Napredno strukturiranje problema odlučivanja

Da bi metode odlučivanja koje su do sada obrađene dale što bolji krajnji rezultat, potrebno je da budu primenjene nad što je moguće bolje definisanim polaznim problemom (ulaznim podacima). Suština dobro postavljenog problema odlučivanja, uz odabir najprikladnije metode odlučivanja za datu situaciju, može se svesti na razrešenje sledećih pitanja:

1. **Da li postoji zavisnost među atributima?** Ako su atributi međusobno zavisni, trebalo bi odabrati skup najnezavisnijih kako bi se smanjio njihov broj, čime bi se umanjio negativan efekat suvišnog ponavljanja informacije i omogućilo se preciznije ocenjivanje težina od strane DO
2. **Na koji način su atributi međusobno uređeni?** Kroz pravilan izbor strukture atributa može se doći do daleko kvalitetnijih odluka, u odnosu na situaciju u kojoj je ovaj aspekt zanemaren.
3. **Kako proceniti vrednosti alternativa za kriterijume koji nam nisu dostupni, a bitni su za proces odlučivanja?** Moguće je proceniti vrednosti alternativa za kriterijum koji nam nije poznat, a imamo istorijske podatke odakle možemo proceniti ove vrednosti.

U nastavku skripte detaljno su obrađena ova tri aspekta kvalitetnog strukturiranja u procesu odlučivanja.

Merenje zavisnosti među atributima

Zamislimo da imamo dva kriterijuma koja su međusobno zavisna, odnosno, da postoji zakonitost u kojoj veća vrednost jednog kriterijuma, dovodi do veće vrednost drugog kriterijuma. Neka svaki kriterijum ima svoje pondere. Tada prilikom računanja najprihvatljivije alternative dolazimo u situaciju da istu informaciju računamo dva puta, odnosno, ako uzmemo težine (pondere) u obzir, to znači da istu informaciju dvostruko jače vrednujemo. Stoga, navodimo da je poželjna osobina problema odlučivanja da su kriterijumi međusobno nezavisni. Zavisnost između kriterijuma možemo da ispitamo pomoću kovarijanse, koeficijenta korelacije i putem linearne regresije.

Kovarijansa

Kovarijansa predstavlja meru zajedničkog varijabiliteta između promenjivih i računa se preko formule:

$$cov(x, y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

gde su x i y kriterijumi iz matrice odlučivanja, N veličinapopulacije, odnosno broj alternativa, a $\bar{x}$ i $\bar{y}$ prosečne vrednosti kriterijuma x i y. Ukoliko se za iznadprosečne vrednosti kriterijuma x javljaju iznadprosečne vrednosti kriterijuma y (dakle, u istom redu su iznadprosečne vrednosti i za x i za y), tada je kovarijansa pozitivna i kažemo da su kriterijumi pozitivno linearno zavisni. U suprotnom, ako većim vrednostima kriterijuma x odgovaraju niže vrednosti kriterijuma y, tada je kovarijansa negativna i kažemo da su kriterijumi negativno linearno zavisni. Zajednički varijabilitet kriterijuma sa samim sobom predstavlja varijansu tog kriterijuma. Glavni nedostatak kovarijanse je taj što dobijene vrednosti ne mogu intuitivno da se tumače . Međutim, ako je vrednost kovarijanse 0, onda možemo da kažemo da ne postoji linearna zavisnost između kriterijuma x i y. Takođe, možemo primetiti da su vrednosti simetrične, odnosno $cov(x, y) = cov(y, x)$.

Korelacija

Kako bismo dobili tumačenje iz kovarijanse, dobijene vrednosti „normalizujemo“. Ovaj postupak predstavlja postupak računanja koeficijenta korelacije. Koeficijent korelacije predstavlja meru koja pokazuje smer zajedničkog variranja dva kriterijuma, izraženu u intervalu [-1, 1]. Vrednost koeficijenta korelacije -1 predstavlja negativnu linearnu povezanost dva kriterijuma, odnosno, što je viša vrednost jednog kriterijuma, to će vrednost drugog kriterijuma biti niža. Sa druge strane, vrednost koeficijenta korelacije 1 predstavlja pozitivnu linearnu povezanost dva kriterijuma, odnosno, rast vrednosti jednog kriterijuma, uzrokovaće i rast vrednosti drugog kriterijuma. Koeficijent korelacije računa se preko formule:

$$cor(x, y) = \frac{cov(x, y)}{stdev(x) * stdev(y)}$$

Kao i kod kovarijanse, vrednosti su simetrične. Drugim rečima, $cor(x, y) = cor(y, x)$. Iako ne postoji formalno pravilo za tumačenje korelacija, možemo ih tumačiti na sledeći način. Ukoliko je koeficijent korelacije veći/manji od +/- 0,70, tada kažemo da postoji jaka pozitivna/negativna linearna zavisnost. Ako je koeficijent korelacije veći/manji od +/- 0,50, a manji/veći od +/- 0,70, tada kažemo da postoji umerena pozitivna/negativna linearna zavisnost. Ako je koeficijent korelacije veći/manji od +/- 0,30, a manji/veći od +/- 0,50 tada kažemo da postoji slaba pozitivna/negativna linearna zavisnost. Ako je koeficijent korelacije u intervalu (-0,30, 0,30), tada kažemo da je linearna zavisnost zanemarljiva, odnosno da nema linerane zavisnosti.[1]

Linearna regresija

Zavisnost između dva kriterijuma možemo ispitati i linearnom regresijom, tako što ćemo vrednosti jednog kriterijuma pokušati da opišemo preko vrednosti drugog; odnosno, želimo da vidimo kako promena vrednosti jednog kriterijuma utiče na promenu drugog. To se postiže rešavanjem sledeće jednačine:

$$y = b(x) * x$$

gde je y vrednost „zavisnog“ kriterijuma (posledice), x vrednost „nezavisnog“ kriterijuma (uzroka), a $b(x)$ koeficijent koji treba odrediti i koji definiše smer i intenzitet zavisnosti. Za razliku od kovarijanse i korelacije, vrednosti $b(x)$ za dva različita smera posmatranja, nisu simetrične (iste). Koeficijent $b(x)$ se dobija kada se kovarijansa podeli varijansom uzročne promenjive:

$$b(x) = \frac{cov(x, y)}{var(x)}$$

[1] Triola, M. F. (2005). *Essentials of statistics*. Pearson/Addison Wesley.

U idealnom slučaju problema odlučivanja, kriterijumi bi trebalo da budu nezavisni, odnosno, da vrednosti i kovarijanse i korelacije i koeficijenta $b(x)$ budu jednake 0 (ne treba ih sve računati). Kada su kriterijumi nezavisni, tada je lakše i jednostavnije (i ispravnije) oceniti težine kriterijuma. Drugim rečima, kada su kriterijumi međusobno zavisni, otežava se sposobnost DO da realno oceni težine kriterijuma.

Vrednosti kovarijanse, korelacije i regresije, koje se dobiju, služe za uočavanje problema zavisnosti.

Primer:

Ispitati zavisnosti koje postoje između kriterijuma u tabeli odlučivanja preko kovarijanse, korelacije i regresije. Da li postoje atributi koji su kandidati za uklanjanje?

	Cena (eur)	Internet	Udaljenost od grada	Čistoća
	min	max	min	max
A	55	3	0,7	4
B	65	1	0,4	3
C	40	0	0,7	4
D	25	2	4	3
E	40	1	2	5
Ponderi	0,35	0,2	0,3	0,15

Prvo ćemo izračunati matricu kovarijansi. Za računanje matrice potrebne su nam prosečne vrednosti za svaki kriterijum.

Cena (eur)	Internet	Udaljenost od grada	Čistoća
45	1,4	3	3,8

Zatim, pravimo matricu veličine kxk, gde je k broj kriterijuma koju ćemo popunjavati preko gore navedene formule. Kovarijansu kriterijuma *Cena* i *Internet* računamo na sledeći način:

$$cov(Cena, Internet) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$
$$= \frac{1}{5}\big((55-45)*(3-1{,}4) + (65-45)*(1-1{,}4) + (40-45)*(0-1{,}4) + (25-45)*(2-1{,}4) + (40-45)*(1-1{,}4)\big) = 1$$

Kovarijansa	Cena (eur)	Internet	Udaljenost	Čistoća
Cena (eur)	190	1	-15,7	-1
Internet	1	1,04	0,316	-0,12
Udaljenost	-15,7	0,316	1,7944	-0,168
Čistoća	-1	-0,12	-0,168	0,56

