

Metodologie pro ISK II

# Analýza kardinálních proměnných

# Kardinální proměnné

- Všechny hodnoty z daného intervalu
- Zjišťujeme:
  - Centrální míry
  - Variabilitu
  - Šikmost, špičatost
  - Percentily (decily, kvantily...)
- Zobrazení: **histogram**

# Míry centrální tendence

**MODUS** je hodnota, která se v datech vyskytuje nejčastěji.

**MEDIÁN** dělí řadu výsledků seřazených podle velikosti na dvě stejně početné poloviny

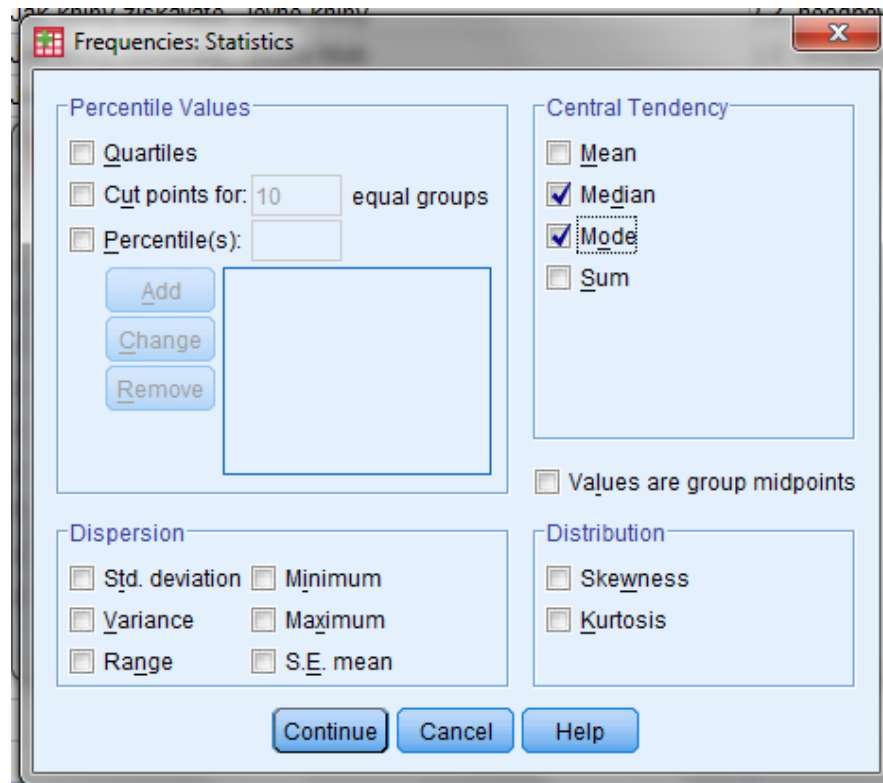
**MEDIÁNOVÁ KATEGORIE** je ta, ve které je dosaženo 50% všech údajů, postupujeme-li od první kategorie výše.

**ARITMETICKÝ PRŮMĚR** je součet všech hodnot vydělený jejich počtem

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$$

# Míry centrální tendence

- Analyze → Descriptive Statistics → Frequencies → **Statistics → Mean, Median, Mode**



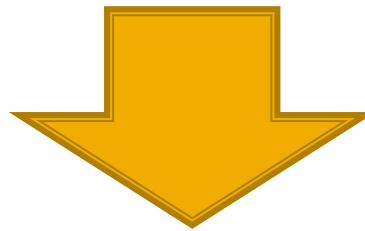
# Aritmetický průměr

- **V roce 2011** dosáhla v ČR průměrná mzda výše **24 319 Kč**. *Můžeme z toho usuzovat, že typický Čech má plat 24 319 Kč?*



# Aritmetický průměr

- **V roce 2011** dosáhla v ČR průměrná mzda výše **24 319 Kč**. *Můžeme z toho usuzovat, že typický Čech má plat 24 319 Kč?*



Průměrný měsíční plat (v Kč)	Medián (Kč)	Rozdíl (v %)
23 300	21 000	11%

Zdroj: [Platy.cz](http://Platy.cz)

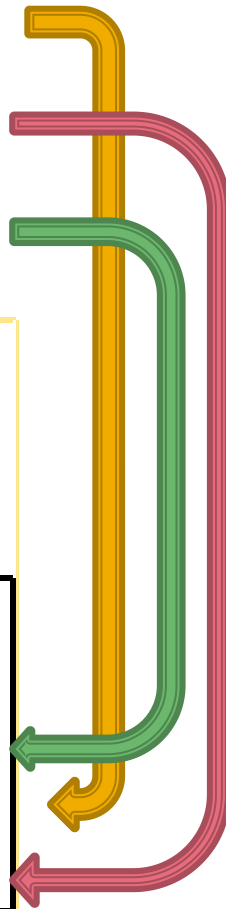
# Co všechno zjišťujeme?

- MINIMUM
- MAXIMUM
- ROZPĚTÍ

## Statistics

Pokuste se prosím odhadnout, kolik knih (odborných, vědeckých i beletrie) jste přečetl za rok 2011.

N	Valid	613
	Missing	0
Range		300
Minimum		0
Maximum		300



Analyze – Frequencies -  
Statistics

The screenshot shows the 'Frequencies: Statistics' dialog box. The 'Dispersion' section has the following options checked:  Range,  Minimum, and  Maximum. The 'Central Tendency' section has the following options checked:  Mean,  Median,  Mode, and  Sum. The 'Percentile Values' section has the following options checked:  Quartiles,  Cut points for: 10 groups, and  Percentile(s):. The 'Continue', 'Cancel', and 'Help' buttons are visible at the bottom.

