

VALICON

IBM SPSS STATISTICS

Predstavitev programskega pakete in brezplačne razširitve Predictive Solutions Imago

Naslov

Ime: Uroš Valant

E: uros.valan@valicon.net

T: 041 365 598

IZJAVA O VAROVANJU OSEBNIH PODATKOV ANKETIRANCEV

Družba Valicon je v skladu s kodeksom ESOMAR in AAPOR zavezana k varovanju osebnih podatkov anketirancev. Valicon tako v vseh pogledih preprečuje, da bi bilo mogoče prepoznati identiteto anketirancev. Vse spremenljivke oziroma polja, ki bi lahko neposredno kazala na identiteto anketiranca, so iz podatkovnih baz in poročil odstranjena. Prav tako so odgovori anketirancev fizično ločeni od podatkov anketirancev. Vsak poskus namerne identifikacije anketiranca ali razkritje identitete anketiranca s strani naročnika ali družbe Valicon pomeni kršitev zgoraj omenjenih kodeksov.

ZAUPNOST PONUDBE

Ponudba je last raziskovalne družbe Valicon. Naročnik ne sme razkriti informacij iz te ponudbe tretji osebi brez vnaprejšnjega pismenega dovoljenja družbe. Edina izjema je naročnikov naročnik. V nobenem primeru naročnik ne sme razkriti metodologije ali vrednosti projekta drugi raziskovalni agenciji.

IBM SPSS in analitični proces

Načrtovanje

- Sample Power
- SPSS Complex Samples
- SPSS Conjoint

1

Dostop do podatkov

- SPSS Statistics Base

2

Upravljanje in priprava podatkov

- SPSS Data Preparation
- SPSS Missing Value

3

Analiza podatkov

- SPSS AMOS
- SPSS Advanced Statistics
- SPSS Categories
- SPSS Exact Tests
- SPSS Forecasting
- SPSS Neural Networks
- SPSS Regression
- SPSS Statistics Base
- SPSS Direct Marketing

4

6

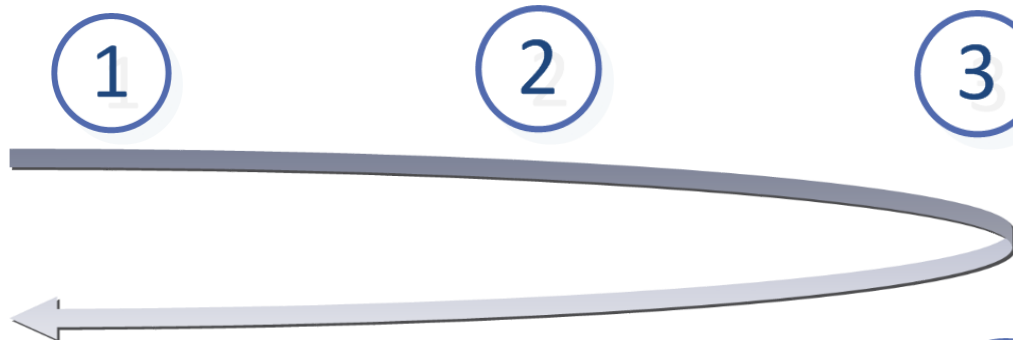
Uporaba/proces

- SPSS C&DS

Poročanje

- SPSS Custom Tables

5



IBM SPSS paketi in moduli



IBM SPSS Base Modul

- **IBM SPSS Base** ima osnovne značilnosti, ki so potrebne za izvedbo celotnega analitičnega procesa, od njegovega začetka do konca:
 - Nudi vam podporo v vsakem koraku analitičnega procesa
 - Analize lahko izvajate s pomočjo **intuitivnega grafičnega vmesnika**
 - Za lažjo in hitrejšo izvedbo specializiranih analiz lahko izbirate med več kot desetimi integriranimi proizvodi
- **Poslovne prednosti:**
 - Podpora poslovnim odločitvam s pomočjo analitike
 - Bodite prepričani v svoje rezultate, s pomočjo povezovanja podatkov iz različnih virov in uporabo preverjenih analitičnih metod
 - Z avtomatizacijo procesov prihranite čas in trud
 - Svojim rezultatom dodajte večjo težo s pomočjo različnih vizualnih prikazov
 - Modul IBM SPSS Statistics Base nudi tudi možnost programiranja s pomočjo SPSS Syntax, SPSS Sax Basic, Python plug-in, R plug-in, .Net plug-in, GPL.