Posmatrajući kovarijacionu matricu, možemo primeriti da vrednosti nisu jednake 0, iako su vrednosti kovarijanse između nekih kriterijuma blizu 0 (vrednosti na glavnoj dijagonali predstavljaju varijansu atributa i treba ih zanemariti, jer ne možemo da tumačimo zavisnost kriterijuma sa samim sobom); stoga, ne možemo zaključiti da su kriterijumi međusobno nezavisni. Razlika je posebno izražena između kriterijuma *Cena* i *Udaljenost*, gde vidimo da postoji negativna linearna zavisnost. Međutim, intenzitet kovarijanse ne možemo direktno tumačiti. Kako bismo mogli da eksplicitno tumačimo zavisnost, računamo koeficijent korelacije. Za koeficijent korelacije bitna nam je standardna devijacija (koren varijanse, odnosno elemenata na dijagonali kovarijacione matrice) svakog kriterijuma.

Cena (eur)	Internet	Udaljenost od grada	Čistoća
13,7841	1,0198	1	0,7483

Korelacija između kriterijuma *Cena* i *Internet* se računa preko date formule i dobija se:

$$cor(Cena, Internet) = \frac{cov(x,y)}{stdev(x) * stdev(y)} = \frac{1}{13{,}7841 * 1{,}0198} = \frac{1}{14{,}057} = 0{,}07$$

Korelacija	Cena (eur)	Internet	Udaljenost	Čistoća
Cena (eur)	1	0,07	-0,85	-0,10
Internet	0,07	1	0,23	-0,16
Udaljenost	-0,85	0,23	1	-0,17
Čistoća	-0,10	-0,16	-0,17	1

Iz matrice korelacija vidimo da postoji jaka negativna linearna zavisnost između kriterijuma *Cena* i *Udaljenost*. Kao i kod matrice kovarijansi (kovarijacione matrice), vrednosti na glavnoj dijagonali treba zanemariti (kod korelacija vrednost će uvek biti 1).

Na kraju, računamo koeficijent regresije i popunjavamo matricu regresije. Koeficijent regresije između kriterijuma *Cena* i *Internet* dobijemo na sledeći način:

$$Cena = b(Internet) * Internet$$

$$b(Internet) = \frac{cov(x,y)}{var(x)} = \frac{1}{1{,}0198^2} = 0{,}96$$

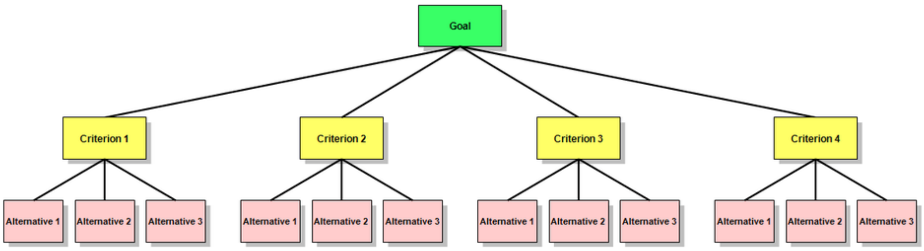
Dobijenu vrednost popunjavamo u matricu u redu gde je *Internet*, a kolonu gde je *Cena*.

Regresije	Cena (eur)	Internet	Udaljenost	Čistoća
Cena (eur)	1	0,01	-0,08	-0,01
Internet	0,96	1	0,30	-0,12
Udaljenost	-8,75	0,18	1	-0,09
Čistoća	-1,79	-0,21	-0,30	1

Vrednosti na glavnoj dijagonali zanemarujemo (vrednost će uvek biti jednaka 1).

