

ÚKOL 2

Jméno a příjmení:

UČO:

Imatrik. ročník:

Úkol 2.1: V souboru EVS99_cvicny.sav zjistěte, zdali rozložení názoru na to, kdo by měl být odpovědný za zajištění bydlení (proměnná q54h), je normální. Řešte graficky i početně.

Řešení:

Analyze – Descriptive statistics – Frequencies

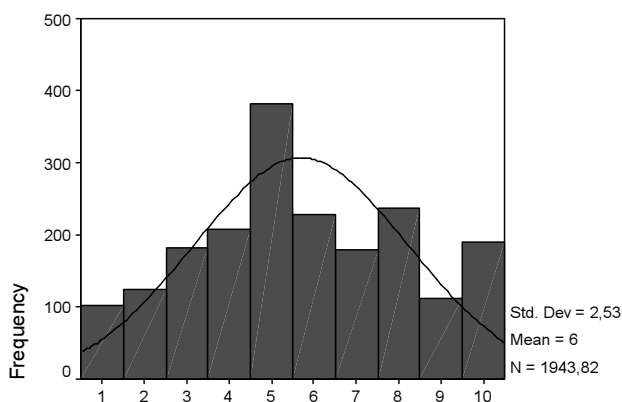
Statistics

Za bydlení má být zodpovědný:

N	Valid	1886
	Missing	22
Mean		5,77
Median		5,00
Mode		5
Std. Deviation		2,531
Skewness		,003
Std. Error of Skewness		,056
Kurtosis		-,869
Std. Error of Kurtosis		,113

Hodnoty šikmosti a špičatosti nejsou příliš vzdáleny od 0, tudíž existuje pravděpodobnost, že rozložení této proměnné bude normální. To si ověříme spočítáním z-skórů pomocí směrodatné odchylky šikmosti a špičatosti. Z-skór šikmosti vyšel 0,05, tj. nižší než 2, z-skór špičatosti je však - 2,8, jeho hodnota je vyšší než 2. Z toho lze vyvodit, že šikmost sice odpovídá normálnímu rozložení, ale naše rozdělení je plošší než normální.

Za bydlení má být zodpovědný:



Za bydlení má být zodpovědný:

Cases weighted by VAHA

Okometricky – podle grafu proloženého křivkou normálního rozložení neodpovídá naše rozložení zcela Gaussově křivce, i když od ní není příliš vzdáleno.

Analyze – descriptive statistics – explore

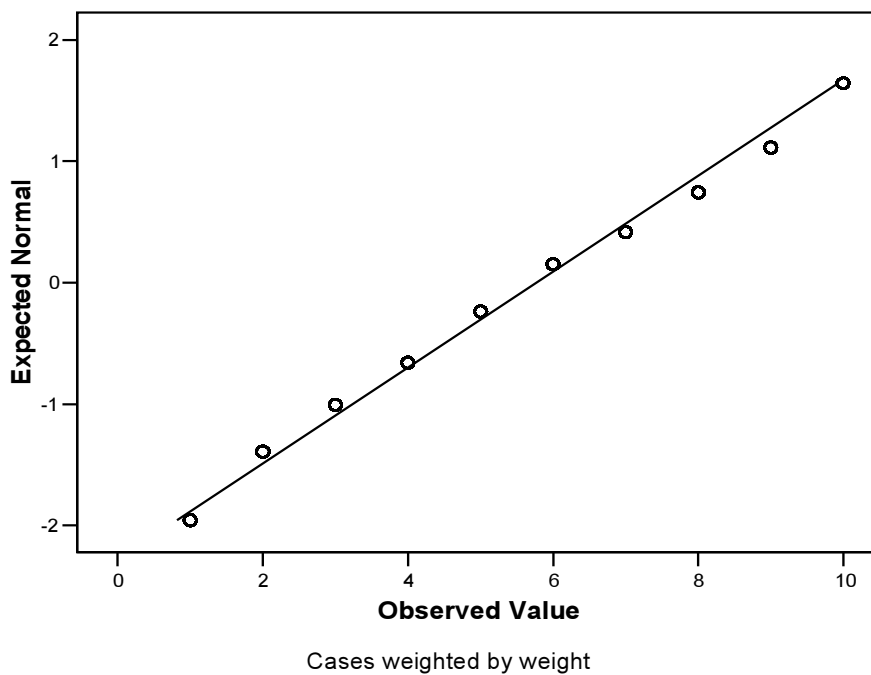
Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
Za bydlení má být zodpovědný:	,123	1886	,000

