

Linearna regresiona i korelaciona analiza

U svakoj naučnoj disciplini osnovni problem je utvrđivanje veza između promenljivih veličina. Te veze mogu biti potpuno određene. Na primer, u fizici se može utvrditi tačna funkcionalna zavisnost između udaljenosti objekata od zemlje i gravitacione sile, ili između gasa u zatvorenoj posudi i temperature.

Međutim, u biološkim i društvenim naukama suočavamo se sa mnogo komplikovanijom situacijom. Ovde imamo daleko manje razloga da očekujemo otkrivanje tačno određene veze između promenljivih veličina. Zato su se, u ovim naučnim disciplinama, morala da koriste statistička izučavanja koja mere prosečne promene jedne veličine izazvane promenama druge veličine. Regresiona analiza upravo ima za cilj da utvrđuje i meri veze takvog tipa, a jačina te veze se procenjuje korelacionom analizom.

Regresiona analiza

U medicinskim istraživanjima često se interes istraživača usmerava prema problemu povezanosti među varijablama. Pri tom je od posebnog interesa mogućnost prognoziranja ili predviđanja vrednosti (ili varijabilnosti) jedne varijable na osnovu drugih varijabli.

Prvi tako formulisani problem potiče od engleskog antropologa Francisa Galtona. On je studirajući zajedno sa Pearsonom naseđivanje u biologiji, mereći visine očeva i sinova ustanovio neku vrstu paradoksa, odnosno, da visoki očevi imaju visoke sinove ali u proseku ne tako visoke kao što su oni sami, i slično, da niski očevi imaju niske sinove ali opet u proseku ne tako niske kao što su oni. Ovu tendenciju proseka neke karakteristike (u ovom slučaju visine) odabrane grupe da u sledećoj generaciji sinova teži ka proseku populacije a ne proseku njihovih očeva, Galton je nazvao regresijom, tačnije, regresijom prema proseku.

Da bi dobio informaciju zavisnosti visine sinova od visine njihovih očeva, Pearson je pretpostavio da se ta zavisnost može izraziti kao funkcija, $y = f(x)$, pri čemu je **y** zavisna varijabla, odnosno varijabla koju želimo da objasnimo ili predvidimo (u Galtonovom primeru visina sinova), a **x** nezavisna varijabla, odnosno varijabla koju koristimo da objasnimo zavisnu varijablu (visina očeva).

Osnovna svrha primene regresione analize je da se na osnovu jedne, poznate promenljive može da predvidi vrednost druge, nepoznate promenljive i to iz jednačine koja pokazuje njihovu zavisnost. Zavisnost može biti linearna i u tom slučaju se izražava jednačinom prave linije ili može biti drugačija (logaritamska, eksponencijalna), pa se izražava odgovarajućim matematičkim izrazima.

Dijagrami rasipanja

Prvi korak u istraživanju zavisnosti između varijabli je crtanje dijagrama rasipanja (scatter dijagram), koji predstavlja grafički prikaz zavisnosti i međuzavisnosti između promenljivih **x** i **y**. Dijagram rasipanja se konstruiše na osnovu dobijenog eksperimentalnog skupa podataka i prikazuje parove vrednosti **x** i **y** u pravougloj koordinatnoj sistemu sa odabranim skalama merenja na apscisnoj i ordinatnoj osi. Na apscisnoj osi se nanose vrednosti nezavisno promenljive **x**, a na ordinatnoj osi vrednosti zavisno promenljive **y**.

Dijagram rasipanja, na očigledan način, omogućava slikovitu predstavu o tome da li postoji ili ne postoji zavisnost i međuzavisnost između promenljivih x i y kao i njen karakter i intenzitet. Tako, na primer, na osnovu nacrtanih eksperimentalnih tačaka može vizuelno da se uoči oblik aproksimativne linije: prava, kriva, rastuća ili opadajuća, tačke maksimuma i/ili minimuma ili prevojne tačke. Prevojne tačke eksperimentalne krive na dijagramu rasipanja mogu značiti granicu između dva različita mehanizma iste pojave ili granicu poremećaja u merenju. Dijagram rasipanja, takođe, na očigledan način otkriva ekstremne vrednosti, a na njemu se mogu lako uočiti grube greške, a često i sistematske i slučajne greške ravnomernim rasipanjem eksperimentalnih podataka oko aproksimativne krive.

Zavisnost i međuzavisnost, na osnovu dijagrama rasipanja, može biti:

- **linearna zavisnost** eksperimentalnih tačaka, koja predstavlja pravolinijski oblik dijagrama rasipanja;

- **nelinearna zavisnost** eksperimentalnih tačaka, koja predstavlja krivolinijski oblik dijagrama rasipanja;

Zavisnost i međuzavisnost, zavisno od oblika dijagrama rasipanja, može biti:

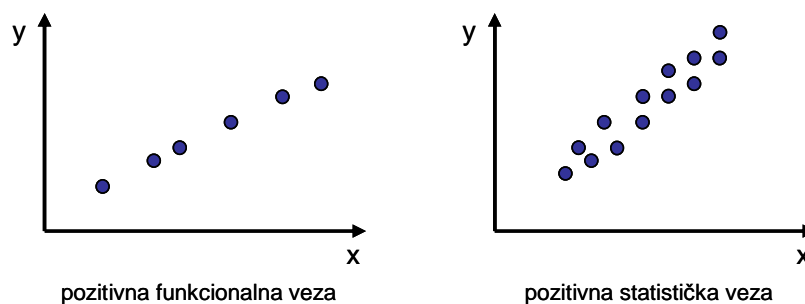
- **rastuća ili pozitivna zavisnost** eksperimentalnih tačaka, koja na dijagramu rasipanja označava upravo proporcionalnu vezu između promenljivih, odnosno sa porastom nezavisno promenljive x zavisno promenljiva y , takođe, raste;

- **opadajuća ili negativna zavisnost** eksperimentalnih tačaka, koja na dijagramu rasipanja pokazuje da je veza između promenljivih obrnuto proporcionalna, odnosno opadanje nezavisno promenljive x dovodi do opadanja i zavisno promenljive y .

U matematici postoje dva oblika zavisnosti :

- funkcionalna zavisnost izražena pomoću matematičkih jednačina tako da svakoj kombinaciji vrednosti nezavisno promenljivih x_1, x_2, \dots, x_k tačno odgovara vrijednost za y i

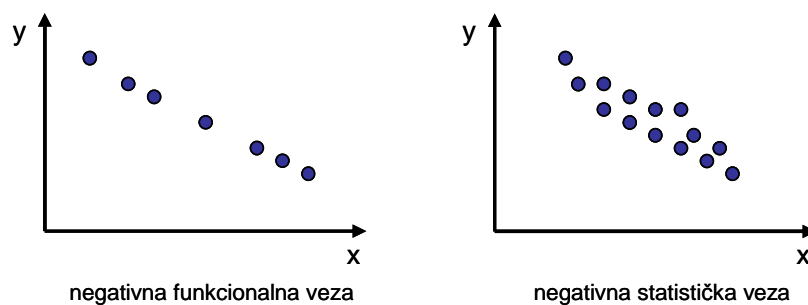
- statistička (stohastička ili slučajna) zavisnost koja se na osnovu eksperimentalnih podataka može izraziti pomoću očekivane zavisnosti ili regresione jednačine (aproksimativne krive), tako da svakoj merenoj kombinaciji vrednosti nezavisno promenljivih x_1, x_2, \dots, x_k odgovara računaska vrednost za y uvećana za grešku eksperimenta (regresije) ε , od čije veličine zavisi preciznost predviđanja regresione jednačine.