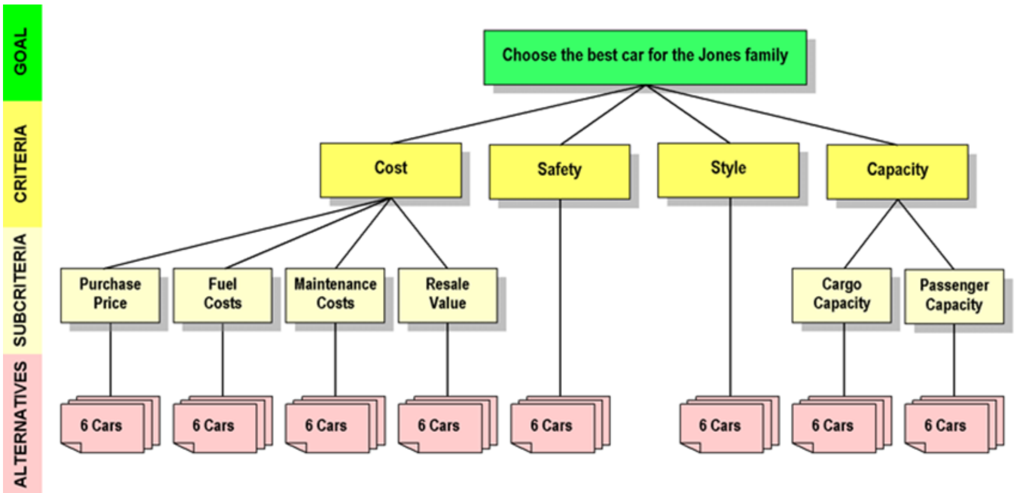
Strukturiranje problema odlučivanja

Problem odlučivanja često može da sadrži veliki broj kriterijuma, što otežava proces strukturiranja problema, odnosno, značajno otežava proces određivanja težina (pondera) kriterijuma (slika 1).



Slika 1. Problem odlučivanja

Kao rešenje problema većeg broja kriterijuma može se napraviti **hijerarhija atributa** (slika 2). Ovim postupkom se problem velikog broja kriterijuma razlaže na veći broj manjih podproblema koje je jednostavnije rešiti. Hijerarhija kriterijuma pravi se u situaciji kada postoji veliki broj kriterijuma (veći od 5). Pravljenje hijerarhije se radi ili domenskim znanjem ili korišćenjem podataka (npr. preko koeficijenta korelacije).



Slika 2. Hijerarhija atributa

Primer:

Za datu tabelu odlučivanja odrediti najbolju alternativu.

	Cena Sobe	Cena Interneta	Cena Kafe	Internet	Udaljenost od grada	Čistoća
	min	min	min	max	min	max
A	55	2	3	3	0,7	4
B	65	3	4	1	0,4	3
C	40	0	2,5	0	0,7	4
D	25	0	3	2	4	3
E	40	3	3	1	2	5

Međutim, DO ne može precizno da direktno odredi težine kriterijuma, te je napravio hijerarhiju kriterijuma. *Cena sobe*, *Cena interneta*, *Cena kafe* se predstavljaju kao kriterijum *Cena*. Dakle, napravljena je hijerarhija. Za tako dobijene kriterijume je odredio težine.

	Cena (eur)			Internet	Udaljenost od grada	Čistoća
	min			max	min	max
	Soba	Internet	Kafa			
A	55	2	3	3	0,7	4
B	65	3	4	1	0,4	3
C	40	0	2,5	0	0,7	4
D	25	0	3	2	4	3
E	40	3	3	1	2	5
Ponderi	0,35			0,2	0,3	0,15

Problem koji se javlja kod raščlanjivanja kriterijuma na podkriterijume, ogleda se u tome što je potrebno definisati težine za podkriterijume na nižim nivoima hijerarhije. Broj nivoa nije ograničen, ali ne treba da bude preveliki. Takođe, težine se dalje integrišu po hijerarhijama.

U datom primeru, DO je odredio sledeće težine unutar kriterijuma *Cena*. Težine iznose: za podkriterijum *Cena Soba* 0,5, za podkriterijum *Cena Internet* 0,3 i za podkriterijum *Cena Kafe* 0,2. Propagiranje težina se radi tako što se težina podkriterijuma pomnoži sa ponderom nadkriterijuma i dobija se:

	Cena Sobe	Cena Interneta	Cena Kafe	Internet	Udaljenost od grada	Čistoća
	min	min	min	max	min	max
Ponderi	0,35 * 0,5 = 0,175	0,35 * 0,3 = 0,105	0,35 * 0,2 = 0,07	0,2	0,3	0,15

Uzmimo ovaj primer da razmislimo koliko bi se, i kakve bi se, matrice procene popunjavale kada bi se težine određivale preko AHP metode. U slučaju kada je sve na jednom nivou hijerarhije (prvi slučaj), razmatrala bi se matrica 6x6 (pošto postoji šest kriterijuma), što lako može dovesti do nekonzistentnosti matrice procene. Sa druge strane, ako se napravi hijerarhija, popunjavaju se dve matrice procene u kojoj je prva veličine 4x4 (imamo četiri kriterijuma na prvom nivou), a druga veličine 3x3.

