TRANSFORMACE A VYTVÁŘENÍ NOVÝCH PROMĚNNÝCH + VÝBĚR PŘÍPADŮ

Deskriptivní analýza kvantitativních dat



MODUS: Nejčastější hodnota 3 4 7 8 10 11

SMĚRODATNÁ ODCHYLKA: Vyjadřuje, jak moc jsou jednotlivé hodnoty rozptýleny kolem průměru PRŮMĚR: 3 4 7 8 10 11 11 M = 7.71

10 11 11

MEDIÁN:

3

Střední hodnota

7

4

SD = 3.25

Jak reportovat základní charakteristiky jednotlivých proměnných v textu/úkolu N (počet validních hodnot) a navíc:

 Nominální/ krátké ordinální: (relativní) četnosti kategorií (tj. Procenta)

– Kardinální (škálové): průměr, SD, minmax

Cvičení

- V datasetu transformace_lekce zkuste vyjet četnosti proměnné, která měří míru důvěry k Televizi Barrandov + její modus
 - Analyze \rightarrow Descriptive Statistics \rightarrow Frequencies
 - Tam kliknout na Statistics a naklikat, co potřebujete (viz minulá hodina)

 Co když tuším, že medián důvěry k Televizi Barrandov by se mohl lišit s ohledem na věk respondenta?

A) transformovat proměnnou věk tak, abychom měli menší množství kategorií
B) zobrazit tuto proměnnou vzhledem k jiné kategorii

Rekódování proměnných (transformace)

- 1. Máme jiný typ próměnné, než potřebujeme
 - Nejčastěji string a potřebujeme numeric
- 2. Máme proměnnou kardinální a chceme z ní udělat ordinální
 - Typicky se tento typ transformace používá u věku shlukujeme do kategorií
- 3. Proměnná je nakódovaná jinak, než chceme
 - Nejčastěji měníme orientaci škály například, kdy chceme, aby se stoupající stupnicí stoupal i postoj k dané proměnné, ale orientaci škály je opačná
 - Nepsaný zvyk: Čím vyšší skór mám, tím více se u mě projevuje měřená charakteristika (názor, postoj, vlastnost…)
- 4. Chceme vytvořit průměrnou hodnotu z "příbuzných" proměnných – tj máme několik položek, které nám měří dohromady jednotlivé dimenze nějakého jevu

Máme jiný typ proměnné než potřebujeme

■ Procedura Transform → recode into different variable (string na numeric)

<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	Direct <u>M</u> arket	ing	<u>G</u> raph:	s	<u>U</u> tili
2				Comput	e Variable alues withir	1 Cases			ĥ	
199 : jazyk_ru 0				Shi <u>f</u> t Val	Shift Values					
			plat	Recode	into Same '	Variables			erne	t
	1	0	14700	Decede	into <u>D</u> iffere	nt\/orighlag				5
	2	2	12700	Recode	into Dillere	nt variables				6
	3	0	12700	🔯 <u>A</u> utomat	ic Recode					5
	4	7	12500	Visual <u>B</u>	inning					5
	5	0	12700	🔀 Opt <u>i</u> mal	Binning					5
	6	2	29600	<u>P</u> repare	Data for M	odeling		•		6
	7	9	14500	Rank Ca	ises					4
	8	1	13300	🗎 Date an	d Time Wiz:	ard				6
	9	9	12900			ar u				5
	10	8	12900		I <u>m</u> e Series					3
	11	2	19300	📲 Replace	Missing <u>V</u> a	alues				5
	12	0	13900	🍘 Random	n Number <u>G</u>	enerators				6
	13	7	13800	Run Per	iding Trans	forms	Ctrl+	G		5
	14	6	13500	2	2 m		1		1	6
		1								

🙀 Recode into Different Variables: Old and New Values

Old Value	New Value
	⊙ Va <u>l</u> ue:
	O System-missing
● <u>S</u> ystem-missing	Copy old value(s)
O System- or <u>u</u> ser-missing	
● Ra <u>ng</u> e:	0 <u>id</u> > New: Y> 1 'm' -> 2
<u>t</u> hrough	Add
Range, LOWEST through value:	Re <u>m</u> ove
Range, value through HIGHEST:	
	Output varia <u>b</u> les are strings <u>W</u> idth: 8
○ All <u>o</u> ther values	Convert numeric strings to numbers ('5'->5)
	<u>Continue</u> Cancel Help

