

Normální rozdělení

Úvod

- Potřebujeme znát pro další statistickou práci
 - Může být použité pro srovnání populací
 - Důležité pro lineární modely
-
- Jak vypadá rozdělení hodnot proměnné?
 - Vypadá podobně jako zvon nebo nějak jinak?

Na jaké znalosti budeme dnes stavět?

- Typy proměnných
- Histogram
- Průměr
- Rozptyl
- Směrodatná odchylka

Co je to rozdělení

- Kolikrát se opakuje nějaká hodnota v datovém souboru
- Různé podoby rozdělení
 - Normální
 - Poissonovo
 - Uniformní

Terminologická vsuvka

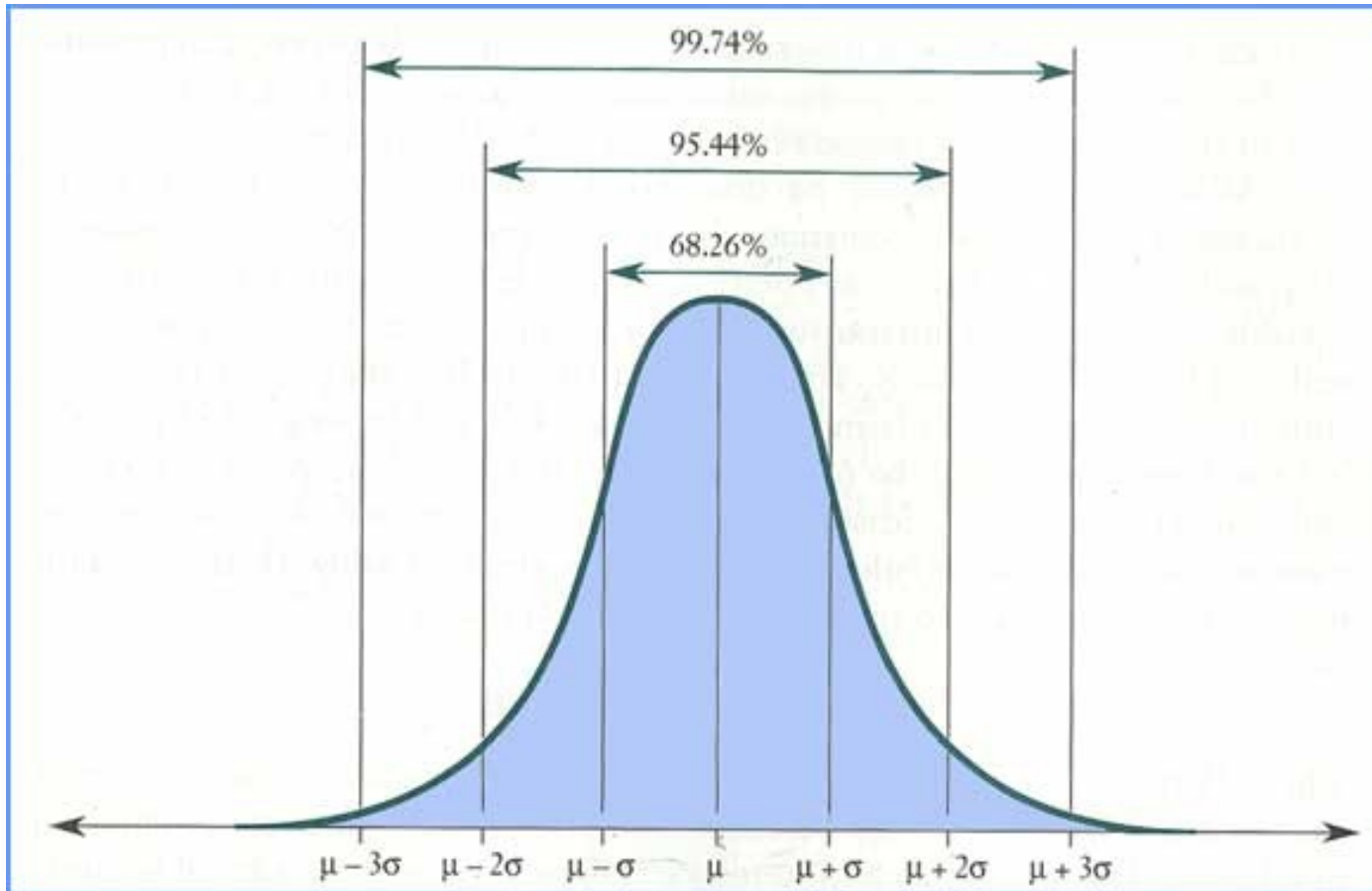
- „Normální rozdělení“ je pouze termín
- Rozdělení odlišné od „normálního“ může být pořád normální
- Počet zlomenin
 - Je normální, že většina lidí neměla žádnou zlomeninu (ale rozdělení se liší do normálního)
- Počet trestních činů, počet vítězství na MS...

