



Univerzitet u Novom Sadu
Prirodno matematički fakultet
Departman za matematiku i informatiku



Daniel Pavlica

Multivarijantna statistička kontrola procesa pomoću Hotelingove T^2 statistike

-Master rad-

Mentor: dr Zagorka Lozanov-Crvenković

Novi Sad, 2016

Sadržaj

Uvod	3
Statistička kontrola procesa i T^2 statistika.....	5
1.1 Ševartove karte (Shewhart charts)	5
1.2 Multivariantne kontrolne procedure	7
1.3 Karakteristike multivariantne kontrolne procedure	10
Osnovni pojmovi T^2 statistike	12
2.1 Statistička distanca.....	12
2.2 T^2 i multivariantna raspodela	16
2.3 Studentova t nasuprot Hotelingove T^2	18
2.4 Distribucionia svojstva T^2	21
Interpretacija T^2 signala za dve promenljive	25
3.1 Ortogonalna dekompozicija	25
3.2 MYT (Mejson-Jang-Trejsi) dekompozicija	31
3.3 Interpretacija signala preko T^2 komponenti	33
3.4 Regresija.....	35
3.5 Raspodela verovatnoća T^2 komponenti.....	38
Interpretacija T^2 signala za opšti slučaj.....	42
4.1 MYT dekompozicija za opšti slučaj.....	42
4.2 Računanje izraza dekompozicije	44
4.3 Bitna svojstva MYT dekompozicije.....	47
4.4 Lociranje signalnih promenljivih	49
4.5 Interpretacija signala na osnovu T^2 komponenti.....	52
4.6 Regresija sa više promenljivih	57
4.7 Šema računanja.....	58
Primer interpretacije T^2 signala za opšti slučaj	60
Zaključak	72
Literatura	73
Biografija	74

Uvod

Industrija se nalazi pod stalnim pritiskom da unapredi proizvodnju i poboljša kvalitet proizvoda. Postoji mnogo načina da se to uradi, ali je jako malo onih koji će pritom i smanjiti troškove proizvodnje. Zato se mnogo truda ulaže u pronalaženje i razvoj novih tehnoloških alatki za rešavanje ovih problema. Zahtev za kvalitetom star je gotovo kao i ljudska civilizacija i javlja se već u prvim oblicima "trgovine", odnosno materijalne razmene. Tokom vremena mnogo npora je uloženo u potvrđivanje kvaliteta i to ne samo proizvoda, već i proizvođača i isporučitelja. Možemo slobodno reći da kontrola kvaliteta ima dugu istoriju. Međutim, statistička kontrola kvaliteta je relativno nova metoda. Počela se uspešno primenjivati dvadesetih godina prošlog veka, kao rezultat prihvatanja teorije uzoraka. Pravilnom primenom na procesne promenljive moguće je smanjiti troškove proizvodnje. Koncept statističke kontrole procesa u proizvodnji prvi je uveo Volter A. Ševart*, koji je 1924. izumeo kontrolne karte.

Statistička kontrola procesa je skup metoda i procedura za prikupljanje, obradu, analizu, tumačenje i prikaz podataka. Koristi se u svrhu osiguranja kvaliteta proizvoda i procesa. Mehanizam statističke kontrole procesa zasniva se na upoređivanju podataka prikupljenih iz procesa sa unapred definisanim kontrolnim granicama za kvalitet proizvoda i proizvodnje, pa na osnovu toga donošenje zaključaka o samom procesu.

Statistička kontrola procesa može da bude delimična i potpuna. Izbor jednog od dva moguća tipa statističke kontrole zavisi, pre svega, od tehničkih uslova i visine troškova. Po pravilu, delimična kontrola se primenjuje u uslovima masovne proizvodnje, gde bi potpuna kontrola uzrokovala izuzetno visoke troškove. Statistička kontrola proizvodnog procesa sprovodi se tokom same proizvodnje, a osnovni cilj je da proizvodi odgovaraju unapred definisanom kvalitetu. U toku njenog izvođenja mogu se razlikovati tri faze :

- postavljanje plana kontrole i uspostavljanje kontrolnih granica
- praćenje procesa proizvodnje na toj osnovi
- preduzimanje odgovarajućih koraka korekcije kada se za to ukaže potreba.

Razlikujemo univariatne i multivariatne kontrolne procedure. Očigledna razika na prvi pogled je u broju promenljivih koje se posmatraju, ali to nije i jedina. Iako univarijante kontrolne procedure imaju široku primenu u industriji one su neadekvantne kada se radi kontrola procesa koji su po svojoj prirodi multivariatni. Ono što je potrebno je metodologija koja omogućava da pratimo veze koje postoje između procesnih promenljivih.

*Walter Andrew Shewhart (1891-1967), američki matematičar i statističar, poznat pre svega, kao autor koncepta statističkog upravljanja kvalitetom pomoću kontrolnih karata

Ovaj rad baviće se upravo tom temom i biće ograničen na strategiju baziranu na Hotelingovoj T^2 statistici. Glavni cilj ovog rada je da se predstavi moderan i celokupan pogled na to kako uspostaviti i raditi multivariantnu kontrolu procesa zasnovanu na konceptualnom pogledu na Hotelingovu T^2 statistiku.

U prvom delu rada će se govoriti o statističkoj kontroli procesa, gde će se spomenuti univariantna i biti obradjena multivariantna statistička kontrola procesa. Takođe, biće navedene njihove razlike i predstaviti T^2 statistika kao idealna kontrolna statistika za multivariantne procese. Potom, će biti izloženi osnovni koncepti T^2 statistike i statističke distance i biće predstavljena primena distributivnih svojstva T^2 statistike na multivariantnu statističku kontrolu procesa uz primere. U početku je rađena interpretacija T^2 signala za dve promenljive, bazirana na MYT (Mejson - Jang - Trejsi) dekompoziciji. Zatim, interpretacija se proširuje na opšti slučaj, tj. na više promenljivih. Pokazano je kako signal može biti izolovan na promenljivu ili grupu promenljivih, koje su odgovorne za signal.

Na kraju, biće uradjen primer u kojem se primenjuje multivariantna statistička kontrola procesa koja koristi Hotelingovu T^2 statistiku. U njemu će se proveriti da li vrednost T^2 statistike za prikupljene podatke upada u granice procesnih varijacija, koje su uspostavljene istorijskim podacima. Ako vrednost T^2 statistike ne upada u kontrolne granice, signal je detektovan. To se može desiti ako se vrednost neke promenljive naglo promeni ili ako se vrednost dve ili više promenljivih ne pridržava linearne korelacije koja je nametnuta istorijskim podacima. Biće analizirani odbačeni slučajevi i kao razlog za njihovo odbacivanje. Sva potrebna računanja i svi predstavljeni grafici urađeni su u programima *Statistica* i *Mathematica* i *Excel*.

GLAVA 1

Statistička kontrola procesa i T^2 statistika

Osnovi koncept statističke kontrole procesa temelji se na upoređivanju podataka prikupljenih iz procesa, koji se kontroliše, sa kontrolnim granicama, koje su izračunate na osnovu istorijskih podataka pa na osnovu toga donošenje zaključaka o samom procesu. Statistička kontrola procesa u proizvodnji koristi se kako bi se osigurao kvalitet proizvoda koji će zadovoljiti zahteve kupca, u pogledu kvaliteta i cene koštanja. Ako se kvalitet proizvoda nalazi u dozvoljenim granicama, smatra se da je proces pod kontrolom. Razlozi za primenu statističke kontrole procesa su sledeći :

- utvrđivanje sposobnosti procesa za proizvodnju proizvoda, koji zadovoljava zahteve kupca
- praćenje procesa kako bi se otkrile promene zbog kojih proces izmiče kontroli
- preduzimanje mera za korekciju procesa i njegovo održavanje pod kontrolom.

Statistička analiza može samo upozoriti na nastale promene, a moguće uzorce treba naknadno utvrditi. Mnogo je uloženo u elektronske sakupljače podataka zbog činjenice da odgovor na najveći deo industrijskih problema leži u prikupljanju uzorka i njihovoj obradi. Još više je uloženo u naprednu kontrolu procesa (NKP). Ovi sistemi su razvijeni i instalirani da se osigura održavanje procesnih promenljivih u predviđenom operacionom rangu. Rade odličan posao u smanjivanju sistemskih varijacija, jer ograničavaju operacioni rang promenljivih. Ipak, ni sistem NKP ne može da se koristi za otkrivanje uzroka nestabilnosti sistema.

Rešenje ovog problema leži u primeni multivarijantne statističke kontrole procesa u skladu sa elektronskim sakupljačem podataka i NKP sistemom. Takva procedura će signalirati procesne nestabilnosti, i u mnogim slučajevima se može koristiti predviđanje nestabilnosti pre nego što se izgubi kontrola. Kada se signali detektuju, procedura omogućava da se signal razloži po promenljivama koji ga uzrokuju.

1.1 Ševartove karte (Shewhart charts)

Prvi teorijski radovi i praktični pokušaji primene matematičke statistike i kontrole kvaliteta datiraju još iz 1924. godine, kada je Volter Ševart iz *Bell Telephone Laboratories* dao skicu prve kontrolne karte, koja zapravo predstavlja univarijantni kontrolni grafik. Iako su njegovi početni grafici pratili procenat oštećenja u procesu proizvodnje, kasnije je proširio ideju na kontrolne grafike za srednju vrednost i standardnu devijaciju proizvodnje. Kontrolne karte su osnovni

instrument pomoću kojega se sprovodi statistička kontrola proizvoda ili proizvodnog procesa. Osnovna uloga kontrolnih karata je u otkrivanju i vizualizaciji poremećaja kvaliteta proizvoda. Kontrona karta se sastoji od tri osnovne kontrolne granice :

- gornja kontrolna granica (*upper control limit - UCL*)
- srednja linija (*central line - CL*)
- donja kontrolna granica (*lower control limit LCL*).

Srednja linija se dobija na osnovu aritmetičke sredine uzorka veličine n , odnosno :

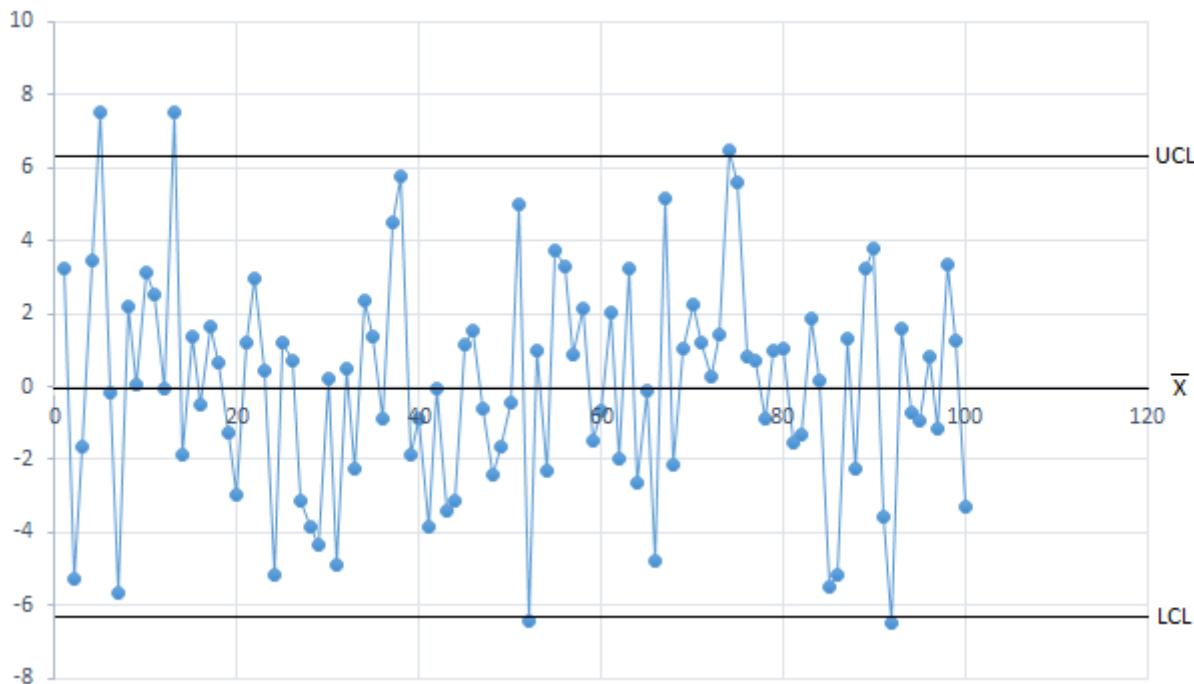
$$\bar{X} = \frac{\sum x_i}{n}$$

Dok se kontrolne granice izračunavaju na pomoću zakona verovatnoće, na osnovu raspodele uzorka. Gornja (*UCL*) i donja (*LCL*) kontrolna granica se računa :

$$UCL = \bar{X} + k * \sigma$$

$$LCL = \bar{X} - k * \sigma$$

gde je σ standarna devijacija uzorka, a k koeficijent. Kada je potrebna jača kontrola procesa za koeficijent se uzima $k = 2$, za koji će oko 5% merenja iz procesa, koji je pod kontrolom, produkovati signal. Najčešće se uzima $k = 3$, jer će se signal javiti samo kod merenja koji su ozbiljno van kontrole, tj. oko 1% merenja iz procesa, koji je pod kontrolom. Na Slici 1.1 prikazan je Ševartov grafik, dizajniran da prati srednju vrednost \bar{X} grupe procesnih merenja baziranih na procesnoj promenljivoj za isti vremenski period. Na grafiku su povučene gornja kontrolna granica (*UCL*) i donja kontrolna granica (*LCL*) kao i srednja linija (*CL*).



Slika 1.1 Ševartov grafik procesne promenljive

Ševartovi grafici se često koriste u otkrivanju neobičnih promena u promenjivama koje su nezavisne i na koje ne utiče ponašanje drugih promenljivih. Ove promene su česte u industriji. Na primer, uzmimo u obzir glavnu laboratoriju velike hemijske industrije. Pretpostavimo da merimo neku jednostavnu i poznatu koncentraciju. Ako rezultat merenja upada u kontrolne granice Ševartovih grafika, smatra se da oprema funkcioniše normalno. Inače, oprema se proverava i ponovo podešava i pravi se novi Ševatov grafik. Može se tvrditi da određeni hemijski proces zavisi od drugih faktora kao što su sobna temperatura ili vlažnost vazduha. Iako, ovi faktori mogu da utiču na određene delove procesa, kompenzacija se postiže kontrolisanjem temperature i vlažnosti. Stoga, ovi uticaji postaju zanemarljivi i proces se posmatra kao nezavisna promenljiva.

1.2 Multivarijantne kontrolne procedure

Postoji mnogo industrijskih oblasti gde se performanse procesa baziraju na ponašanju grupe promenljivih koje su međusobno korelisane. One su dizajnirane da izmene ulazne komponente u neku određenu izlaznu. Na primer, želimo da promenimo prirodni gas, tip energije, u drugi oblik kao što su para ili struja. Ili, možda želimo da konvertujemo slanu vodu u kaustičnu sodu i hlorni gas; pesak u silicijum ili staklo. Naš interes leži u tome da razvijemo kontrolnu proceduru koja će detektovati neobične događaje u tim promenljivama.

U ovim situacijama se ne može koristiti univarijantna kontrolna procedura zato što imamo više promenljivih koje formiraju korelisani skup. Kako promenljive ne variraju nezavisno jedna od druge, moraju se ispitivati zajedno kao grupa, a ne odvojeno.

Multivarijantni procesi se javljaju u mnogim industrijama, kao što je na primer hemijska, gde se ulazne komponente hemijski izmene kako bi se dobio izlazni proizvod. Na primer, već pomenuta proizvodnja hlornog gasa i kaustične sode. Ulazna promenljiva je zasićena slana voda. Pod odgovarajućim uslovima, deo slane vode se razlaže elektrolizom na hlorni gas dok se kaustična soda formira unutar rasola i kasnije se odvaja. Promenljive u ovom procesu prate određene matematičke veze i formiraju visoko korelisani skup.

Postoje dva tipa korelacije između promenljivih multivarijantnog sistema. Prvi tip često javlja zbog efekta neke promenljive koja nije merljiva. Na primer, sečiva gasne ili parne turbine će postati kontaminovana od prekovremenog rada. Iako akumulacija prljavštine nije merljiva, proizvodnja megavata će pokazati negativnu korelaciju sa dužinom vremena od poslednjeg čišćenja turbine. Primer drugog tipa korelacije je veza između temperature i pritiska, jer povećanje temperature dovodi i do promene pritiska. Takva korelacija sprečava da se promenljive ispituju pojedinačno pomoću univarijantne procedure osim ako se ne uzme u obzir i uticaj druge promenljive.

Analiza multivarijantne kontrolne procedure zahteva ispitivanje promenljivih u odnosu na veze koje postoje među njima. Pretpostavimo da analiziramo podatke koji se sastoje od 4 skupa očitavanja temperature i pritiska. Neka su date koordinatne tačke

Tačka 1 = (178, 76), Tačka 2 = (180, 80),

Tačka 3 = (170, 70), Tačka 4 = (172, 74).

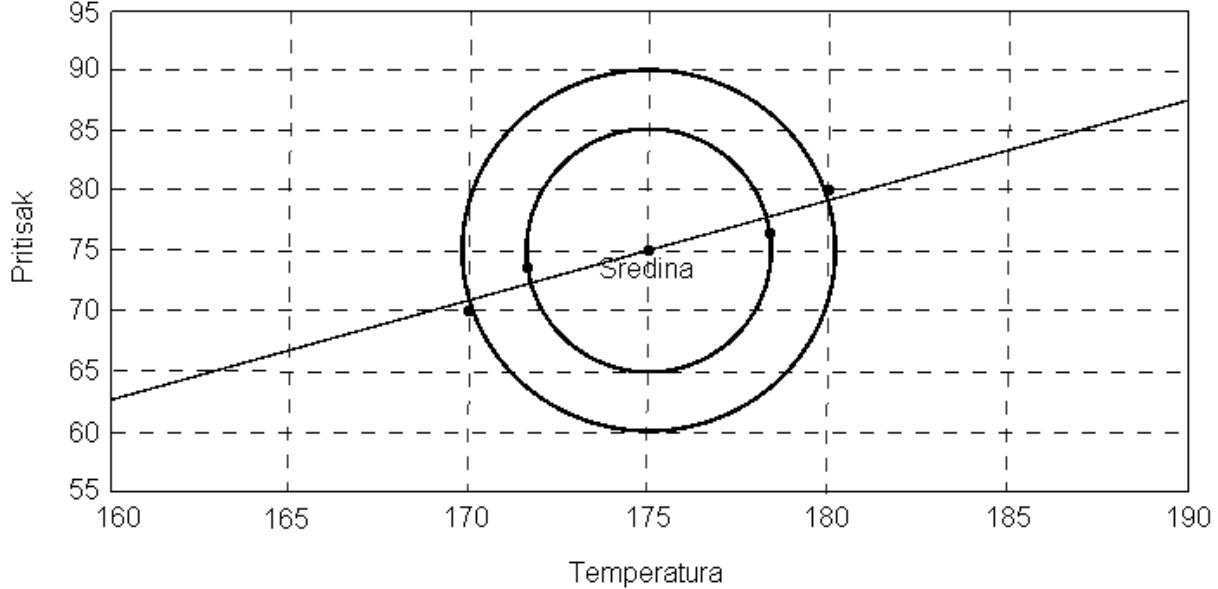
Gde vrednost prve koordinate predstavlja temperaturu, a druga vrednost predstavlja pritisak. Ove četiri tačke su ucrtane u grafik, kao i srednja vrednost (175, 75) na Slici 1.2. Takođe je fitovana linija kroz ove tačke i konstruisane su dve kružnice različite veličine oko srednje vrednosti. Ako se tačka srednje vrednosti posmatra kao tipičan deo uzorka, jedan oblik analize sastoji se od računanja udaljenosti svake tačke od srednje vrednosti. Udaljenost, recimo D, između bilo koje dve tačke, (a_1, a_2) i (b_1, b_2) , je data sa formulom

$$D = [(a_1 - b_1)^2 + (a_2 - b_2)^2]^{\frac{1}{2}}. \quad (1.1)$$

Ovaj tip mere udaljenosti je poznat kao Euklidova, pravolinijska distanca. Distanca koja postoji između svake od četiri tačke sa srednjom vrednosti je izračunata respektivno

$$(3.16, 7.07, 7.07, 3.16).$$

Iz ovih izračunavanja vidi se da su tačke 1 i 4 na jednakoj udaljenosti od srednje vrednosti, tj. nalaze se na kružnici sa centrom u tački srednje vrednosti i poluprečnikom 3.16. Slično tačke 2 i 3 se nalaze na istom rastojanju od srednje vrednosti i nalaze se na većoj kružnici sa poluprečnikom 7.07.



Slika 1.2 Grafik rasturanja ilustruje pravolinijsku distancu

Postoje dve velike kritike ovakve analize. Prvo, varijacije u obe promenljive su potpuno ignorisane. Sa Slike 1.2 vidi se da očitavanja temperature sadrže više varijacija nego očitavanja pritiska, ali ovo može da se desi i zbog razlike u skali između dve promenljive. Međutim, u ovom slučaju očitavanja

temperature zaista imaju veće varijacije. Druga kritika ove analize je da se korelacija između temperature i pritiska takođe ignorisana u potpunosti. Generalno, očekuje se da kada temperatura raste, povećava se i pritisak, što se može videti na osnovu prave data na Slici 1.2. Kada temperatura raste duž horizontalne ose, odgovarajuća vrednost pritiska se povećava duž vertikalne ose. Postavlja se pitanje može li mera udaljenosti biti osmišljena tako da računa na prisustvo linearne veze između odgovarajućih promenljivih i razlike u njihovim varijacijama? Odgovor je da, ipak, ta distanca je statističke, a ne Euklidova i nije laka za izračunavanje.

Za računanje statističke distance (SD) potrebno je znati meru korelacije između promenljivih. Ona je data sa kovarijansom, koja pokazuje kako promenljive variraju zajedno. Uzoračka kovarijansa između temperature i pritiska, označena sa s_{12} , računa se pomoću sledeće formule

$$s_{12} = \left(\frac{1}{n-1} \right) \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

gde x_1 predstavlja temperaturnu komponentu, x_2 pritisak, a broj posmatranja je dat sa n . Vrednost uzoračke kovarijanse između temperature i pritiska je izračunata i iznosi 18.67. Takođe, za računanje statističke distance potrebna je varijansa svake promenljive. Varijansa promenljive x je data sa

$$s^2 = \left(\frac{1}{n-1} \right) \sum_{i=1}^n (x_i - \bar{x}_i)^2.$$

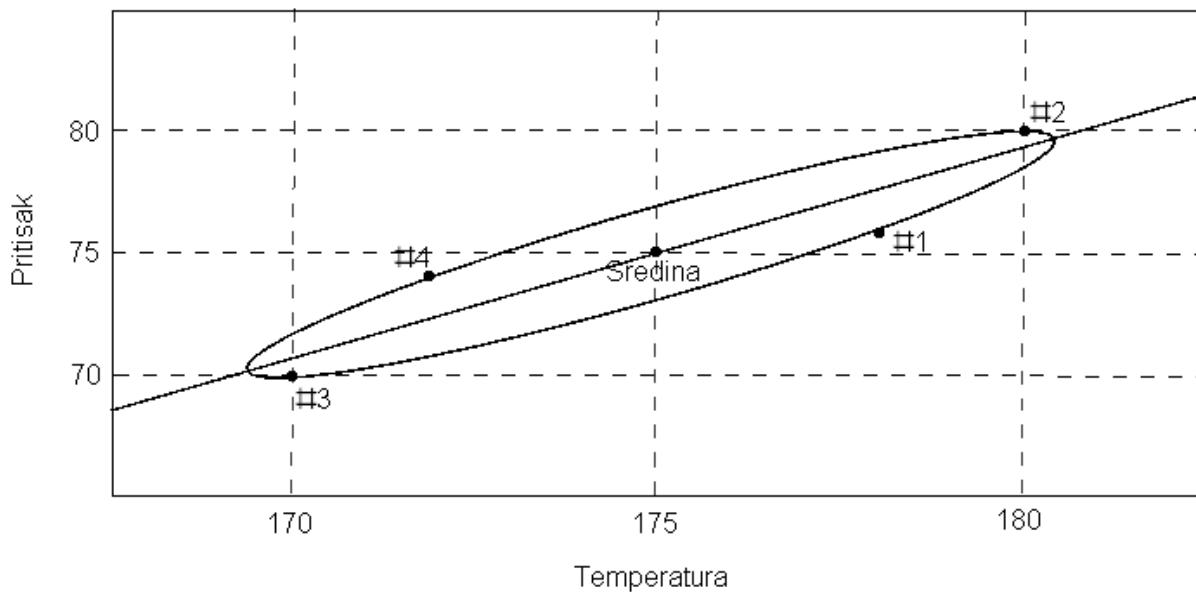
Uzoračka varijansa za temperaturu i pritisak je izračunata i iznosi 22.67 i 17.33 respektivno. Koristeći kovarijansu, uzoračke varijanse, kao i srednju vrednost možemo izračunati kvadratnu statističku distancu

$$(SD)^2 = \frac{1}{(1-r^2)} \left[\left(\frac{x_1 - \bar{x}_1}{s_1} \right)^2 - 2r \left(\frac{x_1 - \bar{x}_1}{s_1} \right) \left(\frac{x_2 - \bar{x}_2}{s_2} \right) + \left(\frac{x_2 - \bar{x}_2}{s_2} \right)^2 \right] \quad (1.2)$$

gde $r = s_{12}/s_1 s_2$ predstavlja koeficijent korelacije. Statistička distanca se dobija korenovanjem obe strane izraza. Kako je izraz (1.2) formula elipse, SD se zove i eliptička distanca. Takođe je poznata i kao Mahalanobisova distanca ili Hotelingova T^2 ili jednostavno T^2 . Računajući $(SD)^2$ za svaku od četiri tačke iz uzorka dobijaju se sledeći rezultati

$$(1.5, 1.5, 1.5, 1.5).$$

Ovom analizom je utvrđeno da su četiri tačke na istom statističkom rastojanju od srednje tačke. Ovaj rezultat je ilustrovan grafički na Slici 1.3. Sve četiri tačke zadovoljavaju jednačinu elipse date na grafiku. Vizualno, ovi rezultati izgledaju nerazumno. Očigledno je da su tačke 1 i 4 bliže tački srednje vrednosti u Euklidovoj distanci nego tačke 2 i 3. Međutim, kada se razlike u varijacijama promenljivih i veze između promenljivih uzmu u obzir, statističke distance su iste. Multivariantne kontrolne procedure predstavljene u ovom radu su razvijene metodama baziranim na konceptu statističke distance.



Slika 1.3 Grafik rasipanja ilustruje statističku distancu

1.3 Karakteristike multivarijantne kontrolne procedure

Postoji najmanje pet poželjnih karakteristika multivarijantne kontrolne procedure. Neke od njih su:

1. Statistika nadgledanja se lako predstavlja na grafiku i pomaže pri identifikovanju kretanja procesa.
2. Kada se javljaju tačke bez kontrole, lako se otkriva promenljiva koja je uzrok toga.
3. Procedura mora biti fleksibilna u primeni.
4. Procedura mora biti osetljiva na male, ali konstantne procesne promene.
5. Procedura mora biti u mogućnosti da prati procese dok je proizvodnja aktivna, ali isto tako i dok nije.

Za jasno razumevanje kontrolne procedure, obrada podataka mora biti globalna, a ne izolowana na određeni skup podataka. Iz tako obrađenih podataka moguće je detektovati čudno ponašanje promenljivih koje dovode do nepravilnosti ili čak haotičnih situacija. Dobra grafička metoda ne samo da pruža brzu detekciju signala nego pomaže pri identifikovanju procesnih kretanja. Statističkim ispitivanjem grafika tokom vremena, posmatra se ponašanje procesa i nepravilnosti se identifikuju u obliku signala, tj. tačke van kontrole.

Kontrolne procedure bi trebalo da funkcionišu sa procesima koji su vremensko zavisni, ali isto tako sa onim koji ne zavise od vremena, trebalo bi da budu primenljive na neprekidne procese kao i na serijske. Trebalo bi da budu dovoljno fleksibilne da rade sa različitim formama kontrolne

statistike, kao što su srednja vrednost uzorka ili individualna posmatranja, i da rade sa različitim ocenjivačima unutrašnje strukture promenljivih.

Iako postoji mnogo multivariantnih kontrolnih procedura, malo je onih koje zadovoljavaju sve navedene karakteristike. Procedure građene na T^2 statistici imaju sve navedene karakteristike. Kao i mnoge multivariantne grafičke statistike, T^2 statistika je univariantna. Ovo je tačno bez obzira na broj procesnih promenljivih koje se koriste za računanje. Ipak, zbog njenih sličnosti sa univariantnim Ševertovim graficima, T^2 kontrolni grafik se ponekad zove i multivariantni Ševertov grafik. Ova veza sa običnim univariantnim grafičkim procedurama olakšava shvatanje ove grafičke metode.

Interpretacija signala zahteva proceduru za izolovanje doprinosa svake promenljive ili određene grupe promenljivih. Kao sa univariantnom kontrolom, situacije bez kontrole mogu se pripisati individualnim promenljivama, koje se nalaze van operativne oblasti. Na primer, ako bi u prethodnom primeru temperatura bila previsoka. Drugi uzrok multivariantnog signala može se pripisati narušenoj vezi između dve ili više promenljivih, kao npr. pritisak nije na onoj vrednosti na kojoj bi trebao da bude za dato očitavanje temperature.