Analiza glavnih komponenti

Drugi način strukturiranja atributa je **automatsko strukturiranje atributa**. Automatsko strukturiranje atributa koristi analizu glavnih komponenti za otkrivanje zavisnosti između kriterijuma i potencijalnu redukciju matrice odlučivanja korišćenjem linearnih jednačina postojećih kriterijuma bez gubitka informacije. Način na koji se vrši automatsko strukturiranje problema je već obrađen u AHP metodi gde se ocenjivala konzistentnost matrice procene, kao i kod Promethee metode kod GAIA ravni.

Kao prednosti primene glavnih komponenti izdvajamo: 1) glavne komponente su međusobno nekorelisane (linearno nezavisne), 2) glavne komponente otkrivaju potencijalno nove osobine alternativa iz postojećih kriterijuma, 3) potencijalno smanjuju dimenzionalnost kriterijuma. Međutim, imaju i negativne osobine kao što su: 1) nemaju mogućnost hijerarhijskog uređenja, 2) ne važi uvek pretpostavka da otkrivene komponente imaju linearnu zavisnost, odnosno moguće je da između kriterijuma postoji nelinearna zavisnost koju glavne komponente ne mogu da otkriju i 3) Često je novodobijene kriterijume teško direktno opisati, s obzirom na to da predstavljaju kombinaciju informacije redukovanih atributa (kriterijuma).

Primer:

Za tabelu odlučivanja sa strane 2, uraditi analizu glavnih komponenti i odrediti broj komponenti koje biste koristili u novom rešenju. Dobijene komponente opisati.

Prvi korak je računanje matrice kovarijansi. Postupak računanja kovarijansi je već opisan na strani 3. Ispod se nalazi matrica kovarijansi.

Kovarijansa	Cena (eur)	Internet	Udaljenost	Čistoća
Cena (eur)	190	1	-15,7	-1
Internet	1	1,04	0,316	-0,12
Udaljenost	-15,7	0,316	1,7944	-0,168
Čistoća	-1	-0,12	-0,168	0,56

Koristeći matricu kovarijansi, računamo sopstvene vrednosti i sopstvene vektore.

Sopst. vrednosti	0,18556	0,00019	0,13307	0,03778
% var	52,04%	0,05%	37,32%	10,59%
Kum. % var	52,04%	52,09%	89,41%	100%

Posmatrajući sopstvene vrednosti, zaključujemo da prva i treća sopstvena vrednost objašnjavaju 89,36% varijanse, te možemo koristiti samo vrednosti koje dobijemo iz prve i treće glavne komponente. Zatim, koristeći sopstvene vektore, dobijamo „nove“ kriterijume koje treba protumačiti. Dakle, treba da posmatramo prvu i treću kolonu iz matrice sopstvenih vektora.

Sopst. vetori	**V1**	V2	**V3**	V4
Cena (eur)	**-0,426**	-0,66374	**0,30187**	0,53558
Internet	**-0,49968**	-0,10516	**-0,85869**	-0,04377
Udaljenost	**0,75273**	-0,48086	**-0,3905**	0,22289
Čistoća	**-0,04735**	-0,56317	**0,13798**	-0,81336

Prvu komponentu možemo da protumačimo kao „Centar grada pre svega“ (glavna komponenta 1), dok treću komponentu (glavna komponenta 2) možemo da tumačimo kao „Odnos cene i kvaliteta usluge“.

Nakon toga, računamo matricu odlučivanja u prostoru glavnih komponenti, čime dobijamo preslikavanje inicijalne tabele odlučivanja u prostor glavnih komponenti. Nakon računanja glavnih komponenti, računaju se ponderi i vrednosti alternativa, te se preko IKOR metode podaci normalizuju i bira se najprihvatljivija alternativa. Kompletan postupak računa se nalazi u skripti Modelovanje preferencija.

	GK1	GK2
A	-0,30158	-0,83393
B	0,39392	-0,47743
C	0,12569	0,07608
D	-0,71242	-0,22713
E	-0,32944	-0,03739

Određivanje vrednosti alternativa za kriterijume koji nam nisu dostupni (linearna regresija)

Način na koji možemo da odredimo vrednosti alternativa za kriterijume koji nam nisu dostupni je putem linearne regresije. Za ovaj postupak potrebni su nam istorijski podaci. Kompletan postupak biće objašnjen na primeru.