MIN: minimální  
hodnota  
MAX: maximální  
hodnota  
RANGE: rozpětí  
(maximální minus  
minimální  
hodnota)

# Rozptyl

- Rozptyl je definován jako **střední hodnota kvadrátů odchylek od střední hodnoty** (průměru).
- vyjadřuje variabilitu rozdělení souboru náhodných hodnot kolem její střední hodnoty.



# Směrodatná odchylka

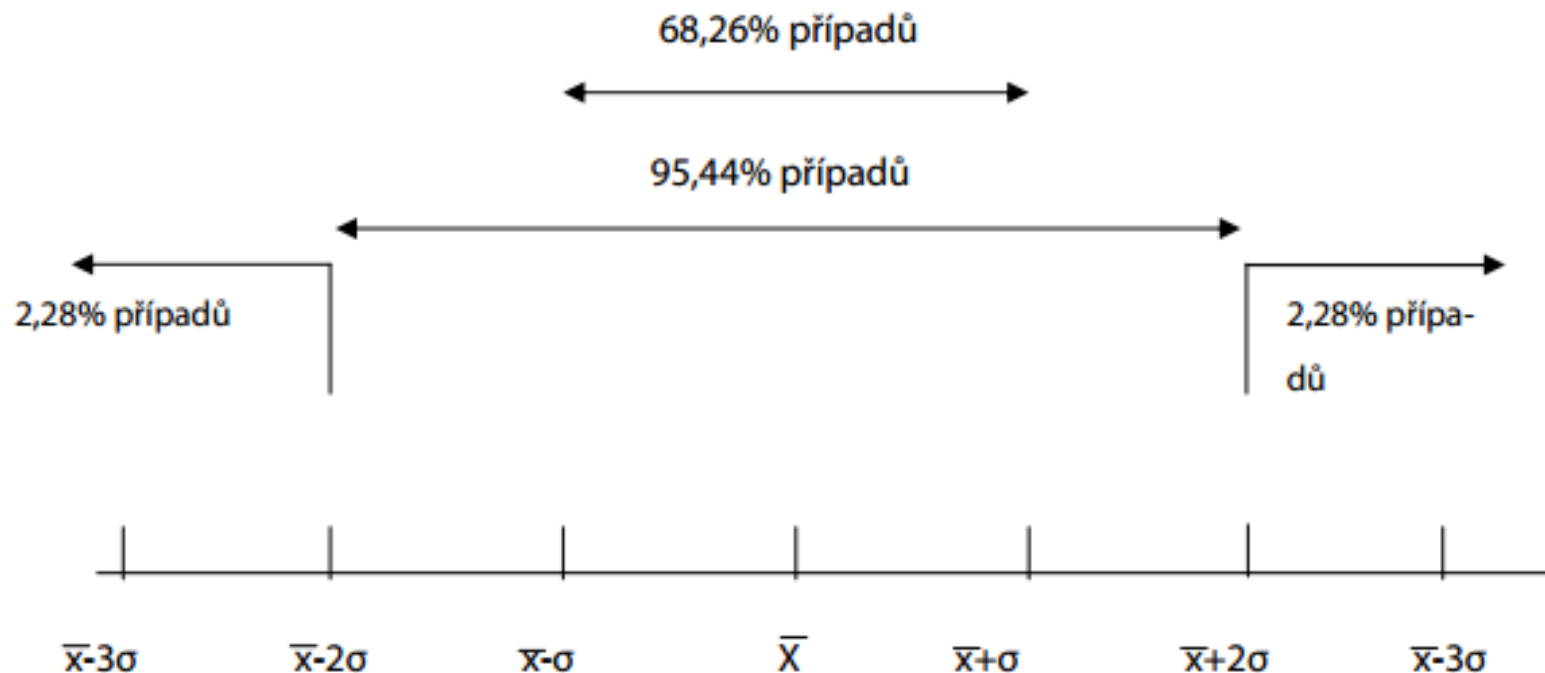
- **SMĚRODATNÁ ODCHYLKA** je druhá odmocnina rozptylu, a v podstatě nám říká, uvnitř jakého intervalu okolo průměru leží zvolené procento případů
- Čím je směrodatná odchylka menší, tím lépe pro aritmetický průměr

# Směrodatná odchylka (příklady)

Zajímavost předmětu	není vůbec zajímavý	.***X(*)**... je velmi zajímavý
Přínosnost předmětu	není vůbec přínosné	***X(*)*... je velmi přínosné
Obtížnost obsahu	velmi snadný	.....(*)**X** velmi obtížný
Náročnost na přípravu	velmi snadný	.....(*)X*... velmi obtížný
Dostupnost studijních zdrojů	velmi špatně dostupné	.....(*)**X* velmi dobře dostupné
Jak učitel učí	velmi špatný	.***X(*)**... vynikající
Učitel jako odborník	není odborníkem	.....(*)***X* je odborníkem