Većina industrija je orijentisana na obim proizvodnje. Sitne promene u efikasnosti mogu napraviti razliku u stvaranju profita ili generisanju gubitka. Male, neprekidne procesne promene na T^2 grafiku mogu se detektovati koristeći određene komponente dekompozicije T^2 statistike. Ovo se postiže praćenjem rezidualnih grešaka povezanih sa ovim uslovima. Procedure za interpretaciju signala obrađene u ovo radu su sposobne da razdvoje T^2 vrednost na nezavisne komponente. Jedan tip komponenti determiniše doprinos individualnih promenljivih za pojavu signala, dok druge komponente proveravaju vezu među grupama promenljivih.

GLAVA 2

Osnovni pojmovi T^2 statistike

U ovoj glavi će biti predstavljene osnovni pojmovi u vezi T^2 statistike, koje će se primenjivati pri konstruisanju multivariantnih kontrolnih grafika. Počinje se sa statističkom distancom i kako je ona povezana sa T^2 statistikom. Važan deo ove glave će biti objašnjavanje razlika između statističke distance i pravolinijske Euklidove distance. Kako je T^2 statistika zapravo uopštenje Studentove t statistike, potrebno je razumeti vezu između univariantne Studentove t i njene analogne multivariantne T^2 statistike. Pomoću te veze moguće je razumeti funkcije verovatnoće koje se koriste za opisivanje T^2 statistike pod različitim uslovima.

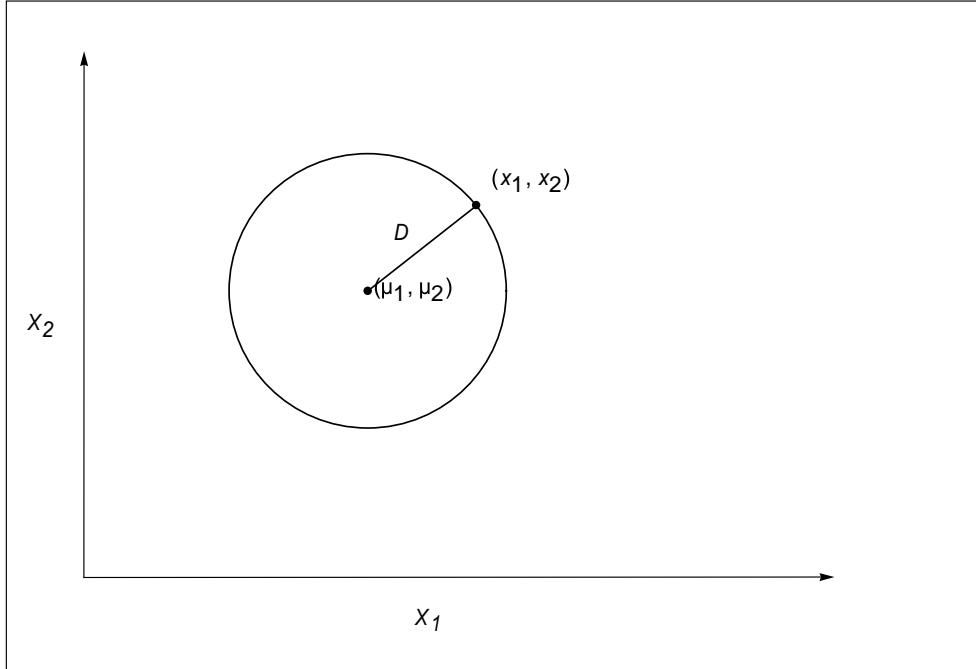
2.1 Statistička distanca

Među prvim dokumentima o proučavanju problema analize koreliranih promenljivih sa perspektive statističke kontrole je Hotelingov dokument iz 1947. godine o korišćenju multivariantnih procedura u analizi bombaških podataka. Njegova kontrolna procedura je bazirana na grafičkoj statistici, koju je predstavio u ranijem radu iz 1931. godine o generalizaciji Studentove t statistike. Statistika je kasnije dobila ime po njemu Hotelingova T^2 . Mahalanobis je 1930. godine predložio korišćenje slične statistike, koja će kasnije biti zvana Mahalanobisova mera distance, za merenje kvadratnog rastojanja između dve populacije. Iako ove dve statistike razlikuje samo konstantna vrednost, T^2 forma je najpopularnija u multivariantnoj kontroli procesa i glavna je tema ovog rada.

Označimo multivariantan vektor sa p promenljivih sa $X' = (x_1, x_2, \dots, x_p)$. Cilj je da se što bolje obrade informacije, koje su sadržane u ovih p promenljivih. Jedan pristup je upotreba grafičke metode, ali ucrtavanje tačaka u p -dimenzionalnom prostoru za $p > 3$ je izuzetno ograničen. Druga metoda za proučavanje informacija iz p -dimenzionalnog opažanja je da redukujemo multivariantni vektor podataka na jednu univariantnu statistiku. Ako rezultujuća statistika sadrži informacije o svakoj od p promenljivih, može se tumačiti i koristiti u donošenju odluka o statusu procesa. Postoje brojne procedure za dobijanje ovog rezultata, a neke od njih su statistička i Euklidova distanca.

Pretpostavimo da u procesu posmatramo nekorelisane bivariantne promenljive (x_1, x_2) . Takođe, pretpostavimo da nam je od interesa da odredimo distancu određene tačke od tačke srednje vrednosti. Distanca između dve tačke je uvek jedan broj bez obzira na dimenziju, odnosno broj promenljivih sa kojima se radi. Uobičajena pravolinijska (Euklidova) distanca meri distancu između

dve tačke po broju jedinica koje ih razdvajaju. Kvadratna pravolinijska distanca, recimo D , između tačaka (x_1, x_2) i populacione srednje vrednosti (μ_1, μ_2) je definisana sa



Slika 2.1 Oblast iste pravolinijske dinstance

$$(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2 = D^2$$

Primetimo da smo uzeli bivarijantni vektor (x_1, x_2) i konvertovali ga u jedan broj D , distanca opažaja se meri u odnosu na tačku srednje vrednosti. Ako je distanca D fiksirana, sve tačke, koje su na istoj distanci od tačke srednje vrednosti, mogu se predstaviti na kružnici sa centrom u tački srednje vrednosti i poluprečnikom D (Slika 2.1). Takođe, svaka tačka koja se nalazi u krugu ima distancu manju od D . Međutim, za većinu statističkog posla Euklidova mera distance nije zadovoljavajuća. Iako, svaka koordinata doprinosi jednakom determinisanju pravolinijske distance, nije se razmišljalo o razlici u varijacijama dve promenljive, merene njihovim varijansama, σ_1^2 i σ_2^2 respektivno. Kako bi se ispravio ovaj nedostatak, uvode se standardizovane vrednosti

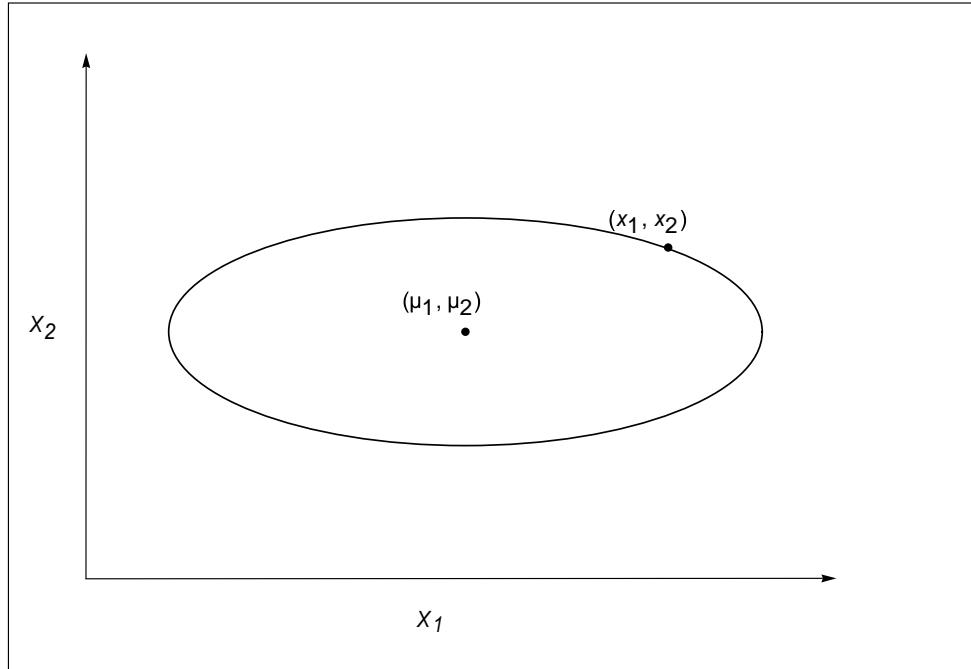
$$\frac{(x_1 - \mu_1)}{\sigma_1}, \frac{(x_2 - \mu_2)}{\sigma_2}$$

I sve tačke zadovoljavaju uslov

$$\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} = (SD)^2. \quad (2.1)$$

Ako se korenuje prethodni izraz dobija se statistička distanca između tačke (x_1, x_2) i tačke srednje vrednosti (μ_1, μ_2) . Za fiksiranu vrednost SD , sve tačke koje zadovoljavaju (2.1) su na istoj statističkoj distanci od tačke srednje vrednosti. Grafik takvog skupa tačaka formira elipsu, kao što je

ilustrovano na Slici 2.2. Svaka tačka unutar elipse imaće manju statističku distancu od SD , dok će svaka tačka van elipse imati veću statističku distancu SD .



Slika 2.2 Oblast iste statističke distance

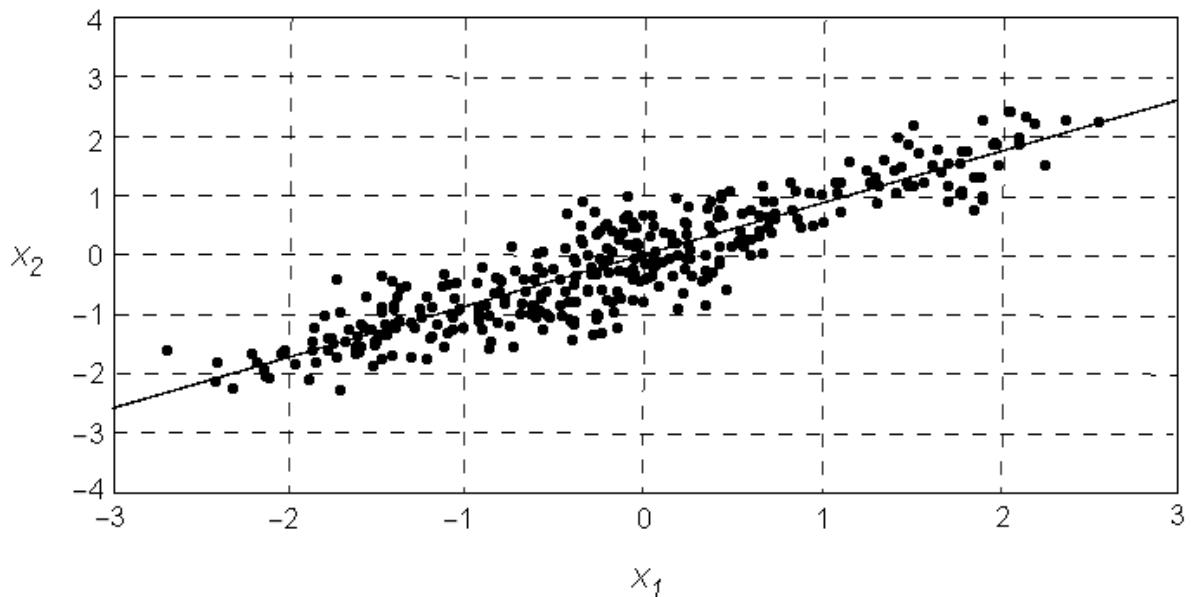
Upređujući statističku distancu sa pravolinijskom, uočavamo da bilo koje dve tačke na elipsi sa Slike 2.2 imaju istu SD , ali mogu da imaju različite Euklidove distance od tačke srednje vrednosti. Ako dve promenljive imaju iste varijanse i nisu korelisane, statistička i Euklidova distanca, ako izuzmemmo konstantnog množitelja, biće iste, u drugim slučajevima su različite. Glavna razlika između statističke i Euklidove distance je ta da dve promenljive korišćene u statističkoj distanci su značajne obrnuto po njihovoj standardnoj devijaciji, dok su obe promenljive jednakog značaja u pravolinojskoj distanci. Zato, promene u promenljivama sa malom standardnom devijacijom će doprinositi više statističkoj distanci nego promene u promenljivama sa velikom standardnom devijacijom kako bi se kompenzovala značajnost u odnosu na veličinu.

Prepostavimo sada da su dve promenljive koje posmatramo korelisane. Grafik rasipanja za dve pozitivno korelisane promenljive je predstavljen na Slici 2.3. Da bismo konstruisali meru statističke distance za ove podatke potrebno je da izvršimo uopštavanje izraza (2.1).

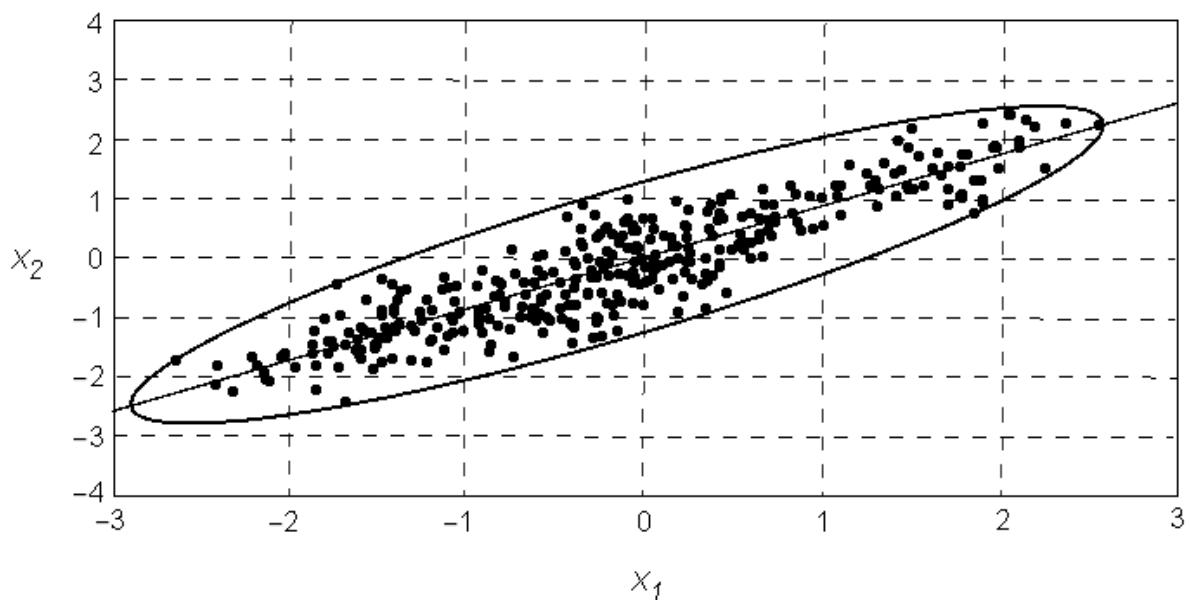
Iz analitičke geometrije, opšta jednačina elipse je data sa

$$a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 = c \quad (2.2)$$

Gde je a_{ij} određena konstanta koja zadovoljava $(a_{12}^2 - 4a_{11}a_{22}) < 0$, dok je c fiksna vrednost. Ako pravilno odaberemo a_{ij} iz (2.2), možemo da rotiramo elipsu kad je rasipanje dve promenljive fiksirano, dok ne dobijemo odgovarajuće poravnjanje. Na primer, elipsa data na Slici 2.4 sa centrom u srednjoj vrednosti dve promenljive je rotirana tako da se reflektuje korelacija među njima.



Slika 2.3 Grafik rasipanja korelisanih promenljivih



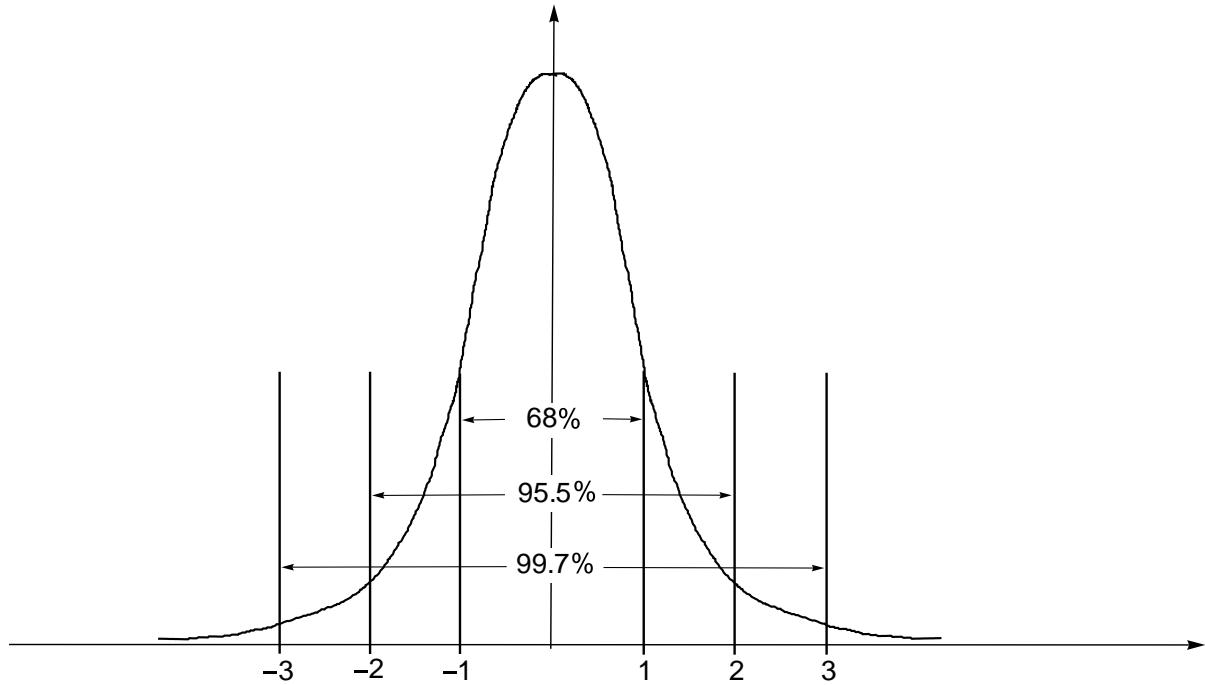
Slika 2.4 Eliptička oblast koja obuhvata sve tačke

2.2 T^2 i multivarijantna raspodela

Slučajna promenljiva X sa gustinom raspodele $f: \mathbb{R} \rightarrow \mathbb{R}_+$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

ima normalnu raspodelu sa parametrima μ, σ , tj. $N(\mu, \sigma^2)$, gde je μ očekivanje i σ standardna devijacija. Ako važi da je $\mu = 0$ i $\sigma = 1$ onda se govori o standardnoj normalnoj raspodeli, koja je prikazana na Slici 2.5. Intervali na rastojanju 1,2 i 3 standardne devijacije od očekivanja, koje je 0, zauzimaju redom 68%, 95.5% i 99.7% površine ispod krive. Isti procenti važe za svaku normalnu raspodelu, bez obzira na matematičko očekivanje i standardnu devijaciju.



Slika 2.5 Grafik normalne raspodele

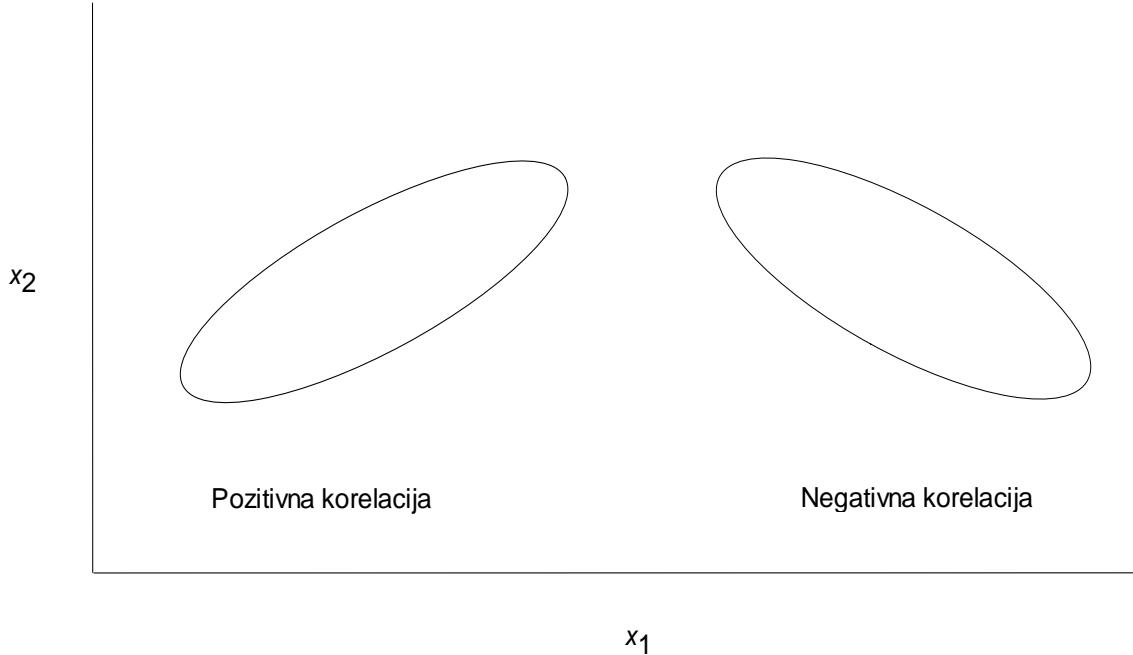
Prepostavimo da su promenljive x_1 i x_2 korelisane i da se vektor (x_1, x_2) može zajednički opisati pomoću bivarijantne normalne raspodele. Pod ovom prepostavkom, statistička distanca između ove tačke i vektora srednje vrednosti (μ_1, μ_2) , je promenljivi deo eksponenta funkcije gustine verovatnoće bivarijantne normalne raspodele

$$f(x_1, x_2) = \left[\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \right] e^{-\frac{1}{2}(SD)^2} \quad (2.3)$$

gde je $x_i \in \mathbb{R}$ za $i = 1, 2$, i gde $\sigma_i > 0$ predstavlja standardnu devijaciju x_i . Vrednost $(SD)^2$ je data sa

$$\frac{1}{(1 - \rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] = (SD)^2 \quad (2.4)$$

gde je $\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$ i predstavlja korelaciju između dve promenljive, $-1 < \rho < 1$.



Slika 2.6 Korelacione njihove elipse

Zbog činjenice da su dve promenljive x_1 i x_2 korelisane, u izrazu (2.4) postoji međusobni proizvod, tj. $-2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right)$. Ako je korelacija pozitivna, elipsa će biti nagnuta na gore udesno, a ako je negativana, elipsa će biti nagnuta na dole udesno. Ovo je predstavljeno na Slici 2.6. Neka je $\rho = 0$, tj. ne postoji korelacija između x_1 i x_2 , elipsa će biti orijentisana isto kao ona na Slici 2.2.

Jednačina (2.4) može se zapisati u matričnom obliku

$$(X - \mu)' \Sigma^{-1} (X - \mu) = (SD)^2, \quad (2.5)$$

gde je $X' = (x_1, x_2)$, $\mu' = (\mu_1, \mu_2)$. Σ je matrica kovarijansi između x_1 i x_2 , a Σ^{-1} je njena inverzna matrica

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix},$$

$$\Sigma^{-1} = \frac{1}{1 - \rho^2} \begin{bmatrix} \sigma_{22} & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_{11} \end{bmatrix},$$

gde su σ_1^2 i σ_2^2 varijanse promenljivih, a $\sigma_{12} = \sigma_{21} = \rho\sigma_1\sigma_2$ kovarijansa između x_1 i x_2 . Izraz (2.5) je oblik Hotelingove T^2 statistike.

Jednačine za konture gustine normalne bivarijantne raspodele dobijaju se fiksiranjem vrednosti SD u jednačini (2.4). Put po kojem se kreće tačka $X' = (x_1, x_2)$ kroz funkciju verovatnoće pri konstantnoj visini je elipsa. Elipse sa konstantnom gustinom se zovu i konture i mogu se determinisati matematički da sadrže fiksani nivo verovarneće.. Eliptička kontura predstavlja sve tačke koje imaju istu statističku distancu ili vrednost T^2 statistike.

Za uopšteni slučaj gde je vektor $X' = (x_1, x_2, \dots, x_p)$ opisan pomoću p-varijantne normalne funkcije verovatnoće date sa

$$f(X) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}T^2}, \quad (2.6)$$

gde $x_i \in \mathbb{R}$ za $i = 1, 2, \dots, p$. Vektor srednje vrednosti za X' je dat sa $\mu' = (\mu_1, \mu_2, \dots, \mu_p)$ i matrica kovarijansi je data sa

$$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}.$$

Dijagonalni elementi σ_{ii} matrice Σ predstavljaju varijanse i-tih promenljivih, a elementi van glavne dijagonale predstavljaju kovarijansu između i-te i j-te promenljive. Primetimo da je Σ nesingularna, simetrična, i pozitivno definitna matrica. Sa ovom postavkom, jednačina za elipsoidnu konturu multivariatne normalne raspodele u (2.6) je data kao

$$T^2 = (X - \mu)' \Sigma^{-1} (X - \mu),$$

gde je T^2 oblik Hotelingove T^2 statistike. Kao i u bivarijantnom slučaju elipsoidna oblast sadrži fiksni procenat multivariatne normalne raspodele i može se tačno odrediti.

2.3 Studentova t nasuprot Hotelingove T^2

Studentova t statistika je jedna od najpoznatijih statistika. Ako je da se uporedi sa Studentovom t raspodelom, statistika je dobijena iz proizvoljnog uzorka veličine n , uzetih iz populacije koja ima normalnu raspodelu sa srednjom vrednosti μ i varijansom σ^2 . Data je formulom

$$t = \frac{(\bar{x} - \mu)}{s/\sqrt{n}}, \quad (2.8)$$

gde je $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ srednja vrednost uzorka, dok je $s = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / (n - 1)}$ je standardna devijacija uzorka. Kvadrat t statistike je

$$\begin{aligned} t^2 &= \frac{(\bar{x} - \mu)^2}{\frac{s^2}{n}} \\ &= n(\bar{x} - \mu)(s^2)^{-1}(\bar{x} - \mu), \end{aligned} \quad (2.9)$$

i njena vrednost je definisana kao kvadratna statistička distanca između srednje vrednosti uzorka i srednje vrednosti populacije. Ako pogledamo prethodni izraz, vidimo da je brojilac zapravo kvadratna Euklidova distanca između \bar{x} i μ , pa se zbog toga zove mera bliskosti srednje vrednosti uzorka i srednje vrednosti populacije i kako se \bar{x} približava μ tako se vrednost t^2 približava nuli. Deljenjem kvadratne Euklidove distance sa ocjenjom varijansom srednje vrednosti uzorka \bar{x} , odnosno s^2/n , dobija se kvadratna statistička distanca. Hoteling je 1931 proširio univarijantnu t statistiku na multivarijantni slučaj koristeći T^2 statistiku baziranu na uzoračkoj oceni, umesto na poznatim vrednostima, kovarijansne matrice. Njegovo izvođenje je opisano u daljem tekstu.

Recimo da imamo uzorak veličine n , X_1, X_2, \dots, X_n , gde je $X'_i = (x_{i1}, x_{i2}, \dots, x_{ip})$, $i = 1, 2, \dots, n$ uzeto iz p-varijantne normalne raspodele sa srednjim vektorom μ' i matricom kovarijansi Σ . Uopštavanje t^2 statistike je dato sa

$$T^2 = n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu), \quad (2.10)$$

gde su \bar{X} i S ocenjivači za μ i Σ i definisani su sa

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i \\ S &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'. \end{aligned} \quad (2.11)$$

U pogledu raspodele verovatnoće, kvadrat t statistike u (2.9) ima oblik

$$\begin{aligned} t^2 &= (\text{normalna slučajna promenljiva}) * \left(\frac{hi - \text{kvadrat slučajne promenljive}}{df} \right)^{-1} \\ &\quad * (\text{normalna slučajna promenljiva}), \end{aligned}$$

gde df predstavlja $n - 1$ stepen slobode hi-kvadratne promenljive, $(n - 1)s^2/\sigma^2$, i normalna slučajna promenljiva je data sa $\sqrt{n}(\bar{x} - \mu)/\sigma$. U ovakvoj postavci, slučajna promenljiva x , i slučajna promenljiva s^2 su statistički nezavisne. Slično T^2 statistika iz (2.10) može se izraziti kao

$$T^2 = (\text{multivariantni normalni vektor})' * \left(\frac{\text{Višartova matrica}}{df} \right)^{-1} * (\text{multivariantni normalni vektor}), \quad (2.12)$$

gde df predstavlja $n - 1$ stepen slobode Višhartove promenljive, $(n - 1)S$, i multivariantni normalni vektor je $\sqrt{n}(\bar{X} - \mu)$. Proizvoljni vektor \bar{X} i proizvoljna matrica S su statistički nezavisni.

