

LEKCE12 FAKTOROVÁ ANALÝZA vzorový výsledek cvičení

JMÉNO A PŘÍJMENÍ:

UČO:

IMATRIK. ROČNÍK:

4.1 Udělejte 15 náhodných výběrů ze souboru EVS-CR1999.sav tak, abyste v každém výběru náhodně vybrali vždy přesně 400 případů z celkového počtu případů. Sledujte, jak se mění průměr a směrodatná odchylka proměnné VEK v těchto výběrech a srovnajte tyto hodnoty s celkovým průměrem. Zobrazte hodnoty těchto průměrů zaokrouhlené do grafu a zjistěte, zdali tato distribuce má charakteristiky normálního rozložení. Budou výsledky vašich řešení identické s řešením vašich kolegů, nebo se budou lišit? Zdůvodněte svou odpověď.

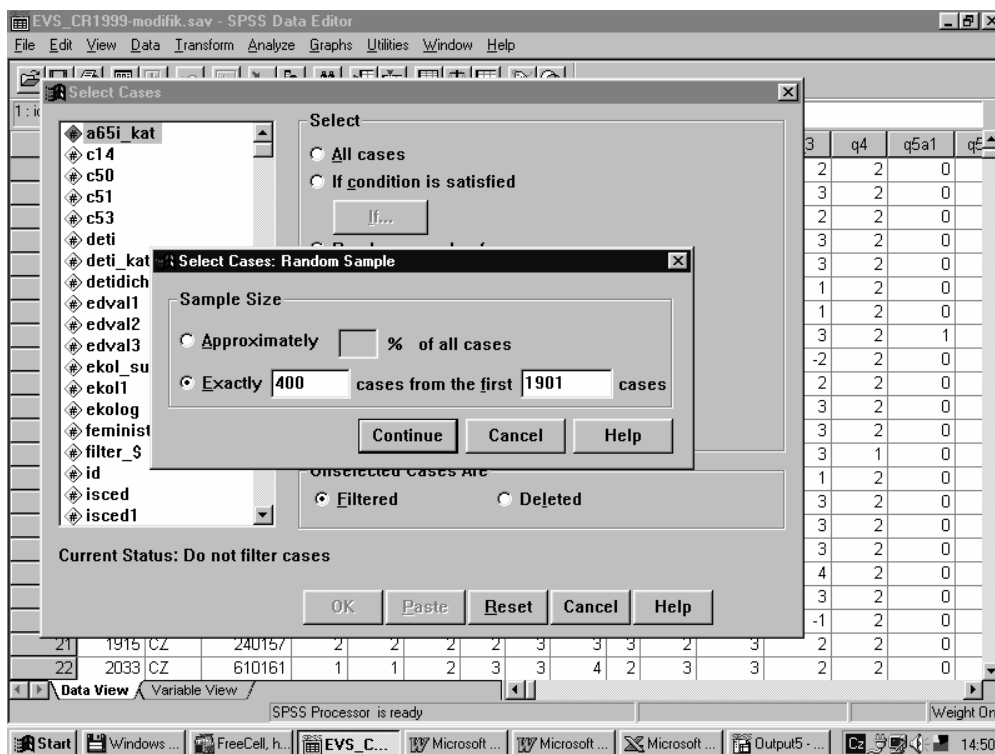
1. Výpočet parametru (nezapomínejme, že soubor, s kterým pracujeme, je pro nás v tomto případě ZÁKLADNÍ SOUBOR, z kterého provádíme NÁHODNÉ VÝBĚRY (o velikosti 400 případů).

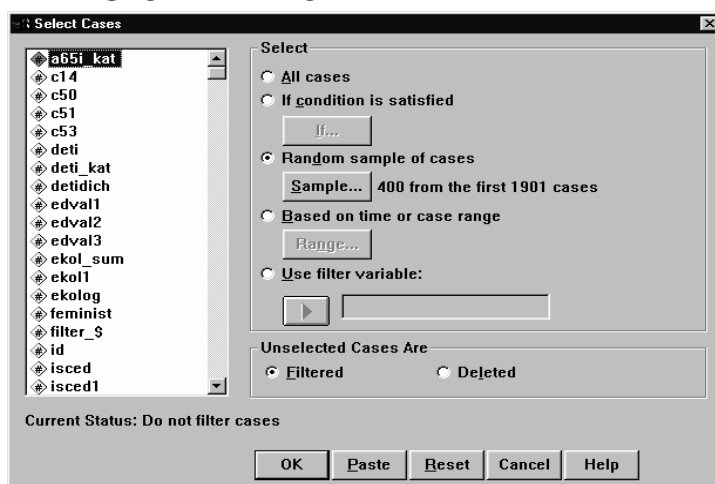
Statistics

VEK		
N	Valid	1901
	Missing	7
Mean		45.69
Std. Deviation		16.83

Parametr je tedy 45,69 let.