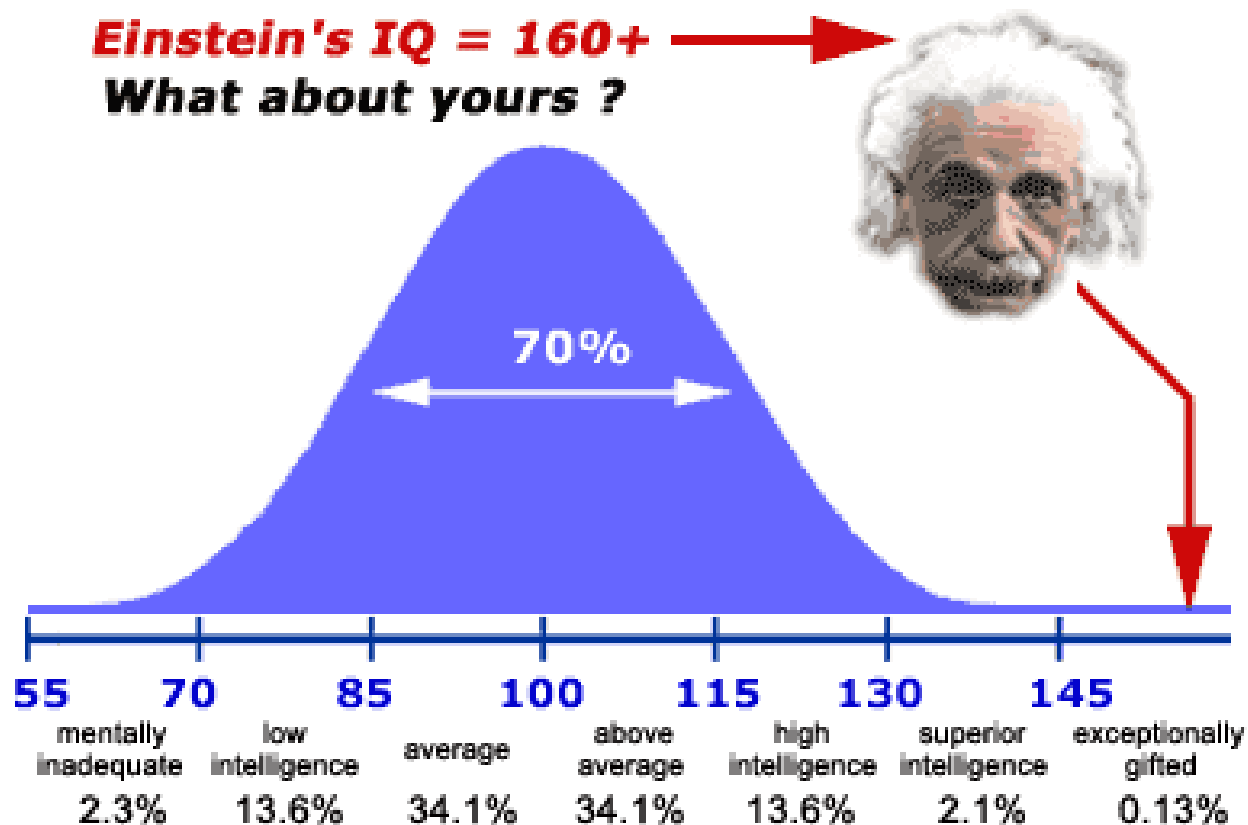
Zajímavost předmětu	není vůbec zajímavý	.....(*)...*X je velmi zajímavý
Přínosnost předmětu	není vůbec přínosné	.....(*)...*X* je velmi přínosné
Obtížnost obsahu	velmi snadný	**X**(.)..... velmi obtížný
Náročnost na přípravu	velmi snadný	*X**(.)..... velmi obtížný
Dostupnost studijních zdrojů	velmi špatně dostupné	.....(*)**X** velmi dobře dostupné
Jak učitel učí	velmi špatný	.....(*)...*X vynikající
Učitel jako odborník	není odborníkem	.....(*)...X je odborníkem

# Směrodatná odchylka



Zdroj: Skripta ke kurzu Statistická analýza dat, Petr Mareš & Ladislav Rabušic 2003.

