

# Univariační analýza

# Plán pro dnešek:

1. Otevřete si soubor .sav „univariace“ ve studijních materiálech
2. Postupně si budeme představovat a provádět základní univariaci
3. Dávejte o sobě vědět, pokud budete potřebovat pomoc

# Čištění

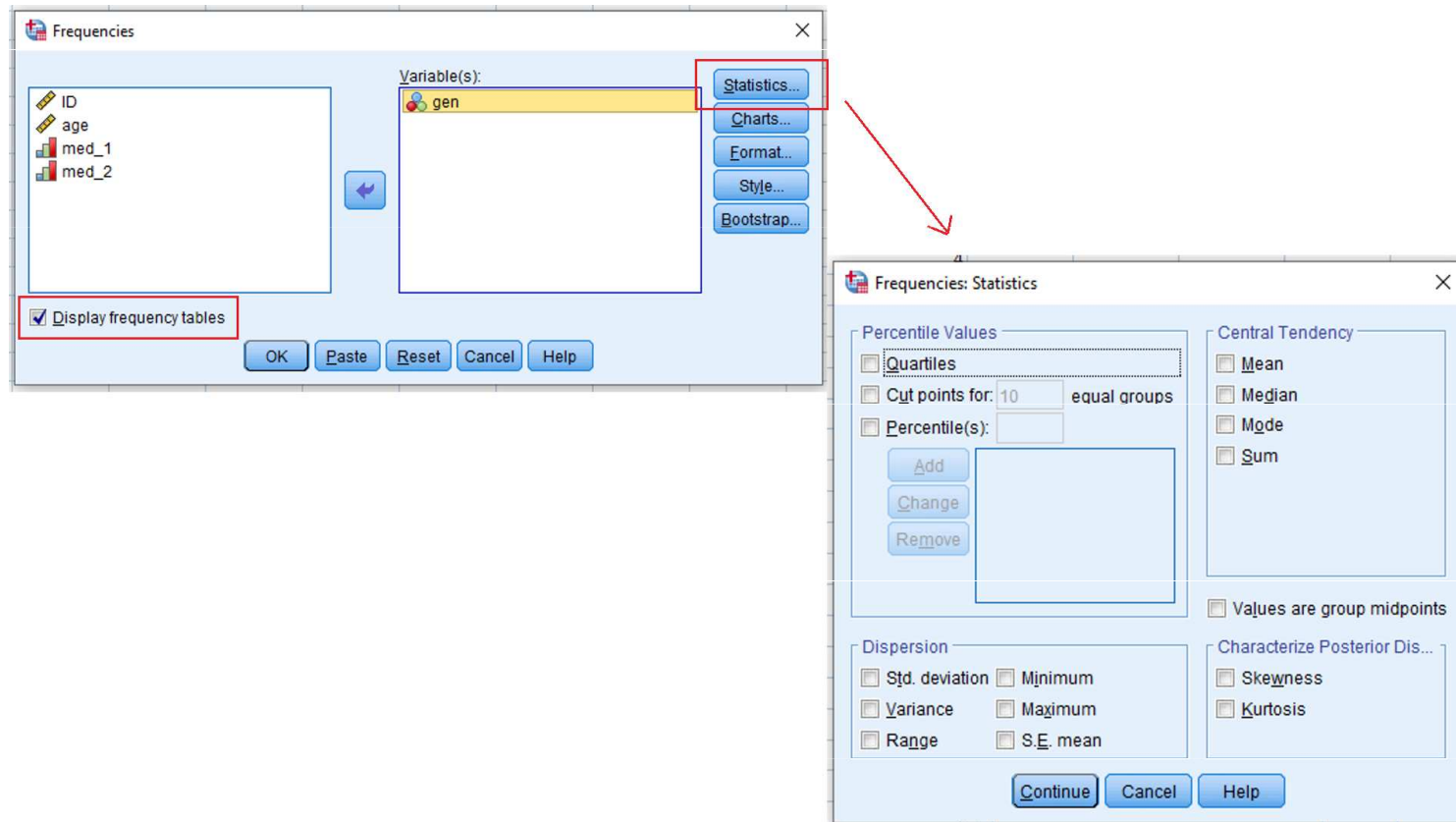
- Ideálně ještě před analýzou (nebo deskripcí) – využívá ale stejné nástroje
- Proč?
  - Data mohou mít celou řadu chyb – překlepy, špatné kódování, invalid data ze strany respondenta...
- V základu nejčastěji
  - Minimum a Maximum (kontrola „outlierů“)
  - Zjištění základních četností hodnot proměnné (správnost hodnot, vyplněnost)  
CTRL+F, případně rekódování (v dalších hodinách)