Slika 1. Dijagrami rasipanja sa pozitivnom pravolinijskom vezom između dve varijable

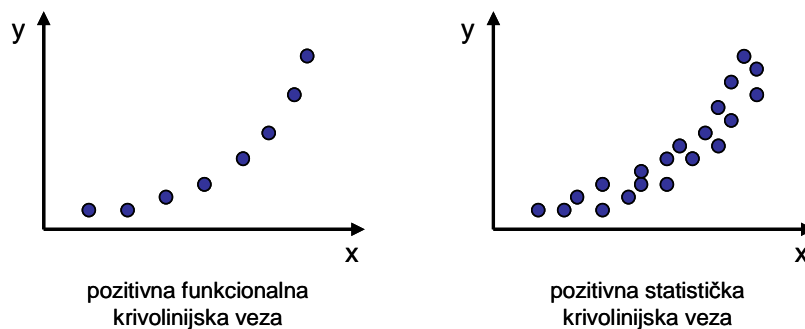
Na Sl. 1 su prikazana dva dijagrama rasipanja; slika levo prikazuje funkcionalnu vezu između varijabli x i y , a zamišljena linija koja povezuje sve tačke je prava linija. Od te linije nema nikakvog odstupanja, stoga se kaže da je ova veza strogo funkcionalna. Zamišljeni pravac je rastući, odnosno porast vrednosti jedne varijable prati porast vrednosti druge posmatrane varijable, pa je stoga ova veza pozitivna.

Grafik desno na Sl. 1 prikazuje slučaj koji je čest u praksi; tačke na ovom dijagramu takođe povezuje prava linija, ali su ovde prisutna pozitivna i negativna odstupanja od te prave linije, što se tumači raznim uticajima drugih varijabli koje se pojavljuju u praksi. Stoga ova veza nije strogo funkcionalna, već se za nju kaže da je statistička (stohastička ili slučajna) veza. Porast vrednosti jedne varijable u proseku prati porast vrednosti druge varijable, pa je i ova veza pozitivna.



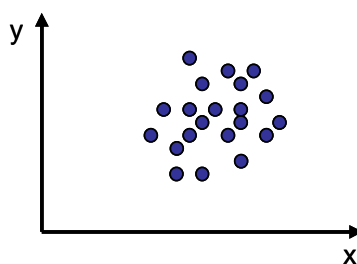
Slika 2. Dijagrami rasipanja sa negativnom pravolinijskom vezom između dve varijable

Na Slici 2 su prikazana druga dva pravolinijska dijagrama rasipanja. Na slici levo je prikazana funkcionalna veza između varijabli x i y (nema odstupanja tačaka od zamišljene linije), ali je zamišljeni pravac padajući, odnosno porast vrednosti jedne varijable prati pad vrednosti druge posmatrane varijable, pa je ova veza negativna. Na grafiku desno je prikazana negativna statistička veza dve varijable, u kojoj porast jedne varijable u proseku prati pad druge varijable.



Slika 3. Dijagrami rasipanja sa pozitivnom krivolinijskom vezom između dve varijable

Međusobne veze između dve varijable ne moraju da budu pravolinijske, već mogu imati i drugi oblik. Na Slici 3 su prikazana dve (funkcionalna i statistička) krivolinijske pozitivne veze između varijabli x i y .



Slika 4. Dijagram rasipanja za varijable među kojima ne postoji povezanost

Na Slici 4. je prikazan dijagram rasipanja koji upućuje na zaključak da nema povezanosti između posmatranih varijabli. Naime, zamišljena linija koja prolazi između tačaka ne postoji i ne može da se kaže da li porast jedne varijable prati rast ili pad druge varijable, jer se pri jednoj vrednosti varijable x javlja više različitih vrednosti varijable y .

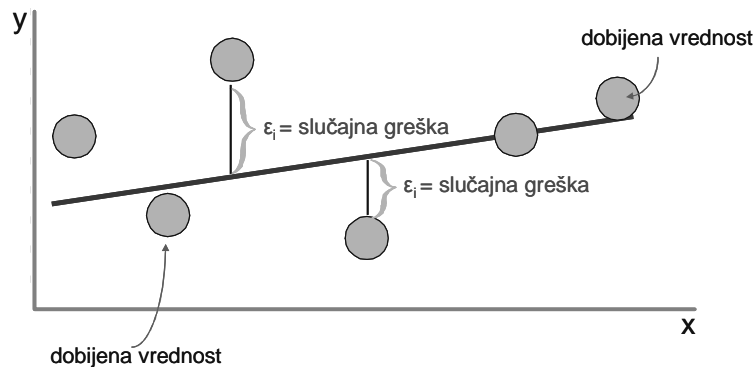
Regresioni modeli

Regresioni modeli se dele na proste, u kojima postoji samo jedna zavisna i jedna nezavisna varijabla i višestruke (multiple) u kojima postoji dve ili više nezavisnih i jedna zavisna varijabla. I jedni i drugi se prema prirodi zavisnosti između varijabli dele na linearne i nelinearne.

Prosti linearni regresioni modeli se izražavaju jednačinom prave, koja za populaciju glasi:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

U ovoj jednačini su: y - zavisna varijabla, x - nezavisna varijabla, a ε - slučajna greška, koja predstavlja razliku između eksperimentalno dobijene vrednosti y i vrednosti y izračunate iz jednačine prave.



Za uzorak jednačina.

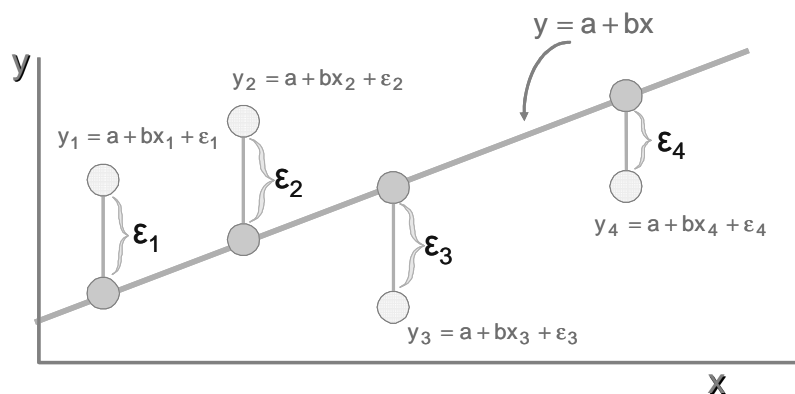
$$y = a + bx$$

gde su a i b nepoznati parametri koje treba oceniti. Konstanta a predstavlja vrednost y kada je $x = 0$ (vrednost u kojoj prava linija seče y osu kada je $x = 0$) i naziva se odsečak na y osi, a služi za procenu populacionog parametra β_0 .

Koeficijent regresije b pokazuje koliko se linearno menja vrednost zavisne varijable y ako se nezavisna varijabla x promeni (poveća ili smanji) za jedinicu mere. a služi za procenu populacionog koeficijenta regresije β_1 . Koeficijent b ima pozitivan predznak kada se sa povećanjem x povećava y . (zavisnost između x i y je upravo proporcionalna), a negativan kada se sa povećanjem x smanjuje y (obrnuto proporcionalna zavisnost). Ako su obe promenljive izražene u istim dimenzijama, b predstavlja tangens ugla α koji prava linija zaklapa sa x osom.

