaritmovaných hodnot souboru.

# Průměr vs medián



## PAMATUJ:

- Průměr je silně ovlivněn extrémními hodnotami (tzv. odlehlá pozorování), medián není ovlivněn vybočujícími pozorováními
- Průměr je vhodný ukazatel středu u normálního/symetrického rozložení, medián je vhodnou charakteristikou středu souboru i v případě veličin s asymetrickým či neznámým rozdělením
- V případě symetrického rozložení jsou jejich hodnoty v podstatě shodné, v případě asymetrického rozložení však nikoliv!



# Charakteristiky variability



## Charakteristiky variability u ordinálních znaků

- **(Inter)kvartilové rozpětí (odchylka; IQR):**  $q = x_{0,75} - x_{0,25}$

## Charakteristiky variability u intervalových a poměrových znaků

- **Rozptyl (variance)** je ukazatelem šířky rozložení získaný na základě odchylky jednotlivých hodnot od průměru 
$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

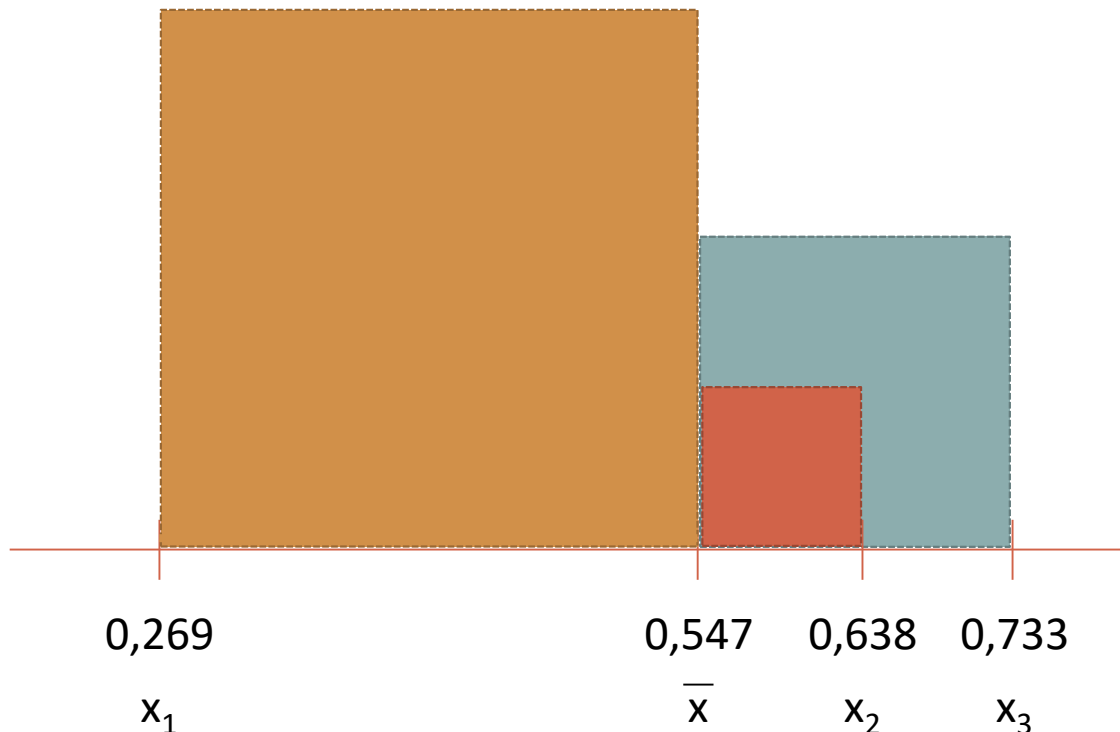
Jeho vypovídací schopnost nejvyšší v případě symetrického/normálního rozložení

- **Směrodatná odchylka** je druhá odmocnina z rozptylu
- **Koeficient variance** - podíl SD ku průměru, u poměrových znaků, umožňuje porovnat variabilitu několika znaků (vyjadřuje se v %)

# Výpočet rozptylu a směrodatné odchylky



- Příklad čtverců odchylek od průměru pro  $n = 3$ .
- Rozptyl je možno značně ovlivnit odlehlými pozorováními.