Višhartova raspodela je generalizacija univariantne hi-kvadratne raspodele na dve ili više promenljivih. Dobila je ime u čast Džona Višarta*, koji je prvi formulisao raspodelu 1928. godine. To je raspodela za simetrične pozitivne semidefinitne matrice, uglavnom kovarijansne matrice, dijagonalni elementi su proizvoljne hi-kvadratne promenljive. Na isti način kako se hi-kvadratna raspodela može konstruisati sumiranjem kvadrata proizvoljnih nezavisnih univariantnih normalnih promenljivih, koje imaju srednju vrednost 0, Višhartova raspodela se može konstruisati sumiranjem unutrašnjih proizvoda proizvoljnih nezavisnih multivariantnih normalnih vektora sa srednjom vrednosti 0. Višartova funkcija gustine je data sa

$$W_{n-1}(S) = \frac{|S|^{1/2(n-p-2)} e^{-\operatorname{tr} S \Sigma^{-1}/2}}{2^{\frac{p(n-1)}{2}} \pi^{\frac{p(p-1)}{4}} |\Sigma|^{\frac{n-1}{2}} \prod_{i=1}^p \Gamma\left(\frac{1}{2}(n-i)\right)}.$$

Matrica S je pozitivno definitna, a $\Gamma(\cdot)$ je gama funkcija. Za razliku od multivariantne normalne raspodele, Wišartova funkcija gustine ima jako malu primenu osim u teoretskim izvođenjima. Za slučaj kada je $p = 1$, može se pokazati Wišartova raspodela može svesti na konstantnu višestruku univariantnu hi-kvadratnu raspodelu. Zbog ovog razloga se smatra da je Wišartova raspodela multivariantni analog hi-kvadratne raspodele. Koristeći dve forme predstavljene u (2.10) i (2.12), moguće je proširiti Hotelingovu T^2 statistiku da predstavlja kvadratnu statističku distancu između različitih kombinacija p-dimenzionalnih tačaka. Na primer, jedna može da koristi T^2 statistiku da nade statističku distancu između individualnog vektora X i njegove poznate populacione srednje vrednosti μ ili njene ocene \bar{X} . Hotelingova T^2 statistika takođe se može izračunati između srednje vrednosti podgrupe uzorka \bar{X}_t i ukupne srednje vrednosti svih podgupa \bar{X} .

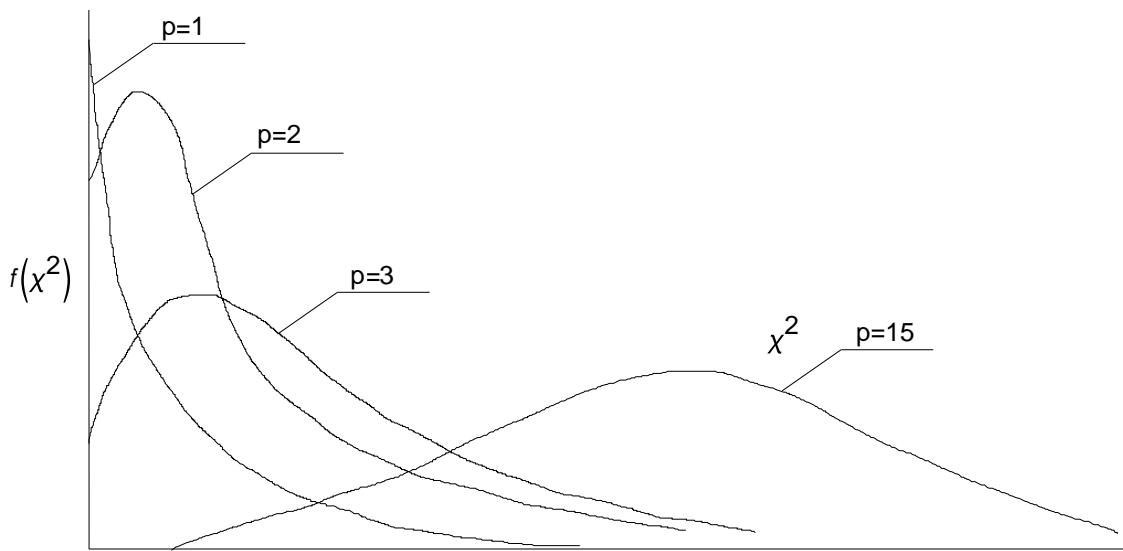
John Wišart (1898-1956) je bio Škotski matematičar i statističar agrikulture. Kao lider u statistici na Univerzitetu Kembridž postao je prvi direktor Statističke Laboratorije 1953. godine.

2.4 Distribucionia svojstva T^2

Jedna od osnovnih prepostavki bilo koje primene Hotelingove T^2 statistike je da se radi o proizvolnjom uzorku p-varijantne normalne populacije, čiji je vektor srednje vrednosti μ , a matrica kovarijansi Σ . Zato se ponašanje nezavisnih elemenata iz uzorka može opisati funkcijom verovatnoće sa parametrima poznatim ili nepoznatim. Ako su parametri nepoznati, prepostaviće se da postoje istorijski podaci, koji su sakupljeni pod stabilnim uslovima kada je proces bio pod kontrolom. Ovi podaci se koriste kako bi se napravile ocene nepoznatih parametara.

Ono što je potrebno uraditi je transformisati ove p-varijantne vektor iz uzorka u jednu Hotelingovu T^2 statistiku. Kako su originalne promenljive proizvoljne, ove nove T^2 vrednosti su takođe proizvoljne i mogu se opisati sa odgovarajućom funkcijom verovatnoće. Na primer, kada su parametri osnovne multivariatantne normalne raspodele nepoznati i moraju biti ocenjeni, koristi se neki oblik univariatne F raspodele za opis proizvoljnog ponašanja T^2 statistike. Ovo je takođe primenljivo i u univariatnom slučaju. Ako se t statistika iz (2.8) može opisati sa t raspodelom sa $(n - 1)$ stepeni slobode, kvadrat statistike, t^2 iz (2.9), može se opisati pomoću F raspodelesa 1 i $(n - 1)$ stepeni slobode. T^2 statistika može se izračunati koristeći jedno posmatranje napravljeno na p komponenti pri fiksnoj uzoračkoj tački, ili se može izračunati koristeći sredinu uzorka veličine m uzete tokom fiksnog perioda.

Određene prepostavke, koje upravljaju raspodelom T^2 statistike, se razlikuju u dva slučaja : parametri μ i Σ osnovne raspodele su ili poznati ili nisu. Drugi slučaj, kada parametri moraju biti ocenjeni, takođe ima dve različite situacije. Prvi se javlja kada je posmatrani vektor X nezavisan od ocenjivača parametara. Nezavisnost se javlja kada X nije uključeno u računanje \bar{X} ili S, uzoračkih ocena za μ i Σ . Druga situacija se javlja kada se X koristi u računanju ocenjivača i zato nije nezavistan od njih.



Slika 2.7 Hi-kvadratna raspodela

Nekoliko različitih funkcija verovatnoće se mogu koristiti za opisivanje T^2 statistike, u ovom radu su predstavljene tri forme, zajednosa uslovima u kojima su primenljive.

- (1) Prepostavimo da su parametri μ i Σ osnovne multivarijantne normalne raspodele poznati. T^2 statistika za individualna posmatranja vektora X ima formu i raspodelu datu sa

$$T^2 = (X - \mu)' \Sigma^{-1} (X - \mu) \sim \chi_{(p)}^2, \quad (2.13)$$

gde $\chi_{(p)}^2$ predstavlja hi-kvadratnu raspodelu sa p stepeni slobode. T^2 raspodela zavisi samo od broja p , koji predstavlja broj promenljivih u posmatranom vektoru X . Grafici hi-kvadratne funkcije raspodele, $f(\chi^2)$, za različite vrednosti p su predstavljene na Slici 2.7. Za manje vrednosti p , posmatramo iskrivljenu raspodelu sa dugim repom udesno; više simetrična forma je posmatrana za veće vrednosti p .

- (2) Prepostavimo da su parametri osnovne multivarijantne normalne raspodele nepoznati i da su ocjenjeni koristeći \bar{X} i S , koji su predstavljeni u (2.11). Ove vrednosti su dobijene koristeći istorijske podatke, koji se sastoje od uzorka veličine n . Forma i raspodela T^2 statistike za individualno posmatranje vektora X , nezavisnog od \bar{X} i S , je data sa

$$T^2 = (X - \bar{X})' S^{-1} (X - \bar{X}) \sim \left[\frac{p(n+1)(n-1)}{n(n-p)} \right] F_{(p,n-p)}, \quad (2.14)$$

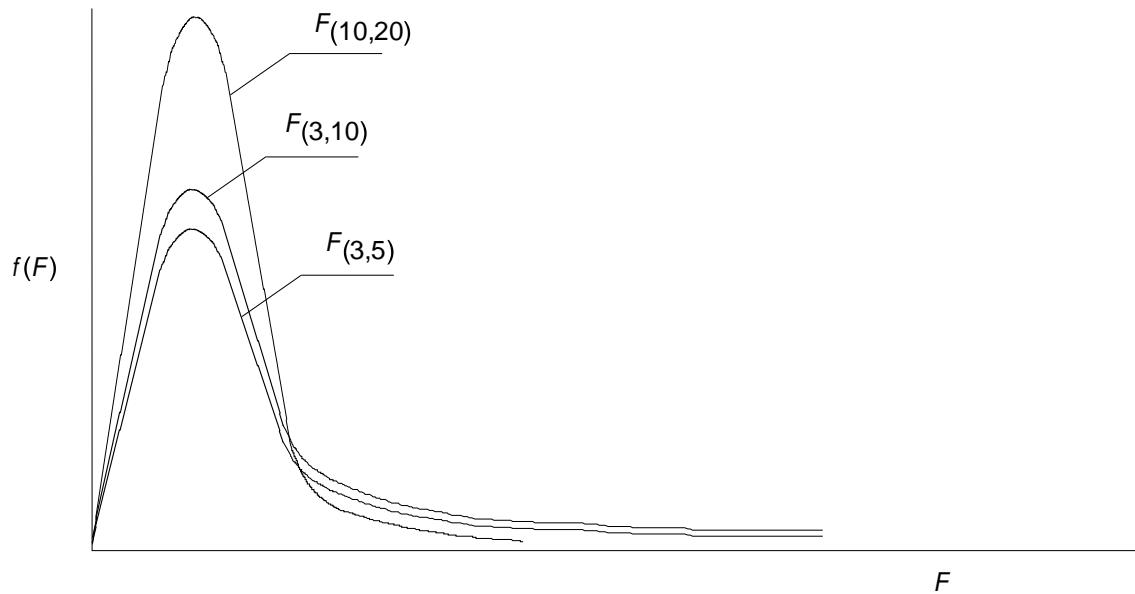
gde $F_{(p,n-p)}$ je F raspodela sa p i $(n-p)$ stepeni slobode. Raspodela iz (2.14) zavisi od veličine uzorka iz istorijskih podataka kao i od broja promenljivih koje se posmatraju.

Grafici F raspodele, $f(F)$, za različite vrednosti stepeni slobode imenioča i brojčića su predstavljeni na Slici 2.8. Opet, posmatrali smo iskrivljenu raspodelu sa dugim repom udesno.

- (3) Prepostavimo da posmatrani vektor X nije nezavisan od ocenjivača \bar{X} i S , već se koristi u njihovom računanju. U ovoj situaciji forma i raspodela T^2 statistike je data sa

$$T^2 = (X - \bar{X})' S^{-1} (X - \bar{X}) \sim \left[\frac{(n-1)^2}{n} \right] B_{(p/2, (n-p-1)/2)}, \quad (2.15)$$

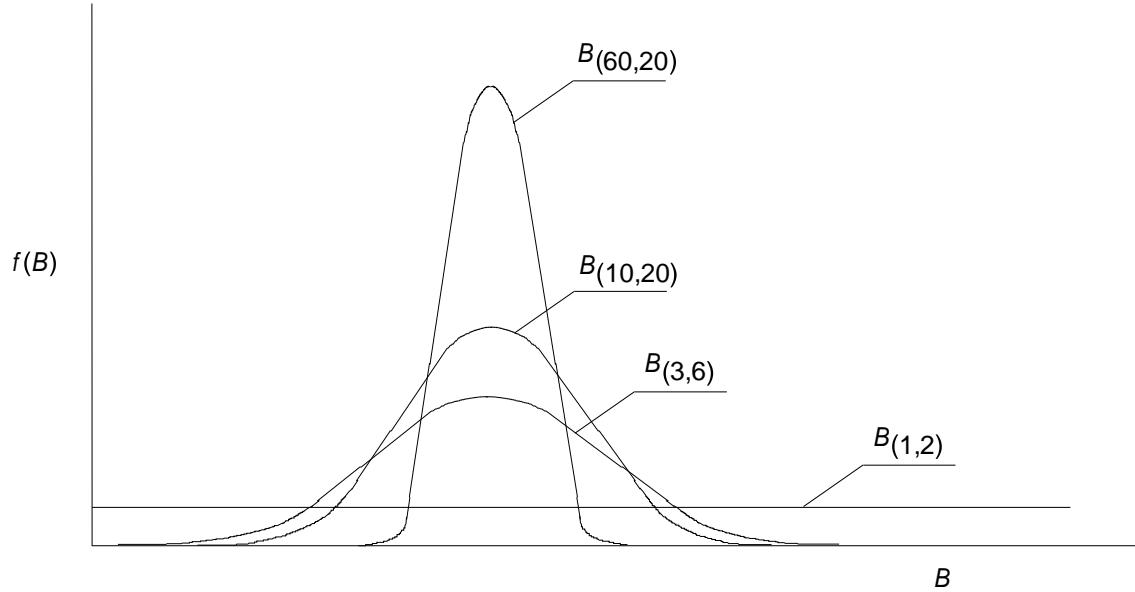
gde $B_{(p/2, (n-p-1)/2)}$ predstavlja beta raspodelu sa parametrima $p/2$ i $(n-p-1)/2$. Raspodela data u (2.15) zavisi od broja promenljivih p , i od veličine uzorka n iz istorijskih podataka. Beta raspodela $f(B)$ vrši restrikciju beta vrednosti na jedinični interval $(0,1)$. I unutar ovog intervala raspodela može da poprими mnoge poznate oblike, kao što su oblici normalne, hi-kvadratne i F raspodele. Primeri beta raspodele za različite parametre su prikazane na Slici 2.9.



Slika 2.8 F Raspodela

U izrazu (2.15) smo koristili beta raspodelu da opišemo T^2 statistiku. Koristeći vezu, koja postoji između F i beta funkcije verovatnoće, T^2 se može izraziti i kao F statistika na sledeći način

$$B_{(p/2, (n-p-1)/2)} = \frac{pF}{(n-p-1)pF}, \quad F \sim F_{(p, n-p-1)}. \quad (2.16)$$



Slika 2.9 Beta raspodela

Kontrola procesa u određenim situacijama je bazirana na praćenju sredine uzorka (tj. podgrupe) od m posmatranja uzeta na svakih k uzoračkih intervala. Raspodela koja opisuje statističku distancu između sredine uzorka i-tog posmatranja vektora \bar{X}_i i srednje vrednosti istorijskih podataka \bar{X} je data sa

$$T^2 = (\bar{X}_i - \bar{X})' S^{-1} (\bar{X}_i - \bar{X}) \sim \frac{(m+n)(n-1)p}{mn(n-p)} F_{(p,n-p)}, \quad (2.17)$$

gde $F_{(p,n-p)}$ predstavlja F raspodelu sa p i $(n-p)$ stepeni slobode. Ova raspodela koja prepostavlja nezavisnost između \bar{X}_i i ocenjivača parametara, zavisi od veličine uzorka m , veličine istorijskih podataka, i broja promenljivih u posmatranom vektoru \bar{X}_i . Iako ocenjivač za S , dat u (2.11), se koristi u statistici (2.17), češće se ubacuju objedinjena ocena za S data sa

$$S_W = \frac{\sum_{i=1}^k S_i}{k}, \quad (2.18)$$

gde S_i predstavlja ocenu uzoračke koarijanse za podatke uzete tokom i-tog perioda, $i=1,2,\dots,k$. Sa ovom ocenom, forma i raspodela T^2 statistike postaje

$$T^2 = (\bar{X}_i - \bar{X})' S_W^{-1} (\bar{X}_i - \bar{X}) \sim \frac{(k+1)(m-1)p}{m(mk-k-p+1)} F_{(p,mk-k-p+1)}, \quad (2.19)$$

GLAVA 3

Interpretacija T^2 signala za dve promenljive

Univarijantna kontrolna procedura uglavnom se odnosi na nadgledanje kontrolnih grafika, na kojima bi se uočile nepravilne varijacije. Na primer, možemo da pratimo promenu srednje vrednosti pomoću \bar{X} grafika ili promene u varijacijama pomoću R grafika. Ako se pojavi signal, to znači da je proces poremećen i/ili da se procesno variranje promenilo. U ovim primerima, interpretacija signala je pojednostavljena, jer se samo jedna promenljiva treba ispitati.

Kada je reč o multivarijantnoj statističkoj kontroli procesa, situacija se komplikuje. Neslaganje sa istorijskim podacima može se pratiti pomoću T^2 statistike. Ako posmatrana T^2 vrednost pada van kontrolnih granica signal je detektovan. Jednostavnost šeme nadgledanja prestaje sa detekcijom signala, jer raznovrsnost među vezama između promenljivih može da proizvede signal. Na primer, uzorak se može identifikovati kao van kontrole zato što se individualna promenljiva nalazi van kontrolnih granica procesa, ustanovljenih istorijskim podacima. Drugi slučaj kada se javlja signal je kada se vrednost dve ili više promenljivih ne pridržavaju linearne korelacije uspostavljene istorijskim podacima. Najgori slučaj je kada je signal uzrokovani kombinacijom ova dva, kada su neke promenljive van kontrole i kada su neke kontrakorelisane.

U ovom poglavlju, predstavljena je metoda za interpretaciju signala bazirana na ortogonalnoj dekompoziciji T^2 statistike. Nezavisne komponente dekompozicije, svaka slična individualnoj T^2 promenljivoj, koriste se da bi se izolovao izvor signala i pojednostavila njegova interpretacija. Diskusija je ograničena na dve promenljive, jer je najlakše goemetrijski zamisliti. Opštiji slučaj sa p promenljivih je predstavljen u glavi 4.

3.1 Ortogonalna dekompozicija

Tipična T^2 statistika za posmatrani vektor $X' = (x_1, x_2, \dots, x_p)$ je data sa

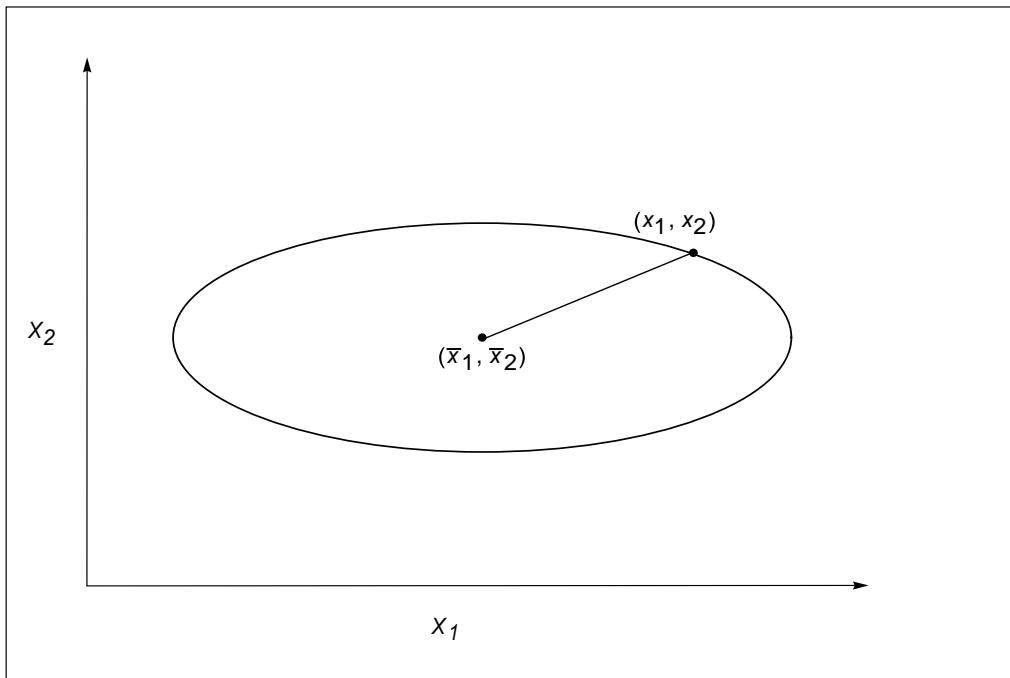
$$T^2 = (X - \bar{X})' S^{-1} (X - \bar{X}). \quad (3.1)$$

Cilj je da se pronađe metodologija pomoću koje će se razložiti T^2 vrednost na ortogonalne komponente, koje se mogu povezati sa svim p procesnih promenljivih. Ortogonalna dekompozicija je standardna u mnogim statističkim procedurama kao što su analiza varianse ili regresiona analiza. Na primer, u analizi varianse, ukupna suma kvadratnog odstupanja odgovarajuće promenljive je razdvojena na dve nezavisne komponente. Jedna komponenta je suma kvadrata koje je objasnila regresija, a druga komponenta je suma kvadrata reziduala. Sličan prilaz dekompoziciji može se primeniti na T^2 statistiku.

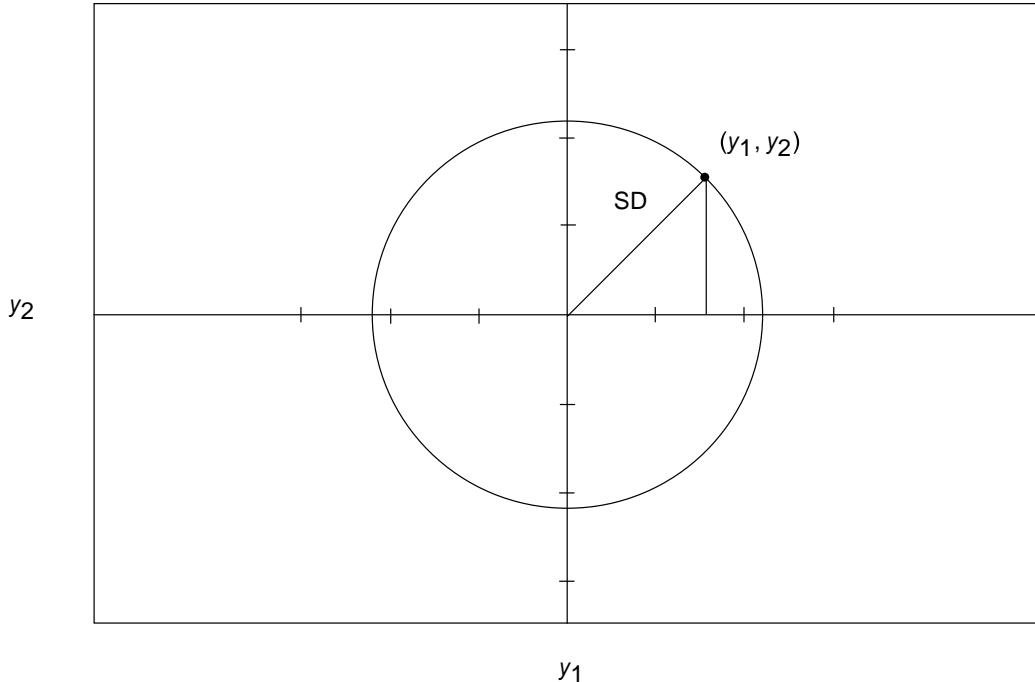
Zamislimo slučaj sa dve promenljive gde je posmatrani vektor dat sa $X' = (x_1, x_2)$ i on ima bivarijantnu normalnu raspodelu sa nepoznatim srednjim vektorom i matricom kovarijansi. Pretpostavimo da su dve promenljive nezavisne pa je u uzorku veličine n njihova međusobna korelacija jednala nula. Takođe pretpostavimo da su dve individualne uzoračke varijanse, s_1^2 i s_2^2 , nejednake i označimo dve odgovarajuće uzoračke sredine sa \bar{x}_1 i \bar{x}_2 . T^2 eliptički kontrolni region se formira pomoću sledeće formule

$$T^2 = \frac{(x_1 - \bar{x}_1)^2}{s_1^2} + \frac{(x_2 - \bar{x}_2)^2}{s_2^2} = c^2, \quad (3.2)$$

gde je c odgovarajući odabrana konstanta koja precizira veličinu kontrolnog regiona. Tipični kontrolni region predstavlja unutrašnjost elipse date na Slici 3.1. Svaka uzoračka tačka locirana na elipsi imaće istu statističku distancu od uzoračke sredine kao i bilo koja druga tačka na elipsi.



Slika 3.1 Bivarijantna nezavisna kontrolna oblast sa različitim varijansama



Slika 3.2 Bivarijantna ortogonalna kontrolna oblast sa jednakim varijansama. SD je statistička distanca

Komponente x_1 i x_2 su nezavisne i to prouzrokuje da elipsa ne bude iskošena. Kako su komponente "težinski" nejednake zbog njihovih nejednakih varijansi, transformisaćemo promenljive u formu koja će im dati istu težinu. Tako dobijamo kružni region i statističku distancu predstavljenu kvadratnim korenom T^2 iz (3.2), ekvivalentnu odgovarajućoj Euklidovoj, pravolinijskoj distanci. Ovo je oblik koji je potreban da bi se interpretirala T^2 vrednost.

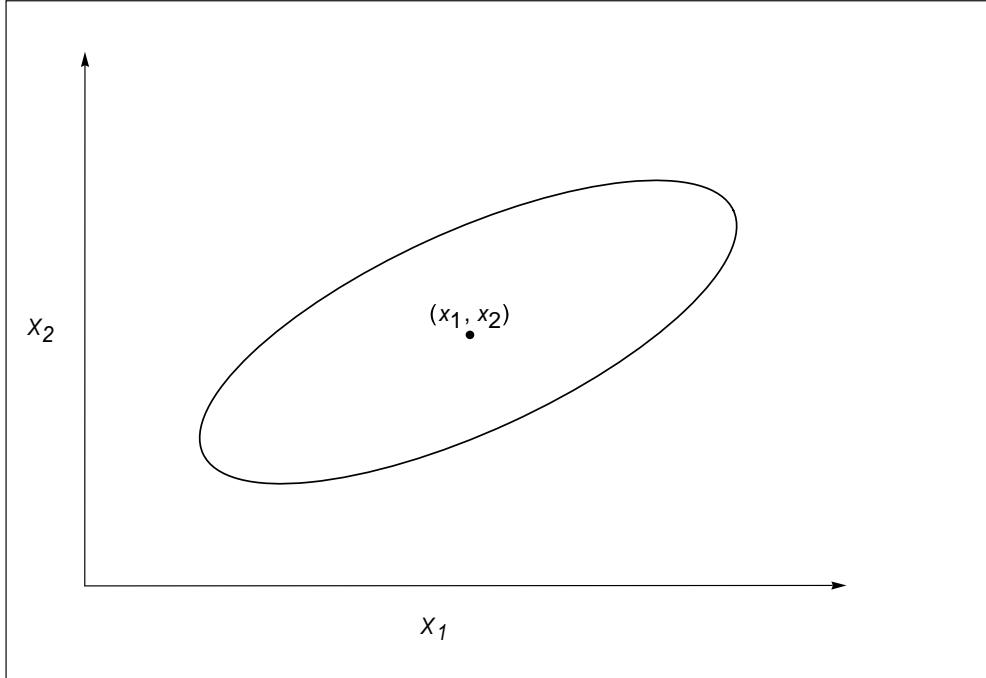
Uvode se smene

$$y_1 = \frac{x_1 - \bar{x}_1}{s_1}, \quad y_2 = \frac{x_2 - \bar{x}_2}{s_2}, \quad (3.3)$$

gde y_1, y_2 predstavljaju predstavlja standardizovane vrednosti za x_1 i x_2 respektivno. Pomoću ovih transformacija možemo da napišemo izraz za T^2 iz (3.1) kao

$$T^2 = y_1^2 + y_2^2 = c^2. \quad (3.4)$$

T^2 vrednost je razdvojena na dve nezavisne komponente, koje sada imaju istu težinu. Prva komponenta y_1^2 meri doprinos x_1 na ukupnu vrednost T^2 , a druga komponenta y_2^2 meri doprinos x_2 za ukupnu vrednost T^2 . Pažljivo proučavanje ovih komponenti će izložiti uzrok signala.