a. Lilliefors Significance Correction

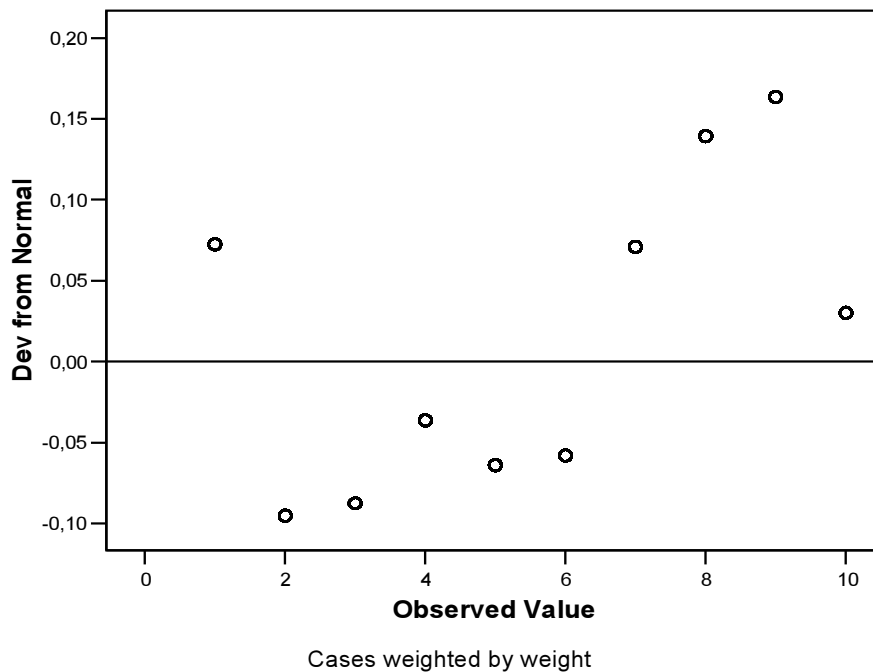
Kolmogorov – Smirnovův test nám vyšel signifikantní – Sig < 0,05, což nám říká, že existuje rozdíl mezi naším rozdělením a rozdělením normálním (zamítáme nulovou hypotézu o neexistenci rozdílu). Podle tohoto testu není naše rozdělení normální. Jelikož však máme velký soubor (mnoho respondentů), je tento test velmi citlivý na jakékoliv odchylky a lze ho pominout.

Normal Q-Q Plot of Za bydlení má být zodpovědný:



Tento graf srovnávající naše rozdělení s normálním nám ukazuje, že se rozdělení proměnné q54h se příliš neliší od normálního rozdělení – pozorované hodnoty jsou téměř všechny na přímce.

Detrended Normal Q-Q Plot of Za bydlení má být zodpovědný:



Zde naše hodnoty nevytvářejí shluky a jsou poměrně blízko přímce (hodnoty na ose y jsou malé) jako by tomu mělo být u normálního rozložení.

U rozložení proměnné q54h existují určité odchylky od normálního rozdělení, ale celkově ze všech testů můžeme říci, že tyto odchylky nejsou příliš velké a naše rozložení se normálnímu podobá. Proto bychom si při další analýze mohli počínat, jako by se jednalo o normální rozložení.

Úkol 2.2: Popište všechny základní charakteristiky věkového rozložení (proměnná vek) v tomto souboru a uveďte, která hodnota věku odděluje 20% nejstarších respondentů.