Rozptyl:

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

Směrodatná odchylka:

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

# Další parametry rozložení



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

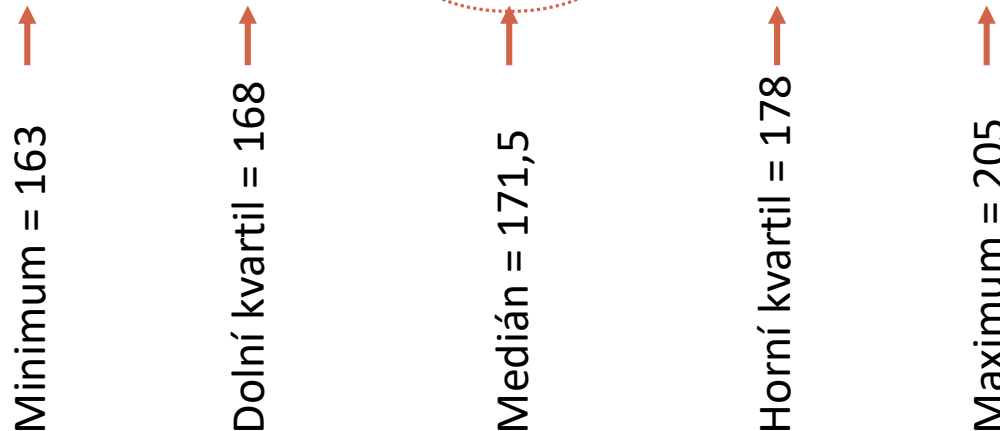


# Popisná sumarizace kvantitativních dat - příklad



Data – výška studentů (v cm): 178, 163, 205, 172, 168, 165, 179, 178, 169, 171

Seřazená data: 163, 165, 168, 169, 171, 172, 178, 178, 179, 205



(Inter)kvartilové rozpětí =  $178 - 168 = 10$

Variační rozpětí (rozsah) =  $205 - 163 = 42$

Průměr:  $\bar{x} = \sum_{i=1}^n x_i / n = (178+163+205+172+168+165+179+178+169+171)/10 = 174.8$

# Ukázka popisu a vizualizace kvalitativních dat



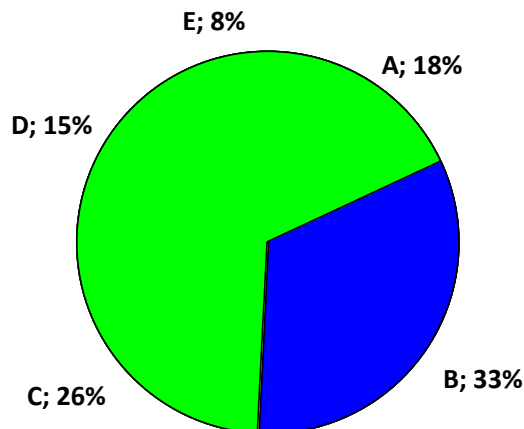
- **Popis kvalitativních dat:** frekvence jednotlivých kategorií
- **Vizualizace kvalitativních dat:** nejčastěji koláčový nebo sloupcový graf