# Co nám může číselná univariace ukázat?

- Četnosti (absolutní, relativní, kumulativní)
- Minimální a maximální hodnoty
- Percentily
- Střední hodnoty
- Míry variability

## – Jak na základní univariaci?

- Analyze – Descriptive Statistics – Frequencies - Statistics



# Percentily

- Jaké hodnoty nám rozdělují na X% vzorku?
- 4 stejné díly vzorku – kvartily
  - 25%, 50%, 75%
- 10 stejných dílů vzorku – decily
  - 10%, 20%, 30%...
- Různě velké části vzorku, například spodních 17% a vrchních 83% (percentily fungují vzestupně!!)

Frequency: Statistics

**Percentile Values**

Quartiles

Cut points for: 10 equal groups

Percentile(s):

Add

Change

Remove

**Central Tendency**

Mean

Median

Mode

Sum

Values are group midpoints

**Dispersion**

Std. deviation  Minimum

Variance  Maximum

Range  S.E. mean

**Distribution**

Skewness

Kurtosis

Continue Cancel Help

# Střední hodnoty

- V širším hledisku obsahuje větší množství ukazatelů – my si budeme povídat o 3:
  - Modus – nejčastější hodnota
  - Medián – hodnota, která výběr rozděluje na 2 početně stejné jednotky
  - Průměr – aritmetický průměr



Statistiky: Frekvence: Statistika

Percentile Values

Quartiles

Cut points for:  equal groups

Percentile(s):

Central Tendency

Mean

Median

Mode

Sum

Values are group midpoints

Dispersion

Std. deviation  Minimum

Variance  Maximum

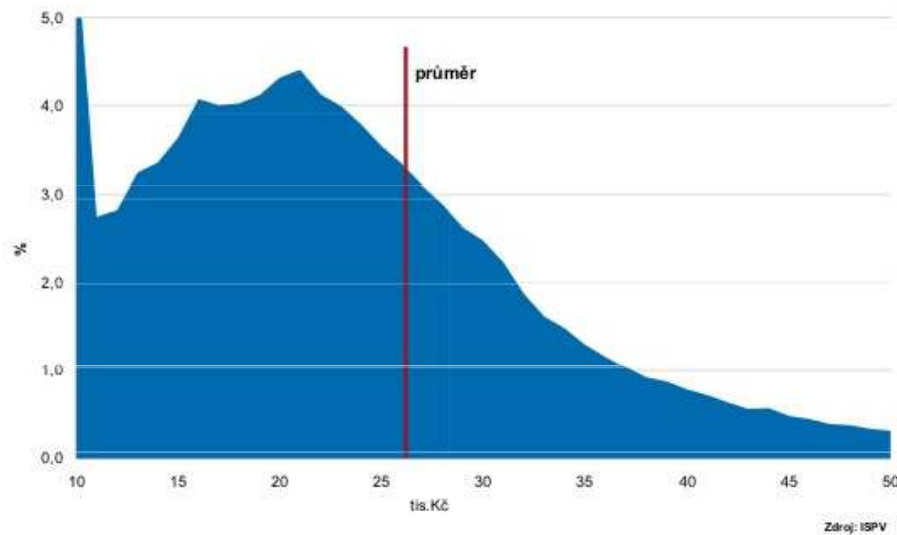
Range  S.E. mean

Distribution

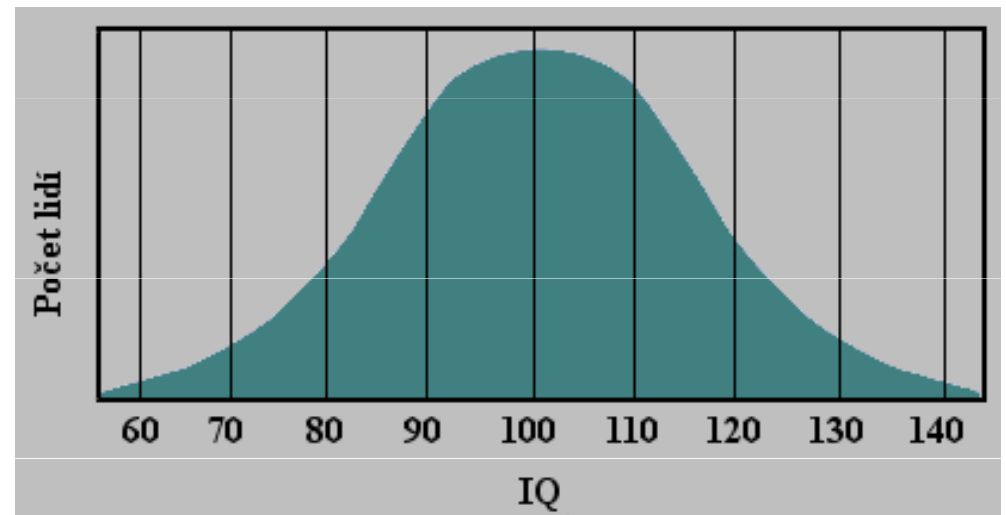
Skewness

Kurtosis

# Střední hodnoty – stejné X rozdílné



Dvě třetiny zaměstnanců jsou pod průměrnou mzdou



# Míra variability

- (minimum, maximum, range...)
- Nominální – koncentrace
- Ordinální – variační rozpětí, ordinální rozptyl
- Obvykle nás zajímají pro průměr (tj. kardinální proměnné)
  - Směrodatná odchylka (průměrná odchylka od průměru)
  - Rozptyl (mocnina ze směrodatné odchylky)

Frequency: Statistics ×

**Percentile Values**

Quartiles

Cut points for:  equal groups

Percentile(s):

**Central Tendency**

Mean

Median

Mode

Sum

Values are group midpoints

**Dispersion**

Std. deviation  Minimum

Variance  Maximum

Range  S.E. mean

**Distribution**

Skewness

Kurtosis

# Frekvenční tabulka

- Vhodná jen pro kategorizované data
- Základní přes Analyze – Desk. Stat. - Frequencies – Display freq. Tables
- Pokročilé přes Analyze – Tables – Custom tables

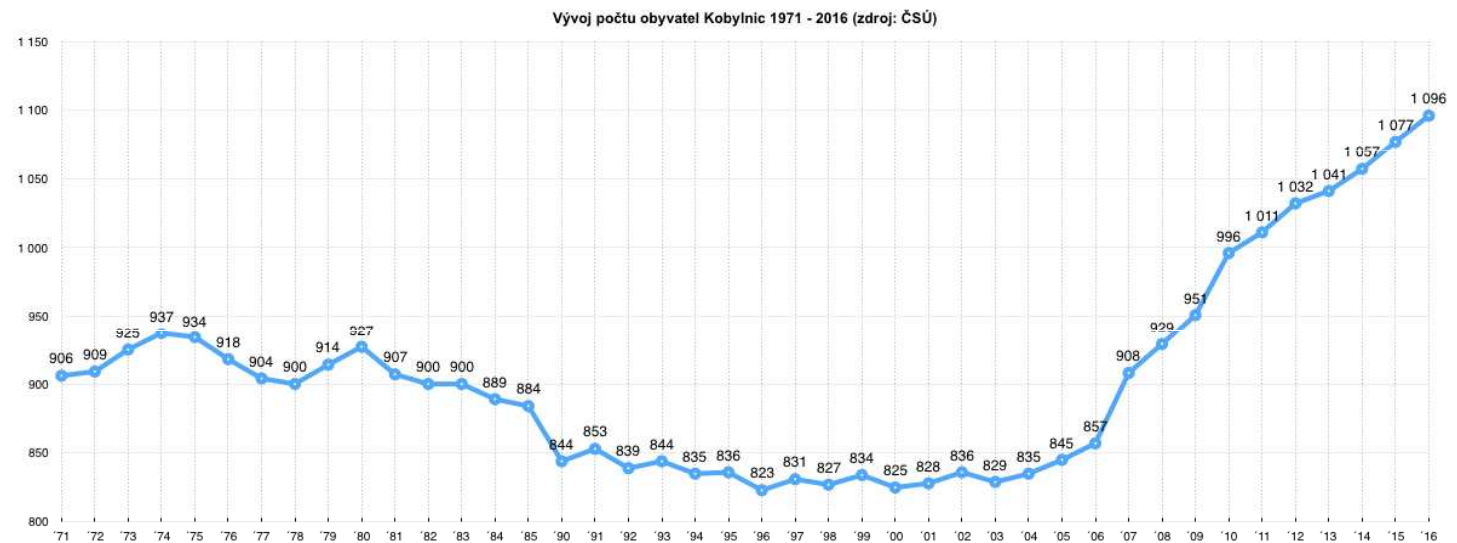
# Co ve frekvenční tabulce máme za údaje?

- Pozorované N (=počet jednotek pro dané hodnoty)
- Procenta – procentuální vyjádření pozorovaného N k celku
  - (Pozor! U custom tables pozor na řádková/sloupcová procenta)
- Validní procenta – procenta ke všem validním (nechybějícím) hodnotám
- Kumulativní procenta – součet procent této a nižších hodnot

# Grafy

## – Několik základních druhů

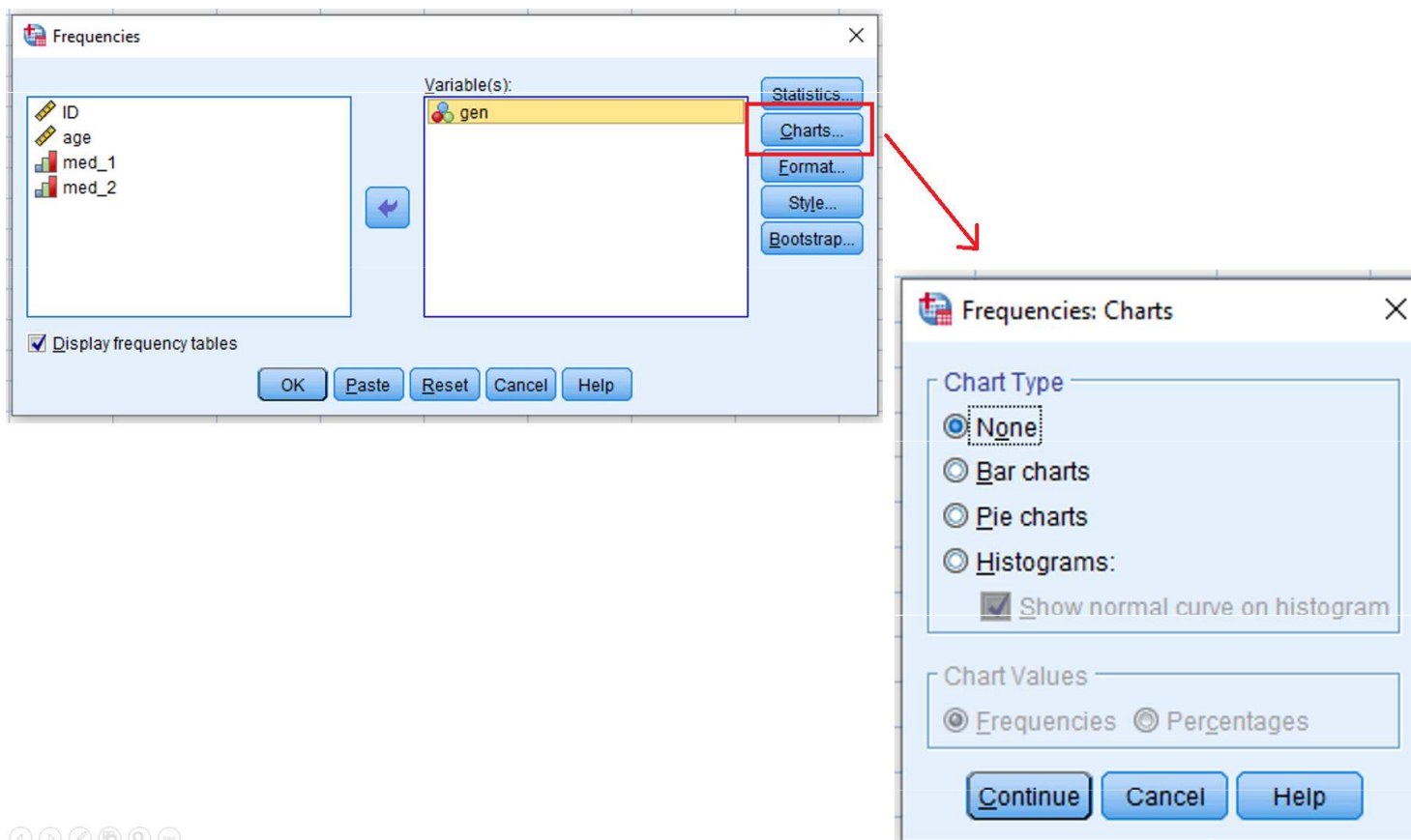
- Sloupcový
- Výsečový (koláčový)
- Histogram
- Boxplot (krabicový)
- Spojnicový



- Sloupcový
  - Vhodný pro srovnání kategorických hodnot
- Výsečový (koláčový)
  - Vhodný pro %
- Histogram
  - Vhodný pro kardinální položky s velkou řadou hodnot
- Boxplot
  - Vhodný pro kardinální položky s informacemi o kvartilech
- Spojnicový
  - Vhodný pro kardinální položky s malou řadou hodnot



- Sloupcový graf, výsečový graf a histogram
  - Analyze – Descriptive Statistics – Frequencies - Charts



## – Box plot, spojnicový

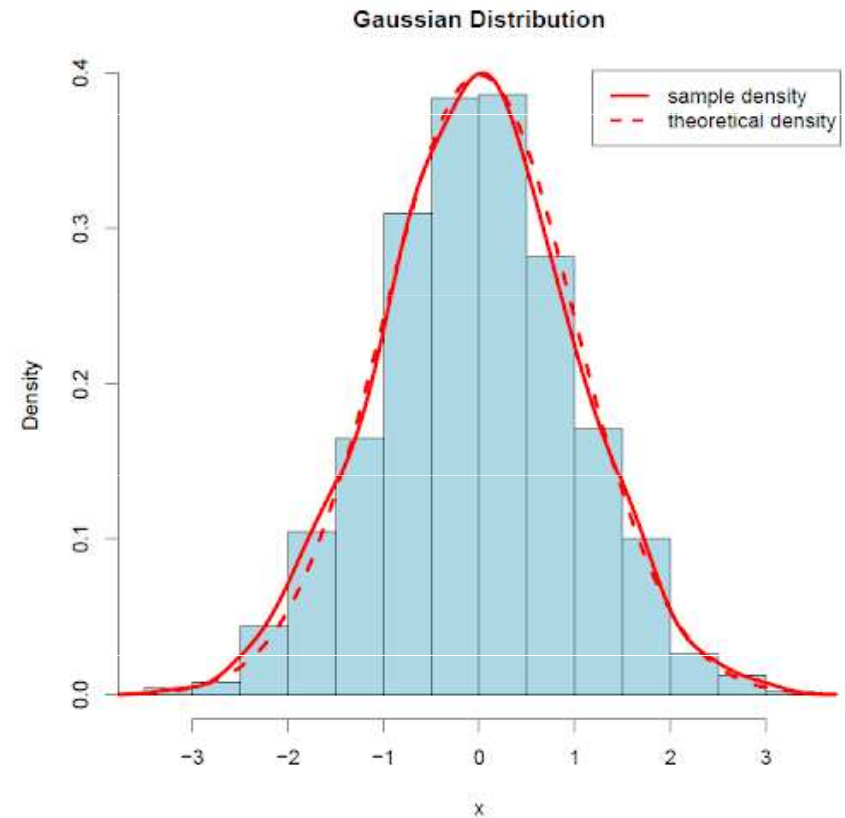
- Graphs – Chart Builder (i alternativa k ostatním druhům grafu)
- Analyze – Descriptive Statistics – Explore – Plots – Boxplots

The screenshot shows the Minitab Chart Builder dialog box. The main window displays a 1-D Boxplot titled "1-D Boxplot of Jaký je Váš věk?". The Y-axis is labeled "Jaký je Váš věk?". The chart shows a blue box with a horizontal line representing the median, and whiskers extending to the minimum and maximum values. The dialog box includes a "Variables:" section with a list of variables: ID, Jaké je Vaše pohlav..., Jaký je Váš věk? [age], Důvěřujete zpravoda..., and Důvěřujete zpravoda... The "Basic Elements" tab is selected, and the "Boxplot" option is chosen from the "Choose from:" list. The "Element Properties" tab is also visible, showing the "Edit Properties of:" section with "Box1" selected, and the "Statistics" section with "Boxplot" selected. The "Error Bars Represent" section is also visible, with "Confidence intervals" selected and "Level (%)" set to 95.

- Poznámka
- 22.10. jsme došli až sem. Normální rozložení a rozdíl mezi deskriptivní a inhereční univariací není součástí lekce univariace. Není ani obsahem kurzu, slouží pouze k doplnění. Následující slidy tedy nebudou v úkolu a také nejsou nutné k ukončení kurzu. Zároveň slouží pouze jako pomůcka k prezentaci vyučujícího. Samotné o sobě to nedává smysl. RZ.

# Normální rozdělení

- Nebo také „Gaussovo rozdělení“
- Modelové rozdělení náhodné veličiny
- V populaci poměrně běžné
- Rychlý test – šikmost a špičatost
  - Šikmost (skewness) – symetrie rozložení; blízká 0, pokud normální rozdělení
  - Špičatost (kurtosis) – míra soustředění hodnot kolem středu; blízká 0, pokud normální rozdělení

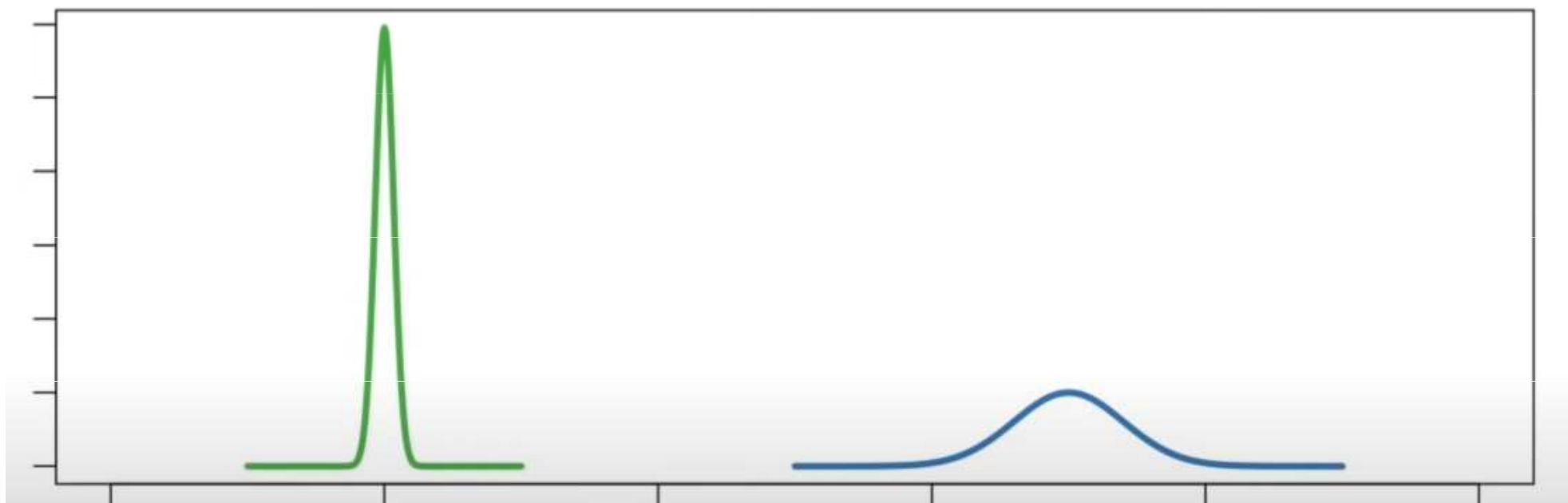
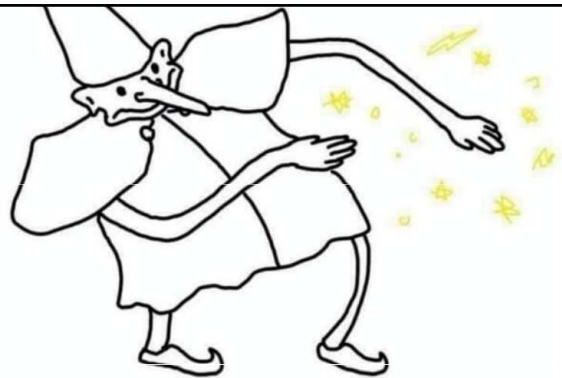


# Okénko do pravděpodobnosti (navíc)



„I cast  
Fireball!“

„Roll 8d6  
for  
damage“



- Cca +/- 2 směrodatné odchylky = 95% případů na normálním rozložení
- Velké vzorky se chovají jako normální rozložení – centrální limitní věta