IBM SPSS Base Modul

- V okviru modula Base lahko izvajate **naslednje analize**:
 - Frekvence
 - Opisne statistike
 - Srednje vrednosti
 - T test
 - Varianca
 - Bivariatne korelacije
 - Delne korelacije
 - Linearna regresija
 - Ordinalna regresija
 - Analiza najbližjega soseda
 - Diskriminantna analiza
 - Faktorska analiza
 - Razvrščanje v skupine
 - ROC krivulja

IBM SPSS Regression Modul

- **IBM SPSS Regression** vam pojasni in pomaga razumeti **odnose med spremenljivkami** ter napovedati njihovo obnašanje. Le s par kliki lahko hitro ustvarite napovedne modele. Neglede na to, s katerim področjem se ukvarjate, lahko na svojih podatkih uporabite različne sofisticirane modele, ki so na voljo v IBM SPSS Regression:
 - **Multinomialna logistična regresija** - Napovedi na osnovi več kot dveh kategorij
 - **Binarna logistična regresija** - Uporablja se, ko imamo opravka z dihotomnimi odvisnimi spremenljivkami
 - **Nelinearna regresija** - Za delo z nelinearnimi modeli
 - **Ponderiranje najmanjših kvadratov** - Metoda, koristna v eksperimentalnih študijah, kadar delamo na manjšem obsegu podatkov.
 - **Probit analiza**

Modul IBM SPSS Advanced Statistics

- S pomočjo **IBM SPSS Advanced Statistics** lahko:
 - Analizirate kompleksne povezave med spremenljivkami
 - Ustvarite fleksibilne modele
 - Z uporabo širokega spektra tehnik modeliranja ustvarite natančne napovedne modele,
 - Analizirate rezultate s pomočjo različnih metod
- Analize kompleksnih povezav med spremenljivkami, dostopne v modulu Advanced Statistics:
 - Multivariatni splošni linearni modeli
 - Komponente variance
 - Linearni modeli
 - Loglinearni modeli
 - Kaplan-Meierjeva analiza preživetja
 - Coxova regresija

IBM SPSS Conjoint

- **Modul IBM SPSS Conjoint** vam omogoča vpogled v proces odločanja pri potrošniku in vam pomaga oceniti, katere značilnosti določenega izdelka ali storitve so potrošniku najpomembnejše. Kadarkoli se soočite s težavo pri izbiranju med različnimi možnostmi, med katerimi bodo morali ljudje izbirati, vam pri sprejemanju prave odločitve lahko pomaga **IBM SPSS Conjoint**:
 - **Orthoplan** - Ustvarite ortogonalen spekter kombinacij, kar bistveno zmanjša število vprašanj, ki jih morate postaviti, da bi dobili informacijo potrebno za izvedbo celovite analize.
 - **Plancards** - Orodje za hitro generiranje kartic, ki se pri anketirancih uporabljajo za lažje razvrščanje značilnosti izdelkov.
 - **Conjoint** - Postopek Conjoint izvaja posebne regresije na vaši kategorizaciji odgovorov. Dobili boste učinkovite rezultate, s katerimi je mogoče na primer videti, katere značilnosti so kupcem najvažnejše in kateri kupci jim dajejo prednost. Prav tako boste lahko izvajali simulacije za ugotavljanje naklonjenosti trga katerikoli značilnosti izdelka.

IBM SPSS Neural Networks

- **Nevronske mreže** delujejo na podlagi učenja povezave med podatki (neodvisne spremenljivke) in rezultati (odvisne spremenljivke). Tipično algoritem razdeli bazo podatkov na podzorec, na katerem poteka učenje, in preostali del podatkov, kjer se preverja veljavnost naučenega. Prednost nevronskih mrež je, da lahko kot odvisne ali neodvisne spremenljivke nastopajo spremenljivke katerihkoli tipov – lestvice, ordinalne ali nominalne spremenljivke. S pomočjo posebne analize pa lahko ugotovimo tudi relativno pomembnost vsake izmed neodvisnih spremenljivk.
- **IBM SPSS Neural Networks** vsebujejo dve metodi:
 - **Multilayer Perception (MLP)** – je uporabna pri reševanju raznolikih problemov, dobro posplošuje rešitve vendar zaradi tega nujno ne najde optimalne rešitve, ter je nekoliko bolj časovno potratna od RBFN metode.
 - **Radial Basis Function Network (RBFN)** – je hitrejša od MLP in se bolje vede pri podatkih, katerih vrednosti so združene v skupine, vendar je slabša pri posploševanju rezultatov in bolj zahtevna za uporabo.

IBM SPSS Data Preparation

- **IBM SPSS Data Preparation** vam omogoča boljšo pripravo podatkov in točnost vaših rezultatov.
- Izberite postopke s katerimi boste hitrejši pri pridobivanju rezultatov ali pa postopke, ki vam bodo pomagale obdelati in pripraviti podatke na popolnoma avtomatiziran način:
 - **Automated Data Preparation (ADP)** - Avtomatska priprava podatkov, ki so koristne tako za izvedbo analiz kot iskanje in odpravljanje napak ter kakovostno zapolnitev manjkajočih vrednosti, vse z enim samim korakom.
 - **Data Validation** - Postopek, ki vam zagotovi konsistentnost podatkov na osnovi pravil, ki definirajo ali je neka vrednost pravilna ali napačna.
 - **Anomaly Detection** - Postopek, ki se nanaša na odkrivanje nenavadni vrednosti v določenem sklopu podatkov, ki niso v skladu z ostalimi podatki.
 - **Optimal Binning** - Postopek, ki vam avtomatsko iz neodvisne razmernostne (scale) spremenljivke naredi kategorično spremenljivko. Število in širina kreiranih kategorij je odvisna od odvisne spremenljivke. Postopek je koristen za vse statistične procedure, ki zahtevajo kategorične spremenljivke kot neodvisne spremenljivke.