## Příklad: Znáмка z biostatistiky (podzim 2014)

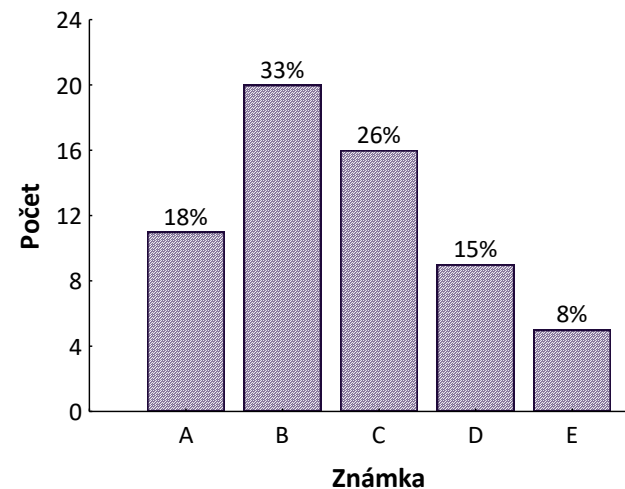
Frekvenční tabulka

Znáмка	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



# Ukázka popisu kvantitativních dat



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

## Příklad: Popis výšky (cm) pacientů

### Popisné statistiky

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



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

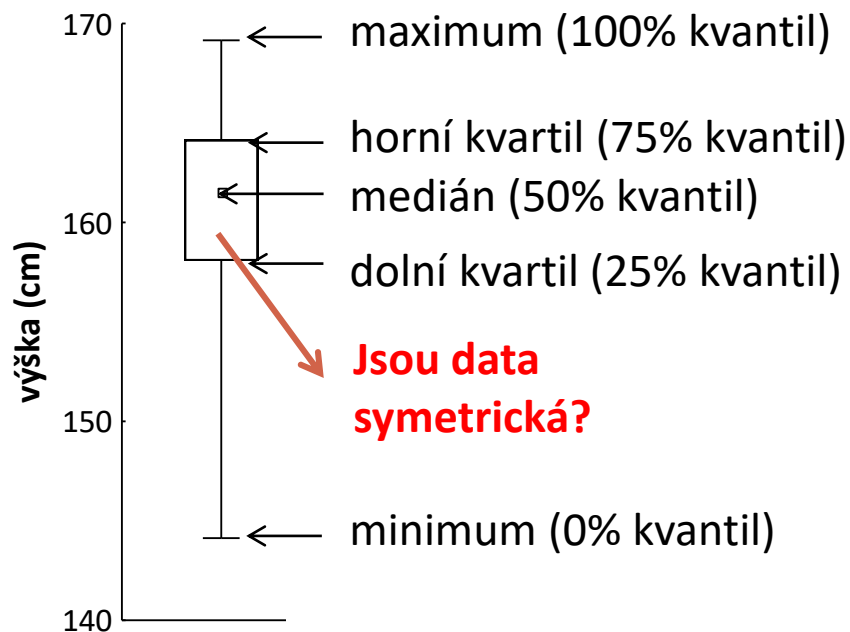
# Ukázka vizualizace kvantitativních dat



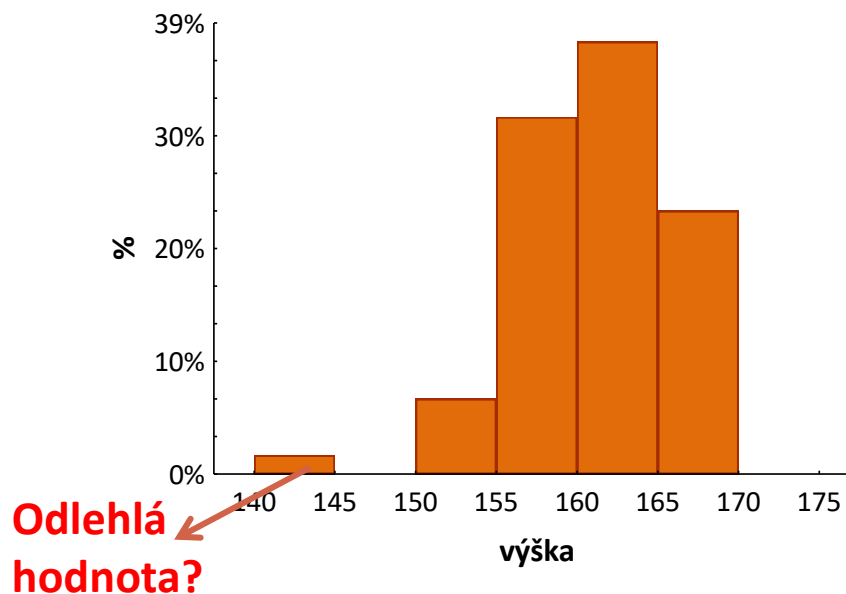
- **Vizualizace kvantitativních dat:** nejčastěji pomocí krabicového grafu nebo histogramu

## Příklad: Popis výšky (cm) pacientů

### Krabicový graf



### Histogram



# III. Cvičení v programu Statistica



**Základní popisné statistiky  
v programu Statistica**

**Datový soubor pacienti.sta**

**Datový soubor studenti.sta**

# Program Statistica



Jak získat program Statistica:

<https://inet.muni.cz>

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

V nabídce kliknout na: **Provozní služby – Software – Nabídka softwaru**

Nalézt: **Statistica 13.3** – kliknout **Získat**

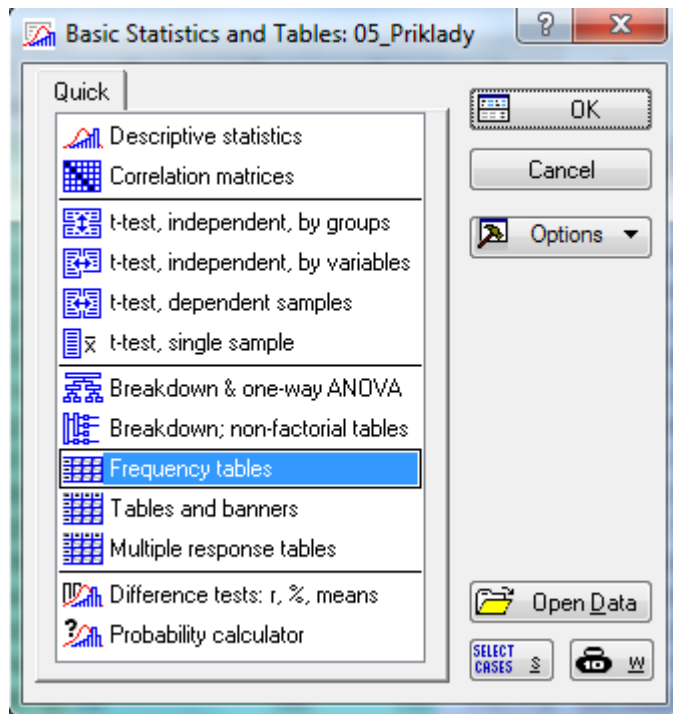
Postupovat dle návodu

# Základy popisné statistiky: soubor pacienti.sta



Načtěte soubor **pacienti.sta**, který obsahuje údaje o 61 pacientech.

- Nejprve budeme pracovat s kategoriální proměnnou.
- Pro proměnnou pohlaví zjistěte: absolutní, relativní četnost, dále absolutní a relativní kumulativní četnost



Postup: Statistics – Basic Statistics – Frequency tables

Category	Count	Cumulative Count	Percent	Cumulative Percent
muz	31	31	50,81967	50,8197
zena	30	61	49,18033	100,0000
Missing	0	61	0,00000	100,0000

# Základy popisné statistiky: soubor patienti.sta

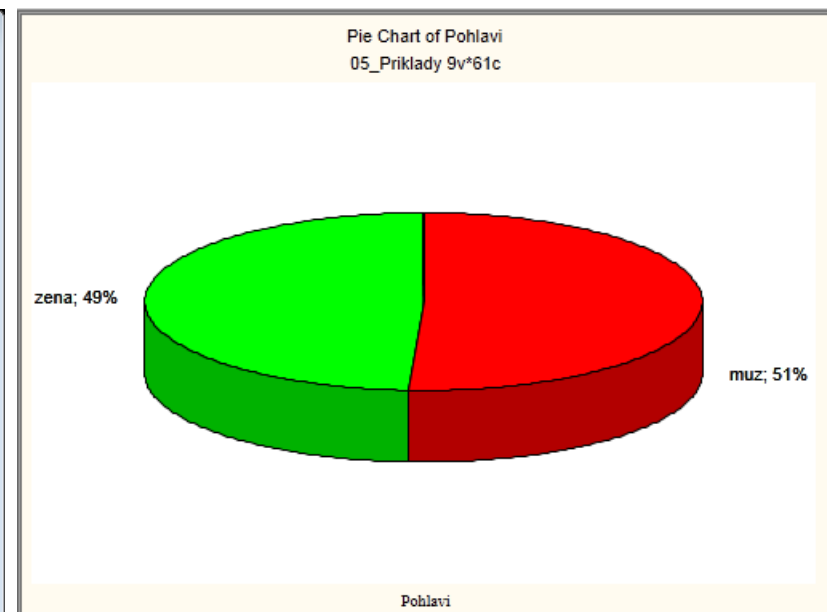


- ***Pomocí výsečového grafu (koláčového grafu) znázorníte proměnnou Pohlaví, doplňte procenta (relativní četnost).***

Postup vytvoření grafu: Graphs – 2D – Pie Charts...

Postup přidání legendy: na záložce Advanced kliknout na „Text and Percent“

The screenshot shows the 'Pie Charts' dialog box in SPSS, with the 'Advanced' tab selected. The 'Graph type' section shows 'Pie Chart - Values' selected for the variable 'Pohlavi'. In the 'Pie legend' section, the 'Text and Percent' option is highlighted with a red arrow. Other options include 'Off', 'Percent', 'Values', 'Text Labels', and 'Text and Value'. The 'Type' section has '3D' selected, and the 'Shape' section has 'Ellipse' selected. The 'Pie labels (values)' section has 'Spreadsheet' selected. The 'Explode slice' checkbox is checked with a value of 1.