Metoda koja se koristi da se izračuna linearna jednačina iz datih tačaka naziva se metoda najmanjih kvadrata, a bazira se na tome da se minimizuje suma kvadriranih razlika (grešaka = ε) između dobijenih i izračunatih vrednosti y . Suština je u tome da se kroz grupu tačaka može da povuče više pravih linija, a "najbolja" je ona kod koje je zbir kvadrata odstupanja tačaka od regresione prave najmanji mogući.

$$\sum_{i=1}^n \varepsilon_i^2 = \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \dots + \varepsilon_n^2$$



Prava linija koja se prema kriterijumu najmanjih kvadrata najbolje uklapa u grupu tačaka naziva se regresiona linija, a jednačina koja je definiše naziva se regresiona jednačina. Regresiona jednačina ima sledeće osobine:

- razlika između stvarne vrednosti y i izračunate vrednosti y je najmanja moguća,
- iz srednje vrednosti x možemo da izračunamo srednju vrednost y ,
- kada x odstupa od srednje vrednosti, možemo da očekujemo i da y odstupa od svoje srednje vrednosti.

Iz regresione jednačine može da se izračuna očekivana vrednost y iz date vrednosti x , odnosno regresiona jednačina može da se koristi za "predviđanje" vrednosti y .

Koeficijenti regresione jednačine izračunavaju se prema sledećim izrazima:

$$b = \frac{\sum x y - N \bar{x} \bar{y}}{\sum x^2 - N \bar{x}^2}; \quad a = \bar{y} - b \bar{x}$$

Standardna greška odstupanja od regresione prave

Veličinu odstupanja tačaka od prave linije izražava standardna greška regresione prave, koja se izračunava prema izrazu:

$$S_{y,x} = \sqrt{\frac{\sum (y_i - \hat{y})^2}{N - 2}} = \sqrt{\frac{\sum y^2 - a \sum y - b \sum x y}{N - 2}}$$

gde je:

$\sum (y_i - \hat{y})^2$ - suma kvadrata odstupanja dobijenih vrednosti y od izračunatih vrednosti \hat{y} i

$N - 2$ - broj stepena slobode (N se umanjuje za dva zato što su a i b izračunati iz istih podataka).

Jaka linearna zavisnost između promenljivih x i y znači da su tačke vrlo blizu prave linije, pa je stoga suma kvadrata odstupanja dobijenih vrednosti y od izračunatih vrednosti \hat{y} , mala, odnosno mala je i standardna greška $S_{y,x}$. I obrnuto, kada je linearna zavisnost između promenljivih x i y slaba, tačke su rasute oko prave, suma kvadrata odstupanja tačaka od prave je velika i velika je standardna greška $S_{y,x}$.

Razlika između dobijenih vrednosti y i izračunatih vrednosti \hat{y} naziva se ostatak (rezidua), a pošto je standardna greška $S_{y,x}$ mera za veličinu ovih rezidua, njen drugi naziv je rezidualna standardna devijacija.

Korelaciona analiza

Regresionom analizom se izračunava jednačina koja pokazuje prirodu zavisnosti između dve promenljive, a korelacionom analizom se pokazuje koliki je stepen te zavisnosti. Drugim rečima, korelaciona analiza objašnjava koliko dobro izračunata jednačina izražava prirodu zavisnosti između promenljivih

Koeficijent determinacije

U regresionoj analizi postoji više varijacija, čija veličina određuje značaj regresione jednačine za predviđanje, a izražavaju se odgovarajućim sumama kvadrata odstupanja:

1. Ukupna suma kvadrata (SK_T) – mera odstupanja dobijenih vrednosti y od srednje vrednosti \bar{y} :

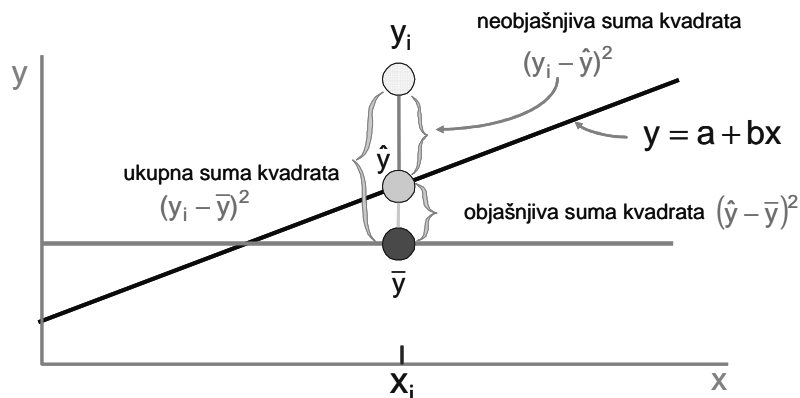
$$\sum (y_i - \bar{y})^2$$

2. Suma kvadrata za koju postoji objašnjenje (objašnjiva suma kvadrata, SK_R) – mera odstupanja izračunatih vrednosti \hat{y} od srednje vrednosti \bar{y} , koja je vezana za relaciju između x i y :

$$\sum (\hat{y} - \bar{y})^2$$

3. Suma kvadrata za koju ne postoji objašnjenje (neobjašnjiva suma kvadrata, SK_E) – mera odstupanja dobijenih vrednosti y od izračunatih vrednosti \hat{y} , koja u suštini predstavlja grešku

$$\sum (y_i - \hat{y})^2$$



Ukupna suma kvadrata (SK_T) jednaka je zbiru objašnjive (SK_R) i neobjašnjive sume kvadrata (SK_E). Odnos objašnjive i ukupne sume kvadrata je koeficijent determinacije, koji se obeležava sa r^2 :

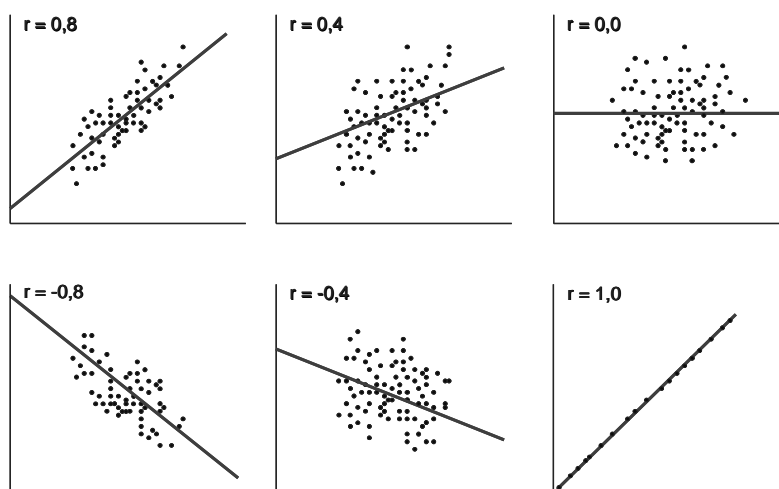
$$r^2 = \frac{SK_R}{SK_T} = \frac{\sum (\hat{y} - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{a \sum y - b \sum xy - N\bar{y}^2}{\sum y^2 - N\bar{y}^2}$$

Kada tačke leže na pravoj, nema neobjašnjive varijacije, pa je ukupna varijacija jednaka objašnjoj varijaciji ($SK_T = SK_R$) i koeficijent determinacije je jednak 1. Kada su tačke rasute oko prave, postoji neobjašnjiva varijacija, pa je ukupna varijacija veća od objašnjive ($SK_T > SK_R$) i koeficijent determinacije je manji od 1. Prema tome, vrednosti koeficijenta determinacije se kreću od 0 (između varijabli x i y nema linearne zavisnosti) do 1 (između varijabli x i y postoji apsolutna linearna zavisnost).

