**MV011 Statistika I – cvičení 4**

1) [R] cv. 8 úkol 1,2,3,4,5,6

1.  Z údajů v tabulce **Sales**, pro které název pozice (***job\_title***) obsahuje řetězec „Rep“, vytvořte html/pdf/rtf obsahující kontingenční tabulku sloupců pohlaví (***gender***) a stát (***country***). Nastavte vhodný nadpis a potlačte výpis datumu. (PROC FREQ)
2. Vytvořte tabulku **Sales1** z tabulky **Sales**, ve které vznikne nový sloupec ***hire\_age*** představující věk zaměstnance v okamžiku nástupu do zaměstnání. Vytvořte formát ***HireAge***, který agreguje zadaný sloupec do kategorií low-<20, 20-<25 a 25-high. Následně vytvořte frekvenční tabulku pro sloupec hire\_age formátovaný pomocí HireAge. (PROC FREQ)
3. Z tabulky **Sales1** z úkolu 2
	1. vypište průměr (mean) a rozsah (range) příjmu (***salary***) pro všechny trojice hodnot sloupců pohlaví (***gender***), stát (***country***) a ***hire\_age*** formátovaného pomocí HireAge z úkolu 2. (PROC MEANS)



* 1. uložte výstup procedury (bez specifikace ukládaných údajů) do tabulky a porovnejte výstup bodu a) a b).



1. Z tabulky **Sales1** z úkolu 2 vytvořte kontingenční tabulku s absolutními četnostmi a řádkově a sloupcově podmíněnými relativními četnostmi. Řádková dimenze bude tvořena kartézským součinem hodnot sloupce ***hire\_age*** formátovaného pomocí HireAge (včetně souhrnu (all)) a hodnot sloupce ***country***. Sloupcová dimenze bude tvořena hodnotami sloupce ***gender***. (PROC TABULATE)



1. Z tabulky **Sales1** z úkolu 2 vytvořte kontingenční tabulku, která bude obsahovat minimum, medián a maximum příjmu (***salary***). Řádková dimenze bude tvořena kartézským součinem hodnot sloupce ***hire\_age*** formátovaného pomocí HireAge a hodnot sloupce ***country***. Sloupcová dimenze bude tvořena hodnotami sloupce ***gender***. U řádkové i sloupcové dimenze včetně všech souhrnů („all“). To vše ve formátu pdf se stylem sasweb. (PROC TABULATE)



1. Analyzujte (zajímá nás základní sada popisných statistik, test pro charakteristiku polohy, kvantily, odlehlá pozorování) sloupec ***salary*** z tabulky **Sales**. Vytvořte výstup ve formátu rtf se stylem sasweb. (PROC UNIVARIATE)



2) V MS Excel (nebo ekvivalentní) nad cs-training (nejlépe neimportované do MS Excel ze SASí tabulky pomocí SAS Add-in for MS Office): spočítat průměr, medián, modus, rozptyl, dolní a horní kvartil, šikmost, špičatost a pearsonův koeficient korelace pro věk a DebtRatio; dále vytvořit kontingenční tabulku NumberofDependents a SeriousDlqin2yrs s absolutními četnostmi, řádkově podmíněnými relativními četnostmi a sloupcově podmíněnými relativními četnostmi.

 



3) [R] cv.9 úkol 1,2

1. Z údajů v tabulce **Sales1** vytvořte bodový graf závislosti ***hire\_age*** na ***birth\_date*** s rozlišením pohlaví (***gender***). Graf doplňte o regresní přímky a upravte vzhled podle vzoru (PROC GPLOT)… formát x-ové osy mmddyy10., tloušťka reg. přímek = 5, font popisu os i legendy = (arial bold, výška 12 bodů, resp. 10 bodů u „regression lines“), font hodnot na osách a hodnot v legendě= (arial bold, výška 10 bodů), výška nadpisu = 12 bodů.



1. Z údajů v tabulce **Sashelp.workers** vytvořte graf počtu elektrikářů (***electric*)** a počtu zedníků (***masonry*)** v *čase(****date***). Upravte vzhled podle vzoru (PROC GPLOT s overlay)… formát x-ové osy mmddyy10., tloušťka křivek = 5, font popisu os = (arial bold, výška 12 bodů), font hodnot na osách a hodnot v legendě= (arial bold, výška 10 bodů), výška nadpisu = 12 bodů, offset legendy = 1%.



* [R] Martin Řezáč – DM1 cvičení <https://www.math.muni.cz/~mrezac/vyuka/DM1_cviceni_v2013.pdf>