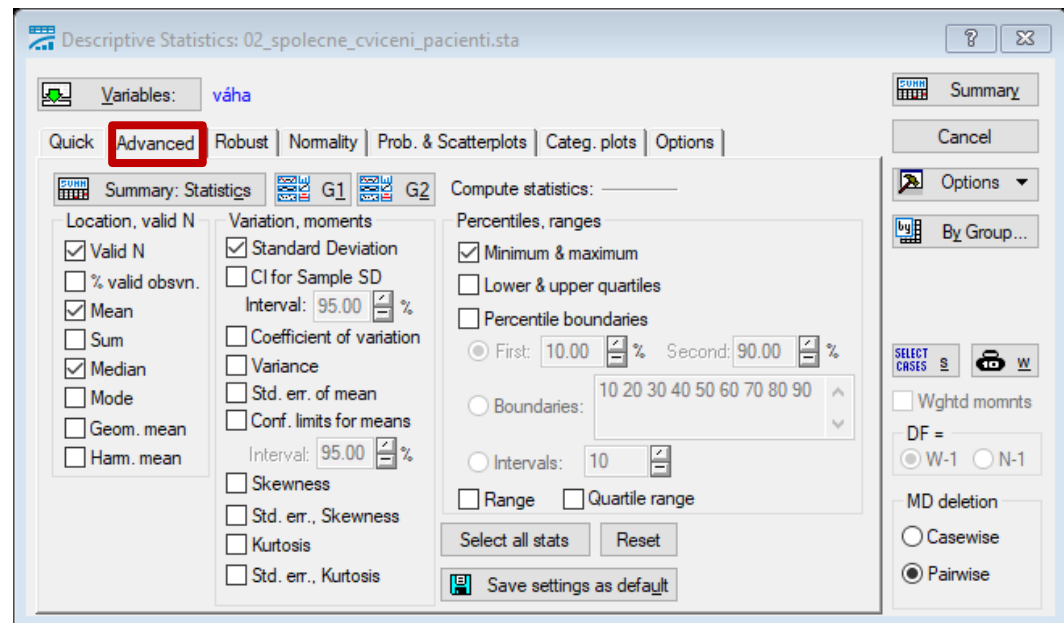
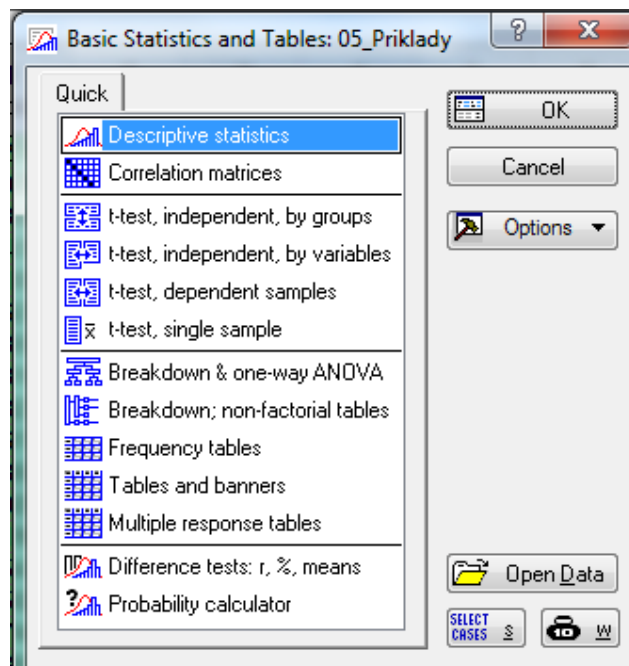




# Základy popisné statistiky: soubor pacienti.sta

- **Nyní budeme pracovat se spojitou proměnnou.**
- **Pro proměnnou váha zjistěte: průměr, medián, minimum, maximum a směrodatnou odchylku**