Koeficijent determinacije pokazuje koji deo varijacije u varijabli y potiče (ili je u vezi, odnosno prati) od varijacije u varijabli x i može da se izrazi i u procentima. Na primer, ako je $r^2 = 0,8$, to znači da 80% varijacije u varijabli y potiče od varijacije u varijabli x.

Koeficijent korelacije

Stepen linearne zavisnosti između dve varijable izražava se Pearson-ovim koeficijentom korelacije ili samo koeficijentom korelacije. Populacioni koeficijent korelacije se označava sa ρ (rho), a njegova vrednost se procenjuje izračunavanjem koeficijenta korelacije uzorka, koji se označava malim slovom r.



Koeficijent korelacije r uvek leži između -1 i 1. Kada je vrednost r blizu -1 ili 1 to znači da između promenljivih x i y postoji jaka linearna zavisnost i da regresiona jednačina može da se koristi za predviđanje sa velikom pouzdanošću. S druge strane, r blizu nule govori da između promenljivih postoji

slaba linearna zavisnost i korišćenje regresione jednačine za predviđanje vrednosti y nije od velike koristi.

Pozitivna vrednost r pokazuje da su promenljive x i y u pozitivnoj linearnoj korelaciji, što znači da se y linearno povećava kako se povećava x, i obrnuto, negativna vrednost r pokazuje da su promenljive x i y u negativnoj linearnoj korelaciji, odnosno da se y linearno smanjuje sa povećanjem x. Koeficijent korelacije r ima isti znak kao i nagib b.

Kod izračunavanja koeficijenta korelacije r treba imati na umu da njegova visoka vrednost ne mora obavezno da znači da su promenljive x i y u uzročno-posledičnoj vezi. Može da se desi da ove promenljive u uzročno-posledičnoj vezi sa nekom trećom promenljivom, pa su na taj način i same u korelaciji. Na primer, ako posmatramo povećanje cena hleba i povećanje ličnih dohodaka u nekom periodu možemo dobiti vrlo visok stepen linearne korelacije, iako ove dve veličine nisu u zavisnosti. Visok koeficijent korelacije r u ovom primeru može biti posledica nekog trećeg faktora sa kojim su ove dve promenljive u zavisnosti, kao što je na primer, inflacija koja uslovljava da i cene hleba i lični dohoci simultano rastu.

Za izračunavanje koeficijenta korelacije potrebna su tri različite sume kvadrata (SK): suma kvadrata varijable x, suma kvadrata varijable y i suma umnožaka varijabli x i y.

$$r = \frac{SK_{xy}}{\sqrt{SK_{xx} SK_{yy}}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{\sum xy - N\bar{x}\bar{y}}{\sqrt{(\sum x^2 - N\bar{x}^2)(\sum y^2 - N\bar{y}^2)}}$$

Koeficijent korelacije može da se izračuna i kao kvadratni koren iz koeficijenta determinacije:

$$r = \sqrt{\text{koeficijent determinacije}}$$

Značajnost koeficijenta korelacije r testira se nultom hipotezom $H_0: r = 0$, a odgovarajuća alternativna hipoteza je $H_A: r \neq 0$. Statistička veličina koju izračunavamo je t i to prema izrazu;

$$t = r \sqrt{\frac{N-2}{1-r^2}}$$

Ukoliko je nulta hipoteza ispravna (r je jednako nuli) izračunata vrednost t biće manja od tablične za izabrani nivo verovatnoće i broj stepena slobode $\phi = N - 2$, što znači da između promenljivih x i y ne postoji značajna linearna korelacija. Ako je izračunata vrednost t veća od tabelarne za navedene kriterijume, prihvata se alternativna hipoteza (r različito od nule) i zaključak je da između promenljivih x i y postoji značajna linearna korelacija.

Primena regresione i korelacione analize u analitici

Regresiona i korelaciona analiza može da se primeni u sledećim slučajevima:

1. Za izračunavanje jednačine standardne krive
2. Za procenu tačnosti metoda, kao i poređenje metoda
3. Za procenu tačnosti metoda uz pomoć metode standardnog dodatka ("recovery") test.

Izračunavanje jednačine standardne krive

Jednačina standardne krive izražava zavisnost između koncentracije (x) i apsorbancije (y) i može da se koristi za izračunavanje koncentracije nepoznatog uzorka uz pomoć njegove apsorbancije. Izračunavanje se izvodi u sledećim fazama:

a) sve dobijene apsorbancije sa odgovarajućim koncentracijama prikažu se grafički u koordinatnom sistemu da bi se videlo ima li nekih tačaka koje se ne uklapaju u linearnu zavisnost. Inspekcijom dobijenog grafika pre izračunavanja jednačine i odbacivanjem tačaka koje se ne

uklapaju, štedi se trud i vreme za ponovno izračunavanje. To se dešava obično kod visokih koncentracija koje se ne uklapaju u linearnost metode.

b) izračuna se koeficijent korelacije r da se vidi da li je stepen zavisnosti zadovoljavajući i da li uopšte ima smisla izračunavati jednačinu prave. Za standardnu krivu uslov je da je $r > 0.99$ i samo u tom slučaju izračunata jednačina prave može da se koristi za izračunavanje koncentracije nepoznatog uzorka. Ako r ne zadovoljava ovaj uslov dalje izračunavanje se prekida i mora da se ponovi analitički postupak za standardnu krivu, jer je greška, verovatno u njemu.

c) izračuna se jednačina standardne krive, odnosno jednačina prave uz pomoć konstanti a i b . Za svaku standardnu krivu uslov je da prolazi kroz nulu, odnosno da se za koncentraciju 0 dobije i apsorbcija 0, što znači da u jednačini prave mora da bude $a = 0$. Izračunavanjem konstante a uvek se dobija neka vrednost (retko kad se dobija da je $a = 0$) i potrebno je da se ispita značajnost tog odsečka, odnosno da se testira nulta hipoteza za $a = 0$. Ako se nulta hipoteza prihvati, odsečak a nije različit od nule i njegova apsolutna vrednost se zanemaruje. S druge strane, kada se nulta hipoteza odbaci, odsečak a je različit od nule i njegova numerička vrednost ne može da se zanemari. U tom slučaju treba ispitati sam analitički postupak (metodu, korišćene reagense itd.) da bi se ustanovilo šta je razlog toj pojavi.