Slika 3.3 Bivarijantna zavisna kontrolna oblast sa različitim varijansama

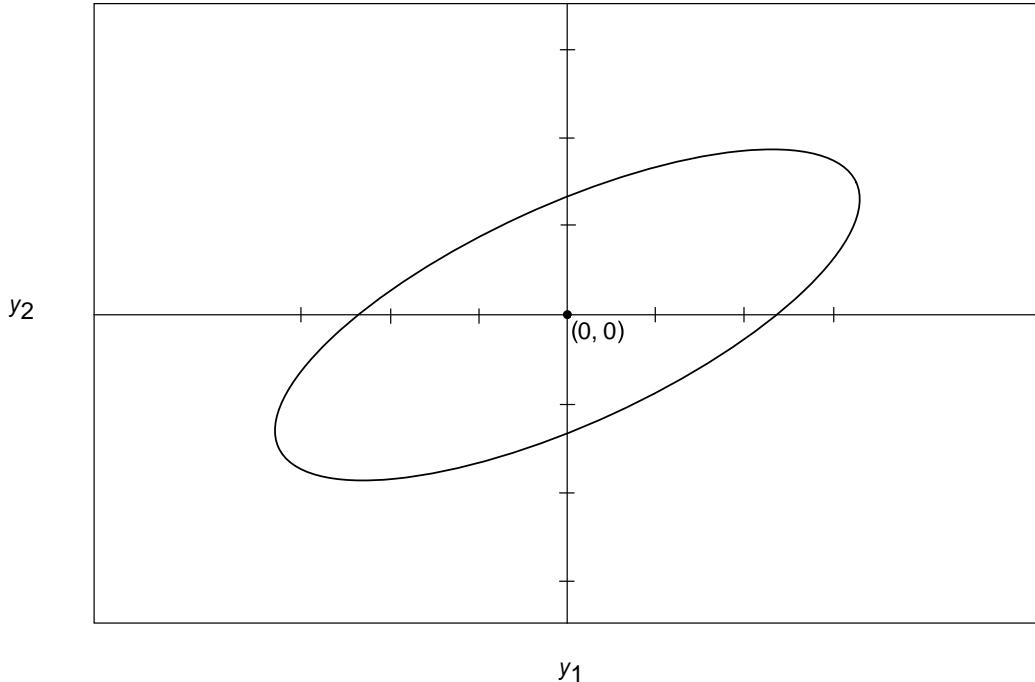
Kontrolni region dobijen pomoću transformacije date u (3.3) je predstavljen unutrašnjošću kruga datog na Slici 3.2. T^2 vrednost u ovom ortogonalno transformisanom prostoru je ista kao kvadratna Euklidova distanca koju tačka (y_1, y_2) ima od koordinatnog početka $(0,0)$, i može se predstaviti sa kvadratom hipotenuze trougla na Slici 3.2. Ovo je jednako statističkoj distanci tačke (x_1, x_2) od srednjeg vektora (\bar{x}_1, \bar{x}_2) . Zbog toga sve tačke sa istom statističkom distancicom su locirana na krugu sa Slike 3.2, kao i na elipsi sa Slike 3.1.

Zamislimo situaciju gde promenljive $X' = (x_1, x_2)$ nisu nezavisne. Kako je međusobna korelacija r između x_1 i x_2 nenula, T^2 vrednost će biti data sa

$$T^2 = \left(\frac{1}{1 - r^2} \right) \left[\frac{(x_1 - \bar{x}_1)^2}{s_1^2} - 2r \left(\frac{x_1 - \bar{x}_1}{s_1} \right) \left(\frac{x_2 - \bar{x}_2}{s_2} \right) + \frac{(x_2 - \bar{x}_2)^2}{s_2^2} \right], \quad (3.5)$$

i odgovarajući kontrolni region će biti nakriven. Ovo je prikazano na Slici 3.3. Opet, ako y_1 i y_2 predstavljaju standardizovane vrednosti za x_1 i x_2 , T^2 vrednost u (3.5) može se zapisati kao

$$T^2 = \left(\frac{1}{1 - r^2} \right) (y_1^2 - 2ry_1y_2 + y_2^2). \quad (3.6)$$



Slika 3.4 Bivarijantna kontrolna oblast

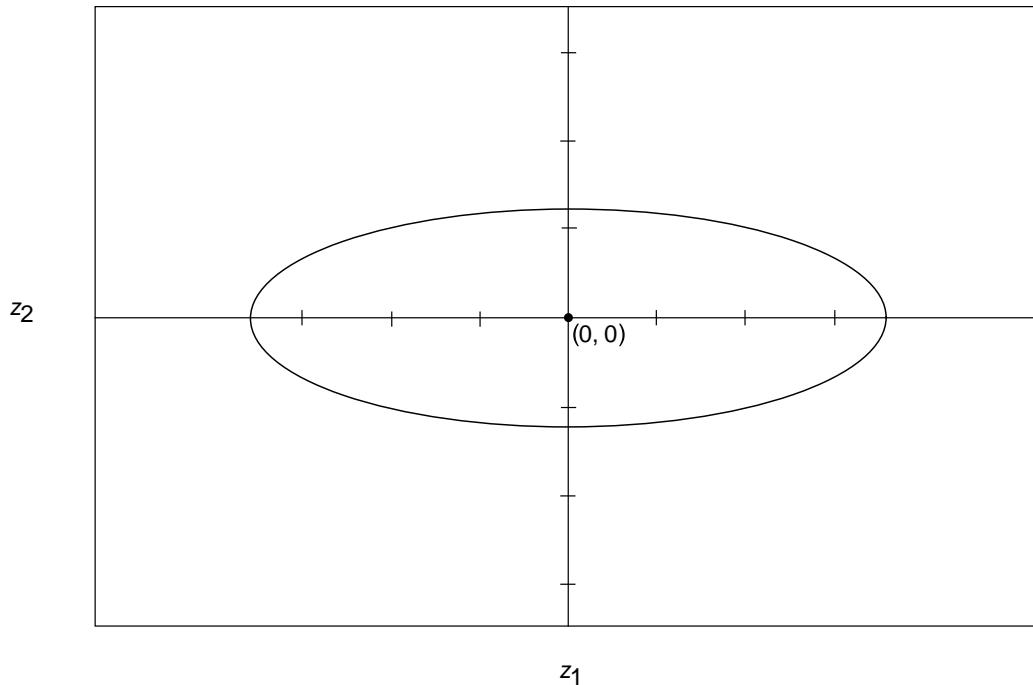
Kako su x_1 i x_2 međusobno korelisani nije moguće razdvojiti doprinos y_1 i y_2 za ukupnu T^2 vrednost, jer se transformacijama standardizovanog prostora ne može ukloniti međusobni proizvod, tj. transformacija nije ortogonalna. Zbog toga je kontrolni region eliptički i nakriven, kao što je prikazano na Slici 3.4. Da bi izrazili statističku distancu iz (3.6) kao Euklidovu distancu, koju možemo da je vizualizujemo, prvo moramo da transformišemo ose ortogonalnog (x_1, x_2) prostora u ose elipse. Ovo nije urađeno ni na Slici 3.4 niti na Slici 3.5. Na primer, ose elipse na Slici 3.4 ne odgovaraju osama (y_1, y_2) prostora. Ose se jedino mogu poravnati kroz ortogonalne transformacije. U ovom slučaju, transformacija u (3.6) je nepotpuna, jer ne pruža rotaciju osa koja je potrebna.

T^2 statistika u (3.6) može se razdvojiti na dve aditivne komponente pomoću sledeće ortogonalne transformacije

$$z_1 = \frac{(y_1 + y_2)}{\sqrt{2}}, \quad z_2 = \frac{(y_1 - y_2)}{\sqrt{2}}. \quad (3.7)$$

Koristeći ove transformacije T^2 vrednost iz (3.6) se može razložiti na

$$T^2 = \frac{z_1^2}{1+r} + \frac{z_2^2}{1-r}. \quad (3.8)$$



Slika 3.5 Bivarijantna kontrolna oblast sa nejednakim varijansama

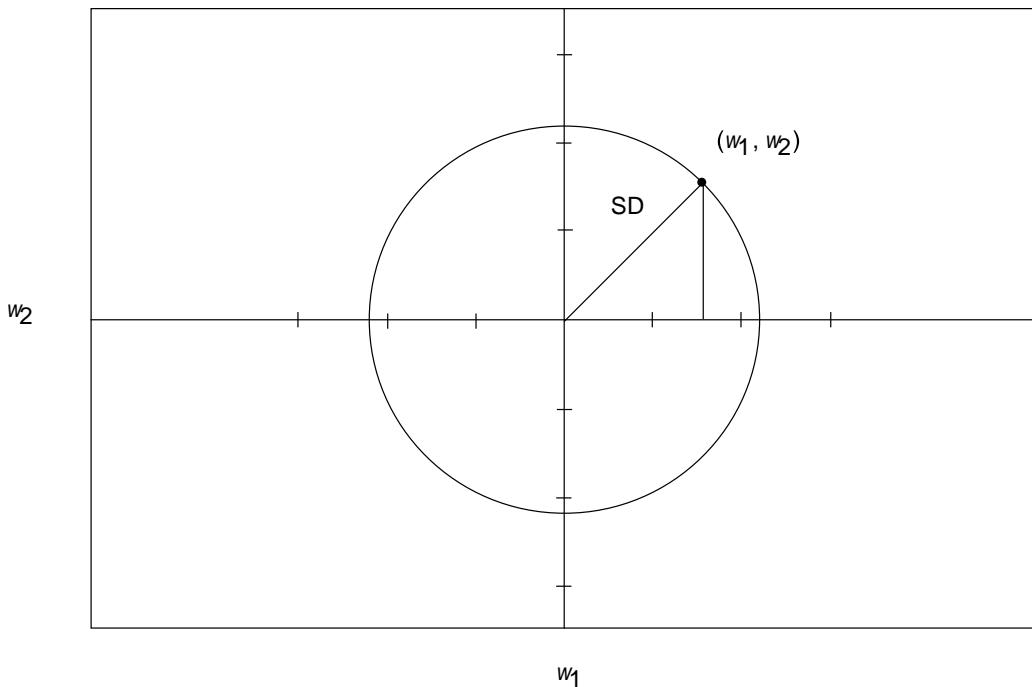
Ipak, grafik kontrolnog regiona za izraz u (3.8) i dalje ne dozvoljava lako računanje statističke distance jer su komponente “težinski” nejednake. Ovo je predstavljeno na Slici 3.5 sa različitim veličinama osa elipse. Prva komponenta ima težinu jednaku recipročnoj vrednosti $\lambda_1 = (1 + r)$, a druga komponenta ima težinu jednaku $\lambda_2 = (1 - r)$. Težine su jednake samo kad je $r = 0$, tj. kad su početne promenljive nezavisne.

Ovaj gore problem može se zaobići transformacijom z vrednosti u (3.7) na novi prostor gde promenljive imaju istu težinu. Jedna takva transformacija je

$$\omega_1 = \frac{z_1}{\sqrt{1+r}}, \quad \omega_2 = \frac{z_2}{\sqrt{1-r}}. \quad (3.9)$$

Koristeći ortogonalnu transformaciju, možemo izraziti T^2 vrednost iz (3.1) kao

$$T^2 = \omega_1^2 + \omega_2^2 = c^2. \quad (3.10)$$



Slika 3.6 Bivarijantna kontrolna oblast sa jednakim varijansama

Rezultujući kontrolni region dat na Slici 3.6 je sada kružni i (kvadratna) statistička distanca je predstavljena hipotenuzom datog trougla. Transformacija data u (3.10) pruža dekompoziciju T^2 vrednosti. Zbog toga će uspešno razdvojiti bivarijantnu T^2 vrednost na dve aditivne i ortogonalne komponente. Ipak, svaka ω_1 i ω_2 komponenta u 3.10 je linearna kombinacija x_1 i x_2 . Kako se svaka komponenta sastoji od obe promenljive, ovo sputava jasnu interpretaciju do izvora signala u obliku individualne procesne promenljive. Ovaj problem postaje ozbiljniji kako broj promenljivih raste. Ono što je potrebno je metodologija koja će pružiti i ortogonalnu dekompoziciju i sredstva za interpretaciju individualnih komponenti. Jedna takva procedura je MYT (Mason-Young-Tracy) dekompozicija, koju su prvu uveli Mason, Young, Tracy (1995).

3.2 MYT (Mejson-Jang-Trejsi) dekompozicija

T^2 statistika je suma kvadrata pa postoji beskonačno mnogo načina da se razdvoji na p nezavisnih komponenti. Međutim, u ovom radu se traži dekompozicija, koja će biti laka za interpretaciju. Problem sa dekompozicijom, rađenom u prethodnom poglavlju, je što može biti teška za interpretaciju jer su njene komponente linearna kombinacija p promenljivih iz posmatranog vektora.

U ovom poglavlju će biti predstavljena MYT procedura za bivarijantni vektor $X' = (x_1, x_2)$, gde su x_1 i x_2 korelsani. Detalji o opštijem slučaju sa p promenljivih biće predstavljeni u glavi 4.

MYT dekompozicija koristi ortogonalnu transformaciju da izrazi T^2 vrednost kao dva ortogonalna i jednakaka (težinski) izraza. Jedna takva dekompozicija je data sa

$$T^2 = T_1^2 + T_{2.1}^2, \quad (3.11)$$

gde su

$$T_1^2 = \frac{(x_1 - \bar{x}_1)^2}{s_1^2}, \quad (3.12)$$

$$T_{2.1}^2 = \frac{(x_2 - \bar{x}_{2.1})^2}{s_{2.1}^2}. \quad (3.13)$$

U ovoj formulaciji, $\bar{x}_{2.1}$ je ocenjivač za uslovnu sredinu promenljive x_2 za datu vrednost x_1 , i $s_{2.1}^2$ je odgovarajući ocenjivač uslovne varijanse promenljive x_2 za datu vrednost x_1 . Kasnije u radu će biti više detalja o ovim ocenjivačima.

Prvi izraz u MYT dekompoziciji iz (3.11) T_1^2 se naziva bezuslovni izraz, jer zavisi samo od x_1 . Dok se drugi izraz ortogonalne dekompozicije $T_{2.1}^2$ zove uslovni izraz, jer je uslovljen sa vrednosti promenljive x_1 . Koristeći ovu notaciju, izraz (3.11) se može zapisati kao

$$T^2 = \frac{(x_1 - \bar{x}_1)^2}{s_1^2} + \frac{(x_2 - \bar{x}_{2.1})^2}{s_{2.1}^2}. \quad (3.14)$$

Kvadratni koren ove T^2 vrednosti je prikazan grafički u T_1 i $T_{2.1}$ prostoru na Slici 3.7. Ortogonalna dekompozicija data u (3.11) je jedna od dve moguće MYT dekompozicije T^2 vrednosti za $p = 2$. Druga moguća dekompozicija je

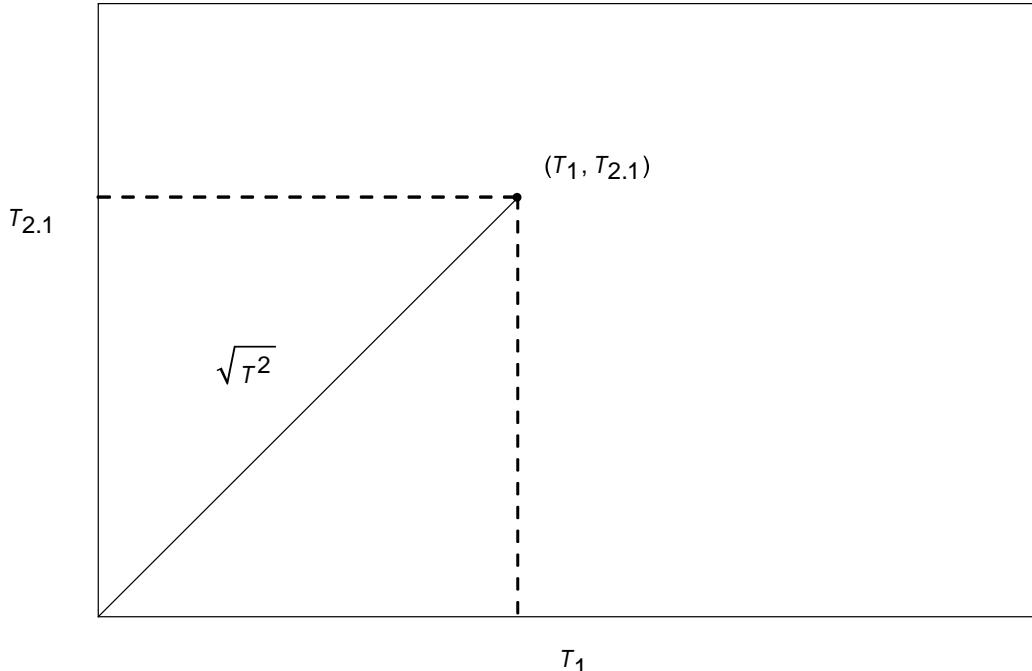
$$T^2 = T_2^2 + T_{1.2}^2, \quad (3.15)$$

gde su

$$T_2^2 = \frac{(x_2 - \bar{x}_2)^2}{s_2^2}, \quad (3.16)$$

$$T_{1.2}^2 = \frac{(x_1 - \bar{x}_{1.2})^2}{s_{1.2}^2}. \quad (3.17)$$

Dekompozicija T^2 vrednosti data u (3.15) se razlikuje od one iz (3.11). Prvi izraz dekompozicije iz (3.15) je bezuslovni izraz za promenljivu x_2 , dok je prvi izraz u (3.11) bezuslovni izraz za promenljivu x_1 . Slično, uslovni izraz iz (3.15) zavisi od uslovne gustine promenljive x_1 za datu vrednost promenljive x_2 , dok uslovni izraz iz (3.11) zavisi od uslovne gustine promenljive x_2 za datu x_1 . Ova dva uslovna izraza nisu isti, osim u slučaju kada su x_1 i x_2 nekorelisani.



Slika 3.7 T^2 grafik za MYT bivarijantnu dekompoziciju

3.3 Interpretacija signala preko T^2 komponenti

Za bivarijantni vektor $X' = (x_1, x_2)$ dve moguće MYT dekompozicije su date sa izrazima (3.11) i (3.15). Uzimajući u obzir obe dekompozicije, imamo četiri jedinstvena izraza : dva bezuslovna T_1^2 i T_2^2 i dva uslovna izraza obeleženih kao $T_{1.2}^2$ i $T_{2.1}^2$. Signalna T^2 vrednost može da proizvede velike vrednosti za bilo koju kombinaciju ovih elemenata. Zamislimo veliku vrednost za bezuslovni izraz

$$T_1^2 = \frac{(x_1 - \bar{x}_1)^2}{s_1^2},$$

ili za drugi bezuslovni izraz

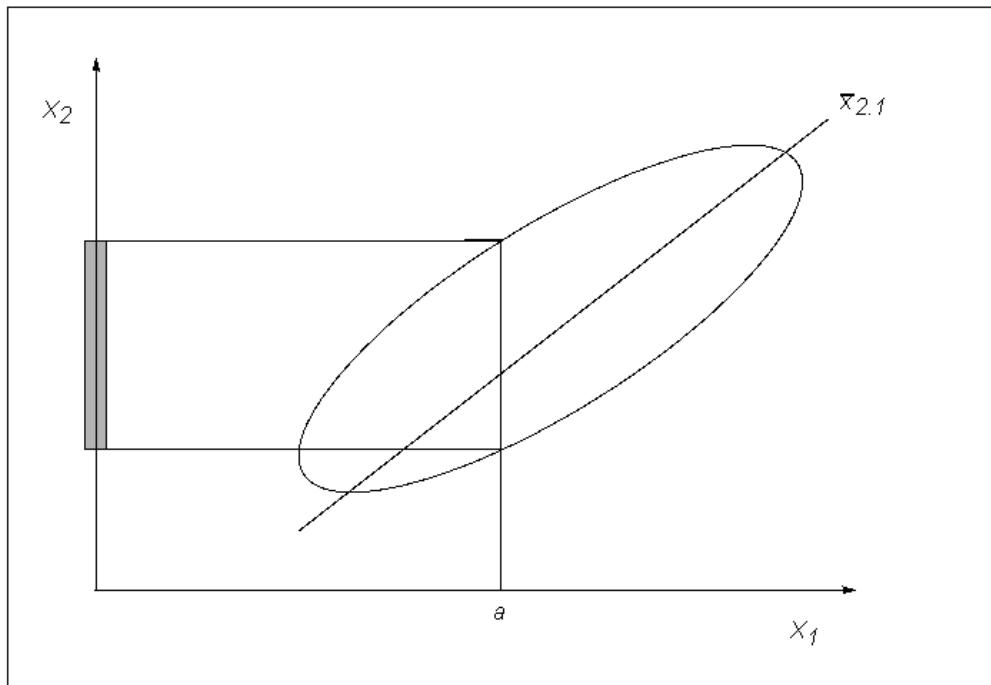
$$T_2^2 = \frac{(x_2 - \bar{x}_2)^2}{s_2^2}.$$

Bezuslovni izraz je kvadrat univarijantne t statistike i meri statističku distancu posmatrane vrednosti x_i od njene srednje vrednosti \bar{x}_i . Ako je posmatrana vrednost promenljive van granica tolerancije (tj. van operacionalnog ranga baziranog na istorijskim podacima), signal se javlja na bezulovnom izrazu.

Stvar postaje komplikovanija kada se radi interpretacija velikih vrednosti na uslovnim izrazima. Uslovni izraz

$$T_{2.1}^2 = \frac{(x_2 - \bar{x}_{2.1})^2}{s_{2.1}^2},$$

koji je mera kvadratne statističke distance posmatrane vrednosti x_2 od uslovne sredine $\bar{x}_{2.1}$. Ova distanca može se videti grafički, zamislimo prezentaciju kontrolnog regiona u prostoru promenljivih (x_1, x_2) predstavljenog na Slici 3.8. Kako se vrednost x_1 menja, tako se i vrednost $\bar{x}_{2.1}$ menja. Zamislimo fiksiranu vrednost x_1 , neka je recimo $x_1 = a$, što je predstavljeno vertikalnom linijom na x_1 osu u tački a . Da bi se kontrola procesa održala za ovu vrednost x_1 , potrebno je da odgovarajuća vrednost x_2 bude u osenčenom delu x_2 ose. Ovo znači da vrednost x_2 mora da bude sadržana u ovom delu uslovne gustine, ili će se signal pojaviti na $T_{2.1}^2$ komponenti.

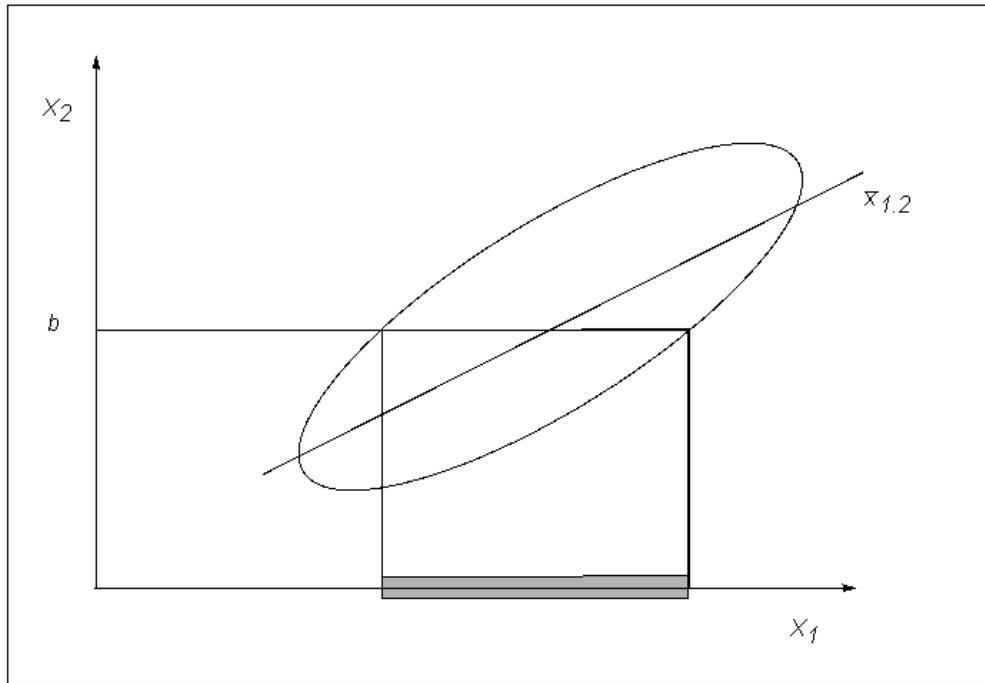


Slika 3.8 Interpretacija $T_{2.1}^2$ komponente

Analogno tome, ako posmatramo drugi uslovni izraz

$$T_{1.2}^2 = \frac{(x_1 - \bar{x}_{1.2})^2}{s_{1.2}^2},$$

na sličan način ilustrujemo signal. Neka je vrednost x_2 fiksirana u tački b , posmatra se organičen interval na x_1 osi, u koji mora da upadne posmatrana vrednost za promenljivu x_1 . Ovo je predstavljeno na Slici 3.9. Ako vrednost x_1 nije sadržana u osenčenom intervalu, signal će se pojaviti na $T_{1.2}^2$ komponenti.



Slika 3.9 Interpretacija $T_{1,2}^2$ komponente

Velika vrednost na uslovnom izrazu implicira da posmatrana vrednost jedne promenljive nije tamo gde bi trebalo da budu u odnosu na posmatrane vrednosti drugih promenljivih. Za vektore (x_1, x_2) iz uzorka koja produkuju signal ovog tipa se kaže da su kontrakorelisani, kao da se poremetio odnos između promenljivih x_1 i x_2 . Kontrakorelacija je česti uzrok multivarijantnog signala.

3.4 Regresija

Kada se signal javlja na uslovnom izrazu to znači da nešto nije u redu sa linearnim odnosima između promenljivih. Proučavanjem uslovnih izraza sa stanovišta regresije dobija se dodatno razumevanje za interpretaciju signala. Na Slici 3.8 ucrtana je regresiona linija $\bar{x}_{2,1}$, koja predstavlja predviđenu vrednost promenljive x_2 na osnovu vrednosti promenljive x_1 . Isto tako regresiona linija $\bar{x}_{1,2}$ data na Slici 3.9 predstavlja predviđenu vrednost promenljive x_1 na osnovu promenljive x_2 . U opštem slučaju $T_{i,j}^2$ je standardizovano posmatranje i-te promenljive korigovano sa ocenama srednje vrednosti i varijanse iz uslovne raspodele promenljive $x_{i,j}$. Za bivarijantni slučaj, opšta forma za uslovni izraz je data sa

$$T_{i,j}^2 = \frac{(x_i - \bar{x}_{i,j})^2}{s_{i,j}^2}. \quad (3.18)$$

Ocenjena sredina promenljive x_i korigovana za x_j , tj. $\bar{x}_{i,j}$ data je sa jednačinom

$$\bar{x}_{i,j} = \bar{x}_i + b(x_j - \bar{x}_j), \quad (3.19)$$

gde su \bar{x}_i i \bar{x}_j uzoračke sredine promenljivih x_i i x_j dobijene iz istorijskih podataka, i b je ocjenjeni regresioni koeficijent koji se odnosi na x_i i x_j u ovim podacima. Leva strana jednačine (3.19) sadrži $\bar{x}_{i,j}$, što je predviđena vrednost za x_i bazirana na odgovarajućoj vrednosti x_j , tj. x_i je zavisna promenljiva i x_j je nezavisna promenljiva. Na osnovu toga brojilac iz jednačine (3.18) je regresioni rezidual, tj.

$$r_{i,j} = (x_i - \bar{x}_{i,j}).$$

Kada se uslovna varijansa napiše kao

$$s_{i,j}^2 = s_i^2(1 - R_{i,j}^2),$$

gde je $R_{i,j}$ višestruka korelacija između x_i i x_j , i zamenujući $r_{i,j}$ za $x_i - \bar{x}_{i,j}$, možemo izraziti $T_{i,j}^2$ kao

$$T_{i,j}^2 = \frac{(r_{i,j})^2}{s_i^2(1 - R_{i,j}^2)}, \quad (3.20)$$

$$T_{i,j}^2 = \left(\frac{r_{i,j}}{s_i}\right)^2 / (1 - R_{i,j}^2).$$

Koristi se ova notacija zbog konzistentnosti sa formulom koja će biti korištena u p-dimenzionalnom slučaju u glavi 4. Uslovni izraz (3.20) objašnjava koliko dobro se buduće posmatranje održene promenljive slaže sa predviđenom vrednosti za tu promenljivu na osnovu druge promenljive. Kada je imenilac iz (3.20) jako mali, što se dešava kada je korelacija blizu jedan, očekivaćemo vrlo blisko slaganje između posmatrane i predviđene vrednosti x_i . Inače, veličina uslovnog T^2 izraza će zavisiti od brojioca, koji je funkcija slaganja između posmatrane i predviđene vrednosti x_i . Značajno odstupanje između ove dve vrednosti će prouzrokovati velike vrednosti za T^2 izraz.

