

StaM-Bladet

Informationsblad för medlemmar i StaM (Statistisk Metodik), sektion inom SFK, Svenska Förbundet för Kvalitet

September 1997

årgång 7 nummer 14

Fjortonde numret

I detta nummer av StaM-Bladet, en skrift som utkommer med stokastiska tidsmellanrum, har vi ånyo fångats och fascinerats av procentsatser och felkvoter och igen situationen med noll fel. För att liva upp det hela har vi låtit Andersson gå en match mot Nilsson. Och det vet ju var och en att retoriken innehåller många knep och rackartyg, men med ett statistiskt slagträ kan man slå världen med häpnad.

För övrigt hittar vi en användbar olikhet samt lite information om nätet. Höstens begivenhet är annars seminariet om multivariata tricks, serverade hos IVF i Mölndal.

Ordförandens ruta

För många av oss som inte arbetar med statistik varje dag, är multivariata tekniker någonting vi möjligen har hört talas om men absolut inte har grepp om. Inom ett flertal branscher börjar det komma till mer och mer användning. Vi tycker därför att tiden är mogen för en första introduktion av dessa metoder till våra medlemmar och andra intresserade. Årets seminarium kommer att avhandla ämnet 'Multivariata tekniker i praktiken'. Föredragshållarna kommer att blanda grundläggande teori med praktiska exempel tagna från olika verksamhetsområden. Tag chansen att få en första idé om vad multivariata tekniker kan användas till i din bransch!

I anslutning till seminariet kommer vi att ha en utställning med blandat material inom statistisk metodik.

Numera finns naturligtvis även SFK på nätet. SFKs hemsida finns under adressen <http://www.sfkvalitet.se> och därifrån kan man ta sig vidare till SFK-StaMs hemsida.

Vi har gjort ett antal sidor ur tidigare StaM-Bladet som vi sedermera skall lägga på vår hemsida. Håll alltså utkik då och då!

Susanna Weinberger, Ovako Steel AB

Förteckning över styrelsen finns på sista sidan

Något mer om olikheter och normalitet

I förra numret av StaM-bladet beskrevs Tjebychevs olikhet. Här utvidgar vi resonemanget och kommer med ytterligare en annan olikhet. Dessutom gör vi en liten jämförelse i tabellform.

Tjebychevs olikhet fungerar för alla former av statistiska fördelningar, oavsett utseende. Detta innebär att den är en mycket konservativ olikhet. I tabellen nedan har vi beräknat Tjebychevs olikhet för $\pm t\sigma$ ($t = 1.0, 1.5, 2.0 \dots 4.0$). Om man vill att Tjebychevs olikhet skall omfatta åtminstone 95% av ytan i fördelningen måste vi ta med åtminstone $\pm 4.47\sigma$:

$$P(|X - \mu| \leq t\sigma) \geq 1 - \frac{1}{t^2} \Rightarrow 1 - \frac{1}{t^2} = 0.95 \Rightarrow t \approx 4.47$$

Från normalfördelningsteorin vet vi att det behövs ungefär $\pm 2\sigma$ (från medelvärdet) för att få 95% av ytan i normalfördelningen. Här ser vi alltså att om man tror på en viss modell kan man ofta göra mer distinkta uttalanden. (Det framgår också från formeln på Tjebychevs olikhet att den kan bara användas för $t > 1$. Om $t = 1$ blir det division med noll (ett verkligt 'ajabaja' i matematiken) och om $t < 1$ blir resultatet negativt och någon negativ sannolikhet finns inte!)

I praktiken har vi ofta en viss modell (fördel-

ning) i tankarna och förhoppningsvis har vi extra kunskap eller data som stöder detta. Vi borde alltså kunna utnyttja denna kunskap och minska osäkerheten i våra uttalanden. Om vi vet att fördelningen är kontinuerlig och unimodial (entoppig) samt att

$$\left| \frac{\mu - \text{mode}}{\sigma} \right| < 1.0$$

vilket betyder att skillnaden mellan fördelningens medelvärde (σ) och dess typvärde (mode) får vara högst 1 standardavvikelse om vi skall kunna använda följande olikhet. (Observera att för alla symmetriska gäller att $\sigma = \text{mode}$):

$$P(|X - \mu| \leq t\sigma) \geq 1 - \frac{4}{9 \cdot t^2}$$

Denna olikhet kallas Camp – Meidells och är alltså något 'bättre' än Tjebychevs olikhet. (Man behöver inte vara orolig om man inte tidigare känner till denna olikhet ty inom matematiken och statistiken finns det nästan en oändlig mängd olikheter och tumregler.) Det är naturligtvis inte rättvist att säga att en viss olikhet är bättre än en annan om man, som i vårt fall, också måste snäva in användningsområdet genom att ställa högre krav.