# Směrodatná odchylka (příklad)



Zdroj: <http://wilderdom.com/intelligence/IQUnderstandingInterpreting.html>

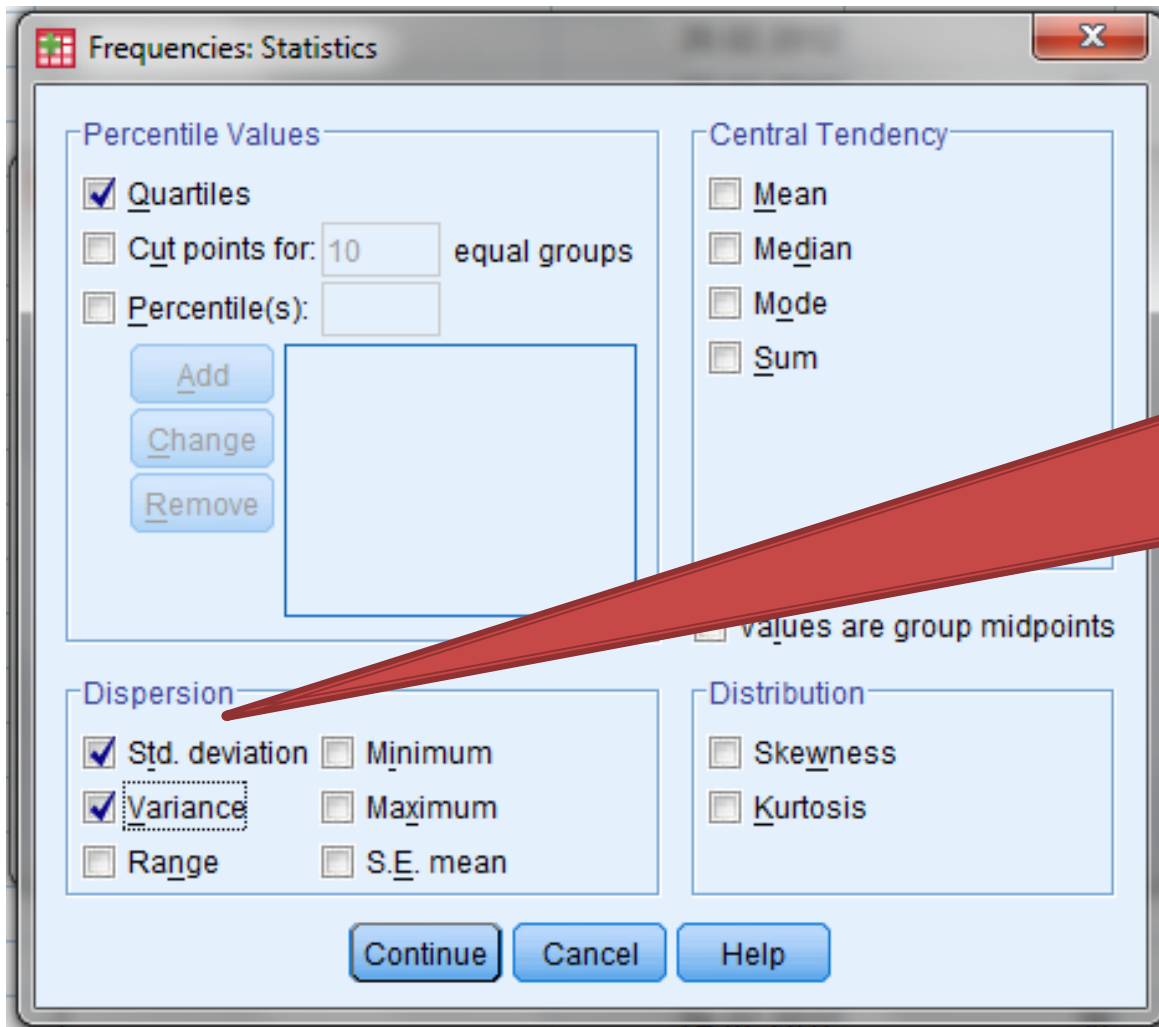
# Rozptyl, směrodatná odchylka

## Koeficient variace

- Rozptyl a směrodatná odchylka mají vždy stejnou jednotku jako daná proměnná (věk v letech, výška v cm, plat v EUR...)
- Chceme-li porovnávat mezi sebou různé proměnné (s různými jednotkami), vypočítáme **variační koeficient**

$$v_x = \frac{s_x}{\bar{x}} \cdot 100 [\%]$$

# Směrodatná odchylka, rozptyl

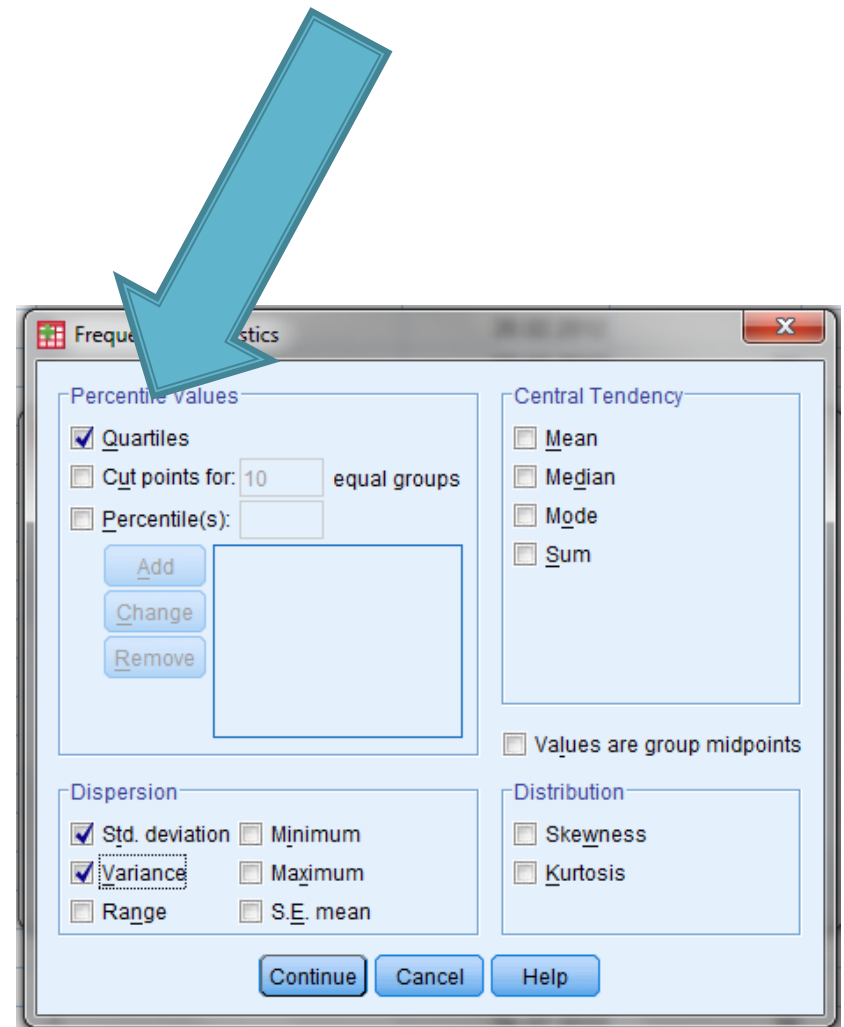


Analyze -  
Frequencies -  
Statistics

# Percentily

Nejčastěji používané:

- **MEDIÁN** (x50)
- **KVARTILY** (x25, x50, x75)
- **DECILY** (x10, x20, x30, x40, x50, x60, x70, x80, x90)