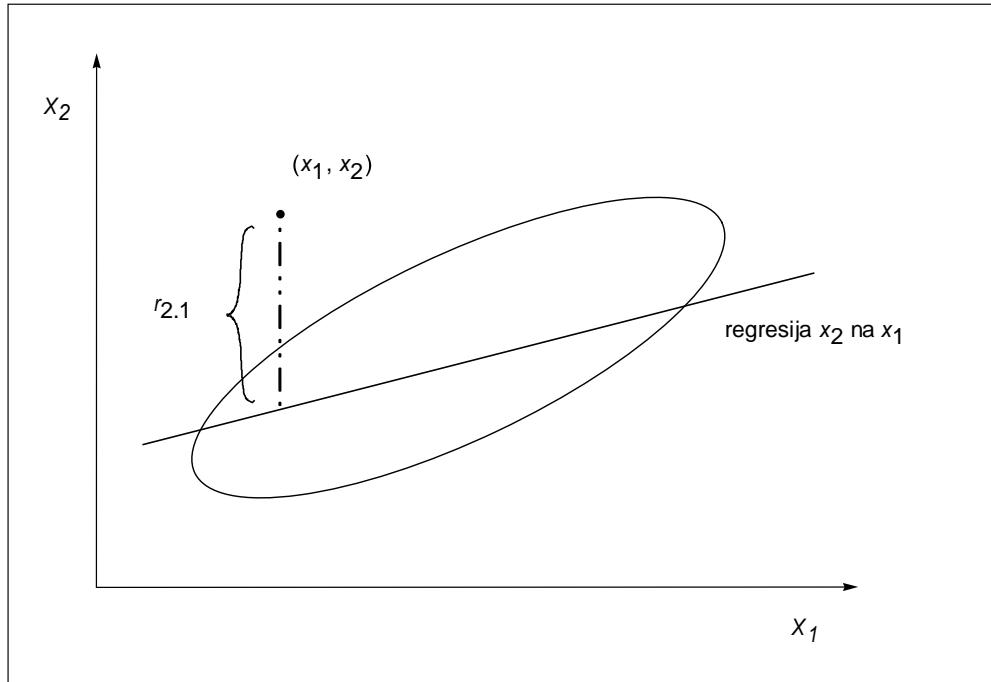
Neka postoji vektor (x_1, x_2) , koji se nalazi unutar Ševartove "kutije" formirane pomoću dve promenljive. Pretpostavimo da postoji značajno ne slaganje u veličinama između posmatrane vrednosti x_1 i odgovarajuće predviđene vrednosti dobijene pomoću vrednosti x_2 i izraza (3.19). Ovo znači da posmatranje na ovoj određenoj komponenti je ispod ili iznad onoga što je predviđeno istorijskim podacima. Zbog boljeg razumevanja izraza (3.20), zamislimo dva uslovna T^2 izraza čiji su kvadratni koreni dati sa

$$T_{2,1} = \frac{r_{2,1}}{s_2 \sqrt{(1 - R_{2,1}^2)}}, \quad (3.21)$$

i

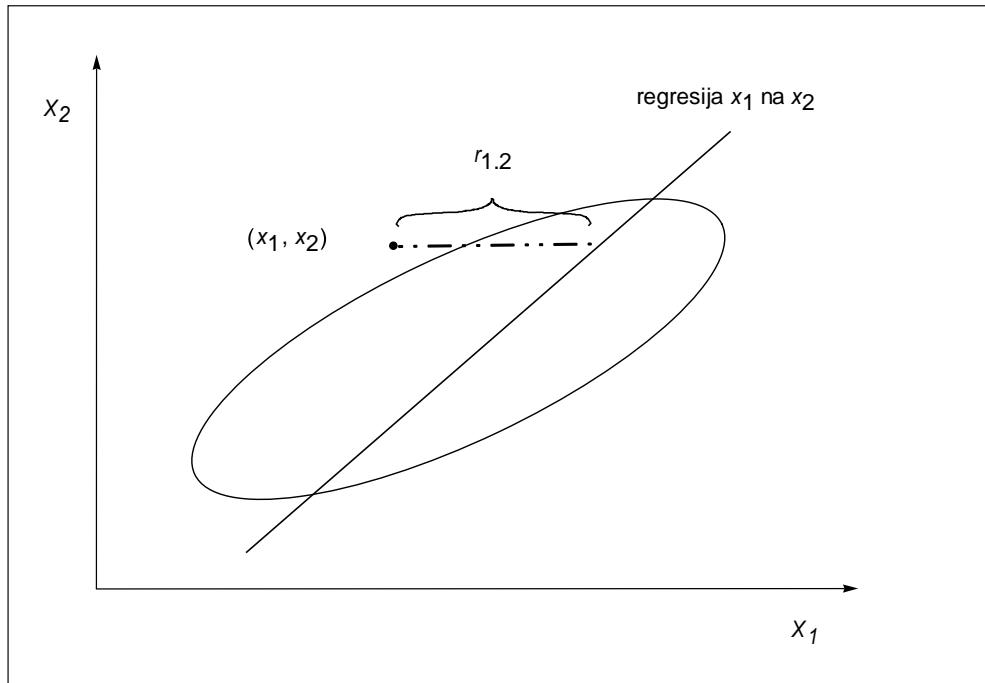
$$T_{1,2} = \frac{r_{1,2}}{s_2 \sqrt{(1 - R_{1,2}^2)}}, \quad (3.22)$$

gde su $r_{2.1} = (x_2 - \bar{x}_{2.1})$ i $r_{1.2} = (x_1 - \bar{x}_{1.2})$ reziduali sa rezidualnih prava $\bar{x}_{1.2}$ i $\bar{x}_{2.1}$.



Slika 3.10 Rezidual za regresiju x_2 u odnosu na x_1

Ovi reziduali su grafički predstavljeni na Slici 3.10 i Slici 3.11. Treba primetiti da dve uslovne vrednosti u (3.21) i (3.22), ako se izuzme $R_{i,j}^2$ izraz, su zapravo standardizovani reziduali koji imaju formu $r_{i,j}/s_i$. Kada su reziduali, dati na Slikama 3.10 i 3.11, posle standardizovanja veliki, javlja se signal na uslovnom T^2 izrazu. Ovo će se desiti samo kada se posmatrana vrednost promenljive x_1 razlikuje od one predviđene sa x_2 , ili kada se posmatrana vrednost promenljive x_2 razlikuje od one predviđene sa x_1 , gde see predviđanje radi na osnovu istorijskih podataka.



Slika 3.11 Rezidual za regresiju \$x_1\$ u odnosu na \$x_2\$

3.5 Raspodela verovatnoća \$T^2\$ komponenti

Bolji uvid i rešavanje nekih problema pri interpretaciju \$T^2\$ komponenti MYT dekompozicije mogu se dobiti određivanjem raspodele verovatnoća za svaki pojedinačni izraz dekompozicije. Pod pretpostavkom da nema signala, svi izrazi MYT dekompozicije, uslovni i bezuslovni, su opisani \$F\$ raspodelom. Bezuslovni izrazi, koji se koriste za determinisanje promenljivih individualno, imaju raspodelu

$$T_j^2 \sim \left(\frac{n+1}{n} \right) F_{(1,n-1)}, \quad (3.23)$$

za \$j = 1, 2\$. Slično, uslovni izrazi \$T_{i,j}^2\$, koji se koriste za proveru linearnih odnosa između promenljivih, imaju raspodelu datu sa

$$T_{i,j}^2 \sim \left(\frac{(n+1)(n-1)}{n(n-k-1)} \right) F_{(1,n-k-1)}, \quad (3.24)$$

gde je \$k\$ jednako broju uslovnih promenljivih. Kada je \$k = 0\$, raspodela iz (3.24) se svodi na raspodelu (3.23). Zbog toga se \$F\$ raspodela može koristiti za određivanje kada je neki, uslovni ili bezuslovni, izraz dekompozicije značajno velik i daje doprinos signalu. Procedura za pravljenje ove determinacije je sledeća. Za određeni nivo \$\alpha\$ i istorijske podatke veličine \$n\$, dobija se \$F_{(\alpha,1,n-k-1)}\$ iz odgovarajuće \$F\$ tabele. Pa se računa donja granica za individualni bezuslovni izraz kao

$$UCL = \left(\frac{(n+1)(n-1)}{n(n-k-1)} \right) F_{(\alpha, 1, n-k-1)}. \quad (3.25)$$

Nakon izračunavanja donje granice UCL svaki bezuslovni izraz se pojedinačno upoređuje sa njom i svaki izraz, koji zadovoljava

$$T_j^2 > \left(\frac{n+1}{n} \right) F_{(\alpha, 1, n-1)},$$

implicira da odgovarajući x_j doprinosi signalu. Isto tako, svaki uslovni izraz veći od njegove donje granice UCL , koja je data sa

$$T_{j,i}^2 > \left(\frac{(n+1)(n-1)}{n(n-k-1)} \right) F_{(\alpha, 1, n-k-1)}, \quad (3.26)$$

implicira da x_i i x_j doprinose signalu.

Pomenuta procedura za lociranje izraza dekompozicije, koji doprinose signalu posmatranog vektora, nije egzaktna. Zamislimo oblast za održavanje kontrole za uslovne i bezuslovne izraze kada je $p = 2$. MYT dekompozicija za T^2 vrednost može se predstaviti sa

$$T^2 = T_1^2 + T_{2,1}^2.$$

Dakle, oblast prihvatanja za T_1^2 je data sa

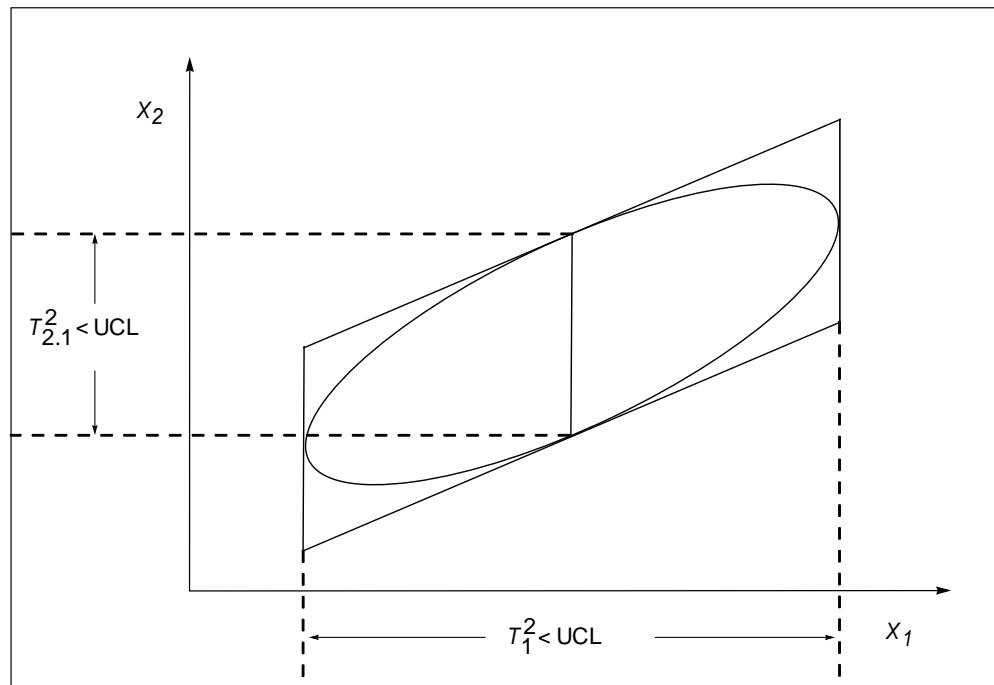
$$T_1^2 < \left(\frac{n+1}{n} \right) F_{(\alpha, 1, n-1)}.$$

Isto tako, oblast prihvatanja za $T_{2,1}^2$ je data sa

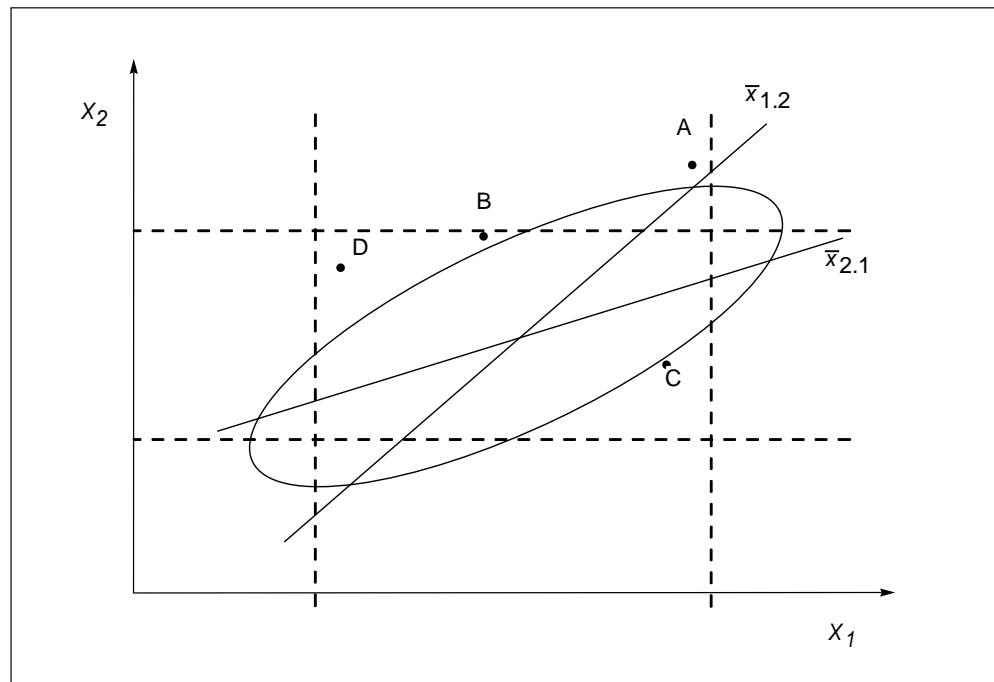
$$T_{2,1}^2 < \left(\frac{(n+1)(n-1)}{n(n-k-1)} \right) F_{(\alpha, 1, n-k-1)}.$$

Odnos između ovih oblasti, definisanih pomoću raspodele individualnih izraza i eliptičkih kontrolnih oblasti, je predstavljen na Slici 3.12. Ovo znači da kada se koristi raspodela individualnih izraza T^2 dekompozicije za detektovanje uzroka signala, zapravo se aproksimira eliptička kontrolna oblast sa paralelogramom. Ono što bi u ovom slučaju odgovaralo je da paralelogram i elipsa sa Slike 3.12 budu slične veličine. Veličina elipse je determinisana izborom opšte verovatnoće, označene sa α_1 , pravljenja greške tipa 1 ili tvrđenja da je proces van kontrole, kada to nije slučaj. Dok je veličina paralelograma kontrolisana izborom određene verovatnoće, označene sa α_2 , korišćene za testiranje "veličine" individualnih izraza. Zato, α_2 predstavlja verovatnoću da se kaže da je komponenta deo signala kada ustvari to nije. Ove dve α vrednosti nisu formalno povezane u ovoj situaciji. Ipak, dvosmislenost se može umanjiti ako uspemo da dve oblasti budu jednake po veličini.

U izrazima (3.23) i (3.24) se koristi F raspodela za lociranje velikih vrednosti među izrazima T^2 dekompozicije. Ovo se radi zato što s obzirom da signal postoji, najverovatniji kandidati među jedinstvenim izrazima totalne MYT dekompozicije su komponente sa velikim vrednostima. Interes u lociranju izraza MYT dekompozicije koji daju signal je laka interpretacija ovih izraza.



Slika 3.12 Oblasti prihvatanja za T_1^2 i $T_{2,1}^2$



Slika 3.13 T^2 kontrolna oblast sa četiri signalne tačke

Posmatrajmo kontrolnu oblast za $p = 2$, datu na Slici 3.13. Primenom izraza (3.26) za $\alpha = 0.05$ dobija se tolerancija za promenljive x_1 i x_2 . Oblasti tolerancije su definisane Ševartovim kontrolnim granicama individualnih promenljivih za određeni α nivo.

Tabela 3.1: Vrednosti izraza MYT dekompozicije

Tačke	T^2 vrednost	T_1^2	T_2^2	$T_{1.2}^2$	$T_{2.1}^2$
A	10.05*	2.78	10.03*	0.02	7.27*
B	6.33*	0.11	3.49	2.83	6.22*
C	6.63*	3.01	0.34	6.29*	3.26
D	9.76*	2.54	1.73	8.03*	7.22*

Pregled vrednosti T^2 dekompozicije za četiri signalne tačke je predstavljen u Tabeli 3.1. Posmatrani vektor A ima veliku vrednost bezuslovnog izraza T_2^2 , jer je promenljiva x_2 van tolerancije. $T_{2.1}^2$ izraz za ovu tačku takođe je velik, jer x_2 nije sadržano u uslovnoj oblasti x_2 za dato x_1 . Ovo se vidi po velikom rastojanju tačke A od regresione linije $\bar{x}_{2.1}$. Isto tako, posmatrani vektor B ima veliku vrednost za njegov $T_{2.1}^2$ izraz, jer se nalazi na velikoj udaljenosti od iste regresione linije. Ipak, njegova dva bezuslovna izraza nisu značajna, jer individualne promenljive su unutar pravougaone oblasti, predstavljene isprekidanim linijama na Slici 3.13. Vektor C ima prihvatljive vrednosti za oba bezuslovna izraza, jer obe promenljive su unutar tolerancije. Ipak, velika vrednost se pojavljuje na $T_{1.2}^2$ izrazu, jer vrednost promenljive x_1 nije sadržana u uslovnoj raspodeli x_1 za dato x_2 . Opet, primetimo ekstremnu distancu tačke C od regresione linije $\bar{x}_{1.2}$. Vrednosti obe promenljive su unutar tolerancije za tačku D, ali vrednost promenljive x_1 ne slaže se onom, koja joj je predviđena na osnovu x_2 promenljive i obrnuto. Ovo produkuje velike vrednosti za ova dva uslovna izraza.

GLAVA 4

Interpretacija T^2 signala za opšti slučaj

U ovoj glavi, proširena je interpretacija T^2 signala na slučaje koji sadrže više od dve promenljive. MYT dekompozicija i dalje ostaje najvažnija stavka ove interpretacije, ali biće predstavljena i neka dodatna svojstva dekompozicije. Na primer, pokazano je da izrazi iz dekompozicije sadrže informacije o rezidualima, generisanim sa svim mogućim linearnim regresijama jedne promenljive u odnosu na bilo koju kombinaciju drugih promenljivih. Takođe, ovo svojstvo je korisno u lociranju izvora signala u izrazima sa jednom promenljivom, ali isto tako i sa grupom promenljivih, ovo svojstvo ima dve važne funkcije. Prvo, može se koristiti da poveća osetljivost T^2 statistike u oblastima sa malim procesnim promenama. Drugo, ova osobina je veoma korisna u razvoju kontrolnih procedura za autokorelisana posmatranja.

4.1 MYT dekompozicija za opšti slučaj

Proširujemo MYT dekompoziciju na opšti slučaj sa p promenljivih. T^2 statistika za p -dimenzionalni posmatrani vektor $X' = (x_1, x_2, \dots, x_p)$ može se predstaviti kao

$$T^2 = (X - \bar{X})' S^{-1} (X - \bar{X}). \quad (4.1)$$

Ako se vektor $(X - \bar{X})$ rastavi na sledeći način

$$(X - \bar{X})' = [(X^{(p-1)} - \bar{X}^{(p-1)}), (x_p - \bar{x}_p)]',$$

gde $X^{(p-1)'} = (x_1, x_2, \dots, x_{p-1})$ predstavlja $(p-1)$ -dimenzionalan vektor promenljivih, bez p -te promenljive x_p i $\bar{X}^{(p-1)}$ predstavlja odgovarajućih $p-1$ elemenata srednjeg vektora. Na sličan način podelimo matricu S tako da je

$$S = \begin{bmatrix} S_{XX} & S_{xx} \\ S_{xX}' & S_p^2 \end{bmatrix}, \quad (4.2)$$

gde je S_{XX} $(p-1) \times (p-1)$ matrica kovarijansi za prvih $(p-1)$ promenljivih, s_p^2 je varijansa x_p , s_{xx} je $(p-1)$ dimenzionalni vektor, koji sadrži kovarijanse između x_p i preostalih $(p-1)$ promenljivih. T^2 statistika iz (4.1) može se podeliti na dva nezavisna dela na sledeći način

$$T^2 = T_{p-1}^2 + T_{p,1,2,\dots,p-1}^2. \quad (4.3)$$

Prvi izraz iz jednačine (4.3)

$$T_{p-1}^2 = (X^{(p-1)} - \bar{X}^{(p-1)})' S_{XX}^{-1} (X^{(p-1)} - \bar{X}^{(p-1)}), \quad (4.4)$$

koristi prvih $p - 1$ promenljivih i sama po sebi je T^2 statistika.

Drugi izraz iz jednačine (4.3) je kvadrat p -te komponente vektora X prilagođenog ocenama srednje vrednosti i standardne devijacije uslovne raspodele promenljive x_p za poznate $(x_1, x_2, \dots, x_{p-1})$. Dat je sa

$$T_{p,1,2,\dots,p-1}^2 = (x_p - \bar{x}_{p,1,2,\dots,p-1})^2 / s_{p,1,2,\dots,p-1}^2,$$

gde je

$$\bar{x}_{p,1,2,\dots,p-1} = \bar{x}_p + B'_p (X^{(p-1)} - \bar{X}^{(p-1)}), \quad (4.5)$$

gde je

$$B'_p = S_{XX}^{-1} S_{xx},$$

$(p - 1)$ -dimenzionalni vektor ocena koeficijenata iz regresije x_p na $(p - 1)$ promenljivu x_1, x_2, \dots, x_{p-1} . Može se pokazati da je ocena uslovne varijanse data sa

$$s_{p,1,2,\dots,p-1}^2 = s_p^2 - S'_{xx} S_{XX}^{-1} S_{xx}.$$

Ponovo, kako je prvi izraz iz jednačine (4.3) T^2 statistika, takođe se može razdvojiti na dva ortogonalna dela

$$T_{p-1}^2 = T_{p-2}^2 + T_{p-1,1,2,\dots,p-2}^2.$$

Prvi izraz T_{p-2}^2 , je T^2 statistika za prve $p - 2$ komponente X vektora, a drugi izraz $T_{p-1,1,2,\dots,p-2}^2$, je kvadrat promenljive x_{p-1} prilagođen ocenama sredine i standardne devijacije uslovne raspodele promenljive x_{p-1} za poznate $(x_1, x_2, \dots, x_{p-2})$.

Nastavljajući dalje iteraciju i razdvajanje na ovaj način dobijamo jednu od mnogih MYT dekompozicija T^2 statistike

$$T^2 = T_1^2 + T_{2,1}^2 + T_{3,2,1}^2 + \dots + T_{p,1,2,\dots,p-1}^2. \quad (4.6)$$

T_1^2 izraz u jednačini (4.6) je kvadrat univarijantne t statistike za prvu promenljivu vektora X i računa se kao

$$T_1^2 = (x_1 - \bar{x}_1)^2 / s_1^2. \quad (4.7)$$

Primetimo da ovaj izraz nije uslovni izraz, jer njegova vrednost ne zavisi od neke uslovne raspodele. Nasuprot, svi drugi izrazi iz (4.6) su uslovni izrazi, jer predstavljaju vrednost promenljive prilagođene sredinom i standardnom devijacijom iz odgovarajuće uslovne raspodele. Ovi uslovni izrazi predstavljeni su sa standardnom tačka notacijom, koja se koristi u multivarijantnoj analizi za označavanje uslovnih raspodela. Pa tako, $T_{i,j,k}^2$ odgovara uslovnoj T^2 za raspodelu promenljive x_i prilagođene, ili uslovljene, sa promenljivama x_j i x_k .

4.2 Računanje izraza dekompozicije

Izrazi MYT dekompozicije mogu se računati na više načina. Iz (4.3) znamo da prvih $(p - 1)$ izraza jednačine (4.6) odgovara T^2 vrednosti podvektora $X'_{(p-1)} = (x_1, x_2, \dots, x_{p-1})$.

$$T_{(x_1, x_2, \dots, x_{p-1})}^2 = T_1^2 + T_{2,1}^2 + T_{3,2,1}^2 + \dots + T_{p-1,1,2,\dots,p-2}^2.$$

Slično, prvih $(p - 2)$ izraza ove formule odgovara podvektoru $X'_{(p-2)} = (x_1, x_2, \dots, x_{p-2})$.

$$T_{(x_1, x_2, \dots, x_{p-2})}^2 = T_1^2 + T_{2,1}^2 + T_{3,2,1}^2 + \dots + T_{p-2,1,2,\dots,p-3}^2.$$

Nastavljajući dalje na ovaj način, može se izračunati T^2 vrednosti svih podvektora originalnog vektora X . Poslednji podvektor, koji se sastoji samo od prve komponente $X'_1 = (x_1)$, se koristi za računanje bezuslovnog T^2 izraza datog u (4.7), tj.

$$T_{(x_1)}^2 = T_1^2.$$

Sve T^2 vrednosti $T_{(x_1, x_2, \dots, x_p)}^2, T_{(x_1, x_2, \dots, x_{p-1})}^2, \dots, T_{(x_1)}^2$ se računaju pomoću opšte formule

$$T_{(x_1, x_2, \dots, x_j)}^2 = (X^{(j)} - \bar{X}^{(j)})' S_{jj}^{-1} (X^{(j)} - \bar{X}^{(j)}), \quad (4.8)$$

gde je $X^{(j)}$ odgovarajući podvektor, a $\bar{X}^{(j)}$ je odgovarajuća sredina podvektora i S_{jj} označava odgovarajuću podmatricu kovarijansi dobijenu iz opšte matrice S , date u (4.2), brisanjem svih nekorištenih vrsta i kolona. Zato se izrazi MYT dekompozicije mogu izračunati na sledeći način

$$T_{p,1,2,\dots,p-1}^2 = T_{(x_1, x_2, \dots, x_p)}^2 - T_{(x_1, x_2, \dots, x_{p-1})}^2,$$

$$T_{p-1,1,2,\dots,p-2}^2 = T_{(x_1, x_2, \dots, x_{p-1})}^2 - T_{(x_1, x_2, \dots, x_{p-2})}^2,$$

.....

$$T_{2,1}^2 = T_{(x_1, x_2)}^2 - T_1^2,$$

$$T_1^2 = (x_1 - \bar{x}_1)^2 / s_1^2.$$

Pokažimo ovaj metod računanja izraza MYT dekompozicije na primeru. Zamislimo situaciju u industriji, gde se neki proces karakteriše sa tri procesne promenljive. Dati su istorijski podaci, dok je proces bio pod kontrolom, sa 23 posmatranja, takođe su date ocena matrice kovarijansi kao i ocena srednjeg vektora

$$S = \begin{bmatrix} 41.075 & 2.938 & 16.221 \\ 2.938 & 4.984 & 4.903 \\ 16.221 & 4.903 & 12.173 \end{bmatrix} \quad \bar{X} = \begin{bmatrix} 525.435 \\ 513.435 \\ 539.913 \end{bmatrix}.$$

Na Slici 4.1 je dat trodimenzionalni grafik sa 95% ($\alpha = 0.05$) kontrolnim elipsoidom. Za svrhe grafika podaci su centrirani u srednjoj vrednosti. T^2 vrednost za novi posmatrani vektor

$X' = (533, 514, 528)$ se računa pomoću formule (4.1) i dobija se T^2 vrednost od 79.9441. Za $\alpha=0.05$ i 23 posmatranja, T^2 kritična vrednost iznosi 11.923. Kako je posmatrana T^2 vrednost veća od kritične vrednosti dolazi se do zaključka da posmatrani vektor produkuje signal. Grafički prikaz signalne tačke je dat na Slici 4.2.

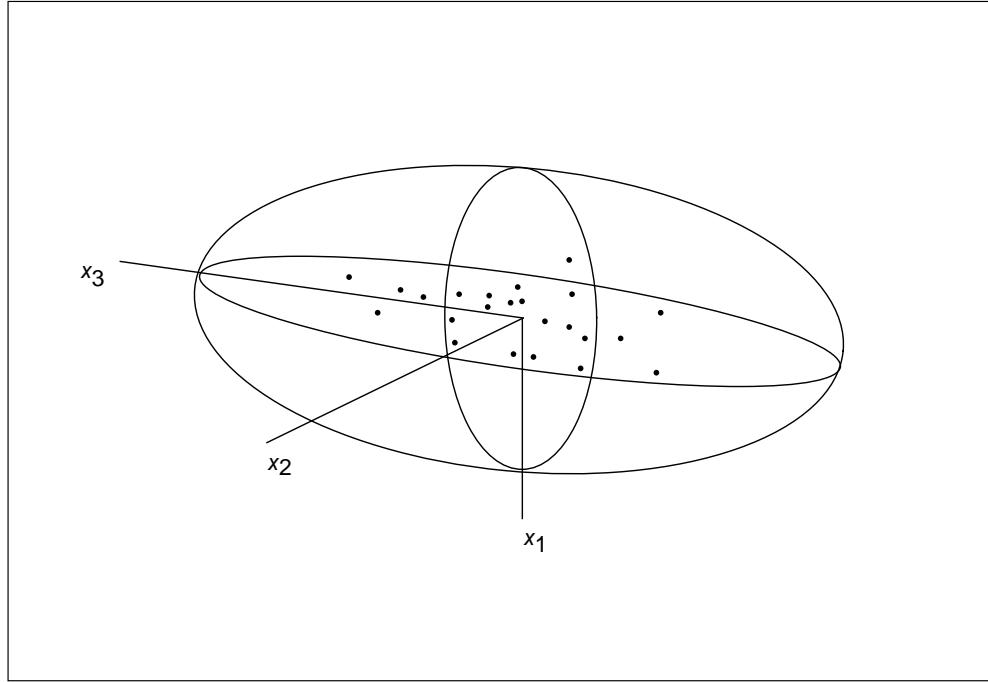
$$T^2 = 79.9441 > 11.923.$$

Iz glave 3 za $p = 2$ znamo da postoje dve različite MYT dekompozicije T^2 statistike. Za $p = 3$ postoji nekoliko različitih MYT dekompozicija, jedna od njih je

$$T_{(x_1, x_2, x_3)}^2 = T_1^2 + T_{2,1}^2 + T_{3,2,1}^2.$$

Da bi se izračunala ova vrednost, prvo treba izračunati vrednosti uslovnog izraza $T_{3,2,1}^2$. Na osnovu rečenog, sledi da je

$$T_{3,2,1}^2 = T_{(x_1, x_2, x_3)}^2 - T_{(x_1, x_2)}^2.$$



Slika 4.1 Elipsoidna kontrolna oblast za tri procesne promenljive

Za dobijanje vrednosti $T_{(x_1, x_2)}^2$ moraju se izbaciti treća kolona i treća vrsta iz ocene matrice kovarijansi S , kao i treća vrsta srednjeg vektora \bar{X} , da bi dobili srednji vektor i matricu kovarijansi podvektora $X^{(2)} = (x_1, x_2)$.