Mats Franzén, Ericsson Cables AB

En jämförelse

I nedanstående tabell visas de olika ytor som faller inom $\pm t\sigma$ för de två olikheterna samt normalfördelningen.

Avvikelse	t=1.0	t=1.5	t=2.0	t=2.5	t=3.0	t=3.5	t=4.0
Tjebychev	-	0.5556	0.7500	0.8400	0.8889	0.9185	0.9375
Camp-Meidell	0.5556	0.8025	0.8889	0.9289	0.9506	0.9637	0.9722
Normalitet	0.6827	0.8664	0.9545	0.9876	0.9973	0.9994	0.9999

Källa: Statistical methods in research and production, O.L Davis, 2nd edition, 1949, Oliver and Boyd, London

Detta har inte hänt

Inköpare Andersson sitter på kontoret och sysslar med några papper. Han har för en tid sedan blivit utsedd av chefen att delta i ett kvalitetsprojekt som går ut på att få bättre kvalitet på några viktiga komponenter som köps in i stora kvantiteter. Andersson var ju inte så där överväldigad entusiastisk över uppdraget. "Jag har ju viktigare saker att syssla med. Och förresten har vi ju en kvalitetsavdelning, om de nån gång fick ändan ur vagnen så skulle vi ha färre problem". Andersson vet dock att han inte bör gnälla, alla måste ställa upp. Och som inköpare har han naturligtvis en viss makt över leverantörerna, och det händer att han får en eller två flaskor whisky runt jul.

Andersson slår på datorn för att titta på de senaste resultaten från SKRAP AB. De brukar leverera ACAT-produkter och felkvoten låg stadigt runt en procent men nu har den minskat gradvis till en mer acceptabel nivå. Andersson får något drömmande i blicken när han minns hur det startade:

Han hade fått både skit och gliringar från produktionen angående problemen med ACAT-produkten. "Om du åtminstone kunde skaffa grejer som dög, vi slösar alltför mycket tid och pengar på omarbete", hade de sagt. "Jojo", tänkte Andersson, "jag har nog sett hur de behandlar grejerna därute. Lite mer aktsamhet skulle inte skada".

Men han hade lovat att ta upp saken med SKRAP AB, trots att han inte tyckte om att ringa och klaga hos den där fähunden Nilsson, en dryg typ. Andersson hade samlat ihop all statistik och alla argument han hade och där fanns det i svart på vitt, ganska tydligt och klart, en procent fel. Andersson morskade upp sig och ringde upp Nilsson. "Ja det var Nilsson", vrålade Nilsson i luren, medan Andersson samlade sig och lyssnade till idiotens flåshurtiga jargong. Som vanligt var det inget speciellt fel på SKRAP-produkterna, och förresten, man är väl inte mer än människa, lite skit får man väl tåla! Andersson kunde känna Nilssons nikotinandedräkt och hans svettiga arm runt sina axlar. "Men visst, för din skull skall vi se över det här". Andersson tvivlade.

En tid senare, en fredagseftermiddag, gick Andersson i tankarna genom helgens fiskeutflykt. Plötsligt störde telefonen hans tankar. Det var Nilsson. "Tjenare, tjenare", tuggade Nilsson, "hur är läget?" Andersson hade ingen lust att berätta hur läget var men spårade en viss överlägsenhet i Nilssons röst.