Primer:

Data nam je tabela odlučivanja gde banka želi da rangira klijente na osnovu javno dostupnih informacija sa njihovih profila na društvenoj mreži, u svrhu određivanja podesnosti za davanje kredita. Za tu potrebu suodređeni kriterijumi *Broj poseta* i *Broj prijatelja* na profilu, jer se smatra da implicitno ukazuju na osobu sa boljim primanjima. Ono što nam je veoma bitno za rangiranje klijenta je njegova plata. Međutim, taj podatak nam je nepoznat, a izuzetno je važan da bi se odlučilo kom klijentu ponuditi kredit (pretpostavka: klijent sa višim primanjima će biti podesniji).

	Broj poseta	Plata	Broj prijatelja
A	23	?	233
B	11	?	151
C	17	?	401
D	9	?	317

Ipak, iznos *Plate* možemo da odredimo preko *Snage auta* (koja je data u drugoj tabeli) koji potencijalni klijent vozi.

	Snaga auta (ks)	Plata (kdin)
A	65	?
B	110	?
C	80	?
D	190	?

Postupak određivanja plate na osnovu snage automobila ćemo uraditi tako što ćemo napraviti linearnu funkciju koja će preslikati vrednost snage automobila u platu koju klijent ima. Za pravljenje funkcije koristićemo istorijske podatke gde imamo informaciju o snazi auta i plati klijenta. Kako bismo ostali dosledni formuli regresije koju smo opisali na početku, slovom X ćemo obeležiti *Snagu automobila*, a sa Y ćemo obeležiti *Platu*. Kolona X se naziva i ulazna ili nezavisna promenjiva, a kolona Y se naziva izlazna ili zavisna promenjiva.

Dakle, pravimo funkciju $Y = b(X) * X$, gde je $b(X) = (cov(X,Y))/(var(X))$.

Kako bismo olakšali računanje, u tabeli istorijskih podataka ćemo dodati kolone Xc i Yc koje predstavljaju vrednosti koje se dobiju kada se od kolona *X* i *Y*, respektivno, oduzme prosečna vrednost posmatrane kolone. One su nam bitne za računanje kovarijanse. Brojilac u formuli kovarijanse ne predstavlja ništa drugo do zbira proizvoda prosečnih odstupanja alternativa od aritmetičke sredine po oba posmatrana kriterijuma. ; Stoga, dodajemo i kolonu Xc * Yc i računamo njenu sumu. Brojevi u ovoj koloni mogu da se tumače na sledeći način: ako je broj koji se dobije množenjem Xc i Yc pozitivan, to znači da se kolone *X* i *Y* „kreću u istom smeru“, odnosno, da su vrednosti kolona *X* i *Y* ili obe pozitivne ili obe negativne, što je indikator korelacije između kolona (kriterijuma). Takođe, potrebna nam je i varijansa kolone X, te računamo sumu kolone Xc^2.

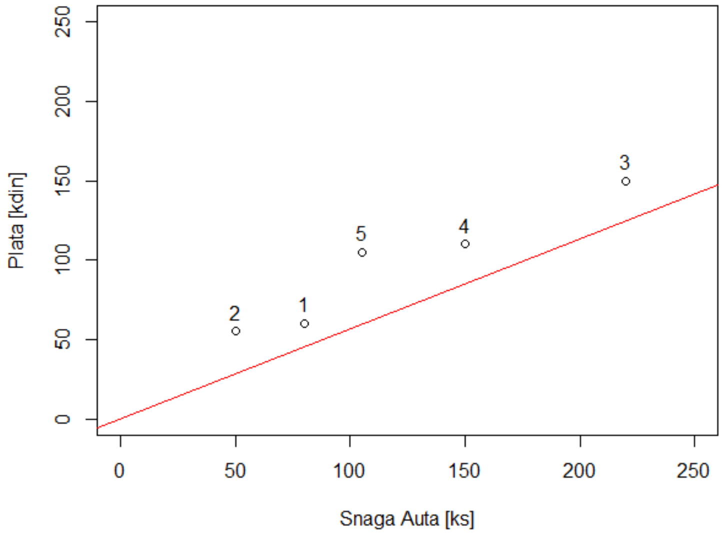
	Snaga auta (ks)	Plata (kdin)	Xc	Yc	Xc * Yc	Xc^2
	X	Y				
1	80	60	-41	-36	1476	1681
2	50	55	-71	-41	2911	5041
3	220	150	99	54	5349	9801
4	150	110	29	14	406	841
5	105	105	-16	9	-144	256
AVG	**121**	**96**			**9995**	**17620**