Značajnost odsečka a ispituje se izračunavanjem standardne greške (S_a) za odsečak i njenim ocenjivanjem uz pomoć Studentovog t -testa. Ako je izračunata vrednost ta veća od tabelarne za izabranu verovatnoću i broj stepena slobode $\phi = N - 2$, odsečak a je značajno različit od nule i obrnuto. Standardna greška za odsečak a izračunava se prema izrazu:

$$S_a = S_{y,x} \sqrt{\frac{\sum x^2}{N(\sum x^2 - N\bar{x}^2)}}; \quad t_a = \frac{a}{S_a}$$

PRIMER 1.: Izračunati jednačinu za standardnu krivu za hemoglobin.

g/L:	50	75	100	125	150	175	200
A:	0,134	0,200	0,272	0,340	0,406	0,482	0,543

$$\begin{aligned} N &= 7 & \Sigma xy &= 345,25 \\ \Sigma x &= 875 & \Sigma y &= 2,377 \\ \bar{x} &= 125 & \bar{y} &= 0,33957 \\ \Sigma x^2 &= 126875 & \Sigma y^2 &= 0,939549 \end{aligned}$$

$$r = \frac{345,25 - 7(125)(0,33957)}{\sqrt{[126875 - 7(125)^2][0,939549 - 7(0,33957)^2]}} = \frac{48,1}{\sqrt{2316,9125}} = \frac{48,1}{48,13432} = 0,9993$$

Vrednost koeficijenta korelacije zadovoljava zahtevani kriterijum da je $r \geq 0,99$, pa može da se pristupi izračunavanju konstanti a i b , odnosno jednačine prave.

$$\begin{aligned} b &= \frac{345,25 - 7(125)(0,33957)}{126875 - 7(125)^2} = \frac{48,1}{17500} = 0,002749 \\ a &= 0,33957 - 0,002749 \times 125 = -0,00406 \end{aligned}$$

Jednačina prave glasi: $y = -0,00406 + 0,00275x$. U jednačini standardne krive odsečak a predstavlja apsorbciju za koncentraciju nula i potrebno je ispitati njegovu značajnost. Kada je, kao u

ovom primeru vrednost odsečka vrlo mala nije neophodno da se značajnost ispituje matematički, već to može da se uradi na osnovu iskustva (apsorbancija 0,004 nije značajna i često je manja od same osetljivosti mernog instrumenta).

$$S_{y,x} = \sqrt{\frac{0,939549 - (-0,00406) 2,377 - 0,002749 (345,25)}{7 - 2}} = \sqrt{\frac{0,00107}{5}} = 0,04626$$

$$S_a = 0,04626 \sqrt{\frac{126875}{7 [126875 - 7 (125)^2]}} = 0,04626 \sqrt{1,35714} = 0,4708$$

$$t_a = \frac{-0,0406}{0,04708} = 0,62$$

Izračunata vrednost t_a manja je od tabelarne za $p = 0,05$ i $\varphi = 4$, što znači i da odsečak a na y osi nije značajno različit od nule, odnosno može da se smatra da je jednak nuli.

Iz jednačine standardne krive može da se izračuna koncentracija nepoznatog uzorka iz njegove apsorbcije.

PRIMER 2.: Iz jednačine standardne krive za hemoglobin iz prethodnog primera izračunati koncentraciju uzorka čija je apsorbcija $A = 0,355$.

Rešenje: U jednačini standardne krive y predstavlja apsorbciju, a x koncentraciju, što znači da jednačina mora da se reši po x :

$$y = -0,00406 + 0,00275 x$$

$$x = \frac{(y + 0,00406)}{0,00275}$$

$$y = 0,355$$

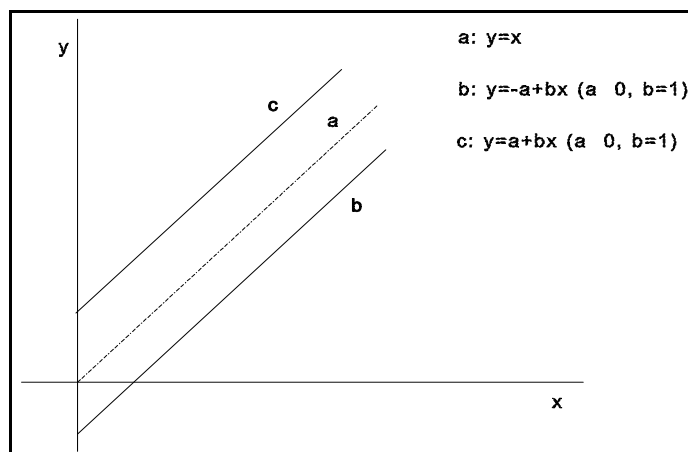
$$x = \frac{(0,355 + 0,00406)}{0,00275}$$

$$x = 130,57 \text{ g/L}$$

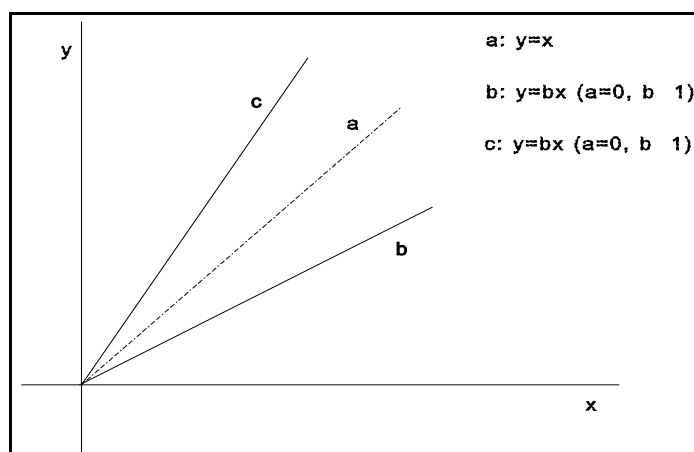
Tačnost metode i poređenje metoda

Regresiona i korelaciona analiza može da se koristi za ispitivanje tačnosti metoda, kao i za njihovo poređenje. Ako se nova metoda primeni na seriju uzoraka poznate koncentracije (na primer standardi), onda se iz dobijenih rezultata može da oceni tačnost metode u odnosu na poznate koncentracije. U ovom slučaju se sa x označavaju poznate koncentracije, a sa y vrednosti dobijene ispitivanom metodom. Tačnost nove metode može da se oceni i ako se ona uporedi sa referentnom metodom, pri čemu se rezultati dobijeni referentnom metodom označavaju sa x , a rezultati dobijeni novom metodom sa y .

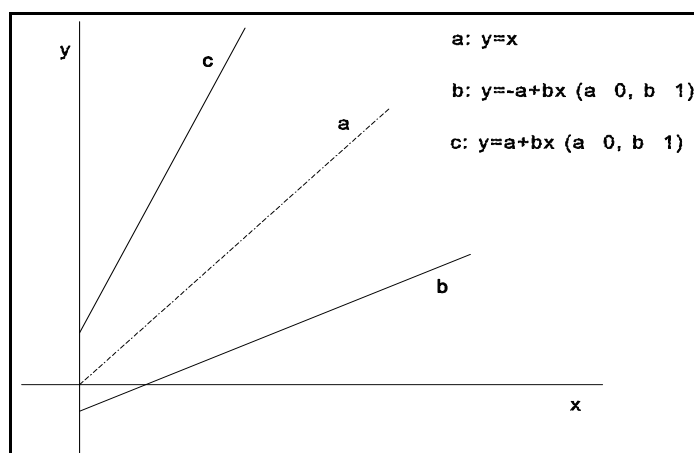
Ako se nova metoda upoređuje sa metodom koja nije referentna, ali ima poznato odstupanje, onda iz toga izvodimo zaključak o odnosu rezultata koji se dobijaju ovim metodama, ali ne i o tačnosti nove metode. Kod ovakvog poređenja se sa x označavaju rezultati dobijeni metodom sa poznatim odstupanjem, a sa y rezultati dobijeni novom metodom.