Nilsson, som tror på MBF (management by fear), hade efter deras tidigare telefonsamtal gormat och skällt på allt och alla inom SKRAP AB men ganska snart hade han glömt allt. Av en ren händelse hade han fått reda på att deras senaste leverans om 150 produkter inte hade innehållit några fel. Så fort han kunde hade han ringt upp Andersson. "Alltså noll felaktiga bland 150 produkter", försökte Nilsson gnugga in, "måste betyda en stor förbättring, va?". "När kommer er nästa order, Andersson?".

Andersson, som hade fått sina tankar störda och ogillade Nilsson än mer, kunde inte komma på något att säga. Noll fel bland 150 detaljer måste väl ändå vara ganska imponerande. För några veckor sedan hade han gått en kurs i statistik och SPS, och försökte nu komma ihåg vad magistern hade sagt:

'Noll fel betyder ju inte att felkvoten nödvändigtvis måste vara noll. Vi kan beräkna ett s.k. konfidensintervall som med en viss sannolikhet täcker den sanna felkvoten. Uttrycket $3/n$ kan användas som en tumregel för att få en övre gräns på ett 95-procentigt konfidensintervall'.

Andersson gjorde snabbt ett överslag. $3/n$ blir alltså här $3/150$ som i decimal form ger 0.02 dvs 2 procent som övre gräns för ett 95-procentigt intervall för felkvoten i Nilssons process. Nu hade han vad han behövde för att tvåla till Nilsson. "Tja", svarade Andersson, "noll felaktiga bland 150 kontrollerade är inte något övertygande resultat om att er felkvot har förbättrats. Vi kräver mer än så". Han hörde hur Nilssons ilska steg. "Va' f-n menar du, duger inte noll fel till er längre, vill ni ha bättre än så?", försökte Nilsson hånande. "Nej", svarade Andersson som kände att han hade ett övertag, "resultatet är bra men det är ingen övertygande information att er verkliga felkvot har förändrats".

Nilsson hade efter ytterligare förtrytelser och förtäckta hot slängt på luren.

Andersson hade senare fått reda på att Nilsson fått ett nytt jobb och att man på SKRAP AB hade startat ett förbättringsarbete i samarbete med sina kunder.

Allt detta minns Andersson. "Han var inte dum den där magistern", tänkte han och lovade sig att vara uppmärksam vid nästa kurstillfälle. Naturligtvis var han glad att hans medverkan blivit en början på ett förbättringsarbete men mest nöjd var han ändå att han fått ge igen på Nilsson.

Se också utredningen på nästa sida!

På föregående sida...

...använder Andersson en relativt okänd tumregel som på engelska kallas *the rule of three*. Den uppkommer ur binomialfördelningen, den statistiska fördelning som är vanligast då man diskuterar felkvoter. Följande rader är en kortfattad härledning. Sannolikheten att få exakt 0 felaktiga i ett stickprov om n detaljer är

$$P(X=0) = (1-p)^n$$

Vi kan få ett övre värde på ett konfidensintervall för p om vi löser ut p i följande uttryck:

$$(1-p)^n \geq \alpha$$

Det övre värdet på p , som vi kallar p_u , blir då:

$$p_u = 1 - \alpha^{1/n}$$

Nu gör vi en så kallade Taylorutveckling av den andra termen:

$$\alpha^{1/n} = 1 + \frac{\ln \alpha}{n} + \frac{(\ln \alpha)^2}{2n^2} + \dots$$

Om vi nu sätter in detta får vi följande resultat där vi också passat på att sätta $\alpha = 0.05$:

$$p_u \approx -\frac{\ln \alpha}{n} \Rightarrow p_u \approx -\frac{\ln 0.05}{n} \Rightarrow p_u \approx \frac{3}{n}$$

Tabellen ger för olika n en jämförelse mellan ovanstående tumregel och mer exakta formler som redovisats i tidigare StaM-Bladet:

n	tumregel	exakt
10	0.300	0.259
50	0.060	0.058
100	0.030	0.030

(Källa: *The American Statistician*, May 1997, Vol. 51, No. 2)

Statistik på Internet

Programvara för multivariat analys är aktuellt med tanke på det förstående seminariet om multivariata metoder i praktiken. En intressant programbas med nedladdningsbara program heter "Clemson continuous quality improvement (CQI) server".