2. Výběr případů provedeme v proceduře SELECT CASES: Random sample of cases (400)





Pro každý takto určený výběr provedeme výpočet průměrného věku (FREQUENCIES) a pak musíme znovu provést výběr a znovu provést výpočet atd. Každý výsledek zapišeme do tabulky. Průměry ve vámi vybraných jednotlivých podsouborech budou odlišné, neboť ani jeden z výběrů není svou strukturou identický se strukturou základního souboru. Vaše výsledky také nemohou být identické ani s tímto vzorovým příkladem, ani s řešením vašich kolegů, protože jde o různé výběry. Pokud bychom ale zvýšili počet výběrů například tím, že bychom do řešení zahrnuly všechny vaše výběry, zvýšení jejich počtu by a) přiblížilo rozložení průměrných věku ve výběrových souborech normálnímu rozložení b) přiblížilo průměr vypočítaný z těchto výběrových souborů parametru. Jeden z možných výsledků:

Statistics

VEK		
N	Valid	406
	Missing	0
Mean		44.99
Std. Deviation		16.31

Rozložení průměrů v 15 výběrech:

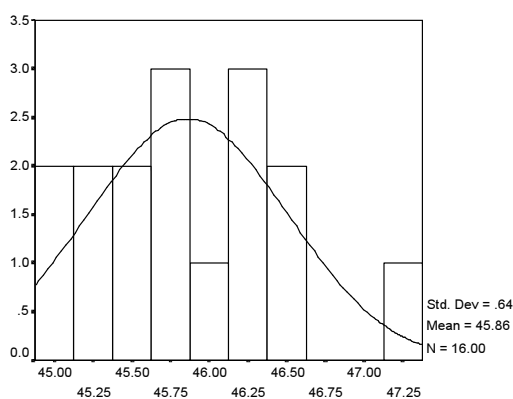
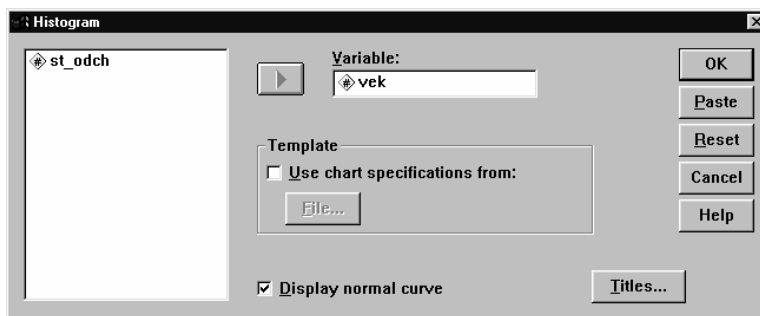
pořadí	průměr	St.odch.
1	44.99	16.31
2	46.09	17.14
3	46.61	17.03
4	46.32	17.40
5	46.46	16.55
6	45.00	16.71
7	45.44	16.99
8	45.22	17.05
9	47.21	16.63
10	46.30	16.37
11	46.30	17.34
12	45.68	16.27
13	45.71	16.74
14	45.22	16.60
15	45.39	16.46
CELKEM	45.86	16.77

Všimněte si:

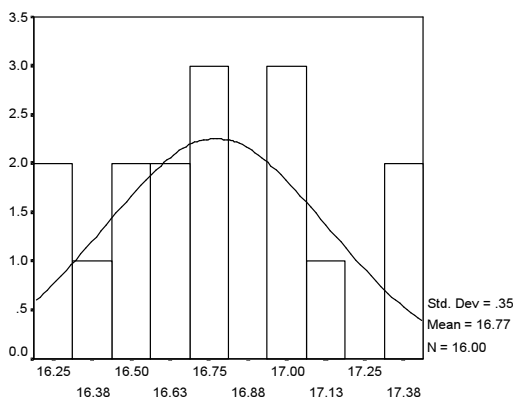
- Průměry (ale také např. standardní odchylky v jednotlivých výběrových souborech se různí (variuji kolem neznámé hodnoty parametru (my ji ovšem v tomto případě známe, protože máme vyčerpávající údaje o základním souboru).
- Jednotlivé výběry mohou mít sice i stejný průměr, ale protože jsou složeny z různých případů, nemusí být v takovém případě shodná standardní odchylka (a naopak).

Další postup:

Tabulku vyplněnou v EXCELU (nastavte si desetinné tečky - vyžaduje to SPSS) zablokujete a vyklopíte do prázdné matice SPSS (jiným způsobem jak převést excelový soubor na systémový soubor v SPSS by bylo použít FILE → OPEN DATABASE → NEW QUERY). Pak si jen popíšete obě proměnné (průměr a standardní odchylku): dáte jim jména, labels a ponecháte 2 desetinná místa. Následně v modulu GRAPHS můžete zadat histogramy a nechat je proložit modelem normálního rozložení.



vek



standardní odchylka

Jak je vidět, k normální rozložení má našich 15 výběrů daleko, **ale to se při tak malém počtu dalo předpokládat**. Při tak malém počtu výběrů se budou také vaše výsledky od výsledků ostatních značně lišit. Avšak čím více výběrů byste udělali tím více by se vaše výsledky blížily normálnímu rozložení a také by byly stále podobnější výsledkům vašich kolegů a kolegyně (za předpokladu, že by i ono zvyšovali stejným způsobem počet svých výběrů).