Postup: Statistics – Basic Statistics  
– Descriptive statistics

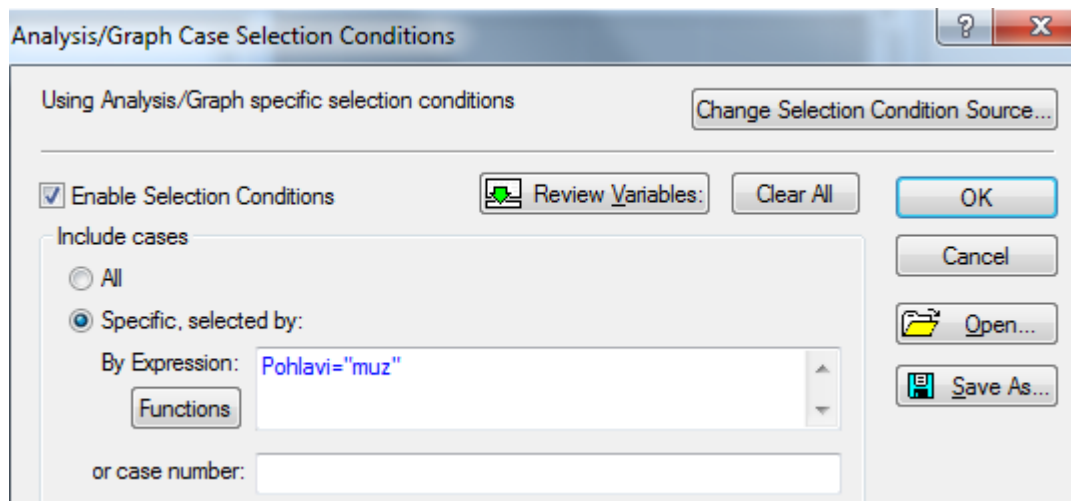


Variable	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.
váha	61	65,63968	66,49219	49,80155	79,20183	4,988461

# Základy popisné statistiky: soubor pacienti.sta



- ***Pokud bychom chtěli zjistit průměrnou váhu pouze u mužů, klikneme na tlačítko SELECT CASES a zvolíme Pohlaví="muz" (nezapomínejte na uvozovky)***



Descriptive Statistics (05_Priklady)					
Include condition: Pohlaví="muz"					
Variable	Valid N	Mean	Minimum	Maximum	Std.Dev.
váha	31	65,37337	49,80155	72,14984	5,034108

# Základy popisné statistiky: soubor pacienti.sta



- **Vytvořte histogram s rozpětím hodnot po pěti, poté zkuste to samé pro muže a ženy.**

Postup: Záložka Graphs -> Histogram -> proměnná váha, záložka Advanced: zatrhnout Boundaries -> Specify Boundaries

2D Histograms

Quick: **Advanced** Appearance Categorized Options 1 Options 2

Variables: Variable: váha

Graph type: Regular

Fit type: Normal

Showing Type: Standard

Intervals: Variable: váha

Number of Intervals: 10

Boundaries: none

Specify Boundaries

Specify Boundaries for váha

Specify boundary range:

Minimum: 45

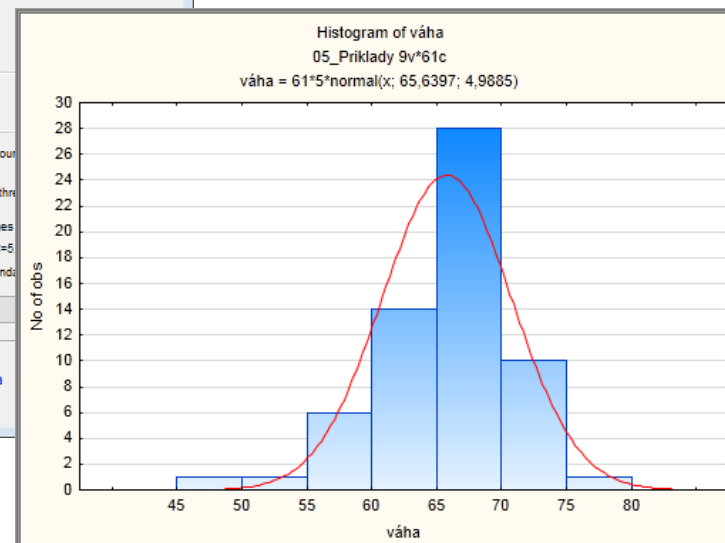
Interval Step: 5

Maximum: 80

Enter Boundaries:

Definition: Upper boundaries

Variables: váha



# Základy popisné statistiky: soubor pacienti.sta



- ***Pokud chceme váhu odděleně pro pohlaví - po boku vpravo By group: vybereme proměnnou pohlaví.***

The image shows two overlapping dialog boxes from the SPSS software. The background dialog is titled "2D Histograms" and has tabs for "Quick", "Advanced", "Appearance", "Categorized", "Options 1", and "Options 2". The "Quick" tab is active. Under "Variables:", the variable "váha" is selected. Under "Graph type:", "Regular" is selected. Under "Fit type:", "Normal" is checked. The "Intervals" section has "Integer mode" selected with "Auto" checked, "Unique values" selected, "Unsorted" selected, "Categories" set to 10, and "Codes" set to "none". The "By Group" sub-dialog box is in the foreground, showing "Grouping Variable(s)..." set to "Pohlaví". It has "Enabled" checked, "Output to single folder" unchecked, "Label Outputs" checked, "Output 'All Groups' results" checked, and "Accumulate tabular results in a single spreadsheet" unchecked. The "Sorting of Groups" section has "Unsorted" selected, "Ascending" and "Descending" unselected. Both dialog boxes have "OK" and "Cancel" buttons.

# Základy popisné statistiky: soubor pacienti.sta



- Pokud chceme histogram váhy pro muže i ženy mít v jenom grafu: vybereme záložku Categorized, zapneme kategorii X a změníme proměnnou na pohlaví.***

2D Histograms

Quick Advanced Appearance **Categorized** Options 1 Options 2

X-Categories

On

Variable: Pohlavi

Integer mode  Auto

Unique values

Unsorted  Asc  Desc

Categories: 10

Boundaries: none

Codes: none

Multiple subsets

Change Variable

Layout

Separate

Overlaid

Y-Categories

On

Variable: none

Integer mode  Auto

Unique values

Unsorted  Asc  Desc

Categories: 10

Boundaries: none

Codes: none

Multiple subsets

Change Variable

OK Cancel

Options

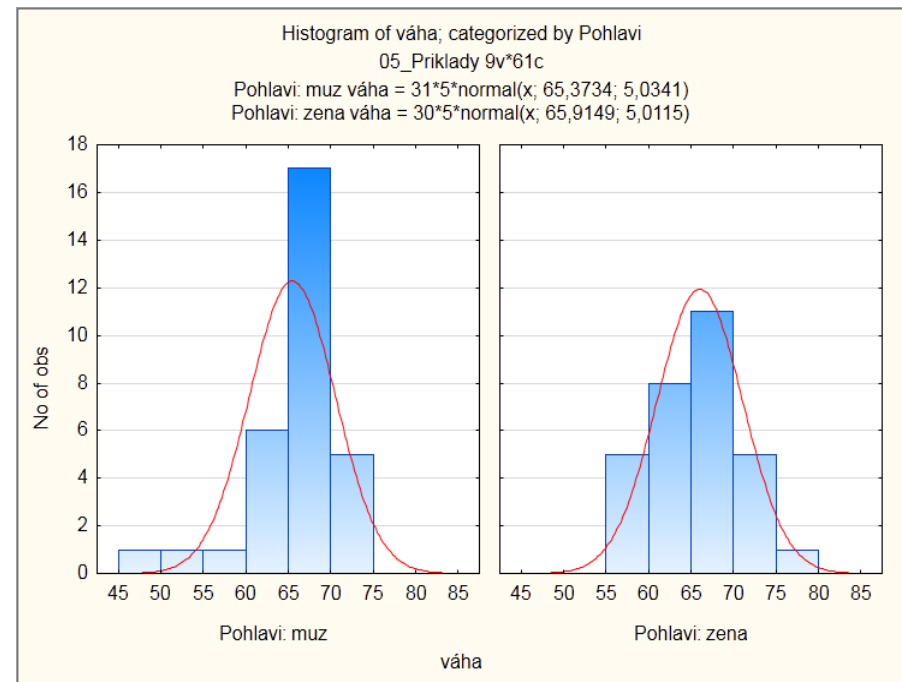
By Group

SELECT CASES Sel Cond

Case Weights

Graphs Gallery

Updating: Auto



# Základy popisné statistiky: soubor pacienti.sta



- **Překódování proměnné:**
- Proměnnou váha překódujete do proměnné vaha\_kategorie tak, aby pacienti pod 60 kg tvořili jednu skupinu a pacienti 60+ druhou skupinu.