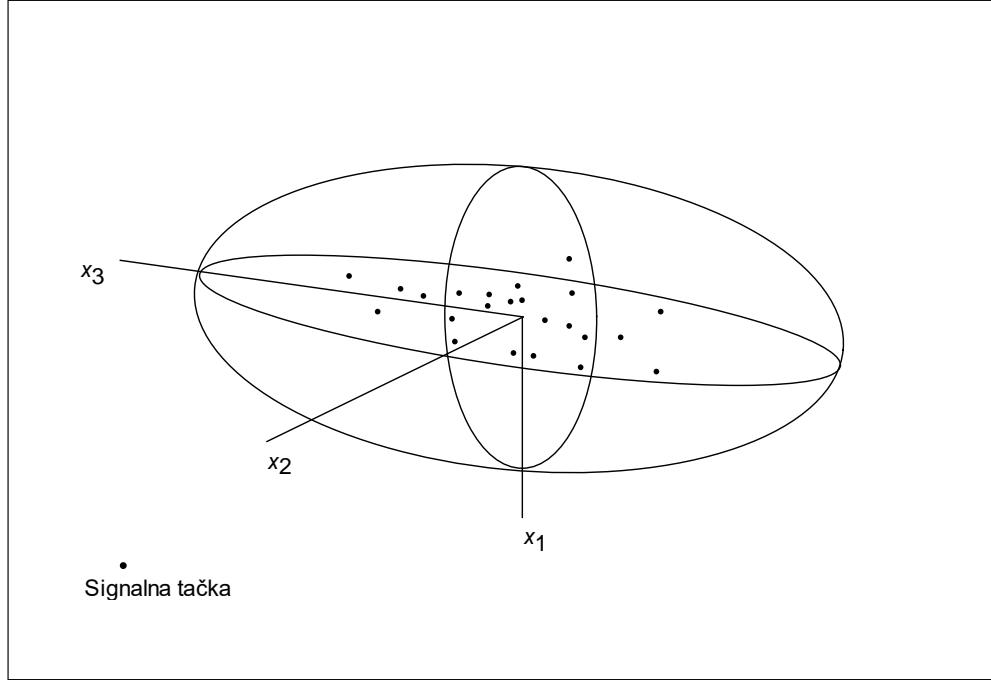
$$S_{22} = \begin{bmatrix} 41.075 & 2.938 \\ 2.938 & 4.984 \end{bmatrix} \quad \bar{X}^{(2)} = \begin{bmatrix} 525.435 \\ 513.435 \end{bmatrix}.$$

$T_{(x_1, x_2)}^2$ se računa na sledeći način

$$T_{(x_1, x_2)}^2 = (X^{(2)} - \bar{X}^{(2)})' S_{22}^{-1} (X^{(2)} - \bar{X}^{(2)}) = 1.3935,$$

dok se $T_{3,2,1}^2$ računa kao

$$T_{3,2,1}^2 = T_{(x_1, x_2, x_3)}^2 - T_{(x_1, x_2)}^2 = 79.9441 - 1.3935 = 78.5506.$$



Slika 4.2 Elipsoidna kontrolna oblast sa signalnom tačkom

Slično, dekompozicija za $T_{(x_1, x_2)}^2$ je data sa

$$T_{(x_1, x_2)}^2 = T_1^2 + T_{2,1}^2.$$

$T_{2,1}^2$ dobijamo računanjem T^2 vrednosti podvektora $X^{(1)} = (x_1)$. Ovo je bezuslovni izraz T_1^2 i računa se na sledeći način

$$T_1^2 = \frac{(x_1 - \bar{x}_1)^2}{s_1^2} = \frac{(533 - 525.435)^2}{41.075} = 1.3934.$$

Na osnovu čega sledi da je

$$\begin{aligned} T_{2,1}^2 &= T_{(x_1, x_2)}^2 - T_{(x_1)}^2 = T_{(x_1, x_2)}^2 - T_1^2 \\ &= 1.3935 - 1.3934 = 0.0001. \end{aligned}$$

Iz ovoga imamo da je

$$\begin{aligned} T_{(x_1, x_2, x_3)}^2 &= T_1^2 + T_{2.1}^2 + T_{3.2,1}^2 \\ &= 1.3934 + 0.0001 + 78.5506 \\ &= 79.9441, \end{aligned}$$

i male vrednosti prva dva izraza T_1^2 i $T_{2.1}^2$, impliciraju da se signal javlja u trećem izrazu $T_{3.2,1}^2$.

Ovo je samo jedna od nekoliko mogućih MYT dekompozicija koju smo izabrali da prikažemo tehniku računanja izraza dekompozicije. Da smo izabrali neku drugu dekompoziciju kao na primer

$$T^2 = T_3^2 + T_{1.3}^2 + T_{2.1,3}^2,$$

drugi izrazi dekompozicije bi imali velike vrednosti. Sa signalnom T^2 vrednosti, garantovano će bar jedan izraz bilo koje dekompozicije biti veliki.

4.3 Bitna svojstva MYT dekompozicije

MYT dekompozicija ima mnogo korisnih svojstva, u ovom podglavlju biće obrađena neka od njih. U p -dimenzionalnom prostoru neka je definisan vektor $X' = (x_1, x_2, \dots, x_p)$. Ako prve dve komponente zamene mesta dobija se novi vektor (x_2, x_1, \dots, x_p) , s tim da jedina razlika između ova dva vektora budu prve dve komponente, koje su permutovane. T^2 vrednosti ova dva vektora su jednakе tj.

$$T_{(x_1, x_2, \dots, x_p)}^2 = T_{(x_2, x_1, \dots, x_p)}^2. \quad (4.9)$$

Dakle, T^2 vrednost se ne menja permutacijom komponenti posmatranog vektora. Invarijantno svojstvo permutovanja komponenti T^2 garantuje da će se za bilo koji raspored komponenti u posmatranom vektoru dati istu T^2 vrednost. Kako postoji $p! = p(p-1)\dots 1$ permutacija komponenti vektora (x_1, x_2, \dots, x_p) , to znači da možemo razdeliti T^2 vrednost na $p!$ načina. Pokažimo ovo na primeru, neka je $p = 3$. Dakle, postoji $3! = 3 * 2 * 1 = 6$ različitih dekompozicija T^2 vrednosti za posmatrani trodimenzionalni vektor.

$$\begin{aligned} T^2 &= T_1^2 + T_{2.1}^2 + T_{3.1,2}^2 \\ &= T_1^2 + T_{3.1}^2 + T_{2.1,3}^2 \\ &= T_2^2 + T_{3.2}^2 + T_{1.2,3}^2 \\ &= T_2^2 + T_{1.2}^2 + T_{3.1,2}^2 \\ &= T_3^2 + T_{1.3}^2 + T_{2.1,3}^2 \end{aligned} \quad (4.10)$$

$$= T_3^2 + T_{2,3}^2 + T_{1,2,3}^2$$

Svaka vrsta iz (4.10) odgovara različitoj permutaciji komponenta posmatranog vektora. Na primer, prva vrsta odgovara vektoru napisanom u originalnoj formi (x_1, x_2, x_3) , dok poslednja odgovara permutaciji (x_3, x_2, x_1) . Svih p izraza iz jedne dekompozicije su nezavisni jedan od drugog, iako to ne mora da bude slučaj sa dekompozicijama uopšte. Sa p izraza u svakoj podeli i $p!$ podela, postoji $p \times p!$ mogućih izraza za računanje u ukupnoj MYT dekompoziciji signalne T^2 statistike. Srećom, ovi izrazi nisu jedinstveni, jer se pojavljuju više od jednom u raspodelama. Na primer, T_1^2 izraz se pojavljuje u prvoj i drugoj dekompoziciji navedenim u (4.10), isto tako $T_{2,1,3}^2$ izraz se pojavljuje u drugoj i petoj dekompoziciji. Uopšteno postoji $p \times 2^{(p-1)}$ različitih izraza među svim mogućim dekompozicijama. Ovi jedinstveni izrazi su oni koji se trebaju ispitati zbog mogućeg doprinosa T^2 signalu. Kada je p veliko, računanje svih ovih izraza može biti dugačak proces. Na primer, kada je $p = 10$, postoji preko 5000 jedinstvenih izraza u MYT dekompoziciji. Kako bi se olakšao ovaj zadatak, uspostavljeno je nekoliko računskih olakšica, o kojima će kasnije biti više reči.

Vratimo se na MYT dekompoziciju datu u (4.6) i prepostavimo da T_1^2 dominira u ukupnoj vrednosti T^2 statistike. Ovo indicira da promenljiva x_1 doprinosi signalu. Ipak, da bismo odredili da li preostale promenljive doprinose signalu, moramo da ispitamo T^2 vrednost za podvektor (x_2, x_3, \dots, x_p) , koji isključuje x_1 komponentu. Ako se dobije mala vrednost T^2 statistike za ovaj podvektor to znači da signal nije detektovan i nije potrebno ispitivati nijedan izraz iz ukupne dekompozicije, koji uključuje neki od ovih $p - 1$ promenljivih.

Sama činjenica da T^2 statistika može da se računa za bilo koji podvektor početnog vektora ima mnoge primene. Na primer, zamislimo situaciju gde je uzorak procesnih promenljivih češće dostupan nego uzorak laboratorijskih promenljivih. T^2 statistika može se izračunati za podvektor procesnih promenljivih kad god uzorak bude dostupan. Druga velika primena ovog rezultata je u situaciji kada neki podaci nedostaju. Na primer, ako se senzor za merenje neke promenljive pokvari, ne mora se prekidati kontrolna procedura već se može izbaciti ta promenljiva iz kontrolne procedure i raditi T^2 statistike na preostalim promenljivama dok se senzor ne zameni.

Vratimo se na primer iz prošlog poglavlja, posmatramo signalni vektor sa tri promenljive $X' = (533, 54, 528)$. Njegova T^2 vrednost je

$$T_{(533,514,528)}^2 = 79.9441.$$

Dalje, računamo T^2 vrednost za podvektor $X'^{(2)} = (533, 514)$

$$T_{(533,514)}^2 = 1.3935.$$

Mala vrednost ovog izraza sugerije da nema problema sa promenljivama x_1 i x_2 i dolazi se do zaključka da je signal uzrokovani sa promenljivom x_3 . Zato svi T^2 izrazi, uslovni i bezuslovni, koji uključuju samo x_1 i x_2 imaju male vrednosti. Te vrednosti su izračunate i zaista sve su male

$$T_1^2 = 1.3934, \quad T_{2,1}^2 = 0.0001, \quad T_2^2 = 0.0641, \quad T_{1,2}^2 = 1.3294.$$

Tabela 4.1: Lista svih mogućih regresija i uslovni T^2 izrazi za $p = 3$

Regresije	Uslovni T^2 izrazi
x_1 u odnosu na x_2	$T_{1,2}^2$
x_1 u odnosu na x_3	$T_{1,3}^2$
x_1 u odnosu na x_2, x_3	$T_{1,2,3}^2$
x_2 u odnosu na x_1	$T_{2,1}^2$
x_2 u odnosu na x_3	$T_{2,3}^2$
x_2 u odnosu na x_1, x_3	$T_{2,1,3}^2$
x_3 u odnosu na x_1	$T_{3,1}^2$
x_3 u odnosu na x_2	$T_{3,2}^2$
x_3 u odnosu na x_1, x_2	$T_{3,1,2}^2$

Još jedno važno svojstvo T^2 statistike je činjenica da $p(2^{p-1} - 1)$ jedinstvenih uslovnih izraza MYT dekompozicije sadrže reziduale iz svih mogućih linearnih regresija svake promenljive na svim podskupovima drugih promenljivih. Na primer, za $p = 3$ lista od 9 (tj. $3(2^{3-1} - 1)$) linearnih regresija za sve promenljive u odnosu na sve moguće podgrupe drugih promenljivih je data u Tabeli 4.1 zajedno sa odgovarajućim uslovnim T^2 izrazima. Ovo svojstvo T^2 statistike pruža proceduru za povećanje osetljivosti T^2 statistike na male procesne promene.

4.4 Lociranje signalnih promenljivih

Zamislimo signalni vektor $X' = (x_1, x_2, \dots, x_p)$ takav da je

$$T_{(x_1, x_2, \dots, x_p)}^2 > UCL.$$

U ovom poglavlju povezaće se interpretacija T^2 signala sa komponentama MYT dekompozicije. Jedna od metoda za lociranje promenljivih koje doprinose signalu je da se razvije iterativna šema unapred. Prvo treba pronaći podskup promenljivih koje ne doprinose signalu. Kako se T^2 statistika može konstruisati za bilo koji podskup promenljivih promenljivih x_1, x_2, \dots, x_p . Konstruišimo T^2 statistiku za svaku promenljivu pojedinačno, $x_j, j = 1, 2, \dots, p$

$$T_j^2 = \frac{(x_j - \bar{x}_j)^2}{s_j^2},$$

gde su \bar{x}_j i s_j^2 odgovarajuće ocene sredine i varijanse koje su dobijene iz istorijskih podataka.

Uporedimo ove T_j^2 vrednosti sa njihovim granicama UCL, gde se

$$\begin{aligned} UCL_{(x_j)} &= \left\{ \frac{p(n+1)(n-1)}{n(n-p)} \right\} F_{(\alpha,p,n-p)} \\ &= \left(\frac{n+1}{n} \right) F_{(\alpha,1,n-1)} \end{aligned} \quad (4.11)$$

računa za prikladan nivo α i za vrednost $p=1$. Isključimo iz originalnog skupa promenljivih sve x_j za koje važi

$$T_j^2 > UCL_{(x_j)},$$

jer ove promenljive definitivno doprinose signalu.

Zatim, iz skupa promenljivih koje ne doprinose signalu računamo T^2 statistiku za sve moguće parove promenljivih, tj. za sve (x_i, x_j) sa $i \neq j$ računamo $T_{(x_i, x_j)}^2$ i upoređujemo ove vrednosti sa gornjom kontrolnom granicom, koja se računa kao

$$UCL_{(x_i, x_j)} = \left\{ \frac{2(n+1)(n-1)}{n(n-2)} \right\} F_{(\alpha,2,n-2)}.$$

Ponovo, sve parove promenljivih za koje važi

$$T_{(x_i, x_j)}^2 > UCL_{(x_i, x_j)},$$

izbacimo iz ove grupe. Isključeni parovi promenljivih, zajedno sa isključenim pojedinačnim promenljivama, obuhvataju grupu promenljivih koje doprinose ukupnom signalu. Nastavimo iteraciju na ovaj način i uspevamo da isključimo iz preostale grupe sve promenljive iz signalne grupe sa tri promenljive, četiri promenljive i tako dalje. Ova procedura stvara grupu promenljivih koje doprinose signalu.

Pokažimo ovo na primeru. Vratimo se ponovo na primer sa trodimenzionalnim vektorom $X' = (533, 514, 528)$, koji ima T^2 vrednosti 79.9441. Bezuslovne T^2 vrednosti za pojedinačne promenljive su date sa

$$T_1^2 = 1.3934, \quad T_2^2 = 0.0641, \quad T_3^2 = 11.6578.$$

UCL se računa na sledeći način

$$UCL_{(x_j)} = \left(\frac{n+1}{n} \right) F_{(\alpha,1,n-1)} = \left(\frac{24}{23} \right) F_{(.01,1,22)} = 8.2956.$$

Iz ovih izraza dobijamo da su T_1^2 i T_2^2 pod kontrolom dok je

$$T_3^2 = 11.6578 > 8.2956,$$

što znači da treća promenljiva doprinosi signalu posmatranog vektora. Delimo originalni vektor na dve grupe (x_1, x_2) i (x_3) . Zatim, računa se T^2 statistika za promenljive x_1 i x_2

$$T_{(x_1, x_2)}^2 = 1.3935$$

i upoređuje sa gornjom granicom datom sa

$$UCL_{(x_1, x_2)} = \left\{ \frac{2(23+1)(23-1)}{23(23-2)} \right\} F_{(.01, 2, 21)} = 12.637.$$

Zaključujemo da nikakav signal nije prisutan u (x_1, x_2) komponenti posmatranog vektora. Stoga, razlog za signal leži u trećoj promenljivoj $x_3 = 528$.

Mana ove metode je da pruža jako malo informacija o tome kako komponente signarnog vektora doprinose signalu. Na primer, u skupu podataka od gore, ne prikazuje kako posmatrana vrednost 528 na promenljivoj x_3 doprinosi ukupnom signalu. Ipak, može se iskoristiti, kao indikator, da locira komponente posmatranog vektora koje doprinose signalu.

Još jedna metoda za lociranje komponenti vektora koji doprinose signalu je da se ispitaju individualni izrazi MYT dekompozicije signarnog vektora i da se pronađu oni koji imaju veliku vrednost. Ovo se postiže upoređivanjem svakog izraza sa odgovarajućom kritičnom vrednosti. Podsetimo se glave 3, raspodela za komponente MYT dekompozicije u situaciji gde nema signala je upravo F raspodela. Za slučaj sa p promenljivih, data je sa

$$T_j^2 \sim \left(\frac{n+1}{n} \right) F_{(1, n-1)} \quad (4.12)$$

za bezuslovne izraze, i sa

$$T_{j, 1, 2, \dots, j-1}^2 \sim \left[\frac{(n+1)(n-1)}{n(n-k-1)} \right] F_{(1, n-k-1)} \quad (4.13)$$

za uslovne izraze, gde je k broj uslovnih promenljivih. U slučaju kad je $k = 0$, raspodela (4.13) se svodi na raspodelu (4.12). Koristeći ove raspodele, kritične vrednosti CV , za određeni nivo α i uzorak istorijskih podataka veličine n za uslovne i bezuslovne izraze se dobijaju na sledeći način:

$$\text{Bezuslovni izrazi: } CV = \left(\frac{n+1}{n} \right) F_{(\alpha, 1, n-1)},$$

$$\text{Uslovni izrazi: } CV = \left[\frac{(n+1)(n-1)}{n(n-k-1)} \right] F_{(\alpha, 1, n-k)}. \quad (4.14)$$

Nakon svega ovoga, sada je lako uporediti svaki individualni izraz dekompozicije sa kritičnom vrednosti i doneti odgovarajuću odluku.

Pokažimo i ovo na primeru, vratimo se ponovo na trodimenzionalni vektor $X' = (533, 514, 528)$ sa T^2 vrednosti od 79.9441 uzet iz primera opisanom u poglavlu 4.3. Tabela 8.2 sadrži 12 jedinstvenih izraza i njihovih vrednosti za ukupnu dekompoziciju ove T^2 vrednosti. Velike komponente se određuju upoređivanjem vrednosti svakog izraza sa odgovarajućom kritičnom vrednosti.

Tabela 4.2: Jedinstveni T^2 izrazi dekompozicije

Izrazi	Vrednosti
T_1^2	1.3934
T_2^2	0.0641
T_3^2	11.6578*
$T_{1,2}^2$	1.3294
$T_{1,3}^2$	28.2305*
$T_{2,1}^2$	0.0001
$T_{2,3}^2$	9.5584*
$T_{3,1}^2$	38.4949*
$T_{3,2}^2$	21.1522*
$T_{1,2,3}^2$	58.7278*
$T_{2,1,3}^2$	40.0558*
$T_{3,1,2}^2$	78.5506*

T^2 vrednosti sa zvezdicom određuju izraze koji doprinose ukupnom T^2 signalu, tj. $T_3^2, T_{1,3}^2, T_{2,3}^2, T_{3,1}^2, T_{3,2}^2, T_{1,2,3}^2, T_{2,1,3}^2, T_{3,1,2}^2$. Svi ovi izrazi sadrže promenljivu x_3 . Ovo je ista promenljiva koja je određena preciznom metodom za detektovanje signalnih promenljivih. Zbog toga, moramo zaključiti da problem leži u ovoj promenljivoj. Ipak, jak argument može se dati da je problem zbog druge dve promenljive, jer četiri signalna izraza sadrže x_1 i četiri sadrže x_2 . Da bi rešili ovaj problem, potrebno nam je više razumevanja o tome šta prouzrokuje signal u izrazima dekompozicije.

4.5 Interpretacija signala na osnovu T^2 komponenti

Uzmemo jedan od p mogućih bezuslovnih izraza dobijenih iz dekompozicije T^2 statistike signalnog vektora. Kao što je rečeno ranije, izraz

$$T_j^2 = \frac{(x_j - \bar{x}_j)^2}{s_j^2},$$

$j = 1, 2, \dots, p$, je kvadrat univarijantne t statistika za posmatranu vrednost j-te promenljive posmatranog vektora X . Da bi se kontrola održala, ova komponenta mora biti manja od njene kritične vrednosti, tj.

$$T_j^2 < \left(\frac{n+1}{n}\right) F_{(\alpha, 1, n-1)}.$$

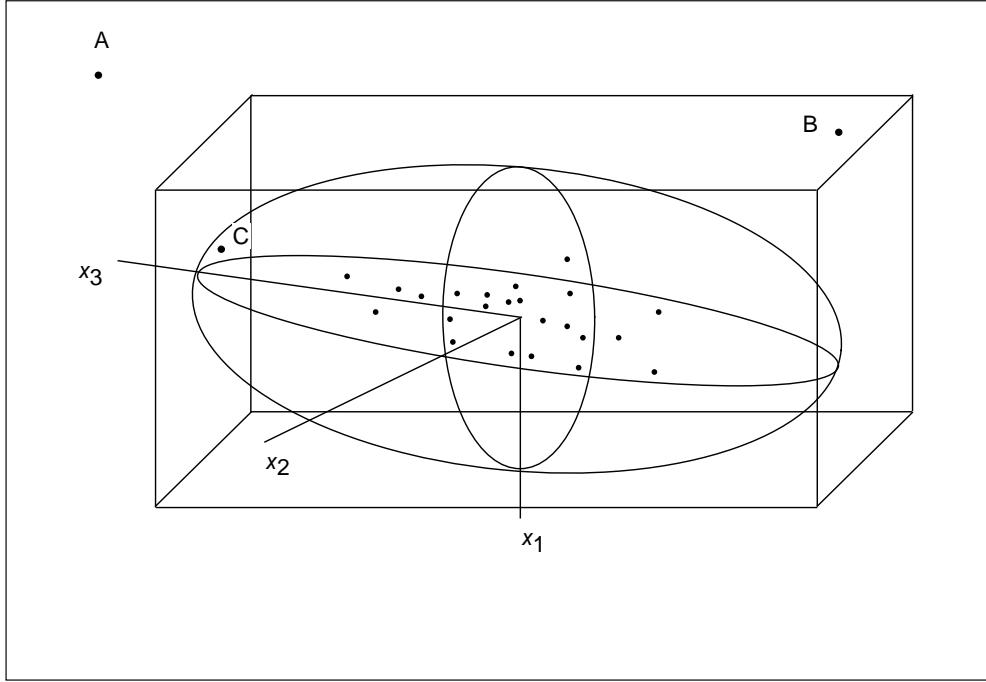
Kako je $t_{(\frac{\alpha}{2}, n-1)} = \sqrt{F_{(\alpha, 1, n-1)}}$, možemo da izrazimo ovaj izraz tako da se T_j nalazi u intervalu:

$$-\sqrt{\frac{n+1}{n}} t_{(\frac{\alpha}{2}, n-1)} < T_j < \sqrt{\frac{n+1}{n}} t_{(\frac{\alpha}{2}, n-1)} \quad (4.15)$$

$$\bar{x} - \sqrt{\frac{n+1}{n}} t_{(\frac{\alpha}{2}, n-1)} s_j < x_j < \bar{x} + \sqrt{\frac{n+1}{n}} t_{(\frac{\alpha}{2}, n-1)} s_j \quad (4.16)$$

gde je $t_{(\frac{\alpha}{2}, n-1)}$ odgovarajuća vrednost t raspodele sa $n-1$ stepeni slobode. Ovo je ekvivalent korišćenju univarijantnog Ševartovog kontrolnog grafika za j -tu promenljivu. Ako se za svaku od p promenljivih konstruišu kontrolni limiti (4.16) i ucrtaju u p -dimenzionalni prostor, dobili bismo hiperpravougaonu kutiju. Ova kutija je ekvivalent Ševartovoj kontrolnoj proceduri na svim individualnim promenljivama. Ipak, pravi kontrolni region, baziran na T^2 statistici, je hiperelipsoid koji se nalazi unutar te kutije. U većini slučajeva, elipsoid neće stati unutar kutije.

Na Slici 4.3 predstavljena je situacija za $p = 3$. Pravougaona kutija predstavlja kontrolne granice za individualne promenljive, izračunate pomoću jednačine (4.16), dok elipsoid predstavlja kontrolnu oblast za ukupnu T^2 statistiku. Takođe, na slici su date i tri signalne tačke, A, B, C . Ako se posmatrani vektor nalazi van granica kutije, na osnovu signalne T_j^2 će se identifikovati promenljiva koja je van kontrole. Kao što je predstavljeno na Slici 4.3 sa tačkom A . Tačka C takođe se nalazi van oblasti kutije, ali ukupna T^2 vrednost za ovu tačku neće biti signalna jer je tačka unutar eliptičke oblasti.



Slika 4.3 Elipsoid unutar kutije

Predimo sada na proučavanje uslovnih izraza dekompozicije, što bi trebalo da na pruži dodatan uvid u lokaciju i uzrok signala. Razmotrimo formu opšteg uslovnog izraza

$$T_{j,1,2,\dots,j-1}^2 = \frac{(x_j - \bar{x}_{j,1,2,\dots,j-1})^2}{s_{j,1,2,\dots,j-1}^2}. \quad (4.17)$$

Ako je vrednost izraza (4.17) manja od kontrolne granice,

$$T_{j,1,2,\dots,j-1}^2 < \left[\frac{(n+1)(n-1)}{n(n-k-1)} \right] F_{(\alpha, 1, n-k-1)},$$

njen brojilac mora biti mali, jer je imenilac ovih izraza fiksiran sa istorijskim podacima. Što znači da je komponenta x_j iz posmatranog vektora $X' = (x_1, x_2, \dots, x_j, \dots, x_p)$ sadržana u uslovnoj raspodeli promenljive x_j za dato x_1, x_2, \dots, x_{j-1} i upada u eliptičku kontrolnu oblast. Signal se pojavljuje na izrazu (4.17) kada x_j nije sadržano u uslovnoj raspodeli promenljive x_j za dato x_1, x_2, \dots, x_{j-1} tj. kad

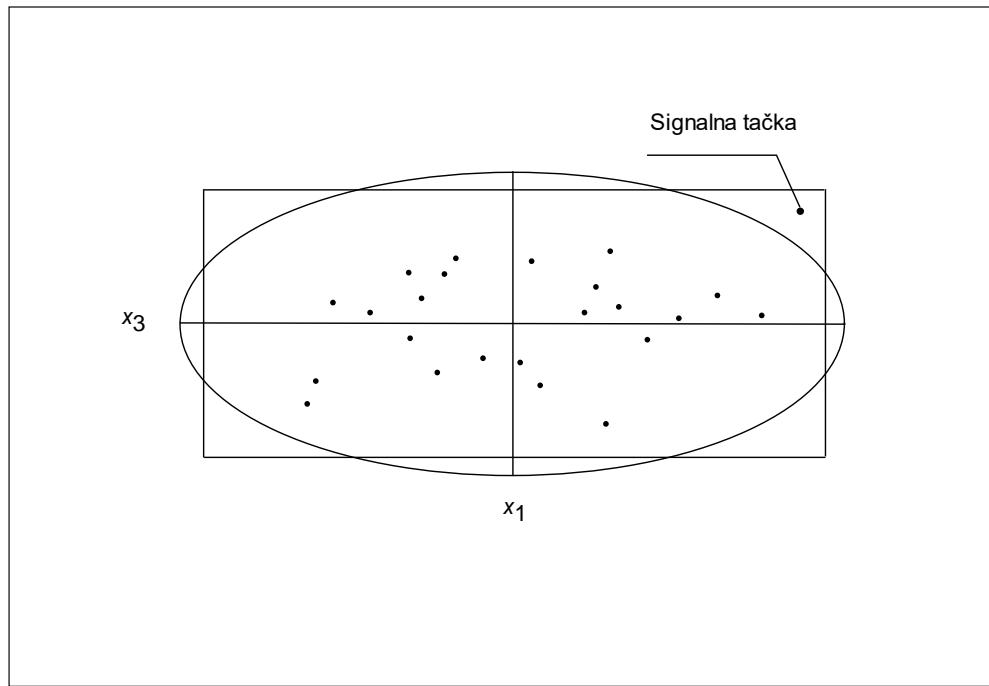
$$T_{j,1,2,\dots,j-1}^2 > \left[\frac{(n+1)(n-1)}{n(n-k-1)} \right] F_{(\alpha, 1, n-k-1)}.$$

Kada se ovo desi, to znači da odnos između promenljivih x_1, x_2, \dots, x_j nije dobar. Na primer, signal na $T_{j,1,2,\dots,j-1}^2$ implicira da posmatranje na x_j nije tamo gde bi trebalo da bude na osnovu vrednost promenljivih x_1, x_2, \dots, x_{j-1} , tj. odnos između x_j i drugih promenljivih nije onakav kakav je predviđen istorijskim podacima. Da bi pokazali kontrakorelaciju, razmotrimo trag kontrolne oblasti na Slici 4.3 u prostoru x_1 i x_3 kao što je predstavljeno na Slici 4.4. Signalna tačka B sa Slike 4.3 se

nalazi u gornjem desnom ugлу na Slici 4.4, unutar operacionog ranga x_1 i x_3 , ali van T^2 kontrolne oblasti. Zato, niti T_1^2 niti T_3^2 neće dati signal, ali će zato izrazi $T_{1,3}^2$ i $T_{3,1}^2$ dati signal.