Dalje, možemo da odredimo $b(X)$, odnosno $b(x) = \frac{9995}{17620} = 0{,}567$. Koeficijent $b(X)$ možemo da tumačimo kao: za koliko hiljada dinara će se iznos plate promeniti, za jediničnu promenu snage automobila izražene u konjskim snagama (ks). Drugim rečima, koeficijent $b(X)$ predstavlja „univerzalnu konstantu“ koja direktno transformiše snagu automobila u iznos plate iz istorijiskih podataka.

Sada kada nam je poznata vrednost $b(x)$, možemo da „odredimo“ vrednosti *Plate* preko *Snage automobila*. To radimo tako što pomnožimo vrednosti snage automobila sa koeficijentom $b(X)$.

	Snaga auta (ks)	Plata (kdin)	$b(X)$*X	Yc
	X	Y		
1	80	60	45,38	-36
2	50	55	28,36	-41
3	220	150	124,80	54
4	150	110	85,09	14
5	105	105	59,56	9

Ukoliko uporedimo kolonu Y i kolonu $b(X)$*X, možemo da primetimo da su vrednosti kolone $b(X)$*X konstantno niže od vrednosti kolone Y. Isti detalj se vidi i na slici 3. Takođe, primećujemo da dobijena funkcija kreće iz koordinatnog početka.



*Slika 3. Y = 0.567 * X*

Ukoliko bismo „podigli" funkciju za određeni broj a, onda bi ona bolje pogađala vrednosti. To ćemo uraditi tako što ćemo promeniti linearnu funkciju. Umesto dosadašnje, pokušaćemo da napravimo sledeću funkciju:

$$Y = a + b(X) * X$$

Koeficijent $b(X)$ nam je poznat, a za računanje vrednosti koeficijenta a, potrebno je da izračunamo grešku procene Y- $b(X)$*X. Stoga, u gornju tabelu dodajemo tu kolonu i računamo njene vrednosti. Nakon toga, računamo prosečnu vrednost kolone greške procene; dobijena vrednost će predstavljati traženi koeficijent a.

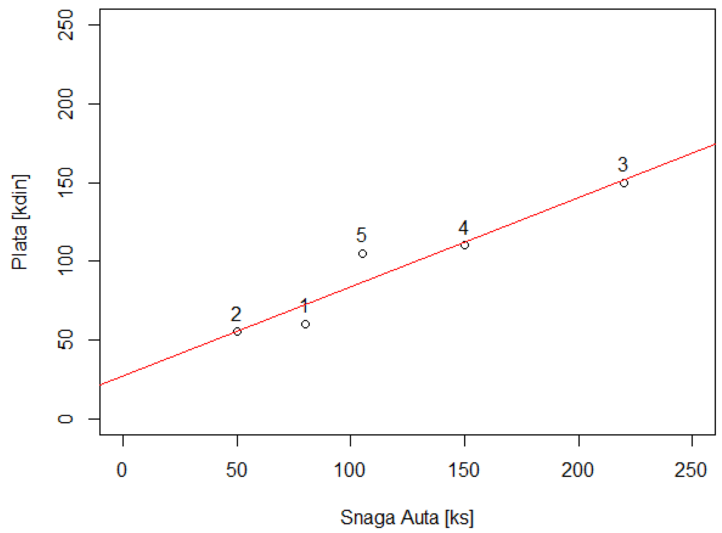
	Snaga auta (ks)	Plata (kdin)	$b(X)$*X	Y - $b(X)$*X
	X	Y		
1	80	60	45,38	14,62
2	50	55	28,36	26,64
3	220	150	124,80	25,20
4	150	110	85,09	24,91
5	105	105	59,56	45,44
				AVG: 27,362

Dakle, sada imamo jednačinu:

$$Y = 27{,}362 + 0{,}567 * X$$

Jednačinu možemo tumačiti na sledeći način: svako jedinično povećanje vrednosti nezavisnog (ulaznog) kriterijuma X (snage motora izražene u ks u našem slučaju), uvećaće vrednost zavisnog (izlaznog) kriterijuma Y (iznosa plate u hiljadama dinara u našem slučaju) za 0,567. Dodatno uvećanje vrednosti promenljive Y, od 27,362 jedinice, odnosi se na količinu informacije koja nije mogla direktno da bude objašnjena vezom između ulaznog i izlaznog atributa. U našem slučaju, to znači da snaga motora automobila koji klijent vozi, ne utiče stoprocentno na to kolika će mu biti plata, već na to utiču i drugi faktori, izraženi ovom konstantom.