[V souboru EVS99_cvicny.sav by měla být proměnná vek již vytvořena (zjistíte to ve variable view), ale pokud není, budete si ji muset nejdříve vytvořit. Jelikož se jedná o proceduru transformace dat, kterou ještě neumíte, dáme vám nyní návod, jak na to. Využijeme k tomu příkazu syntaxe:

```
COMPUTE vek = 99-rok_nar .
EXECUTE .
```

Tento příkaz říká, že abychom vytvořili novou proměnnou *vek*, musíme hodnoty proměnné rok narození (*rok_nar*), který je v datech zaznamenán jako poslední dvojčíslí, odečíst od roku, kdy byl

proveden výzkum EVS (což bylo v roce 1999 a my musíme ve výpočtu použít opět pouze poslední dvojčíslí, aby měl výpočet smysl, tedy údaj 99).

Co s tím? Nyní dejte SPSS příkaz (předpokládám, že již máte otevřený datový soubor EVS99_cvicny.sav), aby otevřel nové okno, okno pro práci se syntaxem:

File – New – Syntax

Do tohoto okna vkopírujte příkaz pro výpočet věku:

```
COMPUTE vek = 99-rok_nar .
EXECUTE .
```

V tomto syntaxovém okně pak klikněte na lištu **Run** a pak na příkaz **All**. Příkaz se provede a vám se na konci matice objeví sloupec s novou proměnnou *vek*.

Proměnnou *vek* lze vytvořit i v datovém souboru (bez syntaxe) pomocí příkazu Transform – Compute Target variable = vek, Numeric expression = 99-rok_nar. Tuto proměnnou si můžeme dále nadefinovat (label apod.) ve variable view.]

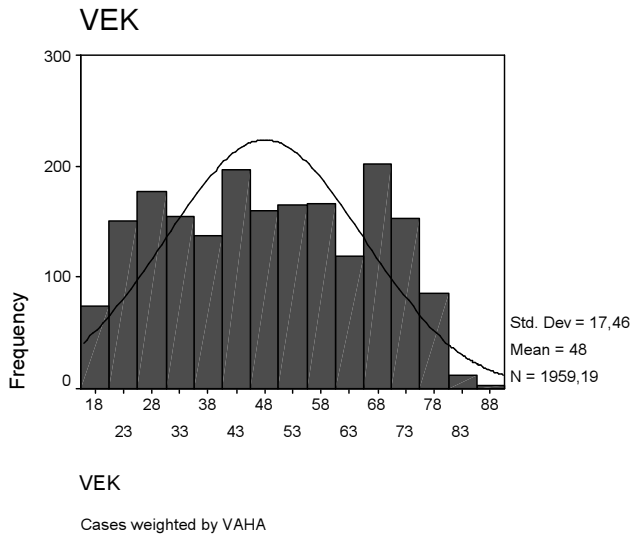
Řešení:

Analyze – descriptive statistics – explore

Descriptives			Statistic	Std. Error
vek	Mean		45,6766	,38611
	95% Confidence Interval for Mean	Lower Bound	44,9194	
		Upper Bound	46,4339	
	5% Trimmed Mean		45,4578	
	Median		45,0000	
	Variance		283,610	
	Std. Deviation		16,84074	
	Minimum		17,00	
	Maximum		88,00	
	Range		71,00	
	Interquartile Range		28,00	
	Skewness		,142	,056
	Kurtosis		-1,030	,112

Průměrný věk v našem výběrovém souboru je 46 let. S 95% spolehlivostí můžeme říci, že v základním souboru se průměrný věk pohybuje mezi 45 a 46 lety. Odlehlé hodnoty na průměr nemají vliv, protože ořezaný průměr je téměř shodný s průměrem pro celý soubor. Medián, který náš soubor pólí, dosahuje hodnoty 45. Směrodatná odchylka, která nám ukazuje míru variability je 17. Z ní lze spočítat variační koeficient (vydělíme ji průměrem), který nám vyjde 0,36, tj. 36 %. To

znamená, že rozložení souboru je poměrně široké a hodnoty se nepohybují jen těsně kolem průměru. Minimální věk v našem souboru je 17 a maximální 88 let. Šikmost i špičatost jsou poměrně nízké, z-skór šikmosti je 2,5 a z-skór špičatosti je 3. Rozložení věku tudíž není normální, což nám ukazuje i graf proložený normální křivkou