Jak normální rozdělení vypadá

- Symetrické rozložení zvonovitého tvaru
- Jeden vrchol
- Průměr = medián

- 67 % případů v intervalu průměr \pm 1 sm. odch.
- 95 % případů v intervalu průměr \pm 2 sm. odch.
- 99 % případů v intervalu průměr \pm 3 sm. odch.

Jak normální rozdělení vypadá



Odpovídá rozdělení proměnné normálnímu rozdělení?

- Ano:
 - Můžeme používat řadu statistických nástrojů bez problémů
 - Aplikace znalostí o normální distribuci
- Ne:
 - Transformování proměnné
 - Rekódování proměnné
 - Přizpůsobení se situaci (záleží na požadované přesnosti)

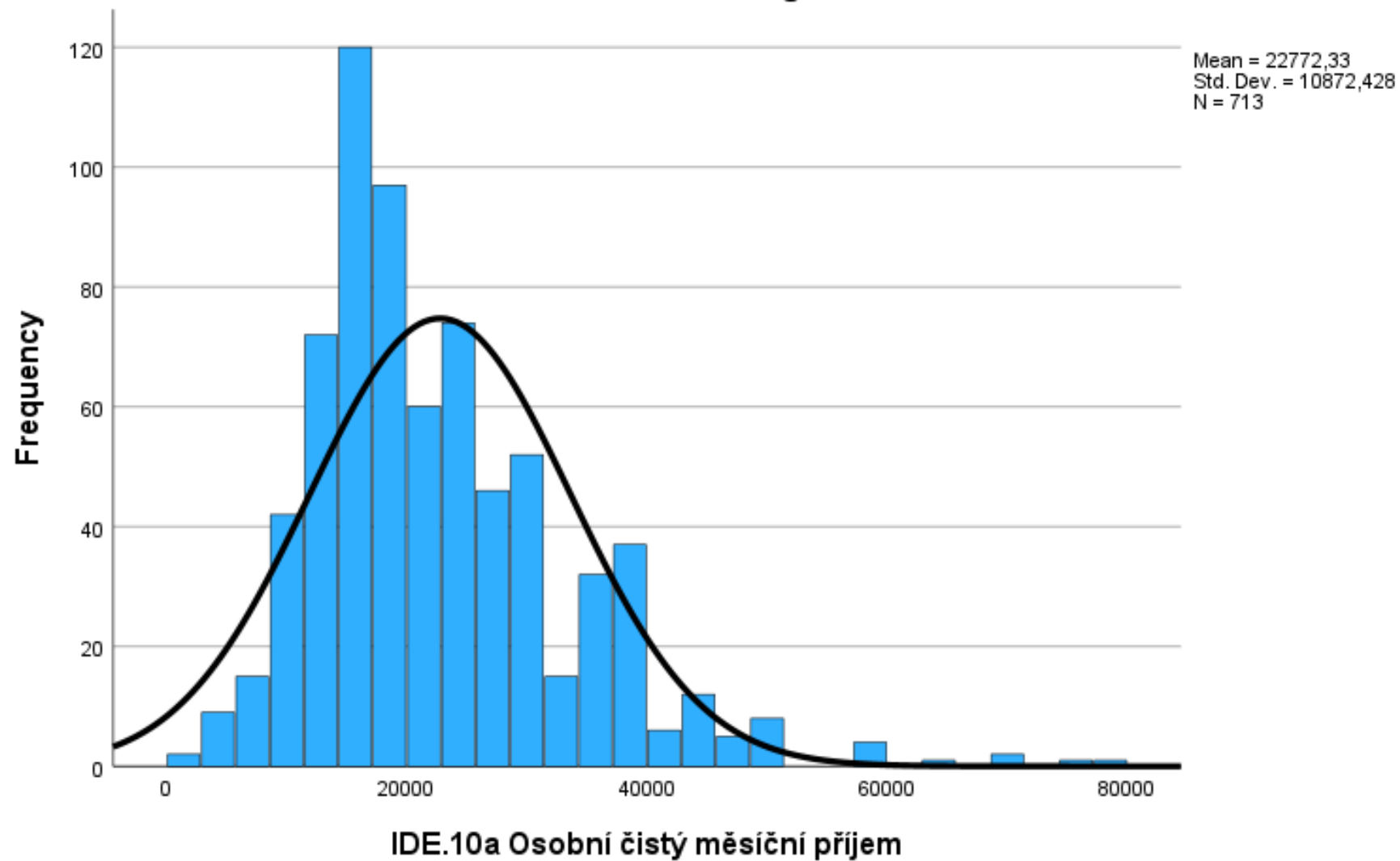
Způsoby posouzení

- Vizually - histogram
- Šikmost a špičatost
- Testy – Kolmogorov-Smirnov a Shapiro-Wilk
- **Pozor na počet pozorování!**

1. Histogram

- Nejjednodušší způsob posouzení normality (a také subjektivní)
- Posouzení tvaru volným okem
- Histogram – graf zobrazující četnosti
- Postup v SPSS:
 - Analyze → Descriptive Statistics → Frequencies
 - Charts – Histograms + Show normal curve on histogram

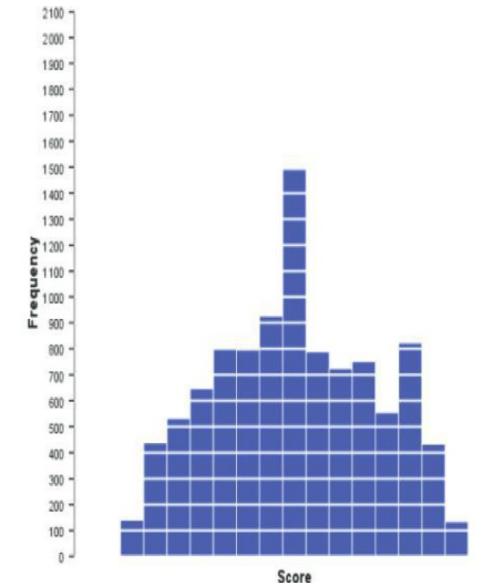
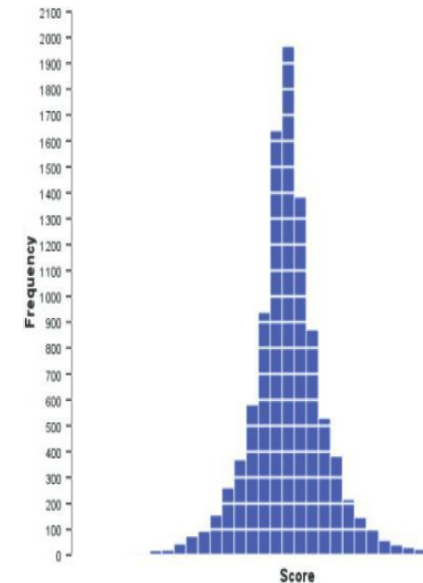
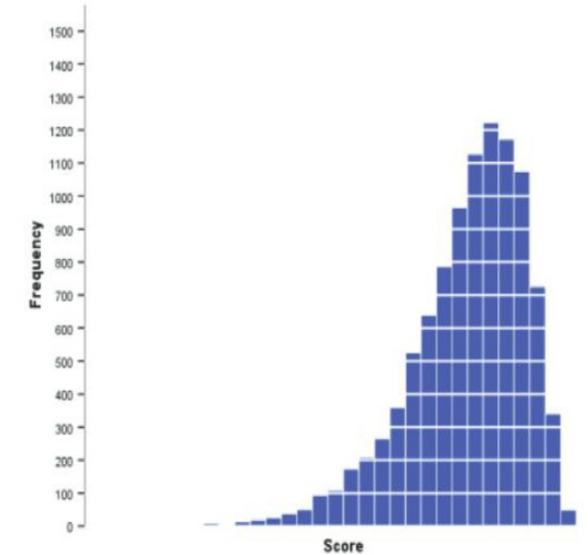
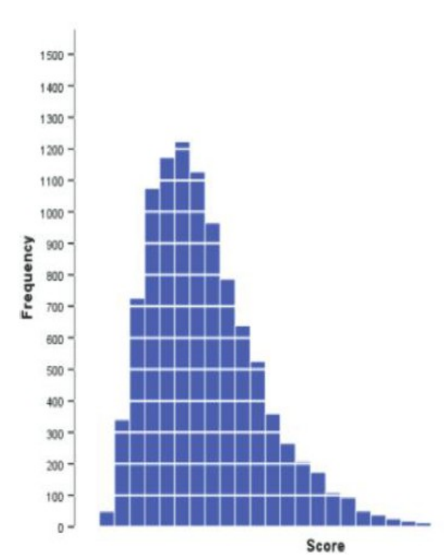
Histogram



2. Šikmost a špičatost

- Šikmost (skewness)
 - Průměr \neq medián (většina hodnot je pod nebo nad průměrem)

- Špičatost (kurtosis)
 - Hodnoty jsou příliš koncentrovány nebo rozptýleny okolo středu



2. Šikmost a špičatost

- Postup v SPSS:
 - Analyze – Descriptive Statistics - Explore
- Odchyly od nuly (kladné i záporné) jsou vychýlením od normální distribuce
- Samotné naměřené hodnoty jsou informativní, pro interpretaci se dělí svou standardní chybou (počítá SPSS)
- Výsledek po dělení $> 1,96/2,58$ => rozdělení proměnné je výrazně špičaté nebo šikmé (tohle má smysl pro vzorky do cca 500 případů)

Descriptives

		Statistic	Std. Error	
IDE.10a Osobní čistý měsíční příjem	Mean	22772,33	407,176	
	95% Confidence Interval for Mean	Lower Bound	21972,92	
		Upper Bound	23571,74	
	5% Trimmed Mean	22067,67		
	Median	20000,00		
	Variance	118209687,83		
	Std. Deviation	10872,428		
	Minimum	2000		
	Maximum	78000		
	Range	76000		
	Interquartile Range	13000		
	Skewness	0,10		
	Kurtosis	0,00		

$0,10 = 13,72$
 $0,00 = 14,55$

3. Testy normálního rozložení

- Kolmogorov-Smirnov test, Shapiro-Wilk test
- „Exaktní“ zhodnocení normality
- Liší se rozložení proměnné signifikantně od normálního rozložení?
- Vhodné spíše pro malé vzorky
- Postup v SPSS:
 - Analyze → Descriptive Statistics → Explore
 - V „Plots“ zvolit Normality plots with tests
 - Příslušné proměnné vložit do „Dependent list“

Hypotézy

- H_1 : Vyšší nezaměstnanost \rightarrow větší podpora krajní pravice
- H_0 (nulová): žádný vztah mezi nezaměstnaností a podporou krajní pravice
- Statistické testy testují **nulové** hypotézy
- Výsledek:
 - Signifikantní - nulovou hypotézu zamítáme \rightarrow opora pro H_1
 - Nesignifikantní - nulovou hypotézu ponecháme v platnosti \rightarrow žádná opora pro H_1

Signifikantnost ve výstupech SPSS

Výstup SPSS	Signifikantnost
,900	10 %
,750	25 %
,500	50 %
,200	80 %
,100	90 %
,050	95 %
,010	99 %
,001	99.9 %
,000	> 99.9 %

$$= (1 - \text{výstup SPSS}) * 100$$

Příklad: $(1 - 0.234) * 100 = 0.766 * 100 = 76.6 \%$

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
IDE.10a Osobní čistý měsíční příjem	,109	713		,922	713	

a. Lilliefors Significance Correction