# Percentily

**Jak vysoký je medián proti průměrné mzdě? (ve vybraných zemích OECD)**

Země	spodních 10 %	medián	horních 10 %
Švédsko	56 %	89,8 %	150,9 %
Finsko	62,3 %	89,5 %	147,9 %
Kanada	44,6 %	89,1 %	166,9 %
Dánsko	60,9 %	89 %	150,4 %
Norsko	63,2 %	88,9 %	149 %
Japonsko	52,4 %	87,6 %	162,7 %
Nový Zéland	51,2 %	87,2 %	160,6 %
Německo	43,4 %	87 %	165,7 %
<b>Česko</b>	<b>49,3 %</b>	<b>85,2 %</b>	<b>153,1 %</b>
Itálie	56,1 %	85,1 %	156,6 %
Švýcarsko	56,6 %	84,9 %	153,4 %
Belgie	60,4 %	84,5 %	153,4 %
Nizozemí	51,7 %	84 %	158,8 %

<http://finexpert.e15.cz/jak-se-lisi-prumerna-mzda-a-median>



# Percentily (příklad)

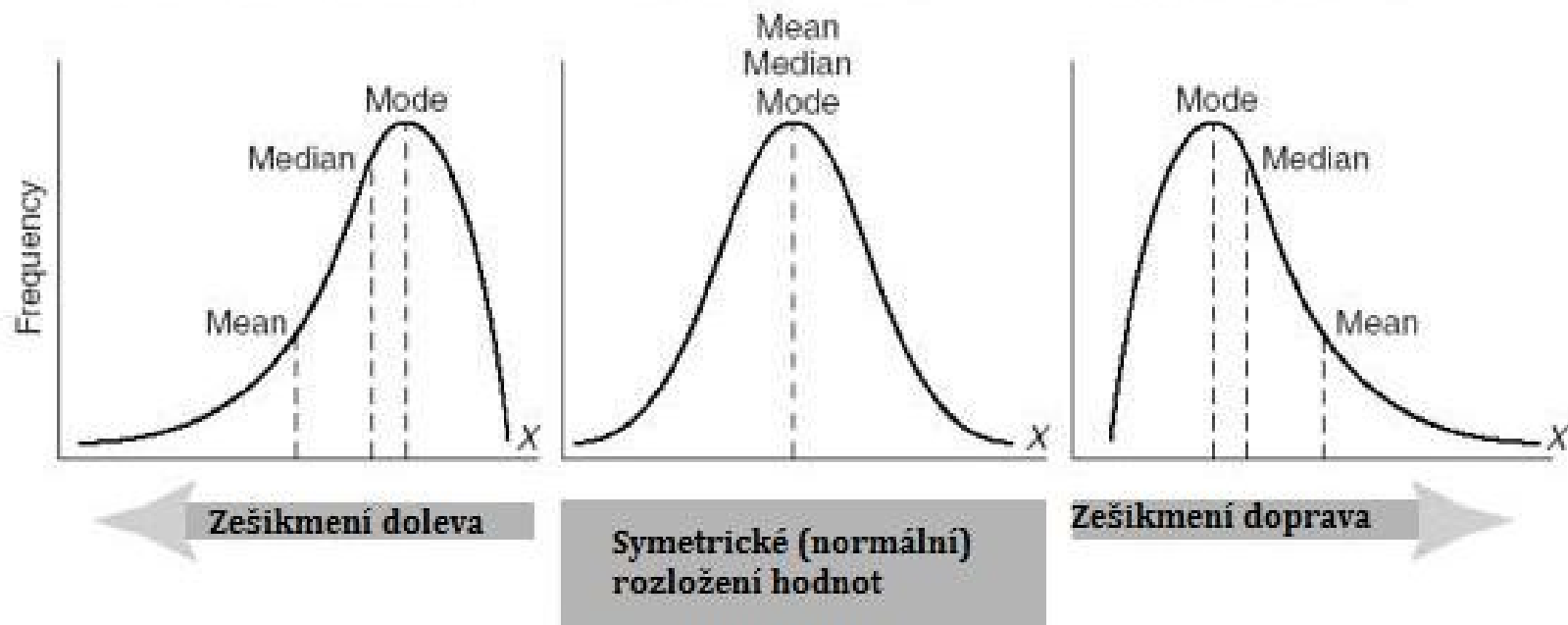
## Statistics

Pokuste se prosím odhadnout, kolik knih (odborných, vědeckých i beletrie) jste přečetl za rok 2011.

N	Valid	613
	Missing	0
Std. Deviation		25,514
Variance		650,939
Percentiles	10	1,00
	20	3,00
	25	4,00
	30	5,00
	40	8,00
	50	10,00
	60	14,00
	70	20,00
	75	20,00
	80	25,00
90	43,60	

- „spodních“ 10 % čtenářů čte jen jednu knihu ročně
- polovina čtenářů přečte do 10 knih ročně
- tři čtvrtiny čtenářů přečte do 20 knih ročně