IBM SPSS Categories

- **Z IBM SPSS Categories** lahko delate različne regresije, ko imate numerične, nominalne ali ordinalne spremenljivke. Podatke lahko vizualno interpretirate, kar pomeni, da lahko zelo razločno vidite, v kakšnem odnosu so vrstice in stolpci v tabelah z rezultati:
 - **Kategorična regresija (CATREG)** – npr. za predvidevanje zadovoljstva z delom na osnovi vrste dela, geografske regije in števila poslovnih potovanj.
 - **Korespondenčna analiza (CORRESPONDENCE)** - omogoča vizualni prikaz odnosa med spremenljivkami, jakosti njihove povezanosti, ipd.
 - **Multipla korespondenčna analiza (MULTIPLE CORRESPONDENCE)** – npr. za raziskovanje odnosa med priljubljeno TV oddajo, starostjo in spolom.
 - **Kategorična analiza glavnih komponent (CATPCA)** – npr. za prikaz odnosa med različnimi znamkami avtomobilov in njihovimi značilnosti kot so cena, teža, poraba goriva, itd.
 - **Nelinearna kanonična korelacijska analiza (OVERALS)** – npr. za analizo značilnosti izdelkov, na primer različnih okusov juh.
 - **Določanje preferenc (PREFSCAL)** – npr. za analizo podobnosti različnih okusov pijač, pri anketirancih različnih starostnih skupin, kjer lahko pridemo do zaključka, da mlajši anketiranci poudarjajo razliko med klasičnimi in novimi okusi, medtem ko starejši anketiranci pogosteje razlikujejo med dietetičnimi in nedietetičnimi okusi.
 - **Večrazsežno lestvičenje (PROXSCAL)** – npr. da skupini voznikov damo nalogo, da ocenijo 26 različnih vrst avtomobilov, glede na deset njihovih značilnosti, ki jih morajo oceniti na 6-stopenjski lestvici. Dobili bomo načrt skupin modelov avtomobilov, ki so si med seboj podobni in ki so najbolj privlačni.

IBM SPSS Exact Tests

- Z modulom **IBM SPSS Exact Tests** lahko namesto na velikih in dragih vzorcih pridete do verodostojnih rezultatov tudi na majhnih ($n < 30$). Obenem ni potrebno, da se odrečete dragocenim informacijam, ki bi jih zgubili z združevanjem kategorij z manj odgovori v večje skupine, s čimer bi lahko uporabljali klasične statistične metode temveč ohranite izvirno obliko baze in njene naravne kategorije ter analizirate, kar ste nameravali.
- Na voljo je več kot **30 različnih testov**, ki pokrivajo celoten spekter neparametričnih in kategoričnih problemov, tako za večje kot manjše podatkovne baze:

- Pearson Chi- square test
- Likelihood ratio test
- Fisher's exact test
- Linear-by-linear association test
- Contingency coefficient
- Phi
- Cramer's V
- Goodman and Kruskal Tau
- Uncertainty coefficient—symmetric or asymmetric
- Kappa
- Gamma
- Kendall's Tau-b and Tau-c
- Somers' D—symmetric and asymmetric
- Pearson's R
- Spearman correlation
- McNemar test
- Sign test
- Wilcoxon signed-rank test
- Marginal homogeneity test
- 2-Sample Kolmogorov-Smirnov test
- Mann-Whitney U or Wilcoxon rank-sum W test
- Wald-Wolfowitz runs test
- Cochran's Q test
- Friedman test
- Kendall's coefficient of concordance
- Kruskal-Wallis test
- Median test
- Jonckheere-Terpstra test
- Sample Chi-square test
- Sample Kolmogorov-Smirnov test
- Sample Wald-Wolfowitz runs test
- Binomial test

IBM SPSS Bootstrapping

- **IBM SPSS Bootstrapping** vam pomaga oblikovati zanesljivejše modele, ki zagotavljajo natančnejše rezultate. Modeli, ki jih oblikujete, igrajo pri odločanju v organizaciji pomembno vlogo. Lahko se uporabljajo za oblikovanje javnih politik, preprečujejo širitev bolezni ali pa določajo ali bo podjetje izpeljalo določeno investicijo. Zato je pomembno, da so **modeli stabilni in vodijo do natančnih in zanesljivih rezultatov**.
- **IBM SPSS Bootstrapping** je modul, ki vam bo pomagal ustvariti **stabilne in zanesljive modele**. Za preverjanje stabilnosti modela ni dovolj, da model preverite le na enem izbranem vzorcu, ki ste ga izbrali iz populacije, saj so rezultati odvisni od tega vzorca. Več vzorčenj vam nudi natančnejše ocene in zanesljivost vaših podatkov.
- Z modulom **IBM SPSS Bootstrapping** svojega modela ne boste preverjali le na enem vzorcu, **temveč bo ta za vas pripravil več tisoč vzorcev** in vam tako omogočil natančen vpogled v celotno populacijo. Pomagal vam bo tudi pri odpravljanju problema z anomalijami, zaradi česar boste na koncu imeli bolj jasno sliko o kakovosti vašega modela.