Když máme hodnoty v jiném měřítku než v jakém je mít chceme

i.e.	0	0	(tan'ini ingula, ao darar	(0, no)	Nene	9 = Diabt d Ordinal	Insut	
IC	0	U	cterymi jazyky se doroz	cumite? francouzstina	{0, ne}	None	Pacada into Different Variables: Old and New Value	~ ~
ric	8	0	Kterými jazyky se doroz	{0, ne}	None	Recode into Different variables: Old and New Value		
	ID průměrní pohlaví (Průměrní Jak často Kterými j Kterými j Kterými j Kterými j	ý měsíční p (gender] ý měsíční p o používáte azyky se do azyky se do azyky se do azyky se do	riables Numeric Varia věk> věk_ka Qld and New Qld and New OK	ble -> Output Variable: t 1. Values case selection condition) <u>Reset</u> Cancel Help	Output Variable <u>N</u> ame: věk_kat Label: věkové kategori C <u>h</u> ang		Old Value Value: System-missing System- or user-missing System- or user-missing Range. 30 through 39 Range, LOWEST through value: Range, value through HIGHEST: All other values	New Value Value Value Value System-missing Copy old value(s) Old> New: 20 thru 29> 1 Add Change Remove Cutput variables are strings Width: 8 Convert numeric strings to numbers ('5'->5) ue Cancel Help

Value Labels				
Value:				Spelling
				opening
Label:				
	1 = "18-24"			
	2 = "25-34"			
	3 = "35-44"			
	4 = "45-54"			
	5 = "55-64"			
<u>A</u> dd	6 = "65 a více"			
<u>C</u> hange				
Remove				
Tto <u>m</u> ove				
		OK Cancel Hel	р	

Když máme proměnnou nakódovanou jinak než jak ji chceme mít

- Standardně platí, že pravidlo, že u delších škál znamená nejmenší hodnota nejmenší míru a největší hodnota největší míru
- Často u škálových proměnných používáme několik položek za sebou, protože díky tomu můžeme docílit přesnějšího změření jevu:
 - R12_1 : Příjmy domácnosti, ve které žiju, pokrývají všechny její potřeby.
 - R12_2 : Myslím, že se mám finančně lépe než většina lidí v této zemi.
 - R12_3 : Obávám se, že na tom v budoucnu budu finančně hůř, než jak na tom jsem nyní.
 - (na 5bodové stupnici, kde 1 znamená "rozhodně nesouhlasím" a 5 znamená "rozhodně souhlasím")

Komputa položek

- Když potřebujeme být trochu "kreativní"
- Umožňuje nám
 - Comupute variable
 - Vytvářet/přetvářet položky podle různých matematických vzorců (sčítání, odčítání, násobení...)
 - Count values
 - Spočítat hodnoty napříč položkama

Interes	farra .	Analyza	<u>C</u> raphs	<u>U</u> tilities	E <u>x</u> tens
🖻 <u>C</u> o	ompute '	Variable			
+ Pr	ogramm	nability Tra	nsformatio	n	
🔀 C <u>o</u>	ount Val	ues within	Cases		
N SI	n <u>i</u> π valu	es			
S 🔤 Re	ecode in	to <u>S</u> ame \	/ariables		-
	ecode in	to Differen	t Variables		
Ν	utomatic	Recode			
S 🕂 Cr	reate Du	mmy Vari	ables		
N 📭 Vi	sual <u>B</u> in	ning			1
N 🔀 O	pt <u>i</u> mal B	inning			
N P	repare D)ata for Mo	deling		>
N 🛃 Ra	an <u>k</u> Cas	es			2
N 🗎 <u>D</u> a	ate and '	Time Wiza	ard		2
N Cr	eate Ti <u>n</u>	<u>n</u> e Series.			à
N N 👯 Re	eplace N	/lissing <u>V</u> a	lues		3
N 🎯 Ra	andom N	lumber <u>G</u> e	enerators		Ē
N R	un Pend	ing Transf	orms	Ctrl+	G
Numeric	4	<u> </u>	U		: кезре

Dvě možná řešení

- Rekódování pomocí procedury "recode" – manuálně převrátíme škálu (1=10, 2=9, …)
- Rekódování pomocí metody compute – spočítáme pomocí vzorečku
 - nová hodnota = nejvyšší hodnota +1 – stará hodnota



Když chceme z "příbuzných" proměnných zjistit průměrnou hodnotu (vytvořit tzv. index)

- Kdy tó potřebujeme: když chceme spočítat skór napříč položkami
- Například máme proměnné, které měří jednotlivé aspekty postoje k sexuální liberalizaci a chceme zjistit, jak průměrně je respondent liberální:
 - Jak moc akceptovatelný je pro vás:
 - VAL1_1 : Potrat
 - VAL1_2 : Homosexualita
 - VAL1_3 : Rozvod
 - VAL1_4 : Sex bez závazku

Numeric Expression:

(VAL1_1 + VAL1_2 + VAL1_3 + VAL1_4)/4

Jak si rozdělit výsledky podle vybrané proměnné



<u>F</u> ile <u>E</u> dit	<u>V</u> iew	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	Direct <u>M</u> arket	ing	<u>G</u> r	
		🏹 Del 🌠 Set	fine <u>V</u> ariable F Measuremer	Properties It Level for U	Jnknown		Z	Selec
1 2 3 4	Nar ID <mark>plat</mark> věk mesto		eate Value Lat py Data Prope w Custom Attr fine date and	n <u>e</u> oronione pels from Da rties i <u>b</u> ute	ata		L ĕrr	
5 6 7	gender internet jazyk_c	➡ Def	fine Date from fine <u>M</u> ultiple R Iidation) Data tesponse S	ets	•	aví Sas Vmi	
8 9 10 11	jazyk_a jazyk_fr jazyk_ru jazyk_n	Ide Ide R Col	ntify D <u>u</u> plicate ntify Unusual m <u>p</u> are Datase	e Cases Cases ets			/mi /mi /mi	Select Cases
12 13 14	jazyk_p prace1 prace2	🔊 S <u>o</u> r 📷 Sor 💐 Tra	t Cases t Varia <u>b</u> les <u>n</u> spose				/mij pr <u>2</u> pr3	 ID průměrný měsíční p věk Jak často používáte
15 16 17 18	prace3 prescas prescas prescas	+ Adj Me + Ca	ust String Wid r <u>g</u> e Files rtesian Produ	lths Across ct	Files	۲	pra čas čas čas	Kterými jazyky se do Kterými jazyky se do Kterými jazyky se do Kterými jazyky se do
19 20 21	gender_	₩ <u>R</u> e:	structure ke Weights pensity Score	Matching			aví 3 9	
22 23 24 25		Ca:	se Control Ma gregate <u>h</u> ogonal Desi it into Filos	tching		•	2	 Svoji praci bych jen Přesčas v pondělí [p Přesčas ve středu [Přesčas v pátek [pre pohlaví numerická [
26 27 28		Col	nt into Files py <u>D</u> ataset it <u>F</u> ile					Current Status: Do not filter cas
29 30 31 32		₩ <u>S</u> el ₩ <u>W</u> e T Sin	ect Cases ight Cases nulate Active E	Dataset				ОК

Select cases

23 _ elect All cases If condition is satisfied <u>I</u>f.... Random sample of cases Sample .. Based on time or case range Range. Use filter variable: utput Eilter out unselected cases Copy selected cases to a new dataset Dataset name: Delete unselected cases es

Reset

Cancel

Help

- 1. Data \rightarrow Select cases
- 2. If condition is satisfied

Co to dělá:

- Vybere z našeho celé vzorku takové respondenty, které splňují naše nastavené podmínky
- Každé další příkazy (statistiky), které SPSS zadáme, pak provádí pouze na vybrané subskupině



Select cases

Chceme zobrazit četnosti pro kategorii nejmladších respondentů

 naklikneme proměnnou Vek_kat pomocí šipky

2. specifikujeme podmínku

Vek_kat je numerická

Vek_kat = 1

3. continue a ok

4. následně znovu vyjedeme četnosti (Analyze → Descriptive Statistics →
Frequencies), které se pvní zobrazí jen pro

Select cases

V datasetu v "Data view" vidíme nepouží respondenty (jsou vyškrtnutí)

- Pokud chceme opět pracovat s celým vzorkem, podmínku zrušíme přes příkaz
- Data \rightarrow Select cases \rightarrow All cases

			_				_
	🔁 H			🛛 🎽		r h	
		ID	plat	věk	mesto	gender	in
		1	22900	31	Vyškov	m	
	2	2	28300	37	Vyškov	m	
	3	3	24300	33	Vyškov	f	
		4	27400	36	Brno	m	
	5	5	29200	37	Vyškov	m	
	6	6	13500	22	Brno	m	
		7	32500	47	Brno	m	
	8	8	90100	45	Brno	m	
1	9	9	17900	26	Brno	f	
	10	10	18300	26	Brno	f	
		11	28200	37	Vyškov	m	
	12	12	25000	34	Brno	m	
	12	12	149100	E1	Pres	-	

View Data Transform Analyze Direct Marketing

Graphs

Split file

- 1. Data → Split file
- 2. Compare groups
- 3. Vložíme proměnnou, podle které se má output rozdělit
 - Tj. Pokud podle věkových kategorií, tak vek_kat

Co to dělá:

- Rozdělí vzorek podle zadané proměnné
- V outputu pak vviede výsledek pro sub-skupinv zvlášt
 Split File

ZRUŠENÍ →





TAKŽE JAKÉ JE ŘEŠENÍ NAŠÍ ÚVODNÍ OTÁZKY, ZDA SE PRŮMĚRNÁ DŮVĚRA K TELEVIZI BARRANDOV LIŠÍ S OHLEDEM NA VĚK RESPONDENTA?