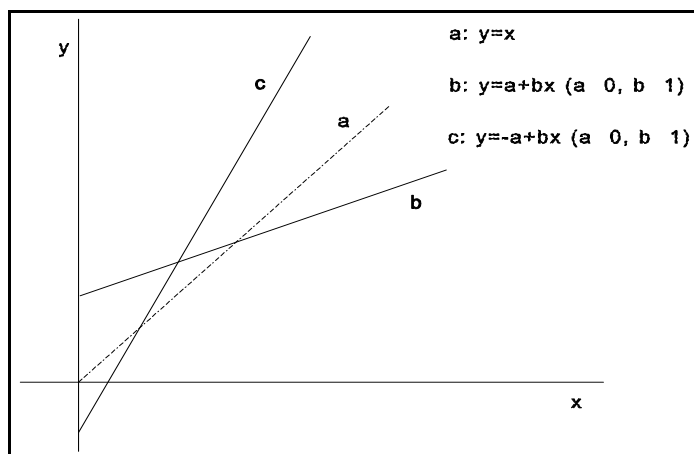
Slika 3. a) $y = x$ $a = 0$ $b = 1$: nema grešaka;
 b) $y = -a + x$ $a < 0$ $b = 1$: negativna sistematska greška, nema procentualne greške.
 c) $y = a + x$ $a > 0$ $b = 1$: pozitivna sistematska greška, nema procentualne greške;



Slika 4. a) $y = x$ $a = 0$ $b = 1$: nema grešaka;
 b) $y = bx$ $a = 0$ $b < 1$: nema sistematske greške, negativna procentualna greška.
 c) $y = bx$ $a = 0$ $b > 1$: nema sistematske greške, pozitivna procentualna greška;



Slika 5. a) $y = x$ $a = 0$ $b = 1$: nema grešaka;
 b) $y = -a + bx$ $a < 0$ $b < 1$: sabiranje negativne sistematske greške i negativne procentualne greške.
 c) $y = a + bx$ $a > 0$ $b > 1$: sabiranje pozitivne sistematske greške i pozitivne procentualne greške;

Slika 6. a) $y = x$ $a = 0$ $b = 1$: nema grešaka;b) $y = a + bx$ $a > 0$ $b > 1$: sabiranje pozitivne sistematske greške i negativne procentualne greške;c) $y = -a + bx$ $a = 0$ $b < 1$: sabiranje negativne sistematske greške i pozitivne procentualne greške.

Kod ovakve primene regresione i korelacione analize odsečak a predstavlja sistematsku grešku metode koja je označena sa y i ima istu dimenziju kao i rezultati te metode, a nagib b predstavlja procentualno odstupanje. Kako su kod poređenja metoda rezultati obe metode izraženi (najčešće) u istim dimenzijama, nagib b je jednak tangensu ugla koji prava zaklapa sa x osom. Ako je slaganje između metoda potpuno ili ako je metoda potpuno tačna, odsečak a je jednak nuli 0 (nema sistematske greške), a b je jednako 1, jer prava zaklapa ugao od 45° sa x osom. Ugao od 45° između prave i x ose znači da se za svaku vrednost x dobija ista vrednost y , odnosno metode daju iste rezultate. U tom slučaju je slaganje 100% jer je $b = 1$. Kada je $b > 1$, metoda označena kao y daje veće vrednosti od metode x za onoliko procenata za koliko je b veće od 1 (na primer: $b = 1,1 = 110\%$) i obrnuto, ako je $b < 1$, metoda y daje manje vrednosti od metode x .

Odnos sistematske i procentualne greške prikazan je grafički na slikama 10.3 do 10.6.

Kako konstante a i b predstavljaju greške, neophodno je da se ispita značajnost njihove razlike od nule. Značajnost odsečka proverava se na način kako je to opisano kod izračunavanja jednačine standardne krive i ako je izračunato ta veće od tabelarne vrednosti za $\phi = \phi - 2$ i izabranu verovatnoću, zaključak je da je značajna sistematska greška metode y .

Značajnost procentualne greške, odnosno značajnost razlike b u odnosu na 1 (100% slaganje) proverava se izračunavanjem standardne greške za nagib S_b i odgovarajuće konstante t_b prema izrazima:

$$S_b = \frac{S_{y,x}}{\sqrt{\sum x^2 - N\bar{x}^2}}; \quad t_b = \frac{1 - b}{S_b}$$

Postupak primene regresione i korelacione analize je isti kao i kod izračunavanja jednačine standardne krive, s tom razlikom što je uslov za vrednost koeficijenta korelacije nešto blaži. Kod poređenja metoda potrebno je da je $r \geq 0.9$.

PRIMER 4.: Dvema metodama određeno je gvožđe u seriji standardnih rastvora. Pokazati kakva je tačnost metoda, kao i kakav je odnos između njih (vrednosti su date u $\mu\text{g}/100 \text{ ml}$)

standardi:	50	100	150	200	250	300
metoda A:	48,5	99,5	150,8	202,4	254,2	308,4
metoda B:	49,5	100,8	150,9	198,9	249,4	302,5

Rešenje:

a) Tačnost metode A

Da bi se ocenila tačnost metode A treba uporediti koncentracije standarda sa vrednostima dobijenim metodom A.

$$\begin{aligned} N &= 6 & \Sigma y &= 1063,8 \\ \Sigma x &= 1050 & \bar{y} &= 177,3 \\ \bar{x} &= 175 & \Sigma y^2 &= 235687,1 \\ \Sigma x^2 &= 227500 & \Sigma xy &= 231545 \end{aligned}$$

$$r = \frac{231545 - 6(175)(177,3)}{\sqrt{[227500 - 6(175)^2][235687,1 - 6(177,3)^2]}} = \frac{45380}{45382,232} = 0,99995$$

$$b = \frac{231545 - 6(175)(177,3)}{227500 - 6(175)^2} = \frac{231545 - 186165}{227500 - 183750} = \frac{45380}{43750} = 1,03726$$

$$a = 177,3 - 1,03726 \times 175 = 177,3 - 181,5205 = -4,2205$$

$$S_{y,x} = \sqrt{\frac{235687,1 - (-4,2205)1063,8 - 1,03726(231545)}{6 - 2}} = \sqrt{\frac{4,3979}{4}} = 1,04856$$

$$S_a = 1,04856 \sqrt{\frac{227500}{6[227500 - 6(175)^2]}} = 1,04856 \sqrt{0,866667} = 0,976156$$

$$t_a = \frac{-4,2205}{0,976156} = -4,324 = 4,324$$

$$S_b = \frac{1,04856}{\sqrt{227500 - (175)^2}} = \frac{1,04856}{\sqrt{43750}} = 0,005013$$

$$t_b = \frac{1 - 1,03726}{0,005013} = \frac{-0,03726}{0,005013} = -7,433 = 7,433$$

Iz dobijenih rezultata može da se zaključi da su i sistematska i procentualna greška značajne, pa bi tačnost mogla da se opiše na sledeći način: metoda A daje za 4,22 $\mu\text{mol/L}$ niže (negativna sistematska greška) i za 3,7% više vrednosti (pozitivna procentualna greška, $b = 1,037 = 103,7\%$). Ovde imamo sabiranje negativne sistematske i pozitivne procentualne greške, tako da dobijena prava seče pravu $y = x$. Tačka preseka se izračunava iz sistema jednačina sa dve nepoznate (jedna je izračunata jednačina, a druga $y = x$).