Návod: Vložíme novou proměnnou vaha\_kategorie za proměnnou váha. Označíme novou proměnnou vaha\_kategorie, záložka Data -> Recode

Recode Values of Variable 10: vaha\_kategorie

Category	Include If:	New Value
Category 1	váha < 60	podváha
Category 2	váha >= 60	normální váha
Category 3		

- **Zjistěte, kolik % žen mělo váhu pod 60 kg?**

# Samostatné cvičení: soubor studenti.sta



**Načtěte soubor studenti.sta, který obsahuje údaje o 26 studentech, získané informace jsou shrnuty v proměnných A,B,C,D.**

Návod: Záložka *Home* → *Open* → vybereme soubor studenti.sta.

**Změňte názvy proměnných: A-jméno studenta, B-známka z biostatistiky, C-pohlaví, D-věk. U proměnných B a C popište jednotlivé varianty (proměnná B odpovídá známce: 1- výborně, 2- velmi dobře, 3- dobře, 4- nedostatečně; proměnná C odpovídá pohlaví:1 - muž, 2 - žena)**

Návod: Vybereme nejprve příslušnou proměnnou A, 2krát klikneme myší → do položky *Name* napíšeme nový název proměnné (*All Specs...* umožní přejmenovat všechny proměnné najednou; *Text Labels* číselným hodnotám přiřadí textový popisek).

**Pojmenujte názvy řádků tabulky jmény studentů, poté proměnnou jméno studenta smažte.**

Návod: Záložka *Data* → *Names* → *Transfer case names from* → *Variable*: Jméno studenta;

smazání-vybereme proměnnou Jméno studenta, pravé tlačítko myši → *Delete Variable*.

# Samostatné cvičení: soubor studenti.sta



**U proměnné Známka zjistěte absolutní, relativní četnost, dále absolutní a relativní kumulativní četnost.**

Návod: Záložka *Statistics* → *Basic Statistics* → *Frequency tables* → *Variables: známka* z biostatistiky → *Summary*

**Zjistěte průměr, medián pro proměnnou Věk. U proměnné pohlaví zjistěte modus.  
Pro proměnnou známka zjistěte medián, modus.**

Návod:

Způsob 1: Označíme proměnnou věk, pravé tlačítko → *Statistics of Block Data* → *Blocks columns* → *All*

Zbůsob 2: Záložka *Statistics* → *Basic Statistics* → *Descriptive statistics* → *Variables: věk* → záložka *Advanced* → vybereme *Mean, Median*.



# Samostatné cvičení: soubor studenti.sta



Proměnnou věk překódujte pomocí následujících 5 intervalů:  $\langle 20,22 \rangle$ ,  $(22,25 \rangle$ ,  $(25,28 \rangle$ ,  $(28,31 \rangle$ ,  $(31,33 \rangle$  do proměnné Věk 2.

Návod: Vložíme novou proměnnou Věk 2 za proměnnou Věk. Označíme novou proměnnou

Věk 2, záložka *Data* → *Recode* → *Category 1: věk*  $\geq 20$  and  $\leq 22$ , *New Value: 1* atd.

Pomocí koláčového grafu znázorněte proměnnou Známku a Pohlaví, doplňte procenta (relativní četnost).

Návod: Záložka *Graphs* → *2D* → *Pie Charts* → Záložka: *Quick: Variables: Známka, Pohlaví*; Záložka: *Advanced* → *Pie legends* vyber *Text and Percent*.

Pomocí sloupcového grafu znázorněte věk pouze pro muže.

Návod: Záložka *Graphs* → *2D* → *Bar/Column Plots* → *Variables: Věk*, v tomtéž okně napravo klikneme na *Select Cases* → zaškrtneme možnost *Enable Selection Conditions* → *Specific* → *selected by Expression: Pohlaví=1*.

# Samostatné cvičení: soubor studenti.sta



**Pro proměnnou Věk vytvořte histogram s intervaly širokými dva roky, poté zkuste to samé zvlášť pro muže a ženy.**

Návod: Záložka *Graphs* → *Histogram* → *Variables: věk*, záložka *Advanced: Intervals Boundaries* → *Specifies boundaries* po boku vpravo *By group*: vybereme proměnnou pohlaví