Adress: <http://deming.ces.clemson.edu/pub/tqmbbs/software>

Under denna adress kan man finna ett program som heter MapStat (Multivariate Statistical analysis package). Detta program innehåller bl a klusteranalys, faktoranalys, multipel regression samt MANOVA. Programmet är DOS-baserat. En finess är att källkoden (Turbo Pascal) följer med och den som är händig kan på egen risk modifiera funktionerna i programmet. Det finns även mycket annat smått och gott i denna programbas.

Prova er fram för att finna multivariata programvaror, algoritmer, artiklar med de olika sökverktyg som Internet erbjuder eller sök i t ex StatLib, adress: <http://www.stat.cmu.edu/>

Mats Franzén, Ericsson Cables AB

Något om multivariata analysmetoder

Mycket av den statistiska analysen går ut på att undersöka samband mellan flera variabler. Ett mycket enkelt exempel är att människans vikt beror på individens längd, ålder, kön m.m. Vi säger att vikten är den beroende variabeln medan längd, ålder och kön är förklarande variabler. Ett sådant samband studeras ofta med linjär regression vilket inte brukar betraktas som en multivariat teknik trots att flera variabler är inblandade. Anledningen är att i regressionsanalysen betraktas de förklarande variablerna som 'exakta', dvs som observerade utan fel och därigenom oberoende av variabeln vikt som innehåller en slumpmässighet.

Alla multivariata tekniker bygger på att de ingående variablerna är statistiskt oberoende av varandra (vilket är mer rimligt). Sådana situationer dyker upp t.ex. när man med moderna analysinstrument kan bestämma en mängd olika mått på ett fåtal undersökta prov. Alla dessa variabler blir givetvis beroende av varandra. Vilka möjligheter har vi då att dra slutsatser ur en sådan undersökning?

Ett knep är att bilda speciella linjärkombinationer av alla variablerna så att de nya variablerna blir oberoende. Sådana skapade variabler kallas principalkomponenter och tekniken förkortas med PCA (principal components analysis). Dessa komponenter kan sedan t.ex. användas i en regressionsanalys som förklarande variabler. En liknande multivariat teknik är partial least square (PLS). Denna bygger på att man uppmätt många variabler X_1, X_2, \dots, X_p och Y på

ett fåtal prov och att man vill prediktera värdet på Y på prov på vilka man endast uppmätt X_1, X_2, \dots, X_p . Y kan t.ex. vara för svår eller dyr att mäta. Denna teknik har utvecklats inom kemometrin men har givetvis användning inom många andra områden.

Andra tekniker är diskriminantanalys där man bildar en diskriminerande funktion av de ingående variablerna och med detta resultat kan man klassa ett prov till olika grupper.

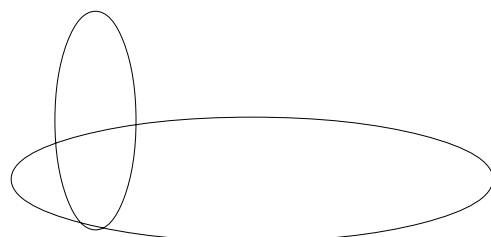
Klusteranalys separerar data i ett antal kluster så att data som liknar varandra kommer i samma kluster.

Multidimensional scaling bildar en karta över data med utgångspunkt från ett avståndsmått mellan data.

Hotelling's T^2 -test är en generalisering av envariabeltestet t-test. Man vill t.ex. testa om två olika behandlingar har samma genomsnittseffekt i alla responsvariablerna som mäts.

Ytterligare multivariata tekniker är kanonisk korrelation och faktoranalys men de beskrivs inte här.