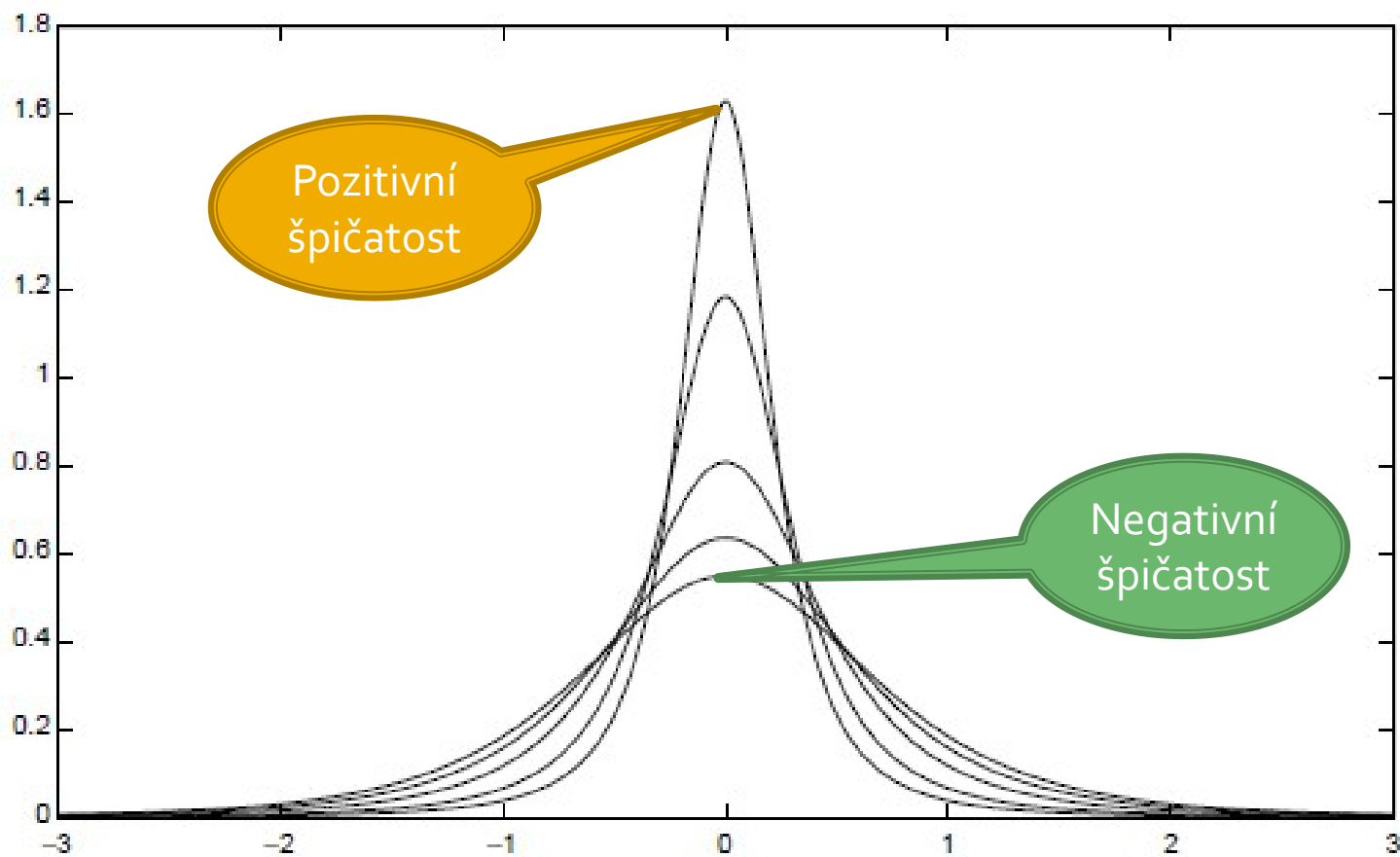
# Šikmost



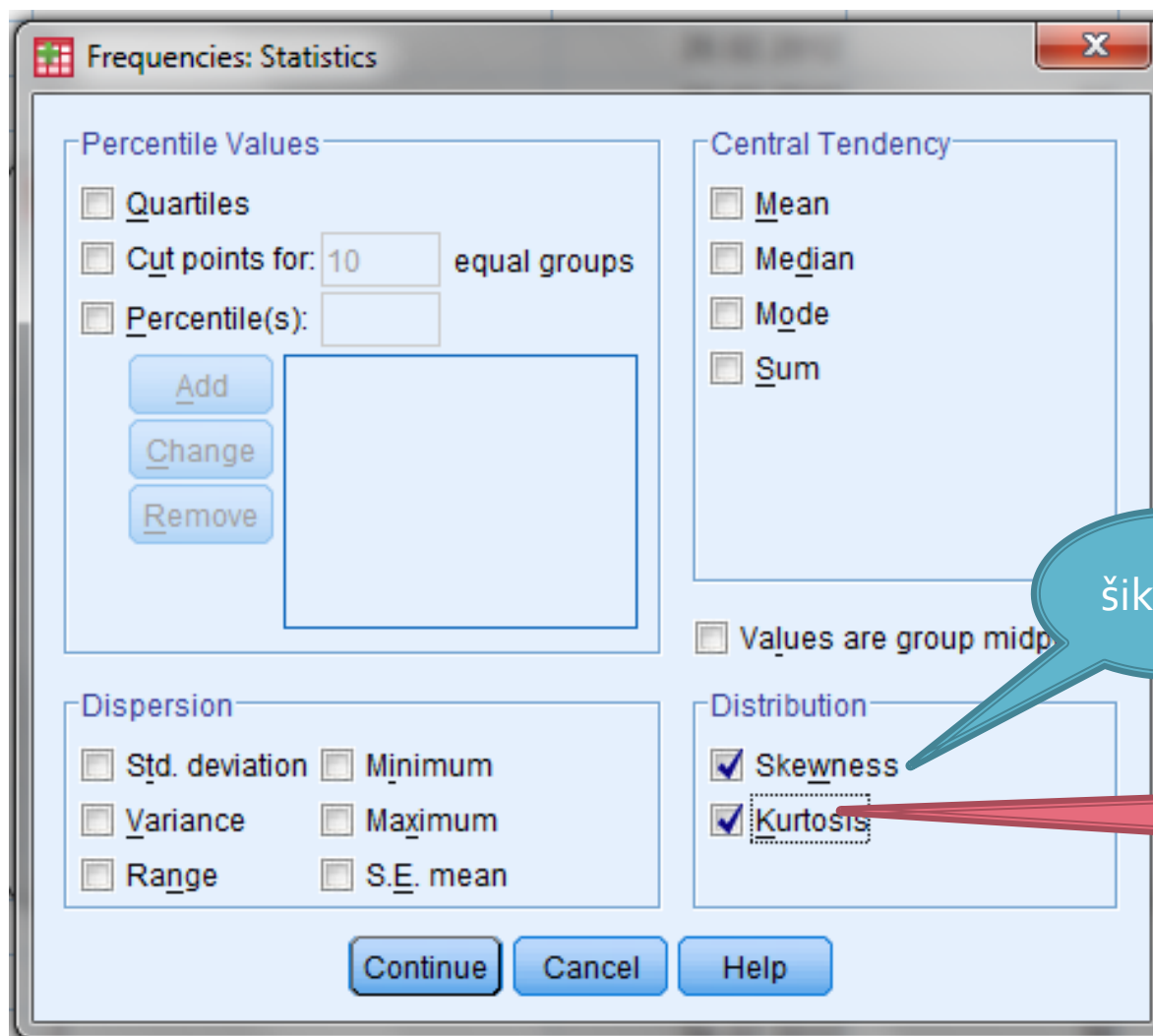
# Šikmost

- **Symetrické (normální) rozložení** - aritmetický průměr, medián a modus mají stejné nebo velmi podobné hodnoty. (o)
- Pokud je aritmetický průměr větší než medián, který je zase větší než modus, znamená to, že je více případů menších než průměr a naše **rozložení je šikmé doprava**. (+)
- Třetí možností je, že je více případů větších než aritmetický průměr. Ten je pak menší než medián a ten je menší než modus. Naše **rozložení je šikmé doleva**. (-)