IBM SPSS Missing Values

- S pomočjo modula **IBM SPSS Missing Values** lahko dopolnite manjkajoče podatke in pridete do veljavnih zaključkov.
- S pomočjo različnih modelov vam orodje napravi nabor različnih baz (običajno 3 do 5), vsako z drugačnimi nadomestnimi vrednostmi pri spremenljivkah pri katerih ste se odločili za nadomestitev manjkajočih vrednosti z izračunanimi ocenami. Nato lahko modelirate različne skupine podatkov z uporabo običajnih tehnik, kot je na primer linearna regresija in napravite ocene parametrov za vsako skupino podatkov ter končno oceno parametrov:
 - **Analiza vzorca** - Tabela vzorcev v podatkih, na osnovi katere se lažje odločimo, pri katerih spremenljivkah bomo nadomestili manjkajoče vrednosti z ocenami in pri katerih ne.
 - **Statistične analize** - Univariatne, listwise in pairwise.
 - **Multiple imputacije** - Avtomatična, monotona in FCS (določena s pogoji)
 - **Analize:**
 - Deskriptivne: frekvenca, deskriptivna analiza, križne tabele, korelacije, neparametrične korelacije, parcialne korelacije;
 - Primerjava povprečij: povprečna vrednost, t-test, neparametrični testi, one-way ANOVA, univariatna ANOVA;
 - Modeli: splošni linearni modeli (GLM), linearna regresija, multinominalna logistična regresija, binarna logistična regresija, ordinalna regresija, linearni mešani modeli;
 - Coxova regresija
 - **Pooling** - Sestavljanje končne rešitve imputiranih podatkov sestavljene iz različnih rešitev imputiranih podatkov.

IBM SPSS Custom Tables

- **IBM SPSS Custom Tables** vam omogoča boljše razumevanje podatkov, učinkovitejšo predstavitev rezultatov ter kakovostnejše posredovanje informacij tistim, ki na njihovi osnovi sprejemajo odločitve:
 - **Interaktivno ustvarjanje tabel** – Omogočen vam je predogled tabel v realnem času, tako lahko v vsakem trenutku veste, kako bo izgledala vaša končna tabela.
 - **Upravljanje s kategorijami spremenljivk** – privarčujte čas in trud z ustvarjanjem novih polj neposredno v IBM SPSS Custom Tables, za potrebe računskih operacij (kot npr. zbiranje, delitev, delitev odstotkov).
 - **Testiranje značilnosti razlik** med kategorijami v tabeli, s testi hi-kvadrat, test neodvisnosti, T-test in Z-test.
 - **Enostaven izvoz podatkov** iz tabel v različne programe.
- Informacije v tabelah lahko predstavite v različnih stilih in oblikah, da bodo ustrezale celotni grafični podobi organizacije ali predpisanim strokovnim standardom. Različni stili se lahko uporabljajo za različne tipe uporabnikov, od začetnika do strokovnjaka in njihovim zahtevam primerno. Rezultati so tako lahko uporabljeni brez dodatnega dela in dodatnih popravkov.

IBM SPSS Forecasting

- Orodje IBM SPSS Forecasting ponuja mnogo možnosti, ki omogočajo hitro razvijanje zanesljivih napovedi z uporabo podatkov časovnih vrst:
 - **TSMODEL** – Preko funkcije Expert Modeler lahko izbor modelov omejite le na ARIMA modele ali na eksponencialno glajene modele. V Expert Modeler-ju lahko tudi sami nastavite vsak posamezen parameter modela. Pri tem lahko uporabite nasvete, ki vam jih ponuja Expert Modeler kot začetno točko za vaše izbore, ali pa za preverjanje svojega dela.
 - **TSAPPLY** – Modeli se shranijo v format PMML. Ko pride do sprememb v podatkih, se lahko napovedi izvedejo avtomatsko, saj skripte omogočajo nadgradnjo brez ponovne nastavitve parametrov in ocenjevanja modela.
 - **SEASON** – Ocenite dodatne sezonske vplive za periodične časovne vrste.
 - **SPECTRA** – Razčlenite časovne vrste na njihove harmonične komponente – niz regularnih periodičnih funkcij z različno dolgimi časovnimi razdobji.

IBM SPSS Complex Samples

- Z IBM SPSS Complex Samples je delo in razumevanje rezultatov raziskav na kompleksnih vzorcih preprosto. Preko intuitivnega vmesnika lahko analizirate podatke in tolmačite rezultate.
- V modulu IBM SPSS Complex Samples vam je na voljo:
 - **Stratificirano vzorčenje** - Povečajte natančnost vašega vzorca ali pa si zagotovite reprezentativni vzorec iz ključnih skupin, z izbiranjem v vzorec iz posameznih podskupin v populaciji. Podskupino lahko predstavlja na primer določeno število moških in žensk, ljudje z določenih delovnih področij ali starostnih skupin, itd.
 - **Vzorčenje v skupinah** - Izberite skupine, ki so lahko šole, bolnice, geografska področja z enotami kot so učenci, pacienti ali občani. Analize na skupinah so pogosto stroškovno učinkovitejše.
 - **Večstopenjsko vzorčenje** - V prvi stopnji na osnovi skupin elementov v populaciji izberete začetni vzorec, nato v drugi stopnji ustvarite vzorec z izbiranjem podvzorca iz vsake enote iz prve stopnje. S ponavljanjem omenjenega postopka, tako lahko izbirate vzorce na višjih stopnjah.