Analyze – descriptive statistics – frequencies – percentils - 80

Statistics

vek		
N	Valid	1902
	Missing	6
Mean		45,6766
Median		45,0000
Mode		18,00
Std. Deviation		16,84074
Skewness		,142
Std. Error of Skewness		,056
Kurtosis		-1,030
Std. Error of Kurtosis		,112
Percentiles	80	63,0000

Nejstarších 20 % respondentů je ve věkové kategorii nad 63 let.

Úkol 2.3: Znázorněte graficky pro jednotlivé vzdělanostní kategorie (proměnná vzdelani) tak, abyste mohli porovnat jejich věkové mediány a interkvartilové rozpětí. Která z nich má nejvyšší medián a která největší interkvartilové rozpětí?

Řešení:

Bud'

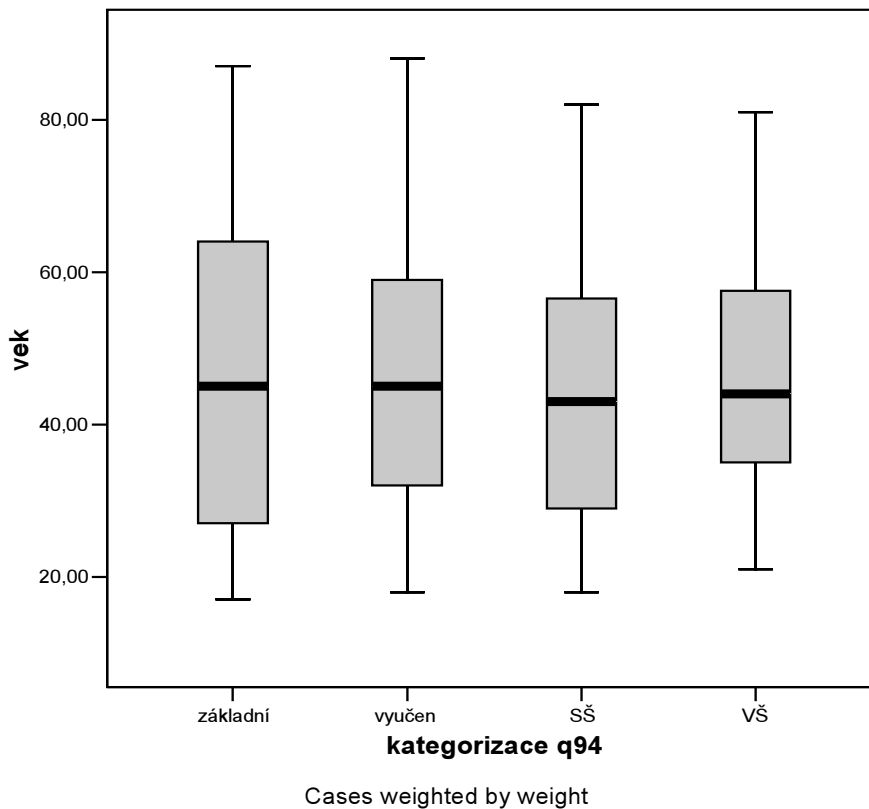
Split file – compare groups – vzdělání

A pak : Analyze – descriptive statistics – explore – vek

Tím uděláme analýzu věku a vykreslí se nám boxplot pro jednotlivé kategorie vzdělání, z něž je možné přečíst medián a interkvartilové rozpětí.

Nebo

Graphs – Boxplot – Simple – Summaries for groups of cases – Define – Variable = vek, Category axis = vzdělání. V tom případě se nám vykreslí jeden obrázek s boxploty jednotlivých věkových kategorií jako je dole.



Nejvyšší věkový medián má kategorie se základním vzděláním: 45,8 let. Největší interkvartilové rozpětí má také kategorie se základním vzděláním: 37 let.