Uslovne raspodele su uspostavljene sa korelacionom strukturom među promenljivama, i uslovni izrazi MYT dekompozicije zavise od ove strukture. Pod pretpostavkom multivarijantne normalnosti, i pod pretpostavkom da nijedna korelacija ne pokazuje zavisnost među promenljivama. Dekompozicija T^2 statistike za ovu situaciju neće sadržati uslovne izraze. Da bismo ovo videli, zamislimo p-varijantnu normalnu raspodelu koja ima poznatu sredinu μ i kovarijansnu matricu datu sa

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma_p^2 \end{bmatrix}.$$



Slika 4.4 Trag kontrolne oblasti za x_1 i x_3

T^2 vrednost za posmatrani vektor X uzet iz ove raspodele će biti data sa

$$\begin{aligned} T^2 &= (X - \mu)' \Sigma^{-1} (X - \mu) = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} + \dots + \frac{(x_p - \mu_p)^2}{\sigma_p^2} \\ &= T_1^2 + T_2^2 + \dots + T_p^2. \end{aligned}$$

Kao što je i očekivano, svi izrazi dekompozicije će biti bezuslovni

$$T_i^2 = \frac{(x_i - \mu_i)^2}{\sigma_i^2}.$$

Dakle, signal na bezuslovnom izrazu implicira da je posmatrana promenljiva van operacionallnog ranga, određenog istorijskim podacima. Vektor koji sadrži signal ovog tipa će na grafiku biti van hiperpravougaone kutije, definisane sa kontrolnim granicama za sve bezuslovne izraze. Ukupani T^2 signal na vektorima unutar kutije, koji imaju velike uslovne izraze, impliciraju da nešto nije u redu sa odnosom između promenljivih koje se nalaze u izrazu. Sve te promenljive će morati da se provere da bi se identifikovao uzrok signala.

Vratimo sa na trodimenzionalni vektor $X' = (533, 514, 528)$, kako bi uradili interpretaciju signala potrebni su još neki podaci. Ti podaci su dati u tabelama 4.3a i 4.3b.

Tabela 4.3a: Pregled statistike

	x_1	x_2	x_3
Srednja vrednost	525.435	513.435	539.913
Minimum	536.000	509.000	532.000
Maksimum	536.000	518.000	546.000

Tabela 4.3b: Korelaciona matica

	x_1	x_2	x_3
x_1	1.000	0.205	0.725
x_2	0.205	1.000	0.629
x_3	0.725	0.629	1.000

Razmotrimo iz Tabele 4.2 vrednost $T_{1,3}^2 = 28.2305$, koja je velika jer premašuje graničnu vrednost $CV = 8.2956$. Veličina $T_{1,3}^2$ implicira da nešto nije u redu sa odnosom između promenljivih x_1 i x_3 . Primetimo iz Tabele 4.3b, gde je na osnovu istorijskih podataka korelacija između ove dve promenljive je 0.725. Ovo znači da te dve promenljive variraju zajedno u istom smeru. Ipak, za posmatrani vektor $X' = (533, 514, 528)$, vrednost za $x_1 = 533$ nešto iznad njene srednje vrednosti od 525.435, dok je vrednost $x_3 = 528$ što je dosta ispod njene srednje vrednosti od 539.913. Da bi

se uspostavila ponovo kontrola nad procesom, ili x_1 se mora smanjiti ako je moguće, ili se x_3 mora povećati.

Da bismo znali koju promenljivu da korigujemo moramo biti dobro upoznati sa procesom i procesnim promenljivama. Ovo uključuje i znanje koja promenljiva je najlakša za kontrolu. Ako se lakše upravlja sa x_1 nego sa x_3 , onda treba smanjiti x_1 i obrnuto. Ako se na obe može uticati, onda možda trebamo da razmišljamo o velikoj vrednosti bezuslovnog izraza T_3^2 u Tabeli 4.2, tj.

$$T_3^2 = 11.6578 > CV = 8.2956.$$

Velika vrednost bezuslovnog izraza implicira da je vrednost promenljive je van Ševartove kutije. Ta promenljiva je $x_3 = 528$, i ona je značajno manja od minimalne vrednosti od 532, navedene u istorijskim podacima. Stoga, da bismo povratili kontrolu u ovoj situaciji trebalo bi da prilagodimo promenljivu x_3 , tj. da je povećamo.

4.6 Regresija sa više promenljivih

Urađena je interpretacija signala na uslovnim izrazima sa aspekta regresije za slučaj sa dve promenljive, sada ćemo je proširiti na opšti slučaj, tj. više od dve promenljive. $T_{j,1,2,\dots,j-1}^2$ je standardizovano posmatranje j-te promenljive prilagođene sa ocenama srednje vrednosti i varijanse iz uslovne raspodele $x_{j,1,2,\dots,j-1}$. Opšta forma ovog izraza je data u (4.17).

Razmotrimo ocenjenu srednju vrednost x_j prilagođenu za x_1, x_2, \dots, x_{j-1} . Ocenjujemo ovu vrednost pomoću jednačine predviđanja

$$\bar{x}_{j,1,2,\dots,j-1} = \bar{x}_j + B'_j(X^{(j-1)} - \bar{X}^{(j-1)}), \quad (4.18)$$

gde je \bar{x}_j uzoračka sredina promenljive x_j dobijena iz istorijskih podataka. Podvektor $X^{(j-1)}$ je sastavljen od posmatranja na $(x_1, x_2, \dots, x_{j-1})$, i $\bar{X}^{(j-1)}$ je odgovarajući ocenjeni srednji vektor dobijen iz istorijskih podataka. Vektor ocjenjenog regresionog koeficijenta B_j se dobija podeljivanjem podmatrice S_{jj} , matrice kovarijansi prvih j komponenti vektora X. Da bi dobili S_{jj} , podelimo prvo matricu S na sledeći način:

$$S = \begin{bmatrix} S_{jj} & S_{j(p-j)} \\ S'_{j(p-j)} & S_{(p-j)(p-j)} \end{bmatrix}.$$

Dalje, radimo podelu matrice S_{jj}

$$S_{jj} = \begin{bmatrix} S_{(j-1)(j-1)} & S_{j(j-1)} \\ S'_{j(j-1)} & S_j^2 \end{bmatrix}.$$

Onda

$$B_j = S_{(j-1)(j-1)}^{-1} S_{j(j-1)}.$$

Kako leva strana izraza (4.18) sadrži $\bar{x}_{j,1,2,\dots,j-1}$, što je predviđena vrednost promenljive x_j za date vrednosti x_1, x_2, \dots, x_{j-1} , brojilac iz (4.17) je regresioni rezidual dat sa

$$r_{j,1,2,\dots,j-1} = (x_j - \bar{x}_{j,1,2,\dots,j-1}).$$

Ako napišemo uslovnu varijansu kao

$$s_{j,1,2,\dots,j-1}^2 = s_j^2(1 - R_{j,1,2,\dots,j-1}^2)$$

i zamenom $r_{j,1,2,\dots,j-1}$ za $(x_j - \bar{x}_{j,1,2,\dots,j-1})$, možemo da izrazimo $T_{j,1,2,\dots,j-1}^2$ kao kvadratni standardni rezidual koji ima oblik

$$T_{j,1,2,\dots,j-1}^2 = \frac{(r_{j,1,2,\dots,j-1})^2}{s_j^2(1 - R_{j,1,2,\dots,j-1}^2)}. \quad (4.19)$$

Uslovni izraz u ovoj formi objašnjava kako dobro će buduće posmatranje određene promenljive biti u skladu sa vrednošću predviđenu skupom vrednosti drugih promenljivih u vektoru, koristeći matricu kovarijansi konstruisanu pomoću istorijskih podataka. Ukoliko imenilac u (4.19) je jako mali, što se dešava kada je R_j^2 blizu 1, velika vrednost uslovnog T^2 izraza će biti zbog brojioca, koji je funkcija slaganja između posmatranih i predviđenih vrednosti x_j . Čak i kad je imenilac mali, što se dešava kad R_j^2 ima veliku vrednost, očekivali bismo veliko slaganje između posmatranih i predviđenih vrednosti x_j . Značajna razlika između ovih vrednosti produkovaće veliku vrednost T^2 izraza.

Rezultat od gore ukazuje da T^2 signal može da se pojavi ako nešto podje naopako sa odnosom između podskupova različitih promenljivih. Ova situacija može se utvrditi ispitivanjem uslovnih T^2 izraza. Signalna vrednost ukazuje da kontradikcija sa istorijskom vezom između promenljivih se pojavila (1) ili zbog vrednosti standardizovane komponente koja je značajno veća ili manja od one predviđene skupom preostalih promenljivih, ili (2) vrednosti standardizovane komponente koja je marginalno manja ili veća od predviđene podskupom preostalih promenljivih kada postoji ozbiljna kolinearnost (tj. Velika vrednost R^2) među promenljivama. Zato signal se pojavljuje kada posmatranje određene promenljive ili skupa promenljivih je van kontrole i/ili kada se posmatranja na skupu promenljivih ne slažu sa odnosom uspostavljenim istorijskim podacima.

4.7 Šema računanja

Primarna stvar za donošenje odluka pri korišćenju multivariatne kontrolne procedure je dužina računanja, odnosno kako nalakše i najbrže doći do rešenja. Prilikom interpretacije može se

zahtevati brojna izračunavanja, i ova činjenica može da obeshrabri. MYT dekompozicija T^2 statistike se pokazala kao odlična vodilja u interpretaciji signalnih T^2 vrednosti, ali broj jedinstvenih izraza može biti ogroman, naročito kad broj promenljivih p prelazi 10. Ovaj problem je vodio do razvoja kompjuterskih programa koji mogu momentalno da pronađu značajne komponente dekompozicije za poprilično velike grupe promenljivih. Ipak, pitanje kako će ove metode računanja raditi kada imamo stotine hiljada promenljivih tek treba da dobije odgovor. Postoje mnogi faktori koji utiču na sposobnosti procedure, one uključuju kapacitete kompjutera, brzinu kompjutera, veličinu podataka, i programske algoritme. Predstavljena je sekvensijalna šema računanja, koja može da smanji broj računanja na neki razuman broj.

KORAK 1. Izračunati pojedinačne T^2 statistike za svaku komponentu X vektora. Ukloniti promenljive čija je T_i^2 vrednost velika. Posmatranje na ovim promenljivama je van pojedinačne kontrole, i nije potrebno da proveravamo kako se odnose prema drugim posmatranim promenljivama. Sa ovim uklanjanjem ovih značajnih promenljivih, smanjili smo skup promenljivih. Proveriti podvektor preostalih k promenljivih za signal. Ako signala više nema, onda smo uspešno locirali problem.

KORAK 2. Ako je signal ostao u podvektoru od preostalih k promenljivih, izračunati sve $T_{i,j}^2$ izraze. Izbaciti iz razmatranja sve parove promenljivih (x_i, x_j) , koji imaju značajan $T_{i,j}^2$ izraz. Ovo znači da nešto nije u redu sa bivarijantnim odnosima. Kada se ovo desi smanjiće se opet skup promenljivih koji se posmatra. Ispitati sve ukljonjene promenljive za uzrok signala. Izračunati T^2 vrednost za preostali podvektor. Ako nema signala, uzrok problema je u bivarijantnim odnosima u tim promenljivama koje su van individualne kontrole.

KORAK 3. Ako podvektor preostalih promenljivih i dalje sadrži signal, izračunati sve $T_{i,j,k}^2$ izraze. Izbaciti sve trostrukе kombinacije promenljivih (x_i, x_j, x_k) koje daju značajan rezultat i proveriti preostali vektor za signal.

KORAK 4. Nastaviti sa računanjem izraza višeg reda na ovaj način dok ne preostane promenljivih u podvektoru. U najgorem slučaju moraćemo da izračunamo sve jedinstvene izraze.

GLAVA 5

Primer interpretacije T^2 signala za opšti slučaj

Da bi predstavili interpretaciju signala za T^2 statistiku i šemu računanja za lociranje signala u izrazima MYT dekompozicije, predstavljamo sledeći primer. Zamislimo proces koji proizvodi specijalni proizvod na zahtev. Kvalitetne karakteristike ovog proizvoda moraju da zadovoljavaju rigorozne standarde zahtevane od strane mušterije. Svaka devijacija vodi do ozbiljnih pomenutih u mušterijinom poslovnom procesu. Kvalitet proizvoda se meri na osnovu sedam promenljivih. Prihvativ proizvod je uspostavljen na osnovu karakteristika promenljivih iz istorijskih podataka koji se sastoje od 85 prethodno prihvaćenih proizvoda. Uspostavljanje razloga za odbacivanje proizvoda je najvažnije, u većini slučajeva, moguće je doraditi odbačene proizvode da bi zadovoljili standarde.

U ovom primeru, kontrolna procedura za kvalitet proizvoda ima dve svrhe. Prva, ako proizvod zadovoljava standarde istorijskih podataka smatra se da je spreman za otpremanje do mušterije. Druga, analizom odbačenih proizvoda možemo da odredimo razloge za njegovo odbacivanje kroz upotrebu MYT dekompozicije na signalnu T^2 statistiku. Biti u mogućnosti da se odredi koje promenljive su razlog za odbacivanje i kako one doprinose ukupnom signalu može biti od pomoći u procesu dorade odbačenih proizvoda. Na primer, predvidimo nove ciljne vrednosti regresijom iz istorijskih podataka za one promenljive koje doprinose odbacivanju. Ove nove vrednosti su prilagođene ne samo istorijskim podacima, nego i vrednostima promenljivih koje ne doprinose signalu.

Istorijski podaci o kvalitetu sedam promenljivih su dati u Tabeli 5.1 na kraju poglavlja i karakterizovani su sa summary statistikom datom u Tabeli 5.2 i korelacionom matricom predstavljenom u Tabeli 5.3. Kao što će biti pokazano kasnije, ova statistika će imati važnu ulogu u interpretaciji signala za pojedinačne komponente dekompozicije.

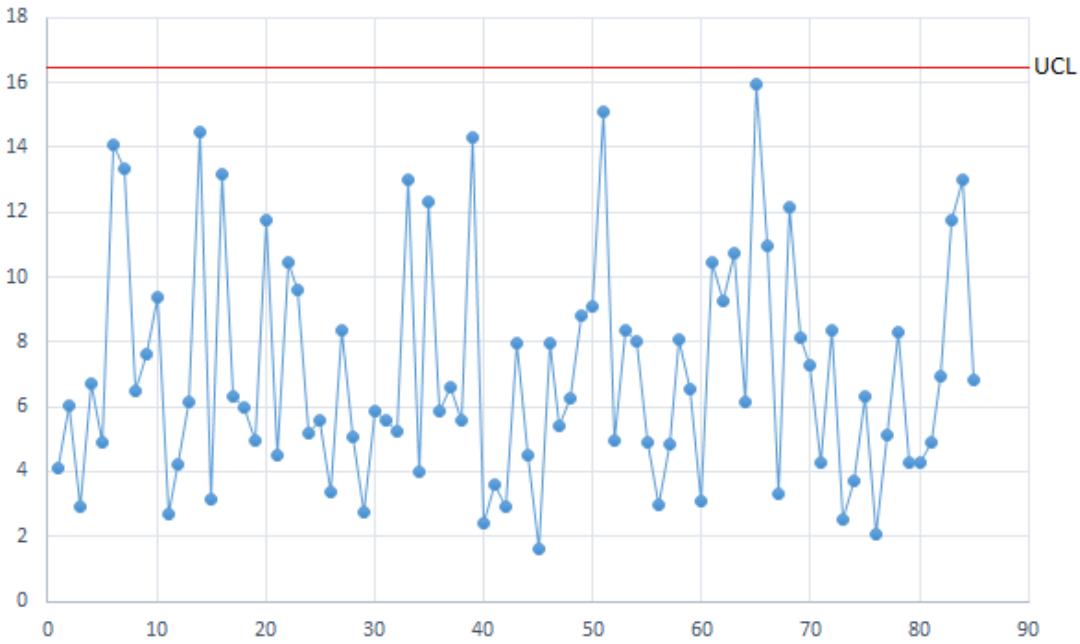
Tabela 5.2: Summary statistika

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
<i>Srednja vrednost</i>	87.14	7.43	3.47	0.45	10.25	18.22	0.98
<i>Minimum</i>	86.47	6.84	3.06	0.12	9.86	17.93	0.57
<i>Maksimum</i>	87.87	7.98	3.93	0.83	10.60	18.56	1.44
<i>St. devijacija</i>	0.24	0.24	0.22	0.16	0.14	0.12	0.18

Tabela 5.3: Korelaciona matica

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	1						
x_2	0.576	1					
x_3	0.719	-0.080	1				
x_4	0.793	0.019	0.989	1			
x_5	0.039	-0.775	0.663	0.585	1		
x_6	0.689	0.113	0.514	0.581	0.276	1	
x_7	0.296	-0.560	0.670	0.645	0.853	0.641	1

Takođe, dat je i grafik T^2 statistike za istorijske podatke na Slici 5.1. Granica za ukupnu T^2 statistiku istorijskih podataka izračunata je pomoću formule (4.11), gde je $n = 85$, $p = 7$ i $\alpha = 0.05$, dobijena je vrednost $UCL = 16.2458$. Vrednosti T^2 statistike za istorijske podatke su dati takođe u Tabeli 5.1. Ovih sedam procesnih promenljivih predstavljaju određenu hemijsku kompoziciju sadržanu u proizvodu. Ne samo da posmatranja na ovim promenljivama moraju biti održana u striktnom operativnom rangu, već moraju da održavaju međusobni odnos određen korelacionom matricom istrijskih podataka. Ova situacija je tipičan primer multivarijantnog sistema.



Slika 5.1 T^2 grafik za istorijske podatke sa označenom gornjom kontrolnom granicom

Neka su pristigli novih 33 podataka, koje treba proveriti za mogući signal. Dati su u Tabeli 5.4, koja se nalazi na kraju poglavlja. Potrebno je izračunati T^2 statistiku za svaki od tih vektora, računanje se radi po formuli (4.1). Međutim, kako nam nije data kovarijansna matrica, moramo naći njenu ocenu na osnovu istorijskih podataka. Data je u Tabeli 5.5.

Tabela 5.5: Kovarijansna matrica

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	0.0560						
x_2	0,0324	0,0564					
x_3	0,0365	-0,0041	0,0461				
x_4	0,0294	0,0007	0,0332	0,0245			
x_5	0,0013	-0,0254	0,0196	0,0126	0,0190		
x_6	0,0196	0,0032	0,0133	0,0110	0,0046	0,0146	
x_7	0,0127	-0,0240	0,0260	0,0182	0,0212	0,0140	0,0326

Dalje, nađemo njenu inverznu matricu i izračunamo T^2 statistike za svaki vektor podataka, one su date u poslednjoj koloni Tabele 5.4. Uporedimo te vrednosti sa gornjom kontrolnom granicom $UCL = 16.2458$ i vidimo da postoji tri signalna vektora, te vrednosti su obeležene sa *. Počnimo sada sa interpretacijom. Pogledajmo prvi signalni vektor $X' = (87.3, 8.4, 3.5, 0.5, 10.3, 18.3, 1.1)$. Njegova T^2 vrednost je

$$T^2 = 2483.934 > 16.2458,$$

kao što vidimo daleko premašuje graničnu vrednost. Pokušajmo da interpretiramo ovu T^2 vrednost pomoću predstavljene šeme računanja. Tabela 5.6 sadrži pojedinačne T^2 vrednosti za sedam bezuslovnih izraza. Značajnost ovih izraza se određuje njihovim upoređivanjem sa kritičnom vrednosti, koja je izračunata iz (4.14), gde je $n = 85$ i $\alpha = 0.05$. Relativno velika stopa greške tipa 1 (tj. $\alpha = 0.05$) se koristi za zaštitu mušterije od primanja loših proizvoda. Analiza rizika, koja je korišćena u određivanju vrednosti stope greške tipa 1, zahteva da bude prihvatljivije da se odbaci dobar proizvod nego da se pošalje loš proizvod mušteriji. Dakle, granična vrednost je

$$CV = 3.99647.$$

Samo bezuslovni izraz T_2^2 prelazi kritičnu vrednost, što ukazuje da promenljiva x_2 doprinosi ukupnom signalu. Kada uporedimo vrednost druge promenljive sa maksimalnom, koja je dobijena iz istorijskih podataka, datom u Tabeli 5.2, saznajemo da je veća od maksimalne vrednosti, tj. $8.4 > 7.98$.

Izbacimo ovu promenljivu iz vektora i testiramo preostali podvektor za signal. Dobija se vrednost

$$T_{(87.3,3.5,0.5,10.3,18.3,1.1)}^2 = 7.93099$$

što je manje od granične vrednosti 14.32966, dobijene formulom (4.11), gde je $n = 85$, $p = 6$, a $\alpha = 0.05$. Kako je vrednost druge promenljive veća od maksimalne potrebno ju je smanjiti.

Prihvatljiva vrednost može da se nađe predviđanjem vrednosti za x_2 pomoću fiksnih vrednosti preostalih promenljivih i regresione jednačine, dobijene iz istorijskih podataka.

Tabela 5.6: Vrednost bezuslovnih izraza za prvi signalni vektor

T^2	Vrednost
T_1^2	0.4826
T_2^2	16.5778*
T_3^2	0.0263
T_4^2	0.0870
T_5^2	0.1122
T_6^2	0.4477
T_7^2	0.4323

Regresiona jednačina je data sa

$$\hat{x}_2 = 1.12535x_1 - 0.15723x_3 - 0.45538x_4 - 0.51851x_5 - 0.35519x_6 - 0.30433x_7 - 77.78694$$

Kada se ubace vrednosti promenljivih iz podvektora dobija se $x_2 = 7.5027$. Ako proverimo sada ukupnu T^2 statistiku, sa novom vrednosti za x_2 , dobija se

$$T^2_{(87.3, 7.5027, 3.5, 0.5, 10.3, 18.3, 1.1)} = 7.930653,$$

što više nije signalna statistika, jer je manja od granične vrednosti 16.2458. Izračunati su i svi uslovni izrazi sa dve promenljive $T^2_{i,j}$ i dati su u Tabeli 5.7. Granična vrednost za ove izraze je 4.04462, dobijena je iz formule (4.14). I zaista, kada pogledamo te vrednosti, signalne su samo one koje sadrže promenljivu x_2 . Dakle, signal na ovom vektoru ćemo rešiti ako smanjimo vrednost promenljive x_2 sa 8.4 na vrednost približnu 7.5. Ukoliko ovo nije moguće postići u procesu dorade, proizvod se odbacuje.

Pređimo sada na drugi signalni vektor $X' = (86.3, 7.1, 3.3, 0.3, 9.9, 18.33, 1.482)$. Njegova T^2 statistika je

$$T^2 = 1729.867 > 16.2458$$

Opet daleko iznad dozvoljene vrednosti. Za početak, izračunajmo sve bezuslovne izraze, dati su u Tabeli 5.8.

Promenljive x_1, x_5, x_7 imaju vrednost veću od granične 3.996471, pa ih izbacujemo iz vektora. Proveravamo preostali podvektor za signal

$$T^2_{(7.1, 3.3, 0.3, 18.33)} = 12.00651.$$

Kada se ova vrednost uporedi sa graničnom vrednosti 10.40844, koja je dobijena iz formule (4.11), gde je $n = 85, p = 4$ i $\alpha = 0.05$, zaključujemo da signal i dalje postoji

Tabela 5.8: Vrednost bezuslovnih izraza za drugi signalni vektor

T^2	Vrednost
T^2_1	12.4816*
T^2_2	1.9664
T^2_3	0.5916
T^2_4	0.9677
T^2_5	6.5735*
T^2_6	0.8423
T^2_7	7.6959*

Izračunamo sve uslovne izraze sa dve promenljive koji sadrže promenljive x_2, x_3, x_4, x_6 , i oni su dati u Tabeli 5.9. U Tabeli 5.7 mogu se naći vrednosti svih uslovnih izraza sa dve promenljive za ovaj signalni vektor i ona se nalazi na kraju poglavlja. Svi ovi izrazi izračunati su na osnovu šeme predstavljene u poglavlju 4.2.

Tabela 5.9: Vrednost uslovnih izraza sa dve promenljive za drugi signalni vektor

T^2	Vrednost	T^2	Vrednost
$T_{2,3}^2$	2.1577	$T_{4,2}^2$	0.9166
$T_{2,4}^2$	1.9154	$T_{4,3}^2$	2.3329
$T_{2,6}^2$	2.2964	$T_{4,6}^2$	3.4723
$T_{3,2}^2$	0.7829	$T_{6,2}^2$	1.1723
$T_{3,4}^2$	1.9569	$T_{6,3}^2$	2.3412
$T_{3,6}^2$	2.0906	$T_{6,4}^2$	3.3469

Dakle, nijedna vrednost nije signalna, što znači da su svi bivarijantni odnosi između ovih promenljivih u skladu sa istorijskim podacima. Nastavljamo dalje sa računanjem svih uslovnih izraza sa tri promenljive, koji sadrže promenljive x_2, x_3, x_4, x_6 . Vrednosti su dati u Tabeli 5.10.

Granična vrednost za uslovne izraze sa tri promenljive je 4.10431 i postoji čak šest signalnih izraza. Kako se u tih šest izraza nalaze sve četiri promenljive, tu je kraj sa računanjem izraza interpretacije jer su izbačene sve promenljive iz podvektora. Ako pogledamo malo bolje signalne izraze zaključujemo da se u svakom od njih nalazi promenljiva x_6 , pa je verovatno u njoj problem. Nađimo predviđenu vrednost za nju preko regresione prave iz istorijskih podataka na osnovu promenljivih x_2, x_3, x_4 .

$$x_6 = -0.17256x_2 - 2.50296x_3 + 3.851997x_4 + 26.42677$$

Tabela 5.10: Vrednost uslovnih izraza sa tri promenljive za drugi signalni vektor

T^2	Vrednost	T^2	Vrednost
$T_{2,3,4}^2$	0.3490	$T_{4,2,3}^2$	0.5243
$T_{2,3,6}^2$	8.2318*	$T_{4,2,6}^2$	3.8177
$T_{2,4,6}^2$	2.6418	$T_{4,3,6}^2$	8.8220*
$T_{3,2,4}^2$	0.3905	$T_{6,2,3}^2$	8.4154*
$T_{3,2,6}^2$	8.0260*	$T_{6,2,4}^2$	4.0733
$T_{3,4,6}^2$	7.4402*	$T_{6,3,4}^2$	8.8303*

Kada u ovu jednačinu ubacimo vrednosti (7.1,3.3,0.3) dobijamo da je $x_6 = 18.09743$. Proverimo ukupnu T^2 statistiku za promenljive x_2, x_3, x_4, x_6 , ali sada za vrednost promenljive x_6 uzimamo novu vrednost, tj. 18.09743.

$$T_{(7.1,3.3,0.3,18.09743)}^2 = 3.27361.$$

Dakle, signala više nema jer je $3.27361 < 10.40844$. Znači, trebalo bi smanjiti promenljivu x_6 na vrednost predviđenu regresijom iz istorijskih podataka. Dalje, preostaje da se reše signali na

bezaslovnim izrazima. Krenemo redom, predvidimo vrednost promenljive x_1 pomoću regresije iz istorijskih podataka na osnovu promenljivih x_2, x_3, x_4, x_6 .

$$x_1 = 0.561946x_2 + 0.356212x_3 + 0.448996x_4 + 0.561795x_6 + 71.28509$$

Kada se ubace odgovarajuće vrednosti dobija se $x_1 = 86.75215$. Odredimo sada pomoću regresije predviđenu vrednost za promenljivu x_5 .

$$x_5 = 0.5239x_1 - 0.70075x_2 + 0.4418x_3 - 0.61244x_4 - 0.1788x_6 - 28.184$$

Kada ubacimo u jednačinu vektor $(86.75215, 7.1, 3.3, 0.3, 18.09743)$ dobija se $x_5 = 10.32929$. Preostaje još da se nađe vrednost za promenljivu x_7 .

$$x_7 = 1.701x_1 - 1.463x_2 - 0.564x_3 - 0.357x_4 - 0.814x_5 - 0.202x_6 - 129.715$$

Ubacimo odgovarajuće vrednosti za nezavisne promenljive u regresiji i dobijamo $x_7 = 0.98284$. Proverimo sada ukupnu T^2 statistiku za nove predviđene vrednosti signalnih promenljivih.

$$T_{(86.75, 7.1, 3.3, 0.3, 10.33, 18.097, 0.98)}^2 = 3.273753$$

Signalna više nema, problem je rešen. Ako je moguće u procesu dorade promeniti vrednost promenljivih x_1, x_5, x_6, x_7 na vrednosti približne onim predviđenim regresijom, proizvod će zadovoljiti kriterijume koje je nametnula mušterija. Ako to nije moguće proizvod se odbacuje.

Preostaje još jedan vektor da se interpretira, a on je $X' = (87.11, 7.15, 3.69, 0.59, 10.2, 18.19, 1.18)$ i njegova ukupna T^2 vrednost je 148.1684 što je svakako veće od granične vrednosti 16.2458. Izračunajmo sve bezaslovne izraze za ovaj vektor, oni su dati u Tabeli 5.11.