Licence, ki so na voljo za SPSS Statistics

- **Uporabniška licenca (Authorisation license)**

- V primeru SPSS Statistics gre za **uporabiške licence**, kjer je vsaka licenca vezana na posamezni računalnik. Licenčna koda je sestavljena iz avtorizacijske kode, ki jo izda IBM in pa konfiguracije računalnika. Zaradi tega licenca ni prenosljiva in je vezana na računalnik, kamor je inštaliran SPSS Statistics. V primeru menjave računalnika je potrebno ponoviti postopek licenciranja. Delovanje programa je vezano na računalnik in deluje, tudi če ni internetne povezave

- **Mrežna licenca (Concurrent license)**

- Mrežna licenca omogoča **poljubno količino inštalacij** znotraj računalniške mreže. Konekcije nadzira poseben strežnik (sentinel). Naenkrat lahko s programom SPSS Statistics dela le **toliko uporabnikov, kolikor mrežnih licenc je kupljenih**. V primeru, da bi radi delali s SPSS izven mreže, recimo doma, se naredi začasni „check-out“, kjer se za določen čas eno mrežno licenco umakne iz uporabe in se z njo dela na drugi lokaciji.

Vrste licenc glede na trajanje

- **Letna licenca ali najem programa:**

- Letna licenca omogoča podjetju oziroma instituciji uporabo licenciranega orodja na **letni osnovi**. Licenco je potrebno podaljšati po preteku enega leta. S podaljšanjem licence program deluje za novo obdobje enega leta, v podaljšanje licence pa je avtomatično vključena tudi tehnična pomoč in nove verzije licenciranih orodij, ki pridejo na tržišče v tekočem letu podaljšanja licence.

- **Neomejena licenca ali nakup programa**

- Gre za **časovno neomejeno** uporabo licenciranega orodja za določeno število uporabnikov. V nakupno ceno je vključeno tudi eno (prvo) leto tehnične, ter tudi popravki in nove verzije licenciranega orodja za prvo leto po nakupu licence. V kolikor želite biti tehnične podpore in vseh pridobitev, ki sodijo zraven, deležni tudi po preteku prvega licenčnega leta, je potrebno mehanizem tehnične podpore ponovno aktivirati. Zaradi neomejenega tipa licence vam programska oprema še vedno nemoteno deluje. V primeru, da se za aktivacijo tehnične podpore po preteku prvega licenčnega leta ne odločite, ostanete na kupljeni verziji programa, a brez brezplačnih nadgradenj in tehnične podpore s strani Valicon in IBM.

Kaj je tehnična podpora?

- Tehnična podpora SPSS uporabnikom omogoča, da smo vam vedno na voljo in v pomoč pri razreševanju vaših **tehničnih težav** s programom IBM SPSS Statistics. Kar v praksi pomeni:
 - podpora pri nameščanju programske opreme
 - interpretacija napak, ki jih sporoča program
 - ocena napake
 - iskanje in detekcija napak in specifičnih problemov.
- V okviru plačane tehnične podpore vam pripadajo tudi vse nove verzije in vsi popravki programa, ki pridejo na tržišče v letu plačane tehnične podpore. **V primeru resnejših zastojev delovanja programske opreme, se povežemo direktno s strokovnjaki družbe Predictive Solutions, ki so sposobni rešiti 99% vseh tehničnih zahtevkov. V ekstra primerih se nato problem prenese na IBM oziroma na njihovo razvojno ekipo. S tem, ko imamo za drugi nivo strokovnjake družbe Predictive Solutions, smo pridobili na odzivnosti in skrajšali čas do rešitve problema.**