Slika 4 pokazuje položaj alternativa u odnosu na novu funkciju. Evidentno je poboljšanje u odnosu na prethodnu sliku, jer se sada vrednosti alternativa nalaze skoro u potpunosti na liniji.



*Slika 4. Y = 27,362 + 0,567 * X*

Sada treba da ispitamo koliko možemo da verujemo proceni. Prvi korak je određivanje greške. Ona se vrši na isti način kao i za funkciju bez koeficijenta a. Primetimo da je prosečna greška 0, a da pojedinačne greške predviđanja ipak postoje. Ova osobina modela nam ukazuje da su greške ravnomerno raspoređene, što je izrazito poželjno. Međutim, kako greška u proceni postoji, nju treba ispitati.

	Snaga auta (ks)	Plata (kdin)	$a + b(X)$*X (predviđena vrednost)	Y – (a + $b(X)$*X) (vrednost greške predviđanja)	$(Y - (a + b(X) * X))^2$ (kvadrat greške predviđanja)
	X	Y			
1	80	60	72,74	-12,74	162,37
2	50	55	55,73	-0,73	0,53
3	220	150	152,16	-2,16	4,66
4	150	110	112,45	-2,45	6,00
5	105	105	86,92	18,08	326,74
				AVG: 0	**SUM: 500,31**

Kako bismo ocenili grešku, računaćemo sumu kvadrata greške (eng. *Sum of squared errors* ili SSE), koja u ovom slučaju iznosi 500,31 Ova mera greške predviđanja ukazuje da postoji greška kod procene vrednosti, međutim, njene vrednosti nije lako tumačiti. Iz tog razloga, uvodimo meru Koren srednje kvadratne greške (eng. *Root mean squared error* ili RMSE), koja se računa preko formule:

$$RMSE = \sqrt{\frac{SSE}{n}}$$

Kako smo SSE izračunali i znamo da je n = 5, dobijamo da je RMSE = $\sqrt{\frac{500,31}{5}} = \sqrt{100,061} =$ $10,00$. Vrednost 10,00 predstavlja prosečnu devijaciju procenjenih vrednosti, odnosnostandardnu grešku procene (eng. *Standard error)*, i ona može da se tumači da je greška procene, u proseku, 10 kdin. Cilj nam je da RMSE bude što niža.

Pored RMSE, može da se izračuna R^2 koji predstavlja procenat varijacije izlazne promenljive Y (u ovom slučaju plate) koji je objašnjen dobijenom funkcijom regresije. Računa se preko formule:

$$R^2 = 1 - \frac{SSE}{n * var(Y)}$$

Opseg vrednosti koje R^2 uzima su između 0 i 1. Vrednost 0 predstavlja model koji apsolutno ne objašnjava varijansu kolone koju procenjujemo. Tada je $SSE = var(Y)$ i model nema upotrebnu vrednost. Ukoliko vrednost R^2 iznosi 1, tada je funkcija koju smo odredili savršeno procenila vrednosti kolone Y, odnosno, tada je $SSE = 0$. Dakle, cilj je da R^2 bude što bliže vrednosti 1.

U našem primeru dobijamo:

$$R^2 = 1 - \frac{500,31}{5 * 1234} = 0,9189$$

Ovim je postupak procene i ocene vrednosti kriterijuma *Plata* završen. Dobijenu jednačinu, $Y = 27,362 + 0,567 * X$, primenjujemo na alternative iz tabele odlučivanja i dobijamo:

	Snaga auta (ks)	Plata (kdin)
A	65	64,23
B	110	89,76
C	80	72,74
D	190	135,14

Nakon toga, vrednosti prepisujemo u tabelu odlučivanja i već poznatim postupkom tražimo najprihvatljiviju alternativu.

	Broj poseta	Plata	Broj prijatelja
A	23	64,23	233
B	11	89,76	151
C	17	72,74	401
D	9	135,14	317