Iako su i sistematska i procentualna greška male po apsolutnoj vrednosti, u ovom primeru su značajne jer je vrednost koeficijenta korelacije r velika. Razlog za to je vrednost standardne greške $S_{y,x}$ koja je u obrnutom odnosu sa koeficijentom r . Što je r bliže jedinici, to je $S_{y,x}$ manje, jer je manje rasipanje tačaka veće, a kada je ova standardna greška mala, male su i standardne greške S_a i S_b , pa su zbog toga velike vrednosti t_a i t_b . Ako je r malo, i velike greške će biti ocenjene kao prihvatljive. Iz tog razloga je postavljen uslov za vrednost koeficijenta r , koji je tako izabran ($r \geq 0,99$).

za standardnu krivu i $r \approx 0,9$ za poređenje metoda) da se odbacuju standardne greške koje nisu prihvatljive zbog svoje velike apsolutne vrednosti.

b) Tačnost metode B

$$\begin{aligned} N &= 6 & \Sigma y &= 1052 \\ \Sigma x &= 1050 & \bar{y} &= 175,333 \\ \bar{x} &= 175 & \Sigma y^2 &= 228649,52 \\ \Sigma x^2 &= 227500 & \Sigma xy &= 228070 \end{aligned}$$

$$r = 228070 - \frac{6(175)(175,333)}{\sqrt{[227500 - 6(175)^2][228649,52 - 6(175,333)^2]}} = \frac{43970}{43974,203} = 0,9999$$

$$b = \frac{228070 - 6(175)(175,333)}{227500 - 6(175)^2} = \frac{43970}{43750} = 1,00503$$

$$a = 175,333 - 1,00503 \times 175 = 175,333 - 175,88025 = -0,54725$$

$$S_{y,x} = \sqrt{\frac{228649,52 - (-0,54725) 1052 - 1,00503(228070)}{6 - 2}} = \sqrt{\frac{8,0349}{4}} = 1,41729$$

$$S_a = 1,41729 \sqrt{\frac{227500}{6[227500 - 6(175)^2]}} = 1,41729 \sqrt{0,886939} = 1,334768$$

$$t_a = \frac{-0,54725}{1,334768} = -0,410 = 0,410$$

$$S_b = \frac{1,41729}{\sqrt{227500 - 6(175)^2}} = \frac{1,41729}{\sqrt{43750}} = 0,006776$$

$$t_b = \frac{1 - 1,00503}{0,006776} = \frac{-0,00503}{0,006776} = -0,742 = 0,742$$

U ovom primeru nisu značajne ni sistematska, ni procentualna greška (obe vrednosti t su manje od tabelarne), tako da je zaključak da metoda B daje tačne rezultate u odnosu na koncentracije standarda.

c) Poređenje metoda A i B

$$\begin{aligned} N &= 6 & \Sigma y &= 1052 \\ \Sigma x &= 1063,8 & \bar{y} &= 175,333 \\ \bar{x} &= 177,3 & \Sigma y^2 &= 228649,52 \end{aligned}$$

$$\Sigma x^2 = 235687,1 \quad \Sigma xy = 232131,91$$

$$r = \frac{232131,91 - 6(177,3)(175,333)}{\sqrt{[235687,1 - 6(177,3)^2][228649,52 - 6(175,333)^2]}} = \frac{45612,665}{45614,797} = 0,99995$$

$$b = \frac{232131,91 - 6(177,3)(175,333)}{235687,1 - 6(177,3)^2} = \frac{45612,665}{47075,36} = 0,968929$$

$$a = 175,33 - 0,968929(177,3) = 175,33 - 171,79105 = 3,53895$$

$$S_{y,x} = \sqrt{\frac{228649,52 - 3,53895(1052) - 0,968929(232131,91)}{6 - 2}} = \sqrt{\frac{7,1955}{4}} = 1,34122$$

$$S_a = 1,34122 \sqrt{\frac{235687,1}{6[235687,1 - 6(177,3)^2]}} = 1,34122 \sqrt{0,834432} = 1,225167$$

$$t_a = \frac{3,53895}{1,225167} = 2,888$$

$$S_b = \frac{1,34122}{\sqrt{235687,1 - 6(177,3)^2}} = \frac{1,34122}{\sqrt{47075,36}} = 0,006182$$

$$t_b = \frac{1 - 0,968929}{0,006182} = \frac{0,031071}{0,006182} = 5,026$$

Standardne greške koje su izračunate kod ovog poređenja statistički su snačajne i zaključak je da nema slaganja metoda. Pozitivna sistematska greška se sabira sa negativnom procentualnom, tako da metoda B daje upočetku veće vrednosti od metode A jer je veći uticaj sistematske greške, a kasnije manje vrednosti, zbog većeg uticaja procentualne greške. Tačka preseka se izračunava iz sistema jednačina sa dve nepoznate:

$$y = 3,53895 + 0,96893 x$$

$$y = x$$

$$x = \frac{3,53895}{0,03107} = 113,9 = y$$

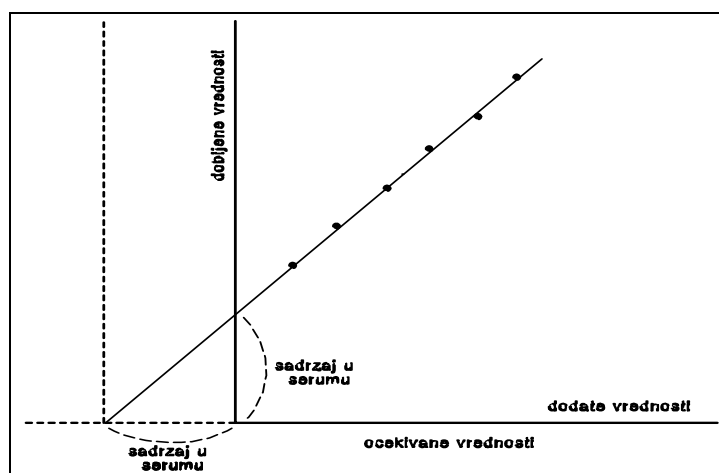
Tačka preseka je 11,39 umol/L, odnosno do te koncentracije metoda B daje veće, a iznad nje manje vrednosti od metode A.

Metoda standardnog dodatka ("recovery" test)

Metoda standardnog dodatka služi za utvrđivanje tačnosti metode kada se koriste uzorci sa biološkim matriksom. Zbog interferencije biološkog materijala nije dovoljno samo utvrditi tačnost sa standardnim rastvorima. "Recovery" test se izvodi tako da se u uzorak biološkog materijala dodaju standardni rastvori određenih koncentracija i potom se metodom čija se tačnost utvrđuje određuje sadržaj željenog konstituenta. Dobijene koncentracije se upoređuju sa dodatim koncentracijama, pri

čemu treba voditi da se u uzorku već nalazi izvesna koncentracija ispitivanog konstituenta. Idealno bi bilo da se u biološkom materijalu uopšte ne nalazi supstanca koju određujemo, tako da se primenom ispitivane metode određuje samo supstanca dodata u obliku standardnog rastvora.