VALICON

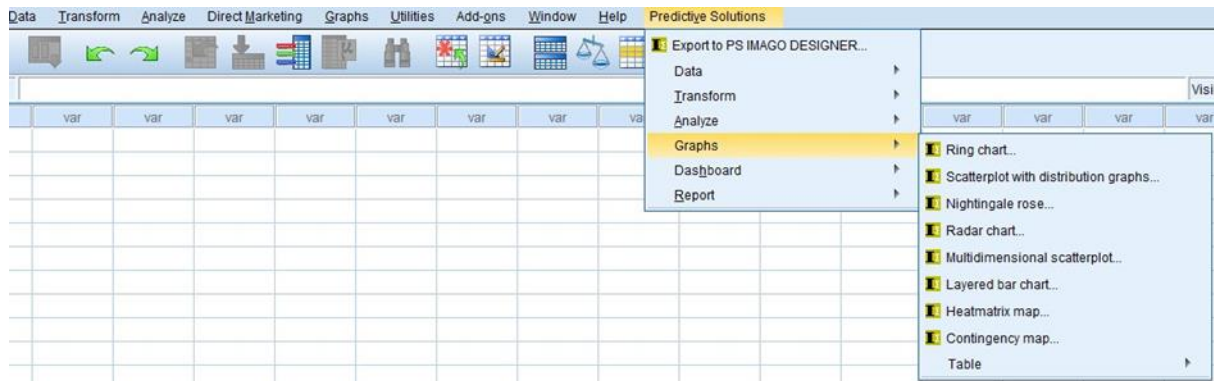
Predictive Solutions Imago

Razširitev IBM SPSS Statistics

Vse lastnosti paketov in modulov ostanejo enake, razširitev je v programu vidna kot dodatna menijska izbira

Razširjena različica – PS Imago

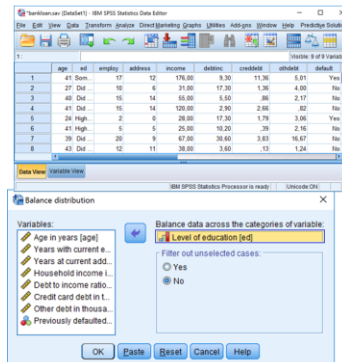
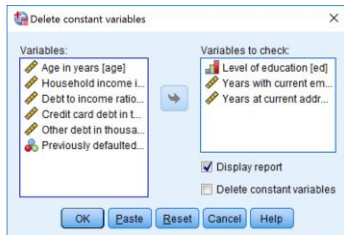
- **Predictive Solutions Imago** je dejansko **IBM SPSS Statistics**, v katerega je vključena dodatna nadgradnja, ki še razširi uporabnost SPSS Statistics programskega paketa. Gre za zbir postopkov za pripravo in transformacijo podatkov, dodatnih statističnih metod in izjemno razširjenih možnosti izdelave poročil ter prikazovalnikov, ki so jih uporabniki pri rednem delu s SPSS Statistics pogrešali.
- Nadgradnja je vključena v vse verzije **IBM SPSS Statistics** in je brezplačna



PS IMAGO – dodatni postopki za pripravo podatkov

- **Data variable duplicates** – pobriše podvojene vrednosti (ne primerov) glede na podobnost
- **Data inventory** – dobimo strukturo datotek v direktoriju
- **Data description** – definiramo lahko podrobne podatke o strukturi izbranih datotek
- **Delete constant variable** – pobrišemo spremenljivke, ki so konstante (vse vrednosti so enake)
- **Balance distribution** – uravnotežene kategorizacijske spremenljivke (distribucija je enakomerna čez vse kategorije)

	age	sex	employ	address	income	debtinc	creditde	otherdeb	default
1	41	Male	17	12	178,00	9,30	11,36	6,01	Yes
2	27	Male	10	6	31,00	17,30	1,36	4,00	No
3	40	Male	16	14	50,00	5,50	2,86	2,17	No
4	41	Male	10	14	100,00	2,90	2,00	8,02	No
5	24	High	2	0	28,00	17,30	1,79	3,06	Yes
6	41	High	5	5	25,00	10,20	2,9	2,16	No
7	39	Male	20	8	67,00	30,60	3,83	16,67	No
8	43	Male	12	11	38,00	3,60	13	1,24	No



		Level of education			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not complete high school	400	54,1	54,1	54,1
	High school degree	235	27,6	27,6	81,8
	Some college	101	11,9	11,9	93,6
	College degree	49	5,8	5,8	99,4
	Post-undergraduate degree	5	,6	,6	100,0
	Total	850	100,0	100,0	

RESULT- balanced data

		Level of education			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not complete high school	5	20,0	20,0	20,0
	High school degree	5	20,0	20,0	40,0
	Some college	5	20,0	20,0	60,0
	College degree	5	20,0	20,0	80,0
	Post-undergraduate degree	5	20,0	20,0	100,0
	Total	25	100,0	100,0	

PS Imago – dodatni postopki za transformacije

- **Recode infrequent categories** – združevanje kategorij z majhnim številom primerov
- **Normalization of variables** – Normalizacija spremenljivk
- **Recode categories monotonically** – avtomatično rekodiranje spremenljivk
- **Compute global value** – Izračun skupne statistike (recimo mediana) in povraten zapis v podatkovni nabor za nadaljnjo analizo
- **Creating dummy variables** – Kreiranje spremenljivk, ki temeljijo na vrednosti kategorij

age	sex	address	income	ethnic	household	vhscale	default	age_MEDIAN	age_MEDIAN_1
1	1	1	12	1	1	1	1	1	1
2	1	1	12	1	1	1	1	1	1
3	1	1	12	1	1	1	1	1	1
4	1	1	12	1	1	1	1	1	1
5	1	1	12	1	1	1	1	1	1
6	1	1	12	1	1	1	1	1	1
7	1	1	12	1	1	1	1	1	1
8	1	1	12	1	1	1	1	1	1
9	1	1	12	1	1	1	1	1	1
10	1	1	12	1	1	1	1	1	1
11	1	1	12	1	1	1	1	1	1
12	1	1	12	1	1	1	1	1	1
13	1	1	12	1	1	1	1	1	1
14	1	1	12	1	1	1	1	1	1
15	1	1	12	1	1	1	1	1	1
16	1	1	12	1	1	1	1	1	1
17	1	1	12	1	1	1	1	1	1
18	1	1	12	1	1	1	1	1	1
19	1	1	12	1	1	1	1	1	1
20	1	1	12	1	1	1	1	1	1