Tabela 5.11: Vrednost bezaslovnih izraza za treći signalni vektor

T^2	Vrednost
T_1^2	0.1178
T_2^2	1.4203
T_3^2	1.0966
T_4^2	0.7574
T_5^2	0.1519
T_6^2	0.0590
T_7^2	1.2123

Ipak, iz tabele se vidi da nijedan od izraza nije signalni. Računamo dalje sve uslovne izraze sa dve promenljive, kako bismo proverili sve bivarijantne odnose između promenljivih. Rezultati svih ovih izraza su dati u Tabeli 5.7, a u Tabeli 5.12 su izdvojeni samo oni signalni.

Tabela 5.12: Vrednost signalnih uslovnih izraza sa dve promenljive za treći signalni vektor

T²	Vrednost
$T_{2.5}^2$	5.5822*
$T_{5.2}^2$	4.3138*
$T_{5.7}^2$	6.4623*
$T_{7.5}^2$	7.5226*

Signal se javlja na četiri izraza, a promenljive koje se pojavljuju u tim izrazima su x_2, x_5, x_7 . Kreiramo novi podvektor koji ne uključuje ove tri promenljive i proveravamo njega za signal.

$$T_{(87.11,3.69,0.59,18.19,)}^2 = 2.73558 < 10.40844$$

Ukupna T^2 vrednost ovih promenljivih je manja od granične vrednosti pa signal više nije prisutan. Zaključujemo da je problem u bivarijantnim odnosima između promenljivih x_2 i x_5 , kao i između x_5 i x_7 . Promenljiva x_2 ima vrednost 7.15 što je manje od srednje vrednosti 7.43 za tu promenljivu dobijene iz istorijskih podataka datih u Tabeli 5.2. Promenljiva x_5 ima vrednost 10.2 što je takođe manje od srednje vrednosti 10.25, koja se nalazi u Tabeli 5.2. To bi značilo da ove dve promenljive variraju u istom smeru, tj. da im je koeficijent korelacije pozitivan. Međutim, ako pogledamo u Tabelu 5.3 vidimo da je vrednost tog koeficijenta -0.775, što nas dovodi u kontradikciju. Dakle, potrebno je povećati vrednost jedne od te dve promenljive preko srednje vrednosti kako bi se rešio problem. Pogledajmo sada promenljivu x_7 , njena vrednost je 1.18, što je veće od njene srednje vrednosti 0.98. Ako to uporedimo sa promenljivom x_5 zaključujemo da ove promenljive variraju u suprotnom smeru, ali kad pogledamo njihov korelacioni koeficijent iz Tabele 5.3 on je pozitivan i iznosi 0.853. Što nas opet dovodi u kontradikciju, da bi se problem rešio potrebno je povećati vrednost promenljive x_5 ili smanjiti vrednost promenljive x_7 . U oba slučaja nam odgovara da povećamo vrednost promenljive x_5 . Pokušajmo pomoću regresije iz istorijskih podataka na osnovu ostalih promenljivih da odredimo vrednost promenljive x_5 .

$$x_5 = 0.564x_1 - 0.734x_2 + 0.427x_3 - 0.620x_4 - 0.183x_6 - 0.024x_7 - 31.128$$

Dobija se vrednost $x_5 = 10.45951$, koja zaista i jeste veća od srednje vrednosti 10.25 za ovu promenljivu. Proverimo ukupnu T^2 statistiku za ceo vektor, ali sa ovom novom vrednosti za promenljivu x_5 .

$$T_{(87.11,7.15,3.69,0.59,10.45951,18.19,1.18)}^2 = 3.12792$$

Dakle, ovaj vektor će prestati da daje signal ako se vrednost promenljive x_5 poveća na približno 10.45951. Ako je to moguće postići u procesu dorade proizvod će postati prihvatljiv za mušteriju, inače se odbacuje.

Tabela 5.1: Istorijski podaci i njihove T^2 vrednosti

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	T^2
1	86.85386	7.252841	3.234104	0.275133	10.2373	18.20367	1.011037	4.131734
2	87.3034	7.852568	3.450834	0.448627	10.08027	18.11264	0.681132	6.055363
3	87.26876	7.747679	3.407813	0.419637	10.11495	18.19237	0.820161	2.900153
4	87.34389	7.483754	3.767677	0.656545	10.36992	18.17876	1.059915	6.718363
5	87.21107	7.565106	3.530978	0.500737	10.1879	18.11587	0.89935	4.902659
6	87.05966	6.844322	3.738548	0.606343	10.59237	18.33762	1.444695	14.08562
7	86.90134	7.503978	3.335876	0.306447	10.17707	17.92639	0.696495	13.33282
8	87.1085	7.211279	3.563373	0.535175	10.40993	18.23721	1.133598	6.511725
9	87.61823	7.746978	3.829061	0.725671	10.24846	18.27816	1.033527	7.602752
10	87.33313	7.695186	3.468408	0.473373	10.14115	18.22513	0.847869	9.364576
11	87.24731	7.461564	3.564052	0.548061	10.27164	18.25941	1.015605	2.697276
12	87.11875	7.155976	3.692788	0.597764	10.43159	18.19724	1.181425	4.203148
13	87.00453	7.365864	3.449209	0.433684	10.24269	18.0486	0.876675	6.121005
14	87.19017	7.141919	3.610591	0.574089	10.49802	18.39832	1.377322	14.45019
15	86.98708	7.536679	3.22023	0.288886	10.08932	18.15842	0.768592	3.148853
16	87.13674	7.977379	3.131059	0.229461	9.856305	18.11039	0.572216	13.16681
17	87.25598	7.329712	3.735829	0.623902	10.38619	18.23397	1.046238	6.292961

18	87.33216	7.35117	3.724853	0.630282	10.42139	18.35563	1.13554	5.987723
19	87.32849	7.498382	3.522684	0.522601	10.25847	18.41746	1.06574	4.94412
20	87.34405	7.860651	3.427631	0.445339	10.10184	18.10906	0.798956	11.76514
21	87.16625	7.245225	3.701337	0.599125	10.39962	18.20705	1.070243	4.518573
22	87.445	7.391427	3.931439	0.800811	10.42324	18.27197	1.099744	10.42306
23	87.10457	7.137599	3.552988	0.495373	10.43838	18.26013	1.273757	9.605358
24	86.94495	7.467078	3.17385	0.252247	10.161	18.16476	0.817805	5.175118
25	86.93105	7.481623	3.251628	0.284337	10.11997	18.04874	0.8144	5.571653
26	87.19226	7.48624	3.511134	0.486734	10.25494	18.23606	0.894652	3.400484
27	86.57515	6.880387	3.216557	0.239539	10.41166	18.09966	1.076586	8.359371
28	86.6915	7.220964	3.176121	0.221467	10.21103	18.06008	0.841746	5.043631
29	87.04367	7.316901	3.465207	0.439181	10.28447	18.16211	0.969746	2.77066
30	87.02234	7.243134	3.545273	0.476091	10.38308	18.17899	0.98616	5.859604
31	86.9565	7.537943	3.147379	0.21492	10.09828	18.18941	0.832394	5.595322
32	86.8356	7.099005	3.368477	0.376181	10.35315	18.12299	1.081422	5.259941
33	87.28493	7.703079	3.496587	0.499366	10.14577	18.13251	0.945768	12.98023
34	87.00821	7.586103	3.28618	0.326086	10.08794	18.11757	0.73436	4.002464
35	87.33176	7.285597	3.857307	0.715867	10.43773	18.3632	1.18225	12.32898
36	86.95513	7.7	3.096162	0.184459	9.995004	18.09173	0.660049	5.841065
37	87.47621	7.860155	3.509667	0.527202	10.09519	18.36042	0.897951	6.619363
38	86.89239	7.02435	3.495363	0.458211	10.44911	18.21219	1.139174	5.556698
39	87.18826	7.647902	3.375557	0.414957	10.10967	18.27761	0.709387	14.32575
40	86.98238	7.255443	3.426661	0.410674	10.28587	18.18285	1.034127	2.430217
41	87.17669	7.221886	3.684605	0.604145	10.3964	18.26246	1.185047	3.608123
42	87.41445	7.618303	3.691303	0.627933	10.27421	18.24335	1.004239	2.935967
43	86.78375	7.407597	3.055469	0.132798	10.11118	18.12294	0.77324	7.955207
44	87.25326	7.369324	3.655893	0.586867	10.38592	18.26096	1.150747	4.490888
45	87.10949	7.348765	3.446557	0.441568	10.30488	18.28831	1.056446	1.610237
46	86.70517	7.042618	3.175756	0.238355	10.30783	18.18945	1.111385	7.978111
47	87.09171	7.402913	3.341285	0.391108	10.19721	18.28754	1.060577	5.412894
48	86.80637	7.187155	3.293368	0.322681	10.26479	18.16445	0.97389	6.276566
49	86.91218	7.396912	3.344305	0.336064	10.26339	17.99106	0.801775	8.794362
50	87.18785	7.321771	3.607163	0.542824	10.35984	18.34234	1.141224	9.098391
51	87.48112	7.680771	3.406657	0.463374	10.15982	18.56327	1.136198	15.08758
52	87.355	7.757428	3.40179	0.447015	10.11461	18.34195	0.937076	4.942343
53	87.5029	7.501989	3.655706	0.613442	10.33025	18.4809	1.270269	8.37568
54	86.86428	7.290312	3.299955	0.288621	10.23882	18.11255	0.920835	8.004127
55	86.97178	7.388624	3.309436	0.352218	10.22234	18.12439	0.955871	4.90234
56	86.99081	7.362128	3.450042	0.427259	10.28187	18.07675	0.902906	2.995322
57	87.17765	7.405349	3.483416	0.468006	10.25567	18.33924	1.099549	4.818345
58	87.63053	7.609742	3.762479	0.710109	10.30241	18.54506	1.255908	8.075988
59	87.25234	7.361682	3.623667	0.561807	10.37736	18.2797	1.217208	6.570328
60	86.88232	7.416107	3.189534	0.262065	10.13097	18.14782	0.832978	3.094823
61	86.90706	7.493654	3.131438	0.252974	10.06818	18.17008	0.784862	10.45796
62	87.14872	6.995106	3.873363	0.714262	10.56137	18.22647	1.327593	9.266838
63	87.87006	7.833046	3.934335	0.828114	10.3032	18.49104	1.195692	10.7589

64	87.2515	7.624781	3.526856	0.51601	10.2051	18.18228	0.805588	6.121642
65	86.75216	7.427558	3.058566	0.148682	10.15492	18.06563	0.643759	15.92231
66	87.29331	7.81387	3.206967	0.30872	9.987311	18.40827	0.898971	10.98755
67	87.2264	7.473567	3.446737	0.447633	10.26392	18.29265	1.060341	3.344379
68	87.09064	7.596381	3.440508	0.443971	10.15868	17.963	0.711692	12.12867
69	87.4118	7.331493	3.904197	0.760904	10.42319	18.30259	1.24341	8.125278
70	87.17022	7.487074	3.382638	0.435444	10.16875	18.29617	0.995921	7.282128
71	87.28507	7.692134	3.402856	0.420173	10.11092	18.25227	0.909442	4.255393
72	87.09621	7.407008	3.419424	0.390964	10.27047	18.27237	0.978835	8.353338
73	87.04796	7.347235	3.484134	0.453915	10.28972	18.10578	0.992335	2.50772
74	87.17171	7.22553	3.644983	0.572051	10.43816	18.28269	1.213621	3.686255
75	87.40294	7.470001	3.680152	0.603891	10.3634	18.36449	1.120372	6.338929
76	87.15707	7.572742	3.374906	0.409918	10.14151	18.24683	0.894678	2.05104
77	87.34042	7.780668	3.471092	0.490728	10.08329	18.23784	0.859415	5.138674
78	87.18848	7.58747	3.317624	0.351563	10.16531	18.2292	0.951083	8.290535
79	87.31816	7.621752	3.504196	0.505681	10.22725	18.27287	0.938522	4.291218
80	87.01752	7.514083	3.364466	0.377721	10.14039	18.10614	0.740192	4.305304
81	87.12089	7.54013	3.397015	0.412274	10.14798	18.14083	0.832456	4.924081
82	87.11821	7.10586	3.730815	0.617822	10.52241	18.23858	1.163822	6.96449
83	87.28742	7.5789	3.639018	0.593841	10.19887	18.2237	0.819501	11.76384
84	86.47239	6.940654	3.056544	0.118848	10.29693	18.03001	0.884766	13.02398
85	86.86139	7.102587	3.276558	0.319952	10.33863	18.23732	1.125095	6.800619

Tabela 5.4: Pristigli podaci za proveru

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	T^2
1	86.80921	7.386307	3.206062	0.2418	10.17429	18.03613	0.749624	3.649442
2	87.0392	7.26017	3.548042	0.48829	10.31064	18.18349	1.007812	5.230268
3	86.97221	7.331177	3.218355	0.310997	10.22815	18.29003	1.059675	9.934352
4	87.62998	8.01931	3.706086	0.65218	10.12408	18.20082	0.722444	13.59453
5	87.45164	7.533382	3.700613	0.656544	10.30052	18.38484	1.200145	6.996052
6	86.8351	7.188651	3.38466	0.365486	10.34611	18.01952	0.860525	9.198223
7	87.1373	7.447972	3.4942	0.480814	10.18754	18.17551	0.945811	8.961562
8	87.3	8.4	3.5	0.5	10.3	18.3	1.1	2483.934*
9	87.32285	7.714859	3.572108	0.526801	10.18057	18.19037	0.861195	7.19137
10	87.18576	7.316317	3.572922	0.524865	10.32788	18.30421	1.201695	5.584417
11	87.26254	7.444198	3.548214	0.506941	10.28032	18.32228	1.094375	3.12205
12	86.88046	7.325524	3.326506	0.344556	10.24124	18.04609	0.896113	4.034107
13	87.11907	7.451967	3.615317	0.543819	10.24596	18.05076	0.788808	9.756148
14	86.88771	7.304812	3.220085	0.257432	10.25898	18.24811	0.913245	10.19149
15	86.84521	7.657049	2.996935	0.122931	9.936985	18.0574	0.573956	8.392171
16	87.3392	7.368613	3.607341	0.606009	10.35266	18.46694	1.23071	12.10473
17	87.09669	7.494277	3.3558	0.378402	10.19618	18.20385	0.892482	1.04575

18	87.20095	7.602824	3.448711	0.42134	10.18704	18.21087	0.857202	6.060831
19	87.21314	7.517715	3.499937	0.472522	10.21857	18.25273	0.981342	2.257857
20	86.6698	7.432043	2.882738	0.02001	10.01415	18.08134	0.669523	12.66086
21	87.38912	7.569049	3.606845	0.574385	10.2435	18.41575	1.049725	5.76977
22	86.3	7.1	3.3	0.3	9.9	18.33	1.482	1729.867*
23	87.56626	7.860584	3.552513	0.562094	10.10763	18.39231	1.021012	6.881133
24	86.97266	7.140768	3.437649	0.41782	10.37739	18.25072	1.231937	7.103391
25	87.06936	6.936815	3.760743	0.629114	10.5789	18.2574	1.388936	8.280774
26	86.91967	7.312671	3.399797	0.366587	10.26982	17.99339	0.933424	9.493802
27	87.29582	7.229815	3.815361	0.697451	10.46415	18.30552	1.231362	3.427371
28	87.11	7.15	3.69	0.59	10.2	18.19	1.18	148.1684*
29	86.82186	7.130045	3.271802	0.306979	10.321	18.23991	1.057765	5.337981
30	86.74294	7.181575	3.23107	0.287746	10.21684	18.13122	0.936939	10.73402
31	86.96626	7.133899	3.421946	0.416131	10.35246	18.2783	1.192191	4.328229
32	87.3353	7.822062	3.340996	0.393696	10.05915	18.29254	0.859019	3.985182
33	87.45471	7.407528	3.694398	0.641599	10.38577	18.5162	1.297937	6.721374

Tabela 5.7: Vrednost uslovnih izraza sa dve promenljive za signalne vektore

	<i>Prvi signalni vektor</i>	<i>Drugi signalni vektor</i>	<i>Treći signalni vektor</i>
$T_{1,2}^2$	4.079803*	11.11505*	0.500073
$T_{1,3}^2$	0.691402	18.37661*	1.53523
$T_{1,4}^2$	0.573572	20.47937*	1.728237
$T_{1,5}^2$	0.465451	9.063288*	0.008746
$T_{1,6}^2$	0.104126	32.98841*	0.006525
$T_{1,7}^2$	0.273703	20.79834*	0.207492
$T_{2,1}^2$	20.17495*	0.599867	1.908576
$T_{2,3}^2$	16.79285*	2.157657	1.234681
$T_{2,4}^2$	16.53847*	1.915376	1.460247
$T_{2,5}^2$	46.93095*	17.88125*	5.582216*
$T_{2,6}^2$	16.17502*	2.29638	1.373255
$T_{2,7}^2$	28.73828*	0.033795	0.481642
$T_{3,1}^2$	0.235137	6.48665*	2.620015
$T_{3,2}^2$	0.241443	0.782879	0.910963
$T_{3,4}^2$	0.788442	1.956866	1.629517
$T_{3,5}^2$	0.006381	-2.47665	3.040476

$T_{3,6}^2$	0.044736	2.09056	1.865335
$T_{3,7}^2$	0.140373	12.51186*	0.174072
$T_{4,1}^2$	0.17795	8.965457*	2.473886
$T_{4,2}^2$	0.047703	0.916649	0.797393
$T_{4,3}^2$	0.849085	2.332917	1.290381
$T_{4,5}^2$	0.014881	-2.89133	1.835146
$T_{4,6}^2$	0.013306	3.47229	1.543605
$T_{4,7}^2$	0.028763	13.18269*	0.043792
$T_{5,1}^2$	0.09504	3.155187	0.148876
$T_{5,2}^2$	30.46539*	22.48833*	4.313843*
$T_{5,3}^2$	0.092235	3.505207	2.095821
$T_{5,4}^2$	0.040092	2.714476	1.229627
$T_{5,6}^2$	0.024496	5.07372*	0.112735
$T_{5,7}^2$	0.186853	66.97633*	6.462262*
$T_{6,1}^2$	0.06918	21.34912*	0.053716
$T_{6,2}^2$	0.044923	1.172269	0.011943
$T_{6,3}^2$	0.466055	2.341227	0.827741
$T_{6,4}^2$	0.373982	3.346906	0.845147
$T_{6,5}^2$	0.359961	-0.65747	0.019796
$T_{6,7}^2$	0.103833	1.259935	1.529512
$T_{7,1}^2$	0.22384	16.01261*	1.407986
$T_{7,2}^2$	12.59327*	5.763239*	0.273633
$T_{7,3}^2$	0.546775	19.61609*	0.289781
$T_{7,4}^2$	0.374522	19.91087*	0.498637
$T_{7,5}^2$	0.507401	68.0987*	7.522626*
$T_{7,6}^2$	0.088916	8.11357*	2.682815

Zaključak

Industrijska kontrola procesa generalno uključuje praćenje skupa koreliranih promenljivih. Takva korelacija otežava tumačenje univariatnih procedura za individualne promenljive. Jedan metod za prevazilaženje ovih problema je korišćenje Hotelingove T^2 statistike. Kao što je prikazano, statistika je bazirana na konceptu statističke distance i ona skuplja informacije prikupljene multivariatnim posmatranjem u jednu vrednost.

Poželjne karakteristike multivarijantne kontrole procesa obuhvataju laku primenu, adekvatnu interpretaciju signala, fleksibilnost, osetljivost na male procesne promene i algoritam koji će koristiti ove podatke. Metoda bazirana na T^2 statistici, koja je opisana u radu poseduje sve ove osobine.

Prikazana je primena MYT dekompozicije u interpretaciji signalnih vektora. Svaka signalna T^2 vrednost ima više moguće MYT dekompozicije u zavisnosti od broja promenljivih i ako je jedna od njih signalna, onda su sve signalne. Svaka od dekompozicija sastoji se od bezuslovnog dela T_i^2 , koji

se koristi za pojedinačnu proveru promenljivih, a uslovni izrazi se koriste za proveru odnosa između promenljivih.

Interpretacija signalnih vektora u smislu promenljivih multivarijantnog procesa je izazovan problem. Da li uključuje puno ili malo promenljivih, problem ostaje isti : kako pronaći signalne promenljive? Rešenje ovog problema leži u primeni Hotelingove T^2 statistike sa MYT dekompozicijom. Na primer, signalne bezuslovne komponente momentalno lociraju uzrok u smislu individualne promenljive ili grupe promenljivih, dok signalni uslovni izrazi lociraju kontra korelisane odnose između promenljivih kao uzrok.

Da bi se uspešno primenjivala ova metoda multivarijantne kontrole procesa potrebno je dobro se upoznati sa procesom i prikupiti ispravne istorijske podatke. Potrebno je poznavati sve promenljive i znati na koje može da se utiče i kako, kao i znati koje mogu da se koriguju u procesu dorade.

Zbog svoje fleksibilnosti i svestranosti pravilna primena Hotelingove T^2 statistike sa MYT dekompozicijom može biti izuzetno moćan alat u multivarijantnoj kontroli procesa.

Literatura

- [1] Zhiqiang Ge, Zhihuan Song (2013), „*Multivariate Statistical Process Controll*“, Springer
- [2] Uwe Kruger, Lei Xie (2012), „*Statistical Monitoring of Complex Multivariate Processes*“, John Wiley & Sons
- [3] Douglas C. Montgomery (2012), „*Statistical Quality Control*“ 7th Edition, John Wiley & Sons
- [4] Sang Wook Choi, Elaine B. Martin, A. Julian Morris, In-Beum Lee (2006), „*Adaptive Multivariate Statistical Process Control for Monitoring Time – Varying Processes*“, American Chemical Society
- [5] Z. Lozanov-Crvenković (2011), „*Statistika*“, Novi Sad

[6] Robert L. Mason, John C. Young (2002), „Multivariate Statistical Process Control with Industrial Application“, American Statistical Association and the Society for Industrial and Applied Mathematics



[7] Paruchuri R. Krishnaiah (1980), „Multivariate analysis – V“, North Holland Pub.

[8] Brian Everitt (2005), „An R and S-Plus Companion to Multivariate Analysis“, Springer

[9] Ingram Olkin (1960), „Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling“, Stanford University

[10] K. V. Mardia, John T. Kent, John M. Bibby (1979), „Multivariate Analysis“, Academic Press

[11] Paruchuri R. Krishnaiah, Calyampudi Radhakrishna Rao (1988), „Quality Control and Reliability“, North Holland Pub.

[12] Ravindra Khattree, Calyampudi Radhakrishna Rao (2003), „Statistics in Industry“, Gulf Professional Publishing

[13] James P. Stevens (2009), „Applied Multivariate Statistics for the Social Sciences fifth edition“, Taylor & Francis Group

[14] Bernard Flury (1997), „A First Course in Multivariate statistics“, Springer

[15] Bryan F. J. Manly (2005), „Multivariate Statistical Methods third edition“, Chapman & Hall/CRC

[16] Martin Bilodeau, David Brenner (2008), „Theory of Multivariate Statistics“, Springer

Biografija

Daniel Pavlica je rođen 24.12.1990. u Novom Sadu. Završio je osnovnu školu „Braća Grulović“ u Beški kao đak generacije. Upisao je gimnaziju „J.J. Zmaj“ u Novom Sadu, smer obdareni učenici u matematičkoj gimnaziji. Nakon završetka srednje škole 2009. godine upisao je osnovne akademske studije Primjenjene matematike na Prirodno-matematičkom fakultetu u Novom Sadu, modul Matematika finansija. 2013. godine završava osnovne akademske studije i upisuje master studije na istom usmerenju. Položio je sve ispite predviđene nastavnim planom i programom u junskom roku 2015. godine.

UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:

RBR

Identifikacioni broj:

IBR

Tip dokumentacije: *Monografska dokumentacija*

TD

Tip zapisa: *Tekstualni štampani materijal*

TZ

Vrsta rada: *Master rad*

MR

Autor: *Daniel Pavlica*

AU

Mentor: *dr Zagorka Lozanov-Crvenković*

MN

Naslov rada: *Multivariantna statistička kontrola procesa pomoću Hotelingove T^2 statistike*

NR

Jezik publikacije: *srpski (latinica)*

JP

Jezik izvoda: *srpski/engleski*

JL

Zemlja publikovanja: *Republika Srbija*

ZP

Uže geografsko područje: *Vojvodina*

UGP

Godina: *2016*

GO

Izdavač: *Autorski reprint*

IZ

Mesto i adresa: *Novi Sad, departman za matematiku i informatiku, Prirodno-matematički fakultet, Univerzitet u Novom Sadu, Trg Dositeja Obradovića 4*

MA

Fizički opis rada: *(5/80/0/17/30/0)*

(broj poglavlja/broj strana/broj citata/broj tabela/broj grafika/broj priloga)

FO

Naučna oblast: *Matematika*

NA

Naučna disciplina: *Statistika*

ND

Predmetna odrednica/ Ključne reči: *Multivariantna kontrola, Euklidova distanca, statistička distanca, Hotelingova statistika, MYT(Mejsen-Jang-Trejsi) dekompozicija*

PO

UDK

Čuva se: *Biblioteka departmana za matematiku i informatiku, Prirodno-matematički fakultet, Univerzitet u Novom Sadu*

ČU

Važna napomena:

VN

Izvod: *Ovaj rad se bavi multivarijantnom statističkom kontrolom procesa, zasnovanom na osnovnim pojmovima Hotelingove T^2 statistike. Na početku rada govori se o statističkoj kontroli procesa, gde će biti spomenuta univarijantna i biti obrađena multivarijantna. Takođe, biće navedene njihove razlike i dat uvod u T^3 statistiku. Potom, će biti izloženi osnovni koncepti T^2 statistike, njena distributivna svojstva, kao i statistička distanca. U nastavku, radiće se interpretacija T^2 signala za dve promenljive, koja će biti bazirana na MYT (Mejson - Jang - Trejsi) dekompoziciji. Pokazano je kako signal može biti izolovati na određene promenljive povezane sa procesom. Zatim se interpretacija proširuje na opšti slučaj, tj. više od dve promenljive.*

Na kraju, biće uradjen primer u kojem se primenjuje multivarijantna statistička kontrola procesa koristeći Hotelingovu T^2 statistiku. U njemu će se proveriti da li vrednost T^2 statistike za prikupljene podatke upada u granice procesnih varijacija, koje su uspostavljene istorijskim podacima. Ako vrednost T^2 statistike ne upada u kontrolne granice, signal će se detektovati. To se može desiti ako se vrednost neke promenljive naglo promenila ili ako se vrednost dve ili više promenljivih ne pridržava linearne korelacije koja je nametnuta istorijskim podacima. Biće analizirani signalni slučajevi korišćenjem MYT dekompozicije i biće obrazložen razlog za njihov signal, biće izračunate nove vrednosti za signalne promenljive i sa tim novim vrednostima signal će nestati.

IZ

Datum prihvatanja teme od strane NN veća: 13. Januar 2016.

DP

Datum odbrane:

DO

Članovi komisije:

KO

Predsednik: *dr Ljiljana Gajić, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu*

Mentor: *dr Zagorka Lozanov-Crvenković, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu*

Član: *dr Ivana Štajner-Papuga, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu*

UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCE
KEY WORDS DOCUMENTATION

Accession number:

ANO

Identification number:

INO

Document type: *Monograph type*

DT

Type of record: *Printed text*

TR

Contents code:

CC

Author: *Daniel Pavlica*

AU

Mentor: *Zagorka Lozanov-Crvenković*

MN

Title: *Multivariate statistical process control using Hotelling's T^2 statistic*

TI

Language of text: *Serbian(Latin)*

LT

Language of abstract: *sebian/english*

LA

Country of publication: *Serbia*

CP

Locality of publication: *Vojvodina*

LP

Publication year: 2016

PY

Publisher: *Author's reprint*

PU

Publication place: *Novi Sad, Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad, Tgr Dositeja Obradovića 4*

PP

Physical description: (5/80/0/17/30/0)
(number of sections/ pages/ references/ tables/ graphs/ appendices)

PD

Scientific field: *Mathematics*

SF

Scientific discipline: *Statistic*

SD

Subject/ Key words: *Multivariate control, Euclidean distance, statistical distance, Hotelling's statistic, MYT(Mason-Young-Tracy) decomposition*

SKW

UC

Holding data: *The library of the Department of Mathematics and Informatics, Faculty Science, University of Novi Sad*

HD

Note:

N

Abstract: This paper deals with multivariate statistical process, based on conceptual view of Hotelling's statistic. In the beginning of the paper we will speak about statistical process control, and few words about univariate will be said, and multivariate as well. Furthermore, we will speak about differences between these two, and introduction in T^2 statistic will be given also. Later on, one can read about basic concepts of statistics, its distributive properties, and statistic distance too. In following part of this paper interpretation of signal for two variables will be processed, which is based on MYT decomposition. It is shown how signal can be isolated on certain variables linked to process.

In the end, one problem will be solved, using multivariate statistical process control, with Hotelling's statistic. In this example it will be tested whether the value of statistic for observed data is in the scope of process limits or not. If the answer is negative, signal will be detected. This is possible if value of some variable is suddenly changed, or if the values of two or more variables are not in linear correlation which is imposed with historical data. Signal cases will be analyzed using MYT decomposition and detailed explanation will be given. Also, new values for signal variables will be calculated and with these new values, signal will disappear.

AB

Accepted by the Scientific Board on: 13. January 2016.

ASB

Defended:

DE

Thesis defend board:

DB

President: *Ljiljana Gajić, PhD, full professor, Faculty of Science, University of Novi Sad*

Mentor: *Zagorka Lozanov-Crvenković, PhD, full professor, Faculty of Science, University of Novi Sad*

Member: *Ivana Štajner-Papuga, PhD, full professor, Faculty of Science, University of Novi Sad*