Da bi se sačuvao biološki matriks, serum i standardi se mešaju u odnosu 9+1 (predhodno se pripreme standardi čija je koncentracija 10 puta veća od željene). Pripremi se i uzorak koji sadrži 9 delova seruma i 1 deo vode i kod izračunavanja se ovaj uzorak tretira kao dodata koncentracija 0. U pripremljenim uzorcima se odredi sadržaj ispitivanom metodom i dobijeni rezultati pretvore u koncentracije. Ovako određene koncentracije označe se kao dobijene koncentracije, a koncentracije standarda koji su dodati u biološki uzorak označe se kao dodate koncentracije (izraz "očekivane" koncentracije nije odgovarajući, jer on predstavlja sumu dodatih koncentracija i već prisutnog sadržaja u serumu, što ovde nije slučaj). I jedne i druge koncentracije se predstave u koordinatnom sistemu i to na x-osu dodate, a na y-osu dobijene koncentracije.



Slika 7. Grafički prikaz metode standardnog dodatka

Iz dobijenih rezultata izračuna se jednačina prave, a zbog toga što nije uzet u obzir sadržaj već prisutne supstance u serumu, prava će imati odsečak a koji je značajan i koji odgovara sadržaju u serumu. Iz tog razloga ne testira se značajnost odsečka a, odnosno ne izračunavaju se standardna greška S_a i t_a . Ako je tačnost metode zadovoljavajuća, odnosno ako se ispitivanom metodom odredi tačno ona koncentracija koja je dodata u obliku standarda, izračunata konstanta b u jednačini prave neće biti značajno različita od 1. Da bi smo to proverili obavezno se testiraju standardna greška S_b i t_b , koji ne smeju da budu značajno veliki. Uslov za r je isti kao i kod poređenja metoda ($r \geq 0,9$). Ako b nije značajno različito od 1, zaključak je da je "recovery" jednak 100%, odnosno odredili smo ono što smo i dodali (tačnost metode je zadovoljavajuća).

Na slici se vidi položaj prave koja je izračunata iz rezultata za "recovery" test. Ako se prava produži do preseka sa x-osom, i ordinata pomeri do te tačke, koncentracije na x-osi sada odgovaraju očekivanim koncentracijama, jer predstavljaju sumu dodatih koncentracija i već prisutnog sadržaja u serumu.

PRIMER 5.: U uzorak seruma dodate su različite koncentracije ureje i to kao standardi čije su koncentracije 50, 100, 150, 200 i 250 mmol/L. Pošto su standardi mešani sa serumom u odnosu 1 + 9, dodate koncentracije standarda su sledeće: 5, 10, 15, 20 i 25 mmol/L. Paralelno se pripremi i uzorak seruma razblažen vodom u istom odnosu i kod njega je dodata koncentracija ureje jednaka 0. U pripremljenim uzorcima se odredi ureja enzimskim postupkom i treba pokazati kakva je njena tačnost.

Rešenje:

dodato (x)	dobijeno (y)	Σx^2	Σy^2	Σxy
0	6,3	0	39,69	0
5	11,6	25	134,56	58
10	15,9	100	252,81	159
15	21,2	225	449,44	318
20	26,0	400	676,00	520
25	31,5	625	992,25	787,5

$$N = 6 \quad \Sigma xy = 1842,5$$

$$\Sigma x = 75 \quad \Sigma y = 112,5$$

$$\bar{x} = 12,5 \quad \bar{y} = 18,75$$

$$\Sigma x^2 = 1375 \quad \Sigma y^2 = 2544,75$$

$$r = \frac{1842,5 - 6(12,5)(18,75)}{\sqrt{[1375 - 6(12,5)^2][2544,75 - 6(18,75)^2]}} = \frac{436,25}{436,43621} = 0,9996$$

$$b = \frac{1842,5 - 6(12,5)(18,75)}{1375 - 6(12,5)^2} = \frac{436,25}{437,5} = 0,99714$$

$$a = 18,75 - 0,99714(12,5) = 18,75 - 12,46425 = 6,2858$$

$$S_{y,x} = \sqrt{\frac{2544,75 - 6,2858(112,5) - 0,99714(1842,5)}{6 - 2}} = \sqrt{\frac{0,36705}{4}} = 0,30292$$

$$S_b = \frac{0,30292}{\sqrt{1375 - 6(12,5)^2}} = \frac{0,30292}{\sqrt{437,5}} = 0,01448$$

$$t_b = \frac{1 - 0,99714}{0,01448} = \frac{0,00286}{0,01448} = 0,1975$$

Odsečak a odgovara sadržaju ureje u serumu i kad tu vrednost saberemo sa dodatim koncentracijama, dobićemo očekivane koncentracije. Nagib je skoro jednog jedinici, što pokazuje i standardna greška S_b koja nije značajna (izračunato t_b manje od kritične vrednosti za $\varphi = 4$ i $p = 0,05$). Zaključak je da ispitivana metoda daje tačne rezultata do koncentracije 31,3 mmol/L (dodato 25 mmol/L i sadržaj u serumu 6,3 mmol/L).

Zadaci za vežbanje

1. Izračunati jednačinu standardne krive za metodu za određivanje propranolol-hidrohlorida i iz nje izračunati koncentraciju nepoznatog uzorka čija je apsorbancija $A = 0,256$.

$\mu\text{mol/L:}$	50	100	150	200	250	300	350	400
A:	0,098	0,204	0,299	0,398	0,505	0,602	0,697	0,805

2. Metodom sa batofenantrolinom određen je slobodni kapacitet vezivanja gvožđa kao razlika ukupnog kapaciteta i serumskog gvožđa (X) i direktno bez određivanja ukupnog kapaciteta (Y). Pokazati da li je razlika između ove dve metode značajna.

X	20,5	42,8	35,4	8,6	55,6	30,8	29,4	41,5	36,2	15,4
Y	21,0	41,7	36,2	8,3	56,8	30,0	28,8	42,4	36,9	15,9

3. U 10 različitih uzoraka seruma određen je bilirubin metodom Jendrassik-Grof, a dobijene apsorbancije su prevedene u koncentracije preko standardne krive napravljene sa alkalnim standardnim rastvorima bilirubina (X) i sa albuminskim standardima bilirubina (Y). Treba pokazati da li je između dobijenih vrednosti razlika značajna, odnosno da li se krive između sebe razlikuju.

X:	17,10	25,24	8,65	9,50	32,40	20,80	16,35	68,72	12,45	48,55
Y:	18,90	27,16	9,60	10,55	35,50	22,98	18,10	75,60	13,65	53,55

4. Za određivanje neorganskog fosfata sa amonijum molibdatom na tržištu se nalaze gotovi test-reagensi dva proizvođača. Da bi se uporedila tačnost ova dva testa urađen je "recovery" dodavanjem standarda različitih koncentracija u serumski "pool". Pokazati da li ima razlike u vrednostima koje se dobijaju korišćenjem ova dva testa.

dodato	0	0,4	0,5	0,6	0,7	0,8	0,9	1,0
test A:	2,31	2,68	2,79	2,87	2,99	3,07	3,15	3,28
test B:	2,18	2,55	2,64	2,73	2,83	2,96	3,07	3,16