Lennart Nilsson, Umeå Universitet



SFK–StaM Seminarium
Tisdagen den 28 oktober 1997
IVF, Mölndal

Multivariata tekniker i praktisk verksamhet

- 09:30–10:00 Registrering, kaffe
- 10:00–10:15 *Inledning*
Ordf. Susanna Weinberger, Ovako Steel
- 10:15–11:00 *Introduktion till multivariat analys*
Ulf Olsson, Sveriges Lantbruksuniversitet, avd för statistik
Verkligheten är mångdimensionell (multivariat) och förkortningar som PCA, PCR och PLS förekommer numera ofta i analysksammanhang. Begreppet multivariat analys är inte begränsat till dessa och med hjälp av exempel ges en översikt över gamla och nya multivariata metoder.
- 11:00–11:15 Paus med frukt
- 11:15–12:00 *Multivariata metoder i industriella tillämpningar*
Håkan Fridén, Umetri AB, Umeå
Kort metodikpresentation samt exempel på användning för övervakning, optimering och prediktion.
- 12:00–13:15 Lunch. Årsmöte StaM
- 13:15–14:00 *Snabbare och billigare kontroll av råvarukvalitet inom bioteknisk industri*
Richard Paldanius, Pharmacia & Upjohn AB
Föredraget beskriver ersättning av immunoassay och fysikaliska mätningar med instrumentell analys och multivariat dataanalys vid kvalitetsbedömning av en kritisk råvara inom allergidiagnostikindustrin.
- 14:00–14:45 *Optimering, modellbygge och kvalitetssäkring – exemplet kväveoxider*
Jonas Persson, Bergström & Öberg AB
Presentationen avser att visa på sambanden mellan praktisk optimering av förbränningsanläggningar, utveckling av prediktionsmodeller för kväveoxider samt kvalitetssäkring av dessa. Optimering görs med simplexmetoden, kalibreringsmodellerna baseras på PLS-regression och kvalitetssäkringen görs med principalkomponentmodeller.
- 14:45–15:15 Kaffe. Diskussion
- 15:15–16:00 *Kan man använda försöksplanering för marknadsundersökningar?*
Prof. Bo Bergman, Kvalitetsteknik, Tekniska högskolan i Linköping
Föredraget behandlar användningen av multivariat analys vid marknadsundersökningar av typ 'Conjoint Analysis'.
- 16:15–16:30 *Diskussion och avslutning*
Ordf. Susanna Weinberger, Ovako Steel AB
Lennart Nilsson, Umeå universitet
- Slut

Anmälan

SFK – StaM Seminarium

Multivariata tekniker i praktisk verksamhet

Tid: Tisdagen den 28 oktober 1997, 09:30 – 16:30

Plats: IVF, Mölndal

Avgift: 2 500 kr (studeranderabatt kan erhållas), inkluderar lunch och kaffe.
Avgiften betalas via faktura som bifogas bekräftelsen. (Vid avbokning senare än 20 oktober debiteras fullt pris.)

Namn: _____

Företag/Organisation
Högskola/Universitet: _____

Adress: _____

Faktuaraadress: _____

Telefon/fax/e-post: _____

Anmälan bör vara oss tillhanda senast den 8 oktober 1997 och skickas till

Susanna Weinberger
Ovako Steel
Fax: 0290 – 259 87
Tel: 0290 – 252 96
e-post: seskfws5@ibmmail.se

Styrelsen

Ordförande:

Susanna Weinberger
Ovako Steel AB
813 82 Hofors
0290 – 252 96

Sekreterare:

Clas Mellby
IVF
Argongatan 30
431 53 Mölndal
031 – 706 61 02

Kassör:

Anders Hynén
ABB Corporate Research
avd R
721 78 Västerås
021 – 32 32 24

Ledamöter:

Ann Brännström-Stenberg
Högskolan i Karlstad
Institutionen för teknik
651 88 Karlstad
054 – 83 82 36

Lennart Nilsson
Matematisk statistik
Universitetet
901 87 Umeå
090 – 16 60 77

Göran Lande
Ericsson Radio Systems AB

Sture Rydin
Stora Cell AB
814 81 Skutskär
026 – 851 07

Mats Franzén
Ericsson Cables AB
824 82 Hudiksvall
0650 – 361 61

Redaktionskommitté:

Lennart Nilsson
Ingemar Sjöström
Susanna Weinberger

Bidrag accepteras gärna via 3.5"-diskett med textmängden i format WordPerfect, Word e.d.

Man blir medlem i SFK–StaM genom att kontakta Svenska Förbundet för Kvalitet telefon 08 – 783 82 54 eller 783 01 71 (fax 661 1967). Kanslisekreterare är Berit Winnsjö.

Några hemsidor

<http://www.sfkvalitet.se>

<http://www.sfkvalitet.se/sektion/sfk-stam.htm>