Compute global value

Compute global values for:

age_MEDIAN_1

OK Cancel Help

Recode into dummy variables

Source variable: age_MEDIAN_1

Number of categories in coded variable: 2

Level of education (cat)

OK Cancel

Dichotomous coding

Source variable: age_MEDIAN_1

Number of categories in coded variable: 2

Level of education (cat)

OK Cancel

income	ethnic	vhscale	e1_CODE_1	e1_CODE_2	e1_CODE_3	e1_CODE_4	age
1	1	1	0	0	1	0	0
2	1	1	0	0	1	0	0
3	1	1	0	0	1	0	0
4	1	1	0	0	1	0	0
5	1	1	0	0	1	0	0
6	1	1	0	0	1	0	0
7	1	1	0	0	1	0	0
8	1	1	0	0	1	0	0
9	1	1	0	0	1	0	0
10	1	1	0	0	1	0	0
11	1	1	0	0	1	0	0
12	1	1	0	0	1	0	0
13	1	1	0	0	1	0	0
14	1	1	0	0	1	0	0
15	1	1	0	0	1	0	0
16	1	1	0	0	1	0	0
17	1	1	0	0	1	0	0
18	1	1	0	0	1	0	0
19	1	1	0	0	1	0	0
20	1	1	0	0	1	0	0

PS Imago – dodatni statistični postopki

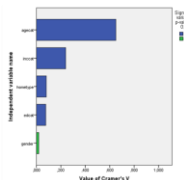
- Izračun **Cramerjeve V** korelirane vrednosti (Cramer's V correlated values) – katere kategorične spremenljivke so najboljši prediktorji za odvisne kategorične spremenljivke
- Mere neenakosti (**inequality measures**) – katere kategorične spremenljivke so najboljši prediktorji za odvisne numerične spremenljivke
- Significant variables **Chi-square** – iskanje prediktorjev za odvisne kategorične spremenljivke z metodo Chi-kvadrat
- Significant variables **CHAID** - iskanje prediktorjev za odvisne kategorične spremenljivke z metodo dreves

1	Age	Sex	Employ	Address	Income	Debt	Credit	Default
1	41	Female	17	12	105,00	9,30	10,30	5,00
2	27	Male	6	5	31,00	17,30	9,30	4,00
3	48	Male	16	14	85,00	5,60	8,00	2,00
4	41	Male	16	14	100,00	2,90	2,80	8,00
5	34	Male	2	8	28,00	17,30	1,70	3,80
6	41	Male	5	5	25,00	10,20	3,00	2,10
7	30	Male	8	8	17,00	20,60	3,60	16,60
8	43	Male	12	11	38,00	3,60	1,00	1,20

List of correlated predictors - Cramer's V

Independent variable name	Independent variable label	Value	Approximate Significance
1 agecat	Age category	.650	.000
2 peccat	Income category in thousands	.239	.000
3 homeage	Building type	.081	.000
4 edcat	Level of education	.076	.000
5 gender	Gender of name on account	.019	.694

Dependent variable --> addresscat 'years at current address'



Cramer's V correlated variables

Variables	Dependent variable
Age in years (age)	Previously defaulted
Years with current add.	Level of education
Household income L.	
Debt to income ratio.	
Credit card debt in L.	
Other debt in thous.	

1	Age	Sex	Employ	Address	Income	Debt	Credit	Default
1	41	Female	17	12	105,00	9,30	10,30	5,00
2	27	Male	6	5	31,00	17,30	9,30	4,00
3	48	Male	16	14	85,00	5,60	8,00	2,00
4	41	Male	16	14	100,00	2,90	2,80	8,00
5	34	Male	2	8	28,00	17,30	1,70	3,80
6	41	Male	5	5	25,00	10,20	3,00	2,10
7	30	Male	8	8	17,00	20,60	3,60	16,60
8	43	Male	12	11	38,00	3,60	1,00	1,20

Inequality measures

Dependent variable: Debt to income ratio (x100)

Independent variables:

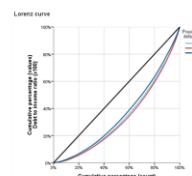
- Age in years (age)
- Level of education
- Years with current add.
- Years at current add.
- Household income L.
- Debt to income ratio.
- Credit card debt in L.
- Other debt in thous.