# Špičatost



# Šikmost a špičatost

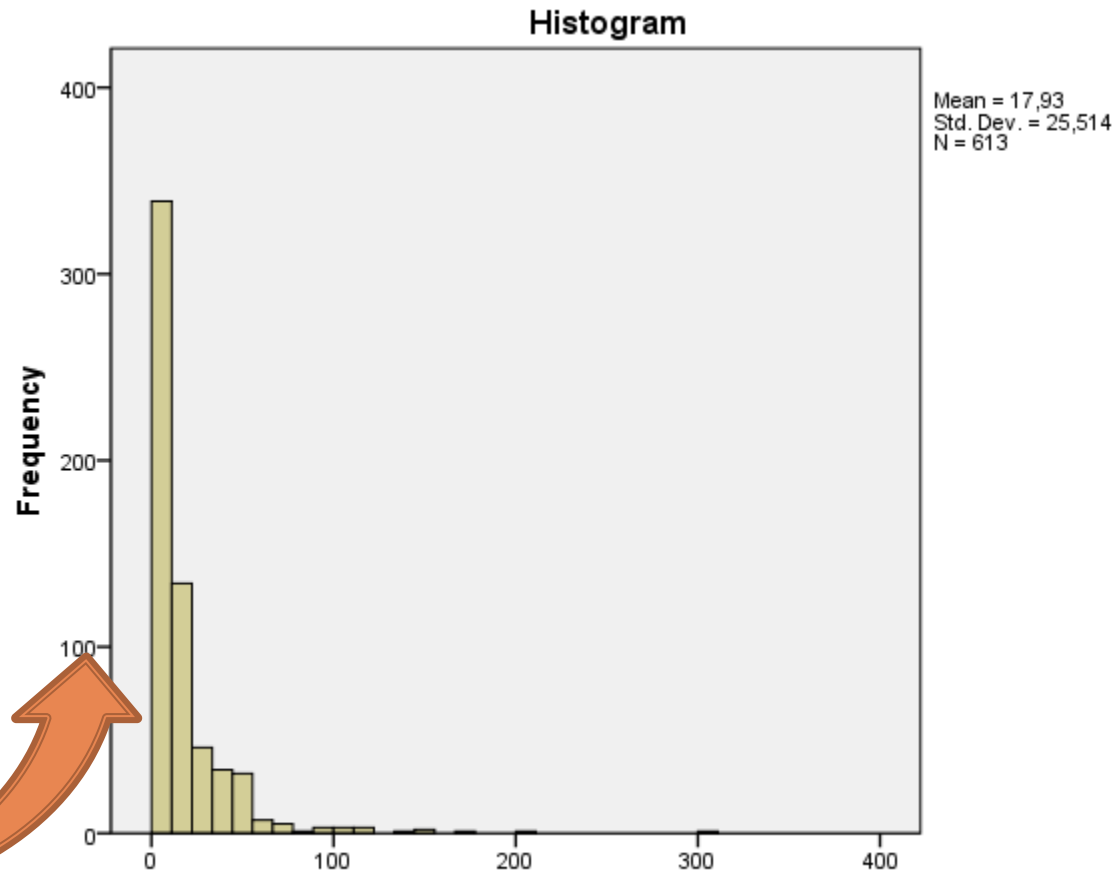


# Šikmost a špičatost (příklad)

## Statistics

Pokuste se prosím odhadnout, kolik knih (odborných, vědeckých i beletrie) jste přečetli za rok 2011.

N	Valid	613
	Missing	0
	Skewness	4,572
	Std. Error of Skewness	,099
	Kurtosis	33,731
	Std. Error of Kurtosis	,197



Pokuste se prosím odhadnout, kolik knih (odborných, vědeckých i beletrie) jste přečetli za rok 2011.

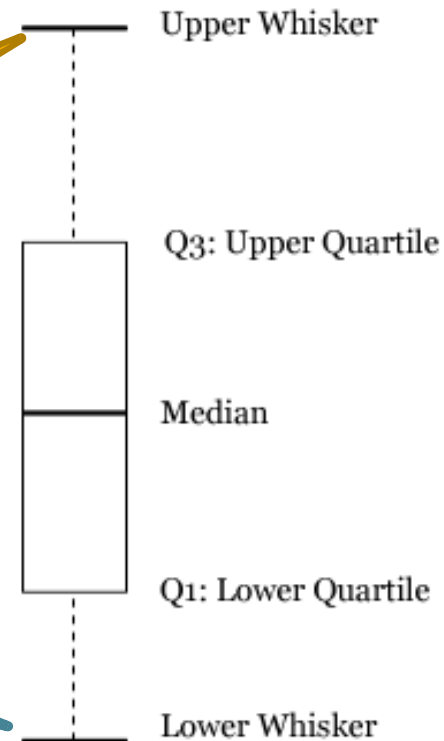
# Boxploty

menší nebo rovna  
horního kvartilu a  
interkvartilového

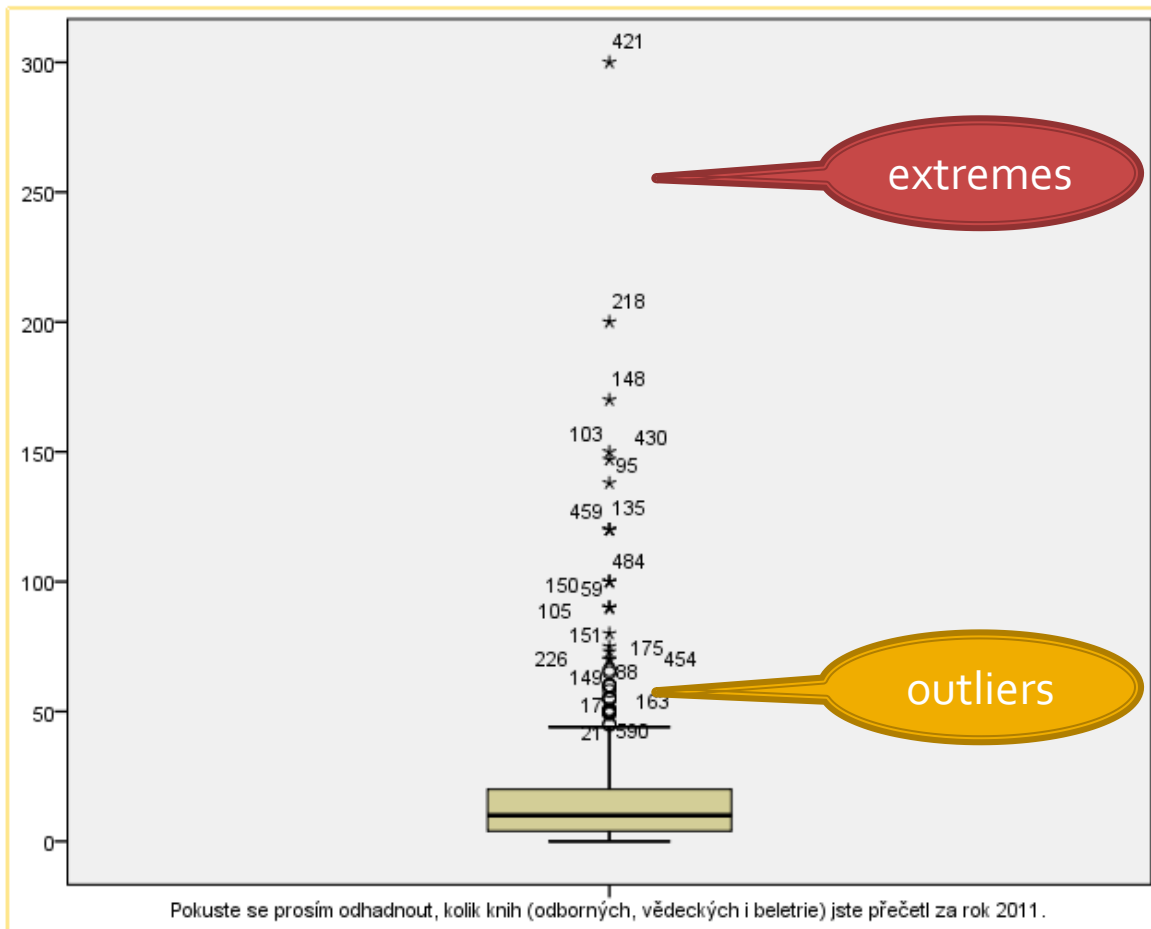
- Velmi podrobné a přehledné vizualizace distribuce spojitých dat

nejvyšší hodnota výběru, která je menší nebo rovna součtu hodnoty horního kvartilu a 1.5 násobku interkvartilového rozpětí

nejnižší hodnota výběru, která je větší nebo rovna rozdílu hodnoty dolního kvartilu a 1.5 násobku interkvartilového rozpětí



# Boxplot (příklad)



Analyze – Explore  
- Plots



# Za DÚ:

- Kolik průměrně naši respondenti kupují knih za rok?
- Je průměr v tomto případě vhodný ukazatel? Zdůvodněte své stanovisko...
- Vytvořte boxploty pro přečtené a nakoupené knihy a porovnejte je.

# Za DÚ z minula:

- Vytvořte novou proměnnou `spokojenost_2`, kde budou jen tři kategorie:
  - Spokojen/a
  - Ani/ani
  - Nespokojen/a
- Vytvořte novou proměnnou „vek“ a okomentujte rozložení (max, min, středové hodnoty)
- Z jakých zdrojů získávají knihy muži a ženy? (vytvořte tabulku četností podle pohlaví)
- Z kolika zdrojů získáváme knihy?