Options:

- Chi-square statistic (checked)
- Get residuals (optional)
- Number of classes: 10

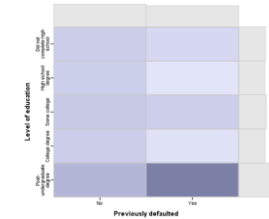
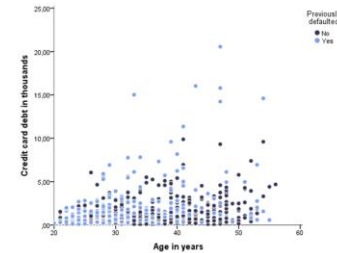
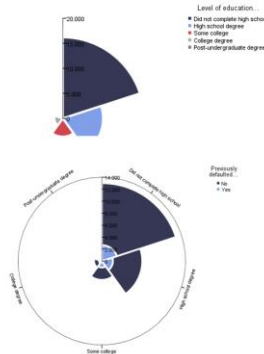
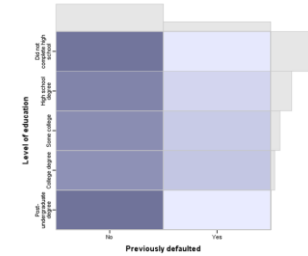
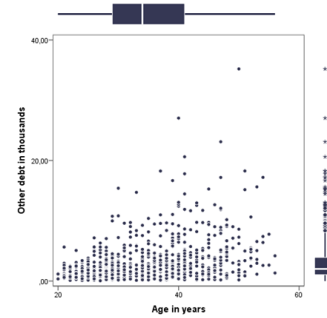
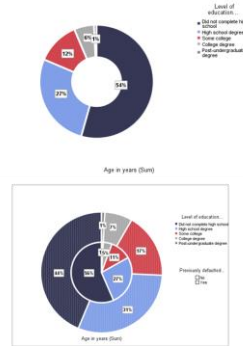
Inequality measures - indexes: Debt to income ratio (x100)

Previously defaulted	Gen	Home	Culture	Threat of OC (cat=1)	Threat of OC (cat=0)
1 No	.34	.24	.04	.21	.19
2 Yes	.35	.26	.03	.21	.20
3 Yes	.30	.21	.03	.16	.15



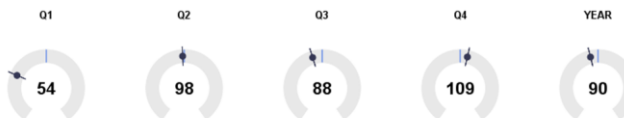
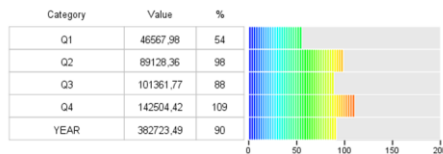
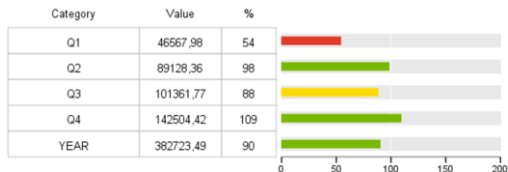
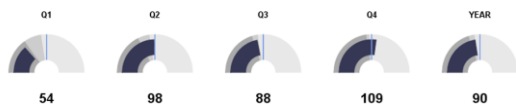
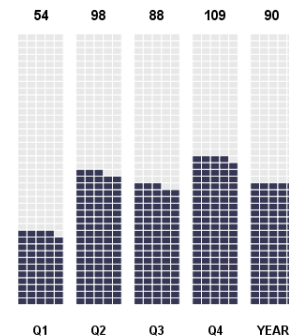
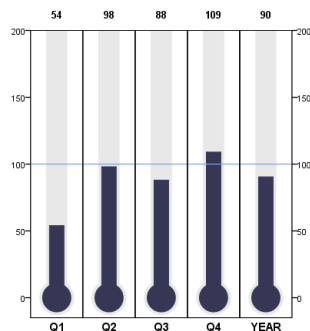
PS Imago – dodatni grafi

- Ring chart
- Scatterplot with distributed graphs
- Nightingale rose
- Multidimensional scatterplot
- Layered bar chart
- Heatmatrix map
- Contingency map
- Različne tablele



PS Imago – nadzorna plošča (Dashboard)

- Odstotki glede na cilj
- Odstotki glede na prag cilja
- Primerno za dnevni pregled napredka na projektih ali za spremljanje podatkov, ki se redno zbirajo



VALICON

ESOMAR
member

Advanced
Business
Partner



Authorized IBM Cloud
Solution Provider - IBM SaaS
Solution Provider - IBM SaaS

**Authorized
Software Value Plus**
Business Analytics

Ponudbe so veljavne do 31.8.2016

VALICON Ljubljana

Kopitarjeva 2
1000 Ljubljana
T: +386 1 420 49 00
F: +386 1 420 49 60
info@valicon.si

VALICON Zagreb

Baruna Trenka16
10000 Zagreb
T: +385 1 640 99 55
F: +385 1 640 99 56
info@valicon.hr

VALICON Sarajevo

Branilaca Sarajeva 20
71000 Sarajevo
T: +387 33 258 655
F: +387 33 258 656
info@valicon.ba

VALICON Beograd

Gavrila Principa 16/2
11000 Beograd
T: +381 11 32 86 978
F: +381 11 30 30